



HAL
open science

Classification par méthodes d'apprentissage supervisé et faiblement supervisé d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie

Romain Cendre

► To cite this version:

Romain Cendre. Classification par méthodes d'apprentissage supervisé et faiblement supervisé d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie. Traitement du signal et de l'image [eess.SP]. Université Bourgogne Franche-Comté, 2021. Français. NNT : 2021UBFCK004 . tel-04085244v2

HAL Id: tel-04085244

<https://theses.hal.science/tel-04085244v2>

Submitted on 28 Apr 2023

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

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

**THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE
FRANCHE-COMTÉ**

PRÉPARÉE À L'UNIVERSITÉ DE BOURGOGNE

École doctorale n°37

Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique

par

ROMAIN CENDRE

**Classification par méthodes d'apprentissage supervisé et faiblement supervisé
d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie**

Thèse présentée et soutenue à Dijon, le Date de soutenance

Composition du Jury :

EL MOATAZ ABDERRAHIM	Professeur, Université de Caen Normandie	Rapporteur
QUAFAFOU MOHAMED	Professeur, Aix-Marseille Université	Rapporteur
DAUL CHRISTIAN	Professeur, Université de Lorraine	Président du Jury
CINOTTI ELISA	Professeure, Université de Sienna	Examineur
PERROT JEAN-LUC	PU-PH, CHU de Saint-Étienne	Examineur
MARZANI FRANCK	Professeur, Université de Bourgogne	Directeur de thèse
MANSOURI ALAMIN	Professeur, Université de Bourgogne	Co-directeur de thèse

Title : Classification of multimodal images by use of supervised and weakly supervised learning methods to provide aid in the diagnosis of lentigo malignant in dermatology

Mots-clés : Supervised learning, Weakly supervised learning, Deep learning, Multimodality, Dermatology, Lentigo Maligna, Lentigo Maligna Melanoma, Image classification, Data fusion, Dermatoscopy, Confocal reflectance microscopy

Abstract:

Carried out in collaboration with the Saint-Étienne University Hospital, this work provides additional information to help the skin diagnosis by providing new decision methods on Lentigo Maligna and Lentigo Maligna Melanoma pathologies. To this end, the modalities regularly used in clinical conditions are made available to this work and are orchestrated within a multimodal process. Among image modalities, may be mentioned the clinical photography, the dermatoscopy, and the confocal reflectance microscopy.

Initially, the first steps of this manuscript focus on reflectance confocal microscopy as the work in computer diagnostic assistance is relatively underdeveloped, in particular on the detection of pathologies of Lentigo Maligna and Lentigo Maligna Melanoma. To this end, several automatic learning methods are described to allow

the separation of this modality at the level of the image data according to the healthy, benign, and malignant annotations. Then, the most image-relevant methods are combined with supervised and weakly supervised prediction methods to enable decision-making at the lesion level.

Secondly, the proposed multimodal process resumes the clinical path, starting with the most affordable and accessible modality, namely clinical photography, dermatoscopy, and finally reflectance confocal microscopy, respectively. According to this principle, two schemes are proposed in this manuscript, one considered as without memory, i.e. requiring only the data made available by the current modality, and the other with memory, i.e. by an accumulation of information acquired previously.

Titre : Classification par méthodes d'apprentissage supervisé et faiblement supervisé d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie

Mots-clés : Apprentissage supervisé, Apprentissage faiblement supervisé, Apprentissage profond, Multimodalité, Dermatologie, Lentigo Maligna, Lentigo Maligna Melanoma, Classification d'images, Fusion de données, Dermatoscopie, Microscopie confocale par réflectance

Résumé :

Réalisé en collaboration avec le CHU de Saint-Étienne, ce travail s'inscrit dans la continuité d'outils d'aide au diagnostic de la peau par l'apport de nouvelles méthodes de prise de décision sur les pathologies de Lentigo Maligna et de Lentigo Maligna Melanoma. À cette fin, les modalités couramment employées en milieu clinique sont mises à disposition de ce travail et sont orchestrées au sein d'un processus de prise en charge multimodal. Parmi ces modalités mises à disposition, peuvent être citées la photographie clinique, la dermatoscopie et la microscopie confocale par réflectance.

Dans un premier temps, ce travail se focalise sur la microscopie confocale par réflectance dont les travaux d'aide au diagnostic par ordinateur sont relativement peu développés, notamment sur la détection des pathologies de Lentigo Maligna et de Lentigo Maligna Melanoma. À cet effet, plusieurs méthodes d'apprentissage automatique

sont présentées pour permettre la séparation de cette modalité au niveau des données images selon les annotations saines, bénignes et malignes. Puis, les méthodes les plus pertinentes au niveau image sont combinées avec des méthodes de prédiction supervisées et faiblement supervisées pour permettre une prise de décision au niveau des lésions.

Dans un second temps, le processus multimodal proposé reprend le cheminement clinique, en débutant par la modalité la plus abordable et accessible, soit respectivement la photographie clinique, la dermatoscopie et enfin la microscopie confocale par réflectance. Selon ce principe, deux schémas sont proposés dans ce manuscrit l'un considéré comme sans mémoire, c'est-à-dire ne requérant que les données mises à disposition par la modalité courante, et l'autre avec mémoire, c'est-à-dire par accumulation de l'information acquise préalablement.

“ *Des nains sur des épaules de géants.* ”

Bernard de Chartres, *Gloses sur Priscien*

REMERCIEMENTS

Comme l'usage le permet, je profite de ces quelques paragraphes pour exprimer par ces éléments subjectifs, mais pourtant cruciaux à l'élaboration d'une thèse, ma gratitude envers celles et ceux qui m'ont soutenu durant la concrétisation de ce projet.

En cela, je tire parti de ces premières lignes pour remercier les diverses personnes rencontrées à diverses étapes de ma vie qui m'ont soutenu dans ma démarche de poursuite d'étude pour ce doctorat et n'ont pas hésité à me recommander lors de ma candidature. Ces quelques lignes sont dédiées à Pierre-Frédéric Villard et Patrick Adelbrecht, rencontrés lors de mes études à l'IUT de Saint-Dié des Vosges ; à Patrick Sabi, rencontré lors de mon parcours à 3iL à Rodez ; ainsi qu'Emilien Micard et Jacques Felblinger, rencontrés durant mes années de passage au CIC-IT de Nancy.

Ainsi, ces travaux ont été rendus possibles grâce à la Région de Bourgogne Franche Comté et le **Fonds européen de développement régional, (FEDER)** dans le cadre de l'initiative des **Jeunes Chercheurs Entrepreneurs, (JCE)**. Je tiens naturellement à leur exprimer mes plus sincères remerciements, sans quoi toute cette aventure n'aurait pas pu voir le jour. En effet, ces trois années auront été l'occasion d'une expérience scientifique, mais également de formations intenses à l'administration des entreprises auprès de l'IAE de Dijon et de la Business School of Burgundy.

Il est important de continuer ces remerciements envers ceux qui ont initié ce travail, Messieurs Franck Marzani et Alamin Mansouri, et qui ont soutenu ma candidature à ce financement et à ce projet commun. Par ailleurs, je tiens à les remercier pour les nombreuses réunions et relectures de ce manuscrit malgré les nombreuses responsabilités qui leur incombent, et de manière individuelle merci à Alamin Mansouri pour ses nombreux encouragements et son apport non négligeable à l'un des articles de conférence sans lequel il n'aurait pu paraître. Par cette même occasion, je dédie mes remerciements à deux médecins, Jean-Luc Perrot et Elisa Cinotti, dont la participation, le temps consacré, ainsi que la bonne humeur et l'énergie dégagée auront été essentiels à la réussite de ce projet. Je profite également de ce paragraphe pour remercier chaleureusement la société Pixience qui n'a pas hésité à soutenir ce sujet de thèse lors de la candidature aux financements **JCE**.

Les dernières étapes de ce manuscrit ne peuvent aboutir sans l'apport de diverses personnes non mentionnées à ce stade des remerciements. Ainsi, je tiens à exprimer ma gratitude à Messieurs Abderrahim El Moataz et Mohamed Quafafou qui ont accepté la tâche de rapporteur dans le cadre de ce manuscrit mais également de membre du jury. Je profite de ce propos pour remercier d'autres membre du jury pour leur participation chaleureuse : Monsieur Christian Daul, ainsi que des médecins déjà mentionnés, Jean-Luc Perrot et Élisabeth Cinotti.

De plus, je remercie tout naturellement ma structure d'accueil, le laboratoire **Imagerie et Vision Artificielle, (ImViA)**, et l'ensemble de ses membres qui m'auront permis de travailler dans un cadre des plus favorables durant ces trois années. Ainsi pour toutes les démarches administratives, je citerai Fabienne Blaise, Anne Fournier, et Khadija Jourani qui m'ont épaulé pour la plupart d'entre elles et qui ont été d'une réactivité à toute épreuve. Notre quotidien serait fortement impacté sans une équipe soudée, ainsi je remercie Cyrille Migniot, Fan Yang, Barthélémy Heyrman, et Julien Dubois qui ont beaucoup contribué à la bonne ambiance de cette structure. J'exprime

également une pensée envers Dominique Gin hac et Barthé lémy Heyrman qui ont gracieusement mis à ma disposition des ressources informatiques sans lesquelles ce travail aurait été difficile à réaliser.

J'ai une pensée toute particulière pour l'ensemble de mes collègues de laboratoire et de Master, qui par leur présence m'ont rendu ces trois années plus conviviales. C'est avec nostalgie que je repense à d'anciennes discussions en compagnie d'Abdelali Douiyek et Mustapha Bouderbane, mes anciens compagnons *d'ilot*, remplacés par une nouvelle équipe tout aussi sympathique avec Abir Zendagui, Marvin Nurit et Yuly Castro. Également, je citerai Margarita Khokhlova, Pierre Bonazza, Alpha Liu et Antoine Léger pour nos discussions passionnées et instructives autour de l'apprentissage automatique et profond. Dans ce registre, un merci à Alpha Liu et David Lewis pour nos échanges sur des thèmes très variés en anglais, ce qui m'a permis d'entretenir cette langue. Enfin je termine des collègues dont le passage m'aura marqué, c'est notamment le cas pour Deivid Botina Monsalve, Axel Moinet, Serge Bobbia, Yoan Marin et Duncan Luguern.

Que serait une thèse sans quelques soupapes pour évacuer le stress accumulé. Je profite de ce paragraphe pour dresser quelques attentions à deux membres : Fan Yang et Yannick Benezeth, qui n'ont jamais hésité à me remonter le moral dans les moments les plus durs. J'ai évoqué des soupapes, la course à pied est l'une d'entre elles, et c'est donc une attention toute particulière que je porte à notre noyau de course soudé composé de Fan Yang, Yannick Benezeth, Cédric Démonceaux, Alexandre Krebs, Abir Zendagui et Rita Meziati. Enfin, c'est grâce à une personne de passage au laboratoire **ImViA** et passionné de cross-country, que j'ai pu découvrir les paysages de Bourgogne, je remercie pour cela Olivier Roblin.

Mes propos sont évasifs à ce sujet, et pourtant le Master aura été contrainte et évasion, et m'aura permis des rencontres impossibles dans d'autres circonstances. Ainsi, je vous remercie Antoine Greisner, Jérémie Vion et Augustin Cognacq d'avoir été de fidèles collègues de travail et amis. Dans une catégorie similaire, mais pourtant lié à ces enseignements de Master, je souhaiterais citer mes collègues de **JCE** dont Alicia Froidurot, Garance Dejoux et Anne Berrou font partie. Je les considère avant tout comme des compagnons de galère car nous avons partagé de nombreuses périodes délicates, tiraillés entre nos responsabilités au laboratoire et envers ces cours.

Bien évidemment, je ne peux finir ces remerciements sans oublier mes proches qui ont été constamment présents à mes côtés, tant durant les moments joyeux que les plus difficiles et ont eus des paroles réconfortantes. Tout d'abord, mes amis avec qui il n'a pas été facile de garder contact malgré la distance et pour lesquels j'ai pu assister aux diverses étapes de leur vie, et ce jusqu'à devenir une famille. Enfin, je terminerais par mon cercle le plus proche, incluant mes parents qui m'ont porté durant toutes ces années et qui m'ont encouragé à aller de l'avant. Et je ne peux oublier ma compagne, Claire, qui m'épaule dans l'ombre depuis maintenant cinq années, et dont la présence me permet de surmonter maintes périodes difficiles. C'est avec joie que je te porte un grand merci.

SOMMAIRE

Introduction	1
I Contexte	7
1 La peau et ses lésions	9
1.1 Présentation, composition et fonctions	10
1.1.1 Présentation	10
1.1.2 Épiderme	11
1.1.3 Jonction dermo-épidermique	12
1.1.4 Derme	13
1.1.5 Hypoderme	14
1.1.6 Types de peau	14
1.2 Lésions de la peau	15
1.2.1 Lésions bénignes	15
1.2.2 Lésions malignes	16
1.2.2.1 Carcinome spinocellulaire (ou épidermoïde)	16
1.2.2.2 Carcinome basocellulaire	16
1.2.2.3 Mélanome	17
1.2.2.4 Une pathologie particulière : le Lentigo maligna	18
2 Des propriétés physiques aux données cliniques	21
2.1 Propriétés de la peau	22
2.2 Modalités d'imagerie non-invasives	25
2.2.1 Modalités par mesure optique	25
2.2.1.1 Imagerie clinique	26
2.2.1.2 Dermatoscopie	27
2.2.1.3 Microscopie confocale par réflectance	29
2.2.1.4 Tomographie en cohérence optique	30
2.2.1.5 Spectroscopie par réflectance	31

2.2.1.6	Imagerie multispectrale	32
2.2.1.7	Microscopie confocale Raman	33
2.2.2	Modalités par mesure non-optique	34
2.2.2.1	Imagerie par résonance magnétique	34
2.2.2.2	Échographie haute fréquence	35
3	Intelligence artificielle et notions complémentaires	37
3.1	Généralités sur l'intelligence artificielle	38
3.2	Approches par apprentissage	39
3.2.1	Apprentissage supervisé	40
3.2.2	Apprentissage non-supervisé	43
3.2.3	Apprentissage semi-supervisé	44
3.3	Apprentissage profond	45
3.3.1	Réseau de neurones artificiels	45
3.3.2	Réseau neuronal convolutif	46
3.3.3	Transfert de connaissances	47
3.4	Paramétrage et évaluation de modèles prédictifs	48
3.4.1	Généralisation de modèles	48
3.4.2	Choix de métriques	49
3.4.3	Réglages des hyperparamètres	52
3.4.4	Méthodes d'évaluation	53
3.5	Fusion d'informations	55
4	Données de travail et méthodes cliniques	57
4.1	Données de travail	58
4.1.1	Présentation des données	58
4.1.2	Structuration des données	61
4.1.3	Interface logicielle	62
4.2	Méthodes cliniques de diagnostic	63
II	Analyse de la microscopie confocale par réflectance	69
	Préambule microscopie confocale par réflectance	71
5	Diagnostic de l'image	77
5.1	Méthodologie	78

5.2	Méthodes d'extraction de caractéristiques	79
5.2.1	Méthodes manuelles d'extraction de caractéristiques spatiales	79
5.2.2	Méthodes manuelles d'extraction de caractéristiques fréquentielles	82
5.2.3	Méthode d'apprentissage par transfert de connaissances	85
5.2.4	Analyse préliminaire des méthodes d'extraction	88
5.3	Pré-traitements de caractéristiques	95
5.3.1	Réduction de dimensions	95
5.3.2	Normalisation de caractéristiques	96
5.3.3	Balancement de données	97
5.4	Processus de prédiction	99
5.4.1	Méthodes de prédiction	99
5.4.2	Procédure d'évaluation	100
5.5	Résultats	103
5.6	Discussion	107
6	Amélioration du diagnostic image	109
6.1	Méthodologie	110
6.2	Approche multi-échelle	111
6.2.1	Décomposition en ondelettes multi-échelle	111
6.2.2	Décomposition spatiale multi-échelle	113
6.3	Approche par fenêtre glissante	115
6.4	Réglage fin de réseaux de convolution	118
6.4.1	Présentation et mise en oeuvre de la méthode	118
6.4.2	Augmentation de données	119
6.4.3	Programme d'apprentissage	120
6.5	Présentation des résultats	121
6.5.1	Protocole d'expérimentation	121
6.5.2	Résultats des expérimentations	122
6.5.3	Méthodes de visualisation	125
6.6	Discussion	128
7	Prise de décision au niveau lésionnel ou patient	131
7.1	Méthodologie	132
7.2	Approches supervisées	133
7.3	Approches faiblement supervisées	135
7.4	Présentation des résultats	137

7.4.1	Protocole d'expérimentation	137
7.4.2	Résultats des expérimentations	138
7.5	Discussion	144
III	Analyse par processus multimodal	147
	Préambule multimodalité	149
8	Vers un processus multimodal d'aide au diagnostic clinique	153
8.1	Méthodologie	154
8.2	Méthodes d'aide au diagnostic par modalité	155
8.2.1	Méthodes d'extraction et de détection	155
8.2.2	Pré-analyse	156
8.3	Calibration de modèles	157
8.4	Processus d'aide au diagnostic multimodal	158
8.4.1	Présentation des processus	158
8.4.2	Modèle d'optimisation des seuils de confiance	161
8.5	Présentation des résultats	162
8.5.1	Protocole d'expérimentation	162
8.5.2	Résultats des expérimentations	163
8.6	Discussion	168
	Conclusion et perspectives	171
	Bibliographie	177
	Liste des figures	193
	Liste des tables	203
	Liste des acronymes	207

INTRODUCTION

La peau est un des organes les plus importants du corps humain et recouvre l'ensemble de celui-ci. Ses lésions sont nombreuses : des plus anodines communément appelées lésions bénignes, aux plus graves d'entre elles regroupées sous le terme de lésions malignes. Ces dernières sont d'origine cancéreuse et engendrent des répercussions dramatiques lorsqu'elles ne sont pas détectées à temps.

Ainsi, ces lésions malignes sont issues d'une division ou d'une mutation anormale de cellules de la peau. Leur cancérogénèse est le fait d'altérations du matériel génétique congénital sous l'effet d'agents extérieurs chimiques, tel que le tabac ou les médicaments, mais aussi viraux. Néanmoins, ces altérations ont le plus souvent pour origine l'action de radiations ionisantes. Celles-ci, impliquées dans la survenance des cancers cutanés sont essentiellement d'origine naturelle (soleil) mais peuvent être aussi d'origine artificielle : lors de loisirs (cabines de bronzage) mais également d'examen (scanners, radiographies) ou de thérapies (radiothérapie).

Afin de limiter l'incidence de ces tumeurs, de nombreuses campagnes de prévention sont menées au sein des pays les plus touchés, intégrant une dimension de sensibilisation et des actions de prévention dans le but d'augmenter le confort de vie futur des individus et de limiter les dépenses de santé en prévenant l'apparition des tumeurs malignes. Bien que souvent sous des formes plus traditionnelles telles que des messages publicitaires, ces campagnes de sensibilisation à but préventif peuvent prendre diverses formes afin de toucher au mieux les personnes cibles. À titre d'exemples, peuvent être citées la mise en place de formations de prévention dès le plus jeune âge au sein des structures scolaires aux États-Unis, ou la multiplication des supports de communication avec, par exemple, l'utilisation de sachets de sucre au Portugal [77, 44].

Néanmoins, les mesures de sensibilisation tendent aujourd'hui à dépasser la prévention. En effet, certaines publications scientifiques proposent une auto-surveillance régulière par des gestes simples, permettant une prise en charge des pathologies dans un meilleur délai par les dermatologues [70]. En parallèle, de plus en plus de campagnes de dépistage sont mises en place par les gouvernements. Ces diverses actions sont particulièrement efficaces pour permettre la détection de mélanomes à des stades moins avancés, et donc plus susceptibles d'être correctement pris en charge et traités par les services cliniques.

Malgré ces nombreuses démarches, le taux d'occurrence de ces cancers ne cesse de progresser dans le monde. Ces dernières années, l'**Organisation Mondiale de la Santé, OMS** (*alt. : World Health Organization, WHO*) dénombre entre 2 et 3 millions de cancers de la peau par an, dont environ 360 000 sont des mélanomes. Ces chiffres sont corroborés par ceux d'études épidémiologiques réalisées en 2015 [205], dont la Table i.1 reprend les diverses données. Ainsi, certaines régions du globe comme l'Australasie ou encore l'Amérique du Nord sont particulièrement touchées par cette recrudescence des maladies de la peau avec de fortes incidences [96]. De même, au niveau des populations touchées, le facteur d'âge semble être un critère de risque important [96].

Organe ou dénomination	Incidence (milliers)	Décès (milliers)
Tous	17 481	8 713
Autre pharynx	161	64
Autres néoplasmes	756	372
Côlon et rectum	1 653	832
Col utérin	526	239
Estomac	1 313	819
Foie	854	810
Larynx	238	106
Leucémie	606	353
Leucémie lymphoïde aiguë	161	110
Leucémie lymphoïde chronique	191	61
Leucémie myéloïde aiguë	190	147
Leucémie myéloïde chronique	64	35
Lèvre et cavité buccale	410	146
Lymphome de Hodgkin	78	24
Lymphome non hodgkinien	666	231
Mélanome	352	60
Mélyome multiple	154	101
Mésothéliome	37	32
Nasopharynx	123	63
Œsophage	1 313	891
Ovaires	251	161
Pancréas	426	412
Prostate	1 618	366
Rein	425	137
Sein	2 422	534
Système nerveux	321	229
Testicules	72	9
Thyroïde	334	32
Trachée, bronches et poumon	2 019	1 722
Utérus	455	90
Vésicule et voies biliaires	188	140
Vessie	541	188

Table i.1 : Statistiques mondiales d'incidence et de mortalité des 32 groupes de cancers (selon **Classification Internationale des Maladies, (CIM-10)** de l'OMS) pour l'année 2015 [65]. Ces chiffres sont exprimés en milliers et ne recensent pas les cancers de la peau hors mélanome.

Pour la France, ce ne sont pas moins de 80 000 nouveaux cas de cancers de la peau estimés en 2015, dont 70 % d'entre eux sont des carcinomes basocellulaires ², soit la 6ème cause de cancer en France. Pour leur part, les mélanomes sont eux estimés à 14 300 cas par an dont 1 773 engendrent un décès [192]. Une détection précoce de ces tumeurs est donc nécessaire en matière de santé publique mais aussi économique, bien qu'il soit difficile d'obtenir des chiffres sur les coûts engendrés par ces pathologies pour la France. Néanmoins à titre d'exemple, les soins générés par les pathologies de la peau se chiffrent annuellement à 8 milliards de dollars aux États-Unis [61].

¹Source : Chiffres issus du site de la [Fondation pour la Recherche Médicale](#).

Le sujet de ce manuscrit « Classification par méthodes d'apprentissage supervisé et faiblement supervisé d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie » s'inscrit dans cette thématique et tente d'apporter des méthodes numériques permettant de faciliter le quotidien des dermatologues dans un contexte clinique de plus en plus contraignant. En effet, cette thèse répond en premier lieu à un besoin sociétal, si l'on considère la baisse constante du nombre de dermatologues en France depuis 2010, visible sur la Figure i.1. Ce travail consiste donc à augmenter la capacité de diagnostic tant de l'expert que du non expert, en optimisant les examens cliniques nécessaires et en réduisant le temps passé à la réalisation des divers examens.

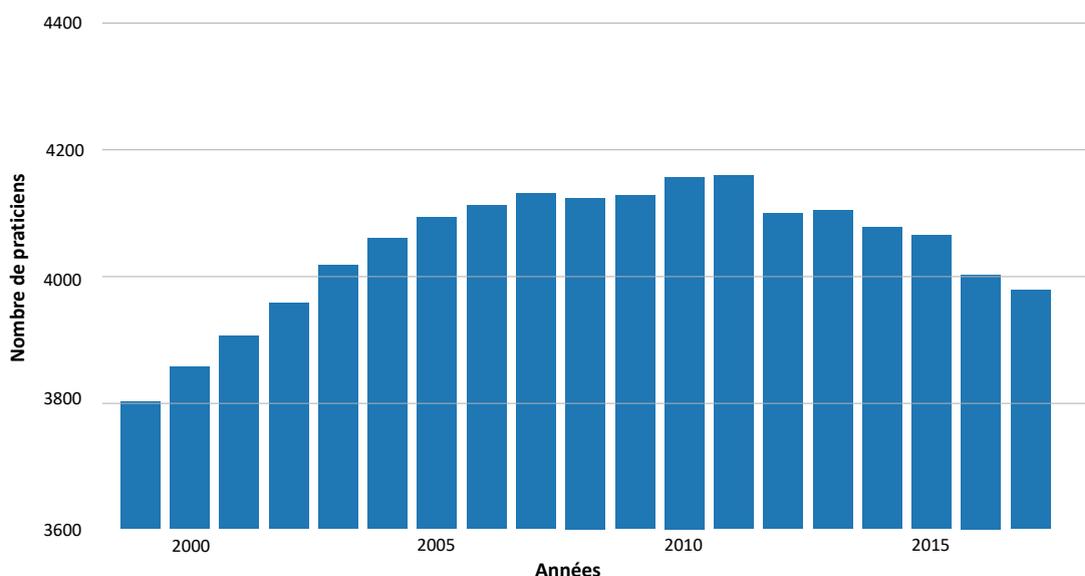


Figure i.1 : Graphique représentant l'évolution du nombre de dermatologues en France entre les années 1999 et 2017 ².

Outre cette réponse à un besoin sociétal, cette thématique ne demeure pas moins intéressante d'un point de vue scientifique. L'intérêt scientifique pour l'utilisation des outils d'aide au diagnostic en pratique dermatologique se traduit notamment par une forte tendance des thématiques de recherche sur le sujet comme le démontre la Figure i.2. Cette tendance est régie par plusieurs facteurs :

- Le **besoin sociétal** de la thématique déjà évoqué lors du précédent paragraphe et **l'intérêt des industriels** de proposer des biens et des services permettant une meilleure efficacité des professionnels du domaine.
- Le **phénomène de l'apprentissage automatique** ou *machine learning* qui résulte d'un engouement stimulé par des avancées constantes et soutenues du domaine mais aussi de la démocratisation des méthodes qui en sont issues.
- Le **défi scientifique** apporté par cette thématique de dermatologie, qui regroupe d'une part les données de travail (les appareils d'acquisition, les conditions d'acquisition, ...) et d'autre part l'information à extraire des données. Ces divers éléments sont approfondis tout au long de ce manuscrit.

²Source : Graphique généré à partir de données en provenance du site du [gouvernement](#) sur la santé.

Ce travail s'inscrit dans une démarche d'aide au diagnostic des lésions de la peau et particulièrement des pathologies de **Lentigo Maligna, LM** (*alt. : Mélanose de Dubreuilh, MD*) et **Lentigo Maligna Melanoma, LMM** (*alt. : Mélanome sur Mélanose de Dubreuilh, MMD*). De nombreux travaux se sont portés sur l'aide à la détection par ordinateur de lésions de la peau à partir d'une modalité unique d'imagerie. Néanmoins, peu d'entre eux s'intéressent à une démarche multimodale de cette thématique, dû à la complexité que génère une telle étude ou bien par manque ou insuffisance de données à leur disposition.

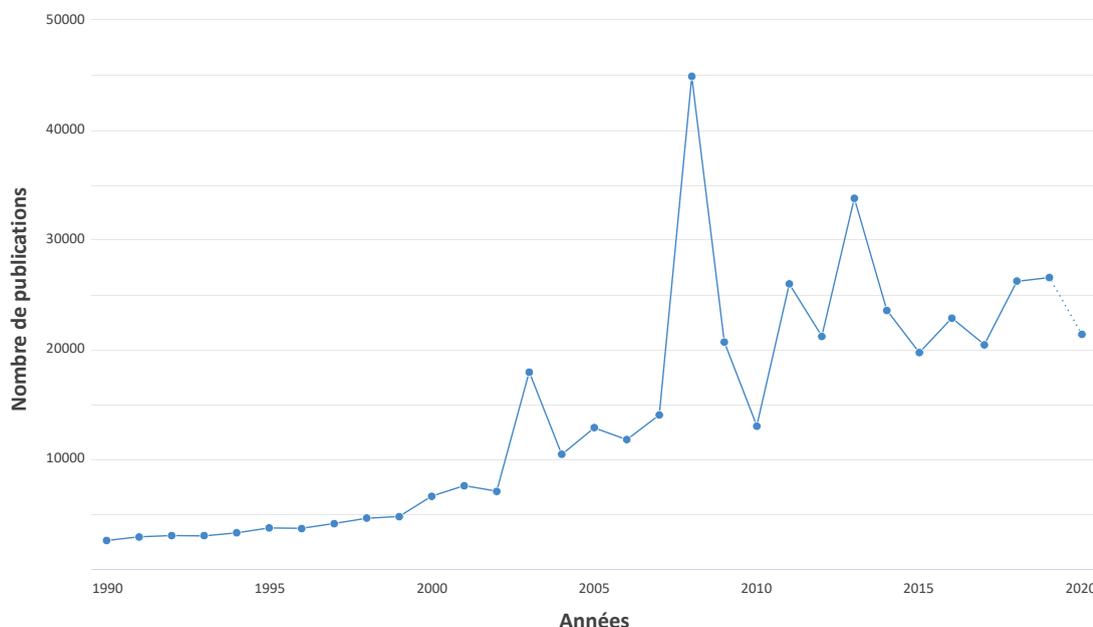


Figure i.2 : Graphique représentant l'évolution du nombre de publications recensées par le moteur de recherche [Google Scholar](#)³ autour des termes de « skin computer diagnosis » entre les années 1990 et 2020.

La matière première mise à la disposition de ce travail permet une orientation de ces travaux dans le sens de la multimodalité. En effet, l'une des problématiques majeures aujourd'hui en dermatologie est ce que les médecins qualifient de zone d'indécision, également appelée *zone grise*. Il s'agit d'une catégorie de cas cliniques pour lesquels le médecin n'a pas à un instant t suffisamment d'informations ou de connaissances pour prendre une décision. Ces cas nécessitent souvent une prise en charge plus importante, avec l'aide d'autres spécialistes ou encore avec notamment de nouveaux examens plus spécifiques.

La finalité de ce travail est de proposer diverses méthodes et outils permettant une aide à la décision dans la gestion de cette zone grise, et ainsi améliorer la prise en charge clinique en service de dermatologie. À l'instar des travaux existants, cette gestion de la zone grise passe d'une part par l'implémentation de méthodes propres à certaines des modalités de ce travail et d'autre part par la mise en place d'un schéma multimodal. Ainsi, un schéma résumant l'objectif de ce processus est visible sur la Figure i.3.

³Source image : Graphique généré à partir du site [Dimensions](#).

Dans ce schéma idéal représenté sur la Figure i.3, les patients jugés avec une certaine certitude comme **bénins** ou **malins**, seront exclus des procédures suivantes. Seuls les patients encore présents dans cette *zone grise* sont pris en charge pour des examens supplémentaires, pour lesquels ce processus est réitéré.

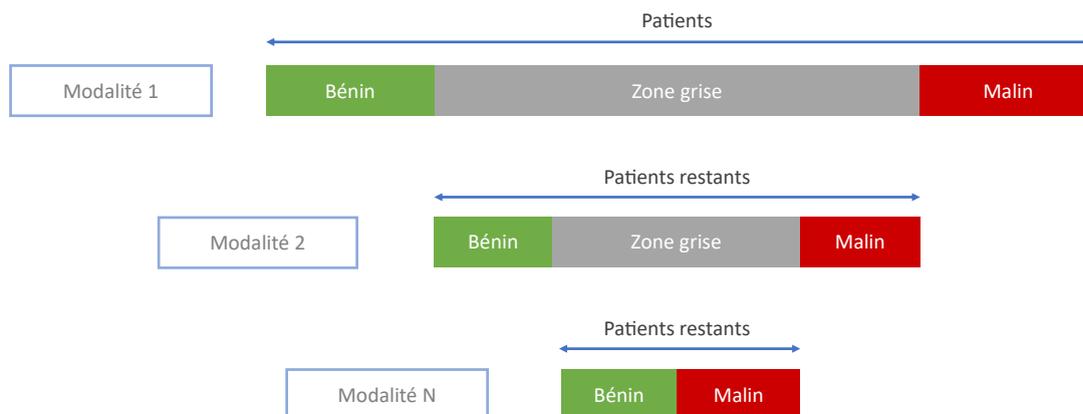


Figure i.3 : Schéma de représentation du processus de réduction de l'indécision, également surnommée zone grise. Ce travail s'inspire du processus cognitif des dermatologues et consiste à réduire l'indécision propre à chacune des modalités cliniques par l'ajout de nouvelles modalités d'imagerie.

À cette fin, ce travail mobilise des connaissances en provenance de divers champs d'applications propres : (i) à la peau d'un point de vue médical, (ii) aux modalités permettant l'observation des tissus, (iii) et à l'intelligence artificielle. Cette complémentarité est représentée sous forme de schéma sur la Figure i.4.

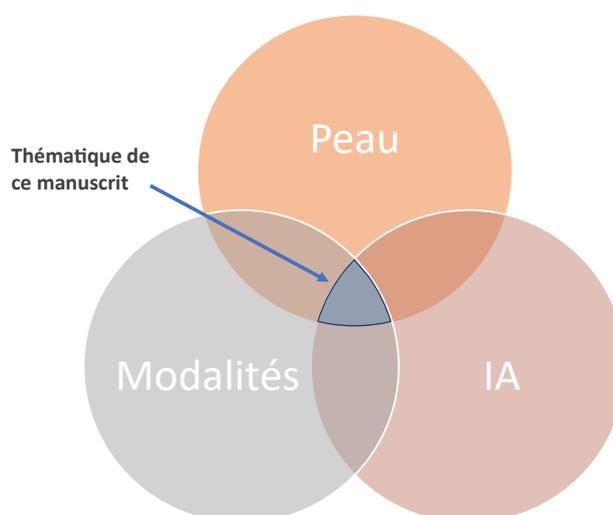


Figure i.4 : Schéma de représentation des divers domaines impliqués dans cette thématique de recherche. Ce travail est au confluent de connaissances de la peau, des modalités d'imagerie et des domaines de l'intelligence artificielle.

Dans un premier temps, le contexte de cette étude est abordé dans la Partie I. Cette partie débute par une présentation de la peau, l'organe majeur de cette étude, réalisée au sein du Chapitre 1. Puis, s'ensuit une mise en évidence des principes d'interaction entre la peau et la lumière, avant de présenter les techniques de visualisation à la disposition des médecins dans le Chapitre 2. Ensuite, le Chapitre 3 présente au lecteur l'ensemble des connaissances d'intelligence artificielle sollicitées dans ce manuscrit. Pour finir cette partie, le Chapitre 4 présente les données mises à disposition de cette étude mais également les principaux travaux en lien avec ce travail.

Dans un second temps, la modalité de **Microscopie Confocale par Réflectance, MCR** (*alt. : Reflectance Confocal Microscopy, RCM*) est abordée par la Partie II. À cette fin, cette partie débute par le traitement des images de cette modalité dans le Chapitre 5 dédié à leur classification de ces images et propose une extension de ces travaux dans le Chapitre 6. Pour finir, le Chapitre 7 propose diverses méthodes permettant à partir des images de remonter à la prise de décision au niveau d'une lésion.

Dans un dernier temps, la finalité de ce travail est traitée par la Partie III dédiée à l'apport de l'information multimodale. Dans ce but, le Chapitre 8 reprend les observations de travaux sur les modalités d'imagerie clinique et de dermatoscopie ainsi que celles issues de la précédente partie sur la modalité de **MCR**, et propose des méthodes d'agrégation séquentielles de ces informations.

Dans la mesure où certains termes d'origine anglaise ne possèdent pas de réel équivalent dans la langue française, ce travail favorise l'utilisation d'acronymes français pour les termes communément acceptés tels que les *modalités d'imagerie* ou encore les *pathologies de la peau*. En revanche, par souci d'homogénéité les *termes techniques sont proposés en anglais* bien que certains aient trouvé leur équivalent dans la langue française. Quel que soit l'acronyme employé, ce travail propose à la première utilisation un format du type : **Langue Cible, LC (alternative : Langue Alternative, LA)**.

I

CONTEXTE

LA PEAU ET SES LÉSIONS

Ce premier chapitre apporte une description de l'organe sujet de cette étude, à savoir la peau. Ainsi, ces quelques pages posent diverses bases en décrivant pas à pas des aspects qui permettent une compréhension de sa physiologie, de son fonctionnement et de ses rôles. De plus, ce chapitre est l'emplacement idéal pour aborder les multiples pathologies pouvant altérer cet organe.

Dans un premier temps, une explication des principales couches et composantes de la peau est réalisée par profondeur croissante. Dans un second temps, ce travail consacre une section dédiée à la présentation de quelques-unes de ses principales lésions. Par ailleurs, ces pages sont l'occasion d'aborder une première fois le Lentigo, pathologie centrale de cette étude, et tout particulièrement ses formes malignes : le **Lentigo Maligna** et le **Lentigo Maligna Melanoma**.

Sommaire

1.1	Présentation, composition et fonctions	10
1.1.1	Présentation	10
1.1.2	Épiderme	11
1.1.3	Jonction dermo-épidermique	12
1.1.4	Derme	13
1.1.5	Hypoderme	14
1.1.6	Types de peau	14
1.2	Lésions de la peau	15
1.2.1	Lésions bénignes	15
1.2.2	Lésions malignes	16

1.1/ PRÉSENTATION, COMPOSITION ET FONCTIONS

La première section de ce chapitre dédié à la peau propose d'introduire de manière synthétique la composition et le fonctionnement de cet organe, avant de présenter ses lésions. Ainsi, les différentes couches la définissant sont abordées par ordre croissant de profondeur lors des prochains paragraphes.

1.1.1/ PRÉSENTATION

Le terme *peau* caractérise dans son sens le plus global, l'enveloppe externe propre aux vertébrés. Présente chez l'homme, elle est l'un de ses organes majeurs mais également l'un des plus lourds : chez un individu adulte d'un poids de 70 kg, elle représente une surface plane d'environ 2 m², soit une masse estimée à 5 kg [133]. Son épaisseur diffère selon la zone du corps étudiée : de 0,5 mm au niveau de zones fines telles que les paupières, à plusieurs centimètres sur les zones adipeuses. La surcharge pondérale influe également de manière importante sur son épaisseur, bien plus élevée chez des personnes souffrant d'obésité.

Bien que la peau soit l'un des premiers éléments visibles d'un individu, son fonctionnement et ses fonctionnalités sont souvent méconnus par la plupart des individus. Ses rôles sont pourtant multiples, parmi lesquels :

- un rôle **physique**, en apportant une protection contre le monde extérieur,
- un rôle **mécanique**, en limitant la perte d'eau ou en offrant une thermorégulation,
- un rôle **sensoriel**, en permettant l'appréhension du monde extérieur.

La dermatologie, branche médicale dédiée à l'étude de la peau, permet de mieux comprendre sa composition, son fonctionnement, ses vulnérabilités et permet d'apporter conseils, préventions et traitements dans le cas de certaines pathologies.

Cette barrière se décompose en diverses couches fondamentales, représentées sur la Figure 1.1, dont : l'**épiderme**, la **jonction dermo-épidermique**, le **derme** et l'**hypoderme**.

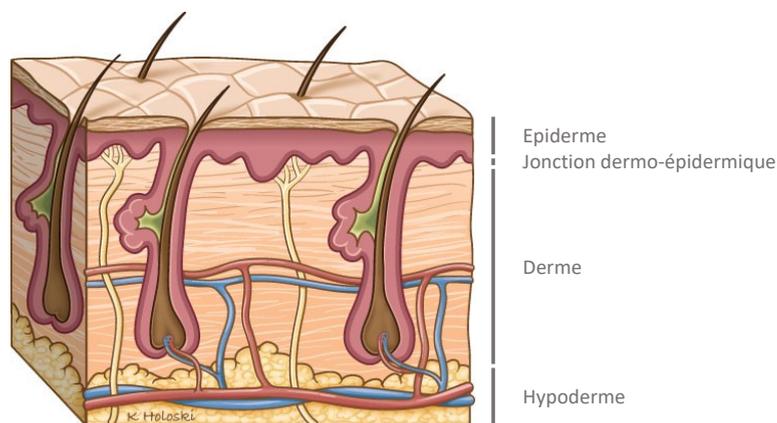


Figure 1.1 : Illustration de la peau et de ses diverses couches ¹.

¹Image source : Dessin par Kellie Holoski – Illustrateur médical.

1.1.2/ ÉPIDERME

L'épiderme correspond à la couche superficielle de la peau et mesure entre 0,01 mm et 0,1 mm d'épaisseur [167]. Un aperçu des différents types de cellules composant cette couche est visible sur la Figure 1.2, parmi lesquelles sont recensés :

- les **kératinocytes** : des cellules dont le rôle est d'offrir une protection contre l'environnement extérieur et d'apporter une imperméabilité. Elles représentent 90 % des cellules de l'épiderme.
- les **cellules de Langerhans** : des cellules dont le rôle est immunitaire.
- les **cellules de Merkel** : des cellules dont le rôle est sensitif.
- les **mélanocytes** : des cellules dont le rôle est d'assurer la protection contre les rayonnements Ultraviolet, (UV) par l'apport d'une pigmentation.

L'une des caractéristiques essentielles de cette couche est le fonctionnement de son processus de différenciation des cellules. Les kératinocytes migrent de la couche inférieure à la couche supérieure en subissant différentes modifications chimiques conduisant à la perte de leur noyau et à leur mort. Ce processus porte le nom de kératinisation, et se subdivise en sous couches aux propriétés variées, respectivement par profondeur croissante : Cornée (*Stratum corneum*), Transition (*Stratum lucidum*), Granuleuse (*Stratum granulosum*), Épineuse (*Stratum spinosum*), Basale (*Stratum basale*). Ces différentes couches peuvent être observées sur la Figure 1.2. En fin de parcours, ces cellules perdent leur cohésion dans un processus de desquamation conduisant la peau à se renouveler.

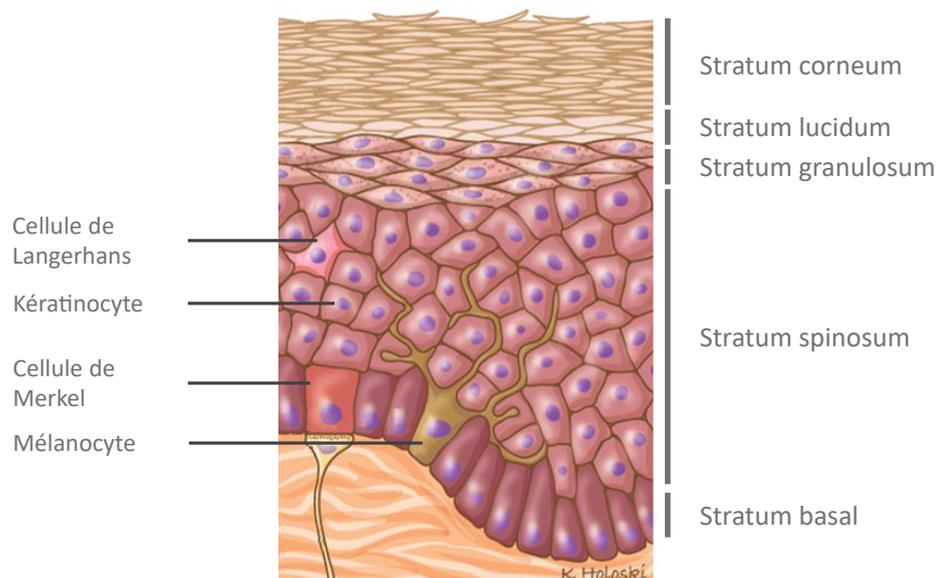


Figure 1.2 : Illustration de l'épiderme reprenant ses divers composés cellulaires ainsi que les diverses couches formées par un processus de différenciation ².

²Image source : Dessin par Kellie Holoski – Illustrateur médical.

1.1.3/ JONCTION DERMO-ÉPIDERMIQUE

La **Jonction Dermo-Épidermique**, JDE (*alt. : Dermo Epidermal Junction, DEJ*), également connue sous le terme de *membrane basale*, se situe à la jonction de l'épiderme et du derme, assurant la cohésion entre ces deux couches. Selon la zone du corps considérée, son épaisseur estimée varie de 60 nm et 300 nm.

Représentée sur la Figure 1.3, cette membrane est scindée en 2 parties :

- la **lame basale** (*lamina basalis*), pouvant être subdivisée en deux nouvelles couches : la *lamina lucida* et la *lamina densa*,
- la **lame réticulaire** ou zone fibrillaire (*lamina reticularis*).

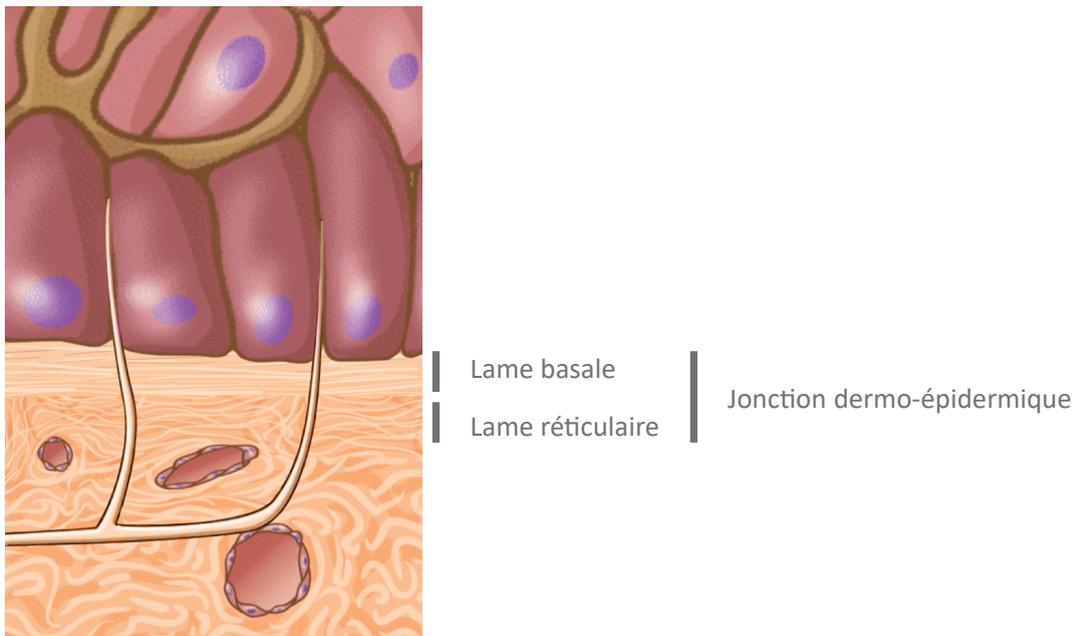


Figure 1.3 : Illustration de la **Jonction Dermo-Épidermique**³. Cette jonction dont le rôle est d'assurer une cohésion entre l'épiderme et le derme peut être subdivisée en deux sous couches : la lame basale et réticulaire.

Il s'agit d'une zone fibreuse, composée de laminine mais également de collagène de type III / IV / VI lui conférant entre autres des propriétés de résistance mécanique et élastique. Ces fibres lui permettent d'assurer une cohésion entre épiderme et derme au travers de points d'ancrage. Sa seconde fonction essentielle est au travers d'un mécanisme de régulation des échanges moléculaires, permettant d'assurer la nutrition des cellules de base de l'épiderme. Pour finir, elle possède un rôle fondamental de ré-épidermisation essentiel à la cicatrisation.

³Image source : Principles of Anatomy and Physiology – John Wiley & Sons.

1.1.4/ DERME

Le derme est une couche d'une épaisseur estimée comprise entre 0,5 mm et 5 mm, composée majoritairement de collagène à hauteur de 80 % (sur la matière sèche), et de fibres élastiques [133]. Cette couche est traversée par de nombreux éléments dont :

- des **vaisseaux sanguins** qui apportent des nutriments,
- des **vaisseaux lymphatiques** qui assurent une fonction immunitaire.

En outre, cette strate contribue de manière essentielle à des aspects de résistance et de contrainte mécanique de la peau, ainsi qu'aux mécanismes de thermorégulation et de cicatrisation de cette dernière. Enfin, cette couche tient un rôle essentiel dans la perception sensorielle (somesthésie) assurant d'une part des propriétés liées à la pression (mécanorécepteurs) mais également liées à la chaleur (thermorécepteurs). Outre ces fonctions, on lui attribue également un rôle nutritif important de par son irrigation sanguine.

La littérature scinde le derme en deux parties majeures, visibles en Figure 1.4, dont :

- le **derme Papillaire** (Papillary dermis), permettant essentiellement la thermorégulation,
- le **derme Réticulaire** (Reticular dermis), assurant la plupart des fonctions mécaniques.

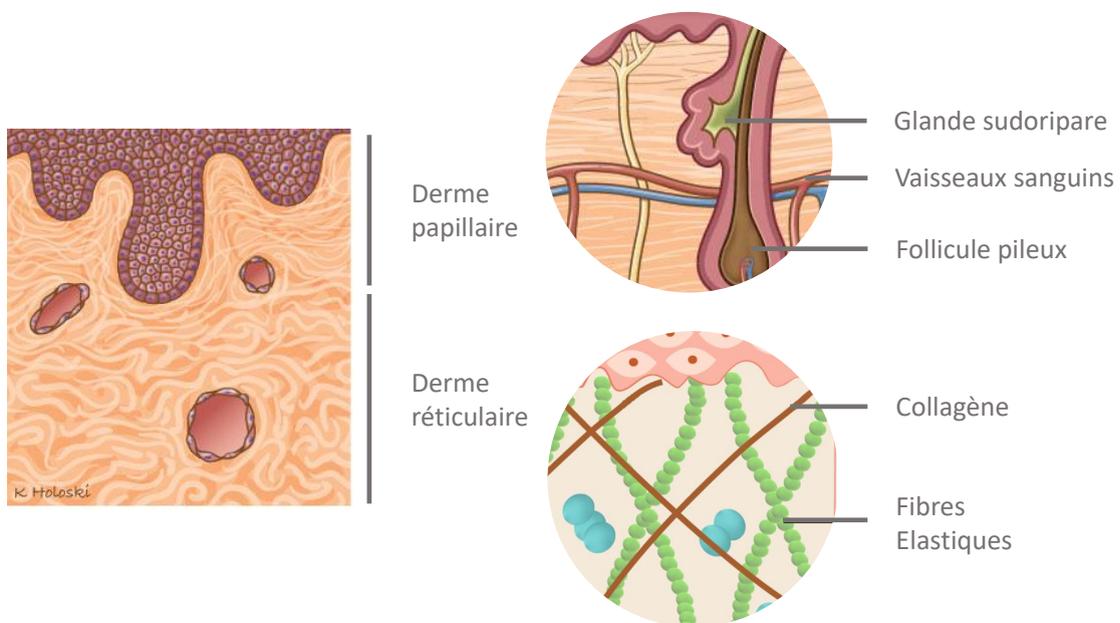


Figure 1.4 : Illustration de représentation du derme⁴. Le derme est subdivisé en deux couches aux propriétés distinctes : le derme papillaire et réticulaire.

⁴Image source : Dessin par Kellie Holoski – Illustrateur médical.

1.1.5/ HYPODERME

L'hypoderme, également appelé couche sous-cutanée, est un tissu présent sous le derme composé essentiellement de graisses (adipocytes, cellules spécialisées dans le stockage de graisses). Son épaisseur est extrêmement variable de 0,1 cm à plusieurs cm, selon :

- la zone considérée,
- l'âge du patient,
- l'alimentation,
- les prédispositions génétiques.

Cette couche de tissu assure diverses fonctions, dont :

- le passage des vaisseaux sanguins et lymphatiques, ainsi que celui des nerfs jusqu'au derme,
- l'interface entre les structures sous cutanées et la peau.

1.1.6/ TYPES DE PEAU

Il est nécessaire afin de bien aborder ce sujet et ses problématiques, d'être conscient des variations inter individus. Pour répondre partiellement à ces variations, Fitzpatrick a établi une classification des profils type de peau, visant à caractériser leur réaction suite à l'exposition au soleil. Ce travail a également permis de créer une association entre les profils type et les risques de cancers associés [66].

Cet aspect est un élément important à prendre en compte lors de l'appréhension de cette problématique. En effet, la plupart des travaux d'aide au diagnostic par ordinateur de la littérature ne décrivent pas les types de peau employés [27, 209, 100]. Néanmoins, les échantillons mis en avant par ces travaux semblent majoritairement axés sur les types I et II selon l'échelle de Fitzpatrick. Certaines pathologies sont ainsi plus propices à se développer sur un type de peau en particulier [141]. De plus, une même pathologie peut diverger en termes de caractéristiques [198]. La Figure 1.5 propose une synthèse visuelle des types de peau identifiés par Fitzpatrick, mais également de la résistance aux UV et du risque de cancer associé à chacun d'entre eux.

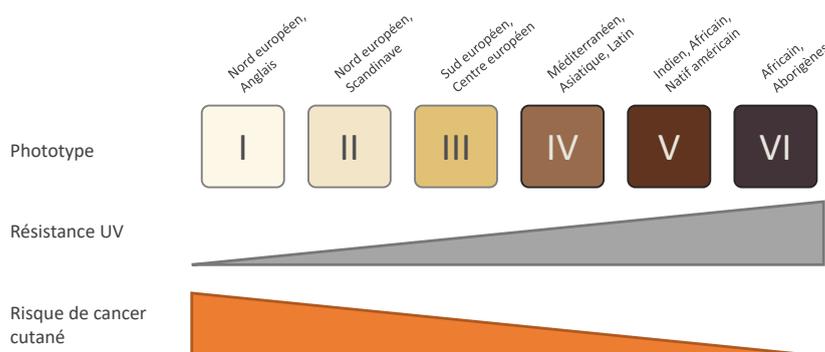


Figure 1.5 : Échelle de Fitzpatrick associant type de peau et propriétés associées [66].

1.2/ LÉSIONS DE LA PEAU

Ce sujet abordant les lésions de la peau, il convient de définir le terme lésion pour bien situer l'objet de cette recherche. Selon PubMed, une lésion « correspond à toute anomalie touchant les tissus d'un organisme, généralement provoquée par une maladie ou par un traumatisme »⁵.

Les lésions de la peau correspondent à un terme générique caractérisant une partie de la peau ayant, en comparaison de la peau l'entourant, une croissance/apparence ou structure qui diffère de tissus considérés comme sains. Ces lésions peuvent ainsi être pré-natales ou post-natales, comporter diverses formes et peuvent présenter différents risques.

Ces lésions sont une catégorie très vaste dont les origines sont multiples, et peuvent être spécifiées par des termes complémentaires. Ainsi, le terme de *lésion pigmentaire* caractérise les lésions affectant la peau par la présence de pigmentations essentiellement liées à la mélanine ou aux globules rouges.

Ces lésions sont sommairement présentées au travers de ces quelques pages. D'une part, les **lésions bénignes** sont abordées à l'aide d'une brève sous-section, et d'autre part, les **lésions malignes** sont traitées plus en détail lors d'une seconde sous-section présentant les pathologies majeures mais également les cas de **Lentigo Maligna** et **Lentigo Maligna Melanoma**.

1.2.1/ LÉSIONS BÉNIGNES

Ces lésions désignent les diverses altérations classées comme sans conséquences graves pour la santé d'un individu. Néanmoins, ces lésions peuvent présenter un terrain pour des lésions plus dangereuses.

L'une des lésions bénignes les plus communes est le **naevus mélanocytaire**, à l'apparence de structures souvent circulaires ou ovales présentes en surface de peau chez l'homme. Les naevus sont susceptibles d'évoluer au cours du temps et peuvent rarement aboutir à un mélanome. La plupart d'entre eux apparaissent durant les trente premières années de vie en moyenne et présentent des couleurs variées, avec des teintes allant d'un rose rougeâtre jusqu'au noir. Parmi ces lésions, peuvent être cités :

- le **naevus commun**, également connu sous le nom de *grain de beauté*, est l'une des lésions les plus fréquentes composée de mélanocytes regroupés en amas.
- le **naevus atypique** se caractérise souvent par une pigmentation variable et des bordures irrégulières.
- le **naevus congénital** apparaît en début de vie, le plus souvent sous forme de tache.
- le **naevus de Spitz** s'apparente souvent à un mélanome de par sa croissance rapide (souvent de 2 à 6 mois) ainsi qu'une couleur brun-rouge.
- le **phénomène de Sutton** caractérisé par une dé-pigmentation aux abords d'un naevus existant sur une durée de quelques semaines à quelques mois.

La meilleure des solutions est de réaliser une surveillance régulière chez les individus afin de s'assurer que ces lésions ne dégèrent pas en pathologies malignes.

⁵Source : Définition par [PubMed Health](#).

1.2.2/ LÉSIONS MALIGNES

Le terme malin qualifie dans le contexte de ce manuscrit une pathologie cancéreuse dont les conséquences peuvent aller jusqu'au décès de l'individu sans l'intervention adéquate d'un spécialiste. Ces cancers de la peau sont issus d'une division ou d'une mutation anormale de cellules de la peau et le risque de décès est principalement causé par des métastases de la tumeur initiale. Certaines de ces tumeurs peuvent être difficiles à prendre en charge car susceptibles de réapparaître sans prendre une marge de sécurité suffisante lors de l'excision. Outre le décès, ces lésions peuvent donc à des stades avancés et selon le type de tumeur engendrer des lésions cicatricielles permanentes à la suite de leur prise en charge chirurgicale, mais également des gênes corporelles importantes. Il est donc crucial de détecter au plus tôt ces lésions et de procéder à leur prise en charge le plus rapidement possible afin de limiter les conséquences, voire les désagréments.

Ces cancers résultent pour la plupart d'une exposition aux UV, principal facteur de mutation. D'autres facteurs tels que le tabagisme, les virus, des prédispositions génétiques ou encore l'utilisation de médicaments immunosuppresseurs peuvent favoriser leur apparition. Les prochaines sous-sections reprennent les pathologies malignes de la peau les plus importantes par ordre d'incidence.

1.2.2.1/ CARCINOME SPINOCELLULAIRE (OU ÉPIDERMOÏDE)

Le Carcinome Spino Cellulaire, CSC (*alt.* : *Squamous Cell Carcinoma, SCC*) se développe à partir de l'épiderme dont la croissance devient incontrôlée. Cette pathologie peut prendre de multiples apparences, variant de simples plaques rouges à des cas de plaies ouvertes. Cette pathologie est à même de provoquer des métastases chez un individu, et donc de se propager.

Ce type de cancer est au second rang des cancers cutanés les plus graves, principalement lié à cette caractéristique de propagation. Ces cancers émergent le plus souvent de zones exposées régulièrement au soleil telles que le visage, les oreilles, le cou, . . . Ces zones dénotent le plus souvent des symptômes liés au dommage du soleil, tels que :

- des rides,
- des tâches,
- une perte d'élasticité.

Les prédispositions génétiques, les conditions de travail et le sexe d'un individu sont autant de facteurs pouvant influencer sur l'apparition de ce type de cancer.

1.2.2.2/ CARCINOME BASOCELLULAIRE

Le Carcinome Basocellulaire, CBC (*alt.* : *Basal Cell Carcinoma, BCC*) correspond à une tumeur maligne développée à partir des cellules basales. Il s'agit de la forme de cancer de la peau la plus fréquente chez l'humain. Ses conséquences médicales sont le plus souvent d'ordre cicatriciel, liées à l'exérèse de cette tumeur. Néanmoins, le coût sociétal engendré est considérable du fait de leur grande fréquence. En effet, ces cancers ont peu de risques de métastases et sont donc moins enclins à se propager à l'ensemble de l'organisme et de provoquer la mort du patient. Ce cancer peut résulter de divers facteurs comme ses homologues, dont les principaux restent la surexposition au soleil ou la défaillance du système immunitaire.

1.2.2.3/ MÉLANOME

Le mélanome est l'un des cancers de la peau les plus agressifs, également qualifié de tumeur mélanocytaire car se développant à partir de mélanocytes. Ce type de cancer apparaît dans environ 70 % des cas sur une peau saine et dans les 30 % restant sur un naevus présent au préalable. Bien que cette tumeur ne représente que 2 % des cas de cancers de la peau [196], il s'agit en contrepartie du type le plus agressif de cancer de la peau. Au niveau mondial et pour l'année 2015, ce cancer a été détecté chez plus de 350 000 personnes et a provoqué le décès de 59 800 individus [96]. Son incidence est fortement aggravée par l'exposition prolongée au soleil et plus particulièrement aux UV. Par ailleurs, les types I et II sur l'échelle de Fitzpatrick sont les types de peau les plus concernés par cette pathologie.

Deux types de catégories de mélanomes sont recensés avec d'une part les pathologies superficielles caractérisant les mélanomes qui présentent une phase d'extension épidermique (extension horizontale) et d'autre part les pathologies qui présentent une phase de développement en profondeur d'emblée sans phase horizontale précessive (extension verticale).

La première catégorie, dont la sévérité est évaluée par l'épaisseur (indice de Breslow) et l'extension en profondeur (indice de Clark [42] - Figure 1.6), distingue trois types de pathologies :

- les mélanomes d'extension superficielle – environ 70 % des mélanomes,
- les mélanomes acro-lentigineux – environ 10 % des mélanomes,
- les **Lentigo Maligna** et de **Lentigo Maligna Melanoma** – environ 10 % des mélanomes [111].

La seconde catégorie est composée des mélanomes nodulaires, dont l'évolution plus rapide est liée à un développement vertical dès les premiers symptômes. L'indice de Clark est également employé dans le but de caractériser les stades de ce type de mélanome [42], représenté sur la Figure 1.6.

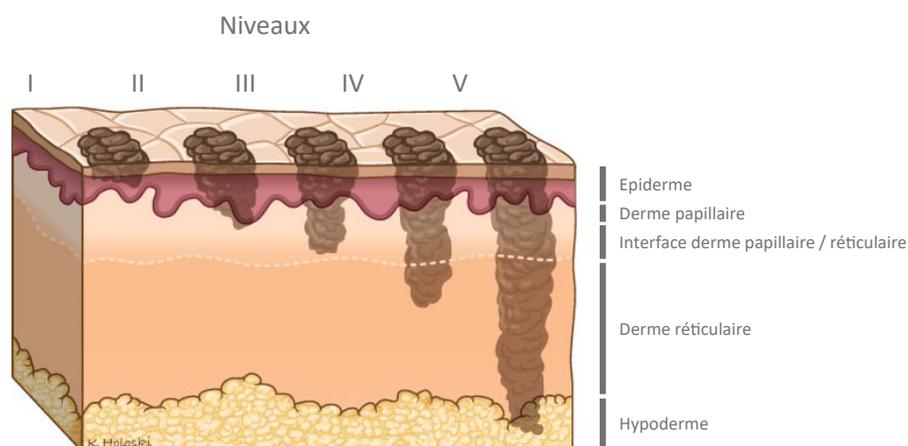


Figure 1.6 : Illustration de l'indice de Clark [42], défini par plusieurs niveaux de progression ⁶.

⁶Image source : Dessin par Kellie Holoski – Illustrateur médical.

1.2.2.4/ UNE PATHOLOGIE PARTICULIÈRE : LE LENTIGO MALIGNA

Les termes de **Lentigo Maligna, LM** (*alt. : Mélanose de Dubreuilh, MD*) et de **Lentigo Maligna Melanoma, LMM** (*alt. : Mélanome sur Mélanose de Dubreuilh, MMD*) correspondent à des tumeurs qui touchent les zones les plus exposées chez un individu tel que le visage chez les populations de plus de 60 ans. Comme précédemment énoncé, ils représentent environ 10 % des cas de mélanomes mais tendent à augmenter ces dernières années et vont désormais jusqu'à toucher les populations considérées comme jeunes. Le facteur explicatif le plus probable à ce jour est l'exposition au soleil croissante chez les différentes populations [9, 111, 57].

Le **LM** résulte d'une prolifération de cellules malignes mélanocytaires au sein de la couche basale de l'épiderme. Son développement progressif est constitué de plusieurs étapes successives, dont : (a) l'apparition de points gris et de globules, (b) l'apparition d'ouvertures asymétriques et de demi-cercles, (c) l'apparition de structures rhomboïdales et de cercles, (d) l'apparition de cercles imbriqués, (e) l'apparition de follicules comblés, (f) et pour finir l'apparition de zones sans structures visibles. Les étapes propres à ce développement sont schématisées sur la Figure 1.7. Ces observations typiques du développement de cette lésion vont dans le sens d'un suivi régulier par un expert permettant de lever toute ambiguïté.

En outre, le **LM** est considéré comme un *stade 0* également considéré sous le terme de *melanoma-in-situ*. Lorsque cette prolifération se propage aux cellules profondes de la peau au-delà de la **JDE** par le biais des follicules pileux, cette pathologie est alors qualifiée de **LMM** et donne lieu à un risque important de métastases résultant de cette infiltration de cellules cancéreuses. Bien que la transition de **LM** à **LMM** soit souvent d'une dizaine d'années, il n'est pas isolé d'observer des cas cliniques dont l'évolution a lieu entre deux examens bi-annuels [135, 111].

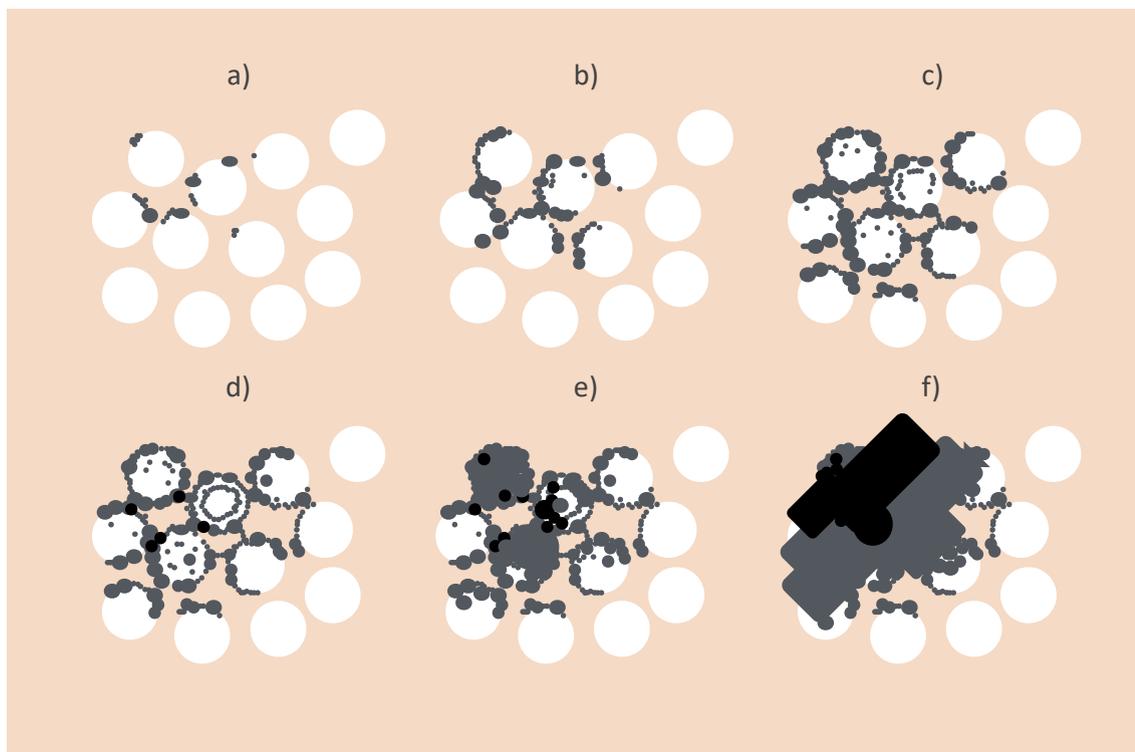


Figure 1.7 : Illustration par Jean-Luc Perrot des stades d'évolution d'un **LM**. Les points blancs permettent de représenter les follicules pileux.

Les pathologies de **LM** sont difficiles à diagnostiquer à l'œil nu et peuvent être confondues avec des pathologies bénignes telles qu'un **Lentigo Solaire, LS** (*alt. : Solar Lentigo, SL*), une **Kératose Séborrhéique, KS** (*alt. : Seborrheic Keratose, SK*) ou une **Kératose Actinique Pigmentée, KAP** (*alt. : Pigmented Actinic Keratosis, PAK*) [57]. Par ailleurs, les **LM** sont très différents des autres types de mélanomes dont les limites peuvent être difficiles à définir. En effet, leur prolifération périphérique peut-être discrète voire non pigmentée [111]. En termes de signes propres, le **LM** se caractérise le plus souvent par une macule pigmentée, et par la présence de bordures et de pigmentations irrégulières dont la taille atteint aisément une dizaine de centimètres. Pour sa part, le **LMM** se présente sous la forme d'une lésion tumorale papuleuse ou nodulaire [135, 111].

Afin de lever cette ambiguïté, il est possible d'avoir recours à l'**histopathologie**. Le dermatologue procède à l'excision par biopsie d'un tissu intra-lésionnel considéré comme pathologique, puis ce même tissu est préparé afin d'être observé sous microscope. D'une part, cette procédure est relativement longue à réaliser et d'autre part représente une charge financière importante une fois rapportée au coût par examen. En addition, il s'agit d'une procédure incommode pour le patient, pouvant échouer si le prélèvement est effectué en dehors d'un foyer mélanocytaire [111]. Des exemples histologiques de pathologie de **LM** peuvent être observés sur la Figure 1.8 - en d) et e).

Cette difficulté à distinguer ces pathologies et les limitations induites par l'histopathologie conduisent à l'utilisation de dispositifs d'imagerie *in-vivo* prévus pour ce domaine d'application. Ainsi, parmi les plus utilisés dans ce domaine peuvent être cités : (i) la dermatoscopie, (ii) et la microscopie confocale par réflectance. Des exemples d'images de pathologies de **LM** issues de ces dispositifs peuvent être observés sur la Figure 1.8 - en a), b) et c). Le chapitre suivant s'oriente en ce sens en apportant des bases quant aux propriétés physiques de la peau et à la manière de l'observer.

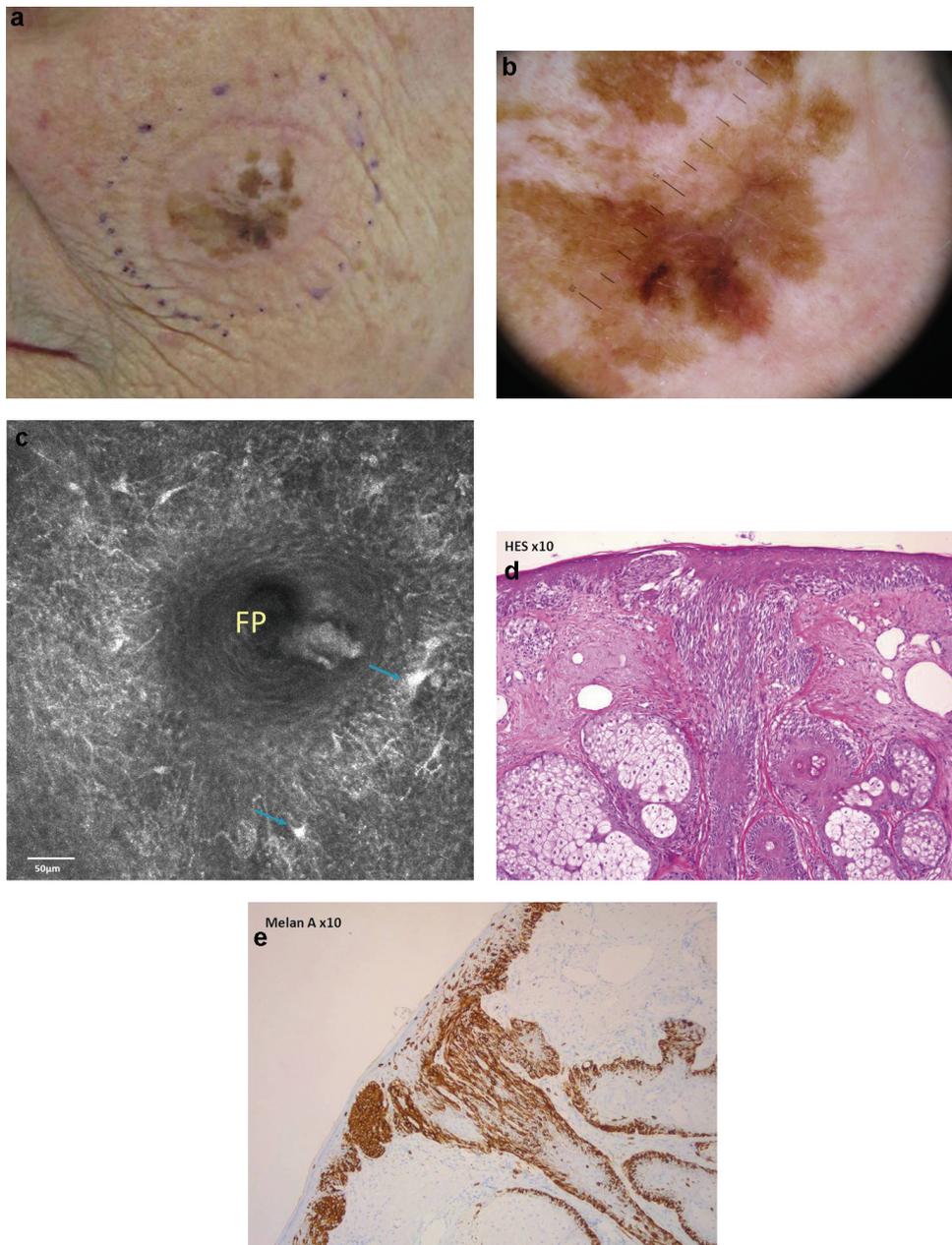


Figure 1.8 : Exemple d'une pathologie de **LM** perçue à l'aide de différents dispositifs médicaux. En a), un aperçu par photographie clinique ; En b), un aperçu par dermatoscopie ; En c), un aperçu par Microscopie confocale par réflectance. En d) et e), les images fournissent un aperçu par histologie.

DES PROPRIÉTÉS PHYSIQUES AUX DONNÉES CLINIQUES

Lors du précédent chapitre ont été présentés la composition et le fonctionnement de la peau ainsi que ses pathologies. De plus, une attention particulière a été portée sur l'une de ces pathologies : le **Lentigo Maligna** mais également sa forme avancée le **Lentigo Maligna Melanoma** et ses caractéristiques permettant son diagnostic.

Dans un premier temps, ce nouveau chapitre permet d'appréhender certaines propriétés physiques de la peau, notamment ses mécanismes d'interaction avec la lumière. Dans un second temps, ce chapitre évoque les principaux dispositifs optiques actuels et le type d'information mis à disposition par chacun d'entre eux pour servir le diagnostic du spécialiste. Par ailleurs, une sous-section est dédiée aux dispositifs basés sur des mesures physiques alternatives telles que l'écho d'ultrasons par exemple.

Sommaire

2.1	Propriétés de la peau	22
2.2	Modalités d'imagerie non-invasives	25
2.2.1	Modalités par mesure optique	25
2.2.2	Modalités par mesure non-optique	34

2.1/ PROPRIÉTÉS DE LA PEAU

La perception de la peau est façonnée par divers phénomènes physiques d'interaction entre la lumière et ses tissus. Ces derniers influent essentiellement de : (i) par leur structure, c'est-à-dire par la disposition de la matière constituant la peau, (ii) par leurs composants, c'est-à-dire par la matière elle-même. Cette section est abordée de manière descendante, en déroulant le processus d'interaction d'un photon qui entre en contact avec la peau. Ce processus est schématisé de manière macroscopique sur la Figure 2.1. La **réflexion spéculaire** est le premier phénomène de structure pouvant se produire à l'interface entre l'air et la peau. Par ce phénomène, la lumière incidente est réfléchiée avec un angle identique à celui de la surface considérée. Ce phénomène est accentué par une peau lisse et uniforme [218]. Cette lumière est ainsi identique en matière de propriété à celle de sa source d'émission, et peut être la plupart du temps considérée comme un parasite car elle ne contient aucune information propre aux composants du milieu observé en apparaissant souvent de manière saturée sur un capteur. La **réflexion diffuse**, est un second phénomène par lequel la lumière est réfléchiée dans une multitude de directions. Ce phénomène peut avoir pour origine de multiples réflexions liées à la structure du composé, ou être issu de diffractions liées au changement de propriétés du milieu [218] (coefficient de réfraction proche de celui de l'eau pour la peau). Enfin, une partie de cette lumière est perdue ou transformée selon la longueur d'onde de celle-ci : (i) son énergie peut-être dissipée sous forme de chaleur, on parle alors de phénomène **d'absorption**, (ii) ou être ré-émise avec le plus souvent une longueur d'onde différente, on parle alors de phénomène de **photoluminescence** regroupant la **fluorescence** et la **phosphorescence** [101].

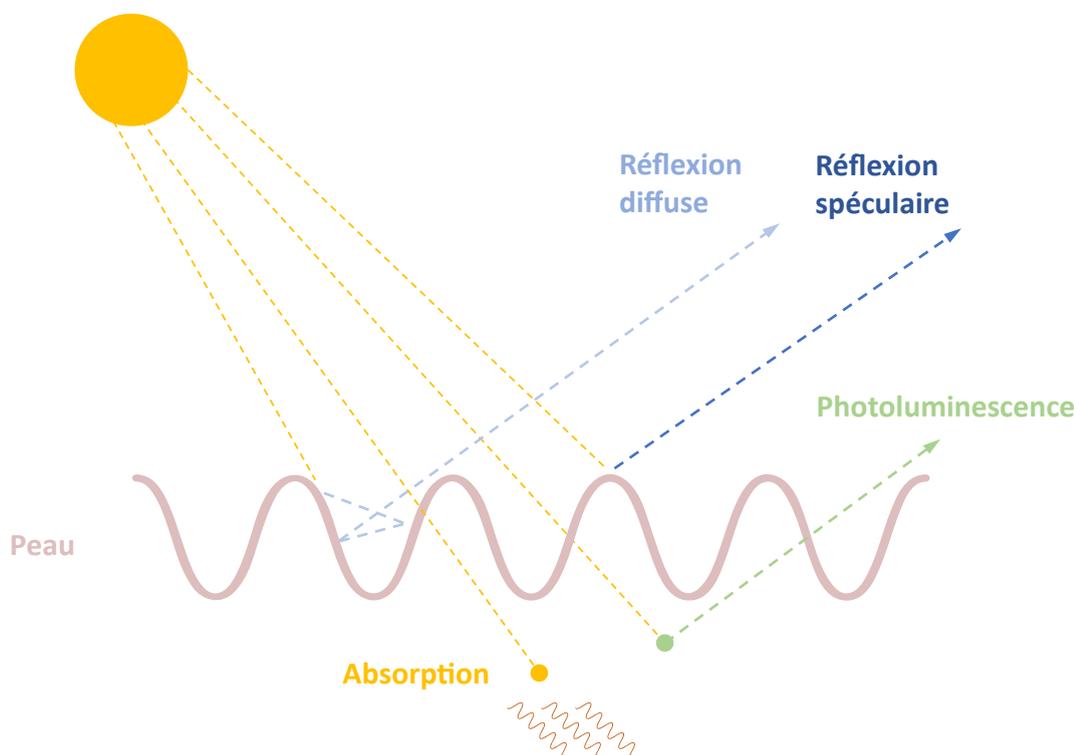


Figure 2.1 : Schéma représentant les principaux mécanismes d'interaction qui peuvent survenir suite à l'interaction de la lumière avec la peau. Les modalités optiques tentent d'exploiter ces divers phénomènes pour extraire une information sur les tissus analysés.

