



**HAL**  
open science

# Amélioration du diagnostic du cancer de la peau par classification des images de dermoscopie à l'aide de techniques d'intelligence artificielle

Lébini Chabi Adjobo

► **To cite this version:**

Lébini Chabi Adjobo. Amélioration du diagnostic du cancer de la peau par classification des images de dermoscopie à l'aide de techniques d'intelligence artificielle. Traitement du signal et de l'image [eess.SP]. Université Bourgogne Franche-Comté; Université d'Abomey-Calavi (Bénin), 2023. Français. NNT : 2023UBFCK103 . tel-04607949

**HAL Id: tel-04607949**

**<https://theses.hal.science/tel-04607949v1>**

Submitted on 11 Jun 2024

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

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



**THESE DE DOCTORAT DE L'UNIVERSITE DE BOURGOGNE FRANCHE-COMTE ET DE L'UNIVERSITE  
D'ABOMEY-CALAVI**

**PREPAREE A L'UNIVERSITE DE BOURGOGNE FRANCHE-COMTE**

École doctorale n°ED37

SPIM

Doctorat en Instrumentation et Informatique de l'Image

Par

CHABI ADJOBO Lébini Firmine Esther

Amélioration du diagnostic du cancer de la peau par la classification des images de dermoscopie à l'aide de techniques d'intelligence artificielle

Thèse présentée et soutenue à Dangbo, le 03/10/2023

Composition du Jury :

Monsieur, EZIN, Cokou Eugène	Professeur Université d'Abomey-Calavi	Président
Monsieur, CHBEIR, Richard	Professeur Université de Pau	Rapporteur
Monsieur, LARABI, Mohamed-Chaker	Professeur Université de Poitiers	Rapporteur
Monsieur, MORAIN-NICOLIER, Frédéric	Professeur Université de Reims	Examineur
Monsieur, ANSARI, Keivan	Docteur Institute of Color Science and Technology	Examineur
Monsieur, GOUTON, Pierre	Professeur Université de Bourgogne Franche-Comté	Directeur de thèse
Monsieur, TOSSA, Joël	Professeur Université d'Abomey-Calavi	Codirecteur de thèse
Monsieur, SANDA MAHAMA, Amadou Tidjani	Docteur Université d'Abomey-Calavi	Encadrant

**Titre :** AMELIORATION DU DIAGNOSTIC DU CANCER DE LA PEAU PAR LA CLASSIFICATION DES IMAGES DE DERMOSCOPIE A L'AIDE DE TECHNIQUES D'INTELLIGENCE ARTIFICIELLE

**Mots clés :** Cancers de la peau, Images de dermoscopie, Réseaux neuronaux convolutionnels, Apprentissage supervisé, Augmentation base de données, Détection automatique.

**Résumé :** Les travaux menés dans le cadre de cette thèse contribuent à l'amélioration du diagnostic des cancers de la peau grâce à des outils d'aide au diagnostic. Ces techniques, plus particulièrement celles basées sur l'apprentissage profond, ont révolutionné le paysage de la vision par ordinateur et s'intéressent à l'analyse des caractéristiques morphologiques, de texture et de couleur dans une image afin de définir le type de lésion qu'elle présente. Les résultats des recherches menées ces dernières années sont prometteurs, notamment pour le diagnostic précoce du mélanome malin. Malgré ces résultats, des contraintes majeures subsistent : la complexité des architectures de modèles, le besoin de grandes bases de données, le déséquilibre des classes observé et la faible représentation des peaux foncées dans les bases de données. Pour répondre à ces limitations et améliorer les performances de prédiction, nous avons proposé : (1) une architecture de réseau convolutionnel adaptée aux petites bases de données.

L'architecture proposée est appelée GCN (Gabor Convolutional Network). (2) une approche d'augmentation d'image pour corriger le déséquilibre de classe observé et la faible représentation des peaux foncées dans les bases de données. L'algorithme proposé combine la transformée de Shearlet non sous-échantillonnée et les réseaux neuronaux à couplage d'impulsions pour augmenter la base de données d'images. (3) Une approche pour la détection automatique des caractéristiques dermoscopiques. L'algorithme proposé est basé sur le modèle Yolo (You look only once) et met en évidence les caractéristiques présentes dans une image dermoscopique. Les solutions proposées ont été validées sur des bases d'images dermoscopiques issues du challenge ISIC (ISIC 2019, ISIC 2018, ISIC 2017) et comparées à d'autres travaux de la littérature. Dans les trois cas, les résultats montrent une amélioration de la précision de la prédiction.

**Title :** IMPROVING SKIN CANCER DIAGNOSIS BY CLASSIFYING DERMOSCOPIIC IMAGES USING ARTIFICIAL INTELLIGENCE TECHNIQUES

**Keywords :** Skin cancers, Dermoscopy images, Convolutional neural networks, Supervised learning, Database augmentation, Automatic detection

**Abstract :** The work carried out as part of this thesis contributes to improving the diagnosis of skin cancers using diagnostic aid tools. These techniques, particularly those based on deep learning, have revolutionized the computer vision landscape and focus on the analysis of morphological, textural and color characteristics in an image in order to define the type of lesion it presents. The results of research carried out in recent years are promising, particularly for the early diagnosis of malignant melanoma. Despite these results, major constraints remain: the complexity of model architectures, the need for large databases, the class imbalance observed and the poor representation of darker skins in databases. To address these limitations and improve prediction performance, we proposed: (1) a convolutional network architecture adapted to small databases.

The proposed architecture is called GCN (Gabor Convolutional Network). (2) an image augmentation approach to correct the observed class imbalance and low representation of dark skins in databases. The proposed algorithm combines the subsampled Shearlet transform and pulse-coupled neural networks to augment the image database. (3) An approach for automatic detection of dermoscopic features. The proposed algorithm is based on the Yolo (You look only once) model and highlights the features present in a dermoscopic image. The proposed solutions have been validated on dermoscopic image bases from the ISIC challenge (ISIC 2019, ISIC 2018, ISIC 2017) and compared with other works in the literature. In all three cases, the results show an improvement in prediction accuracy.



*Dédicaces*

*A mon père, A ma mère,*

*Ce travail représente le fruit de vos années d'investissements et j'en suis reconnaissante.  
Merci d'avoir contribué à faire de moi la personne que je suis. Je vous souhaite une bonne  
santé et que Dieu vous garde.*



## Remerciements

Cette thèse est réalisée en cotutelle entre l'Université de Bourgogne (UB) et l'Université d'Abomey-Calavi (UAC) dont je tiens à remercier les écoles doctorales. Je tiens aussi à remercier le laboratoire Imagerie et Vision Artificielle (ImViA) de UB et l'institut de Mathématiques et Sciences Physiques (IMSP) de l'UAC, dans lesquels les travaux présentés dans ce document ont été menés.

Mes travaux ont bénéficié des financements du programme DAAD In-country In-region et du Centre d'Excellence d'Afrique en Sciences Mathématiques et Applications. Je tiens à remercier tous ceux qui ont initié ces programmes, ainsi que tous ceux qui ont participé à sa mise en œuvre et à sa gestion.

Je voudrais exprimer ma gratitude à mon directeur de thèse Pr. Pierre Gouton. Je lui suis très reconnaissante pour sa disponibilité, son aide, ses conseils et son encadrement qui m'ont permis d'avancer dans mes recherches. J'exprime aussi toute ma reconnaissance à mon co-directeur de thèse Pr Joël Tossa pour son accompagnement et son aide. J'ai eu la chance et l'honneur d'apprendre auprès de chacun d'entre vous. Je remercie également mon encadreur Dr Abdou Tidjani Sanda Mahama pour sa disponibilité et ses conseils tout au long de cette thèse, qui m'ont permis de mener à bien ce travail.

J'adresse mes remerciements aux Professeurs Richard Chbeir, Mohamed-Chaker Larabi, Guy Degla, Frédéric Morain-Nicolier, Sèmèvo Arnaud Roland Martial AHOUANJINO, et au Docteur Keivan Ansari pour avoir accepté d'être rapporteurs et examinateurs de mes travaux. Mes sincères remerciements vont également au Professeur Eugène Ezin qui a accepté d'évaluer ce travail en tant que Président du jury.

Je tiens également à exprimer ma gratitude au Professeur Hugues Adegbidi et au Docteur Franck Olivier Yedomon, tous deux Dermato-Vénérologues, pour leur disponibilité, leur ouverture d'esprit et toutes ces précieuses informations qui ont orienté notre travail.

J'ai une pensée particulière pour tous mes collègues du laboratoire, Frantz Tossa, Elvire Houssou, Vahid Mohammadi, Jhunlyn Lorenzo, Rose Waytehad Moskolai, Jean Kacoutchy Ayikpa, Norbert Hounssou, Guillaume Sodjinou et Kossi Kuma Katapke ; qui ont rendu ces années plus conviviales. Je pense avec nostalgie à nos échanges passionnants lors des pauses café. Un grand merci à mes amis Frantz et Elvire. C'est à eux que je dois des moments inoubliables passés dans une ambiance fraternelle, bienveillante, encourageante et enrichissante.

Je voudrais remercier l'ensemble du personnel administratif de l'UB, de l'IMSP et de l'UAC. Merci pour les informations fournies, la disponibilité, et l'accompagnement tout au long des différentes procédures administratives.

Enfin, je ne remerciais jamais assez mes chers parents, Nestor et Débora, mes frères et sœurs Josias, Marc Aurèle, Florian, Michael, Marie, mes belles-sœurs Madeleine et Marthe pour leur soutien tout au long de ce parcours.