Dans cette configuration, le choix de la source d'émission de lumière est l'un des points importants. En effet, comme démontré dans un travail précédent employant le principe de la **Diode Électro-Luminescente, DEL** (*alt. : Light-Emitting Diode, LED*) en dermatologie [11], **la longueur d'onde d'émission** est l'un des paramètres d'influence possible quant à la profondeur des structures à observer dans la peau. Ainsi, des longueurs d'onde à l'entrée du spectre visible humain (≈ 380 nm) ne parviennent pas à se frayer un chemin en profondeur et ne frôlent que les couches supérieures de l'épiderme, tandis que des longueurs d'onde plus élevées de ce même spectre (≈ 780 nm) tendent à atteindre l'hypoderme et ses vaisseaux sanguins. Une synthèse de la pénétration de la lumière en fonction de la longueur d'onde est visible sur la Figure 2.2.

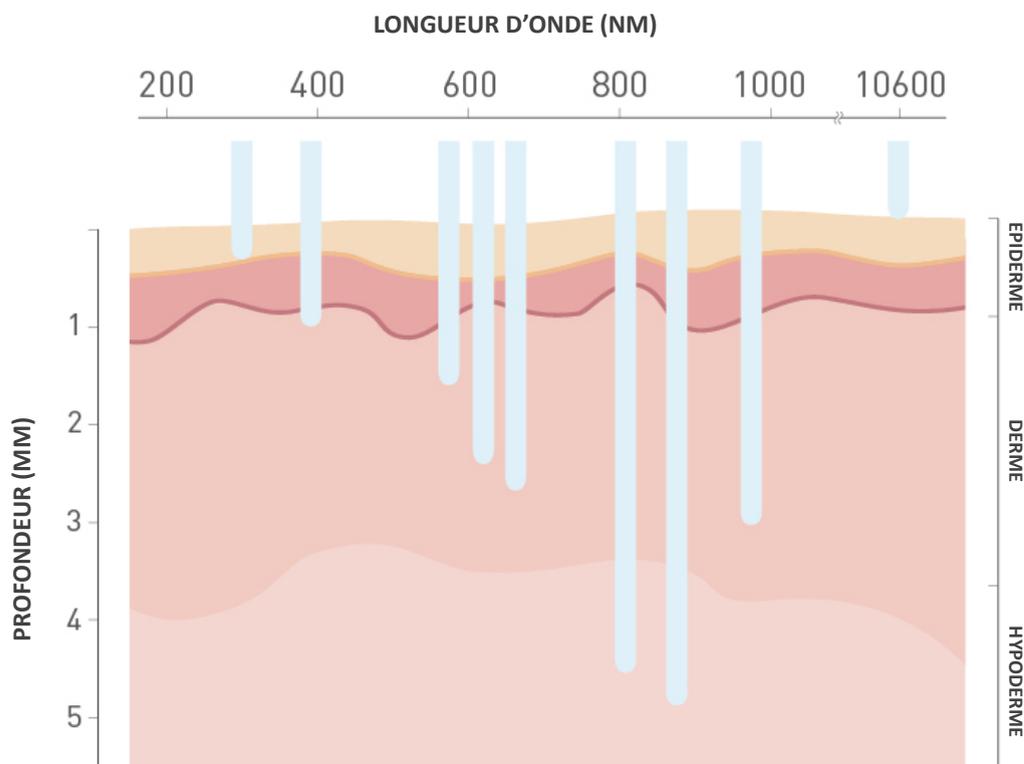


Figure 2.2 : Schéma représentatif de la profondeur de pénétration (en mm) de la peau en fonction de la longueur d'onde de la lumière [11]. Il s'agit d'un ordre de grandeur moyen constaté chez l'être humain, les propriétés de la peau variant d'un emplacement à un autre. Aucune information relative aux types de peaux étudiés, selon l'échelle de Fitzpatrick par exemple, n'a été mentionnée.

Outre la question de la profondeur, le choix de la longueur d'onde permet d'**interagir avec différents types de chromophores** et permet de renseigner localement sur des propriétés de la peau. Certains travaux focalisés sur les systèmes lasers et d'émission en général [182] ont permis de mettre en avant le rapport entre la longueur d'onde et des composants majeurs de la peau tels que **l'eau, l'hémoglobine (oxygénée ou non) ou la mélanine**. Pour illustrer ce propos, la Figure 2.3 synthétise ce rapport entre la longueur d'onde d'émission et l'absorption associée aux principaux composants chromophores de la peau dont les quatre mentionnés. La quantification de ces composants peut présenter un intérêt dans la mesure où certaines pathologies peuvent altérer le métabolisme et donc modifier fondamentalement le ratio de chromophores présents, pouvant par exemple amener à une hypervascularisation. Certains travaux tentent ainsi de revenir à des fonctions de métabolisme [91].

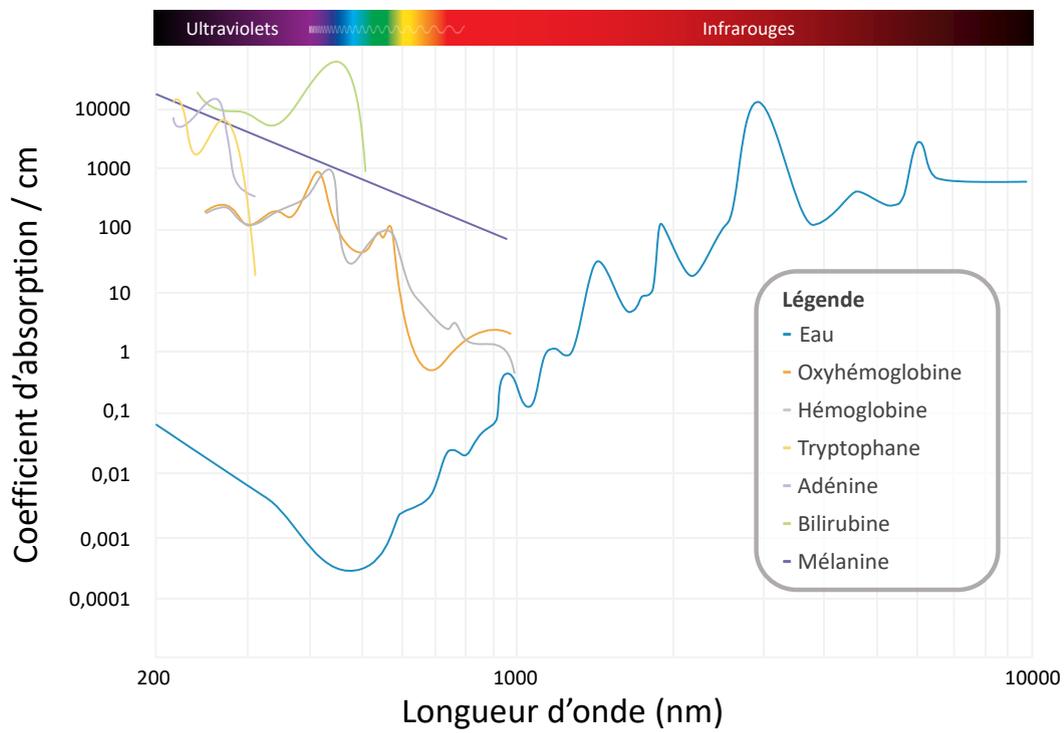


Figure 2.3 : Représentation des coefficients d'absorption en fonction de la longueur d'onde pour les principaux composants chromophores présents dans la peau [162].

2.2/ MODALITÉS D'IMAGERIE NON-INVASIVES

Pour le diagnostic et aussi afin de mieux suivre, voire de conserver une trace de leurs patients, les dermatologues disposent de nombreux outils. Ils peuvent entre autres dans leur pratique clinique se référer à divers types d'examens, distingués dans ce travail sous deux catégories :

- d'une part **les examens non-invasifs**, c'est-à-dire ne provoquant pas de destruction des tissus du corps humain. Ces dispositifs sont capables de restituer une image différente des tissus selon la nature du phénomène physique observé et ne provoquent en général pas d'inconfort particulier pour le patient. Ils peuvent représenter un investissement initial important mais s'amortissent assez bien sur le long terme.
- d'autre part **les examens invasifs**, c'est-à-dire provoquant une effraction des tissus. L'histopathologie est le principal mode d'examen invasif et représente à ce jour le *gold standard* de ce domaine. Cet examen se réalise par biopsie, c'est-à-dire par prélèvement de tissus puis par une analyse histologique de ceux-ci. Cette technique peut être inconfortable pour le patient et requiert plus de temps qu'un examen non-invasif et est généralement plus coûteuse à l'unité.

Après avoir présenté de manière non exhaustive les principales pathologies de la peau et les principes d'interaction majeurs de la lumière avec cet organe, il convient désormais d'aborder les diverses méthodes permettant une observation non-invasive. Ces modalités d'imagerie franchissent en majorité le cap du numérique permettant un suivi plus aisé, voire un échange plus rapide entre professionnels, et l'utilisation de traitements automatiques. À nouveau, deux catégories distinctes sont utilisées pour séparer ces techniques avec d'une part **les techniques de mesure optique** et d'autre part **les techniques de mesure non-optique** basées sur divers phénomènes physiques quantifiables.

2.2.1/ MODALITÉS PAR MESURE OPTIQUE

Cette partie décrit les principaux appareils existants permettant l'observation de la peau. Évidemment, ces modalités sont également transposables à d'autres domaines d'application que celui de la dermatologie. Les dispositifs méconnus ou expérimentaux encore loin d'une application clinique ne sont pas présentés dans cette sous-section. La Table 2.1 met en lumière ces dispositifs et leurs caractéristiques générales à but indicatif.

Technologie	Effet	Résolution spatiale	Profondeur
Imagerie clinique	Réflectance	Lentille dépendant	0,1 mm–1 mm
Dermatoscopie	Réflectance	Lentille dépendant	0,1 mm–1 mm
Microscopie confocale par réflectance	Réflectance	$\mu\text{m}/\text{px}$	< 200 μm
Tomographie en cohérence optique	Réflectance	$\mu\text{m}/\text{px}$	1 mm–2 mm
Imagerie multispectrale	Réflectance	Lentille dépendant	0,1 mm–1 mm
Spectroscopie à réflexion	Réflectance	Diamètre de la fibre	0,1 mm–1 mm
Microscopie confocale Raman	Raman	$\mu\text{m}/\text{px}$	150 μm

Table 2.1 : Table des principales modalités de mesure optique utilisées à but clinique ou à un stade de recherche clinique [101].

2.2.1.1/ IMAGERIE CLINIQUE

L'examen de dépistage classique est associé en dermatologie à une inspection à l'œil nu exercée par une personne compétente ou sensibilisée à la détection de pathologies. L'**imagerie clinique** fait appel à un dispositif de photographie non propre à la dermatologie. Par ailleurs, elle constitue historiquement la première modalité de cette discipline mais également la moins onéreuse car emploie des caméras non spécifiques à la discipline.

Avant l'avènement de l'informatique, cette modalité se basait sur des supports de type argentique. L'évolution de ce matériel vers le numérique, et l'arrivée de systèmes de type **Picture Archiving and Communication System, PACS** (*alt. : Système d'Archivage et de Transmission d'Images, SATI*) ont conjointement motivé une transition vers des données dématérialisées.

Ce type d'imagerie donne à l'observateur un point de vue subjectif, similaire à une observation à l'œil nu, utile dans le cadre d'un diagnostic à distance ou dans le suivi à long terme d'un patient. Néanmoins, les principaux points faibles sont **le manque de standard concernant le format d'image** et **le manque de protocole concernant l'acquisition des images** : pas de spécifications d'éclairage, ni même de spécifications sur le point de vue à adopter au moment de la prise d'image.

Ces dernières années ont vu l'émergence de nouveaux types de dispositifs qui étendent la photographie clinique au corps entier. Ces appareils permettent une traçabilité et un suivi des lésions apparentes sur le corps d'un patient. En effet, l'un des meilleurs critères du suivi des lésions dans le dépistage de pathologies telles que le mélanome, reste le suivi de leur évolution au cours du temps. Cet élément a notamment été mis en avant en 2004, et est encore considéré au moment de cette rédaction comme l'un des critères majeurs [2, 73]. Différents industriels commercialisent des dispositifs de plus en plus performants dont FotoFinder et son système ATBM mais également la société Pixience avec le Body-Mapper® permettant ce suivi ainsi qu'un contrôle de l'éclairage lors de la prise de vue. Une interface guide le médecin lors de la prise d'images et permet d'associer à un modèle humain les diverses parties du corps (voir la Figure 2.4).

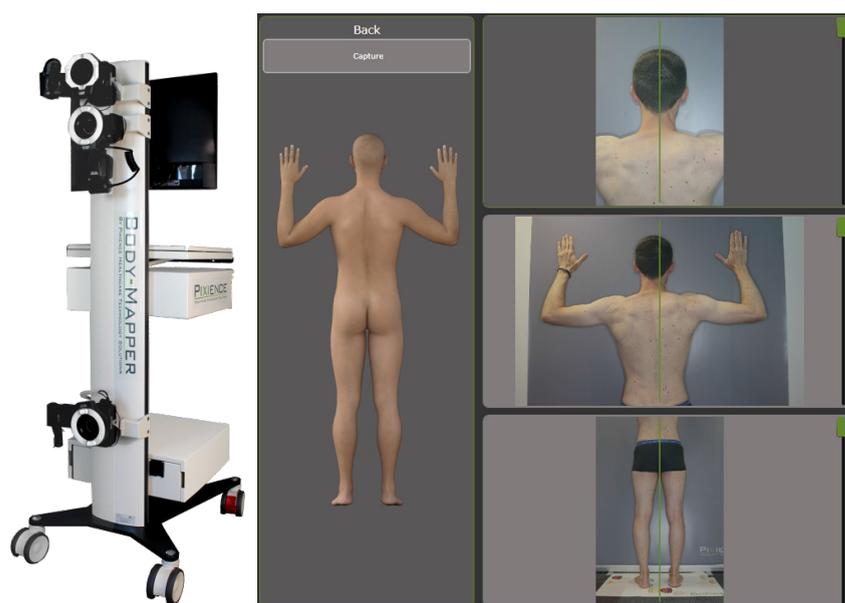


Figure 2.4 : Exemple du Body-Mapper, un appareil de photographie clinique corps entier, proposé par Pixience. À gauche, l'appareil d'acquisition ; À droite, son interface logicielle permettant d'acquérir les diverses parties du corps.

2.2.1.2/ DERMATOSCOPIE

Également appelé *dermoscopie*, *microscopie en épiluminescence* ou encore *microscopie de surface*, ce dispositif permet d'observer les lésions cutanées. Le terme de **dermatoscopie** est préféré aux autres pour la suite de ce manuscrit. Cette technique d'imagerie est partiellement attribuée à Johan Christophorus Kohlhaus dont les travaux menés au xvii^e siècle, auraient grandement contribué à initier cette modalité. Les premiers dispositifs ont émergé en 1971 [128], et ce sont les années 1980 qui ont contribué à son essor chez un grand nombre de praticiens.

Cet outil possède comme première caractéristique, **de réduire les réflexions de lumière et contribue à rendre le stratum corneum translucide** [97], permettant ainsi au praticien de visualiser les couches sous-jacentes : épiderme, **JDE** ou encore le derme papillaire non visible à l'œil nu. Cette réflexion de lumière était initialement supprimée par l'utilisation d'un fluide (eau, gel, ...) entre la peau et la lentille du dispositif (dermatoscope non polarisé). De récentes améliorations ont permis **de rendre l'utilisation d'un fluide obsolète par l'utilisation de lumière polarisée** (dermatoscope polarisé). Un premier filtre polarisé orienté dans une direction est disposé juste après la source de lumière, puis un second filtre polarisé perpendiculairement au premier filtre est mis en place avant la fenêtre d'observation (voir Figure 2.5). Ce dispositif permet ainsi de s'affranchir de la lumière spéculaire directement issue de la source de lumière qui parasite l'observation [53].

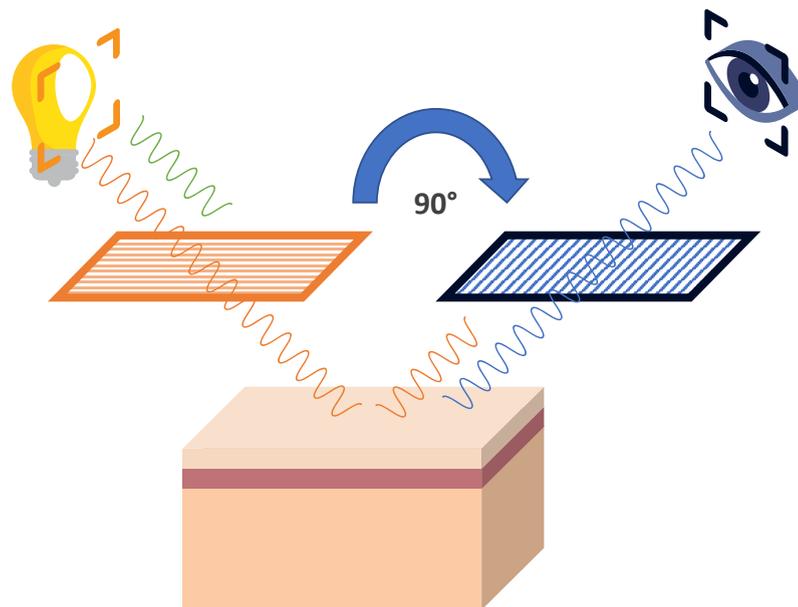


Figure 2.5 : Principe de la dermatoscopie à lumière polarisée. Un premier filtre polarisé est disposé après la source de lumière, puis un second filtre dont la polarisation est perpendiculaire au premier est disposé juste avant la fenêtre d'observation [177].

Une seconde caractéristique non négligeable de ce dispositif, est la mise à disposition d'un zoom permettant de grossir en lumière blanche jusqu'à 400 fois selon la complexité des modèles proposés ¹. Cette dernière caractéristique octroie au praticien la possibilité d'appréhender au mieux la structure de la peau dans ses moindres détails [53].

¹Source : FotoFinder medicam 1000s.

De plus, son faible coût d'achat, sa rapidité de diagnostic et son efficacité face au seul œil humain contribuent largement à sa démocratisation dans la profession [109]. Bien que cette modalité soit utilisée majoritairement dans le dépistage de *lésions pigmentées*, son efficacité semble avérée dans le cas de *lésions non pigmentées* telles que le **CBC** et le **CSC** [109].

Ces dispositifs tendent à s'adapter de plus en plus aux besoins de leur marché, en proposant par exemple une caméra haute résolution et permettant une utilisation polyvalente de l'appareil avec un connecteur générique comme le propose la société *Pixience* (voir la Figure 2.6). L'utilisation de cet appareil dans certaines études, a permis la mise en avant de pathologie de la peau chez la souris [155] et chez l'homme [41]. Certaines sociétés comme *DermLite*, proposent une solution plus nomade et adaptable aux principaux dispositifs de photographie du marché. La solution *DL3N* est visible sur la Figure 2.7.



Figure 2.6 : Exemple du dispositif de dermatoscopie *C-Cube* proposé par la société *Pixience*.



Figure 2.7 : Exemple du dispositif de dermatoscopie *DL3N* proposé par la société *DermLite*. À gauche, le dispositif seul ; Au milieu, ce même dispositif adapté à un appareil photo standard ; À droite, le dispositif adapté à une caméra de téléphone portable.

2.2.1.3/ MICROSCOPIE CONFOCALE PAR RÉFLECTANCE

La **Microscopie Confocale par Réflectance, MCR** (*alt. : Reflectance Confocal Microscopy, RCM*) est une technique d'imagerie décrite durant les années 1950 par Marvin Minsky [170]. Les avancées réalisées durant les années 1990 ont permis de réduire considérablement la taille de cet appareil et de faciliter ainsi son utilisation dans divers domaines. Cette technique de plus en plus répandue dans les services de dermatologie a également connu un regain de notoriété dans les journaux scientifiques : une recherche à l'aide des mots-clés « reflectance confocal microscopy skin » apporte une dizaine de publications vers les années 2000 contre une moyenne de 1500 chaque année depuis 2010².

Cette technique emploie l'utilisation d'un *sténopé* situé devant le capteur et permet **de conserver uniquement les photons issus d'un plan focal**, illustré en Figure 2.8. Ce principe octroie la possibilité **d'ajuster la profondeur** de ces *plans focaux* ou *plans de coupes*, permettant l'acquisition d'images à une même profondeur ou l'acquisition de « piles » d'images dont la profondeur varie [170]. En dermatologie, les dispositifs se basent sur des longueurs d'onde de 830 nm non-invasives pour la peau et les yeux, mais limitent la profondeur du dispositif à 200 μm –300 μm .

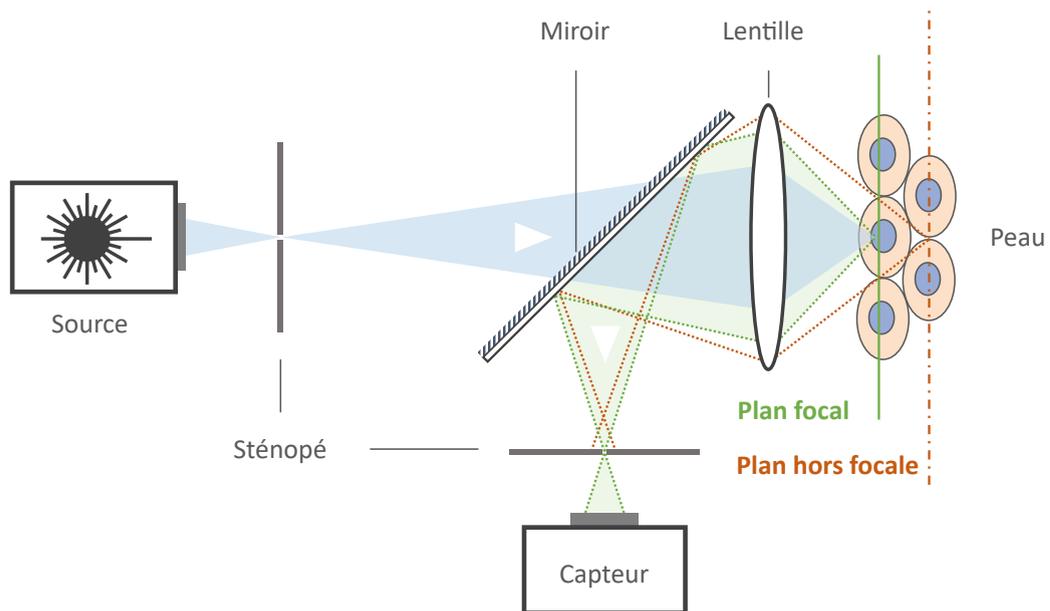


Figure 2.8 : Schéma de fonctionnement du **MCR** par Marvin Minsky. Une source de lumière est émise et focalisée en un point spécifique de la peau puis la lumière est réfléchi et reçue par une caméra au travers d'un sténopé (trou de faible diamètre). La lumière réfléchi en dehors du plan focal défini par l'utilisateur est ainsi arrêtée par ce sténopé.

Enfin, à la manière du dermatoscope, un gel à base d'eau de réfraction proche de celle de l'épiderme est utilisé entre la lentille du dispositif et la peau afin de limiter la perte de photons et de permettre l'obtention d'images du derme.

²Source : Google Scholar.

2.2.1.4/ TOMOGRAPHIE EN COHÉRENCE OPTIQUE

Initialement conçue pour le domaine de l'ophtalmologie, la **Tomographie en Cohérence Optique**, **TCO** (*alt. : Optical Coherence Tomography, OCT*) est une technique récente qui se base sur le principe d'interférométrie et qui tend à se démocratiser à de nombreux domaines d'intervention dont celui de la dermatologie. Schématisé sur la Figure 2.9, ce principe consiste à utiliser **une source de lumière commune divisée en deux faisceaux**, le **premier servant de référence** et le **second servant à l'analyse de l'échantillon**. Ainsi, le faisceau de référence va transiter via un chemin contenant un miroir ajustable permettant de modifier sa distance parcourue. En modifiant cette distance, la profondeur désirée va pouvoir être ajustée au sein de l'échantillon à analyser. En effet, lors de la superposition des deux faisceaux à l'entrée du récepteur, une *interférence constructive* est observée lorsque l'information est en phase, cette information est alors **amplifiée**, tandis qu'une *interférence destructive* est observée lorsque l'information est en déphasage, cette information est alors **atténuée**.

Initialement, ce principe a été exploité pour des dispositifs dits d'interférométrie à faible cohérence temporelle et fournissait une information à une dimension. La **TCO** est une extension de ce principe à deux dimensions, permettant la caractérisation d'une *tranche* de tissu. Cette modalité d'imagerie propose ainsi une information temps réel en profondeur (de l'ordre du millimètre), avec une résolution proche du micromètre. Par ailleurs, il est possible d'obtenir une reconstruction de la peau et de ses structures, par l'acquisition de plusieurs plans de coupe.

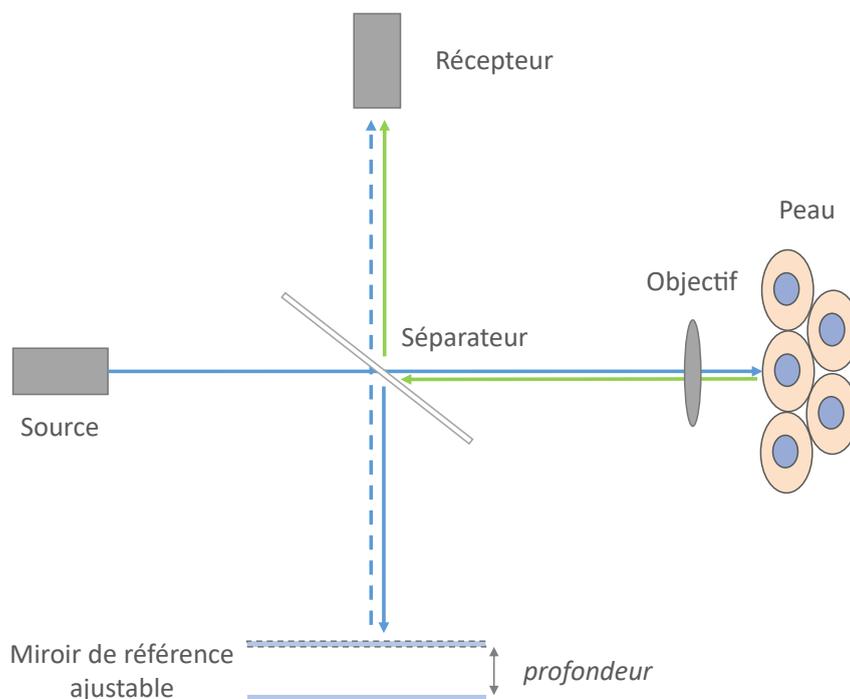


Figure 2.9 : Schéma de fonctionnement de la **TCO**. Une source de lumière génère un unique faisceau divisé en deux faisceaux identiques. L'un de ces faisceaux résultants sert de signal de référence (en bleu), tandis que le second sert à l'analyse de l'échantillon (en vert). Par la suite, ces deux faisceaux sont recombinés à l'entrée du récepteur.

2.2.1.5/ SPECTROSCOPIE PAR RÉFLECTANCE

La spectroscopie de réflectance fournit une information de spectre à un même instant, c'est-à-dire contenant plusieurs valeurs numériques, chacune propre à une longueur d'onde spécifique. L'information fournie est propre à un unique point de l'espace et est obtenue en temps réel. Ainsi, cette modalité ne constitue pas une technique d'imagerie au sens matriciel à deux dimensions.

Pour cela, une source de lumière uniforme et stable dans le temps est utilisée, mais également couvrant l'ensemble du spectre désiré. Cette lumière est ensuite canalisée au sein d'une ou plusieurs fibres optiques afin d'être conduite jusqu'à l'échantillon supposé. Le phénomène de réflectance de l'échantillon est ensuite capté à l'aide d'une ou plusieurs fibres optiques jusqu'à un spectromètre qui va dissocier le signal (phénomène de diffraction), afin de cibler de multiples cellules présentes sur un capteur numérique [138, 130]³.

Ces capteurs font l'objet d'une procédure de calibration à l'aide de *pantones* ou nuanciers dont les propriétés visuelles sont établies, comme le propose par exemple le ColorChecker®⁴. En effet, il est nécessaire d'empêcher une saturation du signal conduisant à une perte d'information tout en exploitant au mieux la quantification octroyée par le capteur. Sa nécessité est le fruit des multiples conditions d'éclairage possibles de l'environnement mais également de la variété des sources de lumière pouvant être employées avec ce type de dispositif. Cette étape de calibration permet de rectifier et compenser le signal obtenu par le capteur, et d'obtenir une version *idéale* de celui-ci dont l'information est maximisée.

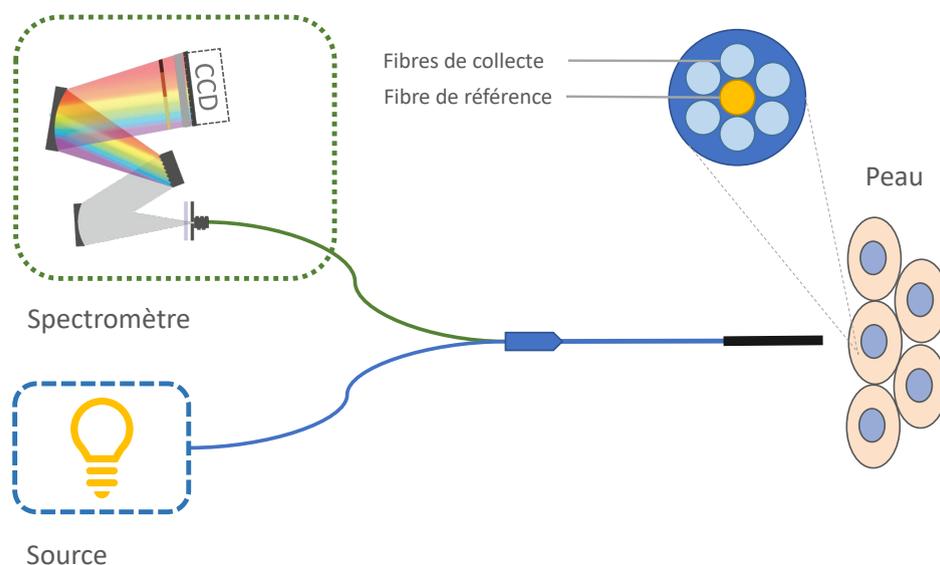


Figure 2.10 : Schéma représentant le fonctionnement d'un spectroscope par réflectance. Une lumière de référence est générée puis acheminée jusqu'à la zone cible. Cette lumière transite le long d'une ou plusieurs fibres optiques. Puis le phénomène de réflectance est capté et envoyé au spectromètre par le biais d'une ou plusieurs fibres. Le signal reçu est ensuite décomposé puis envoyé à la surface d'un capteur numérique.

³Source : Spectromètres AvaSpec®, Avantes, Pays-Bas.

⁴Source : X-Rite, ColorChecker®, Michigan, États-Unis.

2.2.1.6/ IMAGERIE MULTISPECTRALE

L'imagerie multispectrale est l'extension spatiale de la spectroscopie et l'extension spectrale de l'imagerie clinique. Par ce principe d'imagerie, les pixels contenus dans une image ne se résument pas à une valeur pour les images en intensités de gris ou à trois valeurs pour des images RGB, mais à une multitude de valeurs décrivant chacune une longueur d'onde. Ce nombre de valeurs spectrales est ainsi dépendant de la précision recherchée dans un domaine d'application particulier. Les temps d'acquisition entre les diverses valeurs tendent à se réduire fortement avec les évolutions technologiques. Différents principes sont mis en oeuvre pour parvenir à l'obtention de ces images, visible sur la Figure 2.11. Ce travail retient les principes suivants :

- par modification **de la source émettrice de lumière**, soit en utilisant des **DEL** programmables, des lasers ou encore des combinaisons de filtres permettant de modifier les propriétés de la lumière,
- par filtrage **de la lumière reçue**, bien souvent à l'aide de combinaisons de filtres permettant de choisir les longueurs d'onde désirées. Des capteurs multicouches ont émergé ces dernières années et permettent une acquisition quasiment simultanée de l'information sur le capteur en lui-même ⁵.

Ce principe permet de proposer une information plus riche que celle transmise par les capteurs habituels du marché. Ces appareils permettent de mettre cette information en relation avec des modèles d'absorption et de remonter à une estimation de la composition et de la structure du tissu.

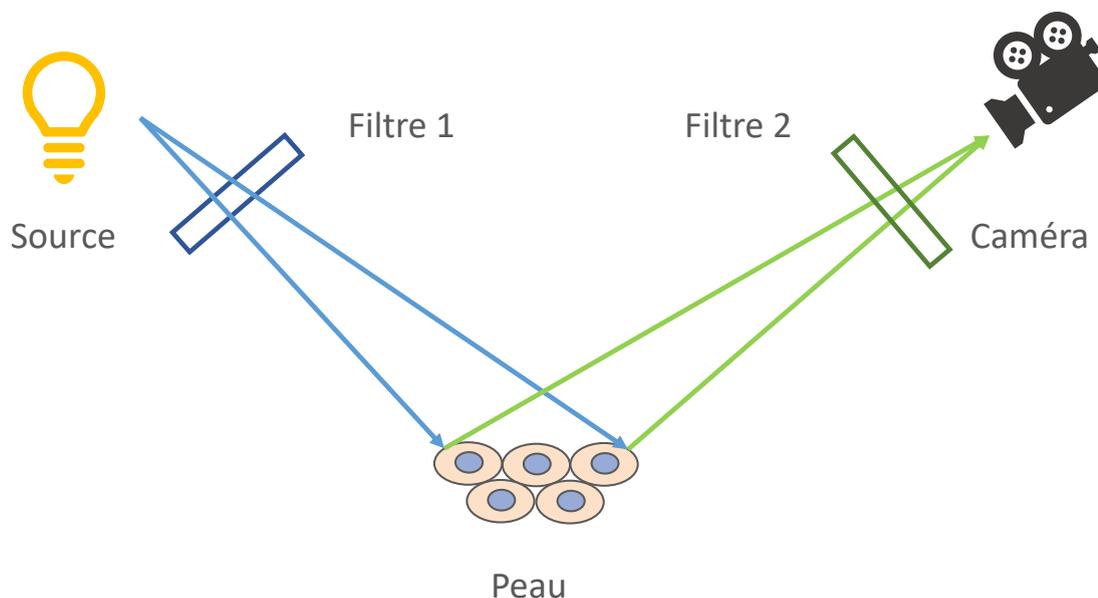


Figure 2.11 : Schéma de fonctionnement d'une caméra multispectrale. Ces dispositifs peuvent adapter la longueur d'onde de la source à partir d'éclairage **DEL**, de lasers ou de filtres (Filtre 1), ou capter différentes longueurs d'onde à partir de la lumière réémise (Filtre 2).

⁵Source : Foveon X3®, Foveon, Santa Clara, Californie.

2.2.1.7/ MICROSCOPIE CONFOCALE RAMAN

Ce type de dispositif porte le nom de l'effet Raman, lui-même inspiré par le nom du physicien Chandrashekhara Venkata Râman à l'origine de cette découverte. Son principe consiste à analyser un décalage de fréquence entre la lumière incidente et la lumière diffusée. Ainsi, ce décalage est propre à chaque matériau et permet de retrouver ce dernier. Cet effet est par son principe non destructif, et a l'avantage de ne pas être dépendant de la longueur d'onde d'excitation. Contrairement à l'effet de fluorescence pour lequel la lumière doit posséder une longueur d'onde spécifique afin d'entrer en *résonance* avec la matière traversée, l'effet Raman s'obtient quelle que soit la longueur d'onde de la lumière. Les différentes interactions et leurs différences sont schématisées sur la Figure 2.12.

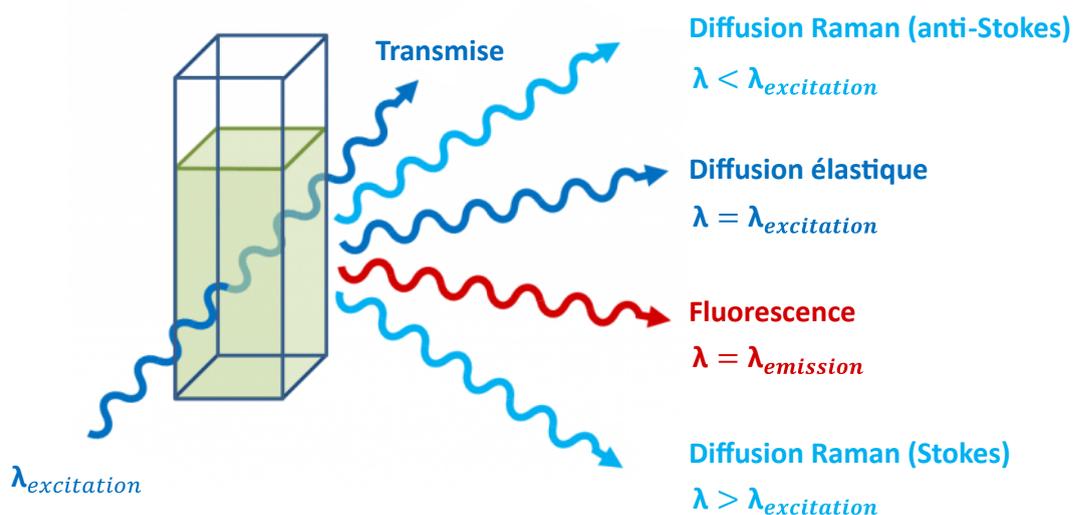


Figure 2.12 : Schéma reprenant les divers effets résultant de l'interaction entre la lumière et la matière⁶. La diffusion Raman se distingue par un décalage de longueur d'onde supérieur (Stokes) ou inférieur (anti-Stokes).

Cet effet a été employé en spectroscopie assez tôt [64] et s'est étendu progressivement à la microscopie confocale [24] permettant d'obtenir non plus une information relative à un point unique, mais d'obtenir des images à deux dimensions, avec une profondeur ajustable par modification du plan focal.

Ses applications dans le milieu de la dermatologie sont variées. En effet, de nombreux travaux ont porté sur la détection des lésions cancéreuses mélanocytaires, dont essentiellement le mélanome avec une sensibilité supérieure à 90 % [125, 168]. Les lésions cancéreuses non mélanocytaires ont également été abordées dont essentiellement les pathologies de CBC et de CSC avec des précisions de détection élevées, supérieures à 90 % [117, 172].

⁶Source : Edinburgh Instruments, Livingston, Royaume-Uni.

2.2.2/ MODALITÉS PAR MESURE NON-OPTIQUE

Cette seconde sous-section se dédie aux modalités d'imagerie offrant une alternative aux techniques optiques employées dans le milieu de la dermatologie. Dans un premier temps, le fonctionnement de l'**Imagerie par Résonance Magnétique, IRM** (*alt. : Magnetic Resonance Imaging, MRI*) est décrit puis dans un second temps celui de l'échographie haute fréquence.

2.2.2.1/ IMAGERIE PAR RÉSONANCE MAGNÉTIQUE

Le principe global de fonctionnement de l'**IRM** est basé sur l'interaction et l'observation des moments magnétiques des noyaux d'hydrogène. À leur état initial, les noyaux d'hydrogène possèdent un moment magnétique non contraint et donc désordonné (la somme de ces moments magnétiques à l'échelle macroscopique est approximativement nulle). Ainsi, la première étape consiste à contraindre le moment magnétique des noyaux d'hydrogène dans une même direction (mais pas forcément un même sens) par un champ magnétique constant, on parle de précession. Dans une seconde étape, des radio-fréquences sont émises pour contraindre, de manière homogène, ces moments magnétiques dans une nouvelle direction (généralement perpendiculairement à ce champ) et ce par excitation. Enfin, l'émission de radio-fréquences est stoppée, puis le temps de retour à l'équilibre est mesuré (temps de relaxation) à l'aide d'une antenne de réception. Ce fonctionnement est résumé à l'aide du schéma disponible sur la Figure 2.13.

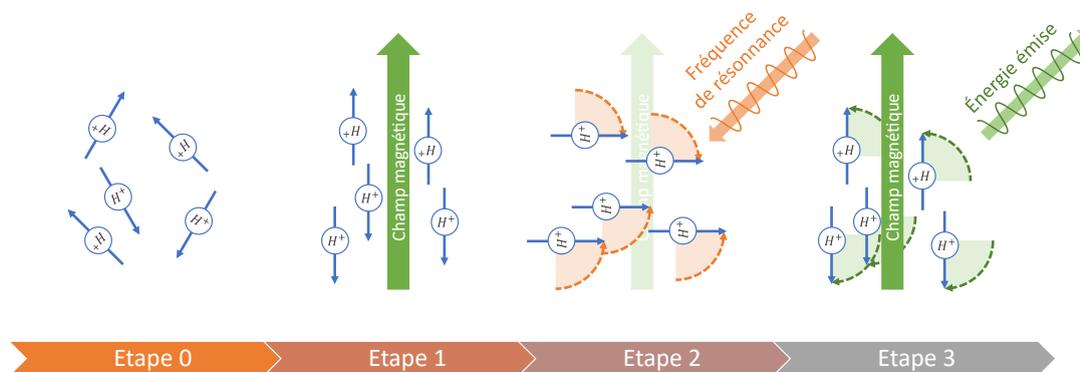


Figure 2.13 : Schéma du principe de fonctionnement de l'**IRM** en quatre étapes. Le champ magnétique constant permet de contraindre les noyaux d'hydrogène, puis l'excitation par radio-fréquence permet d'observer par la suite leur temps de relaxation.

En ce qui concerne l'utilisation de l'**IRM** en dermatologie, le champ d'application semble assez restreint, malgré des études positives concernant son utilisation dans le cadre d'actions de dépistage et de son utilisation à but pré-opératoire du mélanome. De plus, l'utilisation d'agents de contraste comme le gadolinium semble être opportun pour le dépistage de lésions de la peau [221]. Une étude plus récente liée à la dermatologie a été menée sur les mécanismes nerveux de récompense associés aux démangeaisons, mais implique l'utilisation de l'**IRM** fonctionnelle sur la zone du cerveau et non de la peau en elle-même [136]. Néanmoins, l'une des principales contraintes évoquée concerne l'acquisition d'échantillons de faible taille sur des machines conventionnelles rendant le ratio signal sur bruit important et les temps d'acquisition pour compenser ce dernier inexploitable [74]. De plus, les forts coûts associés à ces appareils et leur utilisation contraignante en font des outils difficilement applicables au domaine de la dermatologie.

2.2.2.2/ ÉCHOGRAPHIE HAUTE FRÉQUENCE

Les dispositifs d'échographie sont des appareils d'imagerie basés sur l'utilisation d'ondes acoustiques. Pour cela, une sonde composée de transducteurs électro-acoustiques est employée pour l'émission mais également la réception de ces ondes, par utilisation de l'effet piézoélectrique inverse et direct. Il existe de multiples sondes pouvant émettre à divers intervalles de fréquences et permettant d'atteindre diverses profondeurs. Ces ondes sont ainsi émises par la sonde (effet piézoélectrique inverse) et se propagent au sein des tissus du corps humain, avant d'être partiellement renvoyées à chaque changement de l'impédance de ces tissus. Les ondes renvoyées sont ainsi réceptionnées par la sonde émettrice (effet piézoélectrique direct) et pour former un signal.

Comme énoncé, différentes fréquences sont employées selon la profondeur recherchée, de 2 MHz–10 MHz pour des structures assez lointaines allant de l'abdomen aux artères, et 10 MHz–70 MHz pour des structures plus proches et fines telles que la peau et l'oeil. Ainsi, les travaux de la peau par échographie emploient le terme d'échographie haute fréquence.

Les travaux de recherche mêlant dermatologie et échographie avant les années 2000 ont montré une efficacité de la technique afin de suivre l'évolution de mélanomes [22]. Des travaux plus récents ont permis d'étendre la technique à d'autres pathologies cancéreuses telles que les CBC [49], mais également des CSC [25]. Enfin, cette technique aurait un potentiel pour le suivi de pathologies plus bénignes telles que l'eczéma ou l'arthrose psoriasique [15].



Figure 2.14 : Exemple de la sonde à ultrasons Butterfly iQ⁷, proposant une solution nomade et connectée.

⁷Source : Butterfly Network, New York, États-Unis.

INTELLIGENCE ARTIFICIELLE ET NOTIONS COMPLÉMENTAIRES

Lors du précédent chapitre, les interactions entre la peau et la lumière ont été abordées, ainsi que les dispositifs majeurs permettant l'acquisition d'informations de la peau.

Actuellement, l'informatique est une ressource omniprésente dans de nombreux domaines d'application. Ses facultés à divertir ou encore à guider ses usagers dans leurs choix quotidiens personnels et professionnels en font un précieux allié. De nouvelles pratiques font leurs apparitions, intégrant la machine au cœur des processus décisionnels. Les disciplines médicales évoluent largement et l'ordinateur permet de répondre aux besoins croissants de précision, de reproductibilité intra et inter opérateur, et permet un gain de temps au sein des milieux cliniques. Dans un cadre bien défini, il s'agit également d'un moyen de formaliser la connaissance au travers d'un outil pouvant être utilisé par des non-initiés.

Cette nouvelle partie permet de dérouler un peu plus le plan de ce manuscrit, en abordant cette fois-ci les systèmes d'intelligence artificielle et leur utilité vis-à-vis des données médicales à la disposition de ce travail.

Sommaire

3.1 Généralités sur l'intelligence artificielle	38
3.2 Approches par apprentissage	39
3.2.1 Apprentissage supervisé	40
3.2.2 Apprentissage non-supervisé	43
3.2.3 Apprentissage semi-supervisé	44
3.3 Apprentissage profond	45
3.3.1 Réseau de neurones artificiels	45
3.3.2 Réseau neuronal convolutif	46
3.3.3 Transfert de connaissances	47
3.4 Paramétrage et évaluation de modèles prédictifs	48
3.4.1 Généralisation de modèles	48
3.4.2 Choix de métriques	49
3.4.3 Réglages des hyperparamètres	52
3.4.4 Méthodes d'évaluation	53
3.5 Fusion d'informations	55

3.1/ GÉNÉRALITÉS SUR L'INTELLIGENCE ARTIFICIELLE

La notion d'**Intelligence Artificielle, IA** (*alt. : Artificial Intelligence, AI*) est assez souvent contestée, mais se définit plus généralement comme étant « l'ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine »¹. Ainsi, dans cette définition, ces techniques ne se limitent pas à la reproduction du comportement interactif social humain souvent représenté dans les médias. La mise en pratique sous forme informatique de mécanismes dédiés à la prise de décision à partir de données ou encore, la retranscription du raisonnement d'un expert font partie de cette définition.

Une vision plus complète de ce terme consiste à décomposer l'**IA** en une discipline organisée en sous domaines possédant chacun leurs spécificités, comme schématisé sur la Figure 3.1. Ainsi, le terme d'**apprentissage automatique** ou de *machine learning* caractérise un premier sous champ de l'**IA** et est employé pour qualifier l'utilisation de mécanismes statistiques dans le but de générer des règles, à partir de données d'apprentissage. Ces données étant souvent complexes, il est souvent nécessaire de laisser l'humain intervenir pour extraire l'information pertinente à l'accomplissement de la tâche souhaitée. Enfin, le terme d'**apprentissage profond** ou de *Deep Learning* désigne un sous champ de l'apprentissage automatique, dans lequel des couches successives de prises de décision vont permettre d'augmenter la complexité des relations entre les données d'entrée et les tâches.

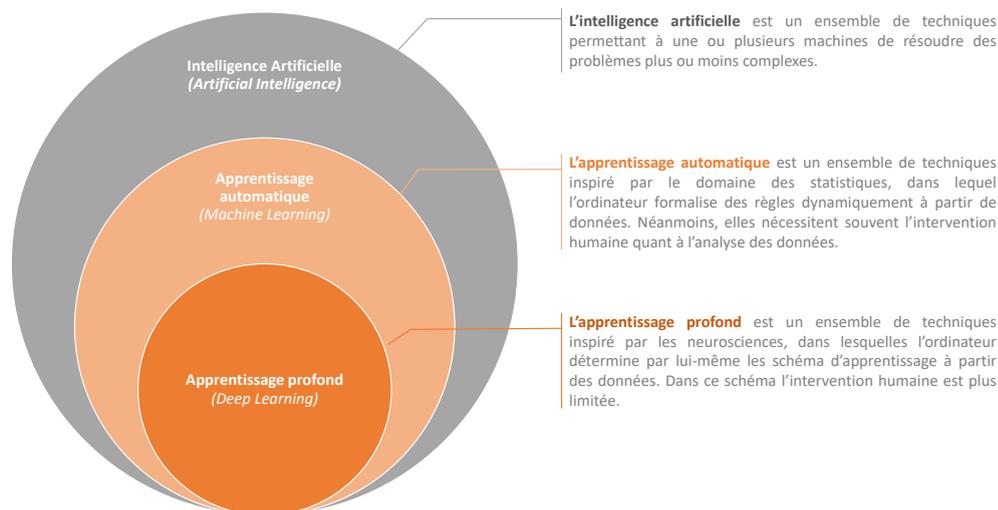


Figure 3.1 : Représentation des relations entre *intelligence artificielle*, *apprentissage automatique* et *apprentissage profond*.

Cette partie a pour vocation d'amener à la compréhension générale des courants d'**IA**, qui permettent la résolution de problématiques variées telles que la dermatologie. Dans une première section, les approches par apprentissage sont décrites, puis par extension amène à la présentation de l'apprentissage profond dans une seconde section. La troisième section décrit les méthodes permettant le paramétrage et l'évaluation de ces modèles. Enfin, la dernière section décrit les principes de fusion utiles à ce travail.

¹Source : Encyclopédie Larousse.

3.2/ APPROCHES PAR APPRENTISSAGE

Ce courant a été *évoqué* en 1948 par Alan Turing ; son idée était de proposer des « machines à apprendre susceptibles de construire elles-mêmes leurs propres codes » [199]. En 1959, Arthur Samuel formule une première définition des approches par apprentissage comme étant un « domaine d'étude dans lequel la possibilité est donnée à l'ordinateur d'apprendre sans avoir été explicitement programmé ».

Les approches par apprentissage automatique peuvent être résumées aux techniques permettant à un système informatique d'adapter son analyse et son comportement en fonction de ses données d'entrée. Une définition plus formelle et récente de ce domaine proposée par Tom Michael Mitchell est qu'un « programme informatique apprend d'une expérience E , en adéquation avec des tâches T et une mesure de performance P , si sa performance sur les tâches T , mesurée par P , s'améliore avec l'expérience E ».

Ainsi au sein des courants d'apprentissage, deux grandes sous catégories synthétisées dans la Figure 3.2 peuvent être distinguées, dont :

- l'apprentissage **supervisé** correspond aux techniques visant à faire correspondre les données d'entrée avec des sorties de valeurs discrètes (on parle aussi de classification) ou continues (on parle alors de régression),
- l'apprentissage **non-supervisé** correspond à des approches exploratoires, dans lesquelles l'ordinateur tente d'identifier des corrélations au sein des données (désigné par le terme de *clustering*),
- l'apprentissage **semi-supervisé** [166] reprend les précédentes théories mentionnées. Ce domaine tente d'extrapoler la connaissance des données annotées à celle des données non annotées afin d'exploiter la totalité de l'information disponible.

Les prochaines sous-sections de ce chapitre abordent chacune de ces catégories respectives de manière plus détaillée.

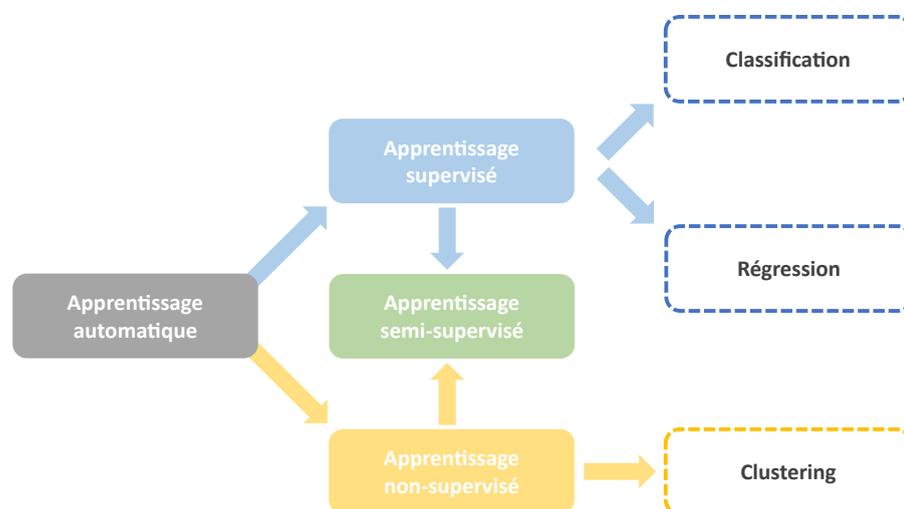


Figure 3.2 : Schéma de décomposition des approches par apprentissage automatique.

3.2.1/ APPRENTISSAGE SUPERVISÉ

L'apprentissage supervisé consiste à déterminer la relation permettant de faire correspondre des données d'entrée x et des données de sorties y et cela, dans le but de produire sur de nouvelles données d'entrée x' , de nouvelles sorties y' .

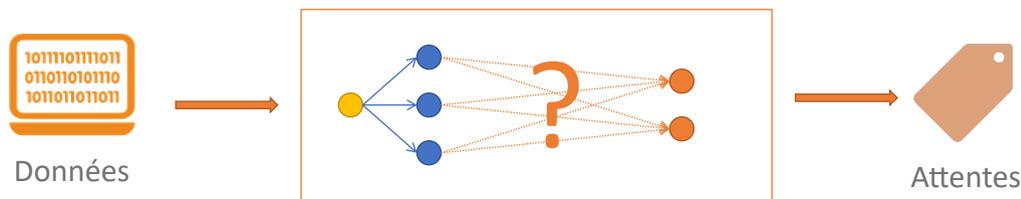
Cette approche implique l'utilisation d'une base d'apprentissage, définie comme *un ensemble de couples entrée-sortie* noté $\{(x_1, y_1), \dots, (x_n, y_n)\}$ avec $n \in N$. Ainsi, l'humain tente d'apporter une signification à ces données sous forme d'annotations (ou étiquettes) pour lesquelles la machine devra être en mesure de déterminer la relation existante.

Dans le but d'achever cette tâche, les variables propres à chaque observation sont supposées suffisamment discriminantes pour permettre l'identification des annotations. Ce problème est défini par la relation $y = f(x)$, où f est une fonction inconnue correspondant au phénomène observé, par une fonction d'approche g appelée fonction de prédiction tel que $y = g(x)$, avec $x = \{x_1, x_2, \dots, x_n\}$ un vecteur de caractéristiques possédant suffisamment d'information [67].

Cet apprentissage, schématisé sur la Figure 3.3, se découpe en deux processus distincts :

- **l'apprentissage** ou *entraînement*, qui consiste à approcher la fonction f par une fonction g compte tenue de données étiquetées,
- **la prédiction** ou *inférence*, qui consiste à prédire sur de nouvelles données à partir de g tel que $g(x') = y'$.

Phase d'entraînement



Phase de prédiction

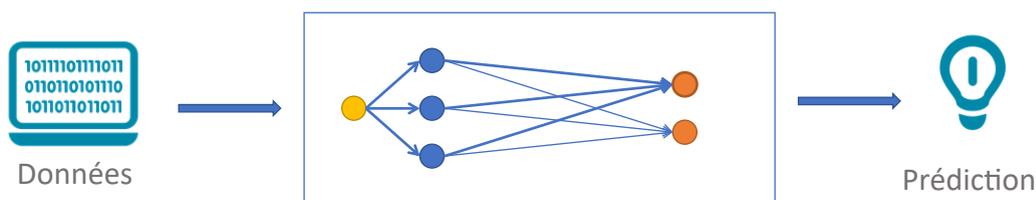


Figure 3.3 : Schéma de l'approche dite supervisée. La phase d'entraînement permet de déterminer les relations entre données et attentes, sur des données connues dites d'apprentissage. La phase de prédiction réutilise les relations déterminées lors de la phase d'entraînement.

Ces approches dites supervisées se regroupent ainsi au travers de deux catégories majeures :

- la **classification**, c'est-à-dire la prédiction de valeurs discrètes, soit l'ensemble des entiers relatifs noté \mathbf{Z} . Sont également utilisés les termes de *classification binaire* lorsque le problème formulé ne comporte que deux classes, et de *classification multi-classes* lorsque la situation nécessite de prédire N classes avec $N > 2$.
- la **régression**, c'est-à-dire la prédiction de valeurs continues, soit l'ensemble des nombres réels noté \mathbf{R} .

Les travaux de ce manuscrit portent sur les méthodes de classification, qui se traduisent par la catégorisation d'images cliniques pathologiques ou de lésions selon divers niveaux de dangerosité. En prenant appui sur l'un des travaux [103], une synthèse de ces méthodes et de leur principe respectif est fournie, dont :

- **les approches logiques** qui correspondent à des approches par succession de décisions [103]. Dans cette catégorie, le principal représentant est celui des arbres de décision [20], également connu sous le terme de **Classification And Regression Trees, CART** (*alt. : Arbre de Classification et de Régression, ACR*), dont le principe peut se résumer à la construction d'un graphe, dans lequel chaque noeud est associé à une décision définie par le choix d'une caractéristique et d'un seuil [160]. Toute la complexité de ces approches repose sur le choix de ces caractéristiques dont la priorité dépend de l'ordre d'importance dans la séparation du problème.
- **les approches statistiques** qui correspondent à des approches probabilistes, qui déterminent la probabilité d'appartenance d'un élément à une classe, pouvant être étendu sous la forme de réseau [102]. Le **Naive Bayes Model, NBM** (*alt. : Modèle Bayésien Naïf, MBN*) [222] est un exemple typique d'approche probabiliste, dans lequel chaque caractéristique est considérée comme indépendante. La relation entre caractéristique et annotation se définit ainsi comme le produit des probabilités de chacune d'entre elles d'appartenir à la classe supposée, régit selon :

$$\hat{y} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \quad (3.1)$$

- **les approches par instances** qui qualifient des méthodes consistant à une accumulation des échantillons afin d'enrichir un espace et non pas de leur interprétation menant à une forme de connaissances. La méthode la plus connue reposant sur ce principe est celle des **k -Nearest Neighbors algorithm, KNN** (*alt. : Méthode des k plus proches voisins, KNN*) [46], dont la théorie majeure repose sur l'appartenance d'une donnée à une classe si celle-ci est suffisamment proche par ses caractéristiques, d'échantillons pré-existants de cette dernière. Il est alors nécessaire de définir un critère de distance, le principal utilisé étant celui de la distance euclidienne.
- **les approches par Support Vector Machine, SVM** (*alt. : Machine à Vecteurs de Support, MVS*) [45] correspondent à des approches assez récentes dont le principe se résume à déterminer une frontière de séparation entre les diverses classes, qui maximise une marge suffisante permettant entre autres une meilleure tolérance au bruit. Afin de déterminer au mieux cette frontière, divers noyaux ont été établis dont le plus simple d'entre eux est le **noyau linéaire**. La présence d'un **noyau de fonction de base radiale** ou *Radial Basis Function* est à souligner, considéré comme un *approximateur universel* si les données à disposition sont suffisantes [206].

- **les approches par ensembles** représentent une catégorie qui qualifie les méthodes qui emploient un ensemble de modèles prédictifs afin de construire un unique modèle plus robuste. L'une de ces méthodes, les **Random Forest, RF** (*alt. : Forêts Aléatoires, RF*) [19] sont une extension du principe d'arbre de décision précédemment évoqué, destiné à réduire le risque de sur-apprentissage du modèle initial. Cette méthode réalise N sélections aléatoires d'observations, pour lequel N arbres de décision sont générés. Afin de diversifier ces arbres, un mécanisme de tirages aléatoires de caractéristiques est également réalisé au niveau de chaque noeud de décision. Une autre de ces méthodes, le **Gradient Boosting, GB** (*alt. : Augmentation du Gradient, AG*) est un mode opératoire généralement constitué par des arbres de décisions. En opposition avec les **RF** où chaque arbre de décision correspond à un modèle de prédiction faible, le **GB** consiste à diminuer de manière séquentielle une fonction de coût à chaque nouvel arbre généré [69].

Devant ces nombreuses méthodes, il est difficile de déterminer un modèle optimal à une situation particulière, notamment dans le cadre de données à forte complexité. En effet, une évaluation de manière empirique de ces diverses méthodes reste encore la meilleure solution afin de déterminer ce modèle optimal, en contrepartie d'un temps de calcul important. Néanmoins, l'un des travaux menés sur la performance des modèles supervisés, a permis de déterminer un certain nombre de critères de comparaison entre ces diverses techniques mettant en valeur en général les avantages et inconvénients de chacune d'entre elles [103].

Pour cela, la Table 3.1 recense la plupart des critères importants évoqués par ce travail de recherche. La *performance de classification* qui semble indéniable à ce champ de travail et qui représente l'un des principaux atouts des modèles **SVM**. La *vitesse d'apprentissage* a également été précisée à titre indicatif. En effet, bien que contraignante lors des expérimentations de ce manuscrit, elle ne constitue pas un frein et reste le principal inconvénient des **SVM**. Enfin, ces derniers sont reconnus comme robustes face à des situations pouvant comporter des caractéristiques non pertinentes ou redondantes.

À la vue de ces éléments, il semble plus que nécessaire de considérer l'utilisation des **SVM** au sein de ces travaux. De même, il est nécessaire de considérer l'utilisation des deux méthodes d'approches par ensemble mentionnées, les **RF** et le **GB**, pour lesquelles aucun travail comparatif généraliste n'a été trouvé. Néanmoins, une étude génomique semble plaider dans le sens des **SVM** et du **GB** [145]. Pour finir, il est également nécessaire de considérer ces modèles de classification au sein de processus proposés par des travaux proches de cette thématique. Ces éléments sont abordés plus en détail dans leur partie respective.

	CART	NBM	KNN	SVM
Performances de classification	**	*	**	****
Vitesse d'apprentissage	***	****	****	*
Tolérance aux caractéristiques non pertinentes	***	**	**	****
Tolérance aux caractéristiques redondantes	**	**	**	***

Table 3.1 : Table de comparaison des méthodes d'apprentissage majeures de la littérature [103]. Ce tableau propose un système de notations allant de une (mauvaise performance) à quatre étoiles (bonne performance).

3.2.2/ APPRENTISSAGE NON-SUPERVISÉ

L'apprentissage non supervisé ou *descriptif* est une seconde approche d'apprentissage dans laquelle l'ordinateur tente de *découvrir* en autonomie, des corrélations au sein de jeux de données. Ces approches émergent de problématiques diverses, telles que :

- la **réduction des coûts** le plus souvent humain nécessaire à l'obtention de données annotées, c'est-à-dire de données pour lesquelles les couples d'entrées et de sorties sont connus.
- la **découverte des diverses relations** pouvant exister au sein d'un amas de données. En effet, une annotation ne correspond qu'à un échantillon de l'information et ne permet pas d'obtenir les relations pouvant régir des modèles d'interactions complexes.
- l'**exploration de nouvelles relations** de type *cause – effet*, en réaction à la masse de données produites par les objets connectés.

Ce principe est concrétisé aux travers de méthodes, comme les k-moyennes qui tentent de déterminer des correspondances en minimisant la différence d'énergie entre les points d'un même groupe, dont un schéma est visible sur la Figure 3.4. Diverses applications peuvent ainsi en résulter, telles que :

- le **regroupement par classes de données**, afin de définir de manière automatique de nouvelles annotations.
- la **réduction de dimensions**, dans le but de ne conserver que l'information essentielle. La méthode du *bag-of-words* est un exemple initialement prévu pour l'analyse de texte [223] et utilisé dans des domaines tels que l'analyse d'images de la peau [174].
- la **découverte de relations** au sein de l'information, et des relations les plus robustes entre variables et dépendances.

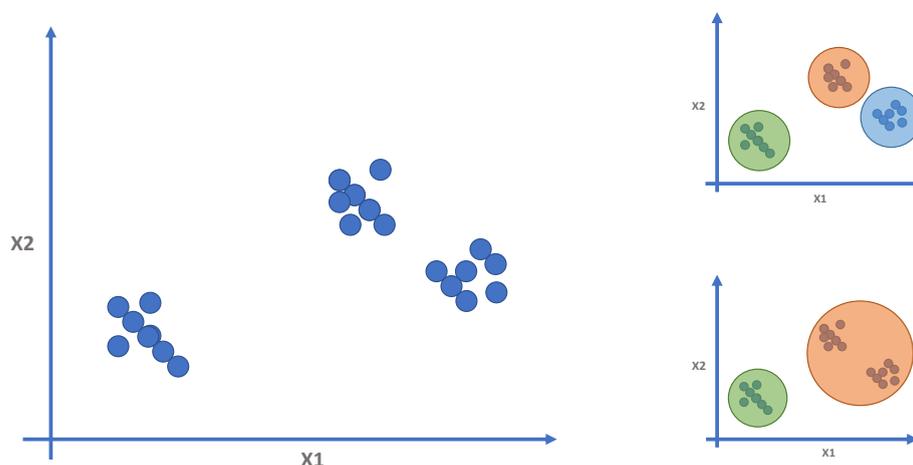


Figure 3.4 : Exemple de données non associées à des annotations préalables, et pour lesquelles des approches non supervisées peuvent être sollicitées afin de découvrir des groupes de données ou *clusters*. Le nombre de groupes le plus souvent défini par l'utilisateur avant traitement, influe fortement la vision du problème.

3.2.3/ APPRENTISSAGE SEMI-SUPERVISÉ

Les techniques d'apprentissage semi-supervisé tentent de combiner les deux principes précédents et sont la conséquence du coût de l'annotation des données qui nécessite souvent le travail d'experts. En effet, il peut être difficile d'avoir accès à une vérité terrain qui couvre la totalité d'un jeu de données sans pour autant perdre l'apport de cette information non-annotée. Ce dernier point est particulièrement intéressant dans le cadre de données à grande dimension pour lesquelles l'espace des dimensions croît de manière exponentielle et dilue les échantillons (connu sous le terme de *fléau des dimensions* [54]).

Ainsi, l'apprentissage semi-supervisé se base sur plusieurs hypothèses auxquelles doivent se confronter les données [226], dont :

- un critère **d'homogénéité**, c'est-à-dire que des données issues d'une zone de haute densité partagent les mêmes annotations,
- un critère **de séparation par faible densité**, c'est-à-dire que s'il existe une séparation entre plusieurs types d'annotations, celle-ci se situant dans une zone à faible densité [33].

En pratique, il est difficile de s'assurer que des données mises à disposition respectent bien ces engagements. Afin de mieux visualiser ce concept, l'exemple en Figure 3.5 démontre l'intérêt de telles approches. Ce principe permet d'éviter d'extrapoler des frontières difficilement définissables dans une situation de faible densité de l'information étiquetée.

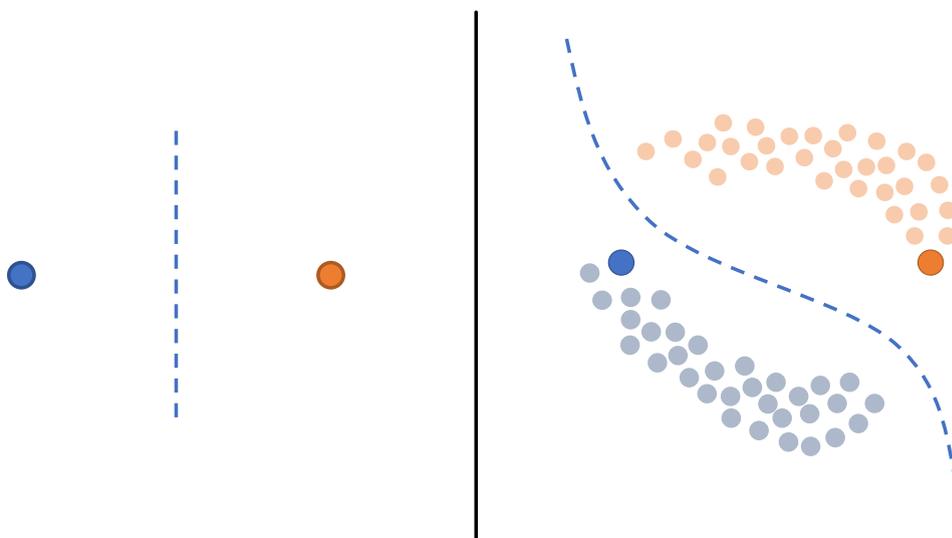


Figure 3.5 : Exemple fréquemment employé pour démontrer l'intérêt de l'apprentissage semi-supervisé du point de vue du principe de densité. À gauche, une classification obtenue à partir de deux données étiquetées ; À droite, la même situation avec l'ajout de données non étiquetées.

3.3/ APPRENTISSAGE PROFOND

L'apprentissage profond correspond à un sous-ensemble de l'apprentissage automatique, encouragé par les récents apports des neurosciences sur le fonctionnement et le rôle des neurones sur la prise de décisions complexes [159, 171]. En effet, le cerveau possède divers niveaux de traitement de l'information, ce que tente d'imiter l'apprentissage profond par l'apport de multiples couches de traitement.

Dans le cadre des méthodes *classiques* d'apprentissage évoquées dans la partie précédente, la dimension de *couches* de traitement n'est pas intégrée. Les approches par apprentissage simple proposent une architecture en trois couches, dont l'une des couches est allouée à la corrélation entre les données d'entrée et de sortie. Par opposition, les approches par apprentissage profond proposent des structures en n couches, avec $n \in \mathbf{N}$ et $n > 1$. Ces couches intermédiaires sont également qualifiées de couches cachées.

Dans une première sous-section, une brève présentation du principe de **Artificial Neural Network**, **ANN** (*alt. : Réseau de Neurones Artificiels, RNA*) est réalisée, puis dans une seconde sous-section est détaillée leur extension au **Convolutional Neural Network**, **CNN** (*alt. : Réseau Neuronal Convolutif, RNC*). Enfin, les aspects liés au principe de transfert de connaissances sont proposés dans une dernière sous-section.

3.3.1/ RÉSEAU DE NEURONES ARTIFICIELS

Les **ANN** sont des structures multicouches composées d'unités basiques appelées neurones. Dans ce schéma, chaque neurone ou unité est connecté à l'ensemble des unités de la couche $N - 1$ et $N + 1$ et forme un *maillage*.

Les neurones se partagent ainsi l'information de point à point de l'entrée vers la sortie, appliquant respectivement l'opération dont ils sont responsables. Cette opération au sein d'une unité est de la forme $y = W\mathbf{x} + \mathbf{b}$ dans laquelle W représente une matrice de poids qui permet de pondérer les signaux des prédécesseurs et b un terme de correction appelé biais [181]. Ainsi, les neurones en entrée d'un réseau reçoivent une information *brute* tandis que, les neurones situés en sortie du réseau reçoivent une information pré-traitée par les prédécesseurs. Le schéma présent sur la Figure 3.6 permet de mettre en évidence un cas simple de résolution par apprentissage pour appréhender cette notion.

Néanmoins, cet exemple met également en avant que ce même modèle pourrait être représenté à l'aide d'une fonction d'ordre suffisante [16]. Ce mécanisme seul ne suffit pas à créer des réseaux exploitant les bénéfices d'un tel agencement. En effet, si une unité de ce réseau est représentée par $y = W\mathbf{x} + \mathbf{b}$, une succession de ces unités pourrait se résumer à la fonction linéaire représentée par l'Equation (3.2)². Afin de tirer parti du potentiel de ce choix d'agencement en couches, ont été introduites des fonctions d'activation (tanh, ReLu, ...) qui permettent d'introduire une non-linéarité au sein de ces réseaux, permettant notamment l'autodétermination de structures de décision complexes.

$$\begin{aligned} y = h(\mathbf{x}) &= \mathbf{b}_n + W_n(\mathbf{b}_{n-1} + W_{n-1}(\dots(\mathbf{b}_1 + W_1\mathbf{x})\dots)) \\ &= \mathbf{b}_n + W_n\mathbf{b}_{n-1} + W_nW_{n-1}\mathbf{b}_{n-2} + \dots + W_nW_{n-1}\dots W_1\mathbf{x} \\ &= \mathbf{b}' + W'\mathbf{x}. \end{aligned} \quad (3.2)$$

²Source : Plateforme de cours en ligne [Deep Learning AI](#), démonstration par Andrew Ng.

Un très bon représentant de ces types de réseaux est le **Multi-Layer Perceptron, MLP** (*alt.* : *Perceptron Multi-Couches, PMC*) utilisé à des fins de classification. Par ailleurs, un inconvénient majeur de ces réseaux est le grand nombre d'hyperparamètres à disposition. Il est ainsi délicat de procéder au choix du nombre de neurones ou même de celui des fonctions d'activation.

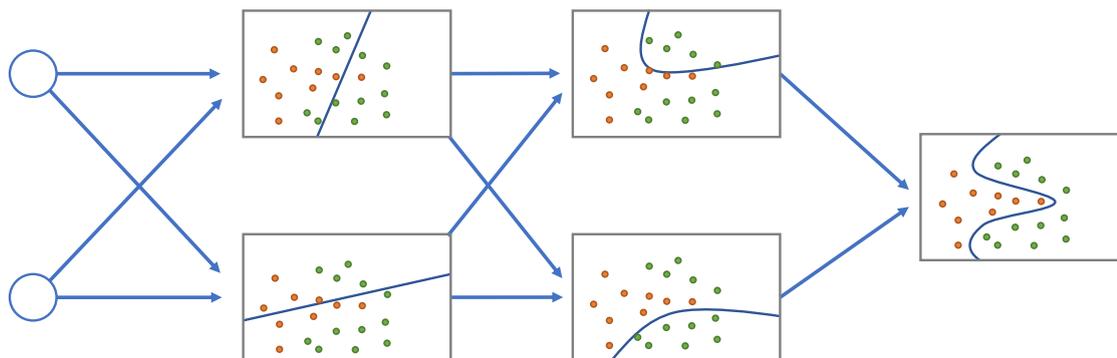


Figure 3.6 : Exemple simplifié d'une structure d'apprentissage profond. Ce modèle se compose de deux couches intermédiaires cachées : la première couche extrait des caractéristiques linéaires simples, tandis que la seconde couche par combinaison avec la première couche conduit à une extraction de caractéristiques plus complexes. Ainsi, la première couche s'apparente à un polynôme de degré 1, tandis que la seconde couche représente des fonctions de degré 2 et ainsi de suite.

3.3.2/ RÉSEAU NEURONAL CONVOLUTIF

Les **CNN** sont une extension des réseaux de neurones profonds, essentiellement utilisés dans un but de traitement de l'image. Moins développé par la littérature, ce principe de réseaux peut être étendu à des données telles que des signaux (1D) ou des volumes (3D).

Ces **CNN** sont une réponse à l'application des **ANN** à des problématiques impliquant des données images et des contraintes multiples qu'ils comportent. Parmi ces contraintes de l'application des **ANN** aux images, peuvent être évoqués : (i) les **nombreuses connexions et paramètres requis**, (ii) l'**absence d'utilisation de la cohérence de l'information spatiale**, (iii) et la **dépendance** du réseau à la taille des informations d'entrée.

Ainsi, ces réseaux font appel à des couches de convolution permettant de résoudre ces trois principales contraintes. D'une part, la convolution permet de réduire fortement le nombre de paramètres d'entraînement, le réseau n'est ainsi plus entièrement *connecté*. Certaines optimisations récentes consistent à remplacer les filtres de convolution 2D par une succession de filtres 1D horizontaux et verticaux, proposant des résultats semblables et réduisant fortement le nombre de paramètres. D'autre part, la convolution permet d'apporter une logique à la dimension spatiale des données ainsi qu'une indépendance à la taille. En effet, le réseau n'est plus entièrement connecté et donc ne dépend plus d'une taille fixe, et peut ainsi traiter des données de tailles différentes mais qui doivent respecter une taille minimum. Le traitement des données par ces couches de convolution porte le nom de carte d'activation, maximum quand le produit de convolution est en phase.

3.3.3/ TRANSFERT DE CONNAISSANCES

L'apprentissage par transfert est une technique connexe à l'apprentissage profond défini comme « faisant référence à la situation où ce qui a été appris dans un contexte . . . est exploité pour améliorer la généralisation dans une autre contexte » [142]. Cette discipline s'inscrit dans le champ de **l'adaptation de domaine**, définie comme le transfert des connaissances d'un domaine D_S , défini par l'hypothèse $h : X_S \rightarrow Y_S$, vers un domaine cible D_T afin de satisfaire $h : X_T \rightarrow Y_T$. Son champ d'application n'est pas limité aux réseaux d'apprentissage profond, mais les contraintes qui leurs sont associées font de ces techniques une ressource utile.

En effet, ces solutions permettent de réduire les difficultés liées à la grande quantité de ressources de calcul, de temps, mais également de données mobilisées nécessaires à un entraînement depuis sa base. Le modèle d'apprentissage attendu est alors initialement plus performant, croît plus efficacement et est plus performant qu'un apprentissage spécifique (Figure 3.7).

Néanmoins, cette solution n'est envisageable que lorsque l'apprentissage, réalisé lors de la première tâche, généralise suffisamment le phénomène observé. Ainsi, il est d'usage courant de ne spécifier lors du transfert que la couche finale de classification du réseau, les premières couches étant employées comme extracteur général de caractéristiques. Toutefois, certaines méthodes de la littérature s'accordent sur l'entraînement conjoint des couches de classification ainsi que des couches de convolutions hautes, considérées comme moins génériques.

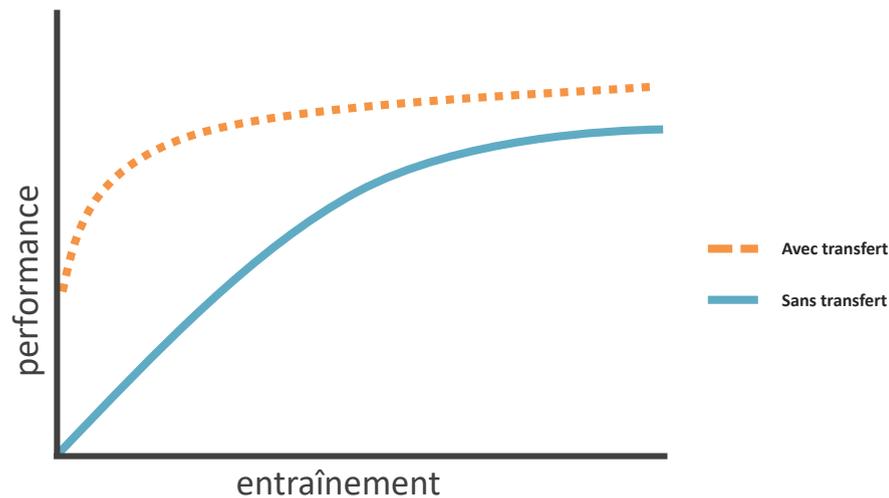


Figure 3.7 : Exemple d'attentes liées à un apprentissage par transfert en opposition à un apprentissage classique.

3.4/ PARAMÉTRAGE ET ÉVALUATION DE MODÈLES PRÉDICTIFS

Cette partie se concentre sur l'évaluation de modèles de prédiction et présente la problématique de **biais**. Ses origines et ses risques associés sont présentés ainsi que les potentielles méthodes permettant de s'en affranchir.

3.4.1/ GÉNÉRALISATION DE MODÈLES

À l'aide de couples d'entrée et de sortie mis à disposition, ces modèles sont capables d'apprendre une relation en la *généralisant* et d'estimer de nouvelles sorties sur de nouvelles données. Il est pour cela nécessaire de parvenir à isoler l'information responsable de ce cheminement, plus ou moins séparable, en cause souvent la qualité de : (i) l'hypothèse formulée, c'est-à-dire si la relation qui lie l'observation et sa conséquence est directe mais également vraie dans toute circonstance, (ii) l'information possédée, c'est-à-dire si l'information est juste pour répondre à la problématique cible (qualité de l'observation, rapport signal sur bruit, ...).

Ce problème, évoqué dans la littérature porte le nom de « dilemme de biais et variance », dans lequel le biais représente le manque de relation pertinente entre données d'entrée et de sortie et où la variance représente l'influence du modèle aux petites fluctuations souvent issues de bruit. Ces termes sont intrinsèquement liés aux problématiques respectives de **sous-apprentissage** et **sur-apprentissage**. La Figure 3.8 reprend de manière visuelle ces problématiques, et les exprime selon un jeu d'entraînement et de test. À noter que le sur-apprentissage peut être limité par l'utilisation de termes de régularisation, de schémas de validation de modèle adaptés ou l'utilisation d'ensembles de données suffisantes.

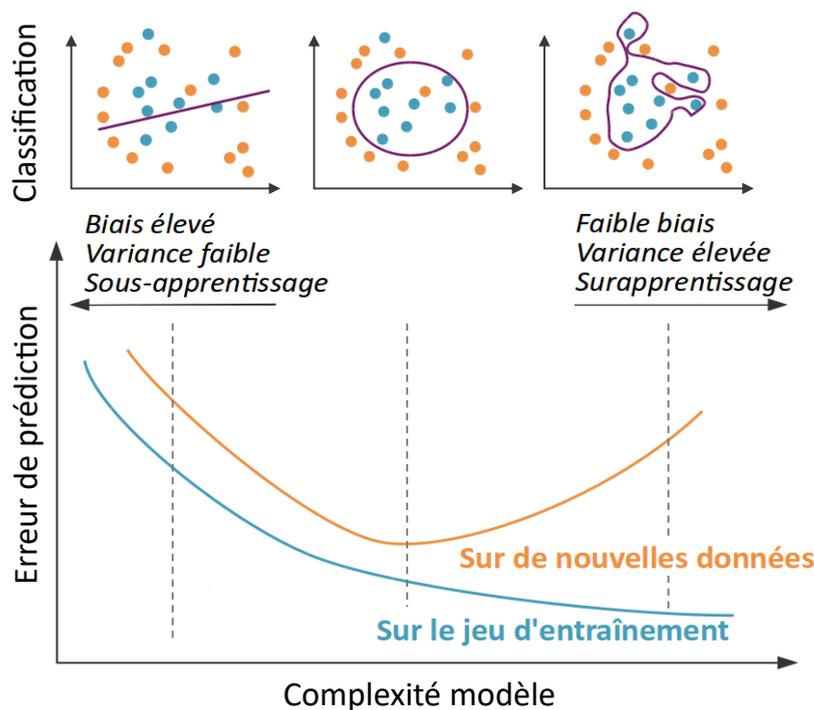


Figure 3.8 : Exemple de dilemme de biais et de variance. À gauche, un cas de sous-apprentissage, ce modèle est une sous approximation du phénomène observé. À droite, un cas de sur-apprentissage, le modèle ne généralise pas suffisamment le phénomène observé.

3.4.2/ CHOIX DE MÉTRIQUES

Le choix des métriques représente une part importante de la problématique de ce manuscrit pour la mise en place de modèles d'apprentissage. D'une part, ce choix permet d'orienter la sélection d'un modèle en particulier et d'autre part, permet de retenir les performances attendues par ce dernier sur des données extérieures au jeu de données employé.

L'un des principaux outils à disposition afin d'évaluer la performance est la matrice de confusion, schématisée sur la Figure 3.9. Elle permet de faire correspondre les classes réelles et prédites : la diagonale de cette matrice fait figurer les prédictions correctement réalisées par le modèle, tandis que les éléments restants renseignent sur les prédictions erronées. Dans une situation à deux classes ou *binaires*, cette matrice fait ainsi ressortir quatre mesures :

- les **vrais positifs**, les éléments de la classe positive détectés comme positif,
- les **vrais négatifs**, les éléments de la classe négative détectés comme négatif,
- les **faux positifs**, les éléments de la classe négative détectés comme positifs,
- les **faux négatifs**, les éléments de la classe positive détectés comme négatifs.

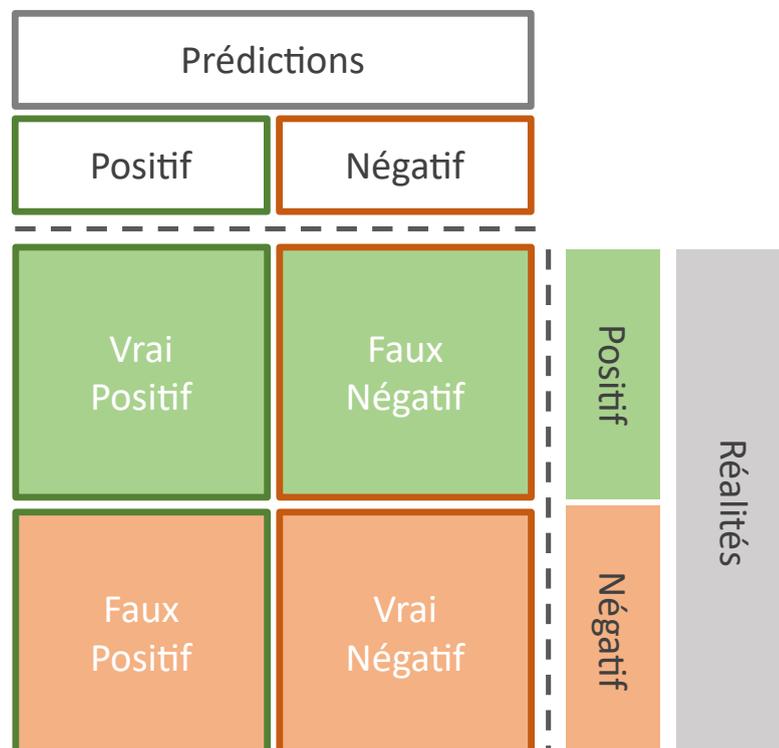


Figure 3.9 : Schéma représentatif d'une matrice de confusion, qui met en avant une situation binaire.

Bien que déjà réduite, l'information de cette matrice est encore très riche et difficile à utiliser comme critère de sélection. Pour cela, diverses métriques éprouvées sont à envisager selon la problématique de recherche. Parmi les plus courantes d'entre elles, les critères de **précision**, **sensibilité** (également appelé **rappel**) ou encore **spécificité** sont pléthores dans la littérature mais ont tendance à ne capter qu'une partie du problème. Par exemple, la sensibilité est un critère très important en médecine ou détection de fraude, mais ne qualifie en rien la spécificité du modèle. De même, la moyenne fait également partie des critères extraits dans le domaine de l'intelligence artificielle, mais peut également manquer de sens en cas de non balancement de l'information [76] ou même ne pas mettre en avant un important gouffre entre spécificité et sensibilité. Ces métriques sont représentées de manière graphique sur la Figure 3.10 ou sous forme mathématique dans l'Equation (3.3).

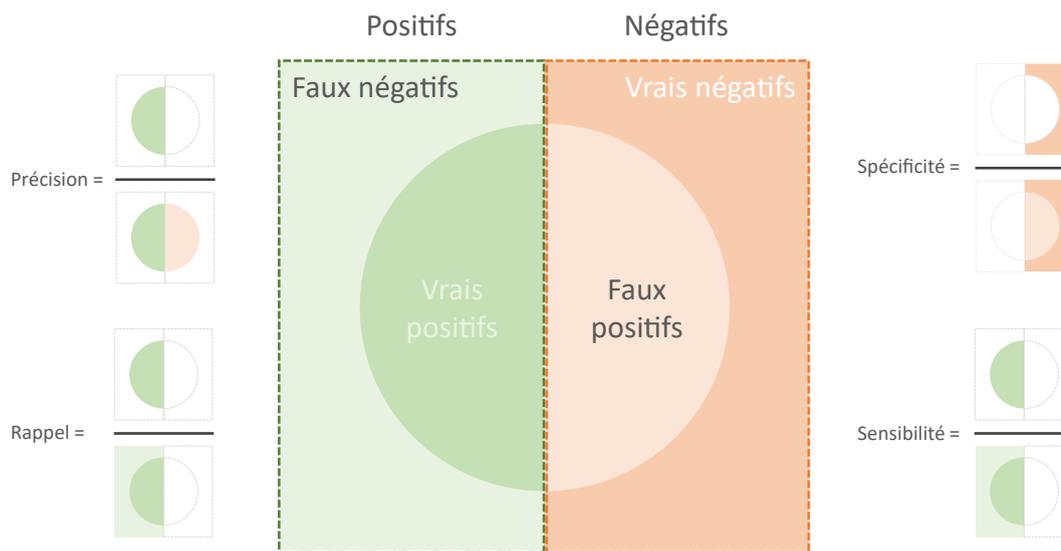


Figure 3.10 : Schéma représentatif des principales métriques obtenues à partir de la matrice de confusion : la *précision*, le *rappel* ou la *sensibilité* et la *spécificité*.

$$\begin{aligned}
 \text{Sensibilité} &= VP/(VP + FN) \\
 \text{Rappel} &= \text{Sensibilité} \\
 \text{Spécificité} &= VN/(VN + FP) \\
 \text{Précision} &= VP/(VP + FP) \\
 \text{Moyenne} &= (VP + VN)/(VP + VN + FP + FN)
 \end{aligned} \tag{3.3}$$

Dans ce cas, il peut être préférable de privilégier des métriques plus robustes, dans le cadre de données non balancées. La plupart des méthodes s'accordent sur l'utilisation de métriques tels que le score F_1 , associant conjointement sensibilité et précision, par principe de la moyenne harmonique [76]. L'Equation (3.4) met à disposition l'expression générique de cette métrique, dont le paramètre β permet de moduler l'importance de la précision ou du rappel. L'influence de ce paramètre est neutre lorsque sa valeur est égale à 1 comme dans le cas du score F_1 .

$$F_{\beta} \text{ score} = (1 + \beta^2) * (\text{Précision} * \text{Rappel}) / ((\beta^2 * \text{Précision}) + \text{Rappel}) \tag{3.4}$$

Une autre métrique permettant de se prémunir de ces effets de non balancement de l'information est celui de l'**AUC** et découlant de la courbe **Receiver Operating Characteristic, (ROC)** dont le principe est résumé sur le schéma Figure 3.11). La courbe ROC est tracée à partir des informations de spécificité / sensibilité et par variation d'une valeur seuil appliquée au modèle de prédiction, tandis que l'**Area Under the Curve, AUC** (*alt. : Aire sous la courbe, ASC*) correspond à l'intégration de l'aire de cette courbe. Ainsi, l'**AUC** est robuste au non-balancement de données mais ses valeurs sont optimistes sur des modèles de prédiction dont la variation du seuil est homogène vis-à-vis des résultats.

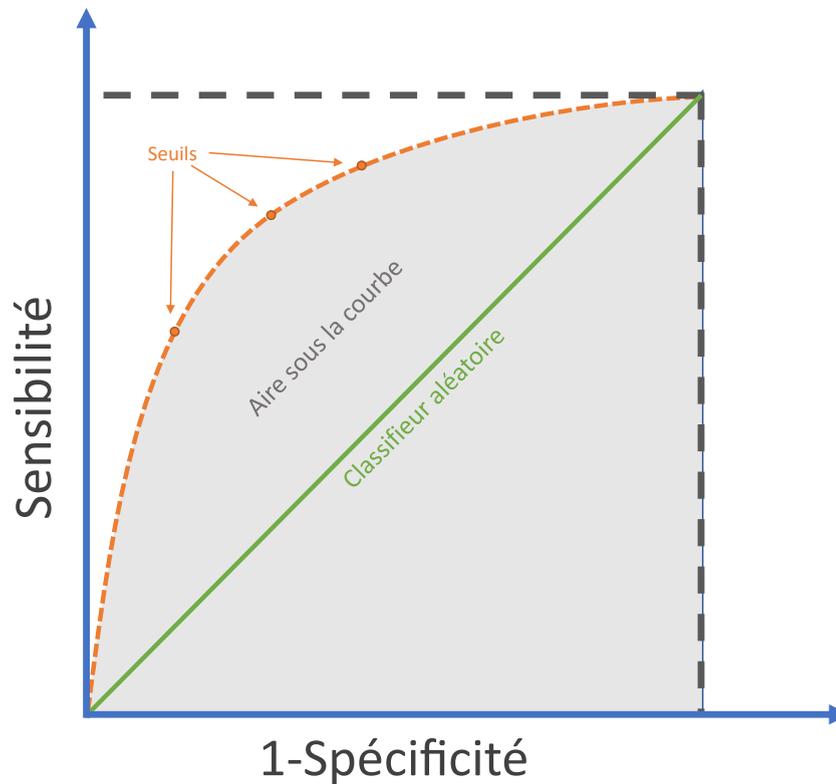


Figure 3.11 : Schéma de courbe ROC en orange et AUC en gris.

D'autres alternatives moins biaisées ont également été formulées [176]. Néanmoins le score F_1 et l'**AUC** sont couramment utilisés dans des articles cliniques et constituent une des bases de ce travail.

3.4.3/ RÉGLAGES DES HYPERPARAMÈTRES

Avant d'entrer plus en profondeur dans le sujet, il est nécessaire de revenir sur le terme **d'hyperparamètres**. Il caractérise les paramètres d'un processus d'apprentissage automatique dont le réglage doit être effectué avant la phase d'entraînement et dont les valeurs dépendent des données à traiter. Le réglage des hyperparamètres peut rarement être déterminé par la simple observation des données, et nécessite ainsi un réglage en fonction de chaque situation rencontrée.

L'une des premières techniques permettant ce réglage consiste à définir une *grille* de ces hyperparamètres, contenant pour chacun, leurs possibles variations [120]. Il s'agit de la méthode la plus équitable puisqu'elle consiste à évaluer chaque combinaison par comparaison des performances finales (voir la Figure 3.12 - Gauche). Néanmoins, elle possède deux limitations majeures, dont :

- son **temps de calcul**, qui va croître de manière exponentielle en fonction du nombre d'hyperparamètres à valider,
- sa **précision**, qui va dépendre du pas choisi pour chacun des paramètres.

La seconde technique consiste à définir un *espace* de recherche dont la distribution de chaque hyperparamètre sera choisie, ainsi qu'un nombre d'évaluations à réaliser pour déterminer la meilleure combinaison possible. Ce nombre d'évaluations est ensuite réalisé sous forme de tirage aléatoire, effectué à l'intérieur de cet espace de recherche défini [14] (voir la Figure 3.12 - Droite). Cette méthode possède l'avantage d'être prévisible en termes de temps puisque le nombre de combinaisons évaluées est défini au préalable, quelles que soient les proportions de l'espace. Néanmoins, cette méthode possède également des inconvénients comme la représentation des tirages au sein de l'espace défini.

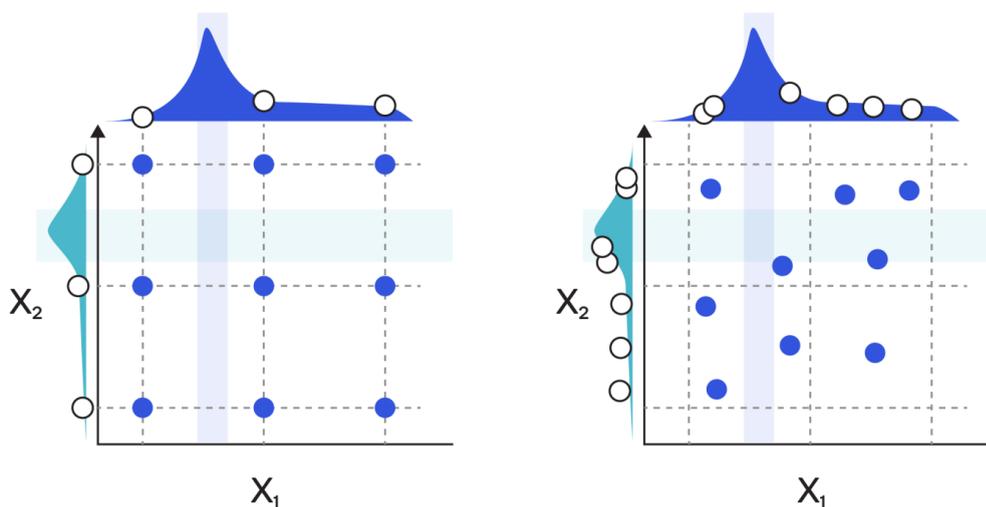


Figure 3.12 : Exemple de recherche sur deux hyperparamètres X_1 et X_2 . À gauche, une recherche sous forme de grille rigide, dans laquelle chaque combinaison est évaluée. À droite, une recherche aléatoire de ces hyperparamètres.

3.4.4/ MÉTHODES D'ÉVALUATION

Divers aspects sont évoqués pour permettre l'ajustement des modèles dans la Section 3.4.3 ainsi que la mesure de leurs performances dans la Section 3.4.2. Il est nécessaire d'orchestrer cet ensemble afin d'obtenir la meilleure attente possible d'un modèle de prédiction, tout en conservant une mesure pertinente permettant de renseigner sur la fiabilité de ce modèle et de ses prédictions futures dans des situations où la vérité terrain n'est pas connue. En effet, l'évaluation d'un modèle de prédiction sur des données identiques à celles utilisées constitue un biais, par l'utilisation de connaissances *a posteriori*. De même, l'ajustement des hyperparamètres rentre dans cette catégorie de l'entraînement, et l'évaluation ne peut se faire sur des données ayant servi à les déterminer. Par ailleurs, une telle évaluation permet de détecter les problèmes de sous-apprentissage et sur-apprentissage évoqués en Section 3.4.1.

La méthode **holdout** est la plus simple d'entre elles et consiste à diviser l'échantillon de données en deux sous-ensembles respectivement d'entraînement et de test. Il est important lors de cette étape de vérifier que les données présentes dans chacun de ces sous-ensembles soient représentatives de chaque classe présente dans la problématique. Les données spécifiées pour le test restent ainsi non utilisées jusqu'à leur confrontation au modèle définitif qui est alors évalué sur ces dernières. Le principal inconvénient de cette méthode est que seule une partie du jeu de données sert d'évaluation à la méthode développée. Néanmoins, cette méthode est rapide à exécuter puisqu'elle ne nécessite l'entraînement que d'un modèle. Cette méthode est schématisée sur la Figure 3.13 - Gauche.

La méthode de **validation croisée** permet d'étendre la méthode précédente à l'ensemble du jeu de données et d'obtenir un score moins biaisé. Pour cela, le jeu de données est subdivisé en k **échantillons**, servant à tour de rôle de jeu de test, le reste de ces données étant réservé à l'entraînement. Ce principe de fonctionnement génère ainsi k modèles de prédictions et autant de scores d'évaluation qui se résument à une unique valeur par application d'une fonction de moyenne. Le principal inconvénient de cette méthode est son temps de réalisation puisqu'elle nécessite l'entraînement d'autant de modèles que d'échantillons. Cette méthode est schématisée sur la Figure 3.13 - Droite.

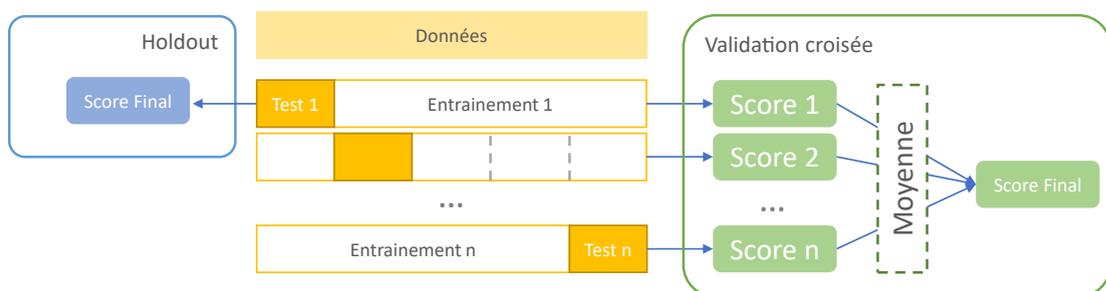


Figure 3.13 : Schéma des principes des deux approches majeures d'évaluation de modèles. À gauche, la méthode de *holdout* ; À droite, la méthode de *validation croisée*.

Ce travail s'intéresse à la technique de validation croisée ainsi qu'à la valeur moyenne renvoyée par celle-ci. Afin de rendre cette information plus complète, la valeur de l'écart-type est conjointement calculée à partir des évaluations réalisées sur les divers échantillons, jugée fiable afin de mesurer la stabilité du processus vis-à-vis des données [98]. La plupart des travaux ayant recours à la validation croisée emploient une valeur k comprise entre 5 et 10, jugée suffisante pour l'obtention de valeurs de moyenne et écart-type dans un temps de calcul raisonnable [92]. Néanmoins, il est

possible d'amener ce principe à l'extrême en choisissant la plus petite valeur possible pour un lot dédié à l'évaluation soit la taille d'une donnée. Cette valeur k devient égale au nombre de données à traiter, devenant un schéma qualifié de *leave-one-out* ou littéralement *en laisser un de côté*. Mais ce type de schéma de validation souffre généralement d'une variance élevée, surtout lorsque les données contiennent des valeurs aberrantes [12].

Ces deux méthodes répondent à la problématique de l'évaluation, mais n'apportent pas de solution quant au choix des hyperparamètres. En effet, utiliser un schéma de validation croisée pour régler et évaluer les modèles conduit inévitablement à surestimer les performances du modèle [197]. Ce type de démarche revient à ajuster au mieux une courbe à un nuage de points tout en affirmant que ces mêmes données correspondent à des données inconnues. L'une des solutions à ce problème est l'utilisation d'un double schéma de validation, dont l'un des plus fonctionnels actuellement est celui de la **validation croisée imbriquée** [26]. Le principe de cette méthode est schématisé sur la Figure 3.14, laquelle se décompose en :

- une boucle **externe**, permettant d'évaluer un modèle à partir de données de **test**,
- une boucle **interne**, permettant de valider un modèle et ses réglages à partir de données de **validation**.

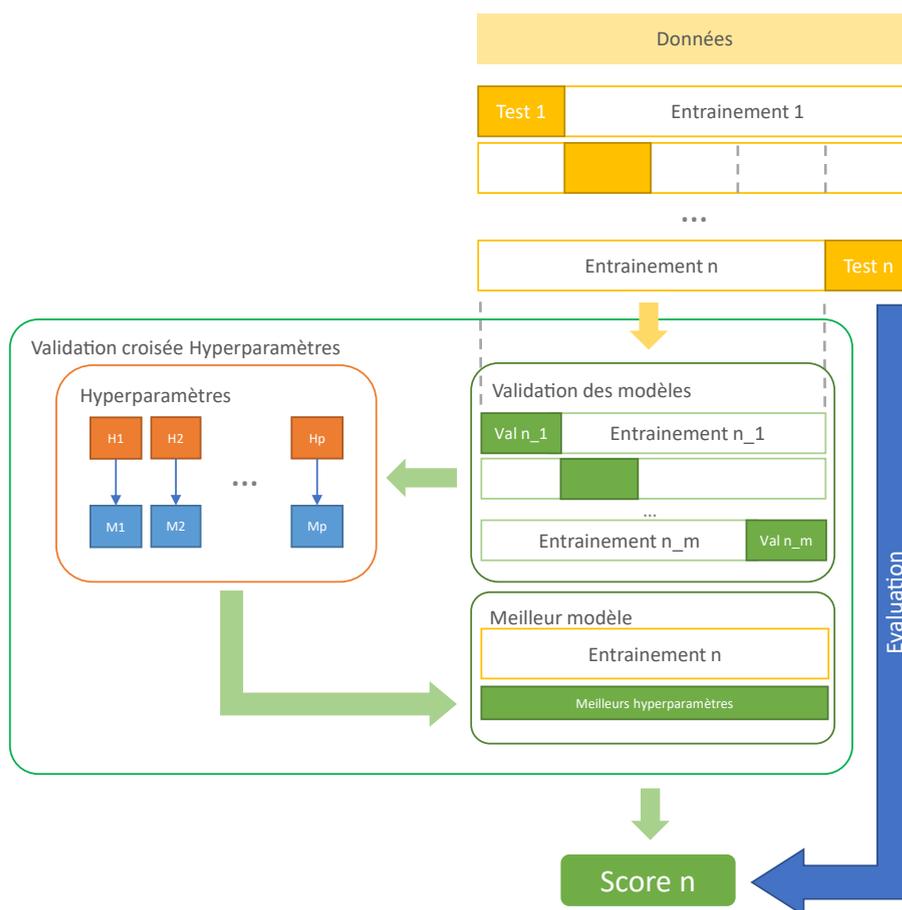


Figure 3.14 : Schéma de la validation croisée imbriquée assurant le choix de la meilleure combinaison d'hyperparamètres et assurant une évaluation robuste des performances du modèle.

3.5/ FUSION D'INFORMATIONS

Dernier aspect de ce sujet, la fusion d'informations se définit comme l'action de « combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision » [17]. Cette fusion se retrouve dans de nombreux domaines d'application dont celui de la santé ou encore de la biométrie, et peut être considérée comme un moyen d'augmenter la fiabilité de systèmes. L'intérêt croissant pour ce domaine correspond, entre autres, à une réponse à l'augmentation et à la diversification du nombre de capteurs présents. L'ouvrage de Bloch [17] justifie la fusion d'informations, par différentes notions auxquelles est soumis chaque capteur, dont :

- **l'incertitude** : c'est-à-dire son « degré de conformité avec la réalité ».
- **l'imprécision** : définie comme une « mesure du défaut quantitatif de connaissance, sur une mesure ».
- **l'incomplétude** : correspond à « l'absence d'information d'une source sur des aspects du problème ».

Cette fusion est indispensable pour permettre la réalisation de certaines tâches. L'effet de McGurk est un exemple concret de la *complémentarité* de la vision et de l'audition dans le rôle de la perception de la parole, mais également de la difficulté à gérer les *conflits* d'informations sur ces deux sources séparées [134]. L'ouvrage de Bloch [17] définit trois termes propres à cette fusion :

- **le conflit** : l'une des notions pouvant conduire à une mauvaise interprétation du système lorsque plusieurs informations contradictoires sont reçues.
- **la redondance** : l'apport multiple d'une même information, pouvant dans l'idéal permettre de réduire les incertitudes et les imprécisions.
- **la complémentarité** : généralement des sources, liée à un apport de caractéristiques différentes permettant l'accès à une information globale et de lever des ambiguïtés.

Il est nécessaire également de caractériser le résultat de la fusion d'informations. Il est relativement difficile d'obtenir une vision commune sur les divers niveaux et termes associés. Ainsi, ce travail se réfère à l'une d'entre elles [47] proche de cette problématique. Trois niveaux de fusion y sont décrits :

- le **niveau bas**, considéré comme étant la fusion sur les données,
- le **niveau moyen**, considéré comme la fusion des caractéristiques ou données plus abstraites,
- le **niveau haut**, considéré comme la fusion des décisions.

Pour ces différents niveaux, le processus de fusion d'informations peut aboutir à : (i) une sélection, (ii) une transformation, (iii) une extraction, (iv) ou encore une classification de l'information. La Figure 3.15 propose une schématisation macroscopique de ce concept.

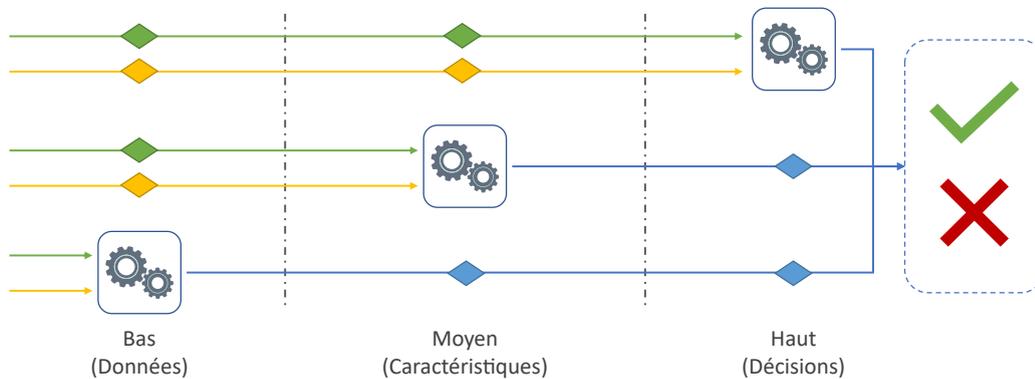


Figure 3.15 : Schéma des différents niveaux théoriques de fusion sur un schéma multimodal à deux capteurs. Le niveau bas représente le niveau le plus proche des données et des capteurs, le niveau moyen représente le niveau au plus proche de l'extraction de caractéristiques avancées, enfin le niveau haut représente le niveau de prise de décision.

La fusion à **bas niveau** s'exerce au niveau du capteur, quand celui-ci bénéficie de plusieurs relevés d'informations physiques, ou après l'acquisition de données provenant de plusieurs capteurs. Cela va nécessiter une cohérence entre ces modalités : un référentiel, un repère commun permettant la mise en parallèle de ces dernières. La fusion à ce niveau consiste essentiellement à fournir une cohérence de ces modalités d'information, voire à réduire cette dernière afin de n'obtenir que celle jugée pertinente dans le cadre de son utilisation.

La fusion à **moyen niveau** s'exerce au niveau de caractéristiques, c'est-à-dire après un premier traitement de chacune des modalités pour n'extraire que les éléments nécessaires à la réalisation du processus. Il s'agit d'une mise en commun des vecteurs d'information dans leur totalité ou réduite, nécessitant dans ce second cas une sélection de l'information pertinente.

Enfin, la fusion à **haut niveau** s'exerce après les mécanismes de prise de décision. C'est-à-dire que les branches correspondant à chaque modalité exercent indépendamment leur prise de décision. Il est donc nécessaire de définir une stratégie permettant de privilégier une décision ou d'extraire une décision à partir de l'information. Deux catégories sont ainsi identifiables : (i) la **sélection dynamique de modèle** de prédiction dans laquelle l'objectif est de définir le modèle pertinent, (ii) et la **fusion de modèles** de prédiction dans laquelle le but est d'obtenir une décision globale tenant compte des différentes décisions.

Plus particulièrement, dans le cadre de problématiques binaires, la fusion de modèles comporte deux niveaux conceptuels de décision basés sur le type d'information mis à disposition par ces modèles. D'une part, le plus haut niveau reprend l'idée d'une fusion appliquée au **niveau décisionnel**. Seules les décisions issues de chaque modèle sont alors considérées selon diverses stratégies : le vote à la majorité ou encore la pondération des votes en sont deux représentants. D'autre part, le plus bas niveau reprend l'idée d'une fusion appliquée aux scores [99]. Dans le cadre de problématiques à plusieurs classes, il est possible d'évoquer une troisième prise de décision au niveau du rang, c'est-à-dire en considérant la position de chaque classe prédite issue des divers modèles.

DONNÉES DE TRAVAIL ET MÉTHODES CLINIQUES

Lors des précédents chapitres ont été présentés les trois domaines sollicités par ce travail. Ce nouveau chapitre se consacre à la présentation des données utilisées pour évaluer ces travaux et procède au recensement des méthodes cliniques sur les modalités concernées.

Dans un premier temps, la *matière* utilisée afin de mener cette recherche est détaillée dans la Section 4.1. En effet, les expérimentations et résultats de ce manuscrit prennent appui sur une base de données initialement conçue pour la réalisation d'une étude clinique. D'une part, cette étude clinique est décrite point par point en reprenant ses objectifs, les conditions d'inclusion permettant la constitution de la base de données, mais également les conclusions sur lesquelles ce travail s'appuie. D'autre part, une structuration de cette base de données est nécessaire ainsi que la mise à disposition d'une interface logicielle permettant l'accès à ses informations, présentées à la fin de cette section.

Dans un second temps, la Section 4.2 est dédiée à l'analyse de critères pertinents sur les trois modalités cliniques mises à disposition de ce travail. L'objectif de cette section est de comprendre les éléments importants utilisés par les dermatologues lors d'un examen clinique.

Sommaire

4.1	Données de travail	58
4.1.1	Présentation des données	58
4.1.2	Structuration des données	61
4.1.3	Interface logicielle	62
4.2	Méthodes cliniques de diagnostic	63

4.1/ DONNÉES DE TRAVAIL

Pour commencer, il est nécessaire de souligner que **cette base a reçu un agrément de la part du comité d'éthique du Centre Hospitalier Universitaire de Saint-Étienne, (CHUSTE)** pour son exploitation au sein de l'étude clinique mentionnée précédemment mais également pour ce travail universitaire (**numéro du comité d'examen institutionnel 672016/CHUSTE**). Par ailleurs, cette étude **est conforme aux recommandations de la déclaration de Helsinki**.

Par cette section, divers aspects de ces données de travail sont présentés. Dans un premier temps, une présentation de sa constitution et des données recueillies est amenée par la Section 4.1.1. Puis dans un second temps, les aspects liés à son organisation sont présentés dans la Section 4.1.2. Pour finir, cette section s'achève par la description de l'interface logicielle permettant l'interaction avec ces données dans la Section 4.1.3.

4.1.1/ PRÉSENTATION DES DONNÉES

Cette base de données compile diverses **lésions faciales** possédant les critères d'inclusion suivants :

- les images proviennent de patients inclus entre les **années 2011 et 2015**, dont les données ont exclusivement été acquises au **Centre Hospitalier Universitaire de Saint-Étienne, (CHUSTE)**,
- les données sont disponibles pour chaque patient sous trois modalités d'image différentes que sont la **photographie clinique**, la **dermatoscopie** et la **MCR**,
- les lésions considérées dans cette base sont celles dont le diagnostic différentiel, c'est-à-dire par élimination méthodique des causes (voir Section 1.2.2.4), était **fortement controversé et supposé comme appartenant à un LM ou LMM**.

Les acquisitions sont réalisées par l'un des trois experts investigateurs de l'étude clinique [40] et **tout cas de collisions de tumeurs au sein est exclu de cette étude**. En termes de composition, la base regroupe 201 patients répartis entre 96 femmes et 105 hommes d'un âge moyen égal à 70 ans compris entre 29 et 97 ans comme présenté sur la Figure 4.1. Cette base comporte 223 lésions uniques avec l'histologie pour diagnostic de référence. Ainsi, sont mis à la disposition de ce travail :

- 137 **lésions malignes**, scindée en 92 **LM**, 23 **LMM**, 20 **CBC** et 2 autres,
- 86 **lésions bénignes**, dont 37 **LS**, 23 **KS**, 15 **KAP**, 8 **nævus** et 3 autres.

D'une part, les données de dermatoscopie sont produites à l'aide d'une caméra *PowerShot® G7*¹ couplée au dispositif proposé par *Fotofinder*². D'autre part, les données de **MCR** sont produites à l'aide d'un *VivaScope 3000®*³. En revanche, aucune information liée à l'acquisition des données de photographie clinique n'a été mentionnée.

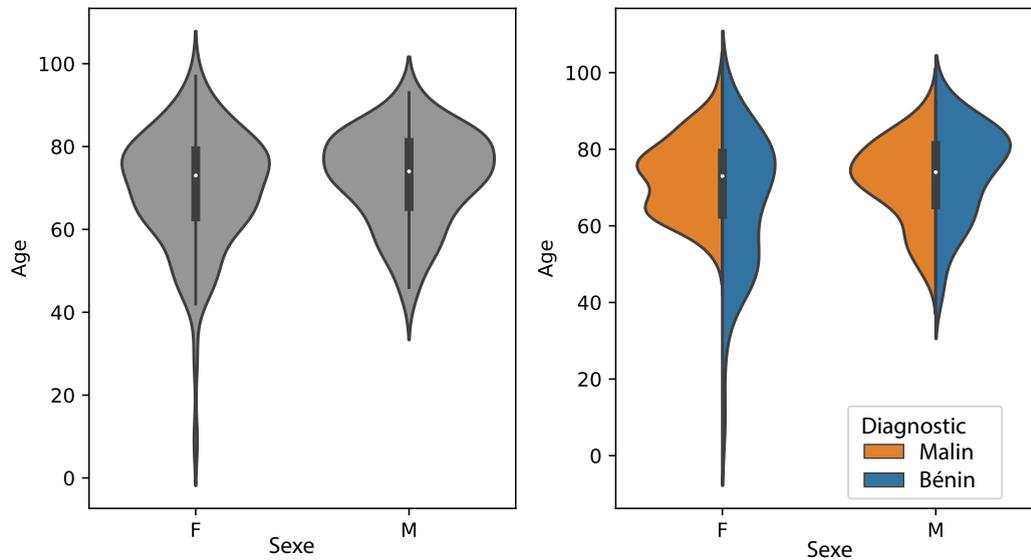


Figure 4.1 : À gauche, répartition en fonction de l'âge et du sexe, tous diagnostics confondus. À droite, répartition entre l'âge et le sexe en tenant compte du diagnostic binaire.

L'intérêt de ces trois modalités est grand dans le contexte de la dermatologie, puisqu'elles représentent le processus *classique* de prise en charge clinique, de la modalité la moins onéreuse avec la **photographie clinique** mais également la moins précise, à des modalités plus onéreuses, également plus riches en information, avec la **MCR**. Ainsi, chaque modalité permet au praticien de réduire, **si nécessaire**, la zone d'incertitude chez un patient souffrant d'une pathologie de la peau. Ce processus type de prise en charge peut être visualisé sur la Figure 4.2.

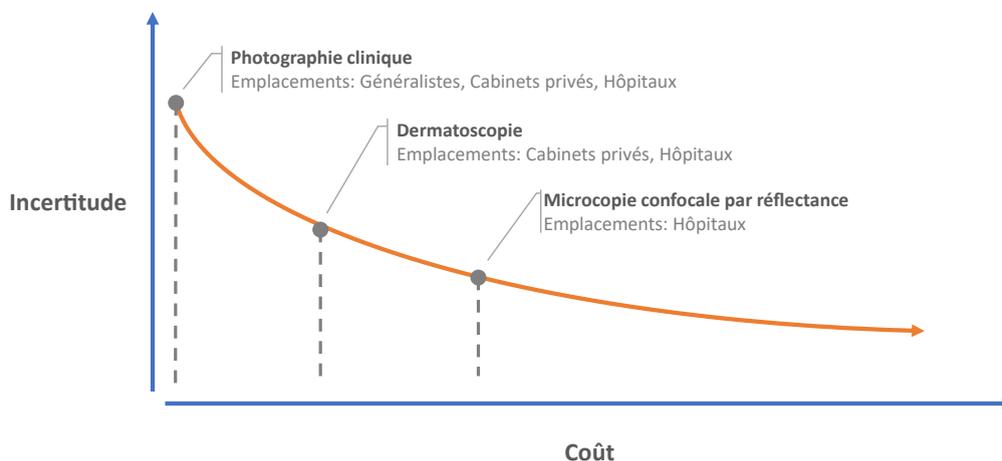


Figure 4.2 : Schéma de représentation du processus type de dermatologie, allant de la modalité la plus incertaine mais également la moins onéreuse, à la modalité la plus précise mais également la plus coûteuse.

Afin d'évaluer les performances de praticiens face à ces lésions, les investigateurs de l'étude de Cinotti et al. [40] ont eu recours à 21 dermatologues détenant une expertise des modalités d'imagerie non-invasives. Ces experts sont répartis de manière homogène selon leurs compétences respectives pour chacun des dispositifs. Ainsi, le panel d'évaluation se décompose de la façon suivante : (i) 6 experts sont soumis à l'ensemble des modalités, (ii) 15 experts sont soumis à l'évaluation de la photographie clinique et de la dermatoscopie (dont 9 uniquement dédiés à ces deux modalités), (iii) et 12 sont soumis à la MCR (dont 6 uniquement dédiés à la MCR) [40]. La répartition de ces experts est résumée en Figure 4.3.

Ces experts ont évalué la gravité des lésions selon les termes **bénin** et **malin** pour chaque cas clinique présenté. Afin d'éviter un biais de la part des experts soumis à l'évaluation des trois modalités différentes, ceux-ci ont été sollicités sur trois journées séparées : (i) une première journée d'évaluation ne comprenant que les images de photographie clinique et de dermatoscopie, (ii) une seconde journée d'évaluation ne comprenant que les images de MCR, (iii) et une dernière journée d'évaluation sur l'ensemble de la base d'images. Par ailleurs, les cas cliniques ont été présentés dans un ordre différent entre chaque journée d'évaluation afin d'éviter des biais liés à l'activité précédente des experts.

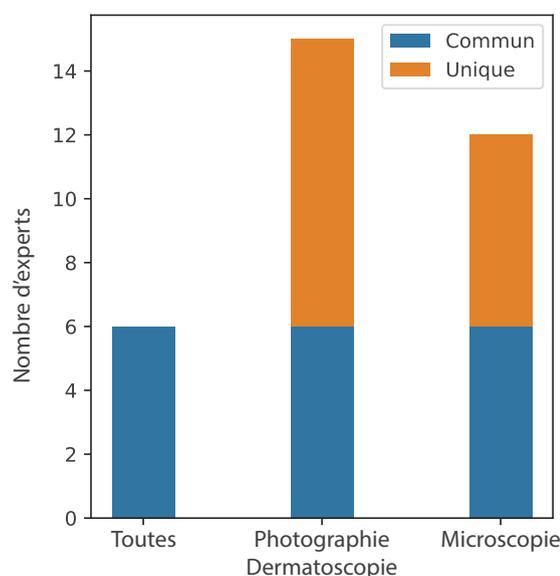


Figure 4.3 : Répartition du panel de 21 dermatologues sur l'évaluation des diverses modalités [40].

Néanmoins, quelques modifications sont à souligner par rapport à ces données initiales. Tout d'abord, une lésion non incluse faute de diagnostic histologique au moment de l'étude clinique est intégrée à la base de données employée amenant à **un total de 224 lésions** dans le cadre de ce manuscrit. De plus, certaines lésions initialement considérées comme bénignes par l'analyse histologique se sont avérées malignes suite à des contrôles annuels ultérieurs. Ainsi, ce manuscrit favorise le diagnostic le plus juste au moment de l'écriture, menant à une base de données comportant en définitive **138 lésions malignes et 86 lésions bénignes**.

¹Source : Canon Powershot®, Canon, Tokyo, Japon.

²Source : FotoFinder Systems GmbH, Bad Birnbach, Allemagne.

³Source : Distribué en Europe par MAVIG GmbH, Munich, Allemagne.

4.1.3/ INTERFACE LOGICIELLE

Malgré cette structuration de la base de données, celle-ci ne permet pas son exploitation en l'état. Ainsi, un second aspect de la gestion de cette base de données a consisté à mettre en oeuvre une interface logicielle permettant d'accéder à celles-ci.

Les données de cette base au format image représentent un volume non négligeable impossible à charger en mémoire vive avec la capacité à disposition de ce travail. Afin de ne pas excéder cette capacité, les données images sont chargées et maintenues en mémoire sous forme d'emplacement sur le disque dur jusqu'à leur utilisation définitive. Diverses méthodes d'accès à ces données sont mises à disposition, dont :

- une méthode `get_images()` avec pour objectif de retourner une structure de données sous forme de table contenant les méta-données et références aux images.
- une méthode `get_sliding_window()` avec pour objectif de retourner une structure de données sous forme de table contenant les méta-données et références pré-extraites par balayage d'une fenêtre glissante *en zone temporaire*.
- une méthode `get_multiple_resolution()` avec pour objectif de retourner une structure de données sous forme de table contenant les méta-données et références pré-extraites par redimensionnement des images *en zone temporaire*.

Chacune de ces méthodes met à disposition des arguments permettant de choisir d'une part la modalité ou encore le type d'annotations désiré et d'autre part des paramètres complémentaires liés au type d'extraction pour les principes de *fenêtre glissante* et de *résolutions multiples*. Cette interface logicielle mise à disposition peut être visualisée sur la Figure 4.5. Par ailleurs, les structures de données renvoyées par ces méthodes sont générées à l'aide de la librairie *Pandas* [163].

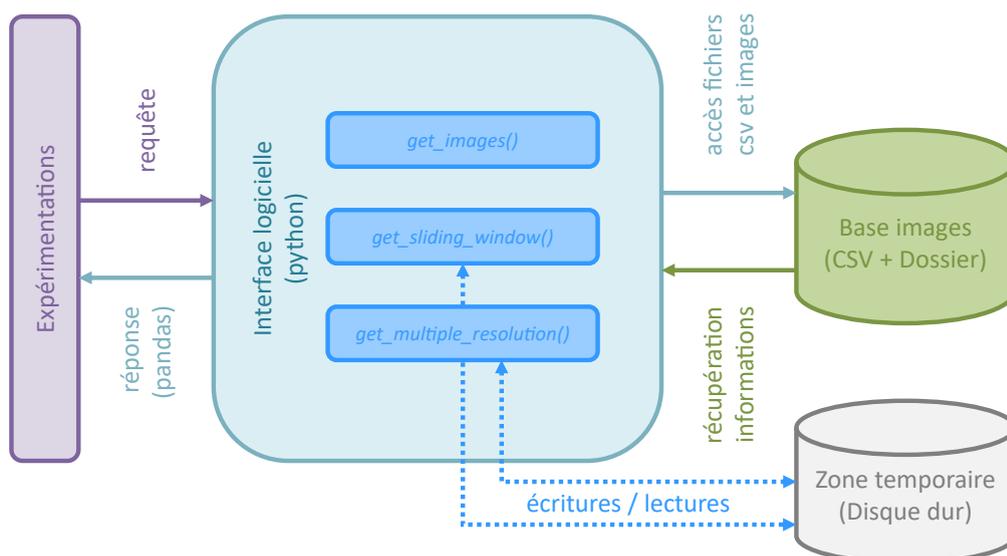


Figure 4.5 : Schéma de représentation de l'interface logicielle mise en place pour permettre les divers accès à la base de données employée dans ce manuscrit. Par cette interface, trois méthodes sont mises à la disposition des expérimentations qui suivent.

4.2/ MÉTHODES CLINIQUES DE DIAGNOSTIC

Pour permettre l'identification de pathologies en dermatologie plus routinière, la mise en avant de critères de diagnostic est une démarche clé mise en place pour rendre cette tâche plus efficace. Ces méthodes ont démarré au milieu des années 1980 pour améliorer la prise en charge du mélanome, sous la forme :

- **d'acronyme** : l'idée est de proposer une méthode de diagnostic facile à mémoriser. Le précurseur de ces méthodes est le travail de Friedman, qui propose l'acronyme **ABCD** pour une observation à l'oeil nu de lésions pigmentées et est utilisable par de nombreuses personnes expertes ou non [70]. Ainsi, cette première version insiste sur l'importance de quatre critères, dont : l'**A**symétrie, la **B**ordure (irrégularités), la **C**ouleur (variations) et le **D**iamètre (supérieur à 6mm). Cet acronyme s'est étendu à un cinquième critère d'**É**volution [2].
- **de liste** : de critères positifs et négatifs, correspondant à des caractéristiques rédhitoires ou non, et amenant à un score final. Ce score varie selon la pathologie considérée et permettra d'en déduire une décision à prendre. L'un des exemples dans le cadre de lésions pigmentées est celui de MacKie en 1986 [129]. Ces critères ont été révisés et se destinent néanmoins à un public expert.

Ces démarches d'identification de critères clés pour parvenir à un diagnostic se sont progressivement étendues à une multitude de pathologies et de dispositifs d'analyse médicaux. Les pathologies de **LM** et **LMM** ne font ainsi pas exception à ce principe.

Dans certains travaux, l'utilisation de la dermatoscopie a permis une prise en charge précoce là où la photographie clinique n'y parvient pas [180]. C'est donc naturellement que des critères sont recherchés sur la modalité de dermatoscopie, afin d'aider le praticien à diagnostiquer cette pathologie. Des caractéristiques propres aux pathologies de **LM** sont ainsi mises en avant telles qu'une importante densité vasculaire, des structures rhomboïdes rouges, des motifs en forme de cibles ou encore un assombrissement lors de l'examen par dermatoscopie [157]. Ces critères sont visualisables sur la Figure 4.6. Un travail de recherche reprend les critères précédemment cités et propose un ensemble de critères positifs et négatifs relatés dans la Table 4.1. En déterminant un seuil de score d'une valeur de 1, ce travail parvient à détecter les pathologies de **LM** avec une sensibilité de 0,95 et une spécificité de 0,75 sur une base de lésions contenant 42 **LM** et 123 lésions bénignes [146].

Type de caractéristiques	Caractéristiques
Positives Mineures (+1 points)	Assombrissement par observation avec dermatoscope
	Présence de cercles gris
	Motifs en forme de cible
	Présence de rhomboïdes gris
Négatives Mineures (-1 points)	Présence de points / globules gris ardoise
	Présence de cercles blancs
	Zones d'hyperkératose
	Présence de rhomboïdes rouges

Table 4.1 : Caractéristiques observables par dermatoscopie jugées pertinentes pour la détection du **LM** [146].

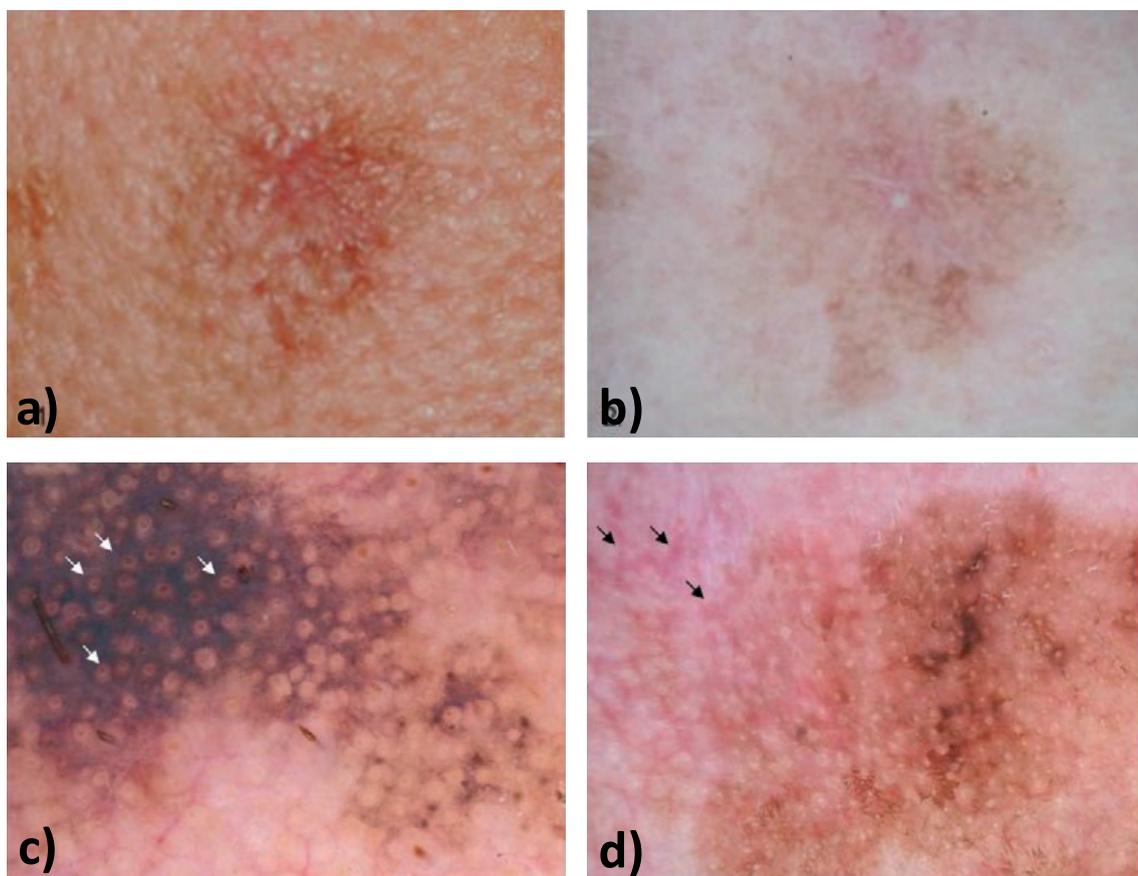


Figure 4.6 : Exemple de cas de données cliniques de dermatoscopie par Pralong et al. [157]. En a) et b), exemple de phénomène d'assombrissement sur un cas de **LM** ; en c), exemple de motifs en forme de cible issue d'un **LM** ; en d), exemple de structure rhomboïdales rouges sur un cas de **LMM**.

La modalité de **MCR** n'est pas exempte de ces démarches simplifiées, notamment pour le diagnostic du **LM** avec les travaux de Pellacani et al. [152] et de Guitera et al. [75]. Ces deux articles proposent sous forme de table de critères ceux jugés positifs et négatifs, dont la Table 4.2 fournit un aperçu. Leur étude comptabilise 81 cas de **LM** et 203 lésions bénignes, sur lesquels une valeur de seuil de 2 a été jugée comme pertinente au regard de l'étude statistique menée. Celui-ci a permis de détecter avec 85 % de sensibilité et 76 % de spécificité les **LM**.

Type de caractéristiques	Caractéristiques
Positives Majeures (+2 points)	Papilles à contour faiblement défini
	Cellules pagétoïdes rondes >20 µm
Positives Mineures (+1 points)	Présence de cellules atypiques dans la JDE
	Combinaison de follicule / cellules atypique / pagétoïde
	Présence de cellules nucléées au sein de papilles
Négatives Mineures (-1 points)	Motif en alvéoles larges

Table 4.2 : Caractéristiques observables par **MCR** jugées pertinentes pour la détection du **LM** [75].

Ainsi, une caractéristique a été déterminée comme pertinente par leurs auteurs lorsque celle-ci s'exprimait de manière plus significative dans l'un des deux groupes pathologiques (**LM** ou bénin). L'un de ces travaux [75] procède à une analyse de ces caractéristiques par profondeur croissante de la peau, dont peuvent être cités les éléments majeurs suivants :

- au niveau de l'**épiderme**, les auteurs ont constaté que les pathologies de **LM** comportent un désordre des cellules de l'épiderme (56 % des lésions **LM** contre 18 % des lésions bénignes). Également, une infiltration pagétoïde a été reportée dans la plupart des cas de **LM** (75 % des lésions **LM** contre 28 % des lésions bénignes). En opposition, un épiderme homogène, caractérisé par des motifs en nid d'abeilles, a été constaté dans 92 % des cas bénins. Ces éléments peuvent être observés sur la Figure 4.7.
- au niveau de la **JDE**, les auteurs ont constaté que les papilles non démarquées étaient observées dans la majorité des **LM** (68 % **LM** and in 17 % des cas bénins). Ces éléments peuvent être observés sur la Figure 4.8.
- au niveau du **derme**, 15 % des cas de **LM** présentent des cellules nucléées larges contre 2 % des lésions bénignes. Ces éléments peuvent être observés sur la Figure 4.9.

Dans l'étude menée par Cinotti et al. [40], les auteurs investigateurs ont demandé d'évaluer certaines caractéristiques récurrentes de la littérature, parmi lesquelles :

- la présence de grandes cellules pagétoïdes arrondies, présentes dans 37 % des **LM** et seulement 5 % des pathologies bénignes,
- la présence de grandes cellules dendritiques dans l'épiderme, présentes dans 81 % des **LM** contre 13 % des pathologies bénignes,
- la localisation au niveau des follicules pileux des cellules atypiques, dans 62 % des **LM** et seulement 7 % des pathologies bénignes.

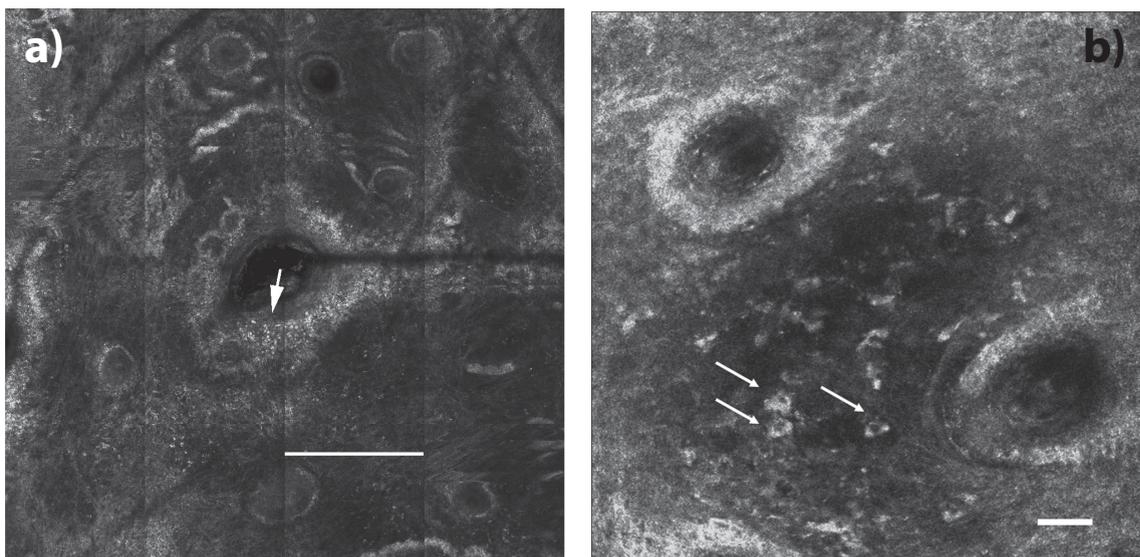


Figure 4.9 : Exemples de cas de données cliniques **MCR** par Guitera et al. [75]. En a), exemple de cellules atypiques à proximité d'un follicule pileux au sein de la **JDE** sur un **LM** ; En b), exemple de cellules nucléées du derme d'un **LM**. Repère : barre = 50 μ m.

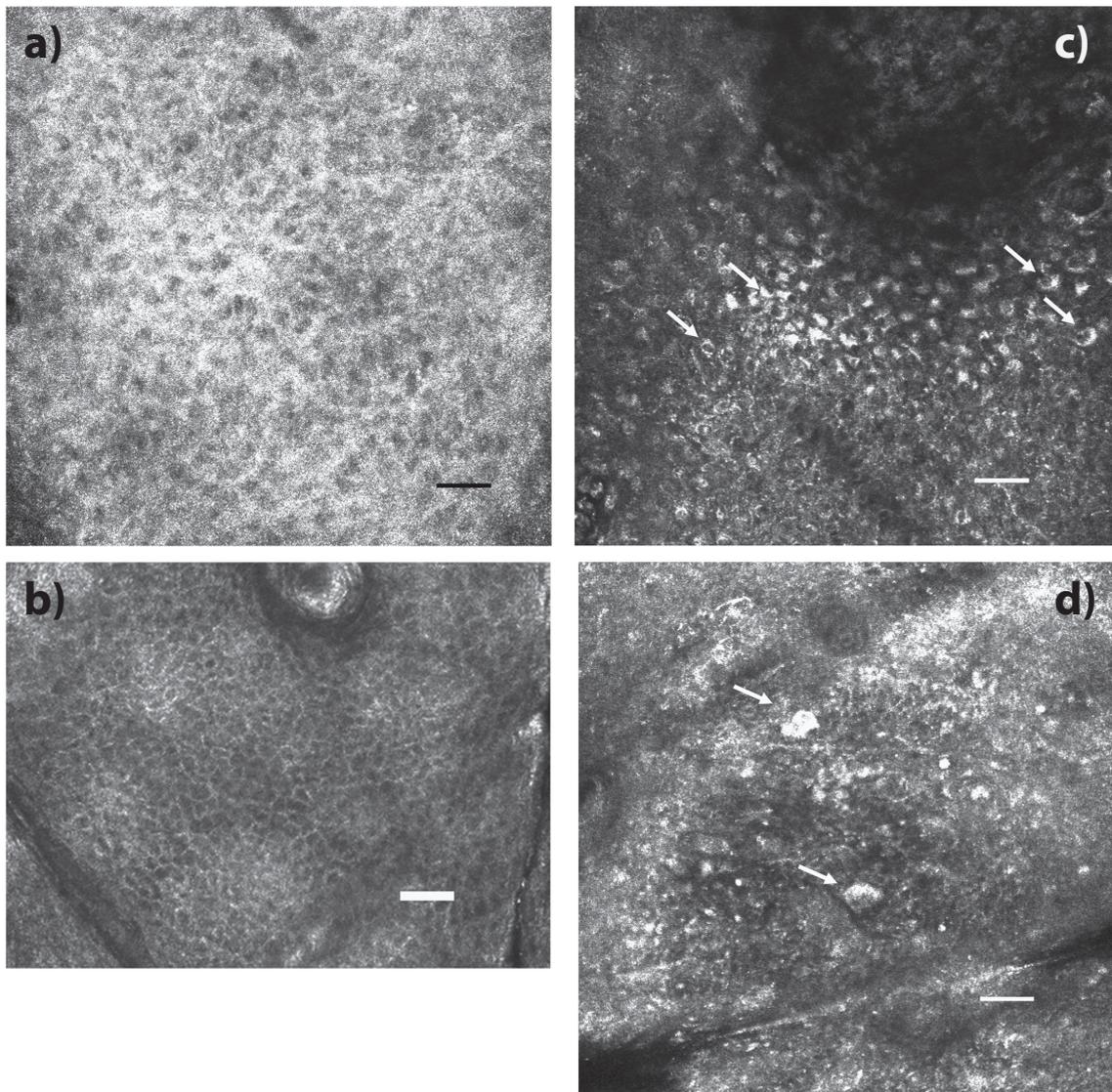


Figure 4.7 : Exemples de cas de données cliniques **MCR** en provenance de l'épiderme par Guitera et al. [75]. En a) et b), exemples de motifs en nid d'abeilles (*broadened honeycomb pattern*), respectivement d'une **KS** et de peau normale ; en c), exemples de motifs de pavés atypiques (*atypical cobblestone pattern*) issus d'un **LM** ; en d), exemples de *grandes cellules pagétoïdes* propres au **LMM**. Repère : barre = 50 μm .

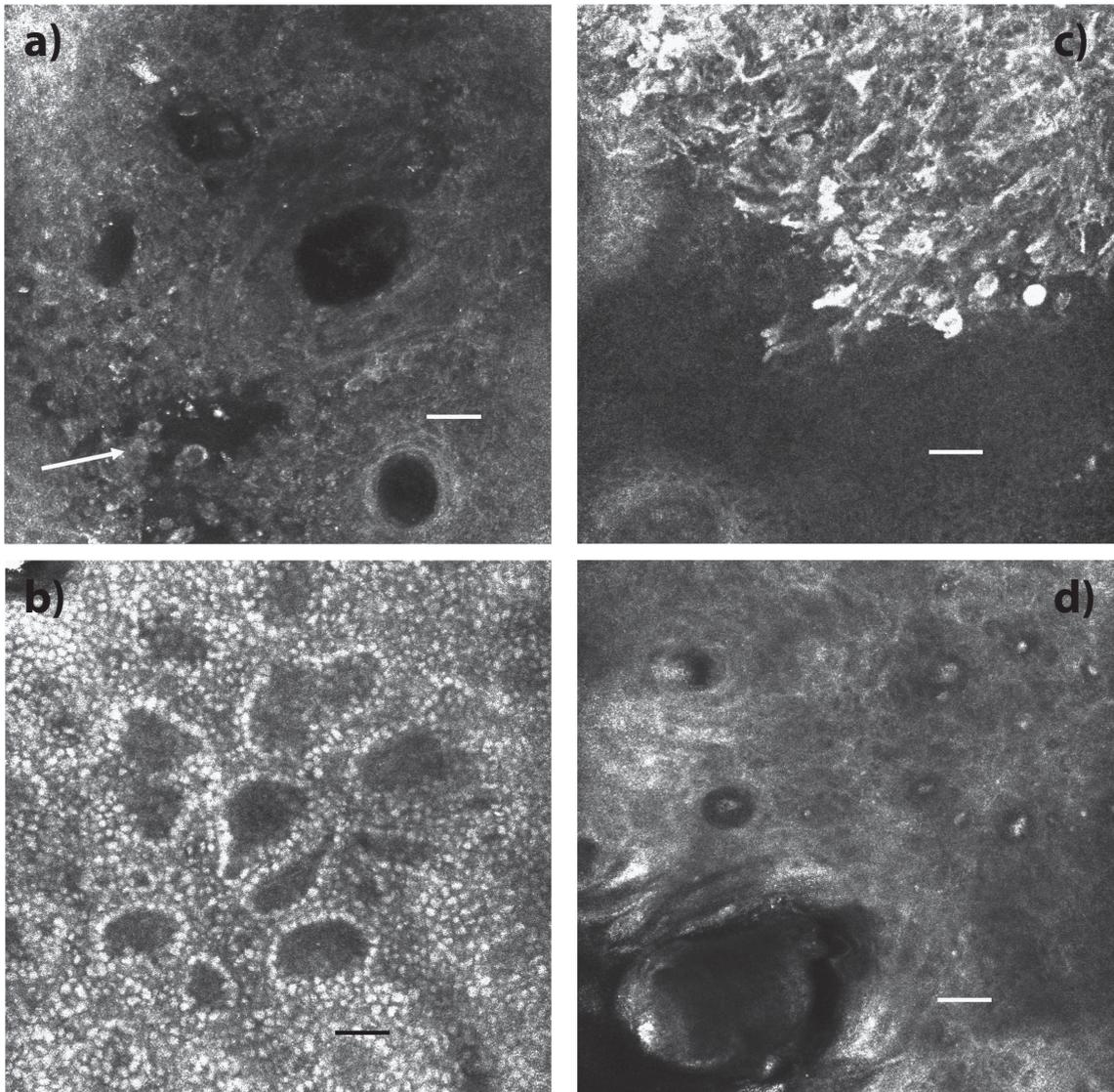


Figure 4.8 : Exemples de cas de données cliniques **MCR** par Guitera et al. [75]. En a) et au niveau de la flèche, exemple de papille à frontières non marquées (*Nonedge papillae*) et de cellules atypiques d'un **LM** acquise à la jonction **JDE** ; En b), exemple de papilles à frontières marquées (*Edge papillae*) d'une pathologie bénigne. En c), exemple de cellules atypiques au niveau de la **JDE** typique de **LM** ; En d), exemple de halo noir autour de cellules atypiques de l'épiderme issu d'un **LM**. Repère : barre = 50 μm .

II

ANALYSE DE LA MICROSCOPIE CONFOCALE PAR RÉFLECTANCE

PRÉAMBULE MICROSCOPIE CONFOCALE PAR RÉFLECTANCE

La précédente partie de ce travail consacrée au contexte, permet une mise en situation des connaissances mobilisées et une présentation des recherches liées à celui-ci. Au sein de cette seconde partie est élaborée une méthodologie destinée à la modalité **Microscopie Confocale par Réflectance**. Cette recherche est motivée par diverses raisons. En premier lieu, il s'agit de l'une des modalités les plus précises à disposition du spécialiste en condition clinique. Cette modalité représente une référence en matière de qualité de diagnostic à réaliser. En second lieu, cette recherche est motivée par le nombre restreint de travaux menés sur le diagnostic de lésion de la peau appliqué à la modalité de **MCR** et assisté par ordinateur. En effet, cette modalité représente un verrou scientifique qu'il est nécessaire de lever pour accomplir le schéma de diagnostic multimodal. Ce préambule a pour objectif de présenter d'une part les données de travail exclusives à la modalité de **MCR** introduites lors du Chapitre 4, et d'autre part de présenter les résultats des experts qui constituent la finalité des méthodes d'aide au diagnostic de cette partie.

Ainsi dans cette partie, seules les données **MCR** des 224 lésions mises à disposition sont considérées **pour l'évaluation** des méthodes et sont exclues les images de photographie clinique et de dermatoscopie. Pour rappel, ces données correspondent à des lésions pouvant porter à confusion lors du diagnostic de **LM** et de **LMM** par le spécialiste. En addition à ces données, 28 patients bénins non présents dans la base d'images initiale sont utilisés afin d'augmenter le nombre d'images annotées comme bénignes **pour l'entraînement** des méthodes développées.

Chaque lésion contient des images **MCR** jugées pertinentes par deux des trois médecins investigateurs dont l'acquisition des images provient de la **JDE**. En effet, ces deux pathologies sont observables à ce niveau de profondeur :

- les **pathologies de LM** résultent d'une prolifération de cellules au niveau de la lame basale (extrémité supérieure de la **JDE**),
- les **pathologies de LMM** résultent de l'intrusion de la mélanine se propageant le long des follicules pileux (extrémité supérieure et inférieure de la **JDE**).

Ces images dont l'acquisition est réalisée au niveau de la **JDE** sont donc ainsi parfaitement adaptées à la caractérisation de ce phénomène par les médecins.

Du point de vue des caractéristiques, ces images arborent une taille constante de 1000 px × 1000 px pour une section mesurée de 920 µm × 920 µm et une résolution en profondeur comprise entre 3 µm et 5 µm. Ces données fournissent ainsi une résolution de 1 µm. Les informations d'acquisition (profondeur, longueur d'onde, ...) contenues dans les images ne pourront être utilisées car non exploitables (procédure de calibration non réalisée, informations manquantes, ...).

À partir de ces données, l'analyse menée par les 12 experts en **MCR** de l'étude [40] permet d'obtenir une valeur moyenne de sensibilité de $0,84 \pm 0,05$ et une valeur moyenne de spécificité de $0,75 \pm 0,06$ sur l'évaluation des pathologies malignes toutes confondues. Sur l'évaluation des données de **LM** et **LMM**, ces scores deviennent plus homogènes avec une valeur moyenne de sensibilité de $0,80 \pm 0,07$ et une valeur moyenne de spécificité de $0,81 \pm 0,05$. Les courbes **ROC** associées à ces performances sont représentées sur la Figure ii.1.

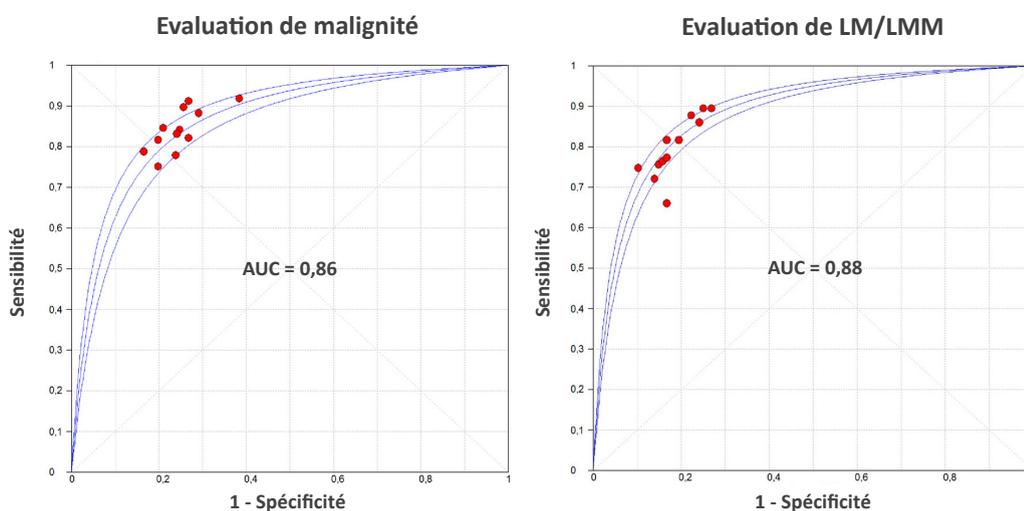


Figure ii.1 : Courbes **ROC** issues de l'évaluation des données **MCR** par les 12 experts [40]. À gauche, le résultat de l'analyse des experts, mené sur la détection d'éléments malins ; À droite, le résultat de l'analyse des experts, mené sur la détection de **LM** et **LMM**.

L'objectif visé par cette partie est d'égaliser ces scores à l'aide de méthodes de diagnostic sur base d'images issues de la **MCR**. À ce titre, les méta-données extraites des informations patient (telles que l'âge, le sexe ou encore l'emplacement de la lésion) ne sont pas exploitées, bien que l'âge ait été mis à disposition des experts en **MCR** durant leur évaluation. Ce choix est justifié car contrairement aux connaissances des experts, ces méta-données ne sont pas représentatives de la population réelle et peuvent induire des biais de corrélations sur des données extérieures à la base employée. Ainsi, les évaluations des méthodes ne portent que sur l'information fournie par les images de ces données **MCR**.

Ainsi, diverses ressources d'intelligence artificielle évoquées lors du Chapitre 3 sont mobilisées lors de ce travail. Des apports ont été réalisés pour permettre le déroulement des expériences présentées lors des chapitres suivants. En effet, la base d'images actuelle ne possède que des annotations au niveau des patients, ce qui limite fortement l'étendue des travaux essentiellement basés sur des méthodes d'apprentissage supervisé. À l'aide d'un outil graphique réalisé pour cette étude (Figure ii.2) et grâce au travail de l'un des dermatologues investigateurs, divers niveaux d'annotations supplémentaires sont obtenus, dont :

- des annotations au **niveau des images** correspondant au plus petit niveau issu des données originelles,
- des annotations au **niveau de sous-parties d'images** également appelés *patches* correspondant à l'extraction d'un nouveau niveau de données à partir des images.

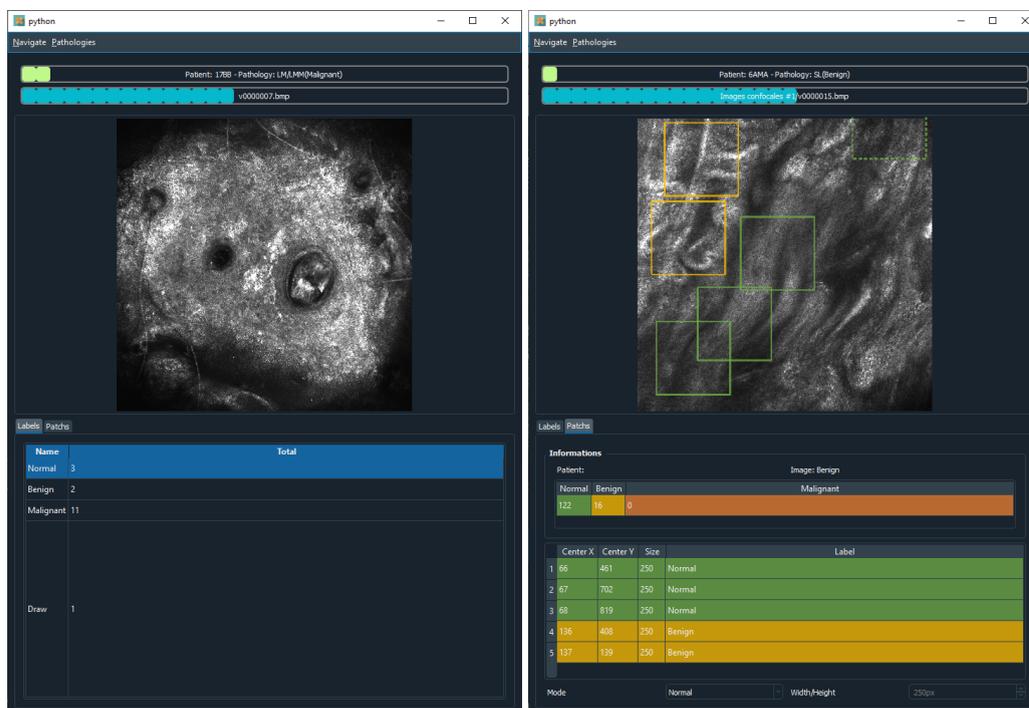


Figure ii.2 : Interface logicielle mise à disposition du spécialiste, afin de procéder aux annotations au niveau des images et des sous-images. À gauche, l'onglet permettant l'annotation au niveau des images ; À droite, l'onglet permettant la création et l'annotation de sous-images.

Les annotations au niveau des lésions bien que initialement sur deux classes, donnent naturellement lieu au niveau des images à des annotations sur **trois classes**. Ces annotations, dont des exemples sont visibles sur la Figure ii.3, sont les suivantes :

- Un premier type d'annotation du nom de *malin* a été apposé *sur des images* contenant **au moins** du tissu de type malin et *sur des sous-images* contenant essentiellement celui-ci.
- Un second type d'annotation du nom de *bénin* a été apposé *sur des images* ne contenant **aucun** tissu malin mais **au moins** du tissu bénin et *sur des sous-images* composées de celui-ci.
- Enfin, un troisième type d'annotation du nom *sain* a été renseigné *sur des images* ne contenant **aucun** des deux types de tissus précédents et *sur des sous-images* composées uniquement de tissus sains.

La répartition complète de cette base annotée est disponible sur la Figure ii.4 dans laquelle peut être visualisée la composition des annotations de la base d'images au niveau des **lésions**, des **images** et des **sous-images**. D'une part, est visible la répartition des données de la *base de référence* ou *base initiale* [40] de laquelle est issue l'**évaluation des experts** mais également servant à l'**entraînement et l'évaluation des méthodes** qui suivent. D'autre part, est visible la répartition des *données additionnelles* qui servent à **corriger le non-balancement lors de la phase d'entraînement**.

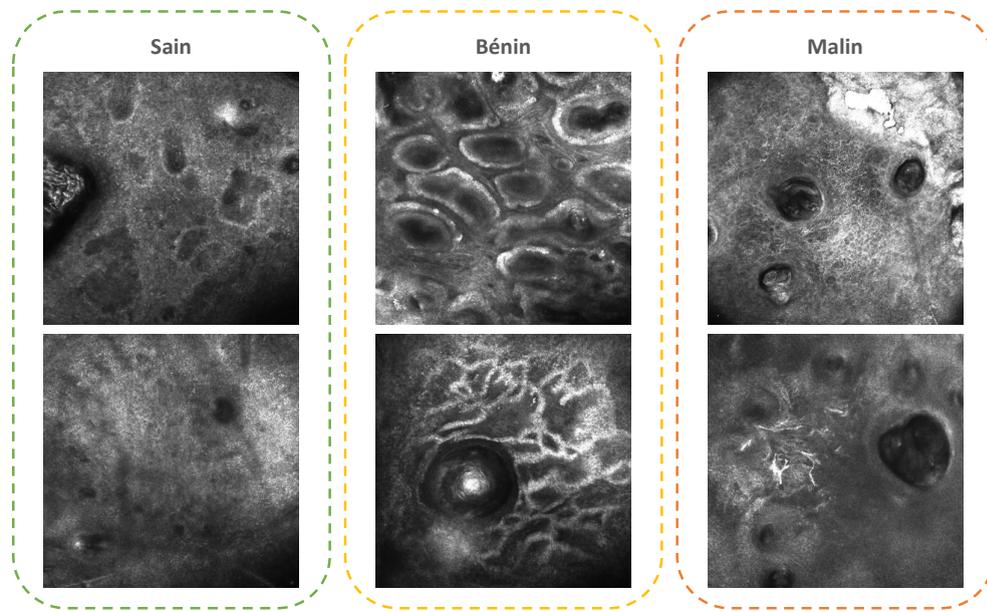


Figure ii.3 : Exemple de tissus et annotations, considérés respectivement de gauche à droite comme *sain*, *bénin* et *malin* par le spécialiste.