# TABLE DES MATIÈRES

<b>Table des figures</b>	<b>xi</b>
--------------------------	-----------

## **LISTE DES SIGLES**

## **INTRODUCTION GÉNÉRALE**

1	Problématique . . . . .	1
2	Contexte . . . . .	2
3	Contributions . . . . .	3
4	Organisation . . . . .	4

## **Chapitre 1**

### **Lésions de la peau**

1.1	Structure et fonctions de la peau . . . . .	6
1.1.1	Structure . . . . .	6
1.1.2	Fonctions . . . . .	10
1.1.3	Phototype cutané . . . . .	10
1.1.4	Particularités de la structure et des fonctions de la peau sur peaux noires . . . . .	11
1.2	Familles de lésions cutanées . . . . .	11
1.2.1	Tumeurs bénignes . . . . .	12
1.2.2	Tumeurs malignes . . . . .	12
1.2.3	Particularités lésions de la peau sur peaux noires . . . . .	13
1.3	Diagnostic . . . . .	13

1.3.1	Examen clinique . . . . .	13
1.3.2	Examen histologique . . . . .	14
1.3.3	Modalités d'imagerie de la peau . . . . .	15
1.3.4	Particularités du processus de diagnostic sur les peaux noires . . . . .	18

## Chapitre 2

### Classification d'images dermoscopiques

2.1	Classification basée sur l'apprentissage automatique traditionnel . . . . .	22
2.1.1	Pré-traitement . . . . .	23
2.1.2	Extraction manuelle de caractéristiques . . . . .	23
2.1.3	Classification . . . . .	25
2.2	Classification basée sur l'apprentissage profond . . . . .	31
2.2.1	Apprentissage à partir de zéro . . . . .	32
2.2.2	Apprentissage par transfert . . . . .	32
2.2.3	Extraction automatique de caractéristiques avec les réseaux de neurones . . . . .	33
2.3	Évaluation des outils d'aide au diagnostic . . . . .	33
2.3.1	Méthodes de validation . . . . .	33
2.3.2	Mesure de performance . . . . .	33
2.4	Comparaison des méthodes de classification . . . . .	36
2.5	Méthode proposée : GCN . . . . .	37
2.5.1	GCN . . . . .	37
2.5.2	Méthode . . . . .	38
2.5.3	Résultats expérimentaux et discussions . . . . .	40

## Chapitre 3

### Augmentation de la base d'images dermoscopiques

3.1	Introduction . . . . .	45
3.2	Augmentation basée sur des approches traditionnelles . . . . .	46
3.2.1	Transformations géométriques . . . . .	46
3.2.2	Recadrage . . . . .	47
3.2.3	Occlusion . . . . .	47
3.2.4	Transformation de couleur . . . . .	48
3.2.5	Injection de bruit . . . . .	48
3.2.6	Filtrage . . . . .	48

3.3	Augmentation basée sur les réseaux générateurs antagonistes . . . . .	48
3.3.1	Généralités sur les réseaux générateurs antagonistes . . . . .	48
3.3.2	Augmentation d'images médicales avec les GAN . . . . .	49
3.4	Augmentation basée sur des approches de fusion d'images . . . . .	50
3.4.1	Fusion d'images dans le domaine spatial . . . . .	50
3.4.2	Fusion d'images dans le domaine transformé . . . . .	51
3.5	Comparaison des méthodes d'augmentation d'images . . . . .	54
3.6	Méthode proposée : NSST PCNN . . . . .	55
3.6.1	PCNN simplifié et adaptif . . . . .	56
3.6.2	Algorithme . . . . .	58
3.6.3	Resultats expérimentaux et discussions . . . . .	59

## Chapitre 4

### Détection automatique de caractéristiques dermoscopiques

4.1	Détection basée sur un apprentissage traditionnel . . . . .	70
4.1.1	Histogramme des gradients orientés . . . . .	71
4.1.2	Méthode de Viola-Jones . . . . .	71
4.1.3	Modèles de pièces déformables . . . . .	72
4.1.4	Correspondance des modèles . . . . .	72
4.2	Détections basées sur l'apprentissage profond . . . . .	73
4.2.1	Architecture des détecteurs d'objets basés sur le deep learning . . . . .	73
4.2.2	Littérature sur la détection de caractéristiques dermoscopiques basée sur l'apprentissage profond . . . . .	74
4.3	Comparaison des méthodes de détection d'objets . . . . .	75
4.4	Méthode proposée : Yolo-dermoscopic-features . . . . .	76
4.4.1	Yolov8 . . . . .	76
4.4.2	Algorithme . . . . .	79
4.4.3	Résultats expérimentaux et discussions . . . . .	81

## CONCLUSION GÉNÉRALE ET PERSPECTIVES

Conclusion générale . . . . .	93
Perspectives . . . . .	94

## Listes des publications

**Bibliographie**

**99**

## LISTE DES TABLEAUX

2.1	Matrice de confusion. . . . .	34
2.2	Étude comparative des méthodes de classification. . . . .	36
3.1	Transformations basées sur l'Équation 3.1 . . . . .	47
3.2	Comparaison de méthodes d'augmentation d'images . . . . .	55
3.3	Algorithme pour la fusion d'images . . . . .	59
3.4	Correspondance entre l'ATI, catégories de peaux et les six phénotypes cutanés de Fitzpatrick. . . . .	61
3.5	Résultats des métriques d'évaluation des différentes méthodes. . . . .	64
3.6	Temps d'exécution. . . . .	64
3.7	Résultats des mesures d'évaluation du Modèle 1. . . . .	65
3.8	Résultats des mesures d'évaluation du Modèle 2. . . . .	65
4.1	Comparaison de méthodes de détection d'objets . . . . .	76
4.2	Algorithme de génération de fichiers d'annotations à partir des d'images binaires . . . . .	81
4.3	Performances de l'algorithme Yolo-dermoscopic-features. . . . .	87
4.4	Comparaison de différentes méthodes de détection de caractéristiques dermoscopiques. . . . .	91
4.5	Performances de différentes méthodes de détection de caractéristiques dermoscopiques. . . . .	91



## TABLE DES FIGURES

1	Évolution du nombre de publications de 2002 à 2022 autour des termes "Skin lesion classification". . . . .	2
1.1	Structure de la peau[1]. . . . .	6
1.2	Couches de l'épiderme et composition cellulaire[1]. . . . .	7
1.3	Structure et composition du derme [1]. . . . .	8
1.4	Annexes de la peau [1]. . . . .	9
1.5	Échelle de Fitzpatrick ou six phénotypes cutanés [2]. . . . .	11
1.6	Exemple de dermatoscopes et d'images dermoscopiques. (a) Molemax HD (de derma medical system, Vienne Autriche)[1], (b) FotoFinder (FotoFinder Systems GmbH, Bad Birnbach, Allemagne)[1], (c) Images dermoscopiques.	16
1.7	(a) Principe de fonctionnement du MCR [3]. (b) Le système Vivascope 1) Vidéodermoscopie ; 2) Microscope confocal VS1500 avec bras articulé ; 3) Microscope confocal manuel VS3000 ; 4) PC ; 5) Écran [3]. . . . .	17
1.8	Principe de fonctionnement de la DRS [1]. . . . .	18
1.9	Principe de fonctionnement de la TCO. . . . .	19
2.1	Schématisation des différentes étapes des méthodes de classification automatique de lésions cutanées. . . . .	22
2.2	a) Exemple d'une image dermoscopique de mélanome asymétrique, à bords irréguliers, et excentrique, (b) Exemple d'une image dermoscopique de mélanome plutôt symétrique, et à bords irréguliers, (c) Exemple d'une image dermoscopique de nevus plutôt symétrique à bords réguliers [4, 5, 6]. . . . .	24

---

2.3	(a) Exemple de réseau pigmentaire atypique sur une image dermoscopique de nevus, (b) Exemple de points irréguliers sur une image dermoscopique de nevus, (c) Exemple de globules irréguliers sur une image dermoscopique de mélanome, (d) Exemple de stries sur une image dermoscopique de mélanome, (e) Exemple de zone sans structure sur une image dermoscopique de nevus [4, 5, 6]. . . . .	25
2.4	Structure d'un neurone artificiel. . . . .	27
2.5	Architecture d'un réseau de neurones artificiels. . . . .	28
2.6	Fonctions d'activation les plus connues.(a) Sigmoides, (b) Tangente hyperbolique (tanh), et (c) ReLU (Rectified Linear Unit). . . . .	28
2.7	Architecture traditionnelle d'un CNN. . . . .	31
2.8	Méthode GCN proposée. . . . .	38
2.9	Architecture du réseau proposé. . . . .	42
2.10	Précisions des GCN et CNN sur une partie de la base de données. . . . .	43
2.11	Précisions des GCN et CNN sur l'ensemble de la base de données. . . . .	43
3.1	Structure de base d'un GAN [7]. . . . .	49
3.2	Classification des méthodes de fusion. . . . .	51
3.3	Procédure de fusion d'images dans le domaine fréquentiel. . . . .	51
3.4	Exemple pyramide d'image sur une coupe d'IRM. [8] . . . . .	52
3.5	Diagramme de la méthode proposée. . . . .	56
3.6	Neurone à couplage d'impulsion de base. . . . .	57
3.7	(a) Source A : caractéristiques dermoscopiques d'un mélanome sur une peau très claire [4, 5, 6]. (b) Source B : image de peau foncée saine. (c) Caractéristiques dermoscopiques d'un nævus mélanocytaire sur une peau mate [4, 5, 6].(d) Caractéristiques dermoscopiques d'un nævus mélanocytaire sur peau foncée [4, 5, 6]. . . . .	60
3.8	(a1)–(a3) Images dermoscopiques de tons de peaux clairs provenant la base de données source A [9, 10, 11]. (b1-f3) Résultats obtenus en utilisant respectivement CT, DWT-SR, DWT-CT, NSCT-SR-SML, et la méthode proposée PCNN-NSST. . . . .	67
3.9	Comparaison des résultats des métriques d'évaluation. (a) Accuracy, (b) Precision, (c) Recall et (d) F1-score. . . . .	68
4.1	Méthodes de détection d'objets. . . . .	70
4.2	Frise chronologique de différentes versions de YOLO. . . . .	77
4.3	Yolo Architecture [12]. . . . .	77
4.4	Pipeline de l'approche Yolo-dermoscopic-feature. Les parties en vert dans la figure représentent les parties sur lesquelles des modifications ont été apportées. . . . .	80
4.5	Structure hiérarchique du fichier json généré. . . . .	82



---

4.6	Structures des caractéristiques et images de l'ensemble de données. (a) Image dermoscopique d'un naevus mélanocytaire, (b) Image dermoscopique d'un mélanome, (c) Image dermoscopique d'un naevus, (d) Image dermoscopique d'un mélanome, (e) Image dermoscopique d'un naevus. . . . .	83
4.7	Graphiques de perte des modèles pendant l'entraînement et la validation. (a) (b) (c) (d) et (e) représentent respectivement les courbes du modèle de détection de globules, du modèle de détection des kystes de type milia, du modèle de détection du réseau pigmentaire, du modèle de détection du réseau négatif et du modèle de détection des stries. . . . .	85
4.8	Mesures de performance des modèles pendant la phase d'apprentissage. (a) (b) (c) (d) et (e) représentent respectivement les mesures de performance du modèle de détection de globules, du modèle de détection des kystes de type milia, du modèle de détection du réseau pigmentaire, du modèle de détection du réseau négatif et du modèle de détection des stries. . . . .	86
4.9	(a) Images dermoscopiques, (b) ) Masques d'annotation des globules par les experts (c) Prédiction des positions des globules. . . . .	87
4.10	(a) Images dermoscopiques, (b) ) Masques d'annotation des MLC par les experts (c) Prédiction des positions des MLC. . . . .	88
4.11	(a) Images dermoscopiques, (b) ) Masques d'annotation des PN par les experts (c) Prédiction des positions des PN. . . . .	89
4.12	(a) Images dermoscopiques, (b) ) Masques d'annotation des NN par les experts (c) Prédiction des positions des NN. . . . .	90
4.13	(a) Images dermoscopiques, (b) Masques d'annotation des streaks par les experts (c) Prédiction des positions des streaks. . . . .	90



## LISTE DES SIGLES

Adam	Adaptive Moment Estimation
ADC	Analog to Digital Converter
AUC	Area Under the ROC Curve
ANN	Artificial Neural Network
CIFAR	Canadian Institute for Advanced Research
CNN	Convolutional Neural Network
CSPDarknet	Cross Stage Partial Network
CT	Color Transfert
DSC	Dice Similarity Coefficient
DPM	Deformable Part Models
DWT	Discrete Wavelet Transform
DCGAN	Deep Convolutional GAN
FPN	Feature Pyramid Networks
GCN	Gabor Convolutional Network
GAN	Generative Adversarial Network
HOG	Histogram of Oriented Gradients

*Table des figures*

---

ITA	Individual Typology Angle
JSC	Jaccard Similarity Coefficient
KNN	K Nearest Neighbors
MNIST	Modified National Institute of Standards and Technology database
NSCT	Nonsubsampled Contourlet Transform
NSST	Non Subsampled Shearlet Transform
PAN	Path Aggregation Network
PCNN	Pulse-Coupled Neural Networks
PGAN	Progressive Growing GAN
R-CNN	Region-based Convolutional Neural Networks
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
RPN	Region Proposal Network
SGD	Stochastic Gradient Descent
SR	Sparse Representation
SSD	Single Shot MultiBox Detector
SURF	Speeded-Up Robust Features
SVM	Support Vector Machine
VGG	Visual Geometry Group
YOLO	You Only Look Once

## Sommaire

<b>1</b>	<b>Problématique</b>	<b>1</b>
<b>2</b>	<b>Contexte</b>	<b>2</b>
<b>3</b>	<b>Contributions</b>	<b>3</b>
<b>4</b>	<b>Organisation</b>	<b>4</b>

## 1 Problématique

Le traitement informatique des données obtenues par les modalités d'imagerie à travers les outils d'aide au diagnostic (Computer-aided diagnosis) a été fortement encouragé ces deux dernières décennies pour améliorer le diagnostic des lésions cutanées. L'amélioration de la précision de diagnostic des lésions cutanées en utilisant les outils d'aide au diagnostic est un sujet de recherche en imagerie médicale qui est en constante évolution. La Figure 1 présente l'évolution du nombre de publications de 2002 à 2022 autour des termes "Skin lesion classification".

Les résultats obtenus par les outils d'aide au diagnostic sont prometteurs notamment pour le diagnostic précoce de mélanome malin. Néanmoins, les architectures des modèles proposés sont complexes et nécessitent des dizaines de milliers à des millions d'images pour la phase d'apprentissage [13]. La constitution de grands ensembles de données d'images médicales est particulièrement délicate en raison l'incidence des maladies, du respect de la confidentialité des informations relatives au patient [14], de la nécessité de recourir à des experts médicaux pour l'étiquetage et du coût élevé des systèmes d'acquisition d'imagerie médicale. Une approche pour contourner cette difficulté est l'apprentissage par transfert. L'apprentissage par transfert une technique où un modèle formé pour une tâche source est partiellement "recyclé" pour une nouvelle tâche cible [15]. Selon [16], le type de données source affecte la précision de l'apprentissage obtenu à partir du nouveau modèle.

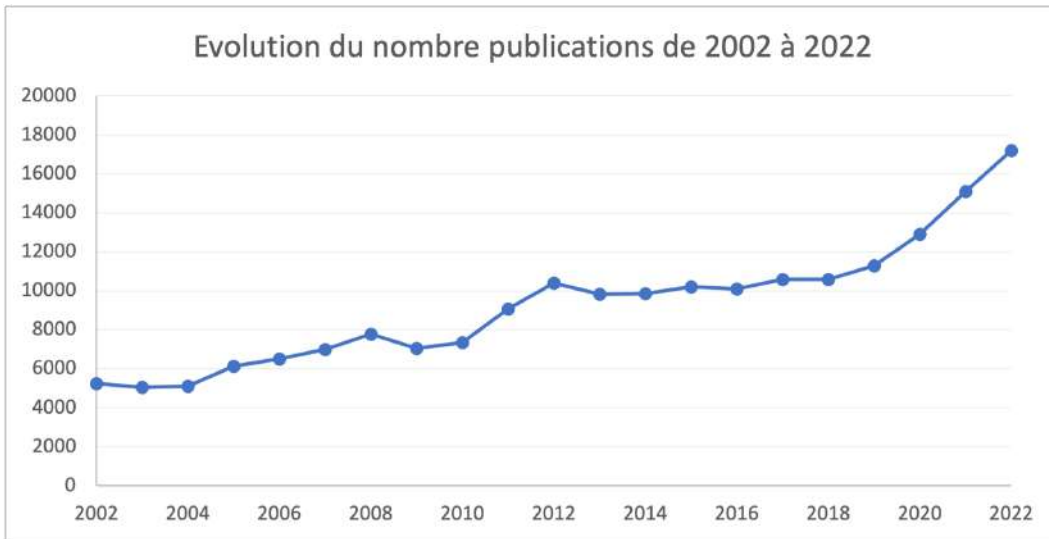


FIGURE 1 – Évolution du nombre de publications de 2002 à 2022 autour des termes "Skin lesion classification".

Aussi, les ensembles d'images du cancer de la peau accessibles au public sont caractérisés par une faible représentation des tous les types de peaux. Des analyses effectuées sur les images des bases de données Fitzpatrick 17k [17] et ISIC 2019 et 2020 [18] ont révélés que ces bases contiennent beaucoup plus d'images de peaux claires que de peaux foncées. La faible représentation de tous les types de peaux empêche toute généralisation des outils d'aide au diagnostic [19].

## 2 Contexte

La peau est le plus grand organe du système tégumentaire qui recouvre et protège le corps. Ce organe peut être affecté d'une part par des lésions bénignes qui sont sans conséquences graves sur l'organisme, et d'autre part par des lésions malignes qui sont des lésions cancéreuses [20].

Les lésions de la peau quand elles sont cancéreuses sont organisées en deux classes [20] : les mélanomes, et les non-mélanomes qui comprennent les carcinomes épidermoïdes et les carcinomes basocellulaires. Selon l'organisation mondiale de la santé [21], l'incidence des cancers de la peau a augmenté au cours des dernières décennies. Chaque année entre 2 et 3 millions cas de non-mélanomes et environ 132 000 cas de mélanomes sont enregistrés. Une étude [22] a révélé que le taux de mélanome aux États-Unis a doublé entre 1982 et 2011. Aussi, une étude [23] sur l'épidémiologie du mélanome cutané en France et en Europe a révélé qu'en France, 14 300 cas de mélanomes sont enregistrés par an dont 1773 décès. L'incidence des cancers de la peau est faible sur des sujets à peaux noires néanmoins le taux de mortalité y est plus élevé à cause du diagnostic tardif [24].

Le risque de cancers de la peau est corrélé aux caractéristiques génétiques (comme le

vitiligo, l'albinisme, le phototype), ainsi qu'au comportement d'exposition aux UV d'une personne [25]. Aussi, la dépigmentation volontaire de plus en plus observée en Afrique subsaharienne peut être un facteur de risque de cancer de peaux [26].

Le cancer de la peau, qu'il s'agisse de mélanomes ou de non-mélanomes, évolue en plusieurs stades. Le stade d'un cancer permet de décrire la quantité de cancer présent dans l'organisme, et de déterminer la gravité du cancer. Lorsqu'il est diagnostiqué à un stade précoce, le cancer de peau est hautement curable.

La première étape de diagnostic de cancers de la peau est l'évaluation visuelle. Dans ce contexte, le dermatologue se base principalement sur l'aspect de la lésion pour déterminer si elle est bénigne ou maligne. L'évaluation visuelle est basée sur la distribution, la taille, la forme, la bordure, la symétrie et la couleur de la lésion. Les diagnostics basés sur la couleur sont subjectifs car la perception de la couleur dépend de la réponse visuelle humaine à la lumière et de l'interaction de la lumière avec la peau [27]. Aussi le diagnostic précoce est très difficile car à ce stade, les mélanomes sont facilement confondus avec des lésions bénignes.

Pour améliorer le diagnostic précoce des cancers de la peau, la dermoscopie, également appelée la microscopie en épiluminescence, a été introduite. La dermoscopie a ouvert une nouvelle dimension dans l'examen des lésions cutanées pigmentées et, en particulier, dans l'identification de la phase précoce du mélanome malin [28]. La dermoscopie est une méthode non invasive qui permet d'évaluer *in vivo* les couleurs et les microstructures de l'épiderme, de la jonction dermo-épidermique et du derme papillaire qui ne sont pas visibles à l'œil nu [13]. En plus de la dermoscopie, il existe aussi d'autres modalités d'imageries comme la microscopie confocale par réflectance (MCR), la spectroscopie par réflectance, et l'imagerie multispectrale qui sont aussi utilisées pour mieux visualiser la structure des tissus cutanés. Néanmoins, l'interprétation des images obtenues de ces modalités nécessite du temps et de l'expérience.