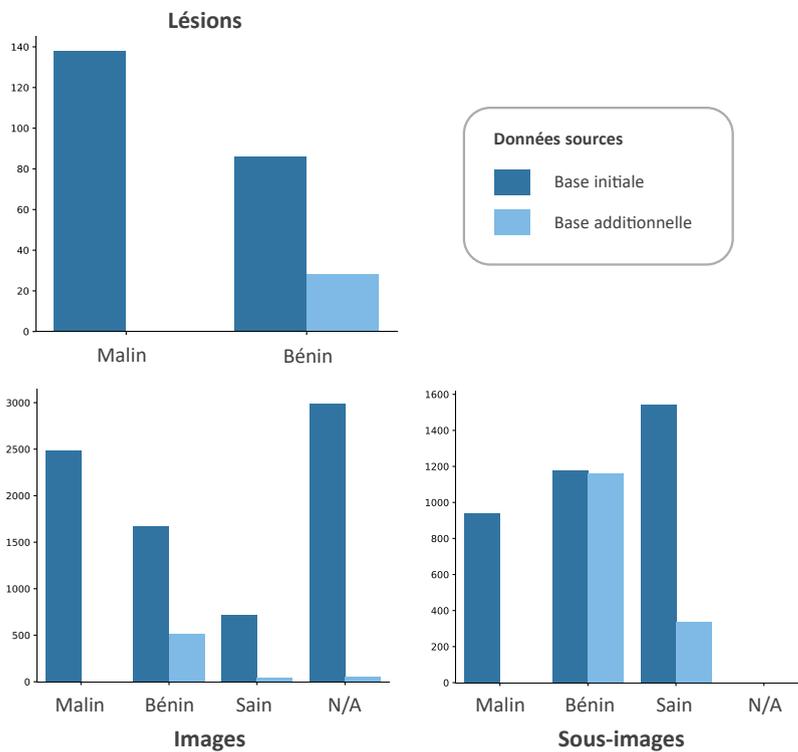


Figure ii.4 : Répartition des annotations réalisées par l'expert sur les lésions, les images et les sous-images. En bleu foncé, la base initiale de Cinotti et al. [40] ; En bleu clair, les données additionnelles.

Afin de répondre à ce besoin de diagnostic au niveau des lésions acquises par **MCR**, cette partie dédiée à cette modalité dans laquelle est développée une méthodologie ascendante se compose de trois chapitres :

- dans un premier temps, le Chapitre **5** aborde la **prise de décision au niveau de l'image** par des schémas simples,
- dans un second temps, le Chapitre **6** propose des méthodes destinées à **améliorer ce diagnostic image** par la mise en oeuvre de schémas avancés permettant à nouveau la prise de décision sur les images **MCR**,
- enfin, le Chapitre **7** se focalise sur une **prise de diagnostic au niveau du patient** par la présentation de processus décisionnel au niveau lésionnel.

DIAGNOSTIC DE L'IMAGE

Comme évoqué durant le préambule, cette partie se veut comme une première étape au processus global de classification sur la modalité de **MCR**. En d'autres termes, ce chapitre regroupe diverses méthodes relatives à la classification des images de cette modalité. Plus précisément, ces méthodes se consacrent à une séparation des images selon les annotations **saines**, **bénignes** ou **malignes**.

Ainsi, ce chapitre propose une approche en plusieurs étapes. La première d'entre elles se destine à l'extraction de caractéristiques pertinentes par des méthodes manuelles mais également auto-déterminées à l'aide de réseaux profonds pré-entraînés, tandis que la seconde explore la mise en œuvre de méthodes de classification éprouvées. Afin de rendre plus riche cette partie et de compléter ses conclusions, la dernière étape aborde divers aspects tels que l'influence de paramètres liés à la normalisation des caractéristiques, la réduction de l'information dans le contexte de méthodes à fort nombre de dimensions ou encore l'influence des données sur les résultats de classification.

Sommaire

5.1	Méthodologie	78
5.2	Méthodes d'extraction de caractéristiques	79
5.2.1	Méthodes manuelles d'extraction de caractéristiques spatiales	79
5.2.2	Méthodes manuelles d'extraction de caractéristiques fréquentielles	82
5.2.3	Méthode d'apprentissage par transfert de connaissances	85
5.2.4	Analyse préliminaire des méthodes d'extraction	88
5.3	Pré-traitements de caractéristiques	95
5.3.1	Réduction de dimensions	95
5.3.2	Normalisation de caractéristiques	96
5.3.3	Balancement de données	97
5.4	Processus de prédiction	99
5.4.1	Méthodes de prédiction	99
5.4.2	Procédure d'évaluation	100
5.5	Résultats	103
5.6	Discussion	107

5.1/ MÉTHODOLOGIE

La classification de lésions de la peau, telles que le mélanome, par l'utilisation de méthodes d'apprentissage est une problématique largement développée dans la littérature. Néanmoins, cette littérature se restreint à quelques articles de recherche [79, 78, 209, 100] lorsque ce cadre se centre en particulier sur la détection des pathologies de **LM** et **LMM** couplées à la modalité de **MCR**.

Dans ce travail, la détection de ces pathologies est réalisée à l'aide de la base d'images mise à disposition, essentiellement centrée autour de lésions malignes de **LM** et **LMM**. Divers angles d'approche sont ainsi menés, avec :

- d'une part sous la forme d'une problématique à trois classes, la séparation des tissus **sains**, **bénins** et **malins**,
- et d'autre part sous la forme d'une problématique binaire, la séparation entre les tissus **malins** et le **reste** des tissus.

Pour cela, ces deux situations sont abordées à l'aide de processus d'apprentissage supervisés dont les diverses étapes sont expliquées lors des prochains paragraphes. En effet, ces approches sont largement développées et démontrées fonctionnelles sur des thématiques variées du domaine de l'imagerie médicale et s'imposent pour la suite de ces travaux [119, 149]. Ainsi, les étapes essentielles des processus de classification *classiques* sont utilisées et adaptées à la problématique de ce manuscrit.

Le pré-traitement des images n'est pas considéré dans ce travail : d'une part les images **MCR** ne présentent que peu d'artefacts ou aberrations, et d'autre part cet aspect est faiblement traité dans la littérature. De ce fait, le processus de classification se compose de trois étapes majeures : (i) l'extraction de caractéristiques, (ii) le pré-traitement avant classification, (iii) et les méthodes de classification. La Figure 5.1 permet une vision globale de ce processus de classification. Ainsi, une section est dédiée à chacune de ces trois étapes, puis à l'aide de deux sections distinctes sont présentés les résultats de ces expérimentations et leur analyse.

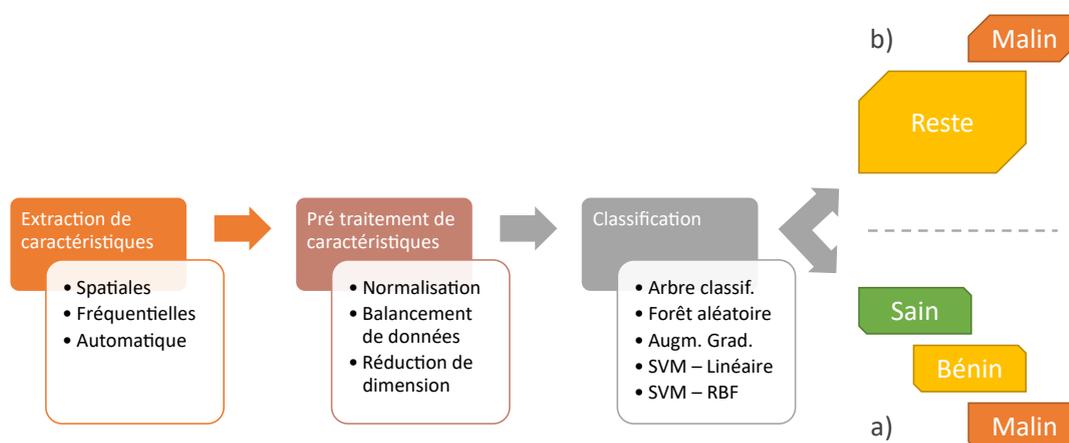


Figure 5.1 : Représentation du processus de diagnostic suivi dans le cadre de l'apprentissage automatique pour les images **MCR**. Ce processus permet une prédiction à trois classes - a), mais également binaire en se focalisant sur la détection de la classe maligne - b).

5.2/ MÉTHODES D'EXTRACTION DE CARACTÉRISTIQUES

L'étape d'extraction de caractéristiques consiste en l'obtention de valeurs dérivées des données brutes dans le but de quantifier un phénomène. Ces valeurs jugées pertinentes sont utilisées pour la suite du traitement, généralement de la classification. Les éléments présentés en préambule visent à montrer l'importance des motifs de tissus, dont l'apparence varie selon :

- **la profondeur de l'acquisition** du tissu considéré, ainsi l'épiderme, la **JDE** et le derme sont très différents du point de vue des tissus qui les constituent,
- **la pathologie à identifier**, les différentes lésions traitées dans ce manuscrit sont assez différentes selon leur état bénin ou malin.

Les nombreux échanges, menés auprès de dermatologues visant à l'introspection de leur propre processus cognitif lors de la décision, aboutissent à une forte présomption de l'importance de **la texture** des tissus présents dans ces images.

Les méthodes d'extraction de caractéristiques sont abordées selon deux axes : (i) d'une part, les **méthodes manuelles d'extraction de caractéristiques** ou déterminées par un humain respectivement séparées en caractéristiques *spatiales* et *fréquentielles*, (ii) d'autre part, les **méthodes non manuelles d'extraction de caractéristiques**. L'appellation *manuelle* s'est ainsi imposée suite à l'apparition de réseaux de convolution profonds, afin de distinguer ces deux types d'approches [140].

5.2.1/ MÉTHODES MANUELLES D'EXTRACTION DE CARACTÉRISTIQUES SPATIALES

L'extraction sur base d'information du domaine spatial est un champ assez développé présentant comme avantage d'être intuitif mais pouvant être affecté par le bruit et autres phénomènes de dégradation de l'information. Ces opérations spatiales peuvent être décrites à deux niveaux d'interprétation, d'une part celui des valeurs brutes dont les descripteurs de premier ordre font partie mais ne fournissent aucune valeur de cohérence spatiale et, d'autre part celui par transformation de l'information initiale dont les descripteurs de second ordre ou plus font partie. Ces dernières méthodes prennent en compte la relation d'une valeur locale à celle de son voisinage [95].

Les descripteurs de premier ordre s'appliquent ainsi aux valeurs d'intensité ou à la distribution de celles-ci, sans traitement préalable. La Table 5.1 recense la plupart de ces caractéristiques issues de la distribution des valeurs brutes. Néanmoins, ces mesures tendent à ressortir des phénomènes globaux de l'image, mais ne suffisent pas à elles seules à analyser des motifs présents [195, 179, 202, 144]. Ces mesures ont notamment été utilisées comme un complément de classification [209].

Mesure	Expression mathématique
Moyenne	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Variance	$v = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Entropie	$e = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$
Kurtosis	$k = \frac{r \sum_{i=1}^n (x_i - \bar{x})^4}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} - 3$
Asymétrie	$a = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right]$

Table 5.1 : Table des différentes mesures obtenues à partir du premier ordre.

Les descripteurs de second ordre, ou plus, utilisent l'information locale et en déduisent une relation à leur voisinage par l'utilisation d'une transformation intermédiaire. Parmi ces transformations, les plus courantes d'entre elles sont : (i) la **Gray Level Co occurrence Matrix**, (**GLCM**), (ii) la **Gray Level Run Length Matrix**, (**GLRLM**), (iii) et la **Gray Level Size Zone Matrix**, (**GLSZM**). L'extraction d'information de second ordre a notamment été traitée en 1973 pour de la reconnaissance sur des images aériennes et satellites, par Haralick et al. [80]. Ce travail est une extension des **GLCM**, dont le principe consiste en un recensement des combinaisons d'intensité présentes dans une image. Cette extraction nécessite une direction dans laquelle les combinaisons sont recherchées. En 2D, ces directions sont au nombre de quatre : (i) horizontale, (ii) verticale, (iii) et les deux diagonales opposées. Le principe de fonctionnement des **GLCM** est schématisé sur la Figure 5.2 pour une extraction selon la direction horizontale. Le travail de Haralick et al. [80] a ainsi démontré la pertinence des **GLCM** dans la différenciation d'images texturées, en calculant selon les quatre directions évoquées précédemment, quatorze critères statistiques. Ces divers critères statistiques sont recensés dans la Table 5.2, ainsi que leurs formules associées. La bibliothèque logicielle *Mahotas* [43] est utilisée afin de réaliser ces diverses extractions de manière optimale et dans un temps raisonnable.

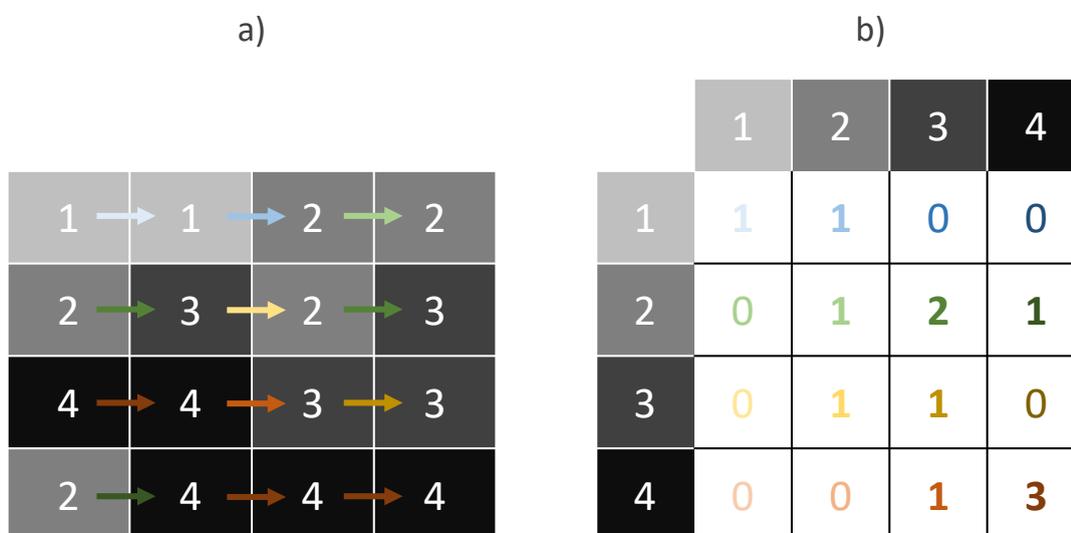


Figure 5.2 : Représentation de la création d'une matrice sur base de **GLCM** selon la direction horizontale et dont la valeur de distance de voisinage est de 1. En a), les flèches représentent les diverses relations de voisinage horizontal, caractérisées par une couleur ; En b), ces diverses relations sont recensées dans la matrice de **GLCM** en employant les mêmes codes de couleurs.

Les travaux menés par l'une des études sur les lésions mélanocytaires en **MCR** [209], complètent ces deux catégories de mesures spatiales par de l'information de texture. En effet, l'étude met en avant l'importance des douze premières caractéristiques formulées par Haralick [80] et adjoint cinq mesures de premier ordre comme complément d'information dont : (i) la moyenne, (ii) l'erreur quadratique moyenne, (iii) l'asymétrie de la répartition de l'histogramme, (iv) le kurtosis (mesure d'aplatissement de la distribution), (v) et l'entropie. Ce même travail met en avant que la texture des lésions de la peau ne possède pas d'orientation particulière. Afin de rendre ces caractéristiques plus robustes aux rotations et, de réduire la quantité d'information extraite, les auteurs préconisent l'utilisation d'une moyenne sur les quatre directions. Ainsi, cette manipulation permet de réduire l'information à un unique vecteur de taille 1×12 pour les caractéristiques d'Haralick et al. [209].

Mesure	Expression mathématique
Moment Angulaire	$\sum_i \sum_j p(i, j)^2$
Contraste	$\sum_{k=0}^{N_g-1} k^2 p_{x-y}(k)$
Corrélation	$\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Différence des Moments Inverse	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i, j)$
Entropie	$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \log(p(i, j))$
Somme - Moyenne	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
Somme - Variance	$\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$
Somme - Entropie	$\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$
Somme des carrés - Variance	$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i, j)$
Différence - Variance	variance of p_{x-y}
Différence - Entropie	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i))$
Mesure de Corrélation 1	$\frac{f_9 - HXY1}{\max(HX, HY)}$
Mesure de Corrélation 2	$[1 - \exp(-2(HXY2 - f_9))]^{1/2}$
Coefficient de Corrélation Maximal	$(\max(\sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}))^{1/2}$

Table 5.2 : Table des différentes mesures proposées par Haralick et al. [80] applicables aux GLCM pour caractériser des images texturées.

Les expériences menées sur base de descripteurs manuels spatiaux se décomposent en plusieurs temps. Tout d'abord, l'importance de l'information de **premier ordre** et de second ordre proposée par **Haralick** et al. [80] selon les *quatre directions* mais aussi la *moyenne* sont évaluées sur leur aptitude à séparer les divers groupes de données. Enfin, la méthode de **Wiltgen** et al. [209] qui propose la combinaison de ces deux précédentes techniques est évaluée en ces mêmes termes [209]. Par ailleurs, la Table 5.3 synthétise l'ensemble des méthodes d'extraction ainsi que le nombre associé de caractéristiques extraites.

Méthode	Nombre de caractéristiques
Mesures premier ordre	14
Haralick [80] - 4 directions	56
Haralick [80] - Moyenne	14
Wiltgen [209] - Spatial	17 (12+5)

Table 5.3 : Liste des méthodes spatiales évaluées dans ce chapitre et leur nombre de caractéristiques extraites associées.

5.2.2/ MÉTHODES MANUELLES D'EXTRACTION DE CARACTÉRISTIQUES FRÉQUENTIELLES

L'extraction sur base d'information du domaine fréquentiel est également un champ de travail assez développé, notamment pour sa robustesse face aux perturbations telles que le bruit, ainsi que la vitesse d'exécution de certaines opérations telles que le filtrage. En revanche, l'information extraite est plus difficile d'interprétation et, nécessite une information spatiale suffisamment conséquente pour être juste [95]. Ce dernier point est négligeable dans ce travail, les données de **MCR** exploitées dans le cadre de ces travaux étant dotées d'une définition spatiale de $1000 \text{ px} \times 1000 \text{ px}$.

Plusieurs courants majeurs de représentation fréquentielle sont utilisés, dont :

- la Transformée de Fourier [201, 175],
- la Transformée en cosinus discrète [178],
- la Transformée en ondelettes [8, 87],
- et la Transformée de Gabor [201].

Ce travail se consacre à deux de ces approches, d'une part à la transformée de Fourier et d'autre part à la transformée en ondelettes, toutes deux traitées sur des problématiques de classification similaires [209, 79, 78]. Également, ces deux méthodes sont deux représentants majeurs des deux grandes familles de représentation en fréquences.

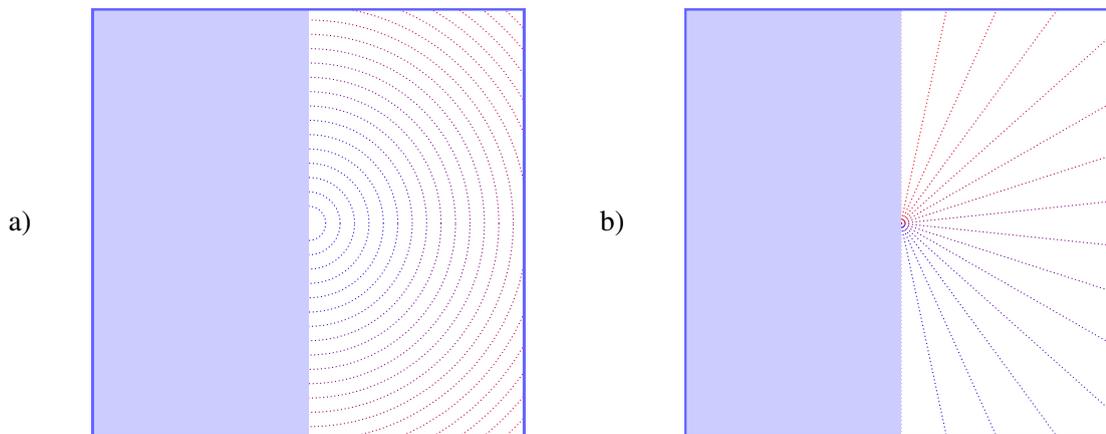


Figure 5.3 : Schéma représentant l'extraction de caractéristiques à partir de l'espace de Fourier. En a), l'utilisation de cercles concentriques depuis l'origine, schématisé en pointillés ; En b), l'extraction selon une direction.

La transformée de Fourier se résume en la décomposition d'une donnée en une somme de fréquences. Dans le cas d'une image, cette transformation donne lieu à une représentation, dans laquelle les basses fréquences, au centre, représentent une forme d'homogénéité au sein d'une image tandis que les hautes fréquences, en extérieur, sont associées à de fortes zones de transition. Cette transformation conserve également l'orientation des fréquences, représentée à l'aide d'un angle formé à partir de l'origine. En revanche, bien que la transformée de Fourier permette de décrire parfaitement la composition fréquentielle de l'image, elle ne peut en localiser la provenance des fréquences [209]. Cet espace reste néanmoins approprié à la caractérisation d'images texturées, et a suscité un grand intérêt au milieu des années 1980 [153]. Différentes mesures ont été proposées dans cet espace afin de caractériser des textures :

- l'extraction d'information à partir **de cercles concentriques** réguliers [175, 209] afin d'isoler l'information propre à chaque intervalle de fréquences (Figure 5.3 - a),
- l'extraction d'information à partir **de directions** en allant du centre vers l'extérieur de la transformée à angles constants [209] (Figure 5.3 - b).

Ces deux travaux préconisent l'extraction d'une valeur de moyenne et d'un écart-type sur différentes régions de l'espace fréquentiel afin de qualifier les phénomènes précédemment présentés [175, 209].

L'une des expériences menée par Wiltgen et al. [209] se base sur la transformée de Fourier, et extrait **38 descripteurs** du spectre de magnitude :

- **22 descripteurs** correspondent à la moyenne calculée sur des rayons répartis de manière équidistante du centre du spectre (Figure 5.3 - a),
- **16 descripteurs** correspondent à la moyenne calculée sur des directions à intervalles constants (Figure 5.3 - b).

La transformée de Fourier est effectuée dans le reste de ce travail par la bibliothèque logicielle *SciPy* [204].

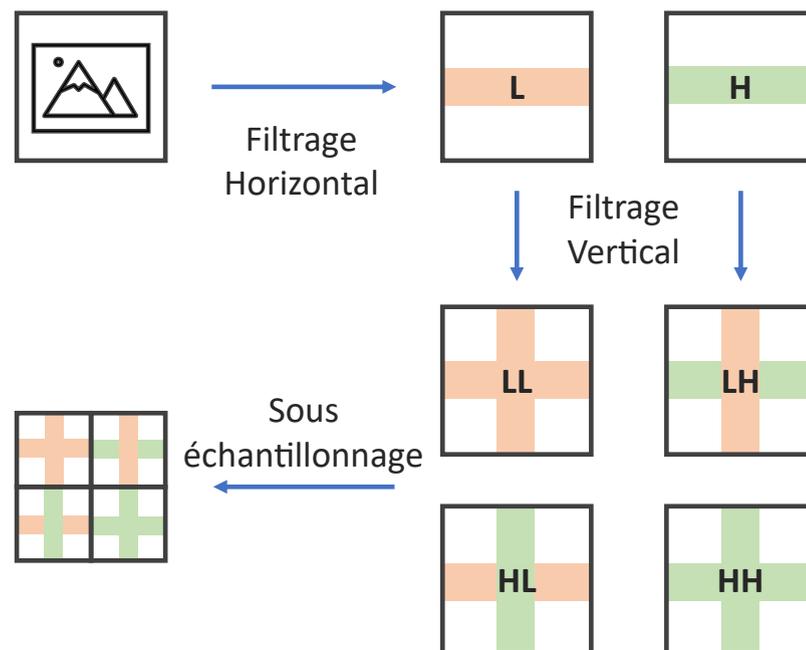


Figure 5.4 : Principe de la décomposition en ondelettes appliquée aux images [123]. La décomposition est réalisée en trois temps. Deux filtres respectivement passe bas (L) et passe haut (H) sont réalisés selon la direction horizontale, puis selon la direction verticale. Enfin, un sous-échantillonnage avec un coefficient de valeur 2 est appliqué afin d'obtenir une matrice de même taille que l'image d'origine.

La transformée en ondelettes possède divers avantages dont la capacité de localisation, par translation de l'ondelette mère, et la possibilité de prendre en considération plusieurs échelles d'analyse,

par homothétie [123, 209]. Concernant le choix de l'ondelette mère, celui-ci ne semble pas affecter la qualité des caractéristiques extraites [62, 123]. Néanmoins, les auteurs préconisent le choix d'ondelettes dont l'enveloppe est *symétrique* ou *anti-symétrique* afin de conserver des propriétés identiques quelle que soit l'orientation de l'image, et une ondelette suffisamment *compacte* pour capter un niveau de détail suffisant [123]. Ainsi, les ondelettes de Daubechies et de Haar sont privilégiées lors de ces diverses expérimentations fréquentielles, les diverses références de l'état de l'art s'orientant vers celles-ci [209, 79]. La décomposition en ondelettes sur images se réalise en trois étapes successives : (i) de filtrage passe-bas et passe-haut horizontal, (ii) de filtrage passe-bas et passe-haut vertical, (iii) puis une étape de sous-échantillonnage afin de réduire la quantité d'informations générées après ces deux précédentes étapes. Ce processus en trois temps est schématisé sur la Figure 5.4 et les opérations de transformée en ondelettes sont effectuées à l'aide de la bibliothèque logicielle *PyWavelets* [113].

Ainsi, les expériences sur la base de descripteurs en fréquences sont réalisées en plusieurs temps. Dans un premier temps, la performance de la transformation de **Fourier** et de la transformation en ondelettes de **Daubechies** mais également de **Haar** est mesurée sur la séparation des divers groupes d'images. Dans un second temps, la méthode de **Wiltgen** et al. [209] employant une combinaison de descripteurs de l'espace de Fourier est évaluée dans ces mêmes conditions [209]. La Table 5.4 synthétise le nombre associé de caractéristiques extraites par ces diverses méthodes d'extraction fréquentielles.

Méthode	Nombre de caractéristiques
Fourier	10 / 20 / 30
Ondelettes de Daubechies	12
Ondelettes de Haar	12
Wiltgen [209] - Fourier	38 (22+16)

Table 5.4 : Liste des méthodes fréquentielles évaluées dans ce chapitre et leur nombre de caractéristiques extraites associées.

5.2.3/ MÉTHODE D'APPRENTISSAGE PAR TRANSFERT DE CONNAISSANCES

Les méthodes d'apprentissage par transfert de connaissances représentent un champ d'application relativement récent de l'apprentissage automatique. Le principe est de réutiliser des connaissances précédemment acquises, dans le but de résoudre de nouveaux problèmes plus rapidement, voire plus efficacement. Pour cela, ce champ vise à proposer des techniques permettant la transposition de connaissances d'une ou plusieurs tâches sources à une tâche cible [147].

Le phénomène conjoint d'engouement pour les **CNN** et la complexité des architectures récentes pour le traitement des images a accentué les recherches en ce sens. En effet, les techniques d'apprentissage par transfert de connaissances sont une réponse efficace à de nombreuses problématiques autour de l'entraînement des **CNN**, comme :

- les **contraintes de suffisance de données**, notamment de données convenablement structurées et annotées disponibles publiquement,
- le **complexité de l'entraînement**, comprenant le réglage des divers paramètres de ces réseaux (augmentations des données, fonction d'optimisation, ...),
- et les **contraintes matérielles nécessaires**, regroupant les aspects matériels dédiés au support de l'entraînement de ces architectures.

Certains domaines d'application ont vu apparaître de nombreuses bases de données publiques, contribuant à résoudre partiellement la contrainte liée aux données, et permettant la mise en compétition de méthodes appliquées à une tâche précise. Parmi ces bases de données publiques, peuvent être citées : (i) la **base MNIST**, pour mesurer l'aptitude de systèmes dédiés à la reconnaissance de chiffres écrits à la main [112], (ii) les **bases CIFAR10 et CIFAR100**, permettant l'évaluation de systèmes de reconnaissance d'objets sur des image miniatures [105], (iii) et plus récemment, la **base ImageNet** pour la reconnaissance d'objets sur des images de taille commune [51].

Actuellement, la base ImageNet est l'une des plus sollicitées pour entraîner et évaluer les architectures de **CNN** actuelles. Outre sa taille, cette base propose des niveaux d'annotations variés, dont 1000 classes sont proposées. De plus, une très grande diversité d'échantillons est mise à disposition pour chaque classe pour un total de plus de 14 millions d'images. Selon certaines études, ce nombre d'images reste néanmoins sur-dimensionné vis-à-vis du nombre d'éléments nécessaires à l'entraînement de ces réseaux [90].

Parallèlement à ce phénomène de mise à disposition de bases de données, sont ouverts publiquement des défis visant à démontrer l'efficacité de chaînes de traitement pour analyser ces données. De nombreuses avancées ont été réalisées sur la base ImageNet, à l'aide d'architectures de **CNN** régulièrement mises à disposition de la communauté. Parmi ces réseaux peuvent être cités les réseaux AlexNet [106], GoogLeNet [185], VGG [173], Inception-V3 [186], ResNet [83] et Inception-ResNet [184], avec respectivement comme précisions 0,54, 0,68, 0,71, 0,76, 0,78, et 0,80 sur la base ImageNet [23].

Néanmoins, la mise à disposition de la communauté de ces bases de données massives et de ces architectures ne permet pas de lever les deux verrous liés aux contraintes d'entraînement et matérielles de ces réseaux. C'est donc particulièrement le partage de modèles de **CNN** pré-entraînés, couplé aux principes d'apprentissage par transfert de connaissances, qui a suscité l'intérêt de nombreux travaux dont certains appliqués à l'imagerie médicale [119]. Ainsi, la majeure partie de ces réseaux pré-entraînés est obtenue à l'aide de la base ImageNet, bien qu'aucune justification formelle n'ait été fournie à ce propos [90].

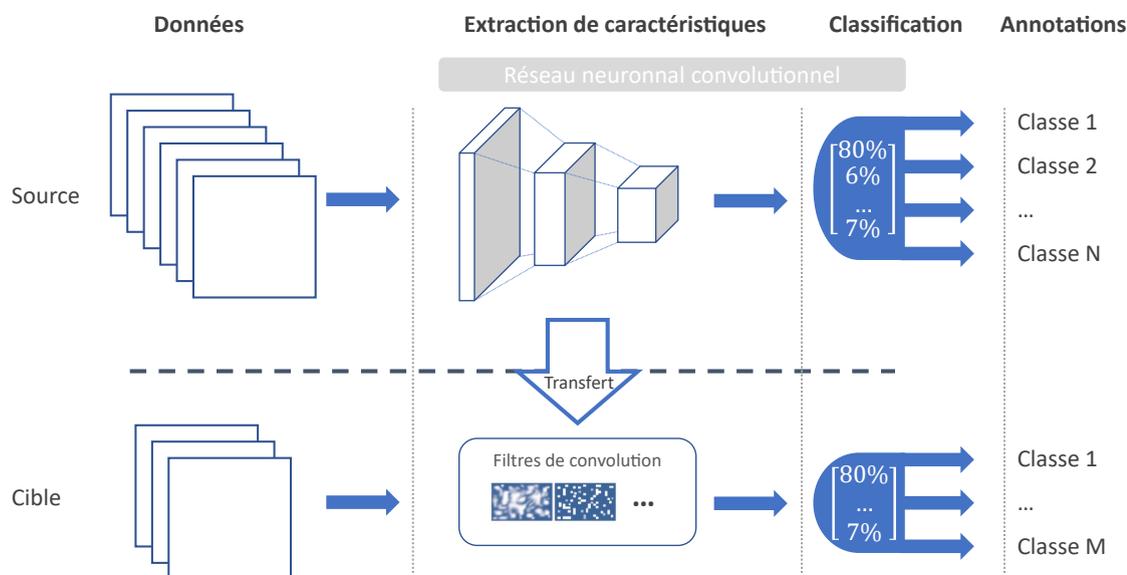


Figure 5.5 : Schéma représentant l'apprentissage par transfert par **CNN**. Le réseau est entraîné sur des données et annotations d'une problématique *source*. Puis, les paramètres de la partie liée à l'extraction de caractéristiques sont réemployés au sein d'une nouvelle problématique *cible*.

L'apprentissage par transfert sur **CNN** se décompose en deux procédés majeurs : (i) d'une part, **l'extraction de caractéristiques**, moins coûteuse en ressources matérielles et en temps, (ii) et d'autre part, **le réglage fin** qui nécessite à l'inverse plus de temps et de ressources matérielles. Au sein de cette partie, le principe d'extraction de caractéristiques est développé afin de limiter les contraintes d'apprentissage, schématisé sur la Figure 5.5. Le principe consiste dans un premier temps à retirer les couches responsables de la classification du problème cible, puisque spécifiques au problème initial [90, 50]. Ces couches sont généralement des couches dites totalement connectées, qui vont pondérer l'importance des zones d'activation jusqu'aux classes respectives. Dans un second temps, les couches responsables de l'extraction de caractéristiques sont conservées puis figées, en partant du postulat que ces caractéristiques sont suffisamment variées et adaptées à d'autres problématiques du traitement de l'image. Enfin, ces couches sont combinées à de nouvelles couches totalement connectées ou à des modèles de classification [119].

À cette fin, une première stratégie consiste à aplatir l'information des couches d'extraction par une étape de *Flatten* (littéralement *Aplatissement*), transformant celle-ci d'un format de matrice 3D à un vecteur 1D (voir Figure 5.6 - Gauche). Néanmoins, l'une des problématiques engendrée concerne la dimension spatiale souvent importante de ces caractéristiques extraites. En effet, la dimension spatiale de ces données sera réduite par les couches convolutionnelles selon leurs paramètres dépendants de l'architecture choisie. Ainsi, les paramètres de décalage, de pas ou encore de taille du noyau de convolution sont autant de paramètres qui affectent la dimension spatiale de sortie. Pour pallier cette information conséquente, des couches dites de *Global Pooling* ont été introduites dans la littérature, et consistent à réduire l'information spatiale d'une carte d'activation à une unique valeur dont le principe est schématisé sur la Figure 5.6 - Droite. Les principales fonctions mises à disposition sont la fonction maximum et moyenne, donnant respectivement lieu à des couches de *Global Maximum Pooling* (ou *Global Pooling - Maximum*) et de *Global Average Pooling* (ou *Global Pooling - Moyenne*). D'autres fonctions sont également développées, mais la littérature s'accorde à préférer l'utilisation des couches de *Global Pooling - Maximum* [39].

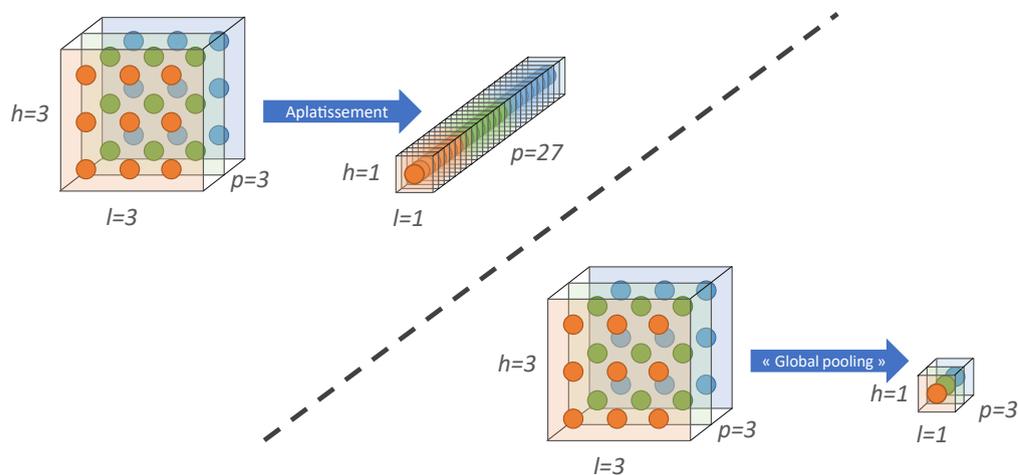


Figure 5.6 : Schéma représentant l'aplatissement et la réduction d'information par *Global Pooling* au sein de **CNN**. Le réseau est entraîné sur des données sources, généralement plus complètes. Puis, les paramètres de la partie liée à l'extraction de caractéristiques sont conservés et réemployés au sein d'une nouvelle problématique cible.

Bien que la majeure partie des travaux en imagerie médicale s'oriente vers l'architecture Inception-V3, son efficacité n'a pas été formellement démontrée au sein des applications médicales. La raison la plus probable expliquant le choix de cette architecture est la mise à disposition de ses modèles pré-entraînés par la plupart des bibliothèques de **CNN** [119]. Les différentes expériences sur l'apprentissage par transfert se focalisent sur des architectures relativement récentes pré-entraînés sur la base d'images ImageNet et proposées par la bibliothèque *Keras Applications* [38]. Afin de traiter les images dans leur intégralité, ces expérimentations ne considèrent que des réseaux dont les couches d'extraction de caractéristiques sont purement convolutionnelles et donc indépendantes de la taille de l'entrée. Ainsi, sont considérées les architectures de **CNN** représentatives des avancées dans le domaine durant la dernière décennie, dont VGG-16, Inception-V3, ResNet-50 et Inception-ResNet. L'ensemble des réseaux employés et nombre de paramètres extraits associés sont résumés au sein de la Table 5.5. Finalement, la manipulation des **CNN** est réalisée à haut niveau grâce à la bibliothèque logicielle *Keras* [37] et est couplée à la bibliothèque de bas niveau *Tensorflow* [1].

Architecture	Global Pooling	Nombre
VGG-16	Maximum	512
	Moyenne	512
Inception-V3	Maximum	2048
	Moyenne	2048
ResNet-50	Maximum	2048
	Moyenne	2048
Inception-ResNet	Maximum	1536
	Moyenne	1536

Table 5.5 : Liste des architectures de **CNN** employées dans ce chapitre et leur nombre de caractéristiques extraites associées.

5.2.4/ ANALYSE PRÉLIMINAIRE DES MÉTHODES D'EXTRACTION

Cette analyse préliminaire de l'information a pour but de montrer l'impact des méthodes d'extraction de caractéristiques précédemment citées sur le jeu de données mis à disposition. D'une part, cette sous-section propose d'observer l'application des méthodes de transformation précédentes sur deux images typiques de tissus **malin** et **bénin**. D'autre part, la capacité de séparation des images de chacune des méthodes précédentes est observée sur l'ensemble des données.

Tout d'abord, la présence de motifs plus lisses et moins contrastés au sein des images bénignes peut-être constaté, à l'inverse des motifs présents dans les images malignes. Un exemple de l'application de **GLCM** de premier ordre sur des images typiques est visible sur la Figure 5.7. Il est possible d'observer une uniformité de cette information quelle que soit la direction observée, allant dans le sens de l'utilisation de la moyenne sur les caractéristiques proposées par Haralick et al. [209]. Enfin, cette transformation montre une certaine cohérence des transitions dans les images avec des valeurs présentes en majorité sur la diagonale. Néanmoins, ce phénomène semble plus intense sur l'image bénigne, traduisant des transitions plus douces que sur l'image maligne.

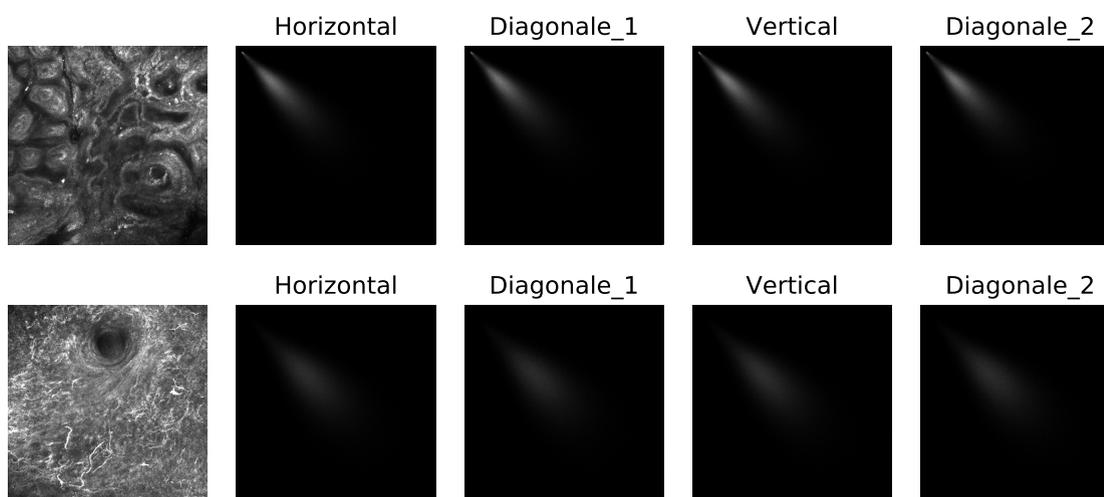


Figure 5.7 : Exemple d'extraction de **GLCM** de second ordre sur deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de **LM**.

La Figure 5.10 met en application la transformée de Fourier centrée sur ces deux mêmes images. L'observation du module de cette transformation permet de constater une information plus présente en intensité sur les basses fréquences que sur les hautes fréquences de l'image bénigne, là où sur l'image maligne cette information au centre forme une tâche plus diffuse. En revanche et après une lecture rapide, la direction ne semble pas apporter d'informations pertinentes comme suggéré par le travail de Wiltgen et al. [209].

Enfin, la Figure 5.8 reflète cette décomposition en ondelettes de Daubechies tandis que la Figure 5.9 donne un aperçu de cette même méthode par ondelettes de Haar. L'application de ces transformations permet de constater que les ondelettes de Haar semblent plus propices à isoler les zones de forte transition que celle de Daubechies. Néanmoins, pour ces deux types d'ondelettes, les décompositions de l'image bénigne semblent plus homogènes que celles de l'image maligne. De même, l'extraction des hautes fréquences sur l'axe vertical et horizontal semble plus adaptée que l'extraction combinée de ces deux axes pour l'image maligne.

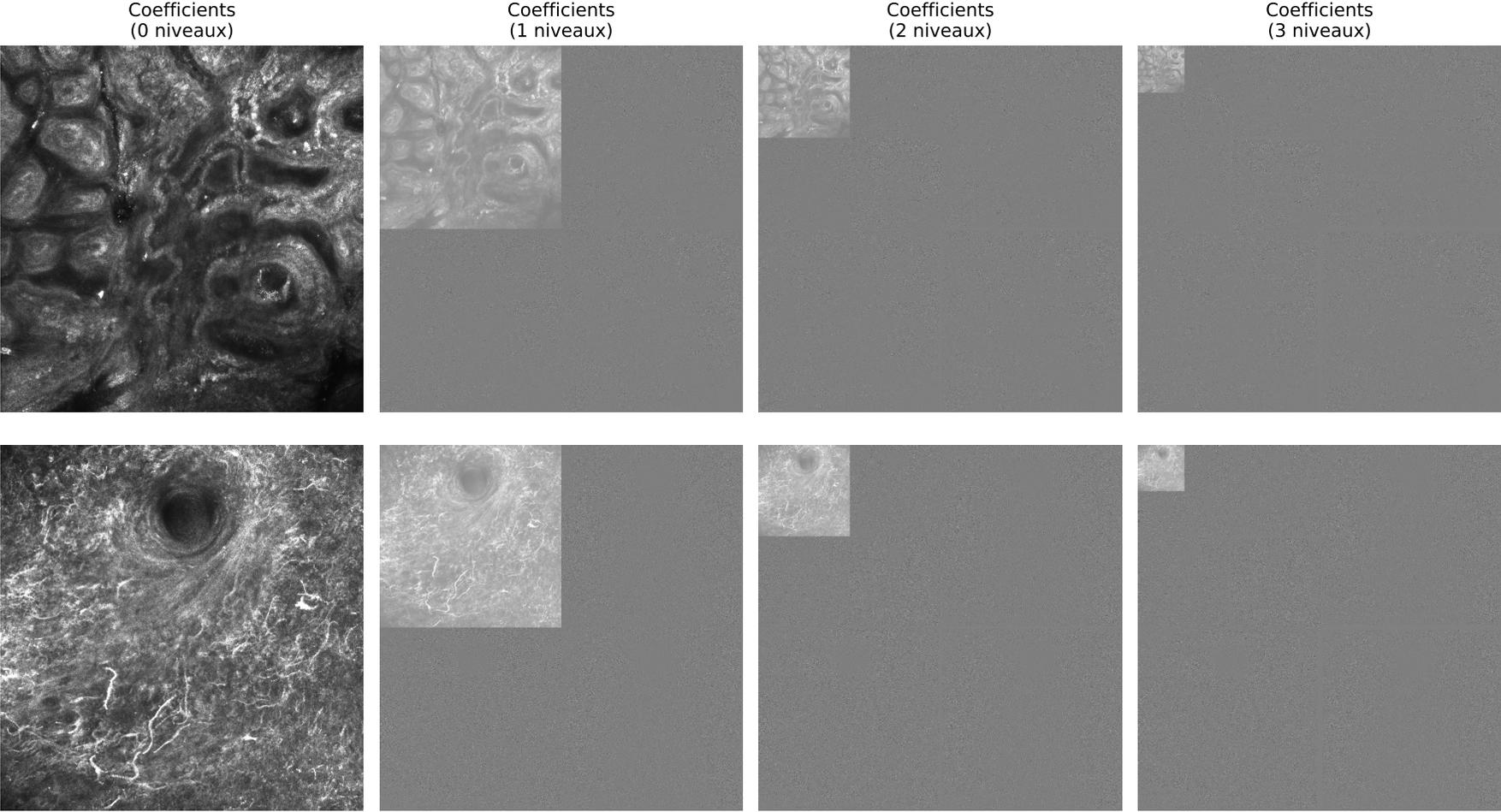


Figure 5.8 : Exemple de transformée en ondelettes de Daubechies appliquée à deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de **LM**.

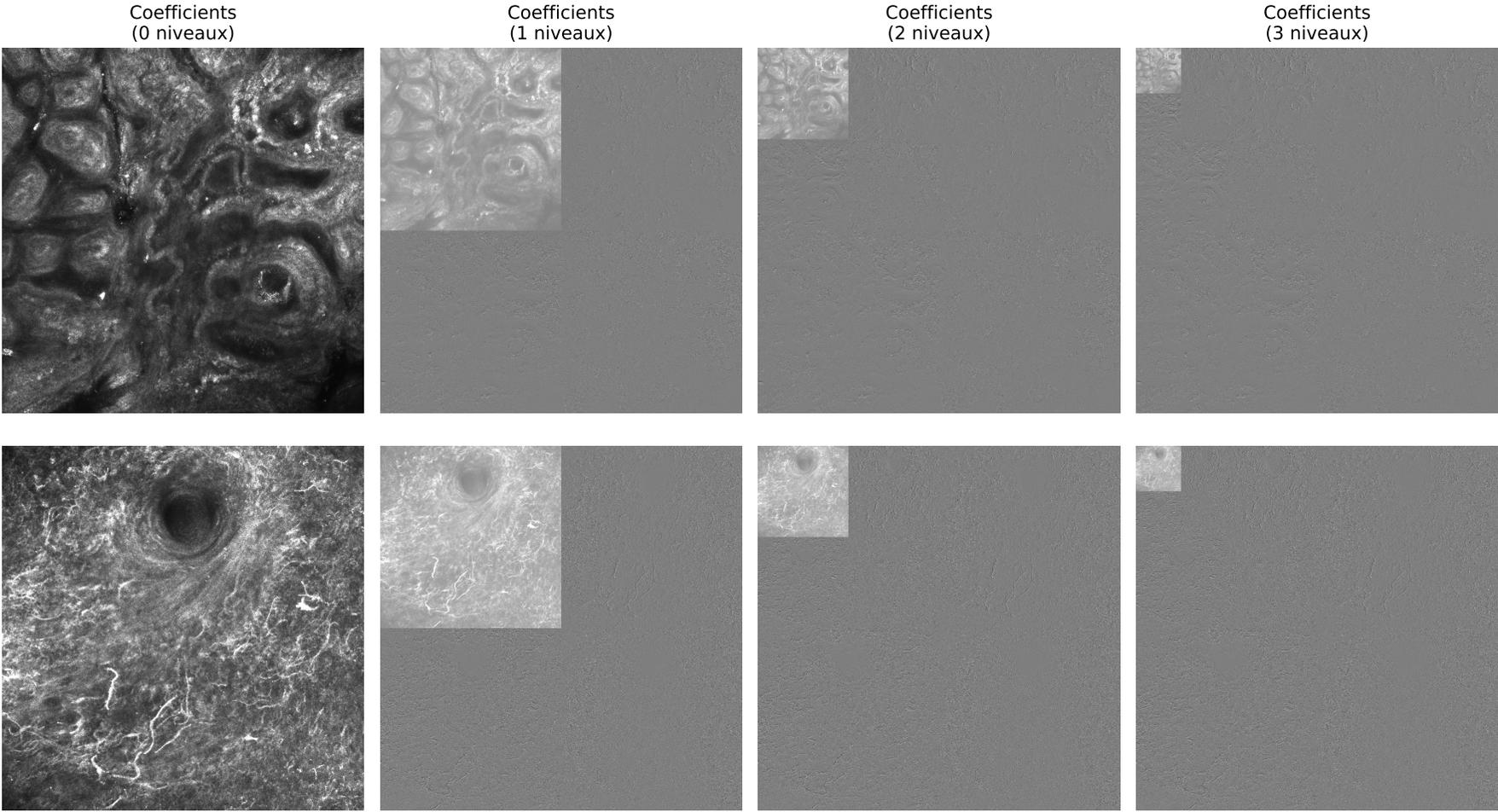


Figure 5.9 : Exemple de transformée en ondelettes de Haar appliquée à deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de **LM**.

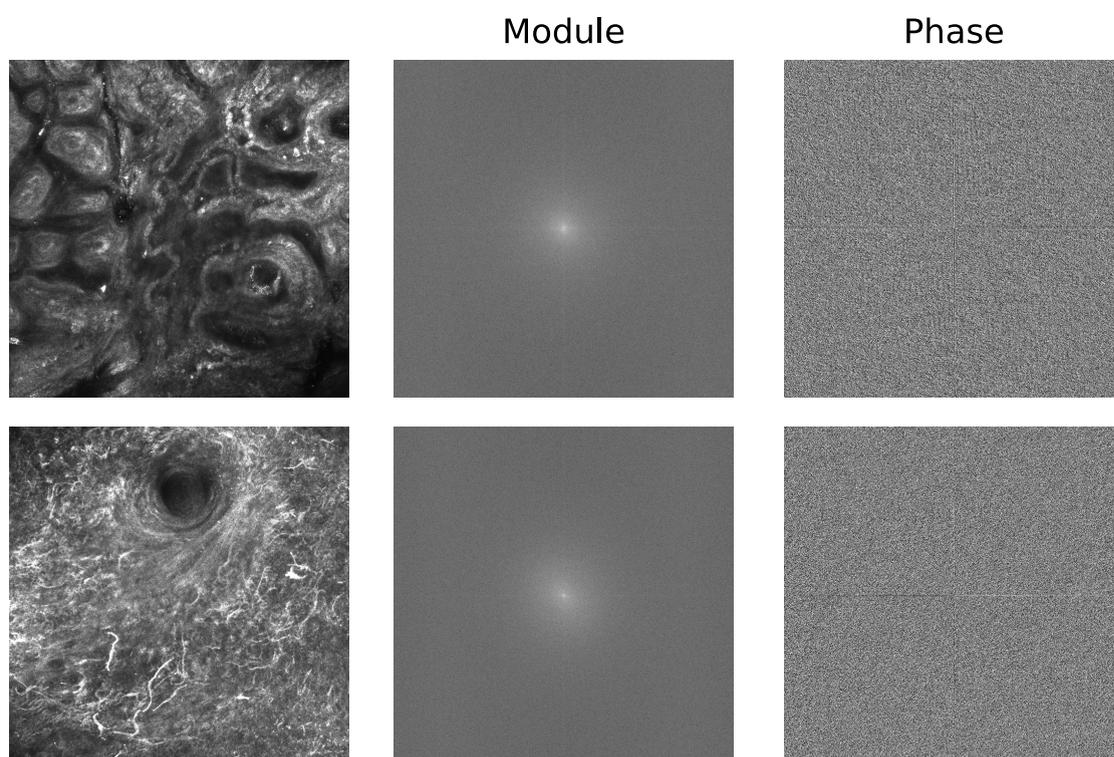


Figure 5.10 : Exemple de transformée de Fourier centrée appliquée à deux images typiques, représentation selon le module et la phase. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de **LM**.

Dans le but d'évaluer la séparation des classes, l'emploi de la réduction par le **Principal Component Analysis, PCA** (*alt. : Analyse en composantes principales, ACP*) est privilégié. En effet, les données utilisées dans ce manuscrit contiennent un grand nombre de dimensions et ce malgré l'extraction de caractéristiques, rendant le problème difficile à analyser visuellement. De nombreuses techniques de réduction de l'information permettant une meilleure visualisation sont proposées dans la littérature, comme par exemple le T-SNE [127] mais dont le principal inconvénient réside dans le nombre de degrés de liberté proposés. Néanmoins, le **PCA** est une technique communément acceptée à cette fin [85] et possède l'avantage de ne pas être dépendante de ce type de paramètres. Les informations de variance, propre à ces deux axes de projection, ainsi que pour chaque classe leur centre et leur dispersion sur la base de l'écart-type sont proposées pour chaque méthode d'extraction.

Ainsi, la Figure 5.11 met en évidence cette technique pour l'ensemble des méthodes spatiales d'extraction. Par ces méthodes, la variance expliquée provenant des deux axes de projection semble à même de fournir l'intégralité de la variance de ces données. Les caractéristiques extraites à l'aide des techniques d'extraction de premier ordre ne semblent pas suffisantes pour séparer les diverses classes. En effet, les centres de masse et leur dispersion semblent assez similaires, notamment pour les données bénignes et malignes. De plus, la séparation des classes ne semble pas marquée dans ces divers schémas d'extraction, et les classes bénignes et saines se chevauchent lorsque les méthodes employant les caractéristiques d'Haralick et al. [80] sont utilisées. Par ailleurs, la distribution des méthodes employant ces caractéristiques semble similaire, suscitant de faibles différences entre elles.

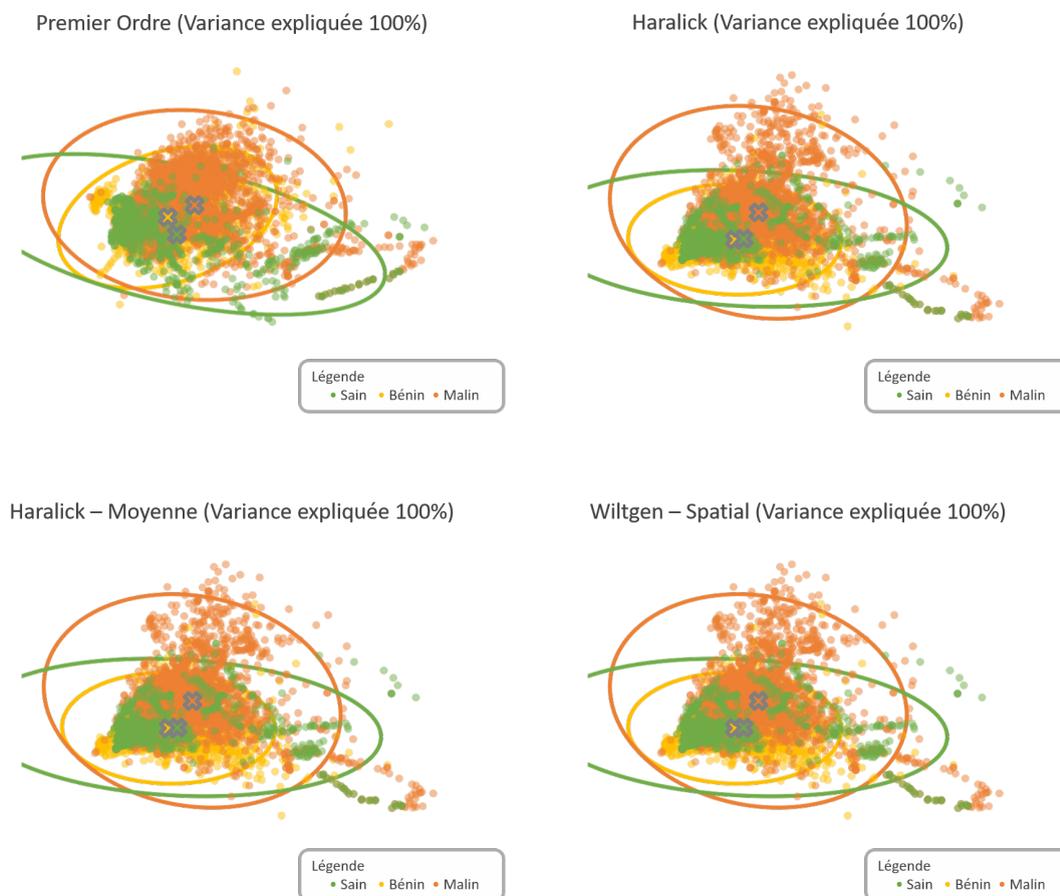


Figure 5.11 : Visualisation des caractéristiques obtenues par techniques d'extraction spatiales et projection PCA sur les deux premières composantes.

La Figure 5.12 permet de mettre en évidence la distribution des données extraites par les méthodes d'extraction fréquentielle. Ainsi, les méthodes basées sur l'extraction de caractéristiques de Fourier semblent posséder une distribution similaire quel que soit le nombre de caractéristiques extraites. De plus, la méthode de Fourier proposée par Wiltgen [209], proposant d'extraire des caractéristiques selon diverses directions, ne semble pas influencer sur cette distribution. Du point de vue des méthodes basées sur le principe d'ondelettes, la séparation des données semble être davantage marquée, et l'impact de l'ondelette mère ne semble pas influencer sur la distribution des données et des classes. Pour l'ensemble de ces méthodes, les deux premières composantes du PCA semblent suffisantes pour exprimer la quasi-intégralité de la variance.

Pour finir, la Figure 5.13 recense les projections par PCA de l'ensemble des méthodes d'extraction basées sur des architectures de CNN pré-entraînées, catégorisées sous le terme de transfert par apprentissage. Une différence notable est à remarquer entre les architectures utilisant une couche de *Global Pooling - Moyenne* et celle utilisant une couche de *Global Pooling - Maximum*. En effet, la séparation entre classes est marquée de manière plus importante lorsqu'une couche de *Global Pooling - Moyenne* est employée. De plus, cette séparation semble être plus prononcée sur les architectures Inception-V3 et ResNet-50. Enfin, par ces méthodes, la variance exprimée sur les deux premières composantes n'est que de 20 % à 40 %. Ce dernier constat s'explique par la grande quantité d'informations initialement proposée par ces architectures, soit entre 512 et 2048 caractéristiques extraites en comparaison d'une ou quelques dizaines pour les méthodes manuelles.

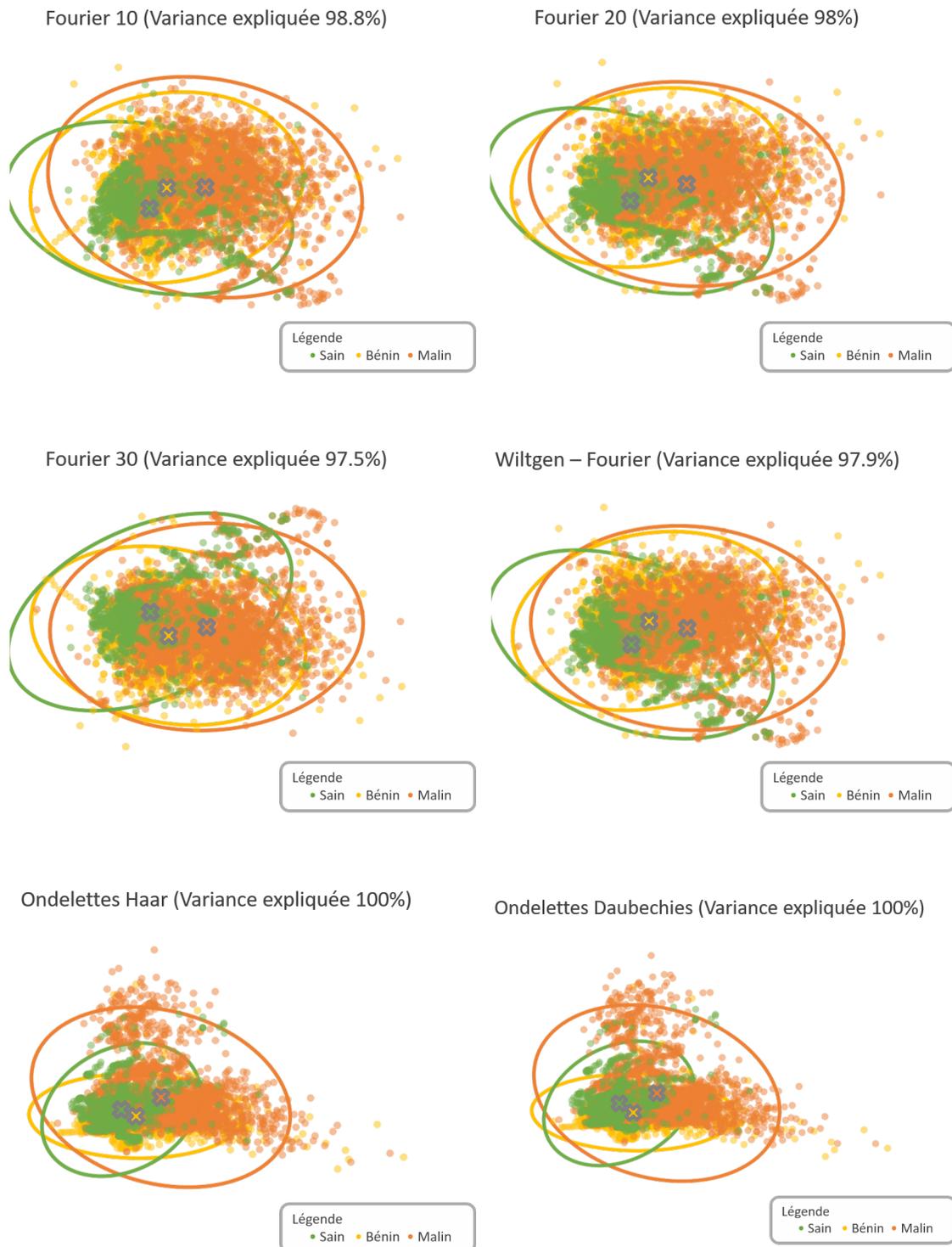
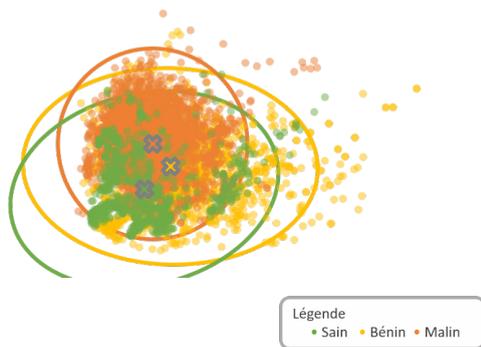
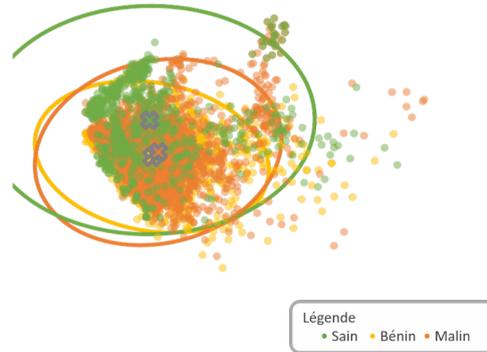


Figure 5.12 : Visualisation des caractéristiques obtenues par techniques d'extraction fréquentielles et projection PCA sur les deux premières composantes.

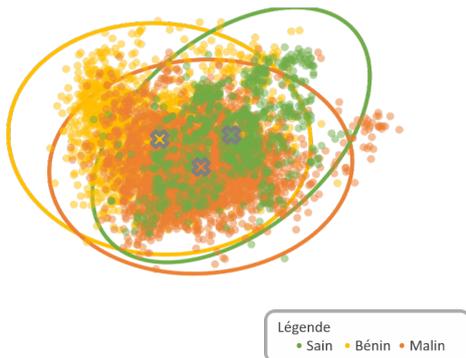
VGG-16 – Moyenne (Variance expliquée 41.4%)



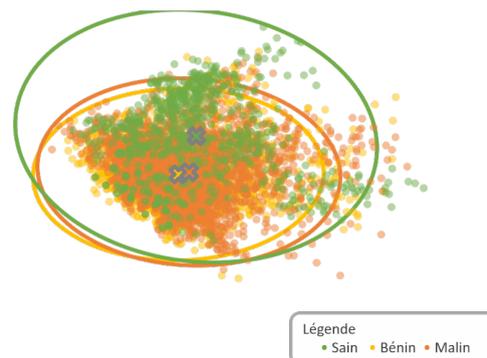
VGG-16 – Maximum (Variance expliquée 39.7%)



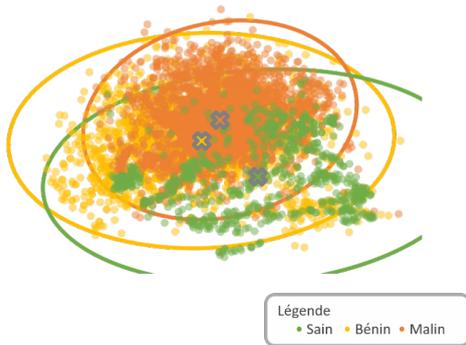
InceptionV3 – Moyenne (Variance expliquée 27%)



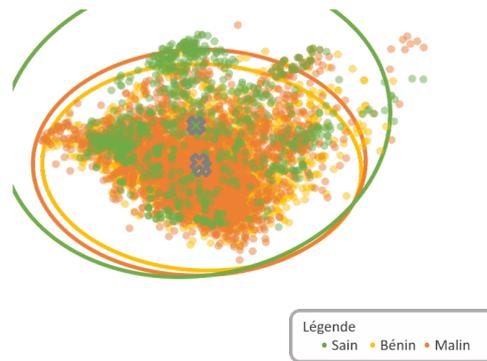
InceptionV3 – Maximum (Variance expliquée 18.7%)



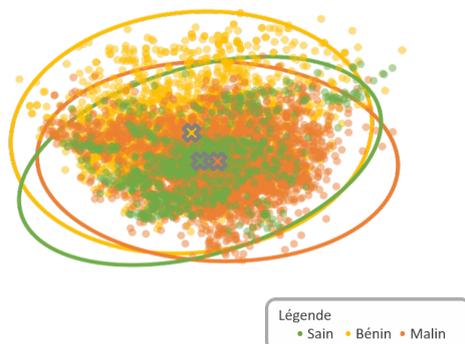
ResNet – Moyenne (Variance expliquée 48%)



ResNet – Maximum (Variance expliquée 39.7%)



InceptionResNet – Moyenne (Variance expliquée 36.2%)



InceptionResNet – Maximum (Variance expliquée 31.1%)

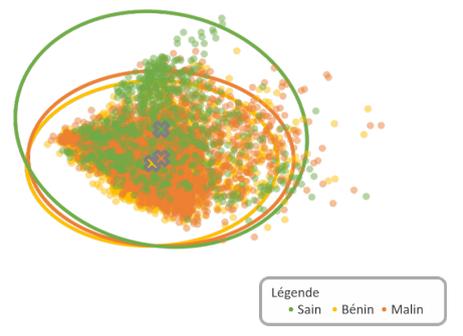


Figure 5.13 : Visualisation des caractéristiques obtenues par techniques d'extraction basées sur des réseaux de type CNN pré-entraînés sur la base ImageNet et projection PCA sur les deux premières composantes.

5.3/ PRÉ-TRAITEMENTS DE CARACTÉRISTIQUES

Cette partie est dédiée à la description de l'ensemble des transformations opérées après extraction de l'information par le biais des techniques évoquées en Section 5.2. Au sein de cette section sont abordées les diverses solutions pouvant améliorer la classification des images MCR. En effet, de nombreuses contraintes peuvent affecter le bon déroulement de la classification :

- un **déséquilibre des annotations** des divers types de tissus à séparer,
- un **nombre de caractéristiques** trop important conduisant à du sur-apprentissage ou à des difficultés à séparer le problème (voir l'extraction de caractéristiques par CNN),
- ou encore **une information non consistante**.

Ainsi, cette partie tente de répondre à ces questions à l'aide de différents domaines d'étude visant à corriger tout ou partie de ces éléments. Ces solutions sont abordées d'un point de vue logique en termes de processus de classification. Dans un premier temps, la **normalisation** des caractéristiques est présentée ainsi que son intérêt au sein du processus de classification. Dans un second temps, les méthodes de **réduction de l'information** sont proposées dans le cadre de l'extraction de caractéristiques par CNN afin de limiter la quantité de ces dernières. Finalement, le **balancement de données** et ses principales stratégies sont mis à disposition du lecteur.

5.3.1/ RÉDUCTION DE DIMENSIONS

Dans le domaine de l'apprentissage automatique, la réduction de dimensions est une technique qui permet de diminuer le nombre de dimensions afin de rendre un problème moins complexe, tout en conservant une information suffisante à la résolution du problème. Également, elle est un moyen de se prévenir du «fléau de la dimension», phénomène qui met en perspective d'une part la croissance exponentielle de l'ajout de nouvelles dimensions à la dilution des divers échantillons dans cet espace. Ce dilemme a pour principales conséquences d'augmenter la complexité d'un problème, la durée nécessaire à sa résolution mais également d'engendrer des risques de sur-apprentissage.

Les caractéristiques extraites par méthodes spatiales et fréquentielles étant assez restreintes (respectivement de 14 à 56 variables et, de 10 à 39 variables), l'impact de la réduction de dimension n'est pas étudié sur ces méthodes. En revanche, le nombre des caractéristiques extraites par méthode de transfert de connaissance est suffisamment important (entre 512 et 2048 variables) pour justifier de l'utilisation de cette réduction et d'étudier leur impact sur ces cas particuliers.

À cette fin, deux catégories de méthodes cohabitent : d'une part les méthodes de sélection de caractéristiques qui réduisent l'espace en décimant certaines dimensions jugées non pertinentes à la résolution du problème ; d'autre part les méthodes par projection ou transformation de l'espace existant qui réduisent et modifient les valeurs existantes. Dans la mesure où ces techniques s'appliquent sur les méthodes d'extraction par transfert de connaissances, cette investigation ne porte que sur cette seconde catégorie de méthodes. En effet, la sélection de caractéristiques est le plus souvent privilégiée lorsque ces caractéristiques font référence à des grandeurs physiques connues. Or, les caractéristiques issues du transfert de connaissances sont auto-déterminées par des CNN dont la signification n'est pas connue.

La méthode de PCA est l'une d'entre elles, dont l'idée majeure est de retirer l'information redondante pour ne conserver que l'information de variance, supposée contenir l'information. Cette

méthode cherche ainsi à identifier les axes selon lesquels la variation des données est maximisée. Il est ensuite nécessaire de déterminer un nombre d'axes suffisant pour séparer le problème (réduction du nombre de caractéristiques). Ainsi, ce travail retient les seuils de la littérature les plus pertinents de cette variance cumulée, à savoir des valeurs de 95 %, 97,5 % et 99 %.

La méthode de **Linear Discriminant Analysis, LDA** (*alt. : Analyse Discriminante Linéaire, ADL*) est employée à une fin de réduction de l'information, pour laquelle sont recherchés des axes de projection qui maximisent la variance inter-groupe. Pour comparaison, la variance est le résultat de la somme de la variance intra-groupe et inter-groupe. De manière similaire à la précédente technique, sont retenus des seuils de variance cumulée de 95 %, 97,5 % et 99 %.

La Table 5.6 reprend ces diverses méthodes ainsi que leurs paramètres mis en oeuvre conjointement à l'utilisation des méthodes d'extraction de caractéristiques sur la base du transfert de connaissances.

Méthode	Paramètre	Valeurs
PCA	Variance expliquée	[95 %, 97,5 %, 99 %]
LDA		

Table 5.6 : Table des méthodes de réduction employées et leurs paramètres associés.

5.3.2/ NORMALISATION DE CARACTÉRISTIQUES

Selon le domaine d'étude visé, la normalisation de caractéristiques peut être l'une des étapes cruciales quant au bon déroulement du processus de classification. Cette recherche a été amenée à considérer cet aspect suite à la dégradation de résultats au sein d'un travail proche de la thématique de ce manuscrit portant sur la classification d'images de dermatoscopie [27].

Il est communément admis que des méthodes telles que les **KNN** dont le principe repose sur des critères de distance, peuvent être fortement affectées par des caractéristiques non normalisées. Pour d'autres méthodes de classification, cette normalisation des caractéristiques fait davantage appel à un bon sens vis-à-vis de leur principe de fonctionnement. Certains travaux mettent en évidence la dégradation des performances sur les **SVM** [94] mais également les réseaux de neurones [27].

Diverses méthodes ont ainsi été proposées afin de remédier à ces défauts de distribution de l'information. L'un de ces travaux propose de réduire l'erreur de classification de modèles **SVM** par l'apport de normalisations tenant compte de la variance, du maximum ou encore conjointement du minimum et du maximum [94]. Le travail mené par Celebi et al. [27] propose une normalisation des caractéristiques à l'aide de la *Cote Z* ou *Standard score*, c'est-à-dire par soustraction de la moyenne au sein d'un même groupe de caractéristiques puis par une division par l'écart-type de ce même groupe.

Afin de couvrir au mieux cette problématique de normalisation, deux des méthodes majeures de la littérature sont envisagées pour la suite de cette sous-section, dont l'Equation ^(5.1) reprend les expressions sous forme mathématique. Ainsi, sont considérées :

- la normalisation par *Minimum / Maximum* [94],
- et la normalisation *Standard* [27].

$$\begin{aligned} \text{Minimum/Maximum} &= \frac{X - \min(X)}{\max(X) - \min(X)} \\ \text{Standard} &= \frac{X - \mu}{\sigma}. \end{aligned} \quad (5.1)$$

5.3.3/ BALANCEMENT DE DONNÉES

L'une des problématiques associées de manière récurrente à l'apprentissage sur des données réelles est celle de la répartition homogène des annotations, en opposition aux domaines impliquant des données synthétiques pouvant provenir de modèles génératifs par exemple. Ces déséquilibres d'annotations sont qualifiés sous le terme anglophone de *class imbalance* [158, 82].

Ces déséquilibres sont propres à de nombreux domaines d'application, mais touchent essentiellement des champs d'activité dans lesquels la tâche visée correspond à la détection d'événements isolés au sein d'une population, comme par exemple : (i) les comportements déviants (fraude bancaire [154]), (ii) le respect de critère de qualité (vérification de pièces industrielles [213]), (iii) ou encore d'anormalités cliniques (cancers ou autres pathologies cliniques [27]).

Certains choix peuvent permettre de limiter l'impact d'un déséquilibre d'annotations. Ainsi, la sélection d'une métrique adéquate peut permettre de limiter la dépendance aux données : la substitution de la précision par des mesures telles que la valeur d'AUC [27] ou de score F_1 sont des solutions pertinentes dans de telles situations. Néanmoins, la plupart des modèles de classification échouent à classer convenablement les données dans des situations de fort déséquilibres. Ces modèles préfèrent prédire constamment la classe majoritaire, pour minimiser le coût de l'erreur [89]. Dans le cadre de modèles basés sur des arbres de décision, la stratégie générale propose de déterminer les critères majeurs de séparation par partitionnements successifs des données. Ce partitionnement conduit inéluctablement à un affaiblissement des classes minoritaires [82]. Deux catégories d'approches coexistent pour solutionner ces déséquilibres [89], d'une part les approches par augmentation ou décimations des diverses catégories d'annotations, et d'autre part les approches algorithmiques qui consistent à pondérer la valeur des annotations afin de prendre en considération le déséquilibre de l'information [193, 82, 191].

En premier lieu, ces déséquilibres de classes peuvent être corrigés par des méthodes simples permettant de pallier ces défauts [158, 82]. Une première catégorie de ces méthodes consiste à **sous-échantillonner** les données jusqu'à obtenir une égalité entre les classes. Son principe le plus simple est celui du **sous-échantillonnage aléatoire** consistant à décimer de manière aléatoire les données des classes majoritaires afin d'égaliser les échantillons de la classe minoritaire. À l'inverse, une autre catégorie consiste à **sur-échantillonner** les données possédées afin d'obtenir à nouveau une égalité entre classes. Le principe le plus simple est celui du **sur-échantillonnage aléatoire** consistant à dupliquer aléatoirement les échantillons des classes minoritaires afin d'égaliser le nombre d'éléments de la classe majoritaire. Le schéma en Figure 5.14 reprend l'idée de ces deux principes.

En second lieu, des approches par correction des données plus avancées ont été proposées dans la littérature afin de compléter les méthodes de sur-échantillonnage et sous-échantillonnage. L'un des exemples majeurs de ces méthodes de sur-échantillonnage est la méthode **Synthetic Minority Over-Sampling**, (**SMOTE**) [34], dont le principe permet de générer par interpolation de nouvelles instances en choisissant aléatoirement une première instance a et de prendre aléatoirement parmi ses plus proches voisins une seconde instance b , afin d'interpoler une nouvelle instance c . En ce qui concerne les méthodes de sous-échantillonnage, peuvent être citées la méthode **Tomek-Link**,

(TL) [194] et Edited Nearest Neighbor, (ENN) [208] utilisées pour enlever les incohérences au sein des données. Néanmoins, ces méthodes seules possèdent diverses lacunes et sont solutionnées à l'aide de nouvelles approches qualifiées d'*hybrides* mêlant ces deux aspects. Parmi cette nouvelle catégorie, ce travail considère l'approche **SMOTE + TL** et l'approche **SMOTE + ENN** dont les principes consistent à synthétiquement augmenter les données par la méthode **SMOTE**, puis à nettoyer les données constituées d'instances réelles et synthétiques par la méthode **TL** ou **ENN**.

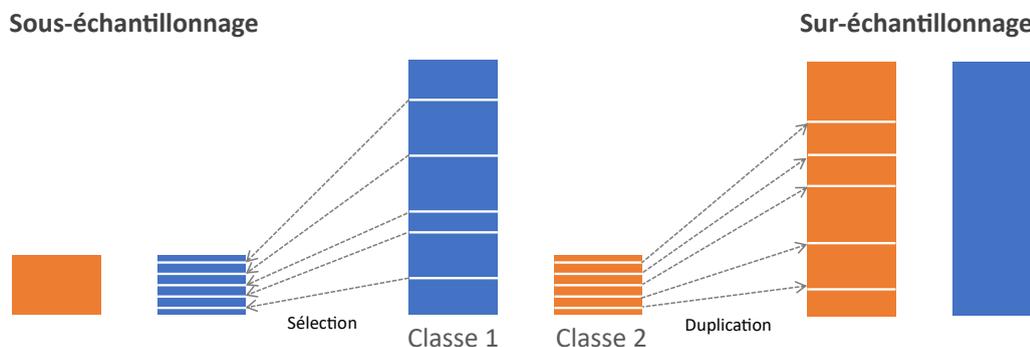


Figure 5.14 : Schéma des deux stratégies principales de balancement de données. À gauche, la stratégie de sous-échantillonnage, dans laquelle des éléments sont aléatoirement choisis jusqu'à réduire l'ensemble des classes au nombre d'éléments de la classe minoritaire ; À droite, la stratégie de sur-échantillonnage, dans laquelle les éléments des classes minoritaires sont dupliqués aléatoirement.