### 3 Contributions

L'objectif principal de cette thèse est d'apporter des solutions aux problèmes posés en vue d'améliorer la précision de diagnostic des cancers de la peau, et particulièrement le diagnostic des mélanomes et des non-mélanomes, pour tous les types de peaux. De manière spécifique, il a été question de :

- proposer une architecture de réseaux convolutionnel adaptée aux petites bases de données : l'architecture proposée est nommée GCN (Gabor Convolutional Network). La particularité de ce modèle est que les neurones de la première couche convolutionnelle sont initialisés avec des filtres de Gabor ;
- proposer un algorithme pour analyser la représentation des tous les types de peaux :

l'algorithme proposé comprends deux étapes. Une première phase de segmentation qui consiste à délimiter la partie saine de la peau de la partie affectée. Ensuite l'ITA (Individual Typology Angle) est appliqué à la partie saine pour déterminer la teinte de peau ;

- proposer une méthode d'augmentation d'images : l'algorithme proposé combine la transformée de Shearlet non sous-échantillonnée et les réseaux de neurones à couplage d'impulsion pour augmenter la base d'images. Cette solution vient corriger le déséquilibre observé dans les ensembles de données de lésions cutanées : les images de sujet à peau mixtes, foncées ou noires sont très peu représentées.
- proposer une méthode de détection automatiques de caractéristiques dermoscopiques : l'algorithme proposé est basé sur le modèle Yolo (You look only once). Les lésions malignes sont caractérisées par la présence de certaines caractéristiques dermoscopiques. La mise en évidence de ces caractéristiques pour les spécialistes améliorera la précision de leur diagnostic.

## 4 Organisation

Ce document est organisé comme suit

- Le chapitre 1 présente en premier la peau, sa structure et ses fonctions. Ensuite les différentes familles de lésions cutanées et le processus de diagnostic des lésions cutanées sont présentés. Enfin un aperçu des modalités d'imagerie utilisées en dermatologie est donné. A chaque étape, les particularités sur peaux noires sont présentées ;
- Le chapitre 2 présente les outils d'aide au diagnostic de cancers de la peau, les méthodes d'évaluation de ces outils et la contribution de cette thèse au diagnostic automatique de cancers de la peau ;
- Le chapitre 3 présente les méthodes d'augmentation de données, les métriques d'évaluation des méthodes d'augmentation, et l'approche proposée pour augmenter la base de données et faciliter une meilleure généralisation de l'outils d'aide au diagnostic ;
- Le chapitre 4 présente les méthodes de détection automatique de caractéristiques dermoscopiques, suivi d'une présentation de la méthode proposée et des résultats d'expérimentation.



# CHAPITRE 1

## LÉSIONS DE LA PEAU

### Sommaire

<b>1.1</b>	<b>Structure et fonctions de la peau</b>	<b>6</b>
1.1.1	Structure	6
1.1.2	Fonctions	10
1.1.3	Phototype cutané	10
1.1.4	Particularités de la structure et des fonctions de la peau sur peaux noires	11
<b>1.2</b>	<b>Familles de lésions cutanées</b>	<b>11</b>
1.2.1	Tumeurs bénignes	12
1.2.2	Tumeurs malignes	12
1.2.3	Particularités lésions de la peau sur peaux noires	13
<b>1.3</b>	<b>Diagnostic</b>	<b>13</b>
1.3.1	Examen clinique	13
1.3.2	Examen histologique	14
1.3.3	Modalités d'imagerie de la peau	15
1.3.4	Particularités du processus de diagnostic sur les peaux noires	18

### Introduction

La peau, est un organe complexe de couverture ou une enveloppe externe que l'on retrouve sur l'homme et chez les vertébrés plus généralement. D'après [1], chez un adulte, la peau pèse près de 5 kg et mesure environ 2 mètres carrés.

L'épaisseur de la peau varie en fonction de la région. En effet la peau observée sur les paumes de mains ou la plante des pieds est plus épaisse que celle observée sur le corps, sur le visage, sur les lèvres ou autour des yeux. La charge pondérale influe également de manière importante sur l'épaisseur de la peau. Une personne souffrant d'obésité va donc présenter une peau plus épaisse [1]. Cette première section présente la peau, sa structure,

ses différentes fonctions, les affections de la peau plus précisément les cancers de la peau, le processus de diagnostic de ces lésions et quelques modalités d'imagerie de la peau.

## 1.1 Structure et fonctions de la peau

### 1.1.1 Structure

Sur le plan anatomique, comme présenté dans la Figure 1.1, la peau est constituée de deux parties principales :

- la partie superficielle externe qui constitue l'épiderme, et une partie interne constituant la jonction dermo-épidermique, le derme et l'hypoderme ;
- l'ensemble des phanères (poils et ongles) et de nombreuses glandes exocrines annexés à la peau.

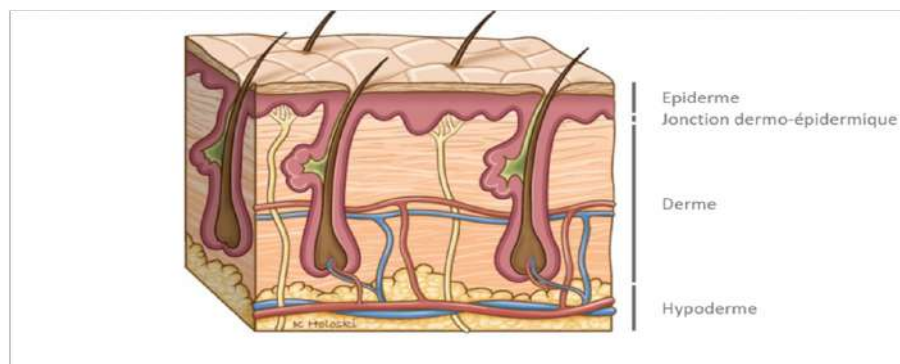


FIGURE 1.1 – Structure de la peau[1].

#### 1.1.1.1 Épiderme

L'épiderme est la partie superficielle externe de la peau et sa surface est principalement constituée de cellules mortes, ou squames, qui sont continuellement renouvelées. Ces cellules mortes proviennent de la couche basale, subissent avec le temps des mutations et migrent respectivement vers les couches épineuse, granuleuse, claire et enfin la cornée. L'épiderme est composé de quatre types cellulaires qui sont :

- les kératinocytes qui sont les cellules principales de l'épiderme (représentent 80% des cellules de l'épiderme). Ce sont elles qui migrent de la couche basale vers la cornée et forment une barrière cutanée qui permet à la peau d'être imperméable et protégée de l'environnement externe ;
- les mélanocytes qui assurent la protection contre les rayonnements Ultraviolet (UV) en synthétisant la mélanine, pigment à l'origine de la couleur de la peau ;

- les cellules de Langerhans et lymphocytes qui sont immunocompétentes. Elles permettent de détecter et capturer les corps étrangers (produits chimiques, bactéries, virus, etc.) qui pénètrent dans l'épiderme ;
- les cellules de Merkel qui sont des cellules sensorielles impliquées dans le sens du toucher.

La Figure 1.2 présente les différentes couches de l'épiderme et sa composition cellulaire.

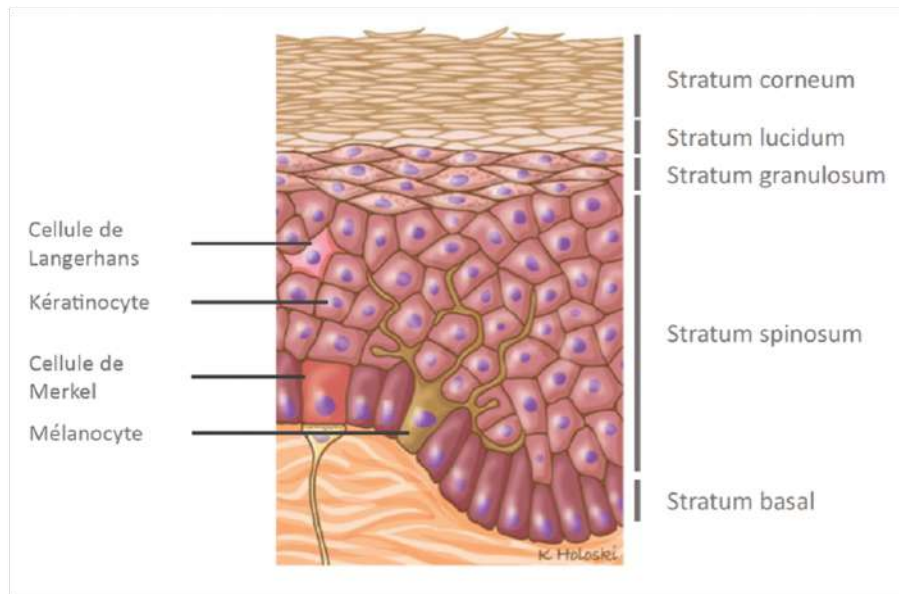


FIGURE 1.2 – Couches de l'épiderme et composition cellulaire[1].

### 1.1.1.2 Jonction dermo-épidermique

La jonction dermo-épidermique (ou membrane basale dermo-épidermique) est l'interface qui se situe entre l'épiderme et le derme et permet d'assurer une solide adhérence de ces deux couches. La jonction dermo-épidermique permet aussi d'assurer la nutrition des cellules de base de l'épiderme au travers d'un mécanisme de régulation des échanges moléculaires [29].

Cette membrane basale est constituée de plusieurs couches. En progressant de l'épiderme vers le derme on a :

- la membrane cellulaire du pôle basal des kératinocytes basaux qui contient les hémidesmosomes ;
- la membrane basale proprement dite, faite de deux feuillets : un feuillet clair aux électrons, la lamina lucida, et un feuillet dense aux électrons, la lamina densa. Les

filaments d’ancrage lient les hémidesmosomes à la lamina densa en traversant la lamina lucida ;

- la région située sous la lamina densa (sublamina densa), qui contient les fibrilles d’ancrage et les faisceaux de collagènes micro-fibrillaires et interstitiels.