En opposition à ces méthodes dédiées, **les approches par pondération** des échantillons consistent à gérer le déséquilibre des annotations en appliquant des poids d'importance différente selon la provenance de l'échantillon lors de l'apprentissage du modèle. Cette approche par défaut est privilégiée puisqu'elle permet de contrer de manière assez intuitive ces déséquilibres.

Dans l'optique de récapituler cette partie, les méthodes utilisées dans le but de corriger le déséquilibre des données sont référencées dans la Table 5.7. Afin de traiter au mieux cette tâche, le balancement de données par les différentes stratégies évoquées est réalisé avec l'aide de la librairie logicielle *Imbalanced-learn* [114]. La correction par pondération est réalisée avec l'aide de la bibliothèque logicielle *Scikit-learn* [151].

Catégorie	Méthode
Sous-échantillonnage	Aléatoire
Sur-échantillonnage	Aléatoire
Hybride	SMOTE + TL
	SMOTE + ENN

Table 5.7 : Table des diverses méthodes de balancement mises en œuvre pour corriger les déséquilibres de données.

5.4/ PROCESSUS DE PRÉDICTION

Cette section dédiée au processus de prédiction se consacre à la dernière étape de traitement des données au sein du processus de classification supervisée. D'une part, son objectif est de présenter les divers modèles de classification utilisés pour la séparation de l'information et d'autre part, de mettre en exergue le processus d'évaluation et les raisons ayant justifié ces choix.

5.4.1/ MÉTHODES DE PRÉDICTION

Afin d'œuvrer à la séparation des échantillons à disposition de ce travail sur la base de leurs caractéristiques, divers modèles de classification sont sollicités et pour lesquels les performances sont évaluées. En effet, bien que les avantages et inconvénients de la plupart d'entre eux aient été évoqués au sein de la Section 3.4, seules les performances de classification mesurées sur le jeu de données de référence sont retenues.

Modèle	Hyperparamètre	Valeurs
CART	Profondeur maximum	[3, ∞]
	Critère de qualité	[Gini, Entropie]
	Caractéristiques maximum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
	Échantillons minimum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
RF	Profondeur maximum	[3, ∞]
	Critère de qualité	[Gini, Entropie]
	Caractéristiques maximum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
	Échantillons minimum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
GB	Profondeur maximum	[3, ∞]
	Critère	[MSE, MAE]
	Caractéristiques maximum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
	Échantillons minimum	[1, 2, 3, 4, 5, 6, 7, 8, 9]
SVM - Linéaire	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]
SVM - RBF	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]
	Gamma	[0,001, 0,01, 0,1, 1, 10, 100, 1000]
MLP	Couche cachées	[(20,), (20,20), (20,20,20,)]
	Activation	[Tanh, ReLu]
	Solveur	[SGD, Adam]
	Alpha	[0,001, 0,01, 0,1, 1, 10, 100, 1000]
	Taux d'apprentissage	[Constant, Adaptatif]

Table 5.8 : Table recensant l'ensemble des modèles de classification et leurs hyperparamètres évalués lors des expériences de ce travail.

Dans un premier temps, le modèle **CART** est utilisé afin de réaliser la séparation des données, son choix étant motivé par une étude dont la thématique est similaire à celle de ce travail [209]. Néanmoins, ces modèles tendent au sur-apprentissage dans des contextes où les dimensions sont nombreuses. Ainsi, sont utilisés de manière complémentaires les modèles **RF** [19] et **GB**, plus aptes à gérer cet aspect. Dans un second temps, les modèles **SVM** à noyaux *linéaires* et à *base radiale*, respectivement abrégés par les termes de **SVM - Linéaire** et **SVM - RBF** dans ce chapitre, sont employés pour cette même tâche de séparation des données, et sont motivés par leur utilisation dans de nombreux contextes de classification pour lesquels ils ont été démontrés comme robustes [175]. Ces modèles sont également employés dans des contextes de classification de lésion de la peau [27]. Enfin, la piste des **MLP** est envisagée afin d'aborder cette problématique de manière non linéaire.

Afin d'obtenir les meilleures performances sur ces modèles, les hyperparamètres associés à chacun d'entre eux sont affinés en respectant la procédure de la prochaine sous-section. Néanmoins, un compromis entre le temps d'exécution et le nombre de ces valeurs est nécessaire, conduisant ce travail à opter pour celles les plus couramment utilisées. L'ensemble des modèles de classification mais également leurs hyperparamètres et leurs valeurs associées sont recensés dans la Table 5.8. Par ailleurs, les modèles de classification employés par ce travail proviennent de la bibliothèque logicielle *Scikit-learn* [151].

5.4.2/ PROCÉDURE D'ÉVALUATION

Pour réaliser l'évaluation, est employée une validation croisée imbriquée déjà discutée dans la Section 3.4. Cette structure de validation est choisie car moins sujette au biais [26]. De plus une valeur de k comprise entre 5 et 10 pour l'évaluation est démontrée empiriquement comme un bon compromis entre biais et variance [92]. Les expériences de ce chapitre sont réalisées avec une valeur k de quatre pour leur évaluation afin de réduire la durée de calcul à un temps raisonnable et de deux lots pour la validation. Leur combinaison représente un bon compromis vis-à-vis des expériences de chapitre. La répartition de ces lots sur les données de **MCR** exploitées dans ce manuscrit est visible sur la Figure 5.15.

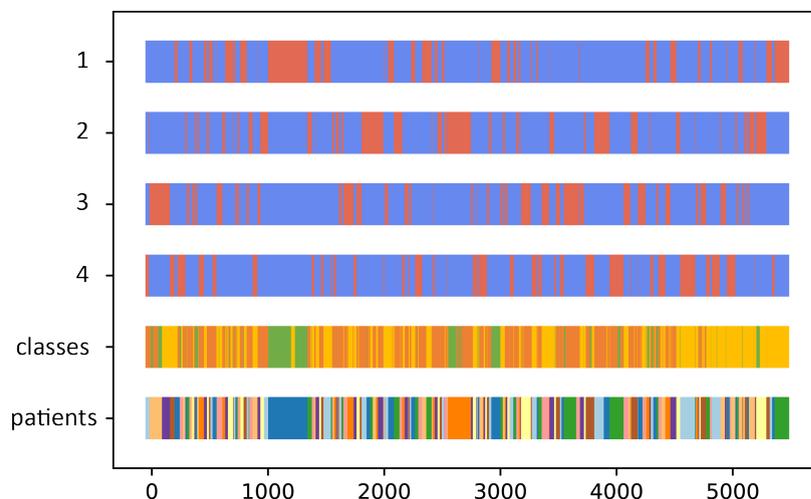


Figure 5.15 : Visualisation des divers lots utilisés pour l'évaluation des processus de classification.

Afin de régler les hyperparamètres et d'évaluer les performances des divers modèles, ce travail opte pour une métrique de type score F_1 également discutée dans la Section 3.4. En effet, cette métrique est robuste face à des déséquilibres d'annotations, tels que ceux rencontrés sur le jeu de données exploité. De plus, elle représente conjointement sous une unique mesure les valeurs de *précision* et de *rappel*. Son expression est binaire, elle sera ainsi calculée pour chaque classe dont la valeur est pondérée par le ratio du nombre d'échantillons disponibles dans chaque classe respective.

Ces expériences sont menées en plusieurs étapes successives, et débutent par l'évaluation de l'ensemble des méthodes d'extraction de caractéristiques conjointement à la normalisation de l'information selon la méthode de *Minimum/Maximum* et la méthode *Standard*. Aucun des tests menés n'est réalisé sur les caractéristiques brutes (sans normalisation), les expériences menées au préalable ne permettant pas de faire converger les modèles de classification dans un temps raisonnable de calcul.

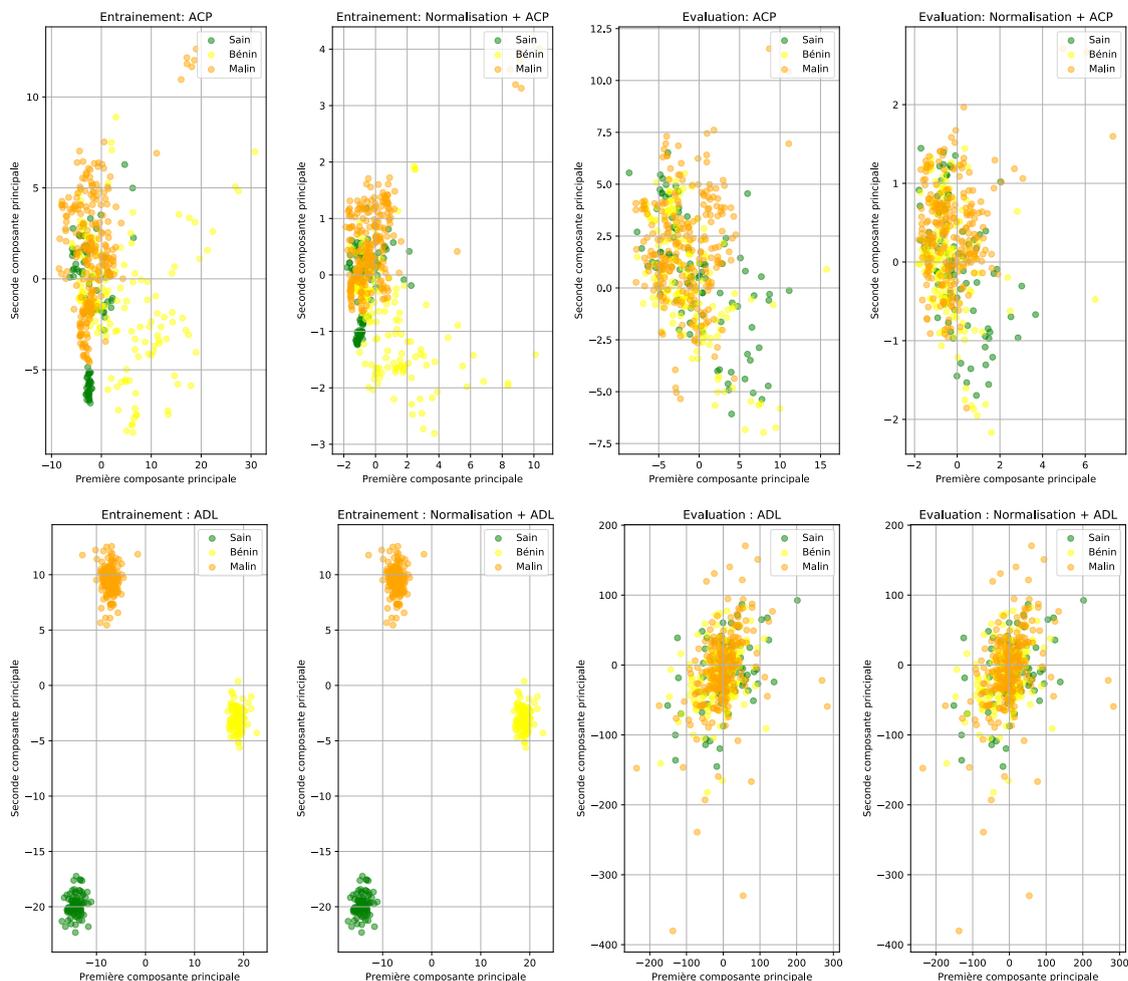


Figure 5.16 : Visualisation de l'influence de la réduction des dimensions, appliquée au deux premières composantes principales, sur un échantillon constitué arbitrairement de 500 images issues des données de MCR. En haut, technique de réduction basée sur le PCA ; En bas, technique de réduction basée sur le LDA. À gauche, un processus procédant à la réduction puis à la normalisation ; À droite, un processus procédant à la normalisation puis à la réduction.

Puis, l'impact de la réduction de dimension est mesuré sur les techniques de transfert de connaissance avant de procéder à leur classification. Contrairement aux recommandations de la librairie ¹, ce travail adopte un schéma de type réduction de l'information suivi d'une normalisation de caractéristiques avant l'étape de classification. En effet, les caractéristiques issues de réseaux convolutionnels possèdent une distribution de valeurs consistante, et n'affectent pas de manière significative les techniques de réduction de dimensions. En revanche, la réduction d'information par techniques de projection peut affecter la consistance de cette distribution des valeurs. Ce phénomène a notamment été observé sur un lot de 500 données choisies arbitrairement, et présentées sur la Figure 5.16.

Pour finir, ces expériences mesurent l'impact du déséquilibre des données sur les résultats de classification et évaluent la pertinence des méthodes qui visent à corriger ce déséquilibre dans ce contexte. Pour cela, les méthodes les plus pertinentes sont soumises à une classification sans correction du déséquilibre d'annotations mais également avec les divers mécanismes mentionnés dans la Section 5.3.3, et leurs résultats sont comparés.

La sélection des modèles les plus pertinents se base sur les résultats à trois classes dans la mesure où ces scores sont représentatifs de la capacité à caractériser la problématique dans son ensemble. Néanmoins dans la dernière partie de cette analyse, seuls les résultats de la détection des images malignes sont détaillés dans la mesure où ceux-ci correspondent à l'objectif visé par ce travail.

¹Source : Scikit-learn - «Importance of Feature Scaling».

5.5/ RÉSULTATS

Afin de limiter ces résultats, ce premier volet s'intéresse aux méthodes les plus performantes propres aux diverses catégories d'extraction de caractéristiques, associées aux techniques de normalisation. Dans un premier temps, en terme de méthode d'extraction spatiale, un score F_1 de 0,64 associé à un écart-type de 0,07 est obtenu par la méthode proposée par **Haralick** et al. [80] et **sans moyenne des axes**. Ce score correspond à l'utilisation conjointe d'une normalisation de type **Minimum / Maximum** et d'un modèle **SVM - RBF**. Dans un second temps, en terme de méthode d'extraction fréquentielle, un score F_1 de 0,67 associé à un écart-type de 0,06 est obtenu par l'utilisation de l'extraction en **ondelettes de Daubechies**. Ce score correspond à l'utilisation conjointe d'une normalisation de type **Minimum / Maximum** et d'un modèle **SVM - Linéaire ou RBF**. De plus, ce score correspond également à l'utilisation conjointe d'une normalisation de type **Standard** et d'un modèle **SVM - linéaire**. Enfin, en terme de méthode par transfert de connaissances, un score F_1 de 0,77 associé à un écart-type de 0,04 est obtenu sur l'architecture **ResNet-50** combinée à une couche de **Global Pooling - Moyenne**. Ce score correspond à l'utilisation conjointe d'une normalisation de type **Minimum / Maximum** et d'un modèle **SVM - Linéaire**. L'ensemble des résultats liés à ces expérimentations est visible sur la Figure 5.17.

Ce second volet s'intéresse aux méthodes d'extraction par transfert de connaissances et à l'impact des méthodes de réduction d'information, dont les combinaisons les plus pertinentes sont retenues afin de limiter la quantité de résultat produite. Seules les architectures couplées à une couche de *Global Pooling - Moyenne* sont évaluées, leurs performances ayant été significativement supérieures aux architectures couplées avec un *Global Pooling - Maximum*. D'une part, un score F_1 de 0,76 associé à un écart-type de 0,03 est obtenu sur l'architecture **ResNet-50** combinée à une couche de **Global Pooling - Moyenne**. Ce score provient de l'utilisation conjointe du **PCA avec une variance cumulée de 99 %**, d'une normalisation de type **Minimum / Maximum** et d'un modèle **SVM - Linéaire**. D'autre part, un score F_1 de 0,73 associé à un écart-type de 0,05 est obtenu sur l'architecture **VGG-16** combinée à une couche de **Global Pooling - Moyenne**. Ce score provient de l'utilisation conjointe du **LDA avec une variance cumulée de 95 %**, d'une normalisation de type **Standard** et d'un modèle **SVM - Linéaire**. Les résultats de ce volet de l'expérience sur les méthodes de réduction de l'information sont reportés sur la Figure 5.18.

Ces mêmes résultats sont également observés du point de vue des performances de la classe maligne. Par ce nouveau point de vue, un score F_1 de 0,82 associé à un écart-type de 0,02 est obtenu par la méthode de transfert de connaissances basée sur l'architecture **ResNet-50** couplé à une couche de **Global Pooling - Moyenne**. Comme précédemment, ce score résulte de l'utilisation conjointe d'une normalisation **Minimum/Maximum** suivi d'un modèle **SVM - Linéaire** dont l'utilisation est combinée à l'utilisation d'une **pondération** dans le but de compenser le déséquilibre de données. Les résultats liés à l'étude du déséquilibre des données, et adoptant le point de vue de la classe maligne, sont disponibles sur la Table 5.10.

	Aucune	Pond.	Sur-éch.	Sous-éch.	SMOTE+ENN	SMOTE+Tomek
Spatial	0,55±0,15	0,61±0,09	0,54±0,15	0,58±0,11	0,60±0,09	0,56±0,16
Fréquentiel	0,59±0,13	0,67±0,06	0,55±0,21	0,58±0,17	0,61±0,12	0,64±0,10
Transfert	0,71±0,08	0,77±0,04	0,74±0,07	0,72±0,05	0,74±0,05	0,74±0,05
Réduction	0,71±0,09	0,76±0,03	0,74±0,06	0,73±0,04	0,73±0,04	0,73±0,06

Table 5.9 : Résultats à trois classes issus de l'expérience mêlant les méthodes les plus performantes associées aux méthodes de balancement de données évoquées.

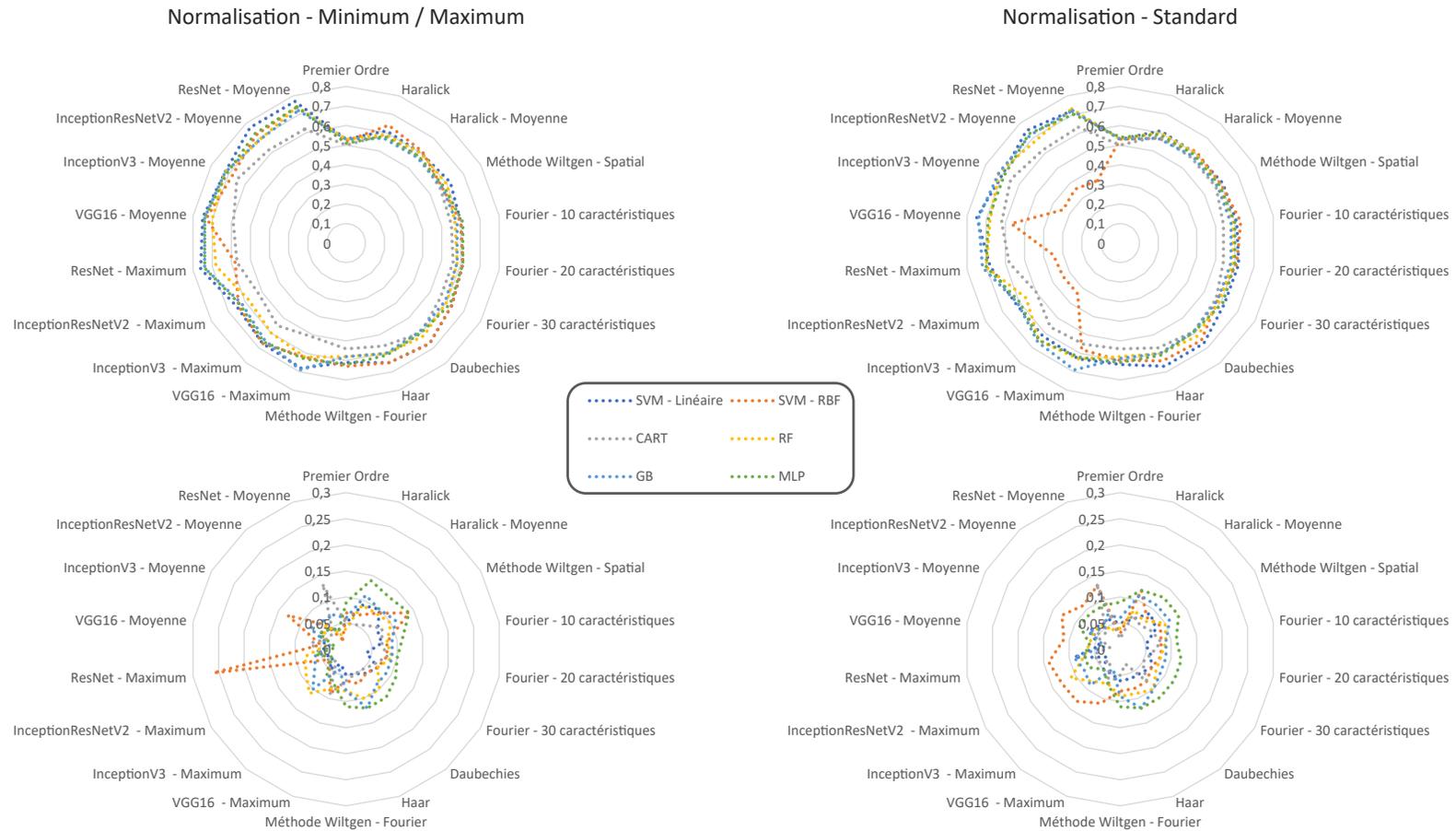


Figure 5.17 : Résultats issus de la classification par les diverses techniques d'extraction et par les modèles de classification mentionnés en couleur. À gauche, les résultats provenant de la normalisation de type *Minimum / Maximum* ; À droite, les résultats provenant de la normalisation de type *Standard*. En haut, la représentation basée sur les valeurs moyennes ; En bas, la représentation tenant compte de l'écart-type.

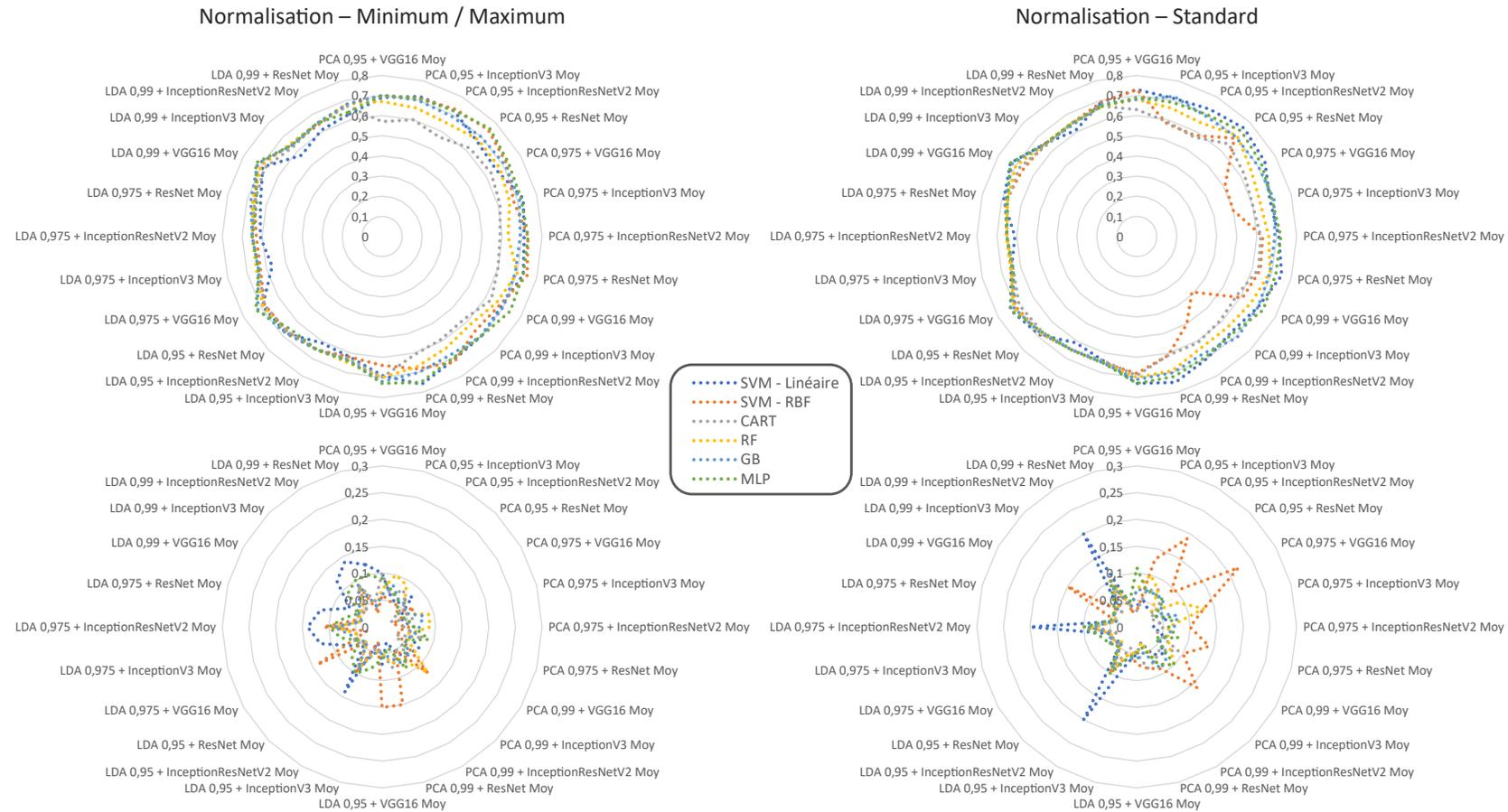


Figure 5.18 : Résultats issus de la classification par l'utilisation de techniques de réduction combinés aux techniques d'extraction par transfert et, par les modèles de classification mentionnés en couleur. Les tracés de couleurs représentent les divers modèles évalués au cours de l'expérience. À gauche, les résultats provenant de la normalisation par Minimum/Maximum ; À droite, les résultats provenant de la normalisation par score Standard. En haut, la représentation basée sur les valeurs moyennes ; En bas, la représentation tenant compte de l'écart-type.

Le dernier volet de ces expérimentations se destine à étudier l'impact du déséquilibre des données sur la qualité des résultats de classification. Son influence est mesurée sur les méthodes les plus performantes relatives aux précédents paragraphes de cette section et couplée aux méthodes de corrections de ces déséquilibres. Ainsi, un score F_1 de 0,77 associé à un écart-type de 0,04 est obtenu par la méthode de transfert de connaissances basée sur l'architecture **ResNet-50** couplé à une couche de **Global Pooling - Moyenne**. Ce score résulte de l'utilisation conjointe d'une normalisation **Minimum/Maximum** suivi d'un modèle **SVM - Linéaire** dont l'utilisation est combinée à l'utilisation **d'une pondération** dans le but de compenser le déséquilibre de données. Les résultats liés à l'étude du déséquilibre des données à trois classes sont visibles sur la Table 5.9.

	Aucune	Pond.	Sur-éch.	Sous-éch.	SMOTE+ENN	SMOTE+Tomek
Spatial	0,65±0,11	0,63±0,07	0,64±0,10	0,64±0,08	0,63±0,09	0,59±0,27
Fréquentiel	0,70±0,09	0,68±0,03	0,66±0,09	0,66±0,09	0,61±0,22	0,66±0,09
Transfert	0,81±0,04	0,82±0,02	0,81±0,03	0,79±0,01	0,80±0,02	0,82±0,03
Réduction	0,81±0,03	0,81±0,01	0,81±0,03	0,79±0,03	0,80±0,01	0,81±0,03

Table 5.10 : Résultats de la détection de tissus malins issus de l'expérience mêlant les méthodes les plus performantes associées aux méthodes de balancement de données évoquées.

Enfin, les courbes **ROC** permettent d'appréhender les performances de prédiction de ce système. Ainsi, les performances propres à l'ensemble de la base d'images annotées sont homogènes par cette technique avec des scores **AUC** dont les valeurs sont comprises entre 0,87 et 0,90. Néanmoins, une analyse **ROC** des lésions de types **LM** et **LMM** comme seules représentantes des classes malignes met en avant des performances moins homogènes entre les diverses classes. Ces scores **AUC** dont les valeurs varient entre 0,79 et 0,92 représentent une limite de ces méthodes à différencier les tissus bénins et les tissus malins. Ces diverses courbes sont visibles sur la Figure 5.19.

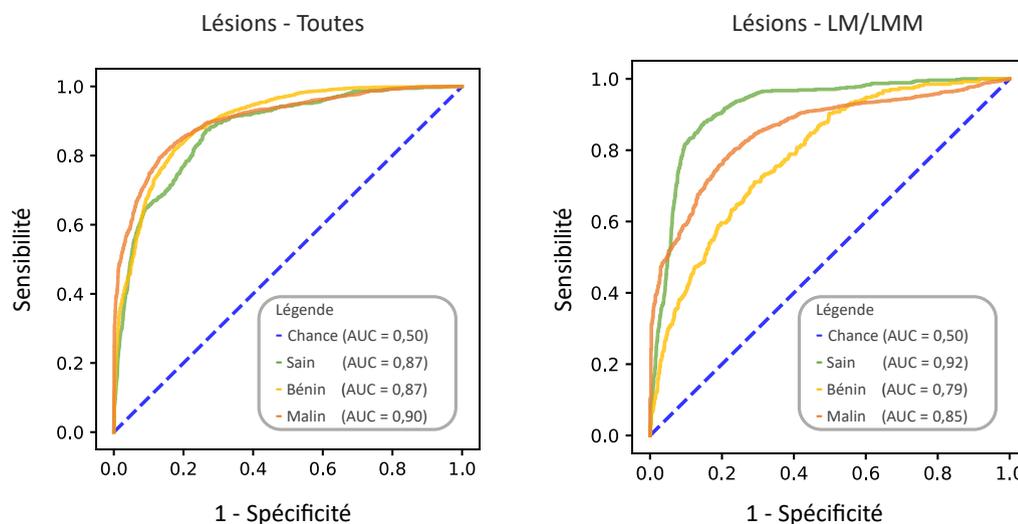


Figure 5.19 : Résultats sous forme de courbes **ROC**. À gauche, résultats sur l'ensemble des lésions d'étude ; À droite, résultats excluant les lésions malignes différentes des types **LM** et **LMM**.

5.6/ DISCUSSION

Suite aux précédents résultats, plusieurs constats peuvent être formulés pour donner suite à la mise en pratique des méthodes d'extraction de caractéristiques. Le premier constat concerne les méthodes d'extraction spatiales, et plus particulièrement l'utilité de la méthode proposée par Haralick et al. [80] sur les données de **MCR** utilisées dans ce travail. En revanche, l'ajout de mesures de premier ordre ne semble pas ou peu impacter les performances. Le second constat concerne les méthodes d'extraction fréquentielles, et en particulier la supériorité sur ces données de l'extraction de caractéristiques par la transformée en ondelettes comparée à l'extraction de caractéristiques par la transformée de Fourier. Ce travail a également confirmé la faible influence de l'ondelette mère sur ces mêmes données, bien que l'ondelette mère de Daubechies soit plus performante dans ce travail. Enfin, le dernier constat concerne la proposition de méthodes sur la base de transfert de connaissance à partir d'architectures de **CNN** pré-entraînées, dont les résultats sont significativement supérieurs aux autres méthodes. Sur l'ensemble des architectures évaluées, la combinaison mêlant l'architecture ResNet-50 pré-entraîné sur ImageNet à une couche de *Global Pooling - Moyenne* produit les résultats les plus pertinents. Cette combinaison est plus amplement examinée dans la suite de ce travail.

En termes de modèles de classification, l'un de ces constats concerne le modèle **SVM - Linéaire** dont les performances et la stabilité sont les plus pertinentes dans la plupart des expériences menées lors de ce chapitre. Par opposition, le modèle **SVM - RBF** a été essentiellement pertinent à l'aide d'une normalisation par Minimum / Maximum, sa performance et stabilité chutant de manière drastique lorsque les caractéristiques sont nombreuses. Du côté des méthodes basées sur les arbres de décision, ces techniques sont globalement peu influencées par les méthodes de normalisation. Comme attendu, les **CART** sont performants dans la plupart des situations, exceptées celles par transfert de connaissances dans lesquelles le nombre de caractéristiques est trop important. À cette fin, les techniques de **RF** et **GB** ont été particulièrement adaptées, les **GB** étant légèrement moins stables que les **RF**. Pour finir cette analyse des modèles, l'utilisation de **MLP** dans le but d'observer l'effet de la non-linéarité semble ne pas avoir aidé à améliorer la classification.

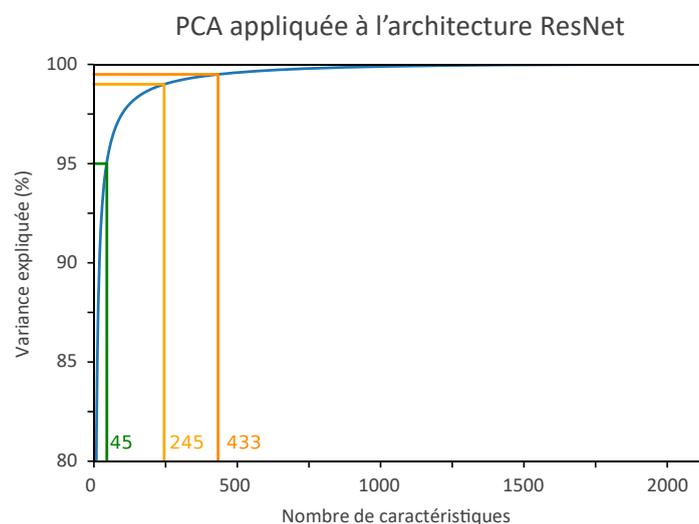


Figure 5.20 : Analyse de la variance expliquée cumulée par l'application du **PCA**. Le nombre de caractéristiques est représenté pour des variances expliquées de 95 % en vert, 99 % en jaune et 99,5 % en orange.

Afin de compenser le nombre élevé de caractéristiques extraites par les **CNN**, des techniques de réduction mises en place afin de pallier ce problème sont évaluées. Dans ce travail, leur utilisation réduit les performances de prédiction, mais également la stabilité en augmentant sensiblement l'écart-type. Bien que moins performants, des résultats intéressants sont obtenus dans une configuration utilisant ResNet-50 et du **PCA** avec une variance expliquée de 99 % permettant de retirer près des trois quarts des caractéristiques. La relation entre variance expliquée et caractéristiques par cette méthode est visible sur la Figure 5.20.

La dernière méthode mise en œuvre pour optimiser les performances de classification des images concerne le balancement des annotations. Dans le contexte de ce manuscrit, la méthode par pondération est préférée aux autres méthodes de sa catégorie car elle propose une correction du non-balancement avec un faible coût de mise en œuvre, tout en garantissant de meilleures performances sur l'ensemble des expériences menées.

En conclusion, ce premier chapitre expérimental a permis de mettre au point un processus permettant de traiter les données **MCR**. Les expérimentations menées permettent d'accomplir une classification avec un score F_1 de 0,77 associé à un écart-type de 0,04 pour 3 classes et de 0,82 associé à un écart-type de 0,02 pour la classification d'éléments malins. De plus ce processus a permis l'obtention d'une **AUC** de 0,90 sur l'ensemble des données et de 0,85 sur les données excluant les lésions différentes des types **LM** et **LMM**. Néanmoins, des améliorations sur la qualité de ce processus de diagnostic sont à prévoir, ainsi que sur la possible prise en compte de l'intersection d'évènements tels que la présence de tissus caractéristiques auprès de follicules pileux.

AMÉLIORATION DU DIAGNOSTIC IMAGE

Les expérimentations conduites lors du précédent chapitre ont tenté d’apporter une première réponse pour la séparation des images de **Microscopie Confocale par Réflectance** et menées selon les annotations saines, bénignes ou encore malignes. Pour cela, ce travail s’est intéressé aux méthodes permettant de caractériser au mieux les aspects de texture de ces images par l’extraction de caractéristiques pertinentes mais également aux méthodes de classification permettant la meilleure séparation des images selon les précédentes annotations.

Néanmoins, ces méthodes abordent la problématique par une compréhension globale des images de **Microscopie Confocale par Réflectance**, dont le principe même est considéré comme une limitation dans ce nouveau chapitre. Ainsi, de nouveaux schémas sont considérés afin d’améliorer le diagnostic de ces images. Des méthodes exploitant le principe d’échelles multiples, de fenêtre glissante, ou encore exploitant les **Convolutional Neural Network** par réglage fin de bout en bout sont mises en œuvre dans ce nouveau chapitre.

Sommaire

6.1	Méthodologie	110
6.2	Approche multi-échelle	111
6.2.1	Décomposition en ondelettes multi-échelle	111
6.2.2	Décomposition spatiale multi-échelle	113
6.3	Approche par fenêtre glissante	115
6.4	Réglage fin de réseaux de convolution	118
6.4.1	Présentation et mise en oeuvre de la méthode	118
6.4.2	Augmentation de données	119
6.4.3	Programme d’apprentissage	120
6.5	Présentation des résultats	121
6.5.1	Protocole d’expérimentation	121
6.5.2	Résultats des expérimentations	122
6.5.3	Méthodes de visualisation	125
6.6	Discussion	128

6.1/ MÉTHODOLOGIE

Lors du précédent chapitre, diverses approches ont été proposées afin de comprendre les composants essentiels d'une image dans son intégralité, et de permettre ainsi la séparation des éléments **sains**, **bénins** et **malins**. Cette approche suppose que l'information extraite afin de caractériser ces images est suffisante pour permettre une séparation selon les trois précédentes annotations. Cette hypothèse n'est que partiellement juste puisqu'une même image de la modalité de **MCR** peut comporter divers types de tissus selon ces trois annotations, dont la hiérarchie d'inclusion entre annotations est visible sur la Figure 6.1.

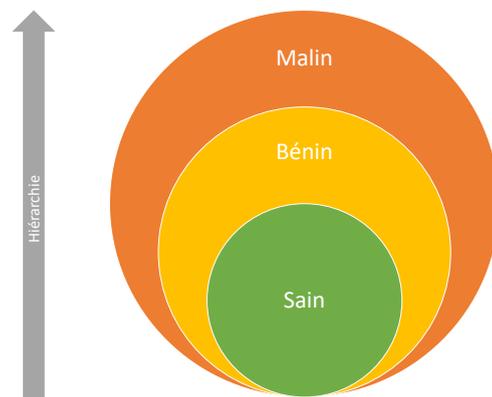


Figure 6.1 : Schéma représentatif de la relation hiérarchique entre les annotations. Les images malignes peuvent ainsi contenir des tissus bénins et sains ; Les images bénignes peuvent contenir des tissus sains ; Les images saines ne contiennent que des tissus sains.

Dans ce nouveau chapitre dédié à l'amélioration du diagnostic de l'image, les observations du Chapitre 5 sont reprises, en privilégiant notamment l'utilisation de la **normalisation par Minimum / Maximum** et le **modèle SVM à noyau linéaire**. Les représentants les plus performants des méthodes d'extraction de caractéristiques sont également conservés dans ce travail, dont :

- pour la **catégorie de méthodes spatiales**, les caractéristiques définies par Haralick et al. [80],
- pour la **catégorie de méthodes fréquentielles**, les caractéristiques sur base d'extraction en ondelettes,
- pour la **catégorie de méthodes de transfert de connaissances**, l'architecture ResNet-50 associée à une couche de Global Pooling - Moyenne.

Ainsi, des schémas alternatifs sont envisagés afin d'extraire une information nécessaire à la séparation des données de la modalité de **MCR**. Ce chapitre débute par la présentation de **méthodes multi-échelle**, en supposant que l'information comporte plusieurs niveaux d'interprétations. Puis, s'ensuit la description de **méthodes par fenêtre glissante** pour extraire localement une information propre aux tissus. Afin d'étendre ce principe d'approche locale, ce travail s'intéresse à l'utilisation du **réglage fin sur des CNN**. Enfin, ce chapitre propose **une analyse des résultats** issus de ces diverses méthodes, lesquels aboutissent à **une discussion** permettant d'approfondir ces résultats.

6.2/ APPROCHE MULTI-ÉCHELLE

Comme évoqué précédemment, les traitements réalisés jusqu'alors ne permettent qu'une compréhension globale de l'image. Cette vision de l'extraction de caractéristiques peut être erronée si l'on suppose une interprétation de l'image à divers niveaux de compréhension.

C'est ainsi que des hypothèses de compréhension de l'image à échelles multiples ont émergé, bien que peu relatées dans la littérature. Ces approches ont été employées à différentes fins : (i) de détection de frontières dans un contexte d'images naturelles [165], (ii) de segmentation d'images [56, 7], (iii) de détection d'objets [63] ou d'actions [150], (iv) ou encore de classification d'images médicales [5, 189]. En termes de traitements appliqués à la modalité de MCR, le travail de Wiltgen et al. [209] semble être le plus proche de la thématique de ce chapitre et propose entre autres des méthodes de prise en charge par échelles multiples. Ainsi, cette section est scindée en deux parties respectives, avec d'une part l'application de ce principe à la décomposition en ondelettes et d'autre part l'application de ce principe à des techniques d'extractions spatiales.

6.2.1/ DÉCOMPOSITION EN ONDELETTES MULTI-ÉCHELLE

La première approche présentée se base sur la décomposition en ondelettes multi-échelle. En effet, ce principe a été démontré comme judicieux dans de nombreux domaines impliquant l'analyse d'images [48]. Cette décomposition peut être effectuée sous la forme de schémas qualifiés soit de **diadique**, soit de **pyramidal**. Ces deux modes sont schématisés sur la Figure 6.2. L'un des articles de référence de ce travail préconise l'utilisation du schéma de décomposition diadique pour l'extraction de caractéristiques de texture sur images de MCR [209]. Pour cela, les auteurs se focalisent sur l'utilisation d'une **ondelette mère de Daubechies** et emploient cette transformée à cinq niveaux de détails successifs sous forme d'arbre diadique. Les mesures associées à chaque bande de fréquences de la transformée sont les mêmes que dans le Chapitre 5 et se basent sur l'extraction de l'**écart-type**, de l'**énergie** et de l'**entropie** pour un total de **39 caractéristiques**. Suite aux résultats obtenus et présentés dans le Chapitre 5, ce procédé est également évalué selon l'**ondelette mère de Haar**.

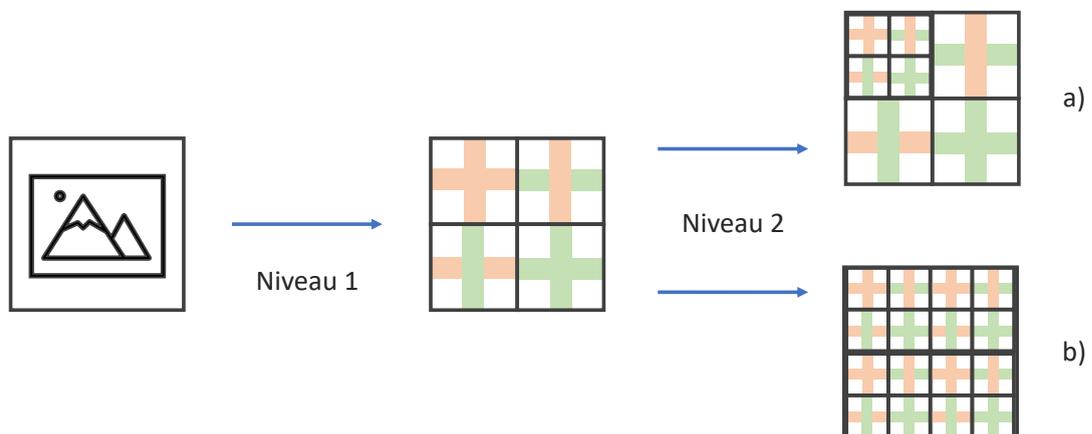


Figure 6.2 : Schématisation des deux principaux types de décomposition successive par ondelettes. En a), schéma de décomposition multi-échelle en ondelettes dit **diadique** ; En b), schéma de décomposition multi-échelle en ondelettes dit **pyramidal**.

La seconde méthode sur laquelle ce travail prend appui est une extension de la transformée en ondelettes de l'article de Wiltgen et al. [79]. Ainsi, cette nouvelle méthode reprend la précédente décomposition selon un schéma diadique, et en modifie le **nombre de décompositions à quatre niveaux**, au lieu des cinq initialement prévus. Selon cette même méthode, une loi normale généralisée centrée, dont la densité de probabilité f est décrite par l'Equation (6.1), est ajustée à la distribution de valeurs de chaque niveau de décomposition. Enfin, les auteurs de l'étude ne retiennent comme caractéristiques que **les paramètres d'échelle α et de forme β** de cette loi normale généralisée pour un total de **24 caractéristiques**, et où Γ représente la fonction gamma.

$$f(x) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(\frac{x}{\alpha})^\beta} \quad (6.1)$$

Afin de gérer ces diverses caractéristiques, ce travail s'emploie à l'utilisation d'un **modèle de type SVM à noyau linéaire**, démontré empiriquement efficace pour la classification de données fréquentielles lors du Chapitre 5. Les hyperparamètres employés afin de tirer le meilleur parti de ce modèle sont listés dans la Table 6.1. L'implémentation du modèle SVM utilisé provient de la bibliothèque logicielle *Scikit-learn* [151].

Modèle	Hyperparamètres	Valeurs
SVM - Noyau linéaire	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]

Table 6.1 : Table reprenant le modèle de classification et ses hyperparamètres employés sur les caractéristiques issues de la décomposition en ondelettes multi-échelle.

Les étapes de transformée en ondelettes et la décomposition en ondelettes multi-échelle de cette expérimentation sont réalisées à l'aide de la bibliothèque logicielle *PyWavelets* [113]. Quant aux opérations d'ajustement de la loi généralisée, celles-ci s'appuient sur la bibliothèque logicielle *SciPy* [204]. Ces expérimentations basées sur la décomposition en ondelettes multi-échelle sont recensées dans la Table 6.2 ainsi que leurs nombres de caractéristiques associées à chacune d'entre elles.

Méthode	Nombre de caractéristiques
Ondelettes - Haar	39 (13×3)
Ondelettes - Daubechies / Wiltgen [209]	39 (13×3)
Halimi [79]	24 (12×2)

Table 6.2 : Liste des méthodes de décomposition en ondelettes multi-échelle et leur nombre de caractéristiques extraites associées.

6.2.2/ DÉCOMPOSITION SPATIALE MULTI-ÉCHELLE

Lors de la précédente sous-section, les approches par décomposition en ondelettes multi-échelle ont été présentées. Toujours dans cette optique d'approche multi-échelle, divers travaux se sont orientés afin de permettre une classification d'images médicales sur la base d'extraction de caractéristiques spatiales [5, 200]. Les prochaines étapes s'inspirent de ces travaux mais également des conclusions apportées par le Chapitre 5 sur l'extraction de caractéristiques spatiales et par transfert de connaissances. Ainsi, sont mis en œuvre deux processus multi-échelle dans ce chapitre, présentés lors des prochains paragraphes.

La première méthode proposée se base sur une approche de type fusion de caractéristiques avant l'étape de classification, déjà rencontrée dans des travaux à but similaire [150, 5]. Afin de permettre l'agrégation de ces caractéristiques, une **normalisation par Minimum / Maximum** associée à un **modèle SVM à noyau linéaire** est employée, démontré efficace de manière empirique lors du Chapitre 5 et robuste dans des contextes à fortes dimensions. À nouveau, l'implémentation du modèle SVM employé dans ce travail provient de la bibliothèque logicielle *Scikit-learn* [151] et les hyperparamètres utilisés par ce modèle sont recensés dans la Table 6.3. Le principe de fonctionnement de cette première méthode est schématisé sur la Figure 6.3.

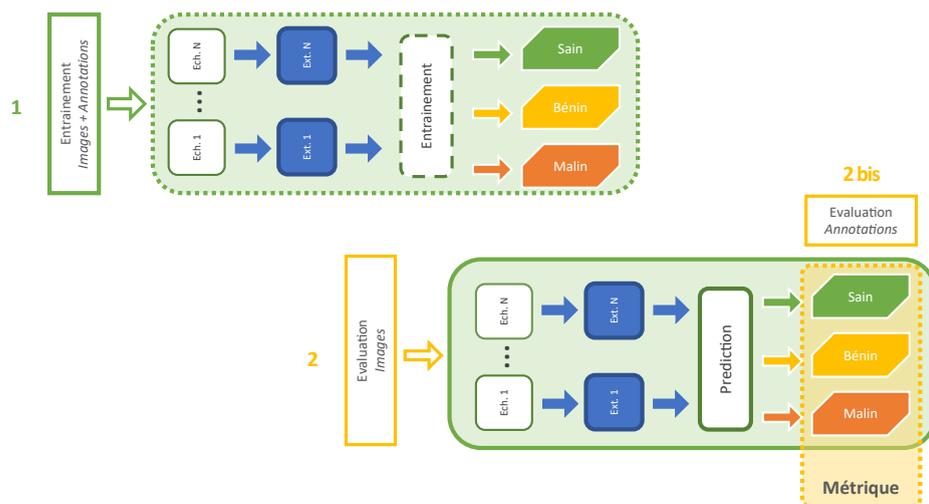


Figure 6.3 : Schéma de représentation du système multi-échelle, sur la base d'une extraction à diverses échelles puis par l'agrégation des caractéristiques avant la procédure de classification.

La seconde méthode correspond à un schéma en deux temps, dans lequel chaque modèle de prédiction est spécialisé à une échelle spatiale précise. À cette fin, est employée pour chacune de ces échelles une **normalisation par Minimum / Maximum** associée à un **modèle de type SVM à noyau linéaire** pour les raisons évoquées précédemment, dont les hyperparamètres sont recensés dans la Table 6.3. Ce faisant, il est alors nécessaire de réunir ou **fusionner** ces informations en une unique décision à l'aide d'**un vote à la majorité** [86, 110], dont le principe ne nécessite aucune phase d'entraînement. Cette technique est ainsi employée : (i) sur les **scores** définis par l'ensemble **R**, (ii) mais également sur les **décisions** définies par l'ensemble **N**. Une première méthode est mise en place à l'aide d'**une fonction de moyenne appliquée aux scores**, de laquelle est extrait le score le plus élevé. Ce schéma permet de notifier d'**une tendance globale** d'une classe par rapport aux autres. Une seconde méthode est mise en place à l'aide d'**une fonction de maximum appliquée**

aux scores, de laquelle est extrait le score le plus élevé. Ce schéma permet d'isoler la classe recueillant la confiance de prédiction la plus élevée. Enfin, la dernière méthode est mise en place par recensement de la classe qui a recueilli **le plus grand nombre de décisions**. Le processus global de cette expérience est présenté sur la Figure 6.4.

Modèle	Hyperparamètres	Valeurs
SVM - Noyau linéaire	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]

Table 6.3 : Table recensant les paramètres du modèle de classification employé pour réaliser la prédiction sur des caractéristiques spatiales multi-échelle. Ce modèle associé à ces hyperparamètres est employé pour la classification par échelle mais également pour l'agrégation des multiples échelles.

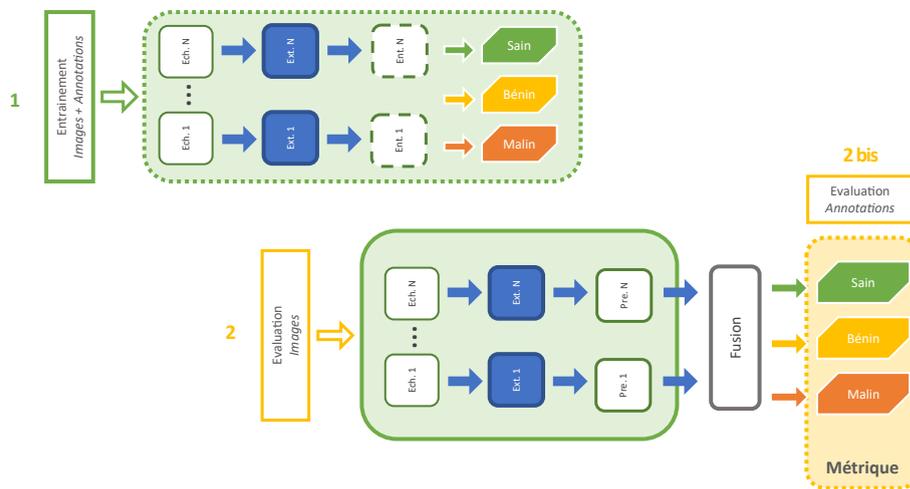


Figure 6.4 : Schéma de représentation du système multi-échelle, sur la base d'extractions et de prédictions à diverses échelles et par agrégation de ces décisions.

Ainsi, les diverses expériences de cette sous-section sont menées à l'aide des modes d'extraction précédemment étudiés dans le Chapitre 5. Dans un premier temps, les descripteurs d'Haralick et al. [80] sont employés et permettent l'extraction de 56 caractéristiques. Dans un second temps, le transfert de connaissances par l'utilisation de l'architecture de CNN ResNet-50 [83] pré-entraîné sur la base *ImageNet* [23] est employé et permet l'extraction de 2048 caractéristiques. Les expériences menées lors de cette sous-section sont recensées dans la Table 6.4.

		Nombre de caractéristiques	
Fusion de caractéristiques	Haralick	224	(56 × 4)
	ResNet-50	8192	(2048 × 4)
Fusion de prédictions	Haralick	56 / échelle	4 (1 × 4) / Décisions
	ResNet-50	2048 / échelle	12 (3 × 4) / Scores

Table 6.4 : Listes des méthodes sur base de décomposition spatiale multi-échelle et leur nombre de caractéristiques extraites associées.

6.3/ APPROCHE PAR FENÊTRE GLISSANTE

Lors de la section précédente, des schémas multi-échelle permettant une compréhension à divers niveaux ont été évoqués. Cette nouvelle section va s'orienter vers une compréhension des données à basse échelle, en partant de l'hypothèse qu'il existe un niveau de détail pour lequel les tissus peuvent être caractérisés avec suffisamment de précision.

Ce type d'approche est qualifié d'approche **par fenêtre glissante** ou par *sliding window*. Néanmoins, le premier terme permet de visualiser le principe de fonctionnement consistant à balayer un espace de plus grande dimension par une fenêtre de dimension inférieure. Cette technique a permis des avancées pour diverses applications, dont :

- la détection et localisation d'objets sur des images naturelles [81] ou encore la détection de cancer et de parties anatomiques sur images de radiologie [84], par motivation d'obtenir l'emplacement précis de certaines informations,
- la détection d'anomalies sur des images histologiques mais encore la classification de données conséquentes [88, 4, 207], motivée par la contrainte matérielle impliquée par la taille de ces images ou par manque d'annotations.

Les données images de **MCR** mises à la disposition de ce travail corroborent ce type de schéma, chacune d'entre elles contenant une composition de tissus divers. Ces labels au niveau de l'image répondent à une hiérarchie mise en place pour qualifier les relations entre ces tissus. Pour rappel, une annotation **maligne** qualifie une donnée comportant au minimum des tissus typiques d'une pathologie maligne mais peut également comporter des tissus bénins et sains ; une annotation **bénigne** ne comporte que des tissus bénins ou potentiellement sains. Ces relations sont représentées sur la Figure 6.1.

Ce travail se distingue de certains travaux [88, 4] dans la mesure où il bénéficie d'une base de *sous-images* d'une taille de **250 × 250 pixels**, annotées pour le besoin par l'un des spécialistes en dermatologie. Cette taille est un compromis entre le niveau de détail des types de tissus observables et les conclusions de Wiltgen et al. [209] faisant mentions de tailles similaires. Néanmoins, afin de s'affranchir partiellement de l'influence de la taille de la fenêtre glissante, **une augmentation de ces sous-images à une taille de 500 × 500 pixels par mosaïquage** est également employée.

Modèle	Hyperparamètres	Valeurs
SVM - Noyau linéaire	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]

Table 6.5 : Table reprenant le modèle et les hyperparamètres du modèle de classification employé pour la détection à basse échelle des tissus dans le processus de fenêtre glissante.

Dans le but de mettre en oeuvre le schéma de détection à basse échelle, les techniques d'extraction de caractéristiques jugées les plus pertinentes du Chapitre 5 sont utilisées, à savoir : (i) la méthode d'Haralick et al. [80] comme représentante des méthodes spatiales, (ii) la méthode par décomposition en ondelettes de Daubechies comme représentante des méthodes fréquentielles, (iii) et enfin la méthode d'extraction par l'architecture ResNet-50 associée à une couche de Global Pooling - Moyenne comme représentante des méthodes par transfert de connaissances. Cette détection à basse échelle est réalisée à l'aide d'**une normalisation par Minimum / Maximum** associée à **un modèle SVM à noyau linéaire**, cette combinaison ayant été jugée la plus pertinente lors du Chapitre 5 et dont les paramètres sont listés dans la Table 6.5.

Afin de gérer les prédictions obtenues au niveau de la fenêtre glissante, plusieurs stratégies sont envisagées sur les décisions et les scores. Ainsi, plusieurs approches sont considérées, dont :

- une approche au niveau des **décisions**, défini par l'ensemble **N**. Cette approche aboutit à un vecteur de taille $1 \times n$ où n correspond au nombre d'éléments extraits dans l'image. Deux méthodes ont été définies :
 - par **priorité**, dans laquelle est privilégiée la détection d'un élément d'une classe hiérarchiquement supérieure comme décision globale à l'image,
 - par seuils **dynamiques**, dans laquelle sont déterminés des seuils optimaux d'éléments nécessaires à chaque classe pour décréter l'appartenance à l'une d'entre elles, et combinés à un système hiérarchique privilégiant les classes prioritaires.
- une approche au niveau des **scores**, défini par l'ensemble **R**. Cette approche aboutit à un vecteur de taille $s \times n$ où s correspond au nombre de scores de classes extrait et n correspond au nombre d'éléments extraits dans l'image. Deux méthodes ont été définies, dont :
 - par seuil **dynamique**, dans un premier temps cette méthode procède à une réduction de l'information à l'aide d'une fonction maximum, permettant de réduire celui-ci à vecteur de taille $1 \times s$ focalisé sur la confiance maximum en chaque classe. Dans un second temps, des seuils optimaux représentant la confiance minimale nécessaire sont déterminés sur ces scores, combinés là encore à un système hiérarchique privilégiant les classes prioritaires.
 - par **classification**, le vecteur de scores est alors considéré comme un vecteur de caractéristiques fourni en entrée d'un modèle **SVM** linéaire, dont les hyperparamètres utilisés sont ceux de la Table 6.5.

Dans ce travail, le scénario portant sur l'extraction de caractéristiques seules à l'aide de la fenêtre glissante, puis à la fusion de ces caractéristiques au niveau de l'image, n'est pas considéré. En effet, ce type de scénario génère bien trop de caractéristiques pour pouvoir être correctement traité au niveau de l'image. De plus, ce type de schéma perd l'une des motivations de cette méthode, à savoir isoler les parties d'images susceptibles de contenir les tissus d'intérêt.

Les paramètres propres à cette méthode sont visibles dans la Table 6.6. Afin de visualiser cette méthode, la Figure 6.5 met à disposition un schéma dans lequel ce processus est représenté.

Taille de fenêtre	Chevauchement
250 × 250	0 %
	25 %
	50 %
500 × 500	0 %
	25 %
	50 %

Table 6.6 : Ensemble des paramètres d'expérimentations menées par l'approche de type fenêtre glissante. L'étude de l'impact de la taille de fenêtre est limitée aux valeurs de 250×250 et 500×500 , mais également du chevauchement sur des valeurs de 0 %, 25 % et 50 %.

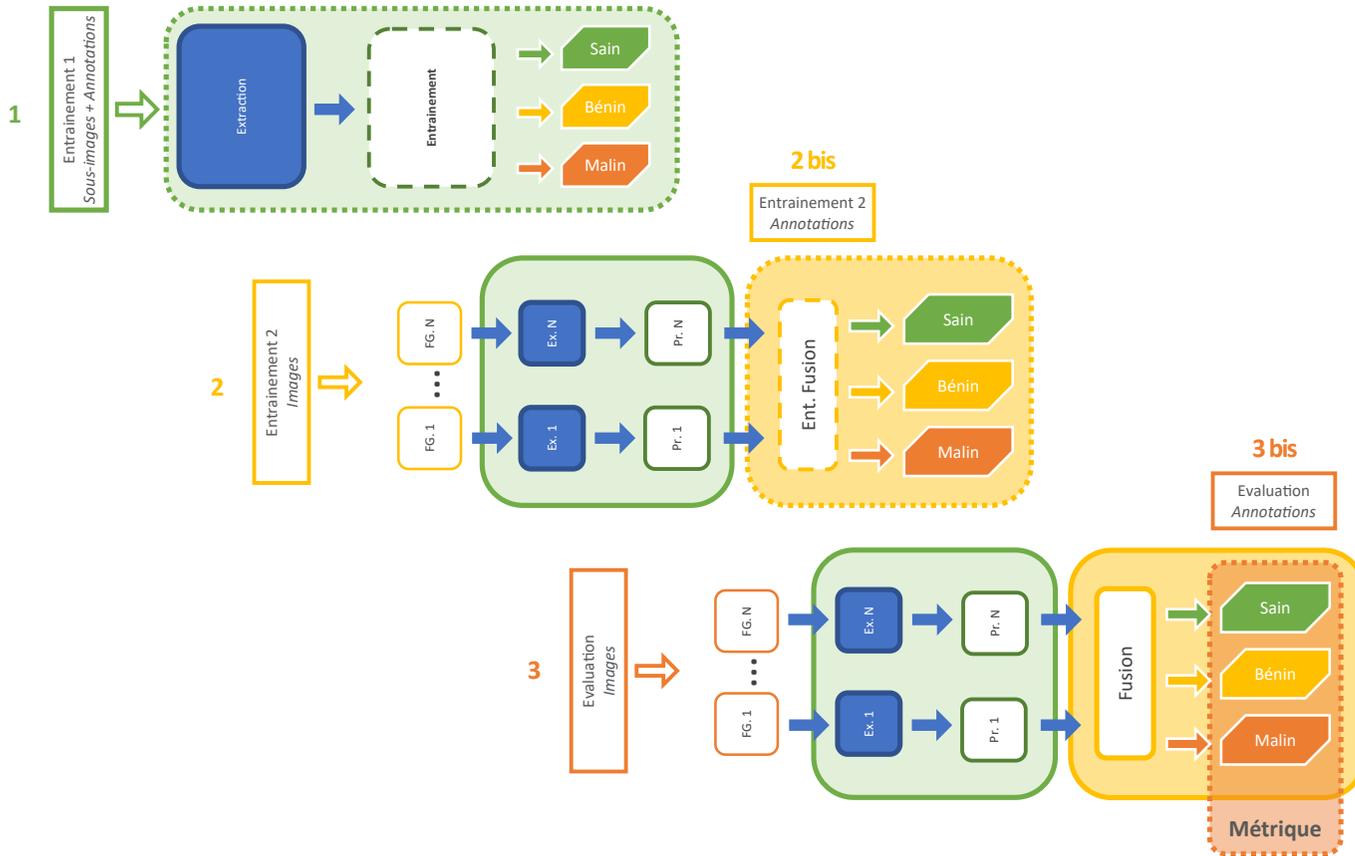


Figure 6.5 : Schéma de représentation du système de prédiction par fenêtre glissante. Ce schéma en deux temps propose de prédire sur une multitude de sous parties de l'image originale puis de fusionner les diverses prédictions afin de fournir une prédiction globale à l'image. Ainsi, une première phase d'entraînement est réalisée à l'aide des sous-images. Puis, le modèle entraîné permet de prédire au niveau des images par principe de fenêtre glissante pour lequel un second modèle permet de définir quels seuils sont critiques.

6.4/ RÉGLAGE FIN DE RÉSEAUX DE CONVOLUTION

Dans cette partie, le réglage fin ou *fine tuning* des CNN est considéré sur cette problématique. En effet, ce principe est sollicité dans de nombreuses thématiques de recherche au moment de la rédaction de ce manuscrit, comme par exemple pour de la détection de plans d'intérêts sur des images par ultrasons [35], mais aussi de plans cardiaques sur des images d'IRM [131]. Également, ce principe est utilisé de manière très encourageante dans des applications visant à de la détection de pathologies sur des acquisitions de rayons X réalisées sur des seins [124] ou encore sur des poumons [71].

C'est de manière naturelle que cette section s'oriente en ce sens, en se consacrant à ses divers aspects à l'aide de sous-sections propres à chaque concept. Tout d'abord, une présentation du principe de réglage fin et des choix opérés est réalisée. Puis, les diverses techniques mises en œuvre pour tirer parti de ce principe sont présentées, dont : (i) l'augmentation de données, (ii) et le programme d'apprentissage.

6.4.1/ PRÉSENTATION ET MISE EN OEUVRE DE LA MÉTHODE

Le réglage fin est une extension de l'apprentissage par transfert, dans lequel sont substituées les couches de classification initiales du réseau par de nouvelles couches adaptées au problème. Cette approche se distingue dans la mesure où tout ou partie du réseau est ré-entraîné à partir des poids existants [187]. Ce mode d'apprentissage est utilisé lorsque les données propres au nouveau problème sont suffisantes, et permet à l'instar du simple apprentissage par transfert d'obtenir un réseau plus adapté au problème cible. Néanmoins, ce mode d'apprentissage **nécessite des ressources et un temps d'entraînement plus conséquent**.

Afin de procéder à ce réglage fin, il est nécessaire de déterminer une architecture cible permettant de gérer au mieux la complexité des données. Par exemple, un travail récent [148] privilégie dans cette démarche l'utilisation d'une architecture *ResNet-50* [83] pour la détection de lésions pulmonaires sur des images de rayons X et propose la localisation de ces dernières. Néanmoins, aucune justification ne semble avoir été apportée sur le choix de cette architecture, et aucune information n'est stipulée sur le type de *Global Pooling* employé par ce travail.

Lors du Chapitre 5, les résultats des architectures de CNN pré-entraînés sur la base *ImageNet* ont été comparés par application du principe de transfert de connaissances. Cette expérience a permis de déterminer de manière empirique la pertinence de l'**architecture de CNN ResNet-50 [83] associée à une couche de Global Pooling - Moyenne** sur cette problématique. Ainsi, ce travail suppose que l'architecture *ResNet-50* [83] associée à une couche de *Global Pooling - Moyenne* couplée au principe du réglage fin est en mesure de spécialiser le réseau à la problématique de lésion de la peau sur des images de MCR, et d'améliorer *in fine* ses performances.

Suivant ce constat, l'initialisation de cette architecture *ResNet-50* est réalisée **à l'aide de poids issus d'un entraînement sur la base ImageNet**. Dans un premier temps, **la dernière couche totalement connectée de ce réseau sur base d'une activation softmax est retirée**, prévue pour supporter les 1000 classes de la base de données *ImageNet*. Dans un second temps, **une nouvelle couche totalement connectée de trois classes sur base d'une activation softmax est ajoutée** afin de supporter les annotations saines, bénignes et malignes. Dans la mesure où ce travail tente de résoudre une problématique à plusieurs classes, **une fonction de coût de type entropie-croisée** est utilisée, suggérée par des travaux dont la problématique est équivalente [10, 148].

En termes de paramètres, la taille des lots ou *batch size* **est réglée à une valeur de 8** pour des

raisons de limitations matérielles. Néanmoins, une taille plus conséquente permet une meilleure stabilité de la convergence des **CNN**. De plus, il est utile de préciser que l'ensemble des données de **MCR** dédié à l'entraînement est utilisé lors de chacune de ces *epochs*. Le nombre d'itérations ou *epochs* n'est pas stipulé ; **un critère d'arrêt en cas de convergence insuffisante après 5 itérations**, dont l'estimation est réalisée sur les données d'entraînement, est jugé suffisant dans ce contexte. Enfin, ce travail emploie une optimisation de type *adam* tel que suggéré par de précédents travaux [10, 148], et dont **le taux d'apprentissage est réglé à une valeur de 0,001** faute de détails sur l'implémentation.

Pour finir sur ces aspects de mise en oeuvre, l'ensemble de ces éléments ont été réalisés à haut niveau à l'aide de la bibliothèque logicielle *Keras* [37] en association à la bibliothèque de bas niveau *Tensorflow* [1].

6.4.2/ AUGMENTATION DE DONNÉES

L'augmentation de données est une technique qui permet de corriger de manière artificielle le nombre d'exemples d'une base de données par application de transformations aux données d'origine [212, 190]. Cette technique est à ne pas confondre avec les méthodes de création d'échantillons virtuels à partir de l'espace des caractéristiques pour corriger les problèmes de balancements. L'augmentation de données, au sens de cette section, permet de corriger un non-balancement des annotations, mais est essentiellement employée dans le but d'enrichir le nombre d'exemples disponibles en contraignant le **CNN** à se focaliser sur des éléments jugés essentiels à la résolution de la tâche visée.

Dans cette sous-section, cette technique est vue selon ces deux aspects mais surtout comme le moyen d'enrichir la base d'images de **MCR** et ainsi d'éviter le sur-apprentissage potentiellement responsable de la dégradation des performances de prédiction. Les techniques considérées à cet effet correspondent à des transformations géométriques linéaires [190], opérées de manière aléatoire à chaque *epochs* lors de l'entraînement. Bien évidemment, ces transformations ne sont appliquées que lors de l'entraînement, les images originales étant employées lors de l'évaluation. Les transformations considérées sont celles susceptibles de se produire dans un contexte clinique lors de l'acquisition des images, dont : (i) la **rotation**, (ii) la **translation**, (iii) ou encore la **symétrie**. De plus, un repliement de l'information est réalisé pour combler le vide généré par ces opérations, jugé plus *réaliste* sur les images traitées, que l'application d'une unique valeur de remplissage.

Ainsi, l'ensemble des paramètres de transformation utilisés lors de l'étape d'augmentation des données sont décrit dans la Table 6.7. Cette étape s'appuie à cette fin sur le module d'augmentation de données de la bibliothèque logicielle *Keras* [37].

Nom	Valeurs
Rotation (en degrés)	[0 à 45]
Translation Horizontale (en pourcentage)	[0 à 20]
Translation Verticale (en pourcentage)	[0 à 20]
Symétrie Horizontale	[Désactivée, Activée]
Symétrie Verticale	[Désactivée, Activée]

Table 6.7 : Table recensant les paramètres de transformation utilisés sur les données images de **MCR** de manière aléatoire lors de l'entraînement.

6.4.3/ PROGRAMME D'APPRENTISSAGE

Le programme d'apprentissage ou *Curriculum Learning* est une démarche visant par analogie à l'humain, à entraîner tout **CNN** en proposant diverses tâches intermédiaires avec une difficulté croissante, afin d'accomplir avec une plus grande efficacité une tâche plus complexe. La raison technique derrière ce principe est de simplifier l'espace de recherche de la tâche cible, afin de trouver plus efficacement un minimum optimal au sein de la fonction de coût, comme suggéré par ses auteurs [13].

De récents travaux menés sur des images médicales, issue de rayons X, ont permis par cette approche d'augmenter les performance de détection de **CNN** sur des cancers du sein [124] ou encore des pathologies pulmonaires [148]. La démarche suivie par ces deux travaux, est de procéder à l'extraction de sous images de taille restreinte autour des dites lésions avérées, en conservant la même propriété d'échelle que les images initiales. Cette extraction permet de réduire la résolution et la complexité du problème en proposant au réseau des images moins volumineuses et davantage focalisées sur des éléments clés afin d'éviter potentiellement des convergences non pertinentes. La gestion de données à échelles multiples sur **CNN** ne donnant pas toujours lieu à des résultats suffisants [203], une seconde étape d'entraînement est ensuite réalisée sur les images de taille réelle afin de former le réseau au problème définitif.

En complément de cette technique, la première phase de l'entraînement du **CNN** est réalisée en deux temps. En effet, l'ajout de nouvelles couches connectées sous-entend une initialisation aléatoire de ces dernières. Ce phénomène peut conduire à une mauvaise optimisation de certains poids, provenant du caractère aléatoire de l'étape d'initialisation. Ainsi, une immobilisation des couches liées à l'extraction des caractéristiques est réalisée dans un premier temps (pré-entraînement), avant de procéder à l'entraînement du réseau dans sa totalité dans un second temps (entraînement). Ce processus d'entraînement en deux temps est schématisé sur la Figure 6.6.

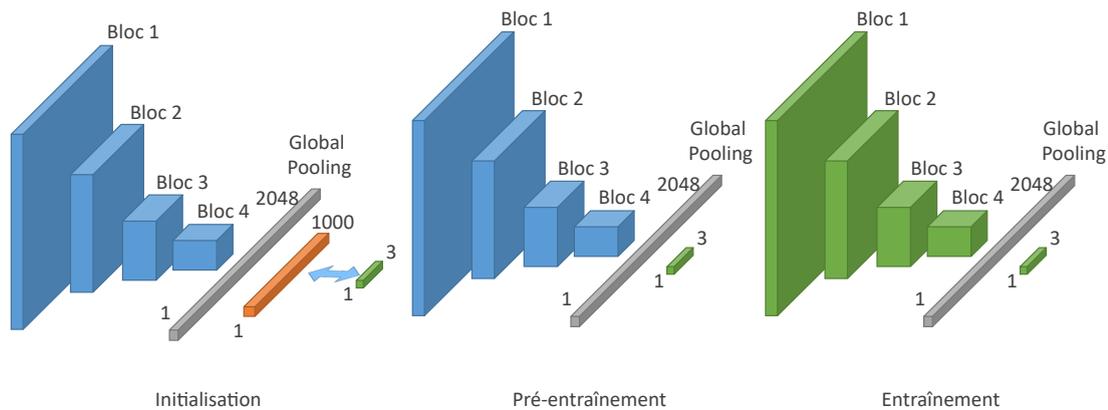


Figure 6.6 : Schéma de représentation du processus de réglage fin utilisé dans ce travail. Celui-ci se décompose d'une étape d'initialisation dans le but de substituer les couches de classification, puis un pré-entraînement permet d'initialiser les poids avant de procéder à l'entraînement.

6.5/ PRÉSENTATION DES RÉSULTATS

Lors des précédentes sections ont été décrites les méthodes envisagées dans le but d'accroître la performance des prédictions au niveau des images de **MCR** mais également leur compréhension. Par cette nouvelle section, les résultats obtenus par ces méthodes sont présentés. Dans une première sous-section, le protocole d'expérimentation suivi pour mesurer les performances de ces méthodes et l'organisation des résultats sont détaillés, puis dans une seconde sous-section, les résultats des diverses méthodes sont présentés.

6.5.1/ PROTOCOLE D'EXPÉRIMENTATION

Le protocole d'expérimentation suivi est similaire à celui du Chapitre 5, afin que les résultats obtenus dans ce chapitre puissent être mis en parallèle et comparés avec ceux précédemment obtenus. Pour rappel, une **validation croisée imbriquée** est employée car les métriques extraites par ce processus sont plus objectives sur les performances du système évalué que des systèmes de validation croisée standards [26]. De plus, cette structure de validation est plus robuste face au biais [26] et démontrée empiriquement moins sensible au biais et à la variance lorsque des valeurs de k sont comprises entre 5 et 10 [92]. Pour des critères similaires à ceux du précédent chapitre, lié au grand nombre d'expérimentations, et de comparaison avec les précédents résultats, **une valeur k de 4** est choisie.

En terme de métrique de validation et d'évaluation, les expérimentations de ce chapitre se basent sur le **score F_1** permettant de représenter les mesures de précision et de rappel au sein d'une même métrique. Par ailleurs, cette métrique a été employée lors du Chapitre 5 et possède une bonne robustesse face à des déséquilibres de données. Concernant la présentation des résultats de l'évaluation, chaque méthode propose deux scores. D'une part, **un score pondéré** dont l'objectif est de fournir une indication de performance sur les trois classes. D'autre part, **un score malin** dont l'objectif est de renseigner sur la capacité d'une méthode à isoler particulièrement les données malignes.

D'un point de vue organisationnel, la prochaine sous-section propose un premier volet dédié aux résultats de classification des méthodes proposées, et met en évidence les combinaisons permettant la maximisation des scores ainsi que les résultats complets à l'aide de tables prévues à cet effet. Lors du premier paragraphe, les résultats des expérimentations menées **par multi-échelle** sont présentés. Puis dans un second et troisième paragraphe, les résultats obtenus **par principe de fenêtre glissante** sont présentés à **une taille de 250×250 pixels puis de 500×500 pixels**. Pour finir, un quatrième paragraphe présente les résultats des expérimentations **de CNN associés au principe de réglage fin**.

Pour finir, ce chapitre propose un second volet consacré à des techniques permettant d'aller au-delà du seul score de classification. Le premier de ces paragraphes propose une solution simple permettant d'observer localement les résultats obtenus par fenêtre glissante. Ainsi, la mise en place d'un code couleur sur les décisions et les scores de prédiction, superposé à l'image d'origine permet de visualiser les éléments déterminants lors de la prédiction. Puis, les paragraphes suivants apportent des pistes sur la visualisation de décisions issues de **CNN** par utilisation du principe de **Class Activation Map, CAM** (*alt. : Carte d'Activation de Classe, CAC*).

6.5.2/ RÉSULTATS DES EXPÉRIMENTATIONS

Cette section suit paragraphe par paragraphe l'organisation énoncée et débute par les résultats issus de la famille de méthodes **par échelles multiples**. Pour la catégorie **par ondelettes** et à l'aide de la **méthode de Wiltgen et al. [209]**, un score F_1 de 0,65 associé à un écart-type de 0,08 est obtenu à trois classes ; un score F_1 de 0,71 associé à un écart-type de 0,06 est obtenu sur la classe maligne. Pour la catégorie de méthodes **spatiale** et à l'aide de la **méthode de fusion de caractéristiques**, un score F_1 de 0,63 associé à un écart-type de 0,06 est obtenu à trois classes ; un score F_1 de 0,67 associé à un écart-type de 0,07 est obtenu sur la classe maligne. Pour finir, la catégorie de méthodes par **transfert de connaissances** et à l'aide de la **méthode de fusion de caractéristiques**, un score F_1 de 0,76 associé à un écart-type de 0,06 est obtenu à trois classes ; un score F_1 de 0,83 associé à un écart-type de 0,02 est obtenu sur la classe maligne. Les résultats de l'expérimentation menée sur la famille de méthodes par échelles multiples sont visibles dans la Table 6.8.

Catégorie	Méthode	Score F_1 - Pondéré	Score F_1 - Malin
Fréquentielle	Haar	0,65 ± 0,08	0,70 ± 0,06
	Wiltgen [209]	0,66 ± 0,06	0,71 ± 0,06
	Halimi [79]	0,38 ± 0,02	0,41 ± 0,04
Spatiale	Fusion de caractéristiques	0,63 ± 0,06	0,67 ± 0,07
	Décisions - Maximum	0,60 ± 0,10	0,64 ± 0,08
	Scores - Moyenne + Maximum	0,60 ± 0,10	0,66 ± 0,07
	Scores - Maximum + Maximum	0,59 ± 0,10	0,65 ± 0,08
	Scores - SVM Linéaire	0,57 ± 0,10	0,65 ± 0,04
Transfert	Fusion de caractéristiques	0,76 ± 0,06	0,83 ± 0,02
	Décisions - Maximum	0,69 ± 0,10	0,78 ± 0,05
	Scores - Moyenne + Maximum	0,69 ± 0,11	0,79 ± 0,06
	Scores - Maximum + Maximum	0,69 ± 0,11	0,79 ± 0,06
	Scores - SVM Linéaire	0,72 ± 0,04	0,79 ± 0,03

Table 6.8 : Résultats issus du processus de classification basé sur les décompositions multi-échelle exprimées à l'aide du score F_1 .