### 1.1.1.3 Derme

Le derme est un tissu conjonctif séparé de l’épiderme par la jonction dermo-épidermique et comprends deux couches (voir Section 1.3). La couche papillaire (ou stratum papillare) qui forme une bordure ondulée bien définie avec la jonction dermo-épidermique. La couche réticulaire (ou stratum reticulare) qui est une zone épaisse et profonde qui forme une transition continue avec l’hypoderme [1, 30].

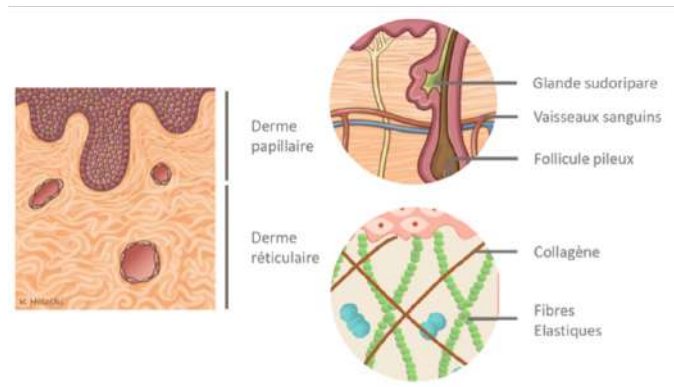


FIGURE 1.3 – Structure et composition du derme [1].

Le derme est constitué de :

- fibres de collagène, élastiques et réticulines qui sont fondamentales au maintien de la structure de la peau et lui donne sa souplesse et son élasticité ;
- cellules dont les fibroblastes qui participent à la synthèse des fibres, et des cellules du système immunitaire comme les lymphocytes, mastocytes, et macrophages tissulaires ;
- protéines glycosaminoglycanes dont l’acide hyaluronique qui possède une très forte capacité hydro-rétentrice et qui contribue à préserver le volume de la peau ;
- vaisseaux sanguins qui nourrissent l’épiderme tout en éliminant les résidus.

Le derme est aussi traversé par un grand nombre de vaisseaux lymphatiques, de glandes et de récepteurs sensoriels. Ces derniers assurent à la peau une sensibilité à la pression et aussi à la chaleur.

#### 1.1.1.4 Hypoderme

L'hypoderme est la couche la plus profonde de la peau et sert d'interface entre les structures sous cutanées et la peau [30]. L'hypoderme est principalement constitué de :

- tissu adipeux : un ensemble de cellules adipeuses (adipocytes) regroupées sous forme de coussins. Le tissu adipeux sert de réserve énergétique (graisses) et les libère au besoin. Ce tissu sert aussi d'amortisseur en protégeant les organes et d'isolant en empêchant les pertes de chaleur ;
- fibres de collagène spéciales (gaine ou cloisons de tissu) : un tissu conjonctif spongieux et lâche qui maintient les cellules adipeuses ensemble.

L'hypoderme est aussi traversé par les vaisseaux et les nerfs qui arrivent dans le derme. Le nombre de cellules adipeuses présentes dans l'hypoderme diffère varie selon la partie du corps et le sexe. Cette remarque s'applique aussi pour la structure des autres parties de la peau.

#### 1.1.1.5 Annexes de la peau

Sur la peau on retrouve aussi des annexes cutanées. Ces annexes prennent naissance au niveau du derme et de l'hypoderme. Les annexes de la peau comprennent les glandes et les phanères [30]. La Figure 1.4 présente les annexes de la peau

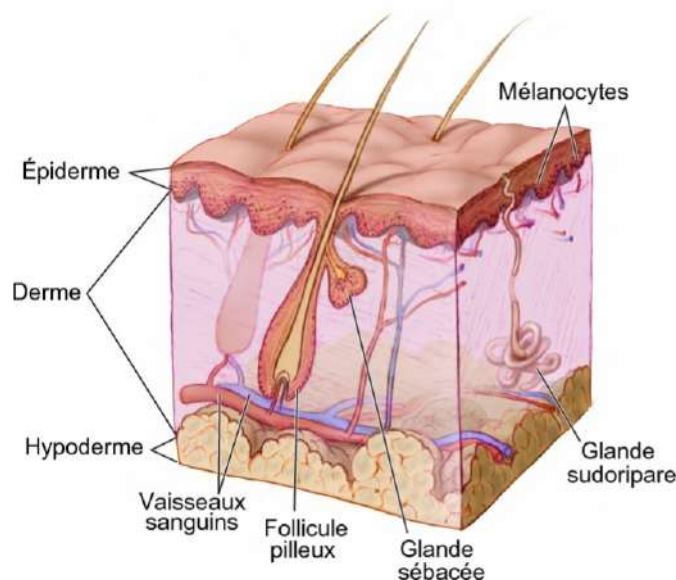


FIGURE 1.4 – Annexes de la peau [1].

Les glandes peuvent être divisées en trois catégories : les glandes sudoripares (ou sudorales) eccrines et apocrines, et les glandes sébacées.

- Les glandes sudoripares eccrines, qui fabriquent la sueur aqueuse ;

- Les glandes sudoripares apocrines, responsables de l’odeur corporelle
- Les glandes sébacées qui secrètent le sébum, ce film hydrolipidique qui protège l’épiderme ;

Les phanères (ou apparent) sont les cheveux, poils et ongles. Généralement, les glandes sébacées sont annexées aux poils, formant les follicules pilo-sébacés qui constituent la racine des poils et des cheveux. Les phanères protègent contre la chaleur, le froid et les ultraviolets [30]. Les phanères ont donc un rôle important également, mais pas vital.

### 1.1.2 Fonctions

La peau et ses annexes (par leur structure et leur composition), assurent plusieurs fonctions fondamentales et complexes [30, 31] dont :

- la protection contre les agressions extérieures (choc, pollution, agents infectieux, substances chimiques), les radiations lumineuses (ultraviolets), les entrées ou les sorties d’eau ;
- la thermorégulation en maintenant la température intérieure à 37°C. En cas de chaleur, la peau transpire pour rafraîchir le corps et en cas de froid, elle contracte le système vasculaire du derme pour conserver la chaleur ;
- la sensibilité à la pression, la chaleur et la douleur. Grâce aux terminaisons nerveuses et de récepteurs présents dans la peau, et qui réagissent en fonction des différents stimuli et renvoient des informations interprétables par le cerveau ;
- la synthèse de substances essentielles à l’organisme. Les kératinocytes sous l’action des UV participent à la synthèse de la vitamine D, et produisent aussi des endorphines qui interviennent dans la régulation de la thymique chez l’humain ;
- la fonction immunitaire grâce aux cellules de Langerhans présentatrices d’antigènes et aux kératinocytes qui produisent de nombreuses cytokines dont les cytokines pro-inflammatoires. Ces dernières interviennent dans la réaction inflammatoire cutanée.

La peau a aussi une fonction psychosociale. En effet la peau à travers sa couleur, sa texture et son odorat, permet de transmettre des messages sociaux comme l’état émotionnel (embarrasement, joie, colère, et stress) [31]. Aussi l’aspect de la peau (couleur, présence d’affection cutanée ou de cicatrice ou ride, éclat, élasticité, et bonne santé) influence négativement ou positivement la perception de la société et la confiance en soi.

### 1.1.3 Phototype cutané

L’épiderme est composé de plusieurs cellules (voir Section 1.1.1.1 dont les mélanocytes. Les mélanocytes produisent la mélanine (phéomélanine et eumélanine) qui donne à la peau

sa pigmentation constitutive. La répartition entre les phéomélanines et les eumélanines varie en fonction des individus et conditionne leur phototype cutané.

Fitzpatrick dans [2] a établi une technique de catégorisation des individus en six phénotypes cutanés. La technique se base sur la couleur constitutive de la peau et de sa réaction sous l'effet des rayons ultraviolets. La Figure 1.5 présente les six phénotypes cutanés, leurs couleurs constitutives et leurs réactions (bronzage, brûlure) aux UV.

					
<b>Phénotype I</b> Peau très claire (avec taches de rousseur)	<b>Phénotype II</b> Peau très claire (sans taches de rousseur)	<b>Phénotype III</b> Peau claire	<b>Phénotype IV</b> Peau Mate	<b>Phénotype V</b> Peau Foncée	<b>Phénotype VI</b> Peau Très Foncée
<b>Réaction aux UV</b> Brûle toujours Ne bronze jamais.	<b>Réaction aux UV</b> Brûle facilement Bronze peu et avec difficulté	<b>Réaction aux UV</b> Brûle peu Bronze progressivement	<b>Réaction aux UV</b> Brûle peu Bronze toujours bien	<b>Réaction aux UV</b> Brûle rarement Bronze intensément	<b>Réaction aux UV</b> Ne brûle jamais Bronze intensément et profondément

FIGURE 1.5 – Échelle de Fitzpatrick ou six phénotypes cutanés [2].

### 1.1.4 Particularités de la structure et des fonctions de la peau sur peaux noires

Les phototypes IV, V, et VI de l'échelle de Fitzpatrick encore appelés peaux noires présentent quelques particularités par rapport aux autres phototypes. Ces particularités sont citées ci-dessous [32, 33] :

- une forte densité de la couche cornée qui explique une meilleure résistance des peaux noires aux coupures et aux piqûres ;
- une plus grande quantité d'eumélanines. L'eumélanine sous l'effet des UV joue un rôle photo protecteur contrairement à la phéomélanine qui est carcinogène sous l'action des UV. Le risque de cancer est donc plus élevé sur les peaux de type I, II et III.

## 1.2 Familles de lésions cutanées

Les lésions cutanées sont des signes anormaux qui surviennent en surface ou sous la peau. Elles peuvent être congénitales ou survenir après la naissance et sont caractérisées par leur morphologie, couleur, texture (rugueux, lisse, squameux, ou cireux). Il existe trois classes de lésions cutanées [20] :

- les lésions élémentaires qui comprennent les macules (rouges, pigmentaires, achromiques), papules, nodules, végétations, et les lésions liquidiennes (vésicules, bulles, pustules) ;

- les lésions élémentaires secondaires ;
- les tumeurs cutanées.

Les tumeurs cutanées se développent en partant de l'épiderme, à partir des éléments constitutifs du derme ou à partir de cellules anormalement présentes dans la peau. Elles peuvent être uniques ou multiples sur un individu et sont bénignes ou malignes.