Ce second paragraphe présente les résultats obtenus à l'aide de l'approche **par fenêtre glissante**, à partir **d'une taille de fenêtre de 250 × 250 pixels**. Les résultats sont maximisés par l'utilisation de **l'extraction par transfert de connaissance** sur la base de sous-images, pour laquelle un score F_1 de 0,91 associé à un écart-type de 0,02 est obtenu à trois classes ; un score F_1 de 0,82 associé à un écart-type de 0,03 est obtenu sur la classe maligne. Par la catégorie de **méthodes spatiales** combinée à la méthode de **Décision - Dynamique** et **sans chevauchement** aboutissent à un score F_1 de 0,62 associé à un écart-type de 0,02 pour trois classes ; un score F_1 de 0,68 associé à un écart-type de 0,05 est obtenu sur la classe maligne. Par la catégorie de **méthodes fréquentielles** combinée à la méthode de **Score - Classification** et **avec un chevauchement de 50 %** aboutissent à un score F_1 de 0,61 associé à un écart-type de 0,07 pour trois classes ; un score F_1 de 0,68 associé à un écart-type de 0,06 est obtenu sur la classe maligne. Pour finir, par la catégorie de **méthodes de transfert de connaissances** combinée à la méthode de **Score - Classification** et **avec un chevauchement de 50 %** aboutissent à un score F_1 de 0,78 associé à un écart-type de 0,03 pour trois classes ; un score F_1 de 0,82 associé à un écart-type de 0,02 est obtenu sur la classe maligne. Les résultats détaillés obtenus par principe de fenêtre glissante dont la taille est de 250 × 250 pixels, sont présentés dans la Table 6.9.

Taille	Chevauchement	Mode	Spatial		Fréquence		Transfert		
			Pondéré	Malin	Pondéré	Malin	Pondéré	Malin	
250	Aucune	Sous-images	0,67 ± 0,10	0,41 ± 0,13	0,72 ± 0,07	0,44 ± 0,10	0,91 ± 0,02	0,82 ± 0,03	
		0 %	Décision - Priorité	0,52 ± 0,02	0,67 ± 0,04	0,44 ± 0,06	0,65 ± 0,05	0,58 ± 0,04	0,71 ± 0,05
			Décision - Dynamique	0,62 ± 0,02	0,68 ± 0,05	0,57 ± 0,06	0,67 ± 0,04	0,75 ± 0,02	0,80 ± 0,03
			Score - Dynamique	0,54 ± 0,02	0,63 ± 0,04	0,50 ± 0,08	0,66 ± 0,05	0,70 ± 0,03	0,76 ± 0,02
	Score - Classification		0,60 ± 0,05	0,66 ± 0,04	0,59 ± 0,08	0,65 ± 0,07	0,78 ± 0,03	0,81 ± 0,03	
	25 %	Décision - Priorité	0,49 ± 0,10	0,67 ± 0,10	0,41 ± 0,10	0,65 ± 0,10	0,53 ± 0,05	0,70 ± 0,09	
		Décision - Dynamique	0,62 ± 0,07	0,68 ± 0,11	0,58 ± 0,08	0,66 ± 0,09	0,75 ± 0,03	0,80 ± 0,04	
		Score - Dynamique	0,52 ± 0,03	0,59 ± 0,06	0,49 ± 0,08	0,66 ± 0,10	0,72 ± 0,04	0,78 ± 0,03	
		Score - Classification	0,62 ± 0,06	0,71 ± 0,06	0,57 ± 0,10	0,66 ± 0,08	0,78 ± 0,03	0,81 ± 0,03	
	50 %	Décision - Priorité	0,46 ± 0,11	0,66 ± 0,08	0,39 ± 0,12	0,64 ± 0,08	0,49 ± 0,10	0,68 ± 0,08	
		Décision - Dynamique	0,62 ± 0,03	0,68 ± 0,07	0,59 ± 0,10	0,67 ± 0,08	0,75 ± 0,03	0,80 ± 0,03	
		Score - Dynamique	0,52 ± 0,06	0,63 ± 0,07	0,49 ± 0,10	0,65 ± 0,07	0,72 ± 0,03	0,80 ± 0,01	
Score - Classification		0,61 ± 0,08	0,69 ± 0,08	0,61 ± 0,07	0,68 ± 0,06	0,78 ± 0,03	0,82 ± 0,02		
500	Aucune	Sous-images	0,67 ± 0,08	0,40 ± 0,10	0,70 ± 0,06	0,50 ± 0,08	0,90 ± 0,02	0,80 ± 0,02	
		0 %	Décision - Priorité	0,58 ± 0,08	0,63 ± 0,05	0,23 ± 0,11	0,00 ± 0,00	0,74 ± 0,05	0,76 ± 0,05
			Décision - Dynamique	0,59 ± 0,04	0,63 ± 0,05	0,28 ± 0,05	0,63 ± 0,05	0,72 ± 0,05	0,76 ± 0,05
			Score - Dynamique	0,50 ± 0,05	0,56 ± 0,15	0,53 ± 0,08	0,66 ± 0,04	0,70 ± 0,03	0,78 ± 0,04
	Score - Classification		0,59 ± 0,08	0,67 ± 0,07	0,36 ± 0,04	0,55 ± 0,17	0,69 ± 0,13	0,74 ± 0,08	
	25 %	Décision - Priorité	0,55 ± 0,08	0,64 ± 0,07	0,23 ± 0,11	0,00 ± 0,00	0,77 ± 0,05	0,80 ± 0,04	
		Décision - Dynamique	0,55 ± 0,10	0,63 ± 0,05	0,28 ± 0,05	0,63 ± 0,05	0,76 ± 0,04	0,80 ± 0,04	
		Score - Dynamique	0,53 ± 0,03	0,55 ± 0,09	0,55 ± 0,07	0,67 ± 0,05	0,73 ± 0,03	0,80 ± 0,04	
		Score - Classification	0,58 ± 0,07	0,67 ± 0,06	0,37 ± 0,04	0,55 ± 0,18	0,77 ± 0,05	0,81 ± 0,04	
	50 %	Décision - Priorité	0,57 ± 0,08	0,67 ± 0,11	0,23 ± 0,07	0,00 ± 0,00	0,77 ± 0,03	0,81 ± 0,05	
		Décision - Dynamique	0,58 ± 0,07	0,67 ± 0,11	0,28 ± 0,12	0,63 ± 0,12	0,76 ± 0,03	0,80 ± 0,05	
		Score - Dynamique	0,52 ± 0,05	0,55 ± 0,06	0,55 ± 0,08	0,67 ± 0,13	0,73 ± 0,04	0,80 ± 0,05	
Score - Classification		0,60 ± 0,06	0,65 ± 0,09	0,34 ± 0,05	0,39 ± 0,23	0,78 ± 0,02	0,81 ± 0,04		

Table 6.9 : Résultats issus du processus de classification basé sur le principe de fenêtre glissante exprimé à l'aide du score F_1 . La première partie de ce tableau reprend les résultats obtenus à l'aide d'une fenêtre glissante de taille de 250×250 pixels tandis que la seconde partie se concentre sur les résultats d'une fenêtre glissante de taille de 500×500 pixels.

Afin de terminer sur cette approche **par fenêtre glissante**, ce paragraphe décrit les résultats obtenus à l'aide **d'une taille de fenêtre de 500×500 pixels**. À nouveau, ces résultats sont maximisés par l'utilisation de **l'extraction par transfert de connaissance** sur la base de sous-images, pour laquelle un score F_1 de 0,90 associé à un écart-type de 0,02 est obtenu à trois classes ; un score F_1 de 0,80 associé à un écart-type de 0,02 est obtenu sur la classe maligne. Pour la catégorie de **méthodes spatiales** combinée à la méthode de **Score - Classification** et **avec chevauchement de 50 %**, un score F_1 de 0,60 associé à un écart-type de 0,06 est obtenu à trois classes ; le chevauchement doit être à une valeur de 25 % pour permettre d'obtenir un score F_1 de 0,67 associé à un écart-type de 0,06 est obtenu sur la classe maligne. Pour la catégorie de **méthodes fréquentielles** combinée à la méthode de **Score - Dynamique** et **avec un chevauchement de 25 %**, un score F_1 de 0,55 associé à un écart-type de 0,07 est obtenu à trois classes ; un score F_1 de 0,67 associé à un écart-type de 0,05 est obtenu sur la classe maligne. Pour finir, la catégorie de **méthodes de transfert de connaissances** combinée à la méthode de **Score - Classification** et **avec un chevauchement de 50 %**, un score F_1 de 0,78 associé à un écart-type de 0,02 est obtenu à trois classes ; un score F_1 de 0,81 associé à un écart-type de 0,04 est obtenu sur la classe maligne. Ces résultats sont détaillés dans la Table 6.9.

Ce nouveau paragraphe clôture à proprement parler les résultats de classification, en énonçant ceux issus des **CNN**. Dans un premier temps, le **processus de réglage fin seul** est évalué afin de mesurer ses performances dans le cadre de son utilisation sur l'architecture **ResNet-50**. Par l'utilisation d'une couche de **Global Pooling - Moyenne**, un score F_1 de 0,56 associé à un écart-type de 0,05 est obtenu à trois classes ; un score F_1 de 0,78 associé à un écart-type de 0,05 est obtenu sur la classe maligne. Dans un second temps, le **processus de réglage fin couplé au programme d'apprentissage** est évalué afin de mesurer l'influence de celui-ci. Par l'utilisation de l'architecture **ResNet-50 couplée à une couche de Global Pooling - Moyenne**, un score F_1 de 0,51 associé à un écart-type de 0,07 est obtenu à trois classes ; un score F_1 de 0,67 associé à un écart-type de 0,05 est obtenu sur la classe maligne. L'ensemble de ces résultats de classification par réglage fin est disponible dans la Table 6.10.

Catégorie	Global Pooling	Score F_1 - Pondéré	Score F_1 - Malin
Réglage fin	Maximum	0,44 ± 0,05	0,69 ± 0,06
	Moyenne	0,56 ± 0,05	0,78 ± 0,05
Programme d'apprentissage	Moyenne	0,51 ± 0,07	0,67 ± 0,05

Table 6.10 : Résultats issus du processus de classification basé sur les **CNN** par réglage fin et programme d'apprentissage exprimés à l'aide du score F_1 .

6.5.3/ MÉTHODES DE VISUALISATION

Enfin, cette section dédiée aux résultats se termine par la proposition d'extensions aux principes de fenêtre glissante et de réglage fin de **CNN**. En effet, ces deux méthodes possèdent par leur nature, la capacité de localiser précisément l'origine d'une information. Dans le cadre de l'approche par fenêtre glissante, c'est de manière directe que ce principe permet de remonter à l'emplacement d'une décision. Deux modes de visualisation sont ainsi proposés, l'un basé **sur les décisions** en utilisant un code couleur propre à chacune des classes mais également **sur les scores** en utilisant le principe de code couleur précédent combiné à l'intensité de la transparence comme indicateur de confiance de la prédiction. Une taille de fenêtre glissante la plus fine permettra par cette technique d'obtenir une information visuelle de meilleure qualité, dualité qu'il est possible d'observer entre les tailles de fenêtre glissante de 250×250 et 500×500 . Des exemples de ces deux propositions de techniques de visualisation sont fournis via la Figure 6.7.

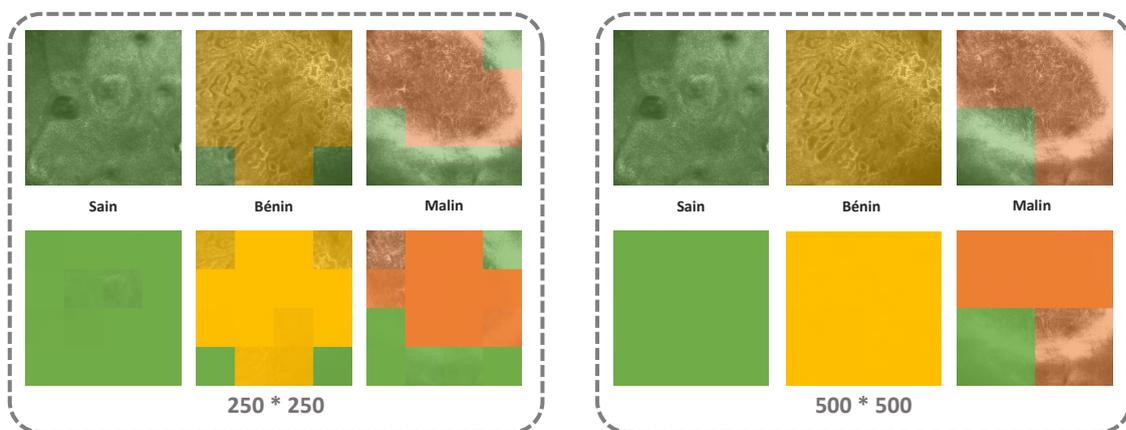


Figure 6.7 : Exemple d'images correctement classifiées à l'aide du principe de fenêtre glissante. À gauche, résultats issus d'une fenêtre de classification de 250 par 250 pixels ; À droite, résultats issus d'une fenêtre de classification de 500 par 500 pixels. En haut, représentations des décisions par code couleur : zones saines en vert, zones bénignes en jaune et zones malignes en orange ; En bas, représentations des scores par les codes couleur précédents et intensités de transparence variables : zones de forte confiance en opaque et zones de faible confiance en translucide.

Le dernier principe de visualisation proposé dans cette sous-section dédiée aux résultats, se base sur les **CNN** en exploitant le principe même de la convolution utilisée au sein de ces réseaux. De nombreux travaux se sont portés sur la compréhension des décisions issues de **CNN** avec **la mise en évidence des filtres auto-déterminés de convolution** [220] ou encore la mise en lumière des zones responsables de l'activation d'une classe par le principe des **Class Activation Map** [224]. La nouvelle méthode proposée dans cette sous-section exploite ce second principe de **CAM**, dont le fonctionnement est présenté sur le schéma de la Figure 6.8. Cette méthode consiste à pondérer le résultat de la dernière couche convolutive par les poids issus de la décision d'une classe. Ces recherches sur les **CAM** ont abouti à divers travaux de détections sur images naturelles dont résultent YOLO [164] ou encore SSD [122]. Ainsi, diverses applications ont vu le jour sur de l'imagerie médicale afin de fournir une visualisation [93] mais également des zones clés de diagnostic [148].

Dans le dernier aspect de ce travail, la méthode des **CAM** est combinée à l'application d'un seuil sur ces valeurs, permettant de définir un masque de segmentation de la zone d'intérêt. Le seuil a été défini de manière arbitraire à une valeur de 85 % du rang entre les valeurs minimum et maximum issues des **CAM**, permettant l'obtention des diverses segmentations présentées sur la Figure 6.9. Sur cette figure, sont représentés des exemples de tissus malins correctement détectés par la méthode, mais également non pertinents sur la partie basse de cette même figure. Ces exemples non pertinents nombreux peuvent s'expliquer par la faible performance obtenue sur l'entraînement des **CNN**. Néanmoins, les intérêts de cette méthode sont multiples, puisque le principe de **CAM** permet par le seul passage d'une image de revenir aux zones d'activations à partir d'une tâche initiale de classification. De plus, cette méthode ne dépend pas de paramètres tel que la taille ou le chevauchement comme cela peut être le cas pour le principe de fenêtre glissante.

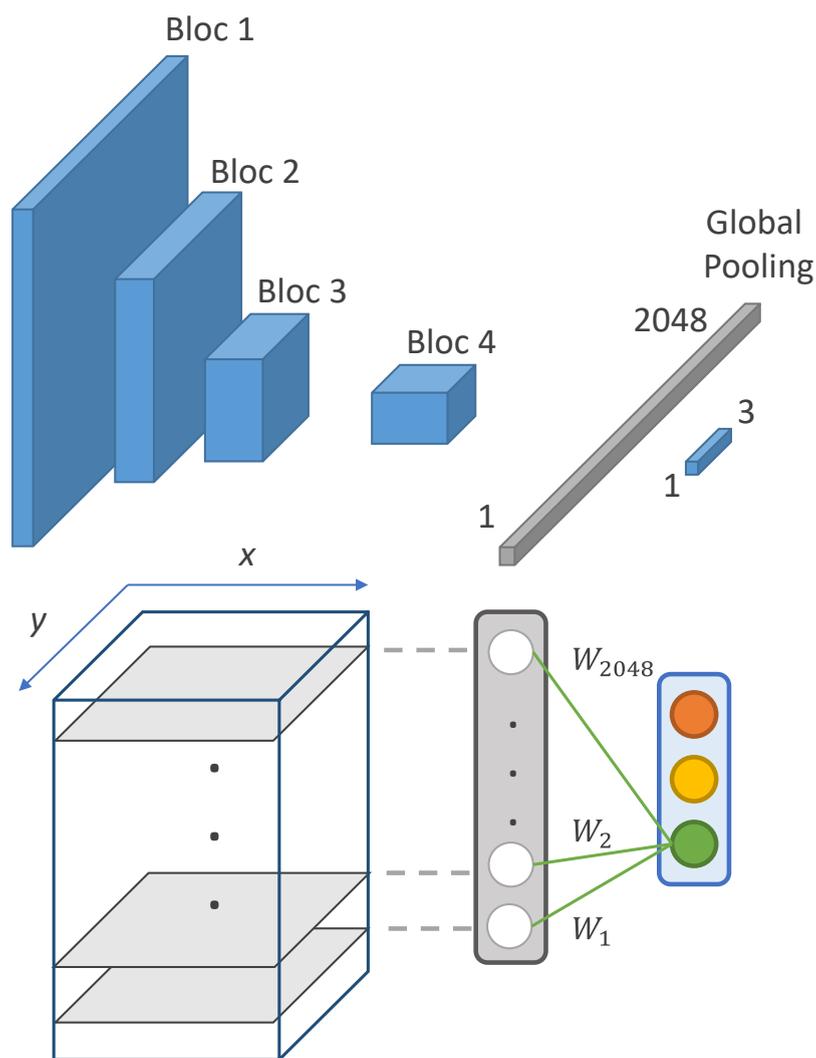


Figure 6.8 : Schéma d'explication du principe de **CAM** employé à but de segmentation des images de **MCR**. Le réseau est entraîné sur la problématique cible, puis *la ou les couches* de classification sont mobilisées dans le but d'obtenir les zones responsables de la décision par pondération des poids de chaque noeud de la classe considérée.

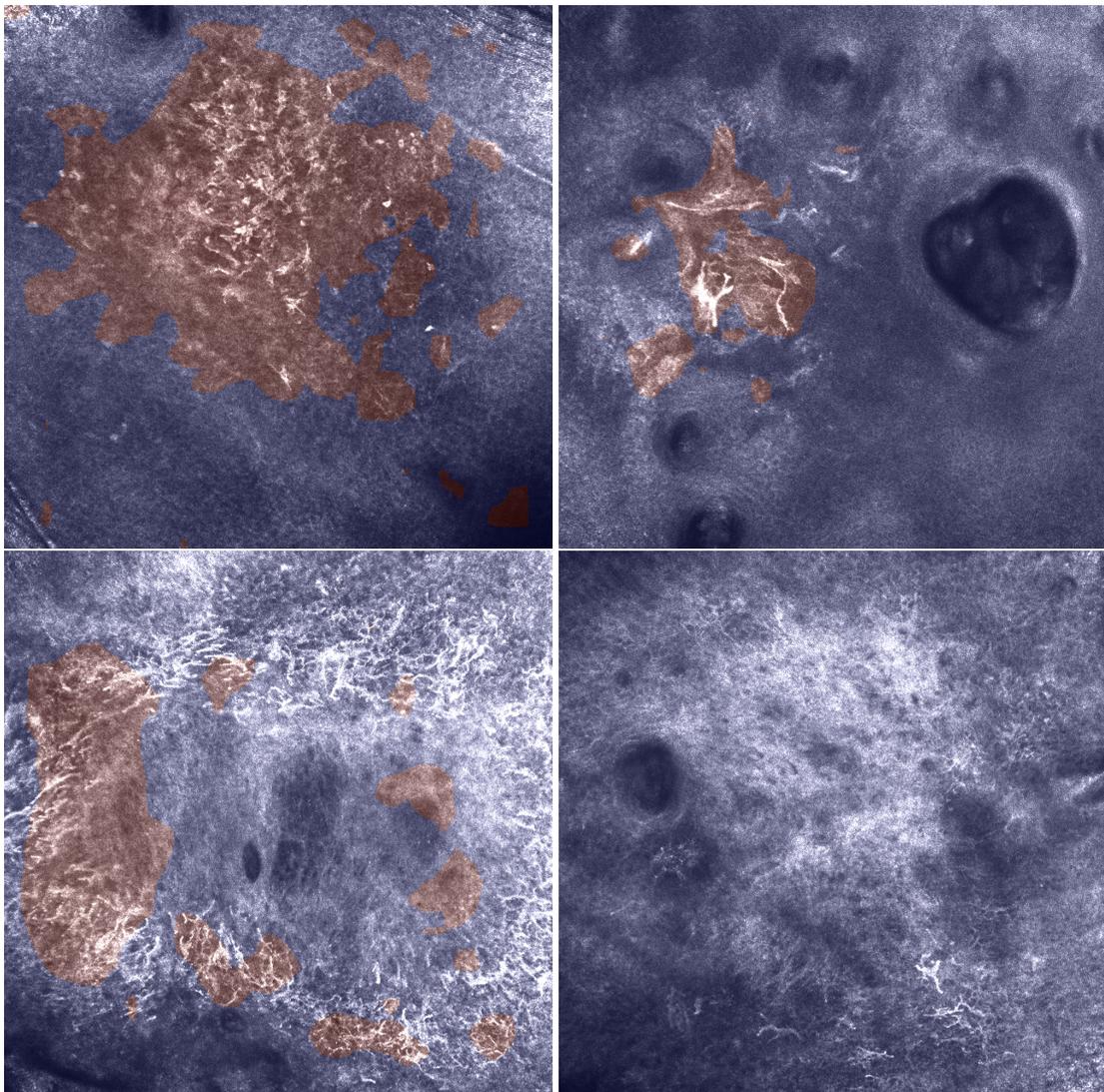


Figure 6.9 : Exemples de l'extraction de zones d'activations par principe de **CAM**, avec application d'un seuil arbitraire à 85 % entre la valeur minimum et maximum. En haut, exemples d'images malignes correctement gérées par ce principe ; En bas, exemples d'images malignes partiellement segmentées ou incorrectement gérées par ce principe.

6.6/ DISCUSSION

En premier lieu, les résultats issus de la décomposition en ondelettes multi-échelle sont sensiblement identiques à ceux présentés lors du Chapitre 5 à trois classes. Une amélioration significative des performances sur la détection d'éléments malins est à noter par la méthode en ondelettes de Wiltgen et al. [209], permettant l'obtention d'un score F_1 de $0,71 \pm 0,06$, initialement à $0,68 \pm 0,03$. L'utilisation de l'ondelette mère de Haar par cette technique ne dégrade que très peu les résultats à deux et trois classes. En revanche, l'évaluation de la méthode proposée par Halimi et al. [79] aboutie à des performances proches de prédictions aléatoires sur les données de MCR exploitées dans ce manuscrit. Des diverses méthodes proposées pour l'utilisation de caractéristiques spatiales et par transfert de connaissances, seule la fusion de caractéristiques s'est réellement avérée fonctionnelle. Cette technique a recueilli par l'utilisation du transfert de connaissances un score F_1 de $0,76 \pm 0,06$ à trois classes et de $0,83 \pm 0,02$ pour la détection d'éléments malins.

En second lieu, une approche par fenêtre glissante a été employée afin d'améliorer la qualité des résultats. Par cette approche, la classification à trois classes obtenue par les diverses techniques d'extraction n'a pas ou peu été améliorée. Sur la détection d'éléments malins, seule la technique employant les descripteurs proposés par Haralick semble avoir recueilli de meilleurs résultats, voyant passer ceux-ci de valeurs de score F_1 $0,68 \pm 0,03$ à $0,71 \pm 0,06$. D'une part, des deux méthodes employant un principe de prédiction sur les décisions, la proposition de méthode sur la base de seuil dynamique semble, comme espéré, la plus judicieuse, mais risque de délaissier certains cas cliniques selon la tolérance de ce seuil. D'autre part, des deux méthodes employant un principe de prédiction sur les scores, la proposition de méthode utilisant un modèle SVM linéaire semble la plus adaptée. Néanmoins, cette méthode risque de favoriser par une pondération plus grande certaines positions de la fenêtre glissante, tandis que la méthode sur seuil dynamique est invariante à cette position. Concernant les hypothèses de la fenêtre glissante, l'une d'entre elles avait été formulée sur le rôle important de la taille de fenêtre dans cette approche, et la tendance des résultats semble meilleure à l'aide d'une fenêtre de taille restreinte de 250×250 mais ces résultats ne sont pas significatifs. Une autre hypothèse supposait l'importance d'un chevauchement des acquisitions en provenance de la fenêtre glissante, et également la tendance des résultats semble plus optimiste sans chevauchement, cependant là encore ces résultats ne sont pas significatifs. Les résultats de cette expérience sont maximisés par l'utilisation de transfert de connaissances avec l'architecture ResNet-50, dans le cadre d'un chevauchement de 50 % mais là encore les résultats ne sont pas suffisamment significatifs pour affirmer l'utilité de ce chevauchement. En addition, ce travail montre qu'il est possible de revenir assez facilement aux décisions à basse échelle par l'utilisation d'un code de couleur ou par opacité. Ce type d'approche pourrait permettre d'expliquer les décisions du système prédictif et d'aiguiller ou de guider le spécialiste à focaliser son attention sur les zones les plus importantes. Il convient néanmoins de solutionner les performances de prédiction de ce système.

Enfin, cette discussion se termine par l'étude des résultats liés au réglage fin dont les expériences ont été réalisées en deux temps, avec un simple réglage fin puis grâce la réalisation d'un programme d'apprentissage à l'aide des sous-images à disposition de ce travail. Par ces méthodes, les résultats ont été fortement impactés dans les deux cas, avec un score F_1 par transfert de connaissances sur ResNet-50 initial de $0,77 \pm 0,04$ à trois classes et $0,82 \pm 0,02$ à deux classes : par réglage fin ces résultats descendent à des valeurs de $0,78 \pm 0,05$ à trois classes et $0,56 \pm 0,05$ à deux classes ; par programme d'apprentissage ces valeurs ne sont plus que de $0,51 \pm 0,07$ à trois classes et $0,67 \pm 0,05$. Un fort sur-apprentissage de ce réseau est supposé malgré le mécanisme d'augmentation de données mis en place. En effet, les diverses phases d'apprentissage réalisées ont abouti à un score F_1 compris entre 0,90 et 0,97 sur les données d'entraînement, démontrant

la présence d'un biais entre les données d'entraînement et l'évaluation sur les données de test. Malgré ces résultats, la possibilité d'obtenir des cartes des zones d'activation indépendamment du choix d'une taille ou d'un chevauchement en un seul passage dans le réseau de l'image, contrairement à la fenêtre glissante, sont encourageants pour élaborer des perspectives futures en ce sens. Néanmoins à ce stade, nombreuses sont les cartes d'activation incorrectes, dont l'origine supposée correspond à un apprentissage insuffisant de ce réseau.

PRISE DE DÉCISION AU NIVEAU LÉSIONNEL OU PATIENT

Les deux précédents chapitres sont consacrés aux diagnostics des images de **Microscopie Confocale par Réflectance** mises à la disposition de ce travail. Diverses solutions permettant la gestion de ces images comme des entités à part entière sont proposées, d'une part à l'aide de techniques propres à la globalité de l'image, et d'autre part à l'aide d'approches multi-échelle ou par instances multiples.

Dans ce nouveau chapitre, une méthodologie est mise en place permettant d'aider les experts lors de la prise de décision au niveau patient sur cette modalité complexe. Cette prise de décision au niveau patient est également nécessaire afin de mettre en place le processus multimodal, finalité de ce manuscrit. Pour cela, les précédentes expérimentations sont sollicitées afin d'extraire l'information des images, et diverses méthodes sont proposées pour permettre l'obtention d'une décision à partir d'un ensemble d'images d'un même patient.

Les résultats de ces expérimentations sont présentés en fin de chapitre, sur l'ensemble des lésions mais également sur les lésions malignes ne comportant que des types **Lentigo Maligna** et **Lentigo Maligna Melanoma**. Ces résultats sont également exprimés en termes de sensibilité et spécificité afin d'être confrontés à ceux des spécialistes sur cette même base de données.

Sommaire

7.1	Méthodologie	132
7.2	Approches supervisées	133
7.3	Approches faiblement supervisées	135
7.4	Présentation des résultats	137
	7.4.1 Protocole d'expérimentation	137
	7.4.2 Résultats des expérimentations	138
7.5	Discussion	144

7.1/ MÉTHODOLOGIE

Lors des deux précédents chapitres, la classification des images de **MCR** a été la principale tâche de ce manuscrit en proposant des traitements permettant leur gestion comme des entités à part entière. Lors du Chapitre 5, ces images ont été abordées dans leur intégralité à l'aide des méthodes d'extraction de caractéristiques sur la totalité de l'image, mais également lors du Chapitre 6 à l'aide de méthodes multi-échelle et niveaux de compréhension.

Néanmoins, ces méthodes dans leur état actuel ne permettent que le diagnostic au niveau de l'image, dont le nombre varie fortement d'une lésion à l'autre soit d'une dizaine à plusieurs centaines d'images pour un même patient. La tâche reste donc en l'état complexe et difficile à trancher par un spécialiste et ne permet pas une aide réaliste à la prise de décisions. De plus, la finalité de ce travail est en l'état impossible, cette information étant inexploitable dans le cadre d'un processus à multiples modalités.

Ce chapitre se dédie à la problématique de la classification des lésions, assimilable à une classification du patient lorsque celui-ci n'est associé qu'à une unique lésion, dont l'objectif est schématisé sur la Figure 7.1. À cette fin, les conclusions formulées lors des précédents chapitres propres aux images et à leur compréhension sont employées pour résoudre cette nouvelle problématique. Ainsi, comme lors du précédent chapitre sont considérées les méthodes d'extraction suivantes :

- pour la **catégorie de méthodes spatiales**, les caractéristiques définies par Haralick et al. [80],
- pour la **catégorie de méthodes fréquentielles**, les caractéristiques sur base d'extraction en ondelettes,
- pour la **catégorie de méthodes de transfert de connaissances**, l'architecture ResNet-50 associée à une couche de *Global Pooling - Moyenne*.

Ce chapitre propose diverses méthodes dans le but d'extraire une décision au niveau patient. Dans cette optique, la première partie de ce travail se consacre à la mise en place d'une classification patient par des mécanismes de prédiction supervisée. Puis, dans la seconde partie de ce travail sont élaborées diverses pistes de prédiction au niveau des lésions employant des principes qualifiés de faiblement supervisés par l'utilisation du **Multi-Instance Learning, MIL** (*alt. : Apprentissage par Instances Multiples, AIM*). Les dernières sections sont consacrées à l'analyse et une discussion des résultats de ces expériences.

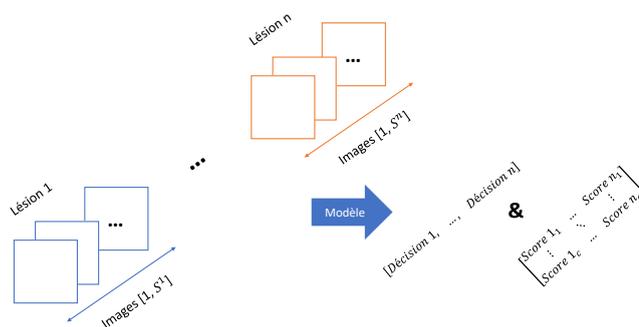


Figure 7.1 : Schéma récapitulatif de la problématique de l'inconsistance du nombre d'images constituant les lésions et de l'obtention de décisions et scores pour chacune de ces lésions.

7.2/ APPROCHES SUPERVISÉES

Cette section dédiée aux approches supervisées se consacre à des schémas d'apprentissages supervisés, c'est-à-dire dont la relation entre les instances et annotations est directe. Les données de **MCR** exploitées dans ce travail possèdent : (i) d'une part des annotations à bas niveau ou **niveau image**, (ii) d'autre part des annotations à haut niveau ou **niveau lésionnel**. L'une des contraintes majeures associées à ces données est l'inconsistance du nombre d'images entre les lésions composant le jeu de données. Ainsi, une lésion peut comporter un nombre d'images compris entre une dizaine et une centaine. Il est donc nécessaire de gérer cette variabilité au sein du processus de classification mis en place.

Afin de résoudre ce problème de manière supervisée, cette section propose un mécanisme de prédiction en deux temps :

- une première étape de prédiction **au niveau image**, dont le but consiste à émettre une étiquette pour chaque image,
- une seconde étape de prédiction **au niveau lésionnel** dont la finalité est d'évaluer la lésion à partir de l'information résultant du processus de classification au niveau image.

À cette fin, les données d'entraînement sont divisées en deux parties : (i) l'une à destination du processus de prédiction image, (ii) et l'autre à destination du processus décisionnel des lésions. Cette séparation des données d'entraînement en deux sous-ensembles permet de contrer un surapprentissage du modèle de prédiction au niveau lésionnel, en évitant une prise de décision sur des images dont les attentes sont déjà connues. Le schéma global mis en oeuvre dans cette section est disponible sur la Figure 7.2.

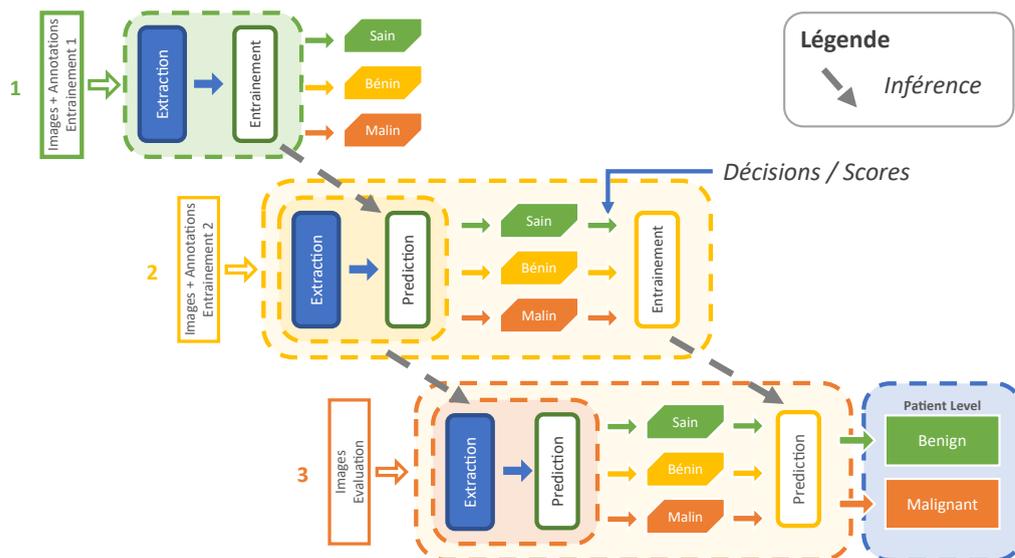


Figure 7.2 : Schéma de représentation du système complet de décision supervisée au niveau lésionnel. Le processus d'entraînement se subdivise en deux étapes : une première étape d'entraînement du modèle de prédiction au niveau des instances images et une seconde étape permettant l'agrégation de l'information au niveau des lésions. La prédiction est ensuite réalisée sur les données de test par inférence des paramètres précédemment entraînés.

Ainsi, la **prédiction au niveau image** de ce processus exploite les méthodes issues des Chapitre 5 et Chapitre 6, dont les expérimentations reposent sur la mise en place d'un processus de détection des images. Cette prédiction est assurée dans un premier temps par les méthodes d'extraction de caractéristiques jugées les plus pertinentes de ces parties, rappelées en Section 7.1. Ainsi, sont utilisées **des méthodes manuelles spatiales et fréquentielles** dont le nombre de caractéristiques extraites est de l'ordre de la dizaine, et **par transfert de connaissances sur des CNN** pré-entraînés dont le nombre de caractéristiques dépasse la centaine. Dans un second temps, ces caractéristiques sont mises à l'échelle par **une normalisation de type Minimum / Maximum** puis séparées à l'aide d'**un modèle SVM à noyau linéaire**. Les hyperparamètres de ce dernier sont rappelés au niveau de la Table 7.1.

Modèle	Hyperparamètres	Valeurs
SVM - Noyau linéaire	C	[0,001, 0,01, 0,1, 1, 10, 100, 1000]

Table 7.1 : Table reprenant le modèle et les hyperparamètres du modèle de classification employé pour la détection bas niveau des images.

La **prédiction au niveau des lésions** de ce processus tente d'établir une relation entre les diagnostics des images réalisés à l'aide de la première partie de ce processus et les annotations des lésions. Le nombre d'instances variant entre chaque lésion, les méthodes présentées ci-dessous sont mises au point pour supporter cette contrainte de fonctionnement. Ainsi, deux types potentiels de données d'entrée sont envisagés dans cette seconde partie de la classification, donnant lieu à deux types d'approches dont :

- une approche au niveau des **décisions** de classification de la première étape, défini par l'ensemble **N**. Ces données sont de la forme $instances \times 1$, dans lequel $instances$ est une variable propre à chaque lésion représentant le nombre d'images. Afin de réduire ces données à un vecteur commun, chaque classe est réduite par la relation $ratio_c = \frac{classe_c}{instances}$, aboutissant à un vecteur réel de dimension $1 \times classes$. À partir de ce vecteur, deux méthodes ont été définies, dont :
 - une méthode dite de **priorité** par laquelle est privilégiée la détection d'une classe hiérarchiquement supérieure comme décision globale à la lésion. Cette méthode est considérée sous le terme de **Décisions - Priorité**.
 - une méthode dite par seuils **dynamiques** par laquelle des seuils optimaux sont déterminés pour chaque classe en privilégiant les classes prioritaires. Cette méthode est considérée sous le terme de **Décisions - Dynamique**.
- une approche au niveau des **scores** de classification de la première étape, défini par l'ensemble **R**. Cette approche aboutit à un vecteur de taille $instances \times classes$, dans lequel $instances$ est une variable propre à chaque lésion représentant le nombre d'images et $classes$ est le nombre de classes prédites soit 2. Afin de réduire là encore ces données à un vecteur commun, chaque classe est réduite soit par une fonction de **moyenne** soit de **maximum** aboutissant à un vecteur réel de dimension $1 \times classes$. Puis, par seuil **dynamique** sur chacune des classes, des seuils optimaux sont définis en privilégiant les classes prioritaires. La combinaison de réduction par moyenne et seuil dynamique est considérée sous le terme de **Scores - Moyenne** et la combinaison de réduction par maximum et seuil dynamique est considérée sous le terme de **Scores - Maximum**.

7.3/ APPROCHES FAIBLEMENT SUPERVISÉES

Une approche est décrite comme étant faiblement supervisée lorsque l'apprentissage de la relation entre les données et les annotations n'est pas directe. Ainsi, ces méthodes faiblement supervisées sont rassemblées en trois catégories majeures [225], dont : (i) la **supervision dite incomplète** lorsque les données d'entraînement sont partiellement annotées, (ii) la **supervision dite inexacte** lorsque les annotations sont insuffisantes vis-à-vis du niveau de détails souhaité, (iii) et la **supervision dite imprécise** lorsque les annotations proviennent de sources manquant d'exactitude. Ces diverses approches ont notamment servi à la résolution de tâches telles que la catégorisation de textes [6, 169] ou d'images [36, 188]. Plus spécifiquement au domaine médical, ces techniques ont permis l'aide au diagnostic d'images médicales sur des pathologies pulmonaires ou du colon [58] mais également du cancer du sein [183].

À cette fin, les annotations et les données images propres à chacune des lésions sont employées pour servir cette tâche faiblement supervisée. Par la nature des données de ce travail, cette nouvelle section s'intéresse au second mode de supervision dite inexacte, par l'utilisation du **Multi-Instance Learning, MIL** (*alt. : Apprentissage par Instances Multiples, AIM*). De manière plus formelles, ces techniques visent à résoudre la relation $f : X \mapsto Y$, à l'aide des données $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_b, y_b)\}$ avec $X = \{x_1, x_2, \dots, x_b\}$ représentant un ensemble ou *bag* d'information composé d'images et y l'annotation associée [67].

Ce principe de méthodes à instances multiples est initialement issu du domaine de la recherche de médicaments, domaine dans lequel sont recherchées des molécules ainsi que le ou les *conformères* (comprendre les formes que peuvent adopter ces molécules) responsables d'une activité biologique [52]. Ainsi dans ce contexte, l'activité biologique observée (**annotation**) est associée à une molécule (**ensemble**) dont certains conformères (**instances**) sont responsables du phénomène observé. Ce faisant, la théorie de base des méthodes à instances multiples stipule qu'un ensemble dont l'annotation est négative ne possède aucune instance relative au phénomène observé, tandis qu'un ensemble dont l'annotation est positive contient au moins une instance relative au phénomène observé [52].

Par cette idée, de nombreuses théories ont émergé étendant le principe de techniques supervisées (principes de régression logistique, d'arbre de décision, ...) à la théorie des instances multiples [132, 216, 18]. Devant ces nombreuses théories et suite aux travaux des Chapitre 5 et Chapitre 6 mettant en avant l'efficacité des **SVM** sur cette problématique d'images de **MCR**, ce manuscrit s'intéresse à leur extension au principe d'instances multiples. L'un de ces travaux [55] développe plusieurs de ces approches et met notamment en avant trois propriétés essentielles des **SVM** : (i) la solidité, (ii) l'exhaustivité, (iii) et la convexité. Ces propriétés sont toutes intrinsèques aux **SVM** et ne peuvent être satisfaites par cette extension au principe d'instances multiples. Il est donc nécessaire de procéder à un compromis de ces propriétés [55].

Le travail de ce chapitre n'approfondit pas ces diverses propriétés, en revanche les divers modes d'approches associés à l'implémentation de ce principe sont envisagés, dont :

- des approches par **instances** dont la méthode consiste à déterminer un hyperplan directement à partir des instances, sans considérer les annotations au niveau des ensembles [6]. La séparation des ensembles découle de l'application du principe de base de la théorie des instances multiples, dans laquelle un ensemble est considéré positif si au moins l'une des instances qui le compose est considérée positive.

- des approches par **ensembles** dont le principe repose sur les annotations des ensembles et consiste à opérer la séparation par hyperplan au niveau des ensembles à partir de noyaux d'ensembles [72].
- des approches dites **hybrides** dont le principe se situe à mi-chemin entre les deux précédentes méthodes [21].

Ainsi, ces divers modes d'approches sont considérés dans les prochains paragraphes de ce travail. Dans le but de ne pas évaluer l'ensemble de ces modèles, ce travail ne se focalise que sur ceux jugés de manière empirique comme les plus pertinents au niveau de la classification d'ensembles par l'étude de Doran et al. [55]. Dans un premier temps, le modèle **Multi-Instance Support Vector Machine, MI-SVM** (*alt. : Machine à Vecteur de Support par Instances Multiples, MVS-IM*) évalué dans ce travail suit une approche par *instances*, dont le principe réside sur l'estimation de la meilleure instance représentative de chaque ensemble puis procède à une classification par **SVM** [6]. Dans un second temps, le modèle **Normalized Set Kernel, NSK** (*alt. : Noyau d'Ensemble Normalisé, NEN*) employé suit une approche par **ensembles**, dont le principe consiste à projeter chaque ensemble et ses instances dans un nouvel espace de caractéristiques sur lequel est réalisé la séparation par **SVM** [72]. Enfin, le modèle **Sparse Balanced Multi-Instance Learning, (sbMIL)** choisi suit un type d'approche dit **hybride**, dont le principe stipule que **les instances d'ensembles négatifs sont négatives et les instances d'ensembles positifs sont négatives ou positives** dont le balancement est géré par une variable inconnue et pour lequel la totalité des instances est projetée avant séparation par un modèle **SVM** [21].

Dans le but d'évaluer la pertinence de ces diverses méthodes, ce chapitre se consacre à l'évaluation de chacune d'entre elles combinée aux représentants des trois catégories d'extraction de caractéristiques formulées lors du Chapitre 5 et rappelées lors de la Section 7.1. De plus, suite aux résultats démontrant l'importance de la normalisation lors du Chapitre 5, l'impact de la normalisation par Minimum et Maximum est évalué en combinaison de ces modèles faiblement supervisés.

L'ensemble des modèles de prédiction exploitant le principe du **MIL** mis en oeuvre dans ce chapitre sont listés en Table 7.2. Afin de mener à bien la réalisation de ces expériences, ce travail se repose essentiellement sur la librairie logicielle *MISVM* prévue à cet effet [55].

Méthode	Paramètre	Valeurs
MI-SVM		
NSK	C	[0,01, 0,1, 1, 10, 100, 1000]
sbMIL		

Table 7.2 : Table reprenant les divers modèles utilisant le principe du **MIL** pour réaliser une classification au niveau lésionnel, ainsi que leurs hyperparamètres associés.

7.4/ PRÉSENTATION DES RÉSULTATS

Les précédentes sections de ce travail sont consacrées à des méthodes aptes à résoudre la problématique de prédiction au niveau des lésions à partir de données de **MCR**. Cette nouvelle section se dédie à la présentation des résultats pour donner suite aux diverses expérimentations mentionnées précédemment. Dans une première sous-section, le protocole d'expérimentation suivi pour permettre l'obtention des résultats est relaté et l'organisation de ces résultats est décrite. Puis dans une seconde sous-section, les divers résultats issus des expériences mentionnées précédemment sont développés.

7.4.1/ PROTOCOLE D'EXPÉRIMENTATION

Afin de réaliser l'évaluation de ces expériences, une **validation croisée imbriquée** est employée, déjà discutée lors de la Section 3.4, pour ses avantages permettant ainsi l'obtention de métriques plus objectives sur les performances du système évalué [26]. Ce type de structure de validation est plus robuste face au biais [26] et est démontré empiriquement moins sensible au biais et à la variance lorsque celle-ci est couplée à des valeurs de k comprises entre 5 et 10 pour l'évaluation [92]. La diversité de combinaisons relativement limitée de cette partie, permet à l'instar des précédents chapitres, d'obtenir des résultats dans un temps raisonnable malgré des valeurs de k élevées. Ainsi **une valeur de k égale à 10** est utilisée pour l'évaluation, tandis qu'une valeur de 2 est conservée pour la boucle interne suffisante pour déterminer la meilleure combinaison d'hyperparamètres. Bien que considéré comme une possibilité, le schéma de validation de type *leave-one-out* pour la boucle externe d'évaluation n'est pas utilisé car plus sujet à une variance élevée lors des mesures notamment en présence de valeurs aberrantes [12]. De plus, un tel schéma de validation est bien plus coûteux à entreprendre puisqu'il nécessite autant de phases d'entraînement et d'évaluation, que d'instances dans le jeu de données considéré, soit 224 pour ce travail.

D'une part, pour procéder au mieux au réglage des paramètres des modèles précédents et, d'autre part, pour retenir une évaluation objective de ces processus malgré les déséquilibres d'annotations, il est nécessaire d'opter pour une métrique adaptée. Ces expériences emploient le **score F_1** permettant de représenter conjointement les mesures de précision et de rappel au sein d'une même métrique. De plus, et afin de compléter l'analyse des résultats et de permettre un parallèle avec ceux des experts en dermatologie, les valeurs de **sensibilité** et de **spécificité** sont renseignées conjointement au score F_1 . Comme pour l'article de Cinotti et al. [40], ce processus décompose la présentation des résultats en deux temps. Dans un premier temps, est présentée l'analyse des mesures précédentes **sur l'ensemble des cas cliniques** du jeu de données, en adoptant comme classe positive les annotations des lésions dites *malignes*. Puis dans un second temps, les mêmes mesures sont extraites **en ne considérant que les cas de LM et LMM parmi les annotations malignes** et en adoptant comme classe positive les annotations des lésions dites *malignes*.

La partie dédiée aux résultats se décompose en diverses étapes, débutant dans un premier temps par une présentation des résultats obtenus **au niveau lésionnel par les diverses méthodes supervisées** présentée lors de la Section 7.2. Dans un second temps, ce travail procède à une présentation des résultats obtenus **au niveau lésionnel par l'utilisation des méthodes faiblement supervisées** listées lors de la Section 7.3. En continuité de ces résultats sur les méthodes faiblement supervisées, ce travail procède à une analyse inversée par la méthode **MI-SVM**. Ainsi, ce travail évalue la qualité des prédictions au niveau des instances images, afin de déterminer la compréhension du phénomène à bas niveau par cette méthode.

7.4.2/ RÉSULTATS DES EXPÉRIMENTATIONS

Dans cette section dédiée aux résultats, les diverses méthodes envisagées pour la prise de décision au niveau des lésions vont être traitées à l'aide de paragraphes respectifs. Cette première analyse à destination des méthodes supervisées se consacre aux résultats obtenus par ces diverses méthodes. Pour les méthodes spatiales, un score F_1 de 0,80 associé à un écart-type de 0,11 est obtenu à l'aide de la méthode **Décisions - Priorité** ; Pour les méthodes fréquentielles, un score F_1 de 0,83 associé à un écart-type de 0,07 est obtenu à l'aide de la méthode **Décisions - Dynamique** ; Enfin, pour les méthodes de transfert de connaissances, un score F_1 de 0,89 associé à un écart-type de 0,08 est obtenu à l'aide de la méthode **Scores - Moyenne**. Ainsi, des différences significatives sont à remarquer entre les méthodes spatiales, fréquentielles et de transfert de connaissances, nettement à l'avantage de cette dernière méthode comme mise en avant lors du Chapitre 5. L'utilisation du transfert de connaissances comme base d'extraction de caractéristiques et la méthode **Scores - Moyenne** permettent l'obtention d'une valeur de *sensibilité* de 0,95 et de *spécificité* de 0,70. Ces résultats peuvent être visualisés dans la Table 7.3.

		Sensibilité	Spécificité	Score F_1
Spatial	Décisions - Priorité	0,92 ± 0,09	0,40 ± 0,16	0,80 ± 0,11
	Décisions - Dynamique	0,80 ± 0,19	0,66 ± 0,22	0,80 ± 0,13
	Scores - Maximum	0,83 ± 0,18	0,57 ± 0,25	0,79 ± 0,12
	Scores - Moyenne	0,83 ± 0,20	0,58 ± 0,30	0,79 ± 0,13
Fréquentiel	Décisions - Priorité	0,93 ± 0,06	0,37 ± 0,25	0,80 ± 0,11
	Décisions - Dynamique	0,82 ± 0,08	0,77 ± 0,10	0,83 ± 0,07
	Scores - Maximum	0,84 ± 0,09	0,58 ± 0,22	0,80 ± 0,09
	Scores - Moyenne	0,80 ± 0,14	0,72 ± 0,25	0,81 ± 0,09
Transfert	Décisions - Priorité	0,98 ± 0,04	0,34 ± 0,14	0,82 ± 0,06
	Décisions - Dynamique	0,91 ± 0,13	0,71 ± 0,24	0,87 ± 0,09
	Scores - Maximum	0,92 ± 0,15	0,57 ± 0,21	0,84 ± 0,12
	Scores - Moyenne	0,95 ± 0,09	0,70 ± 0,22	0,89 ± 0,08

Table 7.3 : Table de résultats issus des processus de classification supervisée employés pour mener une classification au niveau lésionnel, exprimés selon la sensibilité, la spécificité et le score F_1 .

Ce second paragraphe se focalise sur les cas de lésions de types **LM** et **LMM** par ces mêmes méthodes de prédiction supervisées. Pour les méthodes spatiales, un score F_1 de 0,80 associé à un écart-type de 0,14 est obtenu à l'aide de la méthode **Décisions - Dynamique** ; Pour les méthodes fréquentielles, un score F_1 de 0,83 associé à un écart-type de 0,09 est obtenu à l'aide de la méthode **Décisions - Dynamique** ; Pour les méthodes de transfert de connaissances, un score F_1 de 0,87 associé à un écart-type de 0,11 est obtenu à l'aide de la méthode **Scores - Moyenne**. À nouveau, des différences significatives sont à remarquer entre les méthodes spatiales, fréquentielles et de transfert de connaissances, à l'avantage de cette dernière méthode comme mise en avant lors du Chapitre 5. La mise en oeuvre du transfert de connaissances comme base d'extraction de caractéristiques et la méthode **Scores - Moyenne** permettent l'obtention de résultats équivalents avec une valeur de *sensibilité* de 0,95 et de *spécificité* de 0,70 sur les lésions de types **LM** et **LMM**. Les diverses valeurs peuvent être visualisées dans la Table 7.4.

		Sensibilité	Spécificité	Score F_1
Spatial	Décisions - Priorité	0,93 ± 0,11	0,40 ± 0,16	0,78 ± 0,13
	Décisions - Dynamique	0,84 ± 0,18	0,66 ± 0,22	0,80 ± 0,14
	Scores - Maximum	0,83 ± 0,18	0,57 ± 0,25	0,77 ± 0,14
	Scores - Moyenne	0,86 ± 0,18	0,58 ± 0,30	0,79 ± 0,14
Fréquentiel	Décisions - Priorité	0,93 ± 0,07	0,37 ± 0,25	0,77 ± 0,13
	Décisions - Dynamique	0,84 ± 0,12	0,77 ± 0,10	0,83 ± 0,09
	Scores - Maximum	0,87 ± 0,09	0,58 ± 0,22	0,80 ± 0,11
	Scores - Moyenne	0,82 ± 0,14	0,72 ± 0,25	0,81 ± 0,12
Transfert	Décisions - Priorité	0,97 ± 0,05	0,34 ± 0,14	0,79 ± 0,07
	Décisions - Dynamique	0,91 ± 0,16	0,71 ± 0,24	0,86 ± 0,12
	Scores - Maximum	0,90 ± 0,18	0,57 ± 0,21	0,81 ± 0,16
	Scores - Moyenne	0,95 ± 0,11	0,70 ± 0,22	0,87 ± 0,11

Table 7.4 : Table de résultats issus des processus de classification supervisée employés pour mener une classification au niveau lésionnel sur les pathologies de **LM** et **LMM**, exprimés selon la sensibilité, la spécificité et le score F_1 .

		Sensibilité	Spécificité	Score F_1
Spatial	MI-SVM	0,69 ± 0,13	0,57 ± 0,22	0,70 ± 0,12
	NSK	0,81 ± 0,12	0,70 ± 0,12	0,81 ± 0,08
	sbMIL	0,10 ± 0,08	0,99 ± 0,06	0,18 ± 0,13
	MMS + MI-SVM	0,78 ± 0,14	0,77 ± 0,14	0,81 ± 0,12
	MMS + NSK	0,78 ± 0,11	0,69 ± 0,13	0,79 ± 0,08
	MMS + sbMIL	0,88 ± 0,30	0,07 ± 0,30	0,72 ± 0,24
Fréquentiel	MI-SVM	0,72 ± 0,19	0,62 ± 0,26	0,73 ± 0,16
	NSK	0,78 ± 0,15	0,59 ± 0,21	0,77 ± 0,12
	sbMIL	0,27 ± 0,28	0,84 ± 0,29	0,39 ± 0,23
	MMS + MI-SVM	0,83 ± 0,12	0,69 ± 0,16	0,82 ± 0,10
	MMS + NSK	0,80 ± 0,14	0,71 ± 0,09	0,81 ± 0,11
	MMS + sbMIL	0,09 ± 0,06	1,00 ± 0,00	0,16 ± 0,09
Transfert	MI-SVM	0,83 ± 0,12	0,76 ± 0,17	0,84 ± 0,10
	NSK	0,86 ± 0,10	0,77 ± 0,19	0,86 ± 0,11
	sbMIL	0,17 ± 0,11	0,90 ± 0,09	0,28 ± 0,15
	MMS + MI-SVM	0,83 ± 0,10	0,77 ± 0,16	0,84 ± 0,07
	MMS + NSK	0,86 ± 0,10	0,72 ± 0,19	0,85 ± 0,09
	MMS + sbMIL	0,21 ± 0,12	0,91 ± 0,10	0,33 ± 0,17

Table 7.5 : Table de résultats issus des processus de classification faiblement supervisée par **MIL** employés pour mener une classification au niveau lésionnel, exprimés selon la sensibilité, la spécificité et le score F_1 .

Cette seconde partie à destination des méthodes faiblement supervisées par **MIL** va permettre de décrire les résultats obtenus par ces diverses méthodes. Pour les méthodes spatiales, un score F_1 de 0,81 associé à un écart-type de 0,08 est obtenu à l'aide de la méthode **NSK** ; Pour les méthodes fréquentielles, un score F_1 de 0,82 associé à un écart-type de 0,10 est obtenu à l'aide de la méthode combinant **normalisation par Minimum / Maximum et MI-SVM** ; Pour les méthodes de transfert de connaissances, un score F_1 de 0,86 associé à un écart-type de 0,11 est obtenu à l'aide de la méthode **NSK**. Ainsi, de même que pour les méthodes supervisées, des différences significatives sont à remarquer entre les méthodes spatiales, fréquentielles et de transfert de connaissances. Cette dernière méthode comme base d'extraction de caractéristiques et la méthode **NSK** permettent l'obtention d'une valeur de *sensibilité* de 0,86 et de *spécificité* de 0,77. Les valeurs détaillées des résultats d'expérimentation de techniques faiblement supervisées par **MIL** peuvent être visualisées dans la Table 7.5.

Ce second paragraphe à l'intention des méthodes faiblement supervisées par **MIL** se focalise sur les cas de **LM** et **LMM**. Pour les méthodes spatiales, un score F_1 de 0,80 associé à un écart-type de 0,11 est obtenu à l'aide de la méthode **NSK** ; Pour les méthodes fréquentielles, un score F_1 de 0,80 associé à un écart-type de 0,11 est obtenu à l'aide de la méthode combinant **normalisation par Minimum / Maximum au modèle NSK** ; Pour les méthodes de transfert de connaissances, un score F_1 de 0,85 associé à un écart-type de 0,13 est obtenu par la méthode **NSK**. À nouveau, la méthode d'extraction par transfert de connaissances permet l'obtention des résultats les plus pertinents en combinaison de la méthode **NSK**, soit des valeurs de *sensibilité* de 0,87 et de *spécificité* de 0,77. Ces nouvelles valeurs liées à l'évaluation des méthodes faiblement supervisées sur les cas de **LM** et **LMM** peuvent être visualisées dans la Table 7.6.

		Sensibilité	Spécificité	Score F_1
Spatial	MI-SVM	0,69 ± 0,17	0,57 ± 0,22	0,69 ± 0,14
	NSK	0,82 ± 0,15	0,70 ± 0,12	0,80 ± 0,11
	sbMIL	0,11 ± 0,09	0,99 ± 0,06	0,20 ± 0,14
	MMS + MI-SVM	0,78 ± 0,17	0,77 ± 0,14	0,80 ± 0,16
	MMS + NSK	0,81 ± 0,14	0,69 ± 0,13	0,79 ± 0,09
	MMS + sbMIL	0,86 ± 0,30	0,07 ± 0,30	0,67 ± 0,23
Fréquentiel	MI-SVM	0,71 ± 0,21	0,62 ± 0,26	0,71 ± 0,19
	NSK	0,78 ± 0,12	0,59 ± 0,21	0,75 ± 0,10
	sbMIL	0,25 ± 0,28	0,84 ± 0,29	0,37 ± 0,22
	MMS + MI-SVM	0,82 ± 0,13	0,69 ± 0,16	0,80 ± 0,12
	MMS + NSK	0,82 ± 0,14	0,71 ± 0,09	0,80 ± 0,11
	MMS + sbMIL	0,11 ± 0,06	1,00 ± 0,00	0,19 ± 0,10
Transfert	MI-SVM	0,82 ± 0,13	0,76 ± 0,17	0,82 ± 0,12
	NSK	0,87 ± 0,09	0,77 ± 0,19	0,85 ± 0,13
	sbMIL	0,19 ± 0,10	0,90 ± 0,09	0,30 ± 0,15
	MMS + MI-SVM	0,85 ± 0,10	0,77 ± 0,16	0,84 ± 0,08
	MMS + NSK	0,88 ± 0,10	0,72 ± 0,19	0,84 ± 0,11
	MMS + sbMIL	0,25 ± 0,13	0,91 ± 0,10	0,38 ± 0,18

Table 7.6 : Table de résultats issus des processus de classification faiblement supervisée par **MIL** employés pour mener une classification au niveau lésionnel sur les pathologies de **LM** et **LMM**, exprimés selon la sensibilité, la spécificité et le score F_1 .

Pour finir, ce paragraphe se consacre à la présentation des courbes **ROC** au niveau lésionnel des diverses méthodes mentionnées précédemment. Sur l'ensemble des cas cliniques, la méthode supervisée issue du modèle de prédiction **Scores - Moyenne** permet l'obtention d'une valeur d'**AUC** de 0,87 ; la méthode faiblement supervisée par utilisation du modèle **NSK** permet quant à elle l'obtention d'une valeur d'**AUC** de 0,83. Sur les seuls cas de **LM** et **LMM**, la méthode supervisée issue du modèle de prédiction **Scores - Moyenne** permet l'obtention d'une valeur d'**AUC** similaire de 0,88 ; la méthode faiblement supervisée par utilisation du modèle **NSK** permet quant à elle l'obtention d'une valeur d'**AUC** de 0,83 identique à celle constaté sur l'ensemble des lésions. Ces courbes **ROC** et ces valeurs d'**AUC** peuvent être visualisées sur la Figure 7.3.

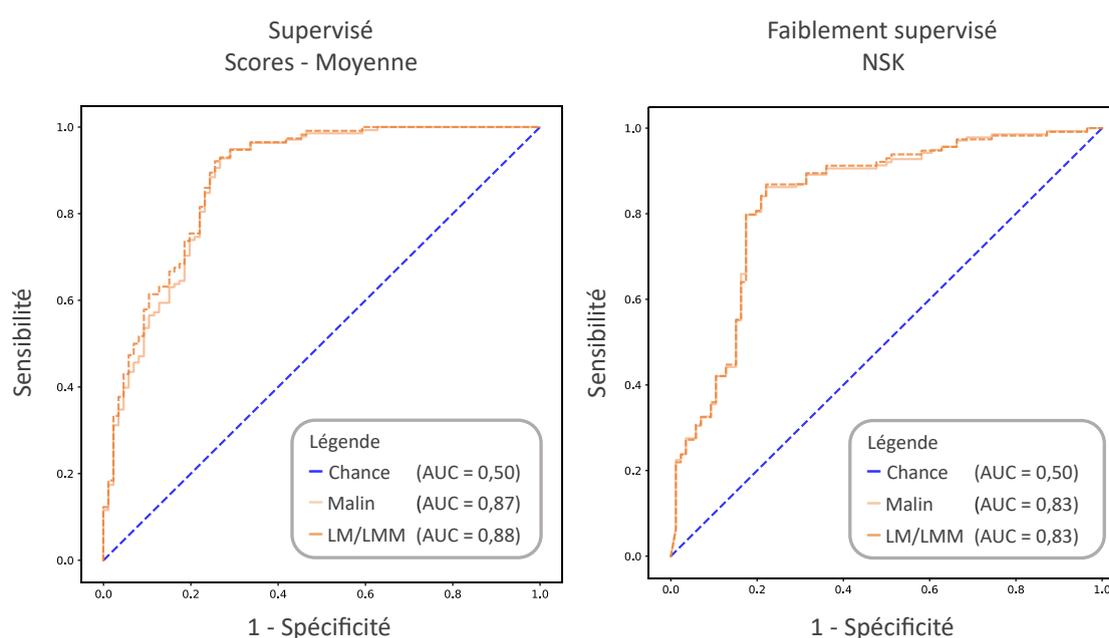


Figure 7.3 : Courbes **ROC** issues des résultats de la prédiction au niveau des lésions sur l'ensemble des cas, figurant à l'aide d'un trait orange estompé tracé en continu, mais également sur les cas de **LM** et **LMM**, figurant à l'aide d'un trait orange tracé en pointillés. À gauche, les courbes issues des prédictions employant la méthode *Scores - Moyenne* ; À droite, les courbes issues des prédictions du modèle **NSK**.

Cette dernière partie se dédie aux résultats de prédiction des méthodes faiblement supervisées par **MIL** dont l'approche se base sur les instances et permet de revenir à un niveau de prédiction sur les instances. À cet effet, seul le modèle **MI-SVM** permet de revenir à un tel niveau de prédiction au sein de ce travail. Pour les méthodes spatiales, un score F_1 de 0,56 associé à un écart-type de 0,14 est obtenu par combinaison de la **normalisation Minimum / Maximum et du modèle MI-SVM** ; Pour les méthodes fréquentielles, un score F_1 de 0,59 associé à un écart-type de 0,15 est obtenu par combinaison de la **normalisation Minimum / Maximum et du modèle MI-SVM** ; Pour les méthodes de transfert de connaissances, un score F_1 de 0,64 associé à un écart-type de 0,16 est obtenu avec le modèle **MI-SVM** seul. Les résultats de cette prédiction au niveau des instances images sont visibles dans la Table 7.7.

		Sensibilité	Spécificité	Score F_1
Spatial	MI-SVM	0,29 ± 0,13	0,90 ± 0,08	0,42 ± 0,14
	MMS + MI-SVM	0,42 ± 0,15	0,93 ± 0,05	0,56 ± 0,14
Fréquentiel	MI-SVM	0,29 ± 0,22	0,89 ± 0,28	0,42 ± 0,20
	MMS + MI-SVM	0,45 ± 0,15	0,93 ± 0,05	0,59 ± 0,15
Transfert	MI-SVM	0,48 ± 0,16	0,97 ± 0,02	0,64 ± 0,16
	MMS + MI-SVM	0,42 ± 0,14	0,97 ± 0,02	0,57 ± 0,14

Table 7.7 : Table de résultats issus de l'application du modèle **MI-SVM** employé pour mener une classification au niveau des instances, sur l'ensemble des images annotées, et exprimés selon la sensibilité, la spécificité et le score F_1 .

Ce second paragraphe reprend les résultats de prédiction des méthodes faiblement supervisées par **MIL** sur les images issues de lésions de type **LM** et **LMM**. Pour les méthodes spatiales, un score F_1 de 0,60 associé à un écart-type de 0,14 est obtenu par combinaison de la **normalisation Minimum / Maximum et du modèle MI-SVM** ; Pour les méthodes fréquentielles, un score F_1 de 0,62 associé à un écart-type de 0,14 est obtenu par combinaison de la **normalisation Minimum / Maximum et du modèle MI-SVM** ; Pour les méthodes de transfert de connaissances, un score F_1 de 0,65 associé à un écart-type de 0,16 est obtenu avec le modèle **MI-SVM** seul. Le détail de ces résultats de prédiction au niveau des instances images à partir des méthodes par **MIL** est disponible dans la Table 7.8.

		Sensibilité	Spécificité	Score F_1
Spatial	MI-SVM	0,31 ± 0,16	0,84 ± 0,19	0,46 ± 0,18
	MMS + MI-SVM	0,44 ± 0,14	0,89 ± 0,18	0,60 ± 0,14
Fréquentiel	MI-SVM	0,28 ± 0,23	0,87 ± 0,32	0,42 ± 0,24
	MMS + MI-SVM	0,47 ± 0,15	0,86 ± 0,25	0,62 ± 0,14
Transfert	MI-SVM	0,49 ± 0,16	0,94 ± 0,07	0,65 ± 0,16
	MMS + MI-SVM	0,43 ± 0,16	0,94 ± 0,08	0,59 ± 0,17

Table 7.8 : Table de résultats issus de l'application du modèle **MI-SVM** employés pour mener une classification au niveau des instances, sur les images annotées issues de pathologies de **LM** et **LMM**, et exprimés selon la sensibilité, la spécificité et le score F_1 .

Pour finir cette présentation des résultats de prédiction au niveau des images à partir de méthodes faiblement supervisées employant le principe du **MIL**, ce paragraphe présente les résultats des courbes **ROC** et des valeurs d'**AUC**. Ainsi sur l'ensemble des cas cliniques, **la méthode MI-SVM** permet l'obtention d'une valeur d'**AUC** de 0,87. Cette même méthode, focalisée sur les cas de **LM** et **LMM**, permet l'obtention d'une valeur d'**AUC** de 0,82. Les courbes **ROC** et les valeurs d'**AUC** obtenues à partir de la prédiction sur des instances images, peuvent être visualisées sur la Figure 7.4.

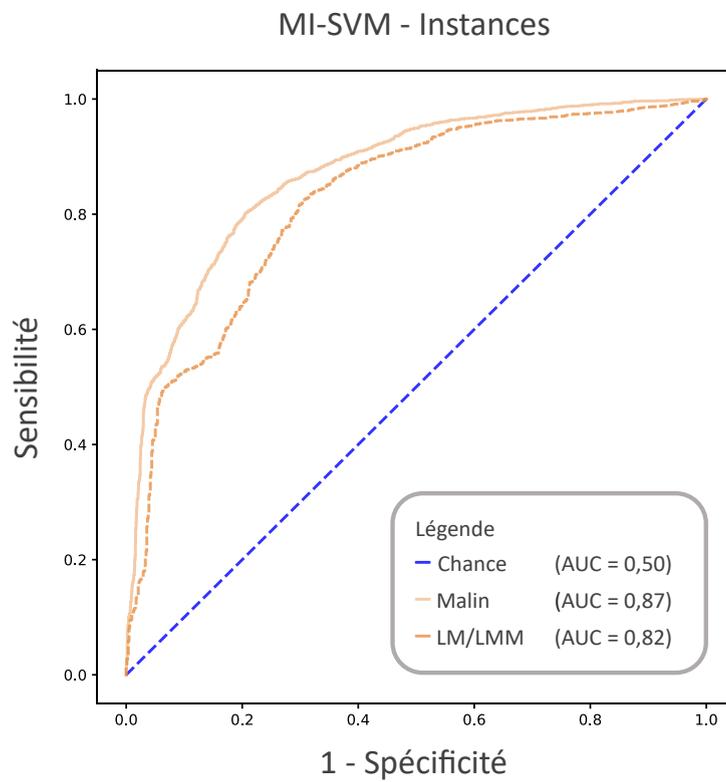


Figure 7.4 : Courbes **ROC** issues de la prédiction au niveau des instances sur l'ensemble des cas, figurant à l'aide d'un trait orange estompé tracé en continu, et en ne considérant que les cas de **LM** et **LMM** parmi les annotations malignes, figurant à l'aide d'un trait orange tracé en pointillés. Les courbes sont issues du mode de prédiction faiblement supervisé sur la base de la théorie du **MIL** par utilisation du modèle **MI-SVM**.

7.5/ DISCUSSION

Cette discussion se dédie à l'analyse des résultats de la précédente section. Ce premier paragraphe détaille les observations issues des méthodes de prédiction supervisées au niveau lésionnel. Les score F_1 sont pour la plupart homogènes si l'on considère chaque méthode d'extraction indépendamment, mais les mesures de sensibilité et spécificité propres à chacune varient fortement entre méthodes. Ainsi, la méthode de **Décisions - Priorité** bénéficie d'une forte sensibilité liée à son principe privilégiant la moindre détection de la classe prioritaire / positive, mais se contente d'une faible spécificité. Ce principe est ainsi plus nuancé par la méthode **Décisions - Dynamique** qui apporte un seuil dynamique d'activation, et dont résultent des valeurs de sensibilité et spécificité plus équilibrées et supérieures en terme de score F_1 . Néanmoins, cette méthode nécessite de laisser de côté certaines détections jusqu'à atteindre un certain seuil d'activation ; ce principe peut être difficile à utiliser en milieu clinique. Afin de modérer ces deux méthodes, ont été proposées deux approches par scores. La première de ces méthodes **Scores - Maximum** initialement conçue pour conserver la sensibilité de la méthode **Décisions - Priorité**, perd une forte partie de celle-ci mais améliore nettement sa spécificité. La seconde de ces méthodes **Scores - Moyenne** initialement conçue pour affiner la décision de la méthode **Décisions - Dynamique**, reste en deçà de celle-ci avec néanmoins un score important sur les caractéristiques issues de l'extraction par transfert de connaissances. En ce qui concerne les résultats sur les pathologies malignes toutes confondues et sur les cas de **LM** et **LMM**, peu de différences sont à remarquer entre ces deux situations, ces méthodes semblent ainsi robustes au type de pathologie traitée.

Ce second paragraphe propose une analyse des méthodes faiblement supervisées par **MIL** dont les résultats sont très hétérogènes, par contraste aux méthodes supervisées. Ainsi, ce paragraphe procède à une analyse méthode par méthode. Dans un premier temps, **le modèle MI-SVM** permet l'obtention de modèles plus sensibles que spécifiques et dont le score global semble assez dépendant du type d'extraction employé, les résultats étant maximisés par l'utilisation du transfert de connaissances. Néanmoins, l'utilisation d'une **normalisation par Minimum / Maximum** semble bénéfique dans la plupart des situations et contribue à baisser cette dépendance à la méthode d'extraction de caractéristiques. Dans un second temps, **le modèle NSK** est choisi pour représenter la catégorie des approches par noyaux d'ensembles, et semble être la méthode la plus adaptée dans ce schéma faiblement supervisé sur le jeu de données traité. Ainsi, cette méthode apporte des résultats assez proches quel que soit le mode d'extraction de caractéristiques choisi, et possède une bonne sensibilité bien que les valeurs de spécificité soient assez proches. Dans un dernier temps, **le modèle sbMIL** est utilisé comme représentant des approches par **MIL** de type hybride, mais dont les performances sont non concluantes dans ce cadre. En effet, ces méthodes ont une grande spécificité mais une sensibilité quasi-inexistante. Pour finir, les conclusions qui régissent les résultats sur les pathologies malignes toutes confondues et sur les cas de **LM** et **LMM** sont similaires à celle des méthodes supervisées ; ces méthodes semblent ainsi robustes au type de pathologie traitée.

Ainsi, les deux approches proposées, c'est à dire par entraînement supervisé et faiblement supervisée, sont adaptées à la problématique actuelle bien que l'approche supervisée soit sensiblement supérieure. Cette supériorité des méthodes supervisées se traduit notamment par des valeurs d'**AUC** de 0,87 pour **le modèle Scores - Moyenne** et de 0,83 pour **le modèle NSK**. Il est à remarquer que toutes ces approches possèdent un écart-type assez élevé situé entre 0,08 et 0,15, à l'exception de **la méthode sbMIL** dont les valeurs sont encore plus conséquentes. Cette tendance assez forte montre une certaine dépendance aux données mais peut être la conséquence du nombre restreint d'échantillons au niveau lésionnel du jeu de données, soit 224 lésions. Un élément supplémentaire à considérer est la complexité nécessaire à l'obtention des résultats supervisés, dont le nombre de paramètres est supérieur aux méthodes faiblement supervisées présentées. Par ailleurs,

les méthodes supervisées utilisées ici nécessitent un processus en deux temps (voir trois si l'on considère l'extraction de caractéristiques), tandis que les méthodes faiblement supervisées nécessitent une étape de moins. Le dernier argument en faveur des méthodes faiblement supervisées est celui de la très faible quantité d'annotations sollicitée pour parvenir à de tels résultats, soit 224 annotations.

Afin de mesurer, d'une part la compréhension au niveau lésionnel de ces modèles faiblement supervisés et d'autre part d'évaluer la qualité d'un système permettant de revenir à des prédictions bas niveau, une évaluation des prédictions au niveau des instances a été réalisée. Ces tests sont menés sur **le modèle MI-SVM**, et on permis d'obtenir une valeur de score F_1 de 0,64 et une valeur d'**AUC** de 0,87 sur l'ensemble des cas cliniques. En comparaison, ces scores étaient respectivement à une valeur de score F_1 de 0,82 et une valeur d'**AUC** de 0,90 au sein du Chapitre 5 et étaient accompagnées de valeurs d'écart-type beaucoup plus restreintes. Il existe de ce fait une différence notable entre ces méthodes, pouvant rendre la prédiction au niveau des instances difficile pour la méthode **MI-SVM**.

Pour finir cette discussion, ce paragraphe réalise un parallèle entre les résultats expérimentaux de ce chapitre et ceux des spécialistes en dermatologie. Pour rappel, l'analyse des résultats des 18 experts en **MCR** renvoie une valeur de sensibilité de $0,84 \pm 0,05$ associée à une valeur de spécificité de $0,75 \pm 0,06$ ainsi qu'une valeur d'**AUC** de 0,86 sur l'ensemble de données initial de l'étude de Cinotti et al. [40]. L'analyse des résultats de ces mêmes experts sur les lésions de type **LM** et **LMM** renvoie une valeur de sensibilité de $0,80 \pm 0,07$ associée à une spécificité de $0,81 \pm 0,05$ ainsi qu'une valeur d'**AUC** de 0,88 issues de ces mêmes données initiales. Sur cet ensemble de données *dont les annotations sont réactualisées*, la méthode supervisée **Scores - Moyenne** parvient à accomplir une sensibilité de $0,95 \pm 0,09$ et une spécificité de $0,70 \pm 0,22$ ainsi qu'une valeur d'**AUC** de 0,87 et **le modèle NSK** parvient à accomplir une sensibilité de $0,86 \pm 0,10$ et une spécificité de $0,77 \pm 0,19$ ainsi qu'une valeur d'**AUC** de 0,83. En ne prenant en compte que les lésions malignes de type **LM** et **LMM**, la méthode supervisée **Scores - Moyenne** parvient à accomplir une sensibilité de $0,95 \pm 0,11$ et une spécificité de $0,70 \pm 0,22$ ainsi qu'une valeur d'**AUC** de 0,88 et **le modèle NSK** parvient à accomplir une sensibilité de $0,87 \pm 0,09$ et une spécificité de $0,77 \pm 0,19$ ainsi qu'une valeur d'**AUC** de 0,83.

III

ANALYSE PAR PROCESSUS MULTIMODAL

PRÉAMBULE MULTIMODALITÉ

Lors de la précédente partie consacrée à la **Microscopie Confocale par Réflectance**, différentes méthodes sont proposées afin d'obtenir des prédictions au niveau des images et des lésions. Cette dernière partie va permettre de mettre au point et d'évaluer des stratégies avec l'objectif de mettre en place des processus multimodaux d'aide au diagnostic des lésions malignes mais surtout des lésions de type **LM** et **LMM**. À cette fin, ce préambule présente lors des prochains paragraphes les données employées à cet effet et reprend les points importants de l'étude clinique de référence.

Pour rappel, les données de photographie clinique et de dermatoscopie ont été mises de côté lors des expériences sur la modalité de **MCR** menées lors de la Partie **II**. Ainsi, cette nouvelle partie intègre l'ensemble des données de lésions obtenues dans le cadre de l'étude de Cinotti et al. [40], soit 224 lésions pour lesquelles sont mises à disposition les acquisitions issues des modalités de *photographie clinique*, de *dermatoscopie* et de **MCR**. Ces données sont focalisées majoritairement autour de lésions malignes de type **LM** et de **LMM** et lésions bénignes proches de ces deux pathologies, susceptibles d'induire le praticien en erreur.