### 1.2.1 Tumeurs bénignes

La catégorie tumeurs bénigne (non cancéreuse) regroupe l'ensemble des lésions de la peau qui ne se propagent pas (pas de métastases) à d'autres parties du corps [20]. Ces tumeurs sont classées comme sans conséquences graves sur l'organisme et ne requièrent généralement pas de traitement, mais dans certains cas, elles sont extraites par chirurgie. Les tumeurs bénignes les plus fréquentes sont :

- la kératose séborrhéique ;
- le lentigo solaire ;
- le nævus mélanocytaire acquis ;
- le dermatofibrome ;
- le kérato-acanthome.

Le nævus mélanocytaire congénital et la kératose actinique (KA) qui sont des formes particulières de nævus et de kératose, présentent un haut risque de transformation en tumeurs malignes. Dans ces cas, une surveillance dermatologique régulière est recommandée pour éviter toute complication possible.

### 1.2.2 Tumeurs malignes

Les tumeurs malignes, regroupent l'ensemble des lésions cancéreuses de la peau [20]. Un cancer cutané est une prolifération excessive et incontrôlée de cellules situées dans la peau. Le type de prolifération, la présentation et le siège initial, sont les caractéristiques majeures qui permettent de classer les tumeurs malignes. On distingue trois types de tumeurs malignes :

- les mélanomes qui se développent à partir de mélanocytes. Ces lésions se propagent de manière horizontale (sur l'épiderme) ou aussi de manière verticale (dans les couches profondes de la peau) [20, 34]. Les mélanomes se métastasent à des stades avancés et présentent un fort taux de mortalité ;
- les carcinomes épidermoïdes (ou carcinomes spinocellulaires) qui se développent à partir de l'épiderme. Ces lésions se propagent très vite à la surface de la peau et



peuvent aussi atteindre d'autres organes [20]. Après les mélanomes, les carcinomes épidermoïdes sont classés en deuxième position des cancers les plus mortels.

- les carcinomes basocellulaires qui se développent à partir des cellules basales. Cette forme de tumeur maligne est la plus fréquente. Les carcinomes basocellulaires présentent un faible risque de métastases et donc un faible risque de mortalité chez le patient atteint [20].

### 1.2.3 Particularités lésions de la peau sur peaux noires

Le cancer de la peau est moins fréquent chez les sujets de peaux noires à cause de la forte teneur en mélanine constatée sur ces types de sujets qui leur offre une meilleure photo-protection (voir Section 1.1.4). Cependant, il est associé à une plus grande mortalité puisque le diagnostic est souvent posé à un stade avancé [32]. Différentes études [32, 35, 36, 37, 38] réalisées par des spécialistes ont analysé la présentation clinique de lésion cancéreuses pour tous les types de peau. Les résultats de ces études révèlent qu'il existe des différences au niveau de la location, de la pigmentation et des facteurs. Les carcinomes basocellulaires chez les personnes à peau noires sont caractérisés par une forte pigmentation, se présentent sous forme de papules et sont localisés sur la tête et le cou et parfois sur le périnée [37]. Chez ces mêmes sujets, les carcinomes spinocellulaires sont caractérisés par une hyperpigmentation ou une hypopigmentation, et se présentent sur les membres inférieurs, la tête, le cou, ou les organes génitaux et résulte généralement d'ulcère, de cicatrice ou de traumatisme inflammatoire chronique [36, 38]. Enfin, en ce qui concerne la position, le mélanome sur peaux noires est généralement acral [35]. Toutes ces spécificités montrent que le diagnostic des lésions cutanées nécessite une certaine expérience, surtout dans le cas des peaux noires.

## 1.3 Diagnostic

Le diagnostic de lésions cutanées se fait suivant une procédure bien définie qui commence par un examen clinique. En cas de suspicion de forme grave de cancer de peau, un examen de seconde intention est recommandé : biopsie. Il existe aussi d'autres techniques d'imagerie aussi appelées biopsies optiques qui sont utilisées pour le diagnostic des lésions cutanées.

### 1.3.1 Examen clinique

L'examen clinique consiste en une analyse sémiologique de la lésion. Cet examen est réalisé par un médecin généraliste ou un dermatologue et comprends deux étapes : l'interrogatoire et l'examen physique dermatologique.

L'interrogatoire permet de recueillir des informations essentielles (voire indispensables) pour poser le bon diagnostic. Ces informations concernent :

- le contexte d'apparition des lésions : signes extra-cutanés associés, comorbidités, expositions (solaire, agents aéroportés, contagio infectieux), prise médicamenteuse ;
- la date de début, le siège initial, le type de douleur, et les modalités évolutives de la lésion ;
- les antécédents personnels et familiaux ;

L'examen physique dermatologique consiste en inspection visuelle (à l'œil nu, ou avec une loupe grossissante) de la lésion. Cet examen est réalisé par un médecin généraliste ou un dermatologue, et nécessite une certaine connaissance des lésions cutanées, de leur configuration et de leur distribution.

Une lésion est suspecte lorsqu'elle pique ou saigne, se répète ou ne guérit pas, ou présente les critères de l'ABCDE [39] du mélanome. L'acronyme ABCDE signifie :

- Asymétrie ;
- Bord irrégulier, encoché, polycyclique ;
- Couleur hétérogène : du brun clair au noir foncé ; des parties blanches non pigmentée, des parties rouges inflammatoires ou des parties cicatricielles bleues ;
- Diamètre supérieur 6 mm ;
- Evolutivité dans sa taille (la lésion s'élargit), sa couleur ou son épaisseur.

Une lésion mélanocytaire est considérée comme suspecte si deux de ces critères sont validés (le critère E étant le critère le plus pertinent) [40]. L'examen clinique est une étape non invasive et indispensable pour une prise en charge médicale du sujet. Néanmoins la précision du diagnostic dépend de l'expérience du spécialiste [41].

### 1.3.2 Examen histologique

L'histologie est un examen de seconde intention réalisé pour confirmer ou infirmer un diagnostic. Cette méthode est invasive, et consiste à prélever du tissu situé au-dessous de la lésion suspecte (c'est la biopsie), puis à analyser au microscope le tissu prélevé [1].

L'histologie bien qu'étant une méthode minimalement invasive qui améliore la précision du diagnostic des cancers de la peau présente quelques inconvénients. En premier, l'analyse en laboratoire des tissus prends du temps ce qui retarde la prise en charge et entraîne de l'anxiété chez le sujet. Aussi, la réalisation de l'examen histologique nécessite d'avoir un matériel adapté, et un savoir-faire ce qui le rends moins accessible. Enfin, d'après [12] près de 60 % des biopsies cutanées réalisées aboutissent à des diagnostics bénins. En effet, selon [42], certains spécialistes réalisent systématiquement des biopsies pour éviter d'avoir des faux négatifs (c'est-à-dire un sujet présentant une affection maligne diagnostiquée à

tort comme étant bénigne). Cette pratique bien que limitant le nombre de faux négatifs, est douloureuse, laisse une cicatrice, et occasionne un stress inutile chez le sujet en cas de faux positif. Une approche de plus en plus courante pour réduire le nombre de biopsies est la biopsie optique, qui utilise différentes modalités d'imagerie de la peau pour observer la peau et sa composition interne.

### 1.3.3 Modalités d'imagerie de la peau

La littérature propose différentes méthodes d'imagerie pour visualiser la structure des tissus. Ces méthodes appliquées en dermatologie ont le principal avantage d'être non-invasives, fournissent plus d'informations sur la structure de la peau et leurs mesures peuvent être numérisées sous forme d'images. La possibilité d'enregistrer ces images facilite donc la documentation et le suivi des lésions cutanées, et encourage le traitement informatisé des données pour améliorer le diagnostic. La dermoscopie, la microscopie confocale par réflectance (MCR) et la spectroscopie par réflectance sont quelques-unes des méthodes qui sont actuellement utilisées pour examiner spécifiquement les lésions de la peau.

#### 1.3.3.1 Dermoscopie

La dermoscopie (ou la microscopie en épiluminescence) est une technique non invasive, utilisée pour l'examen des lésions cutanées [43]. Il existe deux types de dermatoscope :

- le dermatoscope à lumière non-polarisée consiste à appliquer une interface liquide (gel particulier) à la couche cornée pour réduire les reflets de la lumière. Cette interface favorise une meilleure pénétration de la lumière dans la peau. La lumière diffusée par les couches en dessous de la cornée permet l'observation directe des structures dermoscopiques sous-jacentes.
- le dermatoscope à lumière polarisée réduit la visualisation de la lumière réfléchi par la couche superficielle grâce à l'utilisation de deux polariseurs aux axes orthogonaux (se coupant à  $90^\circ$ ). Le polarisateur du détecteur rejette la lumière unidirectionnelle polarisée réfléchi par les couches superficielles et ne laisse passer que les lumières renvoyées des couches profondes. C'est le principe de la polarisation croisée et il permet de visualiser des couches plus profondes de la peau (de l'épiderme, la jonction dermo-épidermique, et aussi du derme papillaire) qui ne sont pas visibles à l'œil.

Le dermatoscope est souvent relié à un appareil photo et à un ordinateur, ce qui permet de sauvegarder les images dermoscopiques et facilite le suivi des lésions (on parle de système de dermoscopie digitale). La Figure 1.6 présente deux systèmes de dermoscopie digitale et quelques images obtenues d'un dermatoscope à lumière polarisée.

Le dermatoscope est largement déployé dans les pays développés à cause de sa portabilité, son coût d'achat relativement abordable, son utilisation à temps réel. Le dermatoscope permet aussi de diminuer le nombre d'excisions de lésions bénignes.

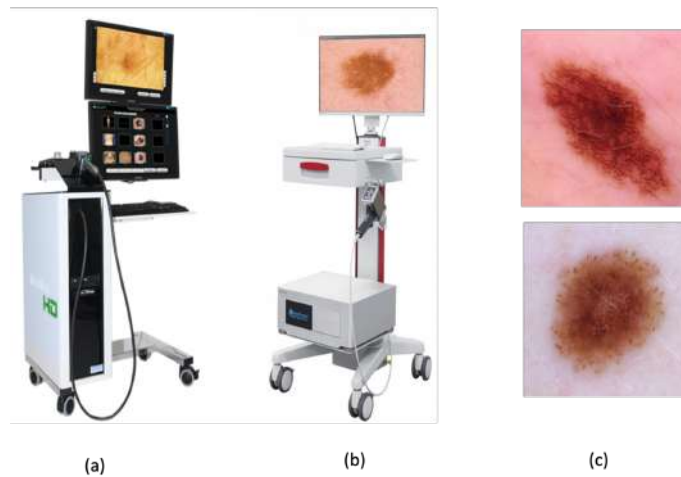


FIGURE 1.6 – Exemple de dermatoscopes et d’images dermoscopiques. (a) Molemax HD (de derma medical system, Vienne Autriche)[1], (b) FotoFinder (FotoFinder Systems GmbH, Bad Birnbach, Allemagne)[1], (c) Images dermoscopiques.

### 1.3.3.2 Microscopie confocale par réflectance (MCR)

Le MCR comme présenté dans la Figure 1.7(a) est principalement composé de :

- sources lumineuses monochromatiques et cohérentes (des lasers émettant à différentes longueurs d’ondes) ;
- un trajet optique (composé de diaphragme qui est un trou de faible diamètre, de lentilles, d’objectif, de détecteurs à faible bruit, de miroirs à balayage rapide, et de filtres) qui permet de recueillir sélectivement la lumière provenant de la région d’intérêt, tout en supprimant la lumière des régions au-dessus et au-dessous du plan focal ;
- un ou plusieurs capteurs (photomultiplicateurs).

Le principe de la MCR repose sur la réflexion de la lumière par les molécules endogènes de la peau. La lumière émise par le laser est réfléchiée par le tissu, puis reçue par l’objectif et transmise à un détecteur, après passage par le sténopé. Le sténopé permet de récupérer uniquement la lumière provenant du plan focal, d’où l’obtention d’images en coupes. L’image complète est ensuite reconstruite point par point en faisant balayer le laser à la surface de la peau.

La MCR permet d’obtenir en temps réel des images horizontales de la peau à une résolution cellulaire depuis de la cornée jusqu’au derme papillaire. Cette technique permet de représenter le tissu dans son état physiologique et favorise le suivi de l’évolution d’une lésion cutanée. La MCR est principalement utilisée pour le diagnostic de lésions mélanocytaires [44].

Le microscope confocal présenté dans la Figure 1.7(b) le plus répandu (Vivascope 1500, Lucid Inc, Rochester, NY, Etats-Unis) est équipé d’un laser diode à 830 nm de faible puis-

sance de 5-10 mW sans risque tissulaire. Il existe aussi un modèle multi laser (488 nm, 658 nm, 785 nm) qui permet l'utilisation de colorants fluorescents tels que la fluorescéine pour améliorer le contraste des images.

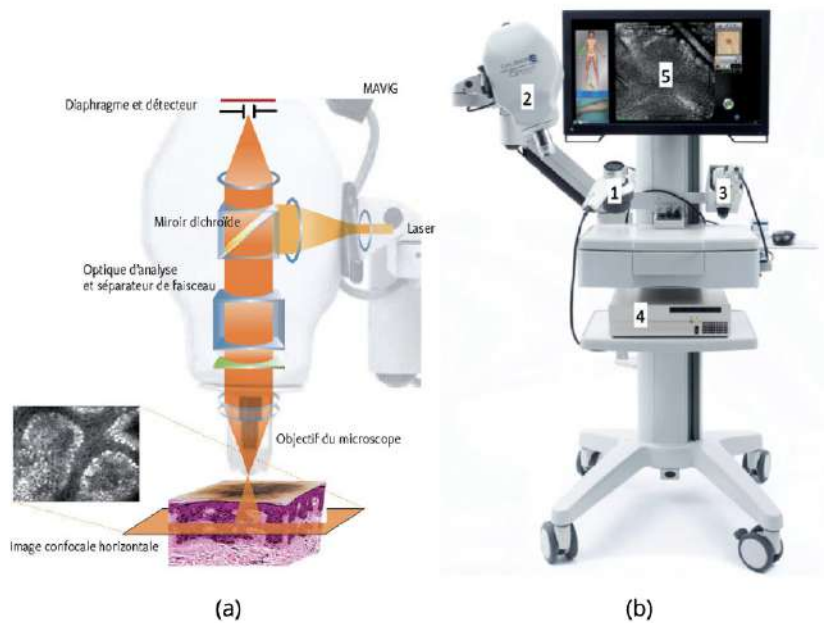


FIGURE 1.7 – (a) Principe de fonctionnement du MCR [3]. (b) Le système Vivascope 1) Vidéodermoscopie ; 2) Microscope confocal VS1500 avec bras articulé ; 3) Microscope confocal manuel VS3000 ; 4) PC ; 5) Écran [3].

### 1.3.3.3 Spectroscopie par réflectance

La spectroscopie par réflectance diffuse ou DRS (Diffuse reflectance spectroscopy) est une méthode non-invasive qui permet de déterminer des informations sur la composition biochimique (la concentration en hémoglobine et l'oxygénation du sang par exemple) et sur la structure des tissus. La Figure 1.8 présente le principe de fonctionnement de la DRS.

La DRS consiste en premier à illuminer la peau avec différentes longueurs d'onde de lumière visible et infrarouge et à détecter ensuite le modèle spectral réfléchi après des événements successifs de diffusion et d'absorption [44, 3]. Les différents composants de la peau produisent des motifs spectraux différents, qui sont analysés par un programme informatique pour construire une image numérique représentative de chaque structure. Cette méthode est principalement utilisée pour différencier les nævus bénins des mélanomes.

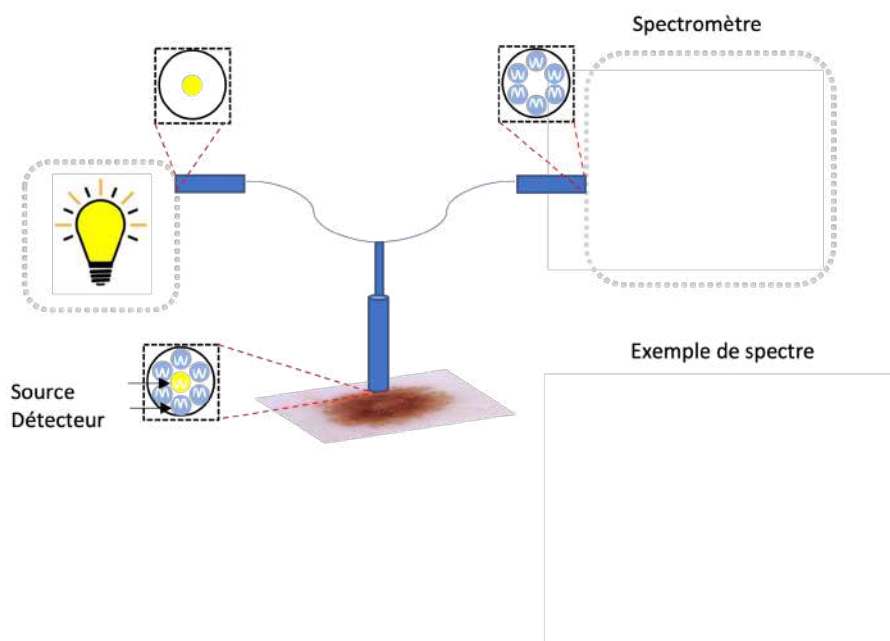


FIGURE 1.8 – Principe de fonctionnement de la DRS [1].

#### 1.3.3.4 Tomographie en cohérence optique

La tomographie en cohérence optique (TCO) est une méthode d'imagerie non-invasive basée sur l'interférométrie à faible cohérence [44, 45]. La Figure 1.9 présente le principe de fonctionnement de la TCO.

La TCO utilise un séparateur pour diviser la source de lumière (infrarouge) en faisceau de référence et en faisceau d'observation. Le faisceau de référence est dirigé vers un miroir ajustable et le faisceau d'observation est balayé la zone d'intérêt de la peau. La mesure des interférences du faisceau de référence avec la lumière réfléchiée par les éléments internes de la peau permet de produire en temps réel une image dont la profondeur de pénétration est de l'ordre du millimètre, avec une résolution proche du micromètre. Les images obtenues sont orientées verticalement comme les coupes histopathologiques.

La TCO est une méthode non invasive qui permet d'avoir en temps réels des informations structurales. La TCO est principalement recommandée pour le diagnostic de kératoses actiniques et les CBC et ne convient pas au diagnostic des lésions pigmentaires comme les mélanomes [46]. L'option dynamique de la TCO permet l'analyse des vaisseaux sanguins en temps réel et peut donc être utile aussi dans le suivi des mélanomes.

#### 1.3.4 Particularités du processus de diagnostic sur les peaux noires

Le processus de diagnostic de lésions cutanées sur peaux noires est le même sur tous les types de peaux. La stratégie de diagnostic commence par l'examen clinique et en cas de suspicion de forme grave de cancer de peau la biopsie est demandée. Néanmoins, une

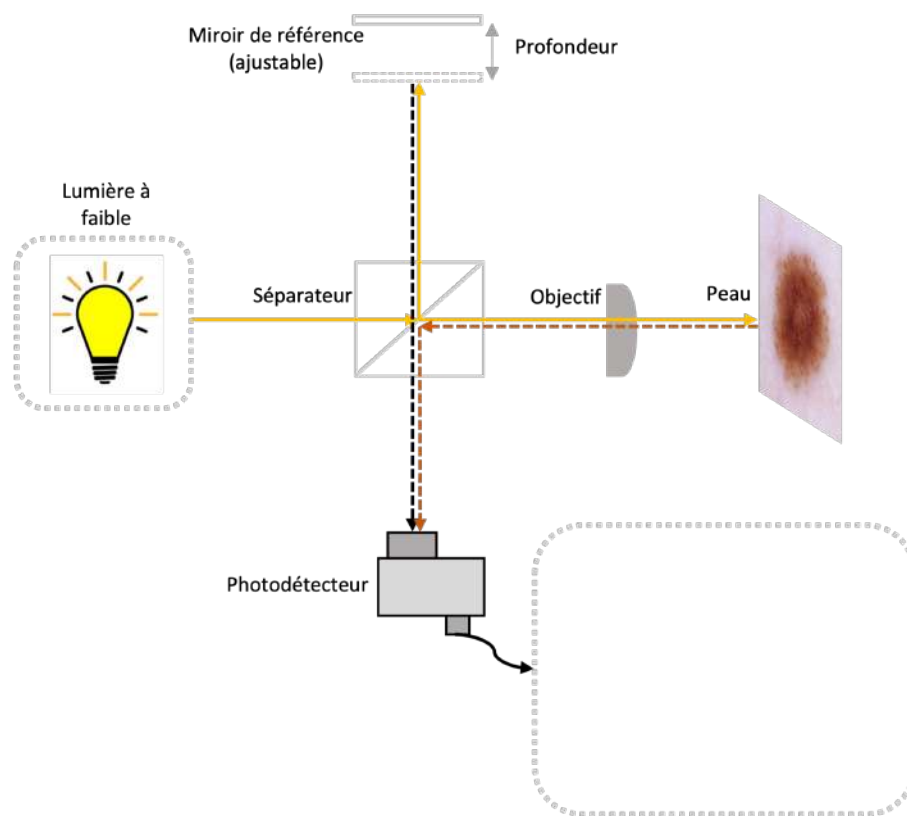


FIGURE 1.9 – Principe de fonctionnement de la TCO.

connaissance des particularités de présentation clinique des lésions cutanées sur peaux génétiquement hyper-pigmentées, est indispensable pour améliorer la précision de diagnostic.