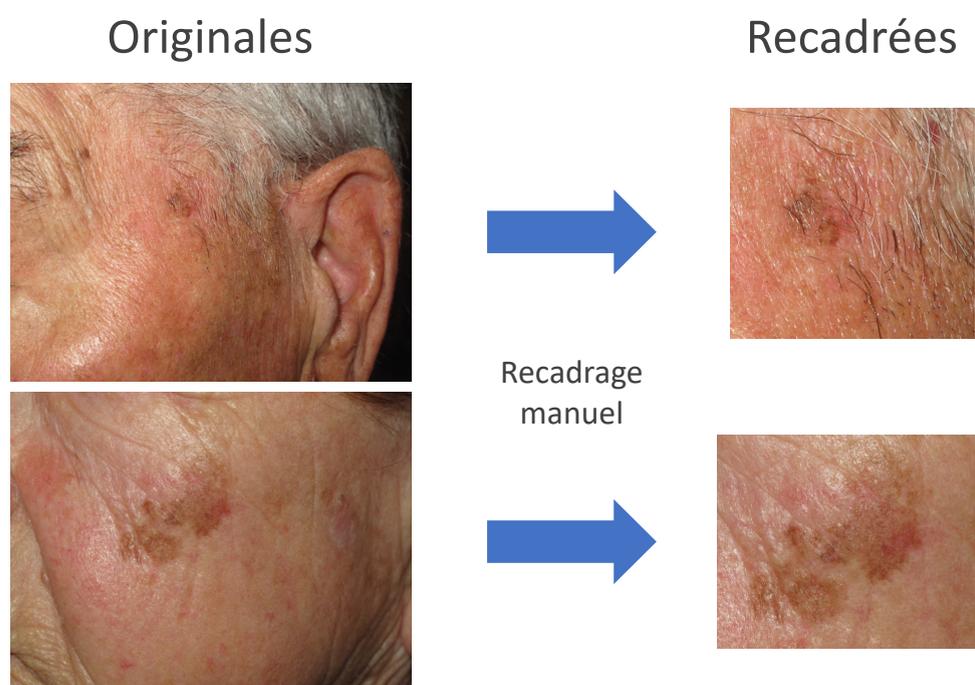


Figure iii.1 : Exemples de données de photographie clinique recadrées manuellement autour de la zone d'intérêt pour le besoin de ce manuscrit.

De manière plus détaillée, chaque lésion issue de ce jeu de données comporte :

- des données de **photographie clinique**, *une image par lésion*. Pour cinq de ces lésions, plusieurs images de photographie clinique étaient mises à disposition, **seule la plus pertinente d'entre elles a été conservée**. De plus, aucun protocole ne régit l'acquisition de ces images ce qui aboutit à une grande diversité, dont la lésion ne représente parfois qu'une faible portion de celle-ci. Afin de limiter cette diversité, **ces images sont recadrées manuellement** pour l'objet de ce travail. La résolution spatiale de ces images recadrées est comprise entre 250×250 pixels pour les plus petites d'entre elles à plus de 2000×2000 pixels pour les plus conséquentes.
- des données de **dermatoscopie**, *une image par lésion*. Ces images sont utilisées de manière brute, pour lesquelles aucun traitement n'a été jugé nécessaire. En effet, le principe de cette modalité permet d'obtenir des images assez homogènes malgré l'absence de protocole d'acquisition. La résolution spatiale de ces images est comprise entre 480×640 pixels pour les plus petites d'entre elles à 3312×4416 pixels pour les plus conséquentes. Cependant, environ 85 % de ces lésions possèdent une résolution spatiale de 3312×4416 pixels.
- des données de **MCR**, dont le nombre varie *de quelques images à plusieurs centaines par lésion*. Pour chaque lésion, ces images sont sélectionnées par deux des trois médecins investigateurs de l'étude [40]. L'acquisition de ces données provient essentiellement de la **JDE**, particulièrement adaptée à l'observation de ce phénomène. Ces données ont été plus longuement décrites lors de la Partie II.

L'évaluation de ces modalités d'images dans leur format initial, par les experts sollicités dans l'étude de Cinotti et al. [40], conduit aux divers résultats suivants :

- sur les données de **photographie clinique**, la participation de 15 experts permet d'obtenir une sensibilité de $0,72 \pm 0,08$ et une spécificité de $0,78 \pm 0,08$ sur l'ensemble des lésions malignes. Aucune information n'est mise à disposition sur les pathologies de **LM** et **LMM** sur cette modalité.
- sur les données de **dermatoscopie**, la participation de 15 experts permet d'obtenir une sensibilité de $0,69 \pm 0,10$ et une spécificité de $0,85 \pm 0,06$. Sur les pathologies de **LM** et **LMM**, ces scores s'écartent de manière significative avec une valeur de sensibilité de $0,61 \pm 0,15$ et de spécificité de $0,92 \pm 0,05$.
- sur les données de **MCR**, la participation 12 experts permet d'obtenir une sensibilité de $0,84 \pm 0,05$ et une spécificité de $0,75 \pm 0,06$. Sur les pathologies de **LM** et **LMM**, ces scores deviennent plus homogènes avec une valeur de sensibilité de $0,80 \pm 0,07$ et de spécificité de $0,81 \pm 0,05$.

En termes de courbes **ROC**, ces précédents résultats sont bien moins nuancés. Ces courbes rapportées au score **AUC**, les experts accomplissent sur la dermatoscopie un score de 0,85 contre 0,86 sur la **MCR** sur l'ensemble des lésions malignes. Sur les lésions de type **LM** et **LMM**, les experts accomplissent sur la dermatoscopie et sur la **MCR** un score de 0,88. Ces résultats de courbes **ROC** et de scores **AUC** apparaissent sur la Figure iii.2.

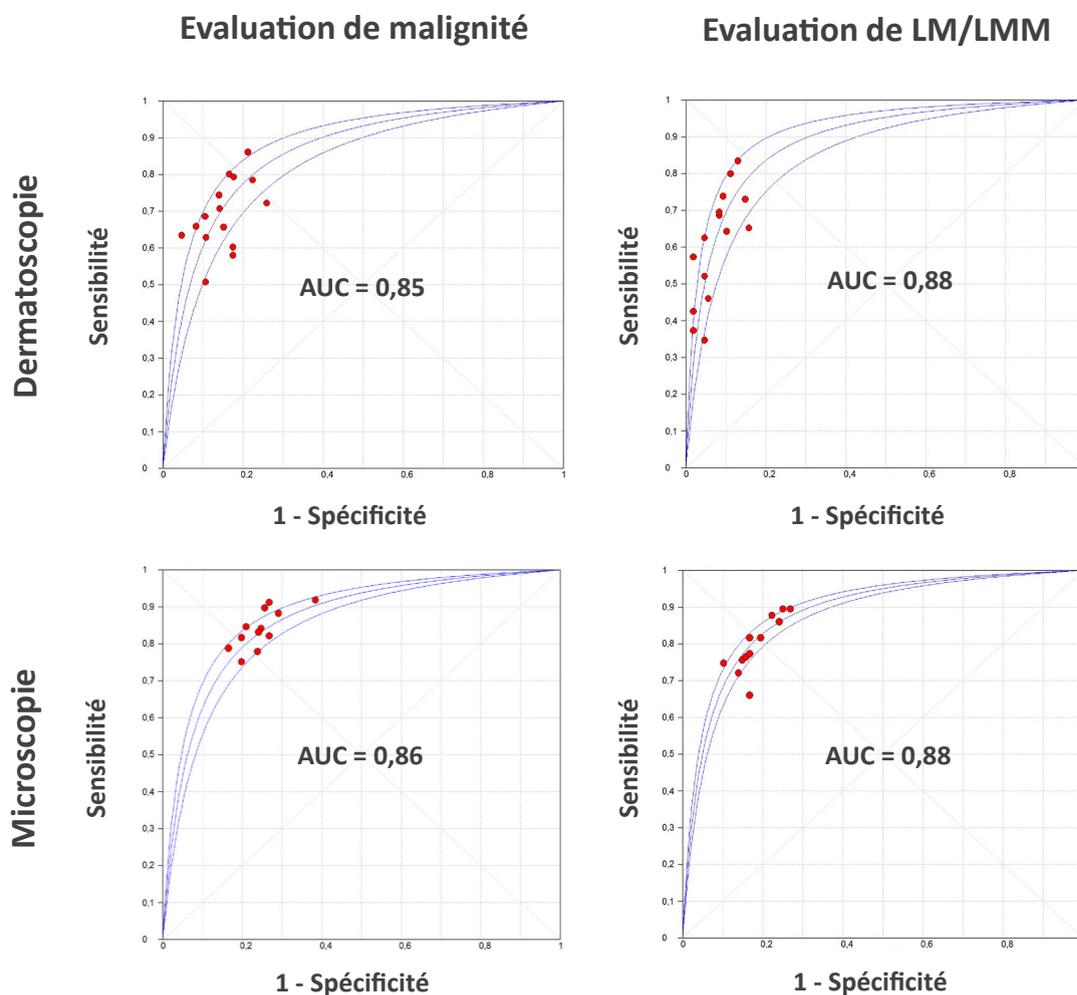


Figure iii.2 : Courbes ROC issues de l'évaluation des données de dermatoscopie et de MCR par les experts [40]. En haut, les résultats des 15 experts menés sur la modalité de dermatoscopie ; En bas, les résultats des 12 experts menés sur la modalité de MCR. À gauche, les résultats sur la détection d'éléments malins ; À droite, le résultat des experts mené sur la détection de LM et LMM.

En revanche, cette étude ne met pas à disposition d'éléments relatifs à une prise en charge séquentielle et ne donne pas lieu à un objectif particulier contrairement à la précédente partie. Néanmoins, l'objectif visé dans cette partie sans quantification exacte est d'optimiser ce processus clinique tout en limitant la dégradation des performances. À nouveau, les méta-données issues des informations patient ne sont pas exploitées, seules les données images sont employées dans la mise en place de ce processus. Pour les mêmes raisons que précédemment, ces méta-données ne sont pas représentatives de la population réelle, et peuvent de ce fait introduire un biais au sein des méthodes présentées par l'apport de corrélations invalides dans le monde réel.

Afin de répondre à ces objectifs et de clore l'objet de ce manuscrit, cette partie dédiée à la multimodalité propose diverses approches en cascade susceptibles de convenir à des applications en milieu clinique. Ainsi, ces éléments sont étudiés au sein du Chapitre 8, par l'étude de processus de décision séquentiels d'aide au diagnostic clinique.

VERS UN PROCESSUS MULTIMODAL D'AIDE AU DIAGNOSTIC CLINIQUE

Les travaux de la précédente partie ont porté sur le diagnostic de données malignes sur la modalité de **Microscopie Confocale par Réflectance**, plus particulièrement de pathologies de **Lentigo Maligna** et de **Lentigo Maligna Melanoma**. Au premier niveau hiérarchique, soit celui des images, ces chapitres ont permis la mise en avant de méthodes de séparation des tissus sains, bénins et malins. Au second niveau hiérarchique, soit celui des lésions, ces méthodes ont permis la prédiction des pathologies bénignes et malignes.

Ce nouveau chapitre remplit l'objectif de ce manuscrit en intégrant une dimension multimodale au processus de prise de décision. Tout d'abord, les aspects d'extraction de caractéristiques et de diagnostic par le biais de modèles de prédiction sont traités pour chacune des modalités présentes dans le jeu de données utilisé. De plus, la calibration de ces modèles de prédiction est envisagée afin de rendre cohérentes les décisions entre chacun d'entre eux. Puis, les processus multimodaux envisagés dans ce travail sont présentés selon deux axes : une première méthode qualifiée comme sans mémoire, puis une seconde méthode qualifiée sous le terme d'avec mémoire. Finalement, les résultats obtenus à l'aide de ces méthodes sont présentés et discutés.

Sommaire

8.1	Méthodologie	154
8.2	Méthodes d'aide au diagnostic par modalité	155
8.2.1	Méthodes d'extraction et de détection	155
8.2.2	Pré-analyse	156
8.3	Calibration de modèles	157
8.4	Processus d'aide au diagnostic multimodal	158
8.4.1	Présentation des processus	158
8.4.2	Modèle d'optimisation des seuils de confiance	161
8.5	Présentation des résultats	162
8.5.1	Protocole d'expérimentation	162
8.5.2	Résultats des expérimentations	163
8.6	Discussion	168

8.1/ MÉTHODOLOGIE

Lors de la Partie II, diverses méthodes de classification ont été proposées afin de gérer les données issues de la modalité de **MCR** soit au niveau des images, soit au niveau des lésions. Outre le besoin de répondre à un certain nombre de verrous scientifiques, la réalisation de ces travaux est une étape nécessaire à la mise en oeuvre du processus multimodal.

Néanmoins, à ce stade les méthodes proposées lors des précédents chapitres ne permettent que la prédiction au niveau de la modalité de **MCR**. De plus, les méthodes de la littérature ne considèrent généralement que le diagnostic d'une modalité en particulier, sans considérer son intégration au sein d'un processus décisionnel plus global. Ces problématiques multimodales sont considérées dans de nombreux champs d'applications tels que l'identification d'individus [126], la conduite autonome [214], la reconnaissance vocale [142] ou encore le diagnostic de pathologies [118, 121]. Ces travaux concernent pour la plupart une prise de décision à partir de sources simultanées d'information, mais ne répondent pas à une problématique de prise de décision multimodale séquentielle.

Ainsi, ce dernier chapitre propose une méthodologie en plusieurs étapes pour permettre la mise en place d'un processus d'aide au diagnostic multimodal en milieu clinique. Ce chapitre ne se focalise pas sur la robustesse des diverses modalités, pour lesquelles d'une part sont reprises les propositions de la littérature propres aux modalités de photographie clinique et de dermatoscopie et d'autre part celle de ce manuscrit sur la **MCR** formulés lors du Chapitre 7. Par ailleurs, ce travail s'inspire pour certaines étapes, de travaux réalisés sur la fusion de données.

Pour parvenir à la réalisation de ce processus de décision multimodal, ce manuscrit propose de détailler les diverses étapes qui le composent, dont une représentation est fournie sur la Figure 8.1. Dans un premier temps, ce chapitre propose une mise au point sur les techniques permettant un diagnostic sur les modalités employées dans ce manuscrit. Dans un second temps, le principe de la calibration de modèle est présenté, dont l'utilisation est recommandée dans des contextes impliquant la collaboration de multiples modèles de prédiction. Dans un troisième temps, les processus multimodaux de prise de décision élaborés pour le besoin de cette problématique sont présentés, avec d'une part la considération d'un processus sans mémoire et d'autre part d'un processus avec mémoire. Pour finir, ce chapitre se termine par une présentation des résultats issus des méthodes proposées, suivi d'une discussion sur ceux-ci.

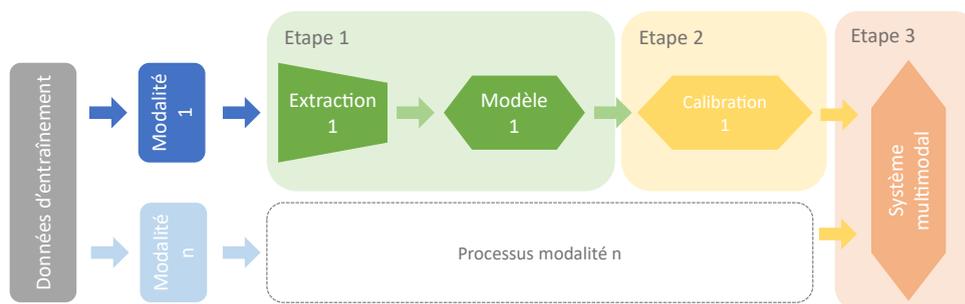


Figure 8.1 : Schéma reprenant la progression de ce chapitre et l'interaction entre les différentes méthodes présentées.

8.2/ MÉTHODES D'AIDE AU DIAGNOSTIC PAR MODALITÉ

Cette section se destine à la présentation des méthodes de diagnostic propres à chaque modalité. L'information brute n'étant pas utilisable en l'état, il est nécessaire de procéder à : (i) sa réduction par l'extraction de caractéristiques pertinentes, (ii) la détection de données malignes à partir de cette information. La première sous-section décrit brièvement les méthodes employées tandis que la seconde sous-section apporte une pré-analyse de ces méthodes.

8.2.1/ MÉTHODES D'EXTRACTION ET DE DÉTECTION

Cette sous-section dédiée aux méthodes d'extraction et de détection des lésions débute par une brève étude des travaux menés sur la photographie clinique et sur la dermatoscopie. Ainsi, la plupart des travaux récents s'orientent autour de mécanismes d'apprentissage profond, dont certains travaux préconisent le recadrage des données employées [116]. Dans ce travail, ce recadrage n'est réalisé que sur les données de photographie clinique jugées moins homogènes, que le reste des données mises à disposition. En termes de gestion de ces modalités d'imagerie, des travaux suggèrent l'utilisation d'architectures telles que Inception-V3 [59, 60] ou encore ResNet-50 [215, 3] associée à des couches de Global Pooling - Moyenne. Ces travaux orientés vers de la classification de lésions de la peau, procèdent à de l'extraction de caractéristiques ou à de la prédiction à partir de ces architectures. En ce qui concerne les méthodes d'extraction et de détection sur la modalité de **MCR**, les méthodes de la Partie II sont mises à contribution dans ce chapitre. Ainsi, la méthode par transfert de connaissances par utilisation de l'architecture ResNet-50 entraîné sur ImageNet couplée à une couche de Globale Pooling - Moyenne permet les meilleurs résultats dans ces conditions. Afin d'uniformiser l'extraction de caractéristiques de ces diverses modalités et par contraintes matérielles, le transfert de connaissances est préféré pour permettre l'extraction de caractéristiques, avec pour choix d'architecture ResNet-50.

D'une part, pour permettre une prédiction sur les données de photographie clinique et de dermatoscopie, ce travail privilégie l'utilisation d'un modèle **SVM** à noyau linéaire à la fois stable et robuste dans la plupart des situations. Ces modalités ne mettant à disposition qu'une unique image, ces deux modalités mettent à disposition un vecteur de taille 1×2048 . D'autre part, pour permettre une prédiction sur les données de **MCR** ce travail reprend les méthodes de prédiction de la Partie II, et notamment l'utilisation d'un modèle par instance multiple de type **NSK** considéré comme le plus performant dans un contexte où le nombre de données est variable. De plus, afin de conserver une dimension de vecteur de caractéristiques régulier pour certains besoins à venir, ce travail a recours aux méthodes de type moyenne, maximum ou de norme p avec des valeurs arbitraires de 4, 6 et 8 comme suggéré par certains travaux de la littérature [115, 217].

Les hyperparamètres associés aux modèles de prédiction cités sont listés dans la Table 8.1. Par ailleurs, le modèle **SVM** employés par ce travail provient de la bibliothèque logicielle *Scikit-learn* [151], tandis que l'implémentation de modèles **NSK** utilise la librairie logicielle *MISVM* [55].

Méthode	Paramètre	Valeurs
SVM	C	[0,01, 0,1, 1, 10, 100, 1000]
NSK		

Table 8.1 : Table reprenant les divers modèles utilisant le principe de **MIL** pour réaliser une classification au niveau lésionnel, ainsi que leurs hyperparamètres associés.

8.2.2/ PRÉ-ANALYSE

Cette pré-analyse des méthodes de classification permet d'obtenir une vision des performances obtenues pour chaque modalité. Pour cela, les évaluations ci-dessous sont réalisées à l'aide d'une valeur de validation croisée de k égale à 10, comme utilisé dans les prochaines expérimentations de ce chapitre et dont les termes sont décrits plus en détail dans la section résultats.

La Table 8.2 met en avant les résultats pour chaque modalité à disposition de ce manuscrit par application des méthodes de la précédente sous-section. Par pondération entre les classes bénigne et maligne, une valeur de score F_1 de 0,62 associée à un écart-type de 0,08 est obtenue sur la photographie clinique ; une valeur de score F_1 de 0,74 associée à un écart-type de 0,11 est obtenue sur la dermatoscopie ; une valeur de score F_1 de 0,80 associée à un écart-type de 0,12 est obtenue sur la **MCR**. Par analyse des résultats de la classe maligne, une valeur de score F_1 de 0,68 associée à un écart-type de 0,09 est obtenue sur la photographie clinique ; une valeur de score F_1 de 0,79 associée à un écart-type de 0,09 est obtenue sur la dermatoscopie ; une valeur de score F_1 de 0,84 associée à un écart-type de 0,11 est obtenue sur la **MCR**. Ces résultats mettent en exergue **une capacité de discrimination croissante des classes bénignes et malignes** selon les modalités de ce travail. Cette capacité de discrimination croissante est exploitée au sein des processus multimodaux proposés dans les prochaines sections.

	Modalité	Précision	Sensibilité	Spécificité	Score F_1
Pondéré	Photographie	0,63 ± 0,08	0,62 ± 0,07	0,59 ± 0,07	0,62 ± 0,08
	Dermatoscopie	0,74 ± 0,11	0,74 ± 0,10	0,70 ± 0,10	0,74 ± 0,11
	Microscopie	0,80 ± 0,11	0,80 ± 0,12	0,81 ± 0,12	0,80 ± 0,12
Malin	Photographie	0,71 ± 0,12	0,66 ± 0,09	0,56 ± 0,14	0,68 ± 0,09
	Dermatoscopie	0,79 ± 0,13	0,79 ± 0,12	0,66 ± 0,22	0,79 ± 0,09
	Microscopie	0,85 ± 0,11	0,83 ± 0,12	0,84 ± 0,11	0,84 ± 0,11

Table 8.2 : Résultats obtenus à partir des méthodes de classification employées sur chacune des modalités utilisées dans ce processus.

Enfin, la Table 8.3 recense les scores par pondération sur les classes bénignes et malignes des méthodes de réduction employées **par l'utilisation d'un modèle SVM à noyau linéaire** sur les données de la modalité de **MCR**. Ainsi, la **méthode de réduction par la moyenne** est à la fois plus performante et la plus stable de ces méthodes avec une valeur de score F_1 de 0,84 associée à un écart-type de 0,07. Dans la suite de ces travaux nécessitant l'utilisation des caractéristiques issues de la modalité de **MCR**, cette **méthode de réduction par la moyenne** est privilégiée afin d'obtenir un vecteur de caractéristiques constant.

Réduction	Précision	Sensibilité	Score F_1
Moyenne	0,84 ± 0,07	0,83 ± 0,07	0,84 ± 0,07
Maximum	0,83 ± 0,12	0,83 ± 0,12	0,83 ± 0,12
Norme $p - 4$	0,82 ± 0,12	0,82 ± 0,12	0,82 ± 0,12
Norme $p - 6$	0,83 ± 0,12	0,83 ± 0,13	0,83 ± 0,13
Norme $p - 8$	0,83 ± 0,12	0,83 ± 0,12	0,83 ± 0,12

Table 8.3 : Résultats pondérés sur la classe bénigne et maligne sur l'efficacité des méthodes de réduction des caractéristiques par classification à l'aide d'un modèle **SVM**.

8.3/ CALIBRATION DE MODÈLES

La majeure partie des modèles de classification mettent à disposition des scores représentant l'appartenance à chacune des classes de la problématique visée. Ces scores sont à l'origine des prédictions de ces modèles et peuvent être obtenus de manière complémentaire à la prédiction. Ainsi, la décision résultante n'est bien souvent que la manifestation de la classe ayant atteint le score de prédiction le plus élevé.

Malheureusement, ces scores n'ont de sens que dans le contexte dans lequel ils ont été obtenus et ne conviennent pas lorsque plusieurs sources d'information différentes sont combinées. Ainsi, une estimation précise de la **probabilité** d'appartenance à chacune des classes est un indicateur plus pertinent dans ce type de situation, ayant un sens en dehors du champ d'action propre à chacun de ces modèles [219].

Diverses méthodes sont proposées afin de transformer ces scores en probabilités, tout d'abord sur des problématiques binaires, puis plus tardivement sur des problématiques à classes multiples. Ce principe de transformation des scores en probabilités porte le nom de processus de **calibration** [137], et cherche à faire converger les scores de sorties $s(x)$ vers une probabilité notée $P(y = 1 | s(x))$. Ces méthodes proposent de corriger ces scores à l'aide de *fonctions de transfert* de différentes formes, dont les performances peuvent être appréhendées à l'aide du **score de Brier**. Ce score dont la formulation est exprimée par l'Equation (8.1) avec p_i la probabilité d'un événement et o_i son observation réelle, s'accompagne de faibles valeurs lorsque la calibration du modèle est supposée parfaite. De plus, la calibration d'un modèle peut être visualisée de manière binaire à l'aide de **courbes de fiabilité**, représentées à l'aide de lots de données constitués de scores semblables, et dont le taux réel d'éléments positifs propre à un lot est exprimé selon les scores moyens de prédiction de celui-ci. Ainsi, la littérature propose diverses méthodes pour permettre de rectifier ces scores, dont les calibrations : (i) *Logistique* [68], (ii) *Sigmoïde* [156, 107], (iii) *Isotonique* [219], (iv) *Bayésienne en Quantiles* [139], (v) ou encore *Bêta* [108].

$$Brier = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2. \quad (8.1)$$

Afin de mesurer l'impact de la calibration, l'hypothèse **sans calibration** tient compte à minima d'une normalisation par minimum et par maximum, les valeurs de scores étant par défaut flottantes et non bornées. Dans la mesure où ce chapitre répond à une problématique binaire, seules les méthodes de calibration binaire sont considérées, dont les méthodes **sigmoïde** et **isotonique** sont les majeurs représentants. Ainsi, **la calibration sigmoïde** se base sur l'hypothèse que les scores issus d'un modèle de prédiction vont suivre une courbe de type *sigmoïde*, et consiste à approcher par une même fonction ces scores afin de les rectifier [143]. Cette fonction est représentée par l'Equation (8.2), dont les deux paramètres de forme A et B sont recherchés par optimisation. À l'instar de la calibration sigmoïde, **la calibration isotonique** est une forme plus généralisée permettant d'approcher une plus grande multitude de tendances, dont l'unique condition requiert d'être *isotonique* ou monotone croissante [143]. Les processus de calibration nécessitent l'utilisation de deux ensembles de données séparés et dédiés à l'entraînement et à la validation. La mise en œuvre de ces calibrations se base sur la librairie logicielle *Scikit-learn* [151].

$$P(y = 1 | s(x)) = \frac{1}{1 + \exp(Af(x) + B)}. \quad (8.2)$$

8.4/ PROCESSUS D'AIDE AU DIAGNOSTIC MULTIMODAL

Cette section se consacre à la mise en place de processus d'aide au diagnostic multimodal, dont l'objectif est de donner une consistance à l'information provenant des diverses modalités d'images. Pour ce faire, cette partie tente de satisfaire le nombre de lésions traitées à chaque modalité tout en conservant une performance de détection maximale. D'une part dans une première sous-section, cet aspect est présenté sous deux différents processus dans le but de résoudre cette problématique. D'autre part, une présentation du modèle permettant la détermination des seuils clés est réalisée.

8.4.1/ PRÉSENTATION DES PROCESSUS

Les processus multimodaux envisagés pour la résolution de la problématique s'inspirent de la procédure actuelle réalisée par les dermatologues. Dans ce processus clinique type, une lésion est prise en charge à partir de modalités simples d'accès à des modalités plus difficile d'accès. Concrètement, le processus type sollicite respectivement la photographie clinique, la dermatoscopie et la **MCR**. Le dermatologue va recourir à une modalité supplémentaire en cas de doute lors du diagnostic, caractérisé dans ce travail par **la zone grise**. Ce travail propose de reproduire ce comportement à l'aide d'une démarche algorithmique, en sollicitant les modalités respectivement dans l'ordre précédent. Ainsi, deux types d'approches sont développés lors des prochains paragraphes, exploitant ce principe de processus **séquentiel** et de *zone grise*.

D'une part, le **processus sans mémoire** consiste à mobiliser les modalités *indépendamment* les unes des autres. Ce type de processus correspond à la prise en charge médicale d'un patient sans communication d'information entre les divers examens menés. Dans un premier temps, l'étape d'entraînement consiste à traiter indépendamment chaque modalité, **sur une fraction des données arbitrairement consacrée à cette étape**, dans le but d'obtenir un modèle de prédiction optimal selon les termes de la Section 8.2. Par ailleurs, c'est également à ce stade de l'entraînement que la méthode de calibration est réalisée selon les termes énoncés lors de la Section 8.3. Dans un second temps, cette même étape d'entraînement consiste à déterminer des seuils de confiance, **sur les données mises à disposition restantes**, permettant de délimiter la zone grise à chaque étape du processus, dont le fonctionnement est détaillé en Section 8.4.2. La Figure 8.2 représente ce concept sous forme de schéma.

D'autre part, le **processus avec mémoire** consiste à mobiliser les modalités en *conservant l'information* recueillie à l'aide des précédentes modalités. Ce type de processus peut être mis en parallèle avec celui d'un praticien qui réalise l'ensemble des examens, ou avec la constitution d'un dossier patient. Dans un premier temps, l'étape consiste à construire des modèles de prédiction cumulatifs **sur une fraction des données, attribuées à l'entraînement**, par agrégation des caractéristiques extraites dont le principe a été adopté dans des travaux de biométrie [104]. À nouveau, il est possible de procéder à la calibration des modèles à ce stade du processus. Dans un second temps, ce processus procède également à la détermination des seuils de confiance, **sur les données mises à disposition restantes**, permettant de délimiter la zone grise à chaque étape du processus et dont le principe de fonctionnement est détaillé en Section 8.4.2. Ce processus est représenté sous forme de schéma à l'aide de la Figure 8.3.

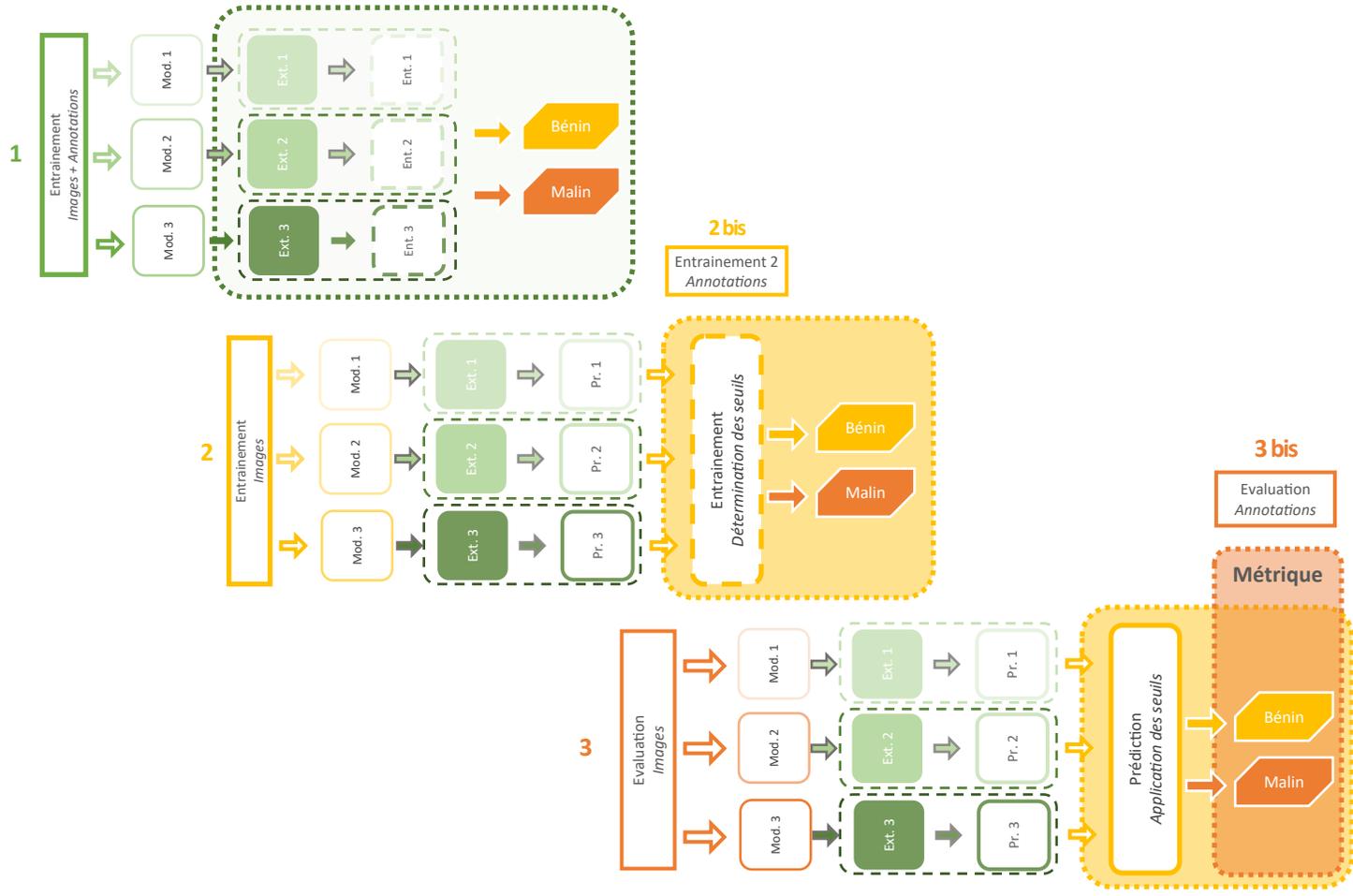


Figure 8.2 : Schéma de fonctionnement du processus multimodal sans mémoire, dont la phase d’entraînement est réalisée en deux temps. Dans un premier temps, un entraînement est réalisé sur les modèles de prédiction au niveau de chaque modalité mise à disposition par le premier jeu de données d’entraînement. Dans un second temps, le modèle de gestion optimise les seuils de confiance à l’aide du second jeu de données d’entraînement. L’entraînement achevé, les performances sont mesurées sur les données d’évaluation.

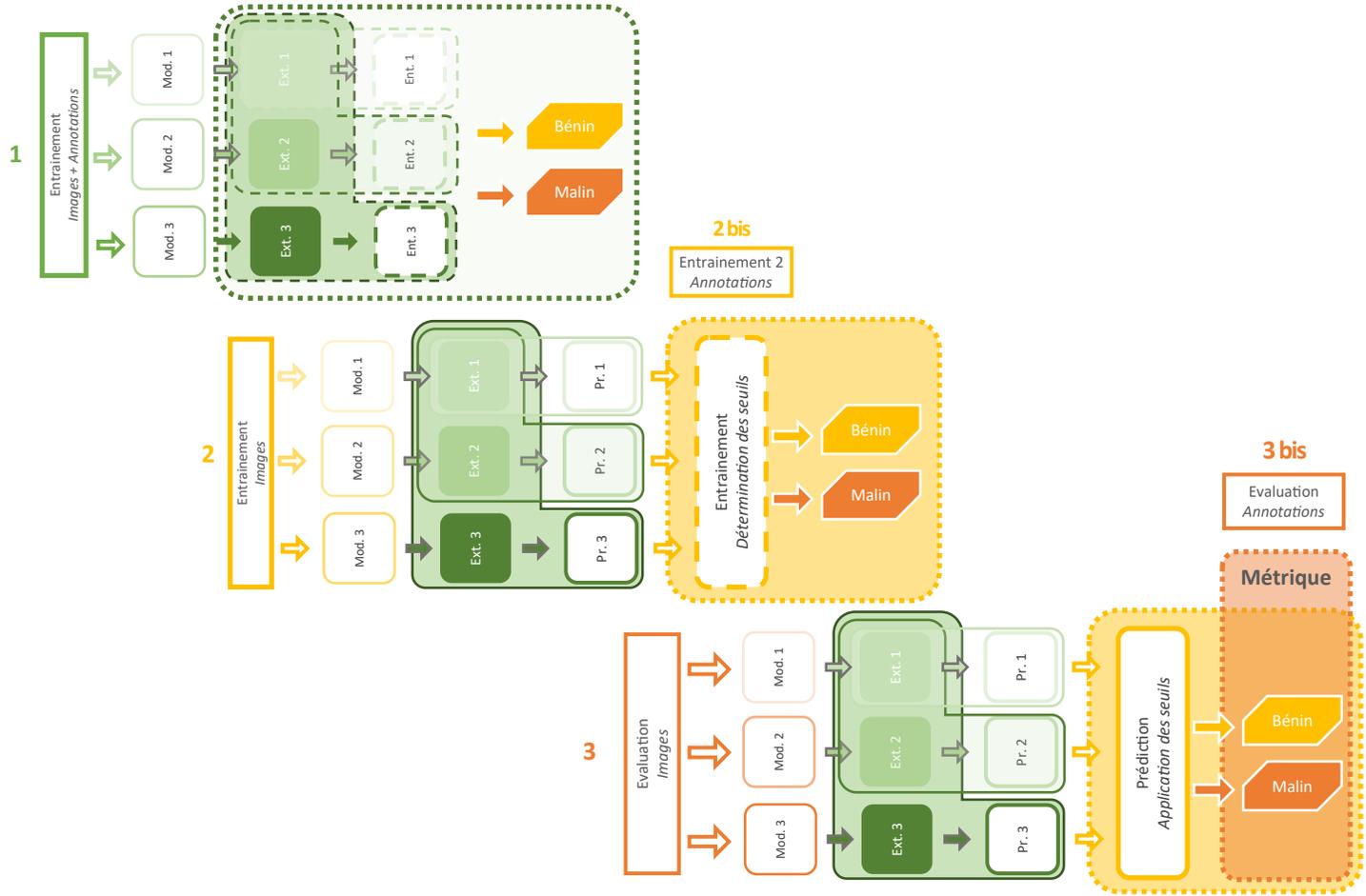


Figure 8.3 : Schéma de fonctionnement du processus multimodal avec mémoire, dont la phase d'entraînement est réalisée en deux temps. Dans un premier temps, un entraînement est réalisé sur les modèles de prédiction propre à chaque étape du processus clinique et pour lesquels les données précédemment acquises sont cumulées. Dans un second temps, le modèle de gestion optimise les seuils de confiance à l'aide du second jeu de données d'entraînement. L'entraînement achevé, les performances sont mesurées sur les données d'évaluation.

8.4.2/ MODÈLE D'OPTIMISATION DES SEUILS DE CONFIANCE

Lors de la précédente sous-section ont été envisagés divers modes de prise en charge multimodal. Néanmoins, ces deux processus en l'état ne permettent pas leur utilisation. Cette nouvelle sous-section clôturera ces aspects de processus multimodal par l'apport de méthodes permettant de déterminer des seuils de confiance pour une prise en charge propre à chacune de ces modalités.

L'entrée nécessaire à l'entraînement de ce modèle est constituée d'un **couple x et y** . D'une part, ce couple se constitue de x , **une matrice de scores de taille $n \times m \times c$** avec n correspondant au nombre d'échantillons, avec m correspondant au nombre de modalités et avec c le nombre de classes. D'autre part, ce couple se constitue de y , **une matrice de sorties attendues de taille $n \times 1$** avec n le nombre d'échantillons. Bien que le but recherché de ce travail soit l'optimisation de la prise en charge des lésions, celui-ci s'inscrit dans une démarche clinique. Ainsi, **une emphase particulière est consacrée à la maximisation d'une métrique**, par défaut le score F_1 bien que celui-ci puisse être substitué par d'autres métriques.

Afin de parvenir à cet objectif, deux hypothèses sont formulées quant à la gestion de ces seuils à chaque modalité. Schématisée sur la Figure 8.4 - Gauche, la première hypothèse considère qu'un seuil de confiance **unique** à toutes les classes est suffisant. Dans ce premier scénario, la zone grise d'une modalité est délimitée par *un seuil global indépendant d'une quelconque classe*. Schématisée sur la Figure 8.4 - Droite, la seconde hypothèse considère que des seuils de confiance **multiples** relatifs à leurs classes respectives sont nécessaires. Dans ce second scénario, la zone grise d'une modalité est délimitée par **des seuils dépendants de chaque classe**, mais néanmoins plus enclins à un sur-apprentissage puisque moins généralisé.

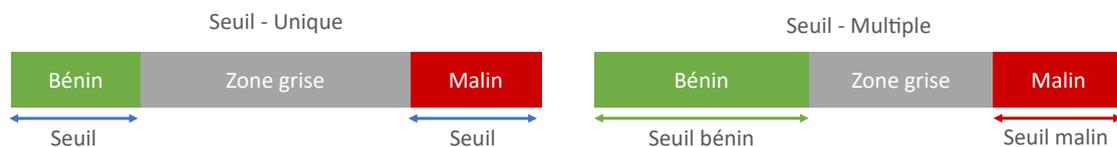


Figure 8.4 : Schéma de représentation des concepts de seuil unique et multiple proposés dans cette section.

Au **niveau des modalités**, l'identification de ces seuils peut être réalisée soit de manière **croissante** à partir de la modalité la moins précise en *améliorant* la métrique vers un score idéal, soit de manière **décroissante** en partant de la modalité la plus précise en *dégradant* la métrique jugée idéale. Au **niveau de la détection** au sein d'une modalité, l'identification de ces seuils peut être réalisée soit de manière **croissante** en *augmentant* la valeur des seuils jusqu'à détecter une baisse de la métrique, soit de manière **décroissante** en *réduisant* la valeur des seuils jusqu'à détecter une baisse de la métrique. Ces divers paramètres sont évalués et synthétisés dans la Table 8.4.

Paramètre	Valeurs
Seuil	[Unique, Multiple]
Modalités	[Croissante, Décroissante]
Détection	[Croissante, Décroissante]

Table 8.4 : Table de recensement des divers paramètres évalués pour la détermination des seuils de confiance optimums au sein du processus multimodal.

8.5/ PRÉSENTATION DES RÉSULTATS

Cette section présente les résultats issus des méthodes précédemment développées dans ce chapitre. Dans un premier temps, cette section procède à la description du protocole d'expérimentation conduisant aux diverses valeurs obtenues. Ainsi, sont présentés les choix concernant le processus de validation, les métriques d'évaluation ainsi que l'organisation des résultats. Puis dans un second temps, les résultats issus des expérimentations mentionnées précédemment sont développés.

8.5.1/ PROTOCOLE D'EXPÉRIMENTATION

Les divers résultats d'évaluation présentés par la suite sont obtenus par l'utilisation d'une **validation croisée imbriquée** dont le principe favorise l'obtention de métriques plus objectives sur les performances du système évalué [26]. De plus, cette structure de validation est démontrée empiriquement moins sensible au biais et à la variance lorsque celle-ci est couplée à des valeurs de k comprises entre 5 et 10 pour l'évaluation [92]. Lors du précédent chapitre, une valeur de k égale à 10 avait été choisie dû à la faible diversité de combinaisons de méthodes proposées. De manière similaire, ce chapitre opte également **pour une valeur de k égale à 10 pour la boucle d'évaluation** pour ces mêmes raisons. Une **valeur de k égale à 2 est utilisée pour la boucle de validation**, jugée suffisante afin de déterminer la meilleure combinaison d'hyperparamètres.

En termes d'apprentissage, l'entraînement des modèles de prédiction au niveau des modalités est jugé plus coûteux que l'entraînement nécessaire à la définition des seuils de confiance sur les modèles proposés. Ainsi, sur les 9 lots mis à disposition du processus d'entraînement, **8 lots sont attribués à l'étape d'entraînement des modalités** tandis que **le dernier lot est attribué au modèle d'optimisation des seuils de confiance**.

Afin de régler au mieux les modèles présentés dans ce chapitre et de donner une évaluation objective malgré les déséquilibres d'annotations, le choix se porte sur le **score F_1** . Cette métrique permet une représentation conjointe des mesures de précision et de rappel au sein d'une même métrique binaire. Dans un premier temps, une valeur pondérée de ce score F_1 selon les classes bénignes et malignes est recensée lors de la présentation des résultats pour chacune des expérimentations. Dans un second temps, la valeur de ce score F_1 selon la classe maligne est relevée pour les systèmes jugés les plus pertinents. Conjointement, les valeurs de **sensibilité** et de **spécificité** sont renseignées de manière complémentaire au score F_1 afin d'obtenir une vision plus complète des performances réelles de ces systèmes. Ces valeurs permettent une mise en parallèle avec les mesures de l'article de Cinotti et al. [40].

Dans un premier temps, ce travail propose de présenter les résultats des expérimentations menées **sur le processus séquentiel sans mémoire** puis **sur le processus séquentiel avec mémoire**. Les résultats pondérés sont présentés pour l'ensemble des données et des expériences de ces deux types de processus. Puis, les résultats sont focalisés sur les expériences les plus performantes de ce manuscrit : d'une part sur l'ensemble des lésions cliniques du jeu de données et d'autre part sur les pathologies de **LM** et **LMM**.

Dans un second temps, les données associées aux tables de résultats sont mises à disposition de manière complémentaire. Ainsi, les données de calibration des processus retenus sont mises à disposition du lecteur, pour permettre une meilleure compréhension de ces derniers. Pour finir, les résultats de simulation de prise en charge des lésions sont présentés avant de finir cette section.

8.5.2/ RÉSULTATS DES EXPÉRIMENTATIONS

Cette section débute par la présentation des résultats du **processus multimodal sans mémoire**. Sans calibration, un score F_1 de 0,81 associé à un écart-type de 0,11 est obtenu par **la méthode à seuils multiples par modalités décroissantes et détection décroissantes**. À l'aide de la calibration sigmoïde, un score F_1 de 0,67 associé à un écart-type de 0,12 est obtenu par **la méthode à seuil unique par modalités décroissantes et détection croissante**. Enfin, à l'aide de la calibration isotonique, un score F_1 de 0,73 associé à un écart-type de 0,08 est obtenu par **la méthode à seuil unique par modalités croissantes et détection croissante**. Les résultats de ce premier processus peuvent être visualisés dans la Table 8.5.

Seuil	Modalités	Détection	Sans calibration	Sigmoïde	Isotonique
Unique	Croissante	Croissante	$0,78 \pm 0,13$	$0,61 \pm 0,18$	$0,73 \pm 0,08$
		Décroissante	$0,76 \pm 0,14$	$0,61 \pm 0,18$	$0,73 \pm 0,12$
	Décroissante	Croissante	$0,80 \pm 0,12$	$0,67 \pm 0,12$	$0,69 \pm 0,17$
		Décroissante	$0,80 \pm 0,12$	$0,65 \pm 0,14$	$0,69 \pm 0,16$
Multiple	Croissante	Croissante	$0,60 \pm 0,14$	$0,61 \pm 0,17$	$0,58 \pm 0,17$
		Décroissante	$0,60 \pm 0,13$	$0,60 \pm 0,17$	$0,58 \pm 0,17$
	Décroissante	Croissante	$0,80 \pm 0,12$	$0,60 \pm 0,12$	$0,71 \pm 0,13$
		Décroissante	$0,81 \pm 0,11$	$0,60 \pm 0,13$	$0,70 \pm 0,15$

Table 8.5 : Table de résultats issus de l'application du processus multimodal sans mémoire. Ces résultats sont exprimés à l'aide du score F_1 et sont issus d'une prise en charge de l'ensemble des cas cliniques mis à disposition de ce manuscrit.

Ce second paragraphe procède à la présentation des résultats du **processus multimodal avec mémoire**. Sans calibration, un score F_1 de 0,82 associé à un écart-type de 0,07 est obtenu par **les méthodes à seuils multiples par modalités décroissantes et détection croissante/décroissantes**. À l'aide de la calibration sigmoïde, un score F_1 de 0,77 associé à un écart-type de 0,10 est obtenu d'une part par **la méthode à seuils multiples par modalités décroissantes et détection croissante**. Enfin, à l'aide d'une calibration Isotonique, un score F_1 de 0,76 associé à un écart-type de 0,10 est obtenu par **la méthode à seuils multiples par modalités décroissantes et détection croissante**. Les résultats liés à ce second processus peuvent être visualisés dans la Table 8.6.

Seuil	Modalités	Détection	Sans calibration	Sigmoïde	Isotonique
Unique	Croissante	Croissante	$0,74 \pm 0,10$	$0,71 \pm 0,14$	$0,62 \pm 0,16$
		Décroissante	$0,72 \pm 0,10$	$0,71 \pm 0,15$	$0,62 \pm 0,16$
	Décroissante	Croissante	$0,80 \pm 0,09$	$0,76 \pm 0,11$	$0,75 \pm 0,09$
		Décroissante	$0,80 \pm 0,08$	$0,77 \pm 0,11$	$0,74 \pm 0,08$
Multiple	Croissante	Croissante	$0,59 \pm 0,13$	$0,59 \pm 0,13$	$0,55 \pm 0,17$
		Décroissante	$0,59 \pm 0,13$	$0,60 \pm 0,14$	$0,55 \pm 0,17$
	Décroissante	Croissante	$0,82 \pm 0,09$	$0,77 \pm 0,10$	$0,76 \pm 0,10$
		Décroissante	$0,82 \pm 0,07$	$0,75 \pm 0,11$	$0,73 \pm 0,13$

Table 8.6 : Table de résultats issus de l'application du processus multimodal avec mémoire. Ces résultats sont exprimés à l'aide du score F_1 et sont issus d'une prise en charge de l'ensemble des cas cliniques mis à disposition de ce manuscrit.

Dans ce troisième paragraphe, les résultats de la prise en charge de la classe maligne sont présentés pour les méthodes dont le score F_1 est maximisé et dont les cellules figurent en gris dans la Table 8.5 et dans la Table 8.6. **Sur l'ensemble des lésions, le processus sans mémoire** obtient une *précision* de 0,81 conjointement à un *rappel (sensibilité)* de 0,91 dont résultent un score F_1 de 0,86 ainsi qu'une *spécificité* assez restreinte de 0,66. **Avec pour seules lésions malignes les pathologies de LM et LMM**, ce processus obtient une *précision* de 0,78 conjointement à un *rappel (sensibilité)* de 0,91 dont résultent un score F_1 de 0,84 et la spécificité reste identique, les lésions bénignes ne changeant pas. **Sur l'ensemble des lésions, le processus avec mémoire** obtient une *précision* de 0,81 conjointement à un *rappel (sensibilité)* de 0,94 dont résultent un score F_1 de 0,87 ainsi qu'une *spécificité* assez restreinte de 0,64. **Avec pour seules pathologies malignes les cas de LM et LMM**, ce processus obtient une *précision* de 0,78 conjointement à un *rappel (sensibilité)* de 0,96 dont résultent un score F_1 de 0,86 et la spécificité reste identique pour les mêmes raisons évoquées. Les résultats de la classe maligne sur l'ensemble des lésions sont recensés dans la Table 8.7, tandis que la Table 8.8 recense ces résultats avec pour seules pathologies malignes les cas de LM et LM.

Type de processus	Précision	Rappel/Sensibilité	Spécificité	Score F_1
Sans mémoire	0,81 ± 0,12	0,91 ± 0,08	0,66 ± 0,20	0,86 ± 0,09
Avec mémoire	0,81 ± 0,10	0,94 ± 0,06	0,64 ± 0,17	0,87 ± 0,05

Table 8.7 : Table reprenant les résultats de la classe maligne, issus de l'application des processus multimodaux sans mémoire et avec mémoire sur l'ensemble des lésions.

Type de processus	Précision	Rappel/Sensibilité	Spécificité	Score F_1
Sans mémoire	0,78 ± 0,14	0,91 ± 0,06	0,66 ± 0,20	0,84 ± 0,10
Avec mémoire	0,78 ± 0,12	0,96 ± 0,05	0,64 ± 0,17	0,86 ± 0,07

Table 8.8 : Table reprenant les résultats de la classe maligne, issus de l'application des processus multimodaux sans mémoire et avec mémoire avec pour seules pathologies malignes les cas de LM et LM.

De manière complémentaire, ce paragraphe quantifie les résultats de calibration à l'aide du score de Brier. Les expérimentations menées sur **le processus sans mémoire**, la modalité de photographie clinique est affecté de manière positive avec un score initial *sans calibration* de 0,27 qui évolue vers une valeur de 0,23 *avec calibration sigmoïde* et vers une valeur de 0,22 *avec calibration isotonique*. En revanche, la courbe de calibration de la dermatoscopie reste relativement inchangée, tandis que celle de la MCR est dégradée. Néanmoins, la stabilité s'accroît sur les trois modalités par l'utilisation des deux méthodes de calibration. Par les expérimentations menées sur **le processus avec mémoire**, aucune différence notable n'est visible sur les scores moyens, et ce, quel que soit le mode de calibration choisi. Néanmoins, contrairement aux précédentes observations, les méthodes de calibration semblent dégrader quelque peu la stabilité des modèles. Les courbes de calibration et scores de Brier des modèles de prédiction au niveau des modalités du processus sans mémoire apparaissent sur la Figure 8.5, tandis que ceux du processus avec mémoire sont indiqués sur la Figure 8.6.

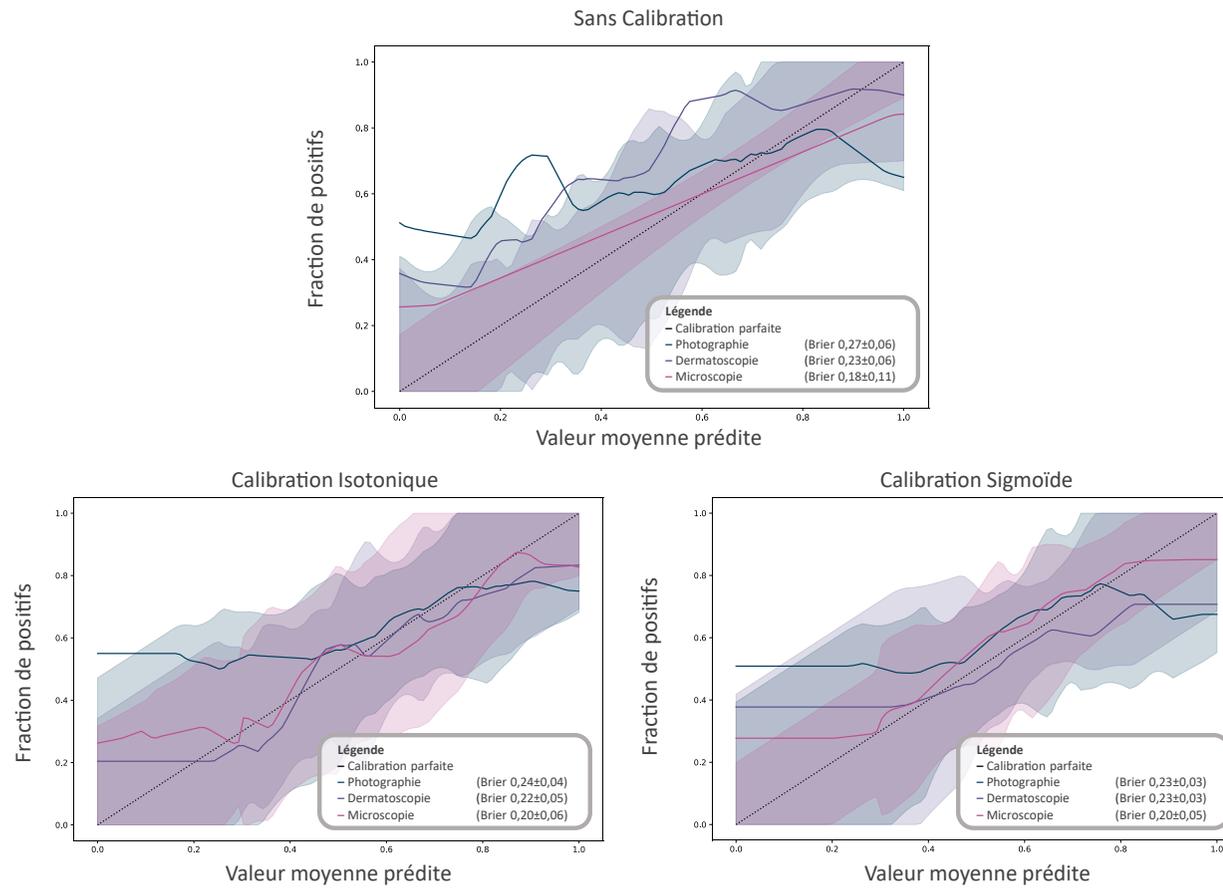


Figure 8.5 : Courbes de calibration et scores de Brier obtenus pour le processus multimodal sans mémoire. L'écart-type calculé sur les différents lots figure en opaque pour les courbes. En haut, le processus multimodal sans calibration ; À gauche, le processus multimodal avec calibration Isotonique ; À droite, le processus multimodal avec calibration Sigmoïde.

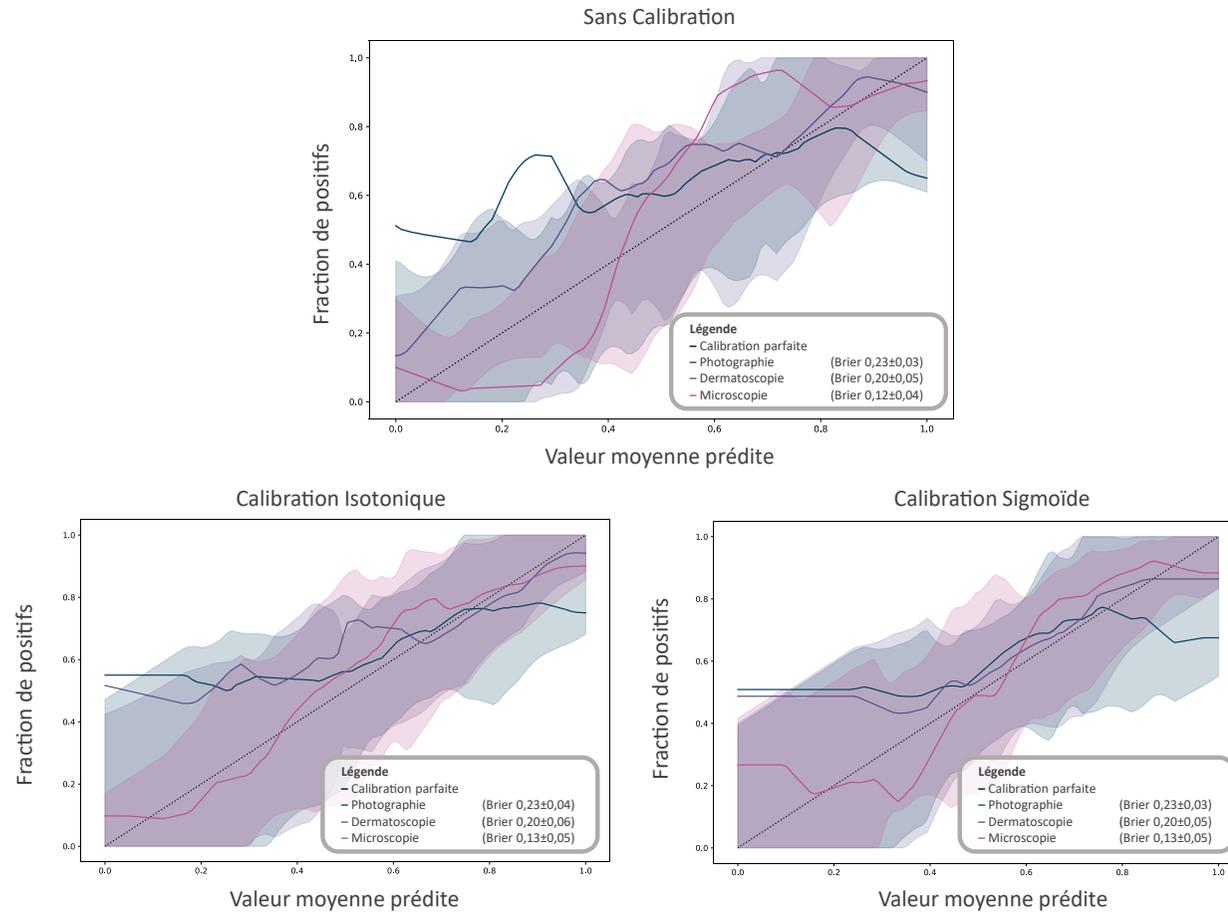


Figure 8.6 : Courbes de calibration et scores de Brier obtenus pour le processus multimodal avec mémoire. L'écart-type calculé sur les différents lots figure en opaque pour les courbes. En haut, le processus multimodal sans calibration ; À gauche, le processus multimodal avec calibration Isotonique ; À droite, le processus multimodal avec calibration Sigmoïde.

Pour finir, les résultats de prise en charge sont l'aboutissement de ce chapitre dont la finalité est de permettre une meilleure répartition des cas cliniques. D'une part, en évitant la réalisation d'examen supplémentaires non nécessaires et d'autre part, afin d'éviter une surcharge des services de dermatologie. Pour cela, à partir des processus multimodaux proposés ayant recueilli le meilleur score F_1 lors du premier et second paragraphe, divers graphiques sont générés automatiquement pour simuler cette prise en charge à partir des seuils déterminés. Dans un premier temps par le principe du **processus sans mémoire**, la photographie clinique permet de prendre en charge 13 lésions considérées comme malignes avec un seuil de confiance fixé à 0,95 ; la dermatoscopie permet de prendre en charge 62 lésions déterminées comme malignes avec un seuil de confiance fixé à 0,56 ; la **MCR** permet de prendre en charge les lésions restantes avec 71 lésions bénignes et 78 lésions considérées comme malignes. Ces résultats de prise en charge des lésions à l'aide d'un processus séquentiel multimodal sans mémoire sont observables sur la Figure 8.7 - Gauche. Par ce processus, 14 lésions malignes sont considérées à tort comme bénignes représentées en vert clair et, 29 lésions bénignes sont considérées à tort comme malignes en rouge clair. Dans un second temps par le principe du **processus avec mémoire**, la photographie clinique permet de prendre en charge une lésion déterminée comme maligne avec un seuil de confiance fixé à 0,99 ; la dermatoscopie permet de prendre en charge 57 lésions déterminées comme malignes avec un seuil de confiance fixé à 0,72 ; la **MCR** permet de prendre en charge les lésions restantes avec 63 lésions bénignes et 103 lésions considérées comme malignes. Ces résultats de prise en charge des lésions à l'aide d'un processus séquentiel multimodal sans mémoire sont observables sur la Figure 8.7 - Droite. Par ce processus, 8 lésions malignes sont considérées à tort comme bénignes représentées en vert foncé et, 31 lésions bénignes sont considérées à tort comme malignes en rouge clair.

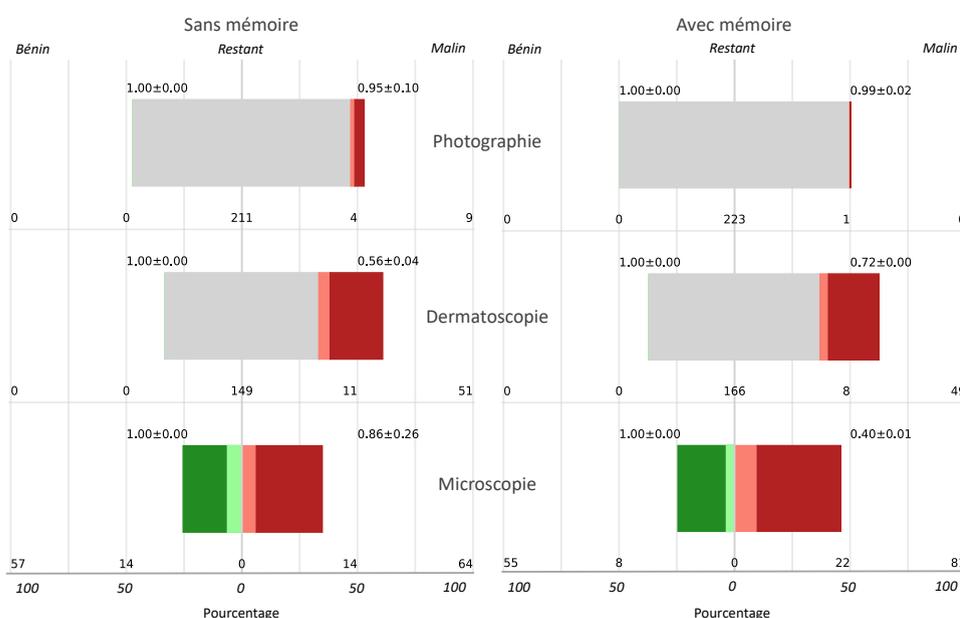


Figure 8.7 : Résultats de la prise en charge des lésions dans un processus séquentiel multimodal. À gauche, selon un processus sans mémoire ; À droite, selon un processus avec mémoire. De gauche à droite, les cas bénins correctement classifiés sont représentés en vert foncé ; les cas bénins incorrectement classifiés sont représentés en vert clair ; les cas malins incorrectement classifiés sont représentés en rouge clair ; les cas malins correctement classifiés sont représentés en rouge foncé.

8.6/ DISCUSSION

Différentes propositions de méthodes sont réalisées dans ce chapitre dédié à la prise en charge de lésions de la peau par multimodalité. Lors de la Section 8.2 consacrée aux méthodes de diagnostic des modalités, le choix s'est porté sur un processus uniforme à l'ensemble des modalités mises à disposition pour ce travail. Ce processus se consacre uniquement à l'extraction de caractéristiques par transfert de connaissances à l'aide de l'architecture de CNN ResNet-50, et à la séparation des données par l'utilisation de modèle SVM à noyau linéaire pour les modalités de photographie clinique et dermatoscopie ainsi que le modèle NSK pour la modalité de MCR. Pour rappel, sur les lésions cutanées de l'étude de Cinotti et al. [41], l'analyse des résultats des experts sur la détection des pathologies malignes conduit aux diverses valeurs suivantes : (i) sur **les données de photographie clinique** une sensibilité de $0,72 \pm 0,08$ et une spécificité de $0,78 \pm 0,08$, (ii) sur **les données de dermatoscopie** une sensibilité de $0,69 \pm 0,10$ et une spécificité de $0,85 \pm 0,06$, (iii) et sur **les données de MCR** une sensibilité de $0,84 \pm 0,05$ et une spécificité de $0,75 \pm 0,06$. Par l'utilisation de ce processus de prédiction uniforme et sur ces données dont les annotations sont réactualisées à la suite de visites de contrôle ultérieures, les résultats suivants sont obtenus sur la classe maligne : (i) sur **les données de photographie clinique** une sensibilité de $0,66 \pm 0,09$ et une spécificité de $0,56 \pm 0,14$, (ii) sur **les données de dermatoscopie** une sensibilité de $0,79 \pm 0,12$ et une spécificité de $0,66 \pm 0,22$, (iii) et sur **les données de MCR** une sensibilité de $0,83 \pm 0,12$ et une spécificité de $0,84 \pm 0,11$.

Ainsi, bien que les prédictions issues de cette méthode soient d'une qualité bien inférieure pour la photographie clinique, en comparaison de la performance des experts, les modalités de dermatoscopie et de MCR sont en revanche assez similaires. En outre, la pertinence de ces modalités s'est avérée suivre le processus de prise en charge clinique, avec un score F_1 dont la valeur s'accroît pour chaque évaluation réalisée selon ce processus. Néanmoins, la stabilité de prédiction semble suivre une démarche opposée, avec une valeur de 0,08 pour la photographie clinique qui atteint 0,12 avec la MCR par les résultats de l'évaluation pondérée. L'un des facteurs à prendre en compte **est celui de la quantité de données**, de l'ordre de la centaine pour la photographie clinique ainsi que la dermatoscopie, et de l'ordre du millier pour les données de la modalité MCR.

Cette pertinence croissante dans la qualité du diagnostic donne lieu à un processus global de prise en charge des lésions dans lequel **une modalité de rang supérieur est sollicitée tant que la confiance dans un diagnostic est jugée insuffisante**. Sur ce principe, deux processus types sont définis pour répondre à cette tâche, le premier *sans mémoire* et le second *avec mémoire*. Dans un premier temps, **aucune différence notable n'est à reporter en termes de score F_1 entre la proposition à seuil unique et à seuils multiples**, avec initialement un risque de sur-apprentissage plus grand pour cette seconde proposition qui ne s'est pas manifesté lors de ces expérimentations. Dans un second temps, plusieurs hypothèses ont été formulées pour permettre la recherche des seuils de confiance. La première de ces hypothèses consiste à rechercher les seuils de confiance optimums soit par amélioration, c'est-à-dire en partant de la modalité la moins précise, soit par dégradation des résultats, c'est à dire en partant de la modalité la plus précise. Les mesures de score F_1 obtenues **par dégradation des résultats sont ainsi sensiblement supérieures à celles obtenues par amélioration des résultats**. La seconde de ces hypothèses consiste à déterminer le sens de recherche des seuils de confiance, aboutissant à déterminer les seuils de manière la plus *restrictive* (seuil haut) ou la plus *laxiste* (seuil bas). Ainsi, cette seconde hypothèse **n'amène aucune différence dans le cadre du processus sans mémoire, tandis que son influence peut être remarquée dans le processus avec mémoire**.

Afin d'apporter une plus grande signification aux scores issus de la classification et dans le but d'améliorer les scores précédents, le principe de la calibration de modèles est également exploré dans ce chapitre. Sur les nombreuses méthodes de calibration existantes, seules les méthodes de calibrations isotonique et sigmoïde sont appliquées aux modèles de prédiction qui composent les deux processus de prise en charge proposés. Ainsi, quel que soit le processus considéré de prise en charge des lésions, **l'utilisation de ces méthodes de calibration a amené une dégradation de la qualité des résultats précédemment obtenus**. Pour la méthode de calibration sigmoïde, l'une des explications possibles est l'inadéquation entre la courbe de calibration initiale de ces modèles et la fonction de correction sigmoïde, visible en Figure 8.5 et Figure 8.6. Pour les deux méthodes de calibration, le faible nombre d'éléments mis à disposition de l'entraînement de ces méthodes suscite un doute non négligeable et particulièrement pour la calibration isotonique plus encline au sur-apprentissage. Ces aspects de calibration sont à améliorer dans de futurs travaux menés sur cette thématique, dans le but d'améliorer la correspondance entre les scores et les probabilités des modèles de classification employés, et de réduire par la même occasion la variabilité des prédictions.

En définitive, bien que le processus avec mémoire soit supérieur au processus sans mémoire, la différence entre ces deux modes de prise en charge n'est pas notable, avec respectivement un score F_1 pondéré de 0,82 et de 0,81. D'une part, il est important de souligner que **ces processus n'augmentent pas la qualité des prédictions réalisées**. En effet, l'une des hypothèses formulées initialement était qu'une restriction des cas les plus fiables pouvait permettre d'augmenter la qualité du diagnostic final, ce que ce travail ne parvient pas à accomplir. D'autre part, les processus proposés **permettent de réduire le nombre d'actes médicaux**. Cependant, dans leur état actuel ces processus de prise en charge multimodaux peuvent être davantage perçus comme un compromis entre l'exactitude des détections et la répartition des cas cliniques tout au long du processus. Comparativement, le processus sans mémoire retenu gère mieux la répartition des lésions tout au long des modalités que celui avec mémoire, mais avec un compromis sur le score F_1 dont la valeur pondérée est de 0,81 contre une valeur de 0,82.

CONCLUSION ET PERSPECTIVES

ÉTAT D'AVANCEMENT

Le travail présenté tout au long de ce manuscrit a proposé une réponse à la problématique de diagnostic des lésions de la peau en dermatologie, par l'utilisation de la multimodalité. Plus spécifiquement, les méthodes proposées dans ce travail ont été orientées autour de la détection des pathologies de **LM** et de **LMM** dont l'incidence ne cesse de croître au fil des années, et pour lesquelles la réalisation d'un diagnostic clinique par le praticien peut être ardue selon leurs configurations respectives ou stade d'avancement. En s'inspirant notamment des avancées sur les méthodes d'apprentissage profond, ce manuscrit a établi une réflexion sur ce sujet permettant la mise en place d'outils d'aide à la détection précoce de pathologies malignes de la peau.

Afin de parvenir à cet objectif, la Partie **I** a proposé une présentation des éléments essentiels à la réalisation de ce sujet dont : (i) une explication clinique **des principales pathologies de la peau mais surtout du LM et de LMM**, (ii) une présentation de **l'interaction de la lumière avec la peau et des modalités d'imagerie existantes**, (iii) et une description **des méthodes d'apprentissage sollicitées** dans ce manuscrit pour permettre la mise en place de méthodes de diagnostic. En outre, cette première partie a servi à présenter **les caractéristiques de la base de données de Cinotti et al. [40]** employée pour évaluer les méthodes d'aide au diagnostic par ordinateur proposées. De même, les apports qui ont été réalisés sur cette base de données telle que la mise à jour des annotations ou encore les modifications d'organisation sont détaillés dans cette même partie. Par ailleurs, cette présentation de la base de données **a servi à comprendre les éléments-clés identifiés par les dermatologues sur chacune des modalités** afin de caractériser ces désordres au sein des tissus. La Partie **II** a mis en avant **les diverses méthodes de diagnostic sur les données de MCR** au niveau des images avant de parvenir au diagnostic d'une lésion de cette modalité. Cette étape du manuscrit fut nécessaire à la réalisation du processus multimodal dans son ensemble, cette modalité étant relativement complexe à traiter en l'état. Enfin, la Partie **III** a permis de répondre à la problématique de ce manuscrit en proposant au lecteur diverses méthodes dont le principe permet une prise en charge séquentielle d'une lésion à l'aide des diverses modalités mises à disposition et couramment utilisées en milieu clinique. Les prochains paragraphes commentent les résultats qui ont été obtenus lors de ces deux dernières parties de contribution.

Contrairement à des modalités en plein essor telles que la dermatoscopie, la modalité de **MCR** est moins développée et ne propose que très peu de travaux en lien avec les pathologies de **LM** et de **LMM**. Ainsi, la Partie **II** a repris les éléments existants de la littérature dont le contenu sert de base de référence en termes de performances aux méthodes proposées. *D'une part au niveau des images*, le processus de classification mêlant simplicité de mise en œuvre et qualité du diagnostic. Ce processus type a été décrit comme étant composé : (i) **d'une extraction de caractéristiques par du transfert de connaissances basé sur l'architecture ResNet-50 pré-entraînée sur ImageNet associé à une couche de Global Pooling - Moyenne**, (ii) **d'une mise à l'échelle des caractéristiques par Minimum / Maximum**, (iii) et **d'un modèle SVM à noyau linéaire**. Cette combinaison a permis l'obtention d'une **valeur de score F_1 pondérée sur les classes saines**,

bénignes et malignes de 0,77 associée à un écart-type de 0,04. Sur la classe maligne, une valeur de score F_1 de 0,82 associée à un écart-type de 0,02 est obtenue, complété par des valeurs de sensibilité de 0,79 et de spécificité de 0,75. Des améliorations non significatives ont été apportées à l'aide de méthodes exploitant le principe d'échelles multiples ou encore le principe de fenêtre glissante, couplées au précédent processus de classification. De ce fait, ces méthodes n'ont pas été considérées dans la suite de ce manuscrit. *D'autre part au niveau des lésions*, les résultats les plus pertinents ont été obtenus à l'aide d'une extraction de caractéristiques par le transfert de connaissances basé sur l'architecture ResNet-50 pré-entraînée sur ImageNet associée à une couche de Global Pooling - Moyenne et du modèle **NSK** basé sur un principe de **MIL**. Une mise à l'échelle étant déjà réalisée par ce modèle, son utilisation en dehors a été rendue caduque. En considérant le diagnostic clinique le plus récent, cette méthode permet d'atteindre **sur l'ensemble des lésions malignes des valeurs de sensibilité de 0,86 et de spécificité de 0,77**, tandis que **sur les lésions de type LM et LMM sont obtenues des valeurs de sensibilité de 0,87 et de spécificité de 0,77**. Ainsi, cette méthode a permis d'obtenir des performances proches de celles obtenues par l'analyse des résultats des experts en **MCR** sur les annotations originales de Cinotti et al. [41], avec sur l'ensemble des annotations malignes des valeurs de sensibilité de 0,84 et de spécificité de 0,75, tandis que sur les lésions de type **LM** et **LMM** sont obtenues des valeurs de sensibilité de 0,80 et de spécificité de 0,81.

Afin de proposer une réponse à la problématique décrite dans ce manuscrit, la Partie **III** a apporté diverses solutions en lien avec la thématique du sujet. Bien que perfectibles, ces méthodes développées sur la modalité de **MCR** permettent la gestion de ces données, qui ont constitué un verrou à la mise en place d'un processus multimodal. Pour les modalités de photographie clinique et de dermatoscopie, des approches par transfert de connaissances ont été préférées aux méthodes par réglage fin représentant un bon compromis entre les performances et ressources matérielles nécessaires. Par la proposition d'un processus sans mémoire, une valeur de score F_1 pondérée sur les classes bénignes et malignes de 0,81 et associée à un écart-type de 0,11 a été mesurée. Cette valeur a découlé d'une prise en charge : (i) sur **la modalité de photographie clinique** : de 4 lésions bénignes incorrectement identifiées et de 9 lésions malignes correctement identifiées, (ii) sur **la modalité de dermatoscopie** : de 11 lésions bénignes incorrectement identifiées et de 51 lésions malignes correctement identifiées, (iii) sur **la modalité de MCR** : de 71 lésions bénignes dont 57 lésions sont correctement identifiées, de 78 lésions malignes dont 64 lésions sont correctement identifiées. Par la proposition d'un processus avec mémoire, une valeur de score F_1 pondérée sur les classes bénignes et malignes de 0,82 associée à un écart-type de 0,07 a été mesurée. Cette valeur a découlé d'une prise en charge : (i) sur **la modalité de photographie clinique** : de une lésion bénigne incorrectement identifiée, (ii) sur **la modalité de dermatoscopie** : de 8 lésions bénignes incorrectement identifiées et de 49 lésions malignes correctement identifiées, (iii) sur **la modalité de MCR** : de 77 lésions bénignes dont 55 lésions sont correctement identifiées et de 89 lésions malignes dont 81 lésions sont correctement identifiées.

Pour finir, ces résultats sont bien évidemment perfectibles et ces méthodes ne se substituent pas à l'avis du dermatologue. Bien que de nombreuses dispositions soient prises pour permettre l'obtention de résultats objectifs, les résultats des évaluations sont issus des données de l'étude de Cinotti et al. [41]. Ainsi, ces données sont une fraction infime des situations cliniques et de la diversité de tissus pouvant être rencontrées. La section suivante propose diverses perspectives envisagées afin d'améliorer les résultats actuels.

PERSPECTIVES

Les résultats obtenus et leur analyse ouvrent sur des perspectives intéressantes à notre sens. Tout d'abord, la classification des lésions **MCR** est encore imprécise, et bien que les performances obtenues par réglage fin de **CNN** ne soient pas suffisantes avec la proposition de ce manuscrit, cette solution semble adaptée à ce type de tâche et permet de localiser efficacement des pathologies responsables au sein des images de manière efficiente. En addition, les travaux actuellement proposés et ceux de la littérature ne considèrent à ce jour qu'une détection des tissus pathologiques au sein des images, pourtant les lésions de **LM** et **LMM** se définissent par une prolifération de ces tissus autour des follicules pileux. Une des pistes consiste en une identification des follicules pileux par des techniques de traitement d'images traditionnelles ou par l'utilisation de **CNN** spécialisés dans la détection d'objets, combinée avec l'identification de tissus pathologiques aux alentours de ceux-ci. Une autre limitation de la modalité de **MCR** concerne l'acquisition des données à disposition. Dans l'état actuel, aucune information de localisation spatiale n'est mise à disposition, ainsi une piste possible est de travailler avec les données brutes fournies par l'appareil afin d'accéder à ces données de localisation et envisager des méthodes volumiques par exemple.

De plus, de nombreux éléments d'améliorations peuvent être apportés au processus multimodal actuel. Les processus d'extraction de caractéristiques et de prédiction proposés ne sont pas suffisamment spécifiques à la photographie clinique et à la dermatoscopie. Également, certains travaux préconisent un recadrage réalisé manuellement autour des lésions pour la photographie clinique, et non considéré pour la dermatoscopie ; l'utilisation de méthodes automatiques de recadrage peut être envisagée pour améliorer ces performances. De manière similaire, les principes de calibration employés sont à revoir, leur utilisation s'étant révélée inefficace et ayant conduit à une dégradation des résultats. Un autre point d'amélioration du processus multimodal concerne la détermination des seuils de confiance. La proposition actuelle ne parvient à concilier l'optimisation que d'une seule contrainte à savoir d'une métrique au choix de l'utilisateur, et délaisse une seconde contrainte liée à l'optimisation du processus de prise en charge des lésions au fil des modalités. De plus, la recherche des seuils de confiance est réalisée de manière itérative dont l'optimalité peut être contestée.

Les données de ce manuscrit sont délicates à confronter aux résultats issus de l'évaluation des diagnostics des divers experts de l'étude de Cinotti et al. [41]. L'une d'entre elles serait de mettre à jour ces résultats à partir des nouvelles annotations issues de visites de contrôle ultérieures de ces mêmes lésions. Par ailleurs, l'une des contraintes majeures des expérimentations liées aux processus multimodaux est la quantité de lésions mises à disposition. L'une des améliorations à court terme consisterait à travailler en collaboration avec les médecins investigateurs sur l'accroissement de ce nombre de cas dédiés à l'entraînement de ce processus afin d'exclure les hypothèses de dégradation des performances liées à ce nombre de lésions. À plus long terme, il est envisagé d'utiliser des données issues de divers centres cliniques pour vérifier la robustesse des méthodes proposées et exclure une dépendance de celles-ci aux opérateurs et appareillages utilisés.

Enfin, les expérimentations menées visent à égaler les annotations issues de la corrélation clinico-pathologique, dont le diagnostic bien que très précis, s'est révélé inexact lors de visites post contrôle sur 8 des 223 lésions initiales. Ainsi ce travail aurait pu consister en la reproduction des résultats issus du diagnostic des experts de l'étude de Cinotti et al. [41] et non du diagnostic issu de la corrélation clinico-pathologique. Une telle problématique aurait nécessité la mise en place d'un processus de classification compatible avec un contexte de travail à annotateurs multiples, dont certains travaux font l'objet [161]. Cette expertise peut néanmoins varier selon le contexte et les lésions abordées. Il est dans ce cas nécessaire de préférer l'utilisation de certaines des annota-

tions. Afin de limiter la sollicitation des experts, il peut être pertinent d'envisager des annotations en provenance d'internes en médecine ou de sources autres de *crowdsourcing* dont la fiabilité varie de manière encore plus forte et que certains travaux tentent de solutionner [211]. À des stades de réflexion plus avancés, il est également important de s'intéresser à la gestion de nouvelles lésions venant compléter la base initiale d'entraînement, pour permettre d'enrichir le processus de prédiction existant. Diverses méthodes permettent alors de recourir à une expertise adéquate selon la difficulté de la lésion rencontrée [210].

VALORISATIONS SCIENTIFIQUES

Pour finir, ces travaux ont donné lieu à de multiples valorisations scientifiques par la publication d'articles de revue, par la participation à des conférences mais également à des groupes de travail. Au moment de l'écriture de ce manuscrit, un article sur la valorisation des travaux de multimodalité présenté dans ce manuscrit est en cours de rédaction. Par ailleurs, un article à orientation médicale sur un projet connexe est prévu sur le thème de la classification de tumeurs des parois buccales et de la gorge par spectrométrie.

Article de revue :

- Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, Elisa Cinotti, and Franck Marzani. Classification of lentigo maligna at patient-level by means of reflectance confocal microscopy data. *Applied Sciences*, 10(8):2830, April 2020

Conférence internationale :

- Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, Elisa Cinotti, and Franck Marzani. Two schemes for automated diagnosis of lentigo on confocal microscopy images. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, July 2019

Conférence nationale :

- Romain Cendre, Alamin Mansouri, Jean-luc Perrot, Elisa Cinotti, and Franck Marzani. Extraction et évaluation de caractéristiques adaptées pour la classification du Lentigo à partir d'images de Microscopie Confocale. In *GRETSI*, pages 2–5, 2019

Groupe de travail international :

- Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, and Franck Marzani. A brief introduction to concepts and perspectives on dermatology images. In *COST, Comulis*, Bruxelles, 2019
- Romain Cendre, Alamin Mansouri, Jean-luc Perrot, Elisa Cinotti, and Franck Marzani. Application to Dermatology - A focus on reflectance confocal microscopy for lentigo detection. In *Colourlab Workshop, Skin Imaging*, Gjøvik, 2020

Code source :

- <https://github.com/RomainCendre/classification>

BIBLIOGRAPHIE

- [1] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Gregory S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian J Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Józefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Gordon Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul A Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda B Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *CoRR*, abs/1603.0, 2016.
- [2] Naheed R Abbasi, Helen M Shaw, Darrell S Rigel, Robert J Friedman, William H Mccarthy, Iman Osman, Alfred W Kopf, and David Polsky. Early Diagnosis of Cutaneous Melanoma Revisiting the ABCD Criteria. *JAMA*, 292(22):2771–2776, 2004.
- [3] Mugahed A Al-antari, Patricio Rivera, Mohammed Al-masni, Edwin Valarezo Añazco, Geon Gi, T.-Y Kim, H Park, and T.-S Kim. An Automatic Recognition of Multi-class Skin Lesions via Deep Learning Convolutional Neural Networks. In *ISIC2018: Skin Image Analysis Workshop and Challenge*, 2018.
- [4] Amin Alqudah and Ali Mohammad Alqudah. Sliding window based support vector machine system for classification of breast cancer using histopathological microscopic images. *IETE Journal of Research*, pages 1–9, March 2019.
- [5] Khaled Alsaih, Guillaume Lemaitre, Join Massich Vall, Mojdeh Rastgoo, Desire Sidibe, Tien Y Wong, Ecosse Lamoureux, Dan Milea, Carol Y Cheung, and Fabrice Meriaudeau. Classification of SD-OCT volumes with multi pyramids, LBP and HOG descriptors: Application to DME detections. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, August 2016.
- [6] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann. Support vector machines for multiple-instance learning. In *Advances in neural information processing systems*, pages 577–584, 2003.
- [7] Pablo Arbelaez, Jordi Pont-Tuset, Jon Barron, Ferran Marques, and Jitendra Malik. Multi-scale combinatorial grouping. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2014.
- [8] S Arivazhagan and L Ganesan. Texture classification using wavelet transform. *Pattern Recognition Letters*, 24(9-10):1513–1521, June 2003.
- [9] M. Baccard and O. Verola. Mélanome de dubreuilh. *Annales de Dermatologie et de Vénérologie*, 136(10):749–751, October 2009.
- [10] Adrian Barbu, Le Lu, Holger Roth, Ari Seff, and Ronald M. Summers. An analysis of robust cost functions for CNN in computer-aided diagnosis. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):253–258, April 2016.

- [11] Daniel Barolet. Light-emitting diodes (LEDs) in dermatology. *Seminars in Cutaneous Medicine and Surgery*, 27(4):227–238, December 2008.
- [12] Yoshua Bengio and Yves Grandvalet. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *J. Mach. Learn. Res.*, 5:1089–1105, 2004.
- [13] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. ACM Press, 2009.
- [14] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(Feb):281–305, 2012.
- [15] Anil Kumar Bhatta, Uma Keyal, and Yeqiang Liu. Application of high frequency ultrasound in dermatology. *Discovery medicine*, 26(145):237–242, dec 2018.
- [16] Christopher M Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [17] Isabelle Bloch. *Fusion d'informations en traitement du signal et des images*, volume 14. 2003.
- [18] Hendrik Blockeel, David Page, and Ashwin Srinivasan. Multi-instance tree learning. In *Proceedings of the 22nd international conference on Machine learning*, pages 57–64, 2005.
- [19] Leo Breiman. *Machine Learning*, 45(1):5–32, 2001.
- [20] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
- [21] Razvan C. Bunescu and Raymond J. Mooney. Multiple instance learning for sparse positive bags. In *Proceedings of the 24th international conference on Machine learning - ICML '07*. ACM Press, 2007.
- [22] Teresa Cammarota, Fabio Pinto, Antonio Magliaro, and Antonino Sarno. Current uses of diagnostic high-frequency US in dermatology. *European Journal of Radiology*, 27:S215–S223, May 1998.
- [23] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. An Analysis of Deep Neural Network Models for Practical Applications. pages 1–7, 2016.
- [24] P.J. Caspers, G.W. Lucassen, and G.J. Puppels. Combined in vivo confocal raman spectroscopy and confocal microscopy of human skin. *Biophysical Journal*, 85(1):572–580, July 2003.
- [25] Orlando Catalano and Alfredo Siani. Cutaneous melanoma: Role of ultrasound in the assessment of locoregional spread. *Current Problems in Diagnostic Radiology*, 39(1):30–36, January 2010.
- [26] Gavin C. Cawley and Nicola L. Talbot. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010.

- [27] M. Emre Celebi, Hassan A. Kingravi, Bakhtiyar Uddin, Hitoshi Iyatomi, Y. Alp Aslandogan, William V. Stoecker, and Randy H. Moss. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6):362–373, September 2007.
- [28] Romain Cendre, Alamin Mansouri, Jean-luc Perrot, Elisa Cinotti, and Franck Marzani. Extraction et évaluation de caractéristiques adaptées pour la classification du Lentigo à partir d'images de Microscopie Confocale. In *GRETSI*, pages 2–5, 2019.
- [29] Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, Elisa Cinotti, and Franck Marzani. Two schemes for automated diagnosis of lentigo on confocal microscopy images. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*. IEEE, July 2019.
- [30] Romain Cendre, Alamin Mansouri, Jean-luc Perrot, Elisa Cinotti, and Franck Marzani. Application to Dermatology - A focus on reflectance confocal microscopy for lentigo detection. In *Colourlab Workshop, Skin Imaging*, Gjøvik, 2020.
- [31] Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, Elisa Cinotti, and Franck Marzani. Classification of lentigo maligna at patient-level by means of reflectance confocal microscopy data. *Applied Sciences*, 10(8):2830, April 2020.
- [32] Romain Cendre, Alamin Mansouri, Jean-Luc Perrot, and Franck Marzani. A brief introduction to concepts and perspectives on dermatology images. In *COST, Comulis*, Bruxelles, 2019.
- [33] Olivier Chapelle and Alexander Zien. Semi-Supervised Classification by Low Density Separation. In *AISTATS*, pages 57–64. Citeseer, 2005.
- [34] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: Synthetic Minority over-Sampling Technique. *J. Artif. Int. Res.*, 16(1):321–357, 2002.
- [35] Hao Chen, Dong Ni, Jing Qin, Shengli Li, Xin Yang, Tianfu Wang, and Pheng Ann Heng. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1627–1636, September 2015.
- [36] Yixin Chen and James Z Wang. Image categorization by learning and reasoning with regions. *Journal of machine learning Research*, 5(Aug):913–939, 2004.
- [37] François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.
- [38] François Chollet et al. Keras applications. <https://keras.io/applications>, 2015.
- [39] Vincent Christlein, Lukas Spranger, Mathias Seuret, Anguelos Nicolaou, Pavel Král, and Andreas Maier. Deep Generalized Max Pooling, 2019.
- [40] E. Cinotti, B. Labeille, S. Debarbieux, C. Carrera, F. Lacarrubba, A.M. Witkowski, E. Moscarella, E. Arzberger, H. Kittler, P. Bahadoran, S. Gonzalez, P. Guitera, M. Agozzino, F. Farnetani, R. Hofmann-Wellenhof, M. Ardigò, P. Rubegni, L. Tognetti, J. Łudzik, I. Zalaudek, G. Argenziano, C. Longo, S. Ribero, J. Malvehy, G. Pellacani, F. Cambazard, and J.L. Perrot. Dermoscopy vs. reflectance confocal microscopy for the diagnosis of lentigo maligna. *Journal of the European Academy of Dermatology and Venereology*, 32(8):1284–1291, February 2018.

- [41] Elisa Cinotti, Jean-Luc Perrot, Bruno Labeille, Frederic Cambazard, Romain Vie, Alexandre Delalleau, Linda Tognetti, and Pietro Rubegni. Season and anatomic site effect on skin color and xerosis quantified using an ultra-high definition videodermoscope. In *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, May 2016.
- [42] Wallace H Clark, Lynn From, Evelina a Bernardino, H Clark, Evelinaa Bernardino, and Martin C Mihm. The Histogenesis and Biologic Behavior of Primary Human Malignant Melanomas of the Skin The Histogenesis Malignant Behavior of Primary Melanomas of the Skin1. *Cancer Research*, (March):705–727, 1969.
- [43] Luis Pedro Coelho. Mahotas: Open source software for scriptable computer vision. *Journal of Open Research Software*, 1(1):e3, July 2013.
- [44] Osvaldo Correia, Bernardo Correia, and Ana Filipa Duarte. A skin cancer prevention campaign. *JAMA Dermatology*, 153(2):129, February 2017.
- [45] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, September 1995.
- [46] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967.
- [47] B.V. Dasarathy. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE*, 85(1):24–38, 1997.
- [48] Luis M.T. de Carvalho, Fausto W. Acerbi, Jan G.P.W. Clevers, Leila M.G. Fonseca, and Steven M. de Jong. Multiscale feature extraction from images using wavelets. In *Remote Sensing Image Analysis: Including The Spatial Domain*, pages 237–270. Springer Netherlands, 2004.
- [49] Elisa de Oliveira Barcaui, Antonio Carlos Pires Carvalho, Paulo Marcos N. Valiante, and Carlos Baptista Barcaui. High-frequency ultrasound associated with dermoscopy in pre-operative evaluation of basal cell carcinoma. *Anais Brasileiros de Dermatologia*, 89(5):828–831, September 2014.
- [50] S. Deepak and P. M. Ameer. Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine*, 111:103345, aug 2019.
- [51] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2009.
- [52] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. Solving the multiple instance problem with axis-parallel rectangles. *Artificial intelligence*, 89(1-2):31–71, 1997.
- [53] Gabriella Campos do Carmo and Marcia Ramos e Silva. Dermoscopy: basic concepts. *International Journal of Dermatology*, 47(7):712–719, July 2008.
- [54] David L Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. In *AMS CONFERENCE ON MATH CHALLENGES OF THE 21ST CENTURY*, 2000.
- [55] Gary Doran and Soumya Ray. A theoretical and empirical analysis of support vector machine methods for multiple-instance classification. *Machine Learning*, 97(1-2):79–102, December 2013.

- [56] Jefersson Alex dos Santos, Philippe-Henri Gosselin, Sylvie Philipp-Foliguet, Ricardo da S. Torres, and Alexandre Xavier Falao. Multiscale classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10):3764–3775, October 2012.
- [57] F. Le Duff, E. Long-Mira, J.-P. Lacour, and P. Bahadoran. Mélanome de dubreuilh en microscopie confocale. *Annales de Dermatologie et de Vénérologie*, 141(8-9):560–564, August 2014.
- [58] Murat Dundar, Balaji Krishnapuram, R B Rao, and Glenn M Fung. Multiple instance learning for computer aided diagnosis. In *Advances in neural information processing systems*, pages 425–432, 2007.
- [59] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, January 2017.
- [60] Xiaoyu Fan, Muzhi Dai, Chenxi Liu, Fan Wu, Xiangda Yan, Ye Feng, Yongqiang Feng, and Baiquan Su. Effect of image noise on the classification of skin lesions using deep convolutional neural networks. *Tsinghua Science and Technology*, 25(3):425–434, June 2020.
- [61] Aaron S. Farberg and Darrell S. Rigel. The importance of early recognition of skin cancer. *Dermatologic Clinics*, 35(4):xv–xvi, October 2017.
- [62] N Fatemi-Ghomi, P L Palmer, and M Petrou. Performance Evaluation of Texture Segmentation Algorithms based on Wavelets, 1996.
- [63] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2008.
- [64] John R. Ferraro, Kazuo Nakamoto, and Chris W. Brown. Instrumentation and experimental techniques. In *Introductory Raman Spectroscopy*, pages 95–146. Elsevier, 2003.
- [65] Christina Fitzmaurice and al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015. *JAMA Oncology*, 3(4):524, April 2017.
- [66] Thomas B. Fitzpatrick. The validity and practicality of sun-reactive skin types I through VI. *Archives of Dermatology*, 124(6):869, June 1988.
- [67] James Foulds and Eibe Frank. A review of multi-instance learning assumptions. *The Knowledge Engineering Review*, 25(1):1–25, March 2010.
- [68] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2):337–407, April 2000.
- [69] Jerome H. Friedman. machine. *The Annals of Statistics*, 29(5):1189–1232, October 2001.
- [70] R. J. Friedman, D. S. Rigel, and A. W. Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: A Cancer Journal for Clinicians*, 35(3):130–151, May 1985.

- [71] Mingchen Gao, Ulas Bagci, Le Lu, Aaron Wu, Mario Buty, Hoo-Chang Shin, Holger Roth, Georgios Z. Papadakis, Adrien Depeursinge, Ronald M. Summers, Ziyue Xu, and Daniel J. Mollura. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(1):1–6, June 2016.
- [72] Thomas Gärtner, Peter A Flach, Adam Kowalczyk, and Alex J Smola. Multi-Instance Kernels. In *Proceedings of the Nineteenth International Conference on Machine Learning, ICML '02*, pages 179–186, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- [73] Alex M. Glazer, Darrell S. Rigel, Richard R. Winkelmann, and Aaron S. Farberg. Clinical diagnosis of skin cancer. *Dermatologic Clinics*, 35(4):409–416, October 2017.
- [74] K.E. Gobel. Magnetic resonance microscopy of skin. In *Imaging in Dermatology*, pages 415–423. Elsevier, 2016.
- [75] Pascale Guitera, Giovanni Pellacani, Kerry A. Crotty, Richard A. Scolyer, Ling-Xi L. Li, Sara Bassoli, Marco Vinceti, Harold Rabinovitz, Caterina Longo, and Scott W. Menzies. The impact of in vivo reflectance confocal microscopy on the diagnostic accuracy of lentigo maligna and equivocal pigmented and nonpigmented macules of the face. *Journal of Investigative Dermatology*, 130(8):2080–2091, August 2010.
- [76] Xinjian Guo, Yilong Yin, Cailing Dong, Gongping Yang, and Guangtong Zhou. On the class imbalance problem. In *2008 Fourth International Conference on Natural Computation*. IEEE, 2008.
- [77] Gery P. Guy, Dawn M. Holman, and Meg Watson. The important role of schools in the prevention of skin cancer. *JAMA Dermatology*, 152(10):1083, October 2016.
- [78] Abdelghafour Halimi, Hadj Batatia, Jimmy Le Digabel, Gwendal Josse, and Jean-Yves Tournet. Statistical modeling and classification of reflectance confocal microscopy images. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, December 2017.
- [79] Abdelghafour Halimi, Hadj Batatia, Jimmy Le Digabel, Gwendal Josse, and Jean Yves Tournet. Wavelet-based statistical classification of skin images acquired with reflectance confocal microscopy. *Biomedical Optics Express*, 8(12):5450, November 2017.
- [80] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973.
- [81] Hedi Harzallah, Frederic Jurie, and Cordelia Schmid. Combining efficient object localization and image classification. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, September 2009.
- [82] Haibo He and E.A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, September 2009.
- [83] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

- [84] Abdulkader Helwan and Dilber Uzun Ozsahin. Sliding window based machine learning system for the left ventricle localization in MR cardiac images. *Applied Computational Intelligence and Soft Computing*, 2017:1–9, 2017.
- [85] J. Himberg, J. Mantyjarvi, and P. Korpipaa. Using PCA and ICA for exploratory data analysis in situation awareness. In *Conference Documentation International Conference on Multisensor Fusion and Integration for Intelligent Systems. MFI 2001 (Cat. No.01TH8590)*. VDI/VDE Soc. Meas. & Autom. Control, 2001.
- [86] Tin Kam Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, 1994.
- [87] Zhang Hong and Zhang Xuanbing. Texture feature extraction based on wavelet transform. In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*. IEEE, 2010.
- [88] Le Hou, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, James E. Davis, and Joel H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.
- [89] Wenhao Huang, Guojie Song, Man Li, Weisong Hu, and Kunqing Xie. Adaptive Weight Optimization for Classification of Imbalanced Data. In Changyin Sun, Fang Fang, Zhi-Hua Zhou, Wankou Yang, and Zhi-Yong Liu, editors, *Intelligence Science and Big Data Engineering*, pages 546–553, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [90] Minyoung Huh, Pulkit Agrawal, and Alexei A Efros. What makes ImageNet good for transfer learning? *arXiv preprint arXiv:1608.08614*, 2016.
- [91] Jay Im and Narasimhan Rajaram. Optical molecular imaging and spectroscopy of oxygenation and metabolism in tumors. *IEEE Journal of Selected Topics in Quantum Electronics*, 22(3):78–87, May 2016.
- [92] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer New York, 2013.
- [93] Xi Jia and Linlin Shen. Skin lesion classification using class activation map. *arXiv preprint arXiv:1703.01053*, 2017.
- [94] Piotr Juszczak, D Tax, and Robert PW Duin. Feature scaling in support vector data description. In *Proc. asc*, pages 95–102. Citeseer, 2002.
- [95] Narendra Kumar Kamila. *Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing*. Information Science Reference - Imprint of: IGI Publishing, Hershey, PA, 2015.
- [96] C. Karimkhani, A.C. Green, T. Nijsten, M.A. Weinstock, R.P. Dellavalle, M. Naghavi, and C. Fitzmaurice. The global burden of melanoma: results from the global burden of disease study 2015. *British Journal of Dermatology*, 177(1):134–140, June 2017.
- [97] Brian Katz and Harold S. Rabinovitz. INTRODUCTION TO DERMOSCOPY. *Dermatologic Clinics*, 19(2):221–258, April 2001.

- [98] Ji-Hyun Kim. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, 53(11):3735–3745, September 2009.
- [99] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.
- [100] S Koller, M Wiltgen, V Ahlgrimm-Siess, W Weger, R Hofmann-Wellenhof, E Richtig, J Smolle, and A Gerger. In vivo reflectance confocal microscopy: automated diagnostic image analysis of melanocytic skin tumours. *Journal of the European Academy of Dermatology and Venereology*, 25(5):554–558, August 2010.
- [101] Nikiforos Kollias and Georgios N. Stamatas. Optical non-invasive approaches to diagnosis of skin diseases. *Journal of Investigative Dermatology Symposium Proceedings*, 7(1):64–75, December 2002.
- [102] I. Kononenko. Bayesian neural networks. *Biological Cybernetics*, 61(5):361–370, September 1989.
- [103] S B Kotsiantis. Supervised Machine Learning: A Review of Classification Techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3–24, NLD, 2007. IOS Press.
- [104] K. Krishneswari and S. Arumugam. Multimodal Biometrics using Feature Fusion. *Journal of Computer Science*, 8(3 SE - Research Article), jan 2012.
- [105] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. CIFAR-10 (Canadian Institute for Advanced Research).
- [106] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F Pereira, C J C Burges, L Bottou, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [107] Meelis Kull, Telmo M. Silva Filho, and Peter Flach. Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration. *Electronic Journal of Statistics*, 11(2):5052–5080, 2017.
- [108] Meelis Kull, Telmo Silva Filho, and Peter Flach. Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 623–631, Fort Lauderdale, FL, USA, 2017. PMLR.
- [109] Aimilios Lallas, Giuseppe Argenziano, Enrico Zendri, Elvira Moscarella, Caterina Longo, Laura Grenzi, Giovanni Pellacani, and Iris Zalaudek. Update on non-melanoma skin cancer and the value of dermoscopy in its diagnosis and treatment monitoring. *Expert Review of Anticancer Therapy*, 13(5):541–558, May 2013.
- [110] L. Lam and S.Y. Suen. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 27(5):553–568, 1997.

- [111] Frédérique-Anne Le Gal, Laurence Toutous-Trellu, Gürkan Kaya, and Denis Salomon. Lentigo maligna : un mélanome particulier. *Revue Médicale Suisse*, 7:765–71, 2011.
- [112] Yann LeCun and Corinna Cortes. {MNIST} handwritten digit database. 2010.
- [113] G Lee, R Gommers, F Wasilewski, K Wohlfahrt, A O’Leary, H Nahrstaedt, and Contributors. PyWavelets - Wavelet Transforms in Python, 2006.
- [114] Guillaume Lemaître, Fernando Nogueira, and Christos K Aridas. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.
- [115] Teng Li, Zhijun Meng, Bingbing Ni, Jianbing Shen, and Meng Wang. Robust geometric ℓ_p -norm feature pooling for image classification and action recognition. *Image and Vision Computing*, 55:64–76, November 2016.
- [116] Yuexiang Li and Linlin Shen. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 18(2):556, February 2018.
- [117] Chad A. Lieber, Shovan K. Majumder, Darrel L. Ellis, D. Dean Billheimer, and Anita Mahadevan-Jansen. In vivo nonmelanoma skin cancer diagnosis using raman microspectroscopy. *Lasers in Surgery and Medicine*, 40(7):461–467, September 2008.
- [118] Liang Lim, Brandon Nichols, Michael R. Migden, Narasimhan Rajaram, Jason S. Reichenberg, Mia K. Markey, Merrick I. Ross, and James W. Tunnell. Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis. *Journal of Biomedical Optics*, 19(11):117003, November 2014.
- [119] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghahfoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, December 2017.
- [120] Ruiming Liu, Erqi Liu, Jie Yang, Ming Li, and Fanglin Wang. Optimizing the hyperparameters for SVM by combining evolution strategies with a grid search. In *Intelligent Control and Automation*, pages 712–721. Springer Berlin Heidelberg, 2006.
- [121] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J. Fulham, and ADNI. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, April 2015.
- [122] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.
- [123] S. Livens. Wavelets for texture analysis, an overview. In *6th International Conference on Image Processing and its Applications*. IEE, 1997.
- [124] William Lotter, Greg Sorensen, and David Cox. A Multi-scale CNN and Curriculum Learning Strategy for Mammogram Classification. In M Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R S Tavares, Mehdi Moradi, Andrew Bradley, Hayit Greenspan, João Paulo Papa, Anant Madabhushi, Jacinto C Nascimento, Jaime S Cardoso, Vasileios Belagiannis, and Zhi Lu, editors, *Deep Learning in Medical*

- Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 169–177, Cham, 2017. Springer International Publishing.
- [125] H. Lui, J. Zhao, D. McLean, and H. Zeng. Real-time raman spectroscopy for in vivo skin cancer diagnosis. *Cancer Research*, 72(10):2491–2500, March 2012.
- [126] Eugen Lupu and Petre G Pop. Multimodal biometric systems overview. *ACTA TECHNICA NAPOCENSIS Electronics and Telecommunications*, 49(3):39–44, 2008.
- [127] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. 9:2579–2605, 2008.
- [128] Rona M. MacKie. Cutaneous microscopy in vivo as an aid to preoperative assessment of pigmented lesions of the skin. *British Journal of Plastic Surgery*, 25:123–129, 1972.
- [129] Rona McLeod MacKie. *An illustrated guide to the recognition of early malignant melanoma*. University of Glasgow, 1986.
- [130] Rupananda Mallia, Shiny Sara Thomas, Anitha Mathews, Rejnish Kumar, Paul Sebastian, Jayaprakash Madhavan, and Narayanan Subhash. Oxygenated hemoglobin diffuse reflectance ratio for in vivo detection of oral pre-cancer. *Journal of Biomedical Optics*, 13(4):041306, 2008.
- [131] J. Margeta, A. Criminisi, R. Cabrera Lozoya, D.C. Lee, and N. Ayache. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 5(5):339–349, August 2015.
- [132] Oded Maron. Learning from ambiguity. 1998.
- [133] J. A. McGrath and J. Uitto. Anatomy and organization of human skin. In *Rook's Textbook of Dermatology*, pages 1–53. Wiley-Blackwell, May 2010.
- [134] Harry McGurk and John MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, December 1976.
- [135] Jeffrey K McKenna, Scott R Florell, Glenn D Goldman, and Glen M Bowen. Lentigo maligna/lentigo maligna melanoma: Current state of diagnosis and treatment. *Dermatologic Surgery*, 32(4):493–504, April 2006.
- [136] Simon M. Mueller, Samuel Hogg, Jannis M. Mueller, Shane McKie, Peter Itin, Julia Reinhardt, Christopher E.M. Griffiths, and Christine Elise Kleyn. Functional magnetic resonance imaging in dermatology: The skin, the brain and the invisible. *Experimental Dermatology*, 26(10):845–853, April 2017.
- [137] Allan H. Murphy and Robert L. Winkler. Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics*, 26:41, 1977.
- [138] Bruce W. Murphy, Rebecca J. Webster, Berwin A. Turlach, Christopher J. Quirk, Christopher D. Clay, Peter J. Heenan, and David D. Sampson. Toward the discrimination of early melanoma from common and dysplastic nevus using fiber optic diffuse reflectance spectroscopy. *Journal of Biomedical Optics*, 10(6):064020, 2005.

- [139] Mahdi Pakdaman Naeini, Gregory F Cooper, and Milos Hauskrecht. Obtaining Well Calibrated Probabilities Using Bayesian Binning. *Proceedings of the ... AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence*, 2015:2901–2907, jan 2015.
- [140] Loris Nanni, Stefano Ghidoni, and Sheryl Brahnam. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172, November 2017.
- [141] Deevya L. Narayanan, Rao N. Saladi, and Joshua L. Fox. Review: Ultraviolet radiation and skin cancer. *International Journal of Dermatology*, 49(9):978–986, August 2010.
- [142] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal Deep Learning. *Proceedings of The 28th International Conference on Machine Learning (ICML)*, pages 689–696, 2011.
- [143] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning - ICML '05*. ACM Press, 2005.
- [144] Khin Nyein Nyein Hlaing. First Order Statistics and Glcm Based Feature Extraction for Recognition of Myanmar Paper Currency. *The IIER International Conference*, (2011):1–6, 2015.
- [145] Joseph O Ogutu, Hans-Peter Piepho, and Torben Schulz-Streeck. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proceedings*, 5(S3), May 2011.
- [146] Ozlem Ozbacivan, Sevgi Akarsu, Necla Ikiz, Fatma Semiz, and Emel Fetil. Dermoscopic Differentiation of Facial Lentigo Maligna from Pigmented Actinic Keratosis and Solar Lentiginosities. *Acta dermatovenerologica Croatica : ADC*, 27(3):146–152, sep 2019.
- [147] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010.
- [148] Beomhee Park, Yongwon Cho, Gaeun Lee, Sang Min Lee, Young-Hoon Cho, Eun Sol Lee, Kyung Hee Lee, Joon Beom Seo, and Namkug Kim. A curriculum learning strategy to enhance the accuracy of classification of various lesions in chest-PA x-ray screening for pulmonary abnormalities. *Scientific Reports*, 9(1), October 2019.
- [149] Sameena Pathan, K. Gopalakrishna Prabhu, and P.C. Siddalingaswamy. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. *Biomedical Signal Processing and Control*, 39:237–262, January 2018.
- [150] Marco Pedersoli, Jordi González, Andrew D. Bagdanov, and Xavier Roca. Efficient discriminative multiresolution cascade for real-time human detection applications. *Pattern Recognition Letters*, 32(13):1581–1587, October 2011.
- [151] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and Others. Scikit-learn: Machine Learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [152] Giovanni Pellacani, Pascale Guitera, Caterina Longo, Michelle Avramidis, Stefania Seidenari, and Scott Menzies. The impact of in vivo reflectance confocal microscopy for the diagnostic accuracy of melanoma and equivocal melanocytic lesions. *Journal of Investigative Dermatology*, 127(12):2759–2765, December 2007.

- [153] Eric Persoon and King-Sun Fu. Shape discrimination using fourier descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(3):388–397, May 1986.
- [154] Clifton Phua, Daminda Alahakoon, and Vincent Lee. Minority report in fraud detection. *ACM SIGKDD Explorations Newsletter*, 6(1):50–59, June 2004.
- [155] Arnaud Pillon, Bruno Gomes, Isabelle Vandenberghe, Valérie Cartron, Patrick Cèbe, Jean-Christophe Blanchet, Vincent Sibaud, Nicolas Guilbaud, Laurent Audoly, Laurence Lamant, and Anna Kruczynski. Actinic keratosis modelling in mice: A translational study. *PLOS ONE*, 12(6):e0179991, June 2017.
- [156] J Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In *Advances in Large Margin Classifiers*, 2000.
- [157] P. Pralong, E. Bathelier, S. Dalle, N. Poulalhon, S. Debarbieux, and L. Thomas. Dermoscopy of lentigo maligna melanoma: report of 125 cases. *British Journal of Dermatology*, 167(2):280–287, June 2012.
- [158] Ronaldo C. Prati, Gustavo E.A.P.A. Batista, and Maria Carolina Monard. Data mining with unbalanced class distributions: Concepts and methods. *Proceedings of the 4th Indian International Conference on Artificial Intelligence, IICAI 2009*, pages 359–376, 2009.
- [159] Steven R. Quartz and Terrence J. Sejnowski. The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20(4):537–556, December 1997.
- [160] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, March 1986.
- [161] Agus Budi Raharjo. *Reliability in ensemble learning and learning from crowds*. PhD thesis, 2019. Thèse de doctorat dirigée par Quafafou, Mohamed Informatique Aix-Marseille 2019.
- [162] Christian Raulin and Syrus Karsai, editors. *Lasertherapie der Haut*. Springer Berlin Heidelberg, 2013.
- [163] Jeff Reback, Wes McKinney, Jbrockmendel, Joris Van Den Bossche, Tom Augspurger, Phillip Cloud, Gyoung, Sinhrks, Simon Hawkins, Adam Klein, Matthew Roeschke, Jeff Tratner, Terji Petersen, Chang She, William Ayd, MomIsBestFriend, Marc Garcia, Jeremy Schendel, Andy Hayden, Daniel Saxton, Vytautas Jancauskas, Ali McMaster, Pietro Battiston, Skipper Seabold, Chris-B1, H-Vetinari, Kaiqi Dong, Stephan Hoyer, Wouter Overmeire, and Martin Winkel. pandas-dev/pandas: Pandas 1.1.3, 2020.
- [164] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.
- [165] Xiaofeng Ren. Multi-scale Improves Boundary Detection in Natural Images. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision – ECCV 2008*, pages 533–545, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
- [166] Christian Robert. Machine learning, a probabilistic perspective. *CHANCE*, 27(2):62–63, April 2014.
- [167] Jane Sandby-Møller, Thomas Poulsen, and Hans Christian Wulf. Epidermal thickness at different body sites: Relationship to age, gender, pigmentation, blood content, skin type and smoking habits. *Acta Dermato-Venereologica*, 83(6):410–413, November 2003.

- [168] Johannes Schleusener, Patrycja Gluszczyńska, Carina Reble, Ingo Gersonde, Jürgen Helfmann, Joachim W. Fluhr, Jürgen Lademann, Joachim Röwert-Huber, Alexa Patzelt, and Martina C. Meinke. In vivo study for the discrimination of cancerous and normal skin using fibre probe-based raman spectroscopy. *Experimental Dermatology*, 24(10):767–772, July 2015.
- [169] Burr Settles, Mark Craven, and Soumya Ray. Multiple-instance active learning. In *Advances in neural information processing systems*, pages 1289–1296, 2008.
- [170] Colin Sheppard. Scanning and confocal microscopy. In *Encyclopedia of Optical and Photonic Engineering, Second Edition*, pages 1–18. CRC Press, September 2015.
- [171] Jeff Shrago and Mark H Johnson. Dynamic plasticity influences the emergence of function in a simple cortical array. *Neural Networks*, 9(7):1119–1129, October 1996.
- [172] Fabricio L. Silveira, Marcos T.T. Pacheco, Benito Bodanese, Carlos A. Pasqualucci, Renato A. Zângaro, and Landulfo Silveira. Discrimination of non-melanoma skin lesions from non-tumor human skin tissues in vivo using raman spectroscopy and multivariate statistics. *Lasers in Surgery and Medicine*, 47(1):6–16, January 2015.
- [173] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. pages 1–14, 2014.
- [174] Ning Situ, Xiaojing Yuan, Ji Chen, and George Zouridakis. Malignant melanoma detection by bag-of-features classification. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, August 2008.
- [175] Fethi Smach, Cedric Lemaître, Jean-Paul Gauthier, Johel Miteran, and Mohamed Atri. Generalized fourier descriptors with applications to objects recognition in SVM context. *Journal of Mathematical Imaging and Vision*, 30(1):43–71, December 2007.
- [176] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In Abdul Sattar and Byeong-ho Kang, editors, *AI 2006: Advances in Artificial Intelligence*, pages 1015–1021, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [177] Sidharth Sonthalia and Feroze Kaliyadan. Dermoscopy Overview and Extradagnostic Applications. 2019.
- [178] Golam Sorwar, Ajith Abraham, and Laurence S Dooley. Texture classification based on DCT and soft computing. In *Meeting the Grand Challenge: Machines that Serve People: Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, volume 2, page 1, United States of America, 2001. IEEE, Institute of Electrical and Electronics Engineers.
- [179] G N Srinivasan and G Shobha. Statistical Texture Analysis. In *International Journal of Computer and Information Engineering*, volume 2, 2008.
- [180] M. Stante, V. Giorgi, I. Stanganelli, B. Alfaioli, and P. Carli. Dermoscopy for early detection of facial lentigo maligna. *British Journal of Dermatology*, 152(2):361–364, February 2005.
- [181] I Stephen. Perceptron-based learning algorithms. *IEEE Transactions on neural networks*, 50(2):179, 1990.
- [182] Nicholas Stewart, Adrian C Lim, Patricia M Lowe, and Greg Goodman. Lasers and laser-like devices: Part one. *Australasian Journal of Dermatology*, 54(3):173–183, April 2013.

- [183] P.J. Sudharshan, Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117:103–111, March 2019.
- [184] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, pages 4278–4284, 2017.
- [185] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [186] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.
- [187] Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, May 2016.
- [188] Jinhui Tang, Haojie Li, Guo-Jun Qi, and Tat-Seng Chua. Image annotation by graph-based inference with integrated multiple/single instance representations. *IEEE Transactions on Multimedia*, 12(2):131–141, 2009.
- [189] Qiling Tang, Yangyang Liu, and Haihua Liu. Medical image classification via multiscale representation learning. *Artificial Intelligence in Medicine*, 79:71–78, June 2017.
- [190] Luke Taylor and Geoff Nitschke. Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, November 2018.
- [191] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme. Cost-sensitive learning methods for imbalanced data. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2010.
- [192] Anne Thuret. L ’ épidémiologie du mélanome cutané en France et en Europe / Box – Epidemiology of cutaneous melanoma in France and in Europe. *Beh*, 18-19:213–214, 2012.
- [193] Kai Ming Ting. An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering*, 14(3):659–665, May 2002.
- [194] Ivan Tomek. Two modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(11):769–772, November 1976.
- [195] Fumiaki Tomita and Saburo Tsuji. Statistical texture analysis. In *Computer Analysis of Visual Textures*, pages 13–36. Springer US, 1990.
- [196] B Tortora, G;Derrickson. *Principles of Anatomy and Physiology*. 2012.
- [197] Ioannis Tsamardinos, Amin Rakhshani, and Vincenzo Lagani. Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization. In *Artificial Intelligence: Methods and Applications*, pages 1–14. Springer International Publishing, 2014.

- [198] Bruna Tuma, Sergio Yamada, Álvaro Nagib Atallah, Fernanda Mendes Araujo, and Sergio Henrique Hirata. Dermoscopy of black skin: A cross-sectional study of clinical and dermoscopic features of melanocytic lesions in individuals with type v/VI skin compared to those with type i/II skin. *Journal of the American Academy of Dermatology*, 73(1):114–119, July 2015.
- [199] A. M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, October 1950.
- [200] Alexia Tzalavra, Kalliopi Dalakleidi, Evangelia I Zacharaki, Nikolaos Tsiaparas, Fotios Constantinidis, Nikos Paragios, and Konstantina S Nikita. Comparison of Multi-resolution Analysis Patterns for Texture Classification of Breast Tumors Based on DCE-MRI. In Li Wang, Ehsan Adeli, Qian Wang, Yinghuan Shi, and Heung-Il Suk, editors, *Machine Learning in Medical Imaging*, pages 296–304, Cham, 2016. Springer International Publishing.
- [201] Ahsan Ahmad Ursani, Kidiyo Kpalma, and Joseph Ronsin. Texture features based on fourier transform and gabor filters: an empirical comparison. In *2007 International Conference on Machine Vision*. IEEE, December 2007.
- [202] Shofwatul Uyun, Sri Hartati, Agus Harjoko, and Subanar. Selection Mammogram Texture Descriptors Based on Statistics Properties Backpropagation Structure. *International Journal of Computer Science and Informatioan Security*, 11(5):1–5, 2013.
- [203] Nanne van Noord and Eric Postma. Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognition*, 61:583–592, January 2017.
- [204] Pauli Virtanen and al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, February 2020.
- [205] Theo Vos and al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The Lancet*, 388(10053):1545–1602, October 2016.
- [206] Junping Wang, Quanshi Chen, and Yong Chen. RBF Kernel Based Support Vector Machine with Universal Approximation and Its Application. In Fu-Liang Yin, Jun Wang, and Chengan Guo, editors, *Advances in Neural Networks – ISNN 2004*, pages 512–517, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- [207] Jason W. Wei, Laura J. Tafe, Yevgeniy A. Linnik, Louis J. Vaickus, Naofumi Tomita, and Saeed Hassanpour. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Scientific Reports*, 9(1), March 2019.
- [208] Dennis L. Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2(3):408–421, July 1972.
- [209] Marco Wiltgen, A. Gerger, C. Wagner, and J. Smolle. Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumors. *Methods of Information in Medicine*, 47(1):14–25, 2008.
- [210] C. Wolley and M. Quafafou. Learning from multiple annotators: When data is hard and annotators are unreliable. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 514–521, Los Alamitos, CA, USA, dec 2012. IEEE Computer Society.

- [211] Chirine Wolley. *Apprentissage supervisé à partir des multiples annotateurs incertains*. PhD thesis, 2014. Thèse de doctorat dirigée par Quafafou, Mohamed Informatique Aix-Marseille 2014.
- [212] Sebastien C. Wong, Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. Understanding data augmentation for classification: When to warp? In *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, November 2016.
- [213] Zhenyu Wu, Wenfang Lin, and Yang Ji. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *IEEE Access*, 6:8394–8402, 2018.
- [214] Yi Xiao, Felipe Codevilla, Akhil Gurram, Onay Urfalioglu, and Antonio M. López. Multi-modal End-to-End Autonomous Driving. (201808390010):1–15, 2019.
- [215] Yutong Xie, Jianpeng Zhang, and Yong Xia. A Multi-Level Deep Ensemble Model for Skin Lesion Classification in Dermoscopy Images. *arXiv preprint arXiv:1807.08488*, 2018.
- [216] Xin Xu and Eibe Frank. Logistic regression and boosting for labeled bags of instances. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 272–281. Springer, 2004.
- [217] Yan Xu, Zhipeng Jia, Liang-Bo Wang, Yuqing Ai, Fang Zhang, Maode Lai, and Eric I-Chao Chang. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics*, 18(1), May 2017.
- [218] Marjorie F. Yang, Valery V. Tuchin, and Anna N. Yaroslavsky. Principles of light-skin interactions. In *Light-Based Therapies for Skin of Color*, pages 1–44. Springer London, 2009.
- [219] Bianca Zadrozny and Charles Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 694–699, 2002.
- [220] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
- [221] Alexander Zemtsov and Lisa Dixon. Magnetic Resonance in Dermatology. *Archives of Dermatology*, 129(2):215–218, 1993.
- [222] Harry Zhang. The Optimality of Naive Bayes. In Valerie Barr and Zdravko Markov, editors, *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*. AAAI Press, 2004.
- [223] Yin Zhang, Rong Jin, and Zhi-Hua Zhou. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52, August 2010.
- [224] B Zhou, A Khosla, Lapedriza. A., A Oliva, and A Torralba. Learning Deep Features for Discriminative Localization. *CVPR*, 2016.
- [225] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2018.
- [226] Xiaojin Zhu, Andrew B Goldberg, Ronald Brachman, and Thomas Dietterich. *Introduction to Semi-Supervised Learning*. Morgan and Claypool Publishers, 2009.

LISTE DES FIGURES

i.1	Graphique représentant l'évolution du nombre de dermatologues en France entre les années 1999 et 2017 ²	3
i.2	Graphique représentant l'évolution du nombre de publications recensées par le moteur de recherche Google Scholar ³ autour des termes de « skin computer diagnosis » entre les années 1990 et 2020.	4
i.3	Schéma de représentation du processus de réduction de l'indécision, également surnommée zone grise. Ce travail s'inspire du processus cognitif des dermatologues et consiste à réduire l'indécision propre à chacune des modalités cliniques par l'ajout de nouvelles modalités d'imagerie.	5
i.4	Schéma de représentation des divers domaines impliqués dans cette thématique de recherche. Ce travail est au confluent de connaissances de la peau, des modalités d'imagerie et des domaines de l'intelligence artificielle.	5
1.1	Illustration de la peau et de ses diverses couches ¹	10
1.2	Illustration de l'épiderme reprenant ses divers composés cellulaires ainsi que les diverses couches formées par un processus de différenciation ²	11
1.3	Illustration de la Jonction Dermo-Épidermique ³ . Cette jonction dont le rôle est d'assurer une cohésion entre l'épiderme et le derme peut être subdivisée en deux sous couches : la lame basale et réticulaire.	12
1.4	Illustration de représentation du derme ⁴ . Le derme est subdivisé en deux couches aux propriétés distinctes : le derme papillaire et réticulaire.	13
1.5	Échelle de Fitzpatrick associant type de peau et propriétés associées [66].	14
1.6	Illustration de l'indice de Clark [42], défini par plusieurs niveaux de progression ⁶	17
1.7	Illustration par Jean-Luc Perrot des stades d'évolution d'un LM . Les points blancs permettent de représenter les follicules pileux.	18
1.8	Exemple d'une pathologie de LM perçue à l'aide de différents dispositifs médicaux. En a), un aperçu par photographie clinique ; En b), un aperçu par dermatoscopie ; En c), un aperçu par Microscopie confocale par réflectance. En d) et e), les images fournissent un aperçu par histologie.	20
2.1	Schéma représentant les principaux mécanismes d'interaction qui peuvent survenir suite à l'interaction de la lumière avec la peau. Les modalités optiques tentent d'exploiter ces divers phénomènes pour extraire une information sur les tissus analysés.	22

- 2.2 Schéma représentatif de la profondeur de pénétration (en mm) de la peau en fonction de la longueur d'onde de la lumière [11]. Il s'agit d'un ordre de grandeur moyen constaté chez l'être humain, les propriétés de la peau variant d'un emplacement à un autre. Aucune information relative aux types de peaux étudiés, selon l'échelle de Fitzpatrick par exemple, n'a été mentionnée. 23
- 2.3 Représentation des coefficients d'absorption en fonction de la longueur d'onde pour les principaux composants chromophores présents dans la peau [162]. 24
- 2.4 Exemple du Body-Mapper, un appareil de photographie clinique corps entier, proposé par Pixience. À gauche, l'appareil d'acquisition ; À droite, son interface logicielle permettant d'acquérir les diverses parties du corps. 26
- 2.5 Principe de la dermatoscopie à lumière polarisée. Un premier filtre polarisé est disposé après la source de lumière, puis un second filtre dont la polarisation est perpendiculaire au premier est disposé juste avant la fenêtre d'observation [177]. 27
- 2.6 Exemple du dispositif de dermatoscopie *C-Cube* proposé par la société *Pixience*. 28
- 2.7 Exemple du dispositif de dermatoscopie *DL3N* proposé par la société *DermLite*. À gauche, le dispositif seul ; Au milieu, ce même dispositif adapté à un appareil photo standard ; À droite, le dispositif adapté à une caméra de téléphone portable. 28
- 2.8 Schéma de fonctionnement du **MCR** par Marvin Minsky. Une source de lumière est émise et focalisée en un point spécifique de la peau puis la lumière est réfléchiée et reçue par une caméra au travers d'un sténopé (trou de faible diamètre). La lumière réfléchiée en dehors du plan focal défini par l'utilisateur est ainsi arrêtée par ce sténopé. 29
- 2.9 Schéma de fonctionnement de la **TCO**. Une source de lumière génère un unique faisceau divisé en deux faisceaux identiques. L'un de ces faisceaux résultants sert de signal de référence (en bleu), tandis que le second sert à l'analyse de l'échantillon (en vert). Par la suite, ces deux faisceaux sont recombinaés à l'entrée du récepteur. 30
- 2.10 Schéma représentant le fonctionnement d'un spectroscope par réflectance. Une lumière de référence est générée puis acheminée jusqu'à la zone cible. Cette lumière transite le long d'une ou plusieurs fibres optiques. Puis le phénomène de réflectance est capté et envoyé au spectromètre par le biais d'une ou plusieurs fibres. Le signal reçu est ensuite décomposé puis envoyé à la surface d'un capteur numérique. 31
- 2.11 Schéma de fonctionnement d'une caméra multispectrale. Ces dispositifs peuvent adapter la longueur d'onde de la source à partir d'éclairage **DEL**, de lasers ou de filtres (Filtre 1), ou capter différentes longueurs d'onde à partir de la lumière réémise (Filtre 2). 32
- 2.12 Schéma reprenant les divers effets résultant de l'interaction entre la lumière et la matière⁶. La diffusion Raman se distingue par un décalage de longueur d'onde supérieur (Stokes) ou inférieur (anti-Stokes). 33
- 2.13 Schéma du principe de fonctionnement de l'**IRM** en quatre étapes. Le champ magnétique constant permet de contraindre les noyaux d'hydrogène, puis l'excitation par radio-fréquence permet d'observer par la suite leur temps de relaxation. 34

2.14	Exemple de la sonde à ultrasons Butterfly iQ ⁷ , proposant une solution nomade et connectée.	35
3.1	Représentation des relations entre <i>intelligence artificielle</i> , <i>apprentissage automatique</i> et <i>apprentissage profond</i>	38
3.2	Schéma de décomposition des approches par apprentissage automatique.	39
3.3	Schéma de l'approche dite supervisée. La phase d'entraînement permet de déterminer les relations entre données et attentes, sur des données connues dites d'apprentissage. La phase de prédiction réutilise les relations déterminées lors de la phase d'entraînement.	40
3.4	Exemple de données non associées à des annotations préalables, et pour lesquelles des approches non supervisées peuvent être sollicitées afin de découvrir des groupes de données ou <i>clusters</i> . Le nombre de groupes le plus souvent défini par l'utilisateur avant traitement, influe fortement la vision du problème.	43
3.5	Exemple fréquemment employé pour démontrer l'intérêt de l'apprentissage semi-supervisé du point de vue du principe de densité. À gauche, une classification obtenue à partir de deux données étiquetées ; À droite, la même situation avec l'ajout de données non étiquetées.	44
3.6	Exemple simplifié d'une structure d'apprentissage profond. Ce modèle se compose de deux couches intermédiaires cachées : la première couche extrait des caractéristiques linéaires simples, tandis que la seconde couche par combinaison avec la première couche conduit à une extraction de caractéristiques plus complexes. Ainsi, la première couche s'apparente à un polynôme de degré 1, tandis que la seconde couche représente des fonctions de degré 2 et ainsi de suite.	46
3.7	Exemple d'attentes liées à un apprentissage par transfert en opposition à un apprentissage classique.	47
3.8	Exemple de dilemme de biais et de variance. À gauche, un cas de sous-apprentissage, ce modèle est une sous approximation du phénomène observé. À droite, un cas de sur-apprentissage, le modèle ne généralise pas suffisamment le phénomène observé.	48
3.9	Schéma représentatif d'une matrice de confusion, qui met en avant une situation binaire.	49
3.10	Schéma représentatif des principales métriques obtenues à partir de la matrice de confusion : la <i>précision</i> , le <i>rappel</i> ou la <i>sensibilité</i> et la <i>spécificité</i>	50
3.11	Schéma de courbe ROC en orange et AUC en gris.	51
3.12	Exemple de recherche sur deux hyperparamètres X_1 et X_2 . À gauche, une recherche sous forme de grille rigide, dans laquelle chaque combinaison est évaluée. À droite, une recherche aléatoire de ces hyperparamètres.	52
3.13	Schéma des principes des deux approches majeures d'évaluation de modèles. À gauche, la méthode de <i>holdout</i> ; À droite, la méthode de <i>validation croisée</i>	53
3.14	Schéma de la validation croisée imbriquée assurant le choix de la meilleure combinaison d'hyperparamètres et assurant une évaluation robuste des performances du modèle.	54

3.15	Schéma des différents niveaux théoriques de fusion sur un schéma multimodal à deux capteurs. Le niveau bas représente le niveau le plus proche des données et des capteurs, le niveau moyen représente le niveau au plus proche de l'extraction de caractéristiques avancées, enfin le niveau haut représente le niveau de prise de décision.	56
4.1	À gauche, répartition en fonction de l'âge et du sexe, tous diagnostics confondus. À droite, répartition entre l'âge et le sexe en tenant compte du diagnostic binaire.	59
4.2	Schéma de représentation du processus type de dermatologie, allant de la modalité la plus incertaine mais également la moins onéreuse, à la modalité la plus précise mais également la plus coûteuse.	59
4.3	Répartition du panel de 21 dermatologues sur l'évaluation des diverses modalités [40].	60
4.4	Schéma organisationnel de la base de données. En haut, la base initiale ; En bas, la base restructurée afin de supporter de nouvelles annotations.	61
4.5	Schéma de représentation de l'interface logicielle mise en place pour permettre les divers accès à la base de données employée dans ce manuscrit. Par cette interface, trois méthodes sont mises à la disposition des expérimentations qui suivent. . . .	62
4.6	Exemple de cas de données cliniques de dermatoscopie par Pralong et al. [157]. En a) et b), exemple de phénomène d'assombrissement sur un cas de LM ; en c), exemple de motifs en forme de cible issue d'un LM ; en d), exemple de structure rhomboïdales rouges sur un cas de LMM	64
4.9	Exemples de cas de données cliniques MCR par Guitera et al. [75]. En a), exemple de cellules atypiques à proximité d'un follicule pileux au sein de la JDE sur un LM ; En b), exemple de cellules nucléées du derme d'un LM . Repère : barre = 50 μm	65
4.7	Exemples de cas de données cliniques MCR en provenance de l'épiderme par Guitera et al. [75]. En a) et b), exemples de motifs en nid d'abeilles (<i>broadened honeycomb pattern</i>), respectivement d'une KS et de peau normale ; en c), exemples de motifs de pavés atypiques (<i>atypical cobblestone pattern</i>) issus d'un LM ; en d), exemples de <i>grandes cellules pagétoïdes</i> propres au LMM . Repère : barre = 50 μm	66
4.8	Exemples de cas de données cliniques MCR par Guitera et al. [75]. En a) et au niveau de la flèche, exemple de papille à frontières non marquées (<i>Nonedge papillae</i>) et de cellules atypiques d'un LM acquise à la jonction JDE ; En b), exemple de papilles à frontières marquées (<i>Edge papillae</i>) d'une pathologie bénigne. En c), exemple de cellules atypiques au niveau de la JDE typique de LM ; En d), exemple de halo noir autour de cellules atypiques de l'épiderme issu d'un LM . Repère : barre = 50 μm	67
ii.1	Courbes ROC issues de l'évaluation des données MCR par les 12 experts [40]. À gauche, le résultat de l'analyse des experts, mené sur la détection d'éléments malins ; À droite, le résultat de l'analyse des experts, mené sur la détection de LM et LMM	72

ii.2	Interface logicielle mise à disposition du spécialiste, afin de procéder aux annotations au niveau des images et des sous-images. À gauche, l'onglet permettant l'annotation au niveau des images ; À droite, l'onglet permettant la création et l'annotation de sous-images.	73
ii.3	Exemple de tissus et annotations, considérés respectivement de gauche à droite comme <i>sain</i> , <i>bénin</i> et <i>malin</i> par le spécialiste.	74
ii.4	Répartition des annotations réalisées par l'expert sur sur les lésions, les images et les sous-images. En bleu foncé, la base initiale de Cinotti et al. [40] ; En bleu clair, les données additionnelles.	74
5.1	Représentation du processus de diagnostic suivi dans le cadre de l'apprentissage automatique pour les images MCR . Ce processus permet une prédiction à trois classes - a), mais également binaire en se focalisant sur la détection de la classe maligne - b).	78
5.2	Représentation de la création d'une matrice sur base de GLCM selon la direction horizontale et dont la valeur de distance de voisinage est de 1. En a), les flèches représentent les diverses relations de voisinage horizontal, caractérisées par une couleur ; En b), ces diverses relations sont recensées dans la matrice de GLCM en employant les mêmes codes de couleurs.	80
5.3	Schéma représentant l'extraction de caractéristiques à partir de l'espace de Fourier. En a), l'utilisation de cercles concentriques depuis l'origine, schématisé en pointillés ; En b), l'extraction selon une direction.	82
5.4	Principe de la décomposition en ondelettes appliquée aux images [123]. La décomposition est réalisée en trois temps. Deux filtrages respectivement passe bas (L) et passe haut (H) sont réalisés selon la direction horizontale, puis selon la direction verticale. Enfin, un sous-échantillonnage avec un coefficient de valeur 2 est appliqué afin d'obtenir une matrice de même taille que l'image d'origine.	83
5.5	Schéma représentant l'apprentissage par transfert par CNN . Le réseau est entraîné sur des données et annotations d'une problématique <i>source</i> . Puis, les paramètres de la partie liée à l'extraction de caractéristiques sont réemployés au sein d'une nouvelle problématique <i>cible</i>	86
5.6	Schéma représentant l'aplatissement et la réduction d'information par <i>Global Pooling</i> au sein de CNN . Le réseau est entraîné sur des données sources, généralement plus complètes. Puis, les paramètres de la partie liée à l'extraction de caractéristiques sont conservés et réemployés au sein d'une nouvelle problématique cible.	87
5.7	Exemple d'extraction de GLCM de second ordre sur deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de LM	88
5.8	Exemple de transformée en ondelettes de Daubechies appliquée à deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de LM	89
5.9	Exemple de transformée en ondelettes de Haar appliquée à deux images typiques. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de LM	90

5.10	Exemple de transformée de Fourier centrée appliquée à deux images typiques, représentation selon le module et la phase. En haut, un cas d'image bénigne ; En bas, une image pathologique typique de LM	91
5.11	Visualisation des caractéristiques obtenues par techniques d'extraction spatiales et projection PCA sur les deux premières composantes.	92
5.12	Visualisation des caractéristiques obtenues par techniques d'extraction fréquentielles et projection PCA sur les deux premières composantes.	93
5.13	Visualisation des caractéristiques obtenues par techniques d'extraction basées sur des réseaux de type CNN pré-entraînés sur la base ImageNet et projection PCA sur les deux premières composantes.	94
5.14	Schéma des deux stratégies principales de balancement de données. À gauche, la stratégie de sous-échantillonnage, dans laquelle des éléments sont aléatoirement choisis jusqu'à réduire l'ensemble des classes au nombre d'éléments de la classe minoritaire ; À droite, la stratégie de sur-échantillonnage, dans laquelle les éléments des classes minoritaires sont dupliqués aléatoirement.	98
5.15	Visualisation des divers lots utilisés pour l'évaluation des processus de classification.	100
5.16	Visualisation de l'influence de la réduction des dimensions, appliquée au deux premières composantes principales, sur un échantillon constitué arbitrairement de 500 images issues des données de MCR . En haut, technique de réduction basée sur le PCA ; En bas, technique de réduction basée sur le LDA . À gauche, un processus procédant à la réduction puis à la normalisation ; À droite, un processus procédant à la normalisation puis à la réduction.	101
5.17	Résultats issus de la classification par les diverses techniques d'extraction et par les modèles de classification mentionnés en couleur. À gauche, les résultats provenant de la normalisation de type <i>Minimum / Maximum</i> ; À droite, les résultats provenant de la normalisation de type <i>Standard</i> . En haut, la représentation basée sur les valeurs moyennes ; En bas, la représentation tenant compte de l'écart-type.	104
5.18	Résultats issus de la classification par l'utilisation de techniques de réduction combinés aux techniques d'extraction par transfert et, par les modèles de classification mentionnés en couleur. Les tracés de couleurs représentent les divers modèles évalués au cours de l'expérience. À gauche, les résultats provenant de la normalisation par <i>Minimum/Maximum</i> ; À droite, les résultats provenant de la normalisation par <i>score Standard</i> . En haut, la représentation basée sur les valeurs moyennes ; En bas, la représentation tenant compte de l'écart-type.	105
5.19	Résultats sous forme de courbes ROC . À gauche, résultats sur l'ensemble des lésions d'étude ; À droite, résultats excluant les lésions malignes différentes des types LM et LMM	106
5.20	Analyse de la variance expliquée cumulée par l'application du PCA . Le nombre de caractéristiques est représenté pour des variances expliquées de 95 % en vert, 99 % en jaune et 99,5 % en orange.	107
6.1	Schéma représentatif de la relation hiérarchique entre les annotations. Les images malignes peuvent ainsi contenir des tissus bénins et sains ; Les images bénignes peuvent contenir des tissus sains ; Les images saines ne contiennent que des tissus sains.	110

- 6.2 Schématisation des deux principaux types de décomposition successive par ondelettes. En a), schéma de décomposition multi-échelle en ondelettes dit **diadique** ; En b), schéma de décomposition multi-échelle en ondelettes dit **pyramidal**. 111
- 6.3 Schéma de représentation du système multi-échelle, sur la base d'une extraction à diverses échelles puis par l'agrégation des caractéristiques avant la procédure de classification. 113
- 6.4 Schéma de représentation du système multi-échelle, sur la base d'extractions et de prédictions à diverses échelles et par agrégation de ces décisions. 114
- 6.5 Schéma de représentation du système de prédiction par fenêtre glissante. Ce schéma en deux temps propose de prédire sur une multitude de sous parties de l'image originale puis de fusionner les diverses prédictions afin de fournir une prédiction globale à l'image. Ainsi, une première phase d'entraînement est réalisée à l'aide des sous-images. Puis, le modèle entraîné permet de prédire au niveau des images par principe de fenêtre glissante pour lequel un second modèle permet de définir quels seuils sont critiques. 117
- 6.6 Schéma de représentation du processus de réglage fin utilisé dans ce travail. Celui-ci se décompose d'une étape d'initialisation dans le but de substituer les couches de classification, puis un pré-entraînement permet d'initialiser les poids avant de procéder à l'entraînement. 120
- 6.7 Exemple d'images correctement classifiées à l'aide du principe de fenêtre glissante. À gauche, résultats issus d'une fenêtre de classification de 250 par 250 pixels ; À droite, résultats issus d'une fenêtre de classification de 500 par 500 pixels. En haut, représentations des décisions par code couleur : zones saines en vert, zones bénignes en jaune et zones malignes en orange ; En bas, représentations des scores par les codes couleur précédents et intensités de transparence variables : zones de forte confiance en opaque et zones de faible confiance en translucide. . . 125
- 6.8 Schéma d'explication du principe de **CAM** employé à but de segmentation des images de **MCR**. Le réseau est entraîné sur la problématique cible, puis *la ou les couches* de classification sont mobilisées dans le but d'obtenir les zones responsables de la décision par pondération des poids de chaque noeud de la classe considérée. 126
- 6.9 Exemples de l'extraction de zones d'activations par principe de **CAM**, avec application d'un seuil arbitraire à 85 % entre la valeur minimum et maximum. En haut, exemples d'images malignes correctement gérées par ce principe ; En bas, exemples d'images malignes partiellement segmentées ou incorrectement gérées par ce principe. 127
- 7.1 Schéma récapitulatif de la problématique de l'inconsistance du nombre d'images constituant les lésions et de l'obtention de décisions et scores pour chacune de ces lésions. 132
- 7.2 Schéma de représentation du système complet de décision supervisée au niveau lésionnel. Le processus d'entraînement se subdivise en deux étapes : une première étape d'entraînement du modèle de prédiction au niveau des instances images et une seconde étape permettant l'agrégation de l'information au niveau des lésions. La prédiction est ensuite réalisée sur les données de test par inférence des paramètres précédemment entraînés. 133

- 7.3 Courbes ROC issues des résultats de la prédiction au niveau des lésions sur l'ensemble des cas, figurant à l'aide d'un trait orange estompé tracé en continu, mais également sur les cas de LM et LMM, figurant à l'aide d'un trait orange tracé en pointillés. À gauche, les courbes issues des prédictions employant la méthode Scores - Moyenne ; À droite, les courbes issues des prédictions du modèle NSK. 141
- 7.4 Courbes ROC issues de la prédiction au niveau des instances sur l'ensemble des cas, figurant à l'aide d'un trait orange estompé tracé en continu, et en ne considérant que les cas de LM et LMM parmi les annotations malignes, figurant à l'aide d'un trait orange tracé en pointillés. Les courbes sont issues du mode de prédiction faiblement supervisé sur la base de la théorie du MIL par utilisation du modèle MI-SVM. 143
- iii.1 Exemples de données de photographie clinique recadrées manuellement autour de la zone d'intérêt pour le besoin de ce manuscrit. 149
- iii.2 Courbes ROC issues de l'évaluation des données de dermatoscopie et de MCR par les experts [40]. En haut, les résultats des 15 experts menés sur la modalité de dermatoscopie ; En bas, les résultats des 12 experts menés sur la modalité de MCR. À gauche, les résultats sur la détection d'éléments malins ; À droite, le résultat des experts mené sur la détection de LM et LMM. 151
- 8.1 Schéma reprenant la progression de ce chapitre et l'interaction entre les différentes méthodes présentées. 154
- 8.2 Schéma de fonctionnement du processus multimodal sans mémoire, dont la phase d'entraînement est réalisée en deux temps. Dans un premier temps, un entraînement est réalisé sur les modèles de prédiction au niveau de chaque modalité mise à disposition par le premier jeu de données d'entraînement. Dans un second temps, le modèle de gestion optimise les seuils de confiance à l'aide du second jeu de données d'entraînement. L'entraînement achevé, les performances sont mesurées sur les données d'évaluation. 159
- 8.3 Schéma de fonctionnement du processus multimodal avec mémoire, dont la phase d'entraînement est réalisée en deux temps. Dans un premier temps, un entraînement est réalisé sur les modèles de prédiction propre à chaque étape du processus clinique et pour lesquels les données précédemment acquises sont cumulées. Dans un second temps, le modèle de gestion optimise les seuils de confiance à l'aide du second jeu de données d'entraînement. L'entraînement achevé, les performances sont mesurées sur les données d'évaluation. 160
- 8.4 Schéma de représentation des concepts de seuil unique et multiple proposés dans cette section. 161
- 8.5 Courbes de calibration et scores de Brier obtenus pour le processus multimodal sans mémoire. L'écart-type calculé sur les différents lots figure en opaque pour les courbes. En haut, le processus multimodal sans calibration ; À gauche, le processus multimodal avec calibration Isotonique ; À droite, le processus multimodal avec calibration Sigmoïde. 165

- 8.6 Courbes de calibration et scores de Brier obtenus pour le processus multimodal avec mémoire. L'écart-type calculé sur les différents lots figure en opaque pour les courbes. En haut, le processus multimodal sans calibration ; À gauche, le processus multimodal avec calibration Isotonique ; À droite, le processus multimodal avec calibration Sigmoidale. 166
- 8.7 Résultats de la prise en charge des lésions dans un processus séquentiel multimodal. À gauche, selon un processus sans mémoire ; À droite, selon un processus avec mémoire. De gauche à droite, les cas bénins correctement classifiés sont représentés en vert foncé ; les cas bénins incorrectement classifiés sont représentés en vert clair ; les cas malins incorrectement classifiés sont représentés en rouge clair ; les cas malins correctement classifiés sont représentés en rouge foncé. . . . 167

LISTE DES TABLES

i.1	Statistiques mondiales d'incidence et de mortalité des 32 groupes de cancers (selon CIM-10 de l'OMS) pour l'année 2015 [65]. Ces chiffres sont exprimés en milliers et ne recensent pas les cancers de la peau hors mélanome.	2
2.1	Table des principales modalités de mesure optique utilisées à but clinique ou à un stade de recherche clinique en dermatologie [101].	25
3.1	Table de comparaison des méthodes d'apprentissage majeures de la littérature [103]. Ce tableau propose un système de notations allant de une (mauvaise performance) à quatre étoiles (bonne performance).	42
4.1	Caractéristiques observables par dermatoscopie jugées pertinentes pour la détection du LM [146].	63
4.2	Caractéristiques observables par MCR jugées pertinentes pour la détection du LM [75].	64
5.1	Table des différentes mesures obtenues à partir du premier ordre.	79
5.2	Table des différentes mesures proposées par Haralick et al. [80] applicables aux GLCM pour caractériser des images texturées.	81
5.3	Liste des méthodes spatiales évaluées dans ce chapitre et leur nombre de caractéristiques extraites associées.	81
5.4	Liste des méthodes fréquentielles évaluées dans ce chapitre et leur nombre de caractéristiques extraites associées.	84
5.5	Liste des architectures de CNN employées dans ce chapitre et leur nombre de caractéristiques extraites associées.	87
5.6	Table des méthodes de réduction employées et leurs paramètres associés.	96
5.7	Table des diverses méthodes de balancement mises en œuvre pour corriger les déséquilibres de données.	98
5.8	Table recensant l'ensemble des modèles de classification et leurs hyperparamètres évalués lors des expériences de ce travail.	99
5.9	Résultats à trois classes issus de l'expérience mêlant les méthodes les plus performantes associées aux méthodes de balancement de données évoquées.	103
5.10	Résultats de la détection de tissus malins issus de l'expérience mêlant les méthodes les plus performantes associées aux méthodes de balancement de données évoquées.	106

6.1	Table reprenant le modèle de classification et ses hyperparamètres employés sur les caractéristiques issues de la décomposition en ondelettes multi-échelle.	112
6.2	Liste des méthodes de décomposition en ondelettes multi-échelle et leur nombre de caractéristiques extraites associées.	112
6.3	Table recensant les paramètres du modèle de classification employé pour réaliser la prédiction sur des caractéristiques spatiales multi-échelle. Ce modèle associé à ces hyperparamètres est employé pour la classification par échelle mais également pour l'agrégation des multiples échelles.	114
6.4	Listes des méthodes sur base de décomposition spatiale multi-échelle et leur nombre de caractéristiques extraites associées.	114
6.5	Table reprenant le modèle et les hyperparamètres du modèle de classification employé pour la détection à basse échelle des tissus dans le processus de fenêtre glissante.	115
6.6	Ensemble des paramètres d'expérimentations menées par l'approche de type fenêtre glissante. L'étude de l'impact de la taille de fenêtre est limitée aux valeurs de 250×250 et 500×500 , mais également du chevauchement sur des valeurs de 0 %, 25 % et 50 %.	116
6.7	Table recensant les paramètres de transformation utilisés sur les données images de MCR de manière aléatoire lors de l'entraînement.	119
6.8	Résultats issus du processus de classification basé sur les décompositions multi-échelle exprimées à l'aide du score F_1	122
6.9	Résultats issus du processus de classification basé sur le principe de fenêtre glissante exprimé à l'aide du score F_1 . La première partie de ce tableau reprend les résultats obtenus à l'aide d'une fenêtre glissante de taille de 250×250 pixels tandis que la seconde partie se concentre sur les résultats d'une fenêtre glissante de taille de 500×500 pixels.	123
6.10	Résultats issus du processus de classification basé sur les CNN par réglage fin et programme d'apprentissage exprimés à l'aide du score F_1	124
7.1	Table reprenant le modèle et les hyperparamètres du modèle de classification employé pour la détection bas niveau des images.	134
7.2	Table reprenant les divers modèles utilisant le principe du MIL pour réaliser une classification au niveau lésionnel, ainsi que leurs hyperparamètres associés. . . .	136
7.3	Table de résultats issus des processus de classification supervisée employés pour mener une classification au niveau lésionnel, exprimés selon la sensibilité, la spécificité et le score F_1	138
7.4	Table de résultats issus des processus de classification supervisée employés pour mener une classification au niveau lésionnel sur les pathologies de LM et LMM , exprimés selon la sensibilité, la spécificité et le score F_1	139
7.5	Table de résultats issus des processus de classification faiblement supervisée par MIL employés pour mener une classification au niveau lésionnel, exprimés selon la sensibilité, la spécificité et le score F_1	139

7.6	Table de résultats issus des processus de classification faiblement supervisée par MIL employés pour mener une classification au niveau lésionnel sur les pathologies de LM et LMM , exprimés selon la sensibilité, la spécificité et le score F_1 . . .	140
7.7	Table de résultats issus de l'application du modèle MI-SVM employé pour mener une classification au niveau des instances, sur l'ensemble des images annotées, et exprimés selon la sensibilité, la spécificité et le score F_1	142
7.8	Table de résultats issus de l'application du modèle MI-SVM employés pour mener une classification au niveau des instances, sur les images annotées issues de pathologies de LM et LMM , et exprimés selon la sensibilité, la spécificité et le score F_1	142
8.1	Table reprenant les divers modèles utilisant le principe de MIL pour réaliser une classification au niveau lésionnel, ainsi que leurs hyperparamètres associés. . . .	155
8.2	Résultats obtenus à partir des méthodes de classification employées sur chacune des modalités utilisées dans ce processus.	156
8.3	Résultats pondérés sur la classe bénigne et maligne sur l'efficacité des méthodes de réduction des caractéristiques par classification à l'aide d'un modèle SVM . . .	156
8.4	Table de recensement des divers paramètres évalués pour la détermination des seuils de confiance optimums au sein du processus multimodal.	161
8.5	Table de résultats issus de l'application du processus multimodal sans mémoire. Ces résultats sont exprimés à l'aide du score F_1 et sont issus d'une prise en charge de l'ensemble des cas cliniques mis à disposition de ce manuscrit.	163
8.6	Table de résultats issus de l'application du processus multimodal avec mémoire. Ces résultats sont exprimés à l'aide du score F_1 et sont issus d'une prise en charge de l'ensemble des cas cliniques mis à disposition de ce manuscrit.	163
8.7	Table reprenant les résultats de la classe maligne, issus de l'application des processus multimodaux sans mémoire et avec mémoire sur l'ensemble des lésions. . .	164
8.8	Table reprenant les résultats de la classe maligne, issus de l'application des processus multimodaux sans mémoire et avec mémoire avec pour seules pathologies malignes les cas de LM et LM	164

LISTE DES ACRONYMES

- ANN** Artificial Neural Network (*alt.* : Réseau de Neurones Artificiels, RNA). 45, 46
- AUC** Area Under the Curve (*alt.* : Aire sous la courbe, ASC). 51, 97, 106, 108, 141, 142, 144, 145, 150, 195
- CAM** Class Activation Map (*alt.* : Carte d'Activation de Classe, CAC). 121, 125–127, 199
- CART** Classification And Regression Trees (*alt.* : Arbre de Classification et de Régression, ACR). 41, 42, 99, 100, 107
- CBC** Carcinome Basocellulaire (*alt.* : Basal Cell Carcinoma, BCC). 16, 28, 33, 35, 58
- CHUSTE** Centre Hospitalier Universitaire de Saint-Étienne. 58
- CIM-10** Classification Internationale des Maladies. 2, 203
- CNN** Convolutional Neural Network (*alt.* : Réseau Neuronal Convolutif, RNC). 45, 46, 85–87, 92, 94, 95, 107–110, 114, 118–121, 124–126, 134, 168, 173, 197, 198, 203, 204
- CSC** Carcinome Spino Cellulaire (*alt.* : Squamous Cell Carcinoma, SCC). 16, 28, 33, 35
- CSV** Comma-Separated Values. 61
- DEL** Diode Électro-Luminescente (*alt.* : Light-Emitting Diode, LED). 23, 32, 194
- ENN** Edited Nearest Neighbor. 98
- FEDER** Fonds européen de développement régional. v
- GB** Gradient Boosting (*alt.* : Augmentation du Gradient, AG). 42, 99, 100, 107
- GLCM** Gray Level Co occurrence Matrix. 80, 81, 88, 197, 203
- GLRLM** Gray Level Run Length Matrix. 80
- GLSZM** Gray Level Size Zone Matrix. 80
- IA** Intelligence Artificielle (*alt.* : Artificial Intelligence, AI). 38
- ImViA** Imagerie et Vision Artificielle. v, vi
- IRM** Imagerie par Résonance Magnétique (*alt.* : Magnetic Resonance Imaging, MRI). 34, 118, 194
- JCE** Jeunes Chercheurs Entrepreneurs. v, vi

- JDE** Jonction Dermo-Épidermique (*alt.* : *Dermo Epidermal Junction, DEJ*). 12, 18, 27, 64, 65, 67, 71, 79, 150, 193, 196
- KAP** Kératose Actinique Pigmentée (*alt.* : *Pigmented Actinic Keratosis, PAK*). 19, 58
- KNN** *k*-Nearest Neighbors algorithm (*alt.* : *Méthode des k plus proches voisins, KNN*). 41, 42, 96
- KS** Kératose Séborrhéique (*alt.* : *Seborrheic Keratose, SK*). 19, 58, 66, 196
- LDA** Linear Discriminant Analysis (*alt.* : *Analyse Discriminante Linéaire, ADL*). 96, 101, 103, 198
- LM** Lentigo Maligna (*alt.* : *Mélanose de Dubreuilh, MD*). 4, 9, 15, 17–21, 58, 63–67, 71, 72, 78, 88–91, 106, 108, 131, 137–145, 149–151, 153, 162, 164, 171–173, 193, 196–198, 200, 203–205
- LMM** Lentigo Maligna Melanoma (*alt.* : *Mélanome sur Mélanose de Dubreuilh, MMD*). 4, 9, 15, 17–19, 21, 58, 63, 64, 66, 71, 72, 78, 106, 108, 131, 137–145, 149–151, 153, 162, 164, 171–173, 196, 198, 200, 204, 205
- LS** Lentigo Solaire (*alt.* : *Solar Lentigo, SL*). 19, 58
- MCR** Microscopie Confocale par Réflectance (*alt.* : *Reflectance Confocal Microscopy, RCM*). 6, 29, 58–60, 64–67, 71, 72, 75, 77, 78, 80, 82, 95, 100, 101, 107–111, 115, 118, 119, 121, 126, 128, 131–133, 135, 137, 145, 149–151, 153–156, 158, 164, 167, 168, 171–173, 194, 196–200, 203, 204
- MIL** Multi-Instance Learning (*alt.* : *Apprentissage par Instances Multiples, AIM*). 132, 135, 136, 139–144, 155, 172, 200, 204, 205
- MI-SVM** Multi-Instance Support Vector Machine (*alt.* : *Machine à Vecteur de Support par Instances Multiples, MVS-IM*). 136, 137, 139–145, 200, 205
- MLP** Multi-Layer Perceptron (*alt.* : *Perceptron Multi-Couches, PMC*). 46, 99, 100, 107
- NBM** Naive Bayes Model (*alt.* : *Modèle Bayésien Naïf, MBN*). 41, 42
- NSK** Normalized Set Kernel (*alt.* : *Noyau d'Ensemble Normalisé, NEN*). 136, 139–141, 144, 145, 155, 168, 172, 200
- OMS** Organisation Mondiale de la Santé (*alt.* : *World Health Organization, WHO*). 1, 2, 203
- PACS** Picture Archiving and Communication System (*alt.* : *Système d'Archivage et de Transmission d'Images, SATI*). 26
- PCA** Principal Component Analysis (*alt.* : *Analyse en composantes principales, ACP*). 91–96, 101, 103, 107, 108, 198
- RF** Random Forest (*alt.* : *Forêts Aléatoires, RF*). 42, 99, 100, 107
- ROC** Receiver Operating Characteristic. 51, 72, 106, 141–143, 150, 151, 195, 196, 198, 200

sbMIL Sparse Balanced Multi-Instance Learning. 136, 139, 140, 144

SMOTE Synthetic Minority Over-Sampling. 97, 98

SVM Support Vector Machine (*alt.* : *Machine à Vecteurs de Support, MVS*). 41, 42, 96, 99, 100, 103, 106, 107, 110, 112–116, 122, 128, 134–136, 155, 156, 168, 171, 205

TCO Tomographie en Cohérence Optique (*alt.* : *Optical Coherence Tomography, OCT*). 30, 194

TL Tomek-Link. 97, 98

UV Ultraviolet. 11, 14, 16, 17