## Conclusion

Ce chapitre présente en premier la peau, les différentes couches qui la compose, et les fonctions qu'elle assure. Ensuite nous avons présenté les deux familles d'affections de la peau, le processus de diagnostic des lésions de la peau et quelques systèmes d'imagerie de la peau. A chacune des trois étapes du chapitre, les particularités observées sur peaux génétiquement hyper-pigmentées sont mises en évidence.





## CHAPITRE 2

# CLASSIFICATION D'IMAGES DERMOSCOPIQUES

### Sommaire

---

<b>2.1</b>	<b>Classification basée sur l'apprentissage automatique traditionnel . . .</b>	<b>22</b>
2.1.1	Pré-traitement . . . . .	23
2.1.2	Extraction manuelle de caractéristiques . . . . .	23
2.1.3	Classification . . . . .	25
<b>2.2</b>	<b>Classification basée sur l'apprentissage profond . . . . .</b>	<b>31</b>
2.2.1	Apprentissage à partir de zéro . . . . .	32
2.2.2	Apprentissage par transfert . . . . .	32
2.2.3	Extraction automatique de caractéristiques avec les réseaux de neurones . . . . .	33
<b>2.3</b>	<b>Évaluation des outils d'aide au diagnostic . . . . .</b>	<b>33</b>
2.3.1	Méthodes de validation . . . . .	33
2.3.2	Mesure de performance . . . . .	33
<b>2.4</b>	<b>Comparaison des méthodes de classification . . . . .</b>	<b>36</b>
<b>2.5</b>	<b>Méthode proposée : GCN . . . . .</b>	<b>37</b>
2.5.1	GCN . . . . .	37
2.5.2	Méthode . . . . .	38
2.5.3	Résultats expérimentaux et discussions . . . . .	40

---

### Introduction

Comme présenté dans la Figure 1 de l'introduction générale, ces deux dernières décennies, plusieurs travaux scientifiques ont contribué à améliorer le diagnostic de lésions cutanées en proposant des outils d'aide au diagnostic. Ces outils permettent de faire une classification automatique des lésions cutanées et aident les experts et les non-experts à se concentrer sur les lésions malignes ce qui favorise le suivi des patients. Les méthodes de classification automatique de lésions cutanées sont principalement basées

sur deux approches de l'intelligence artificielle : l'apprentissage automatique traditionnel (traditional machine learning) et l'apprentissage profond (deep learning).

L'apprentissage automatique est la capacité qu'a l'ordinateur d'apprendre à partir de données à effectuer une tâche sans être explicitement programmés. Lorsque l'algorithme d'apprentissage automatique à une structure simple (comme une régression linéaire ou un arbre de décision), cet apprentissage est aussi appelé apprentissage traditionnel. L'apprentissage profond est basé sur le réseau de neurones artificiel qui a une structure complexe. Il permet de traiter des données non structurées telles que des documents, des images et du texte. La 2.1 illustre les principales étapes de ces deux approches ; avec au-dessus les étapes de l'apprentissage automatique traditionnel et au-dessous, les étapes de l'apprentissage profond.

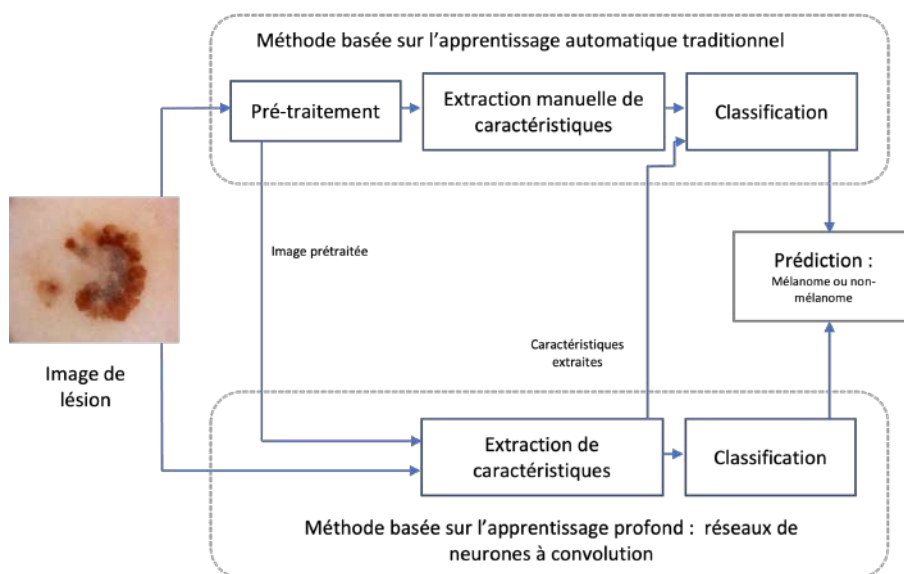


FIGURE 2.1 – Schématisation des différentes étapes des méthodes de classification automatique de lésions cutanées.

Dans la suite de ce chapitre, nous présentons les deux types de méthodes de classification automatique de lésions cutanées, les approches d'évaluation de ces méthodes et la méthode proposée pour améliorer la classification automatique d'images de cancers de la peau .

## 2.1 Classification basée sur l'apprentissage automatique traditionnel

Les principales étapes des méthodes basées sur l'apprentissage automatique traditionnel sont : le pré-traitement, l'extraction manuelle de caractéristiques, et la classification.

### 2.1.1 Pré-traitement

Le pré-traitement de l'image est une étape qui permet de supprimer ou réduire les distorsions indésirables et d'améliorer certaines caractéristiques de l'image importantes pour le post-traitement. Cette étape implique principalement trois éléments :

- améliorer la qualité de l'image par l'élimination des effets d'éclairage variables [47, 48] la conversion de l'image dans un espace couleur différent [9], ou l'amélioration du contraste [10] ;
- restaurer l'image [11] ;
- éliminer les poils gênants présents dans l'image [48].

Une combinaison adéquate des étapes de prétraitement peut jouer un rôle important dans la classification précise des lésions.

### 2.1.2 Extraction manuelle de caractéristiques

L'extraction manuelle de caractéristiques nécessite d'identifier et de décrire les caractéristiques pertinentes pour un problème donné et de mettre en œuvre un moyen d'extraire ces caractéristiques. Cette étape nécessite donc une certaine compréhension du contexte ou du domaine pour mieux sélectionner les caractéristiques utiles.

Le processus d'extraction de caractéristiques peut être global sur toute l'image ou local sur un nombre restreint de pixels appelé zone d'intérêt, et consiste à appliquer des transformations mathématiques sur l'image ou la zone d'intérêt afin d'obtenir des valeurs dérivées qui permettent de mieux rendre compte de certaines propriétés de l'image. Les caractéristiques extraites sont généralement liées à la morphologie, la couleur et la texture des images.

#### 2.1.2.1 Les caractéristiques de morphologie

Décrire la forme d'une lésion cutanée nécessite au préalable une identification de régions affectées dans l'image. L'identification se fait par la segmentation de l'image ou la détection de contours [48]. Décrire la forme d'une lésion cutanée consiste à extraire des caractéristiques comme le diamètre estimé (estimé car les images peuvent être acquises à différents niveaux de grossissement), la symétrie, l'irrégularité des bords, l'excentricité. La règle ABCDE [39], la règle de vérification en 7 points [49], et l'indice de Menzies [50], sont quelques approches utilisées pour extraire des caractéristiques dont celles relatives à la morphologie de lésions cutanées.

La Figure 2.2 présente des images dermoscopiques qui permettent d'observer quelques caractéristiques morphologiques de lésions cutanées

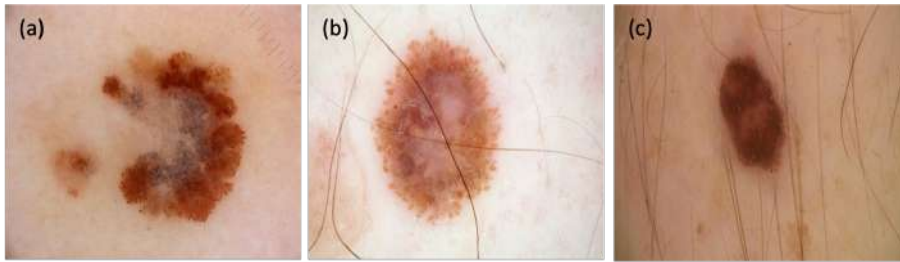


FIGURE 2.2 – a) Exemple d’une image dermoscopique de mélanome asymétrique, à bords irréguliers, et excentrique, (b) Exemple d’une image dermoscopique de mélanome plutôt symétrique, et à bords irréguliers, (c) Exemple d’une image dermoscopique de nevus plutôt symétrique à bords réguliers [4, 5, 6].

### 2.1.2.2 Les caractéristiques de couleur

La diversité de couleur constitue un élément important du diagnostic des maladies cutanées. La présence d’une couleur sur la peau traduit une augmentation de chromophores comme le pigment (brun, noir, gris, ou bleu), les lipides ou la kératine (jaune), le collagène (blanc), ou le sang (rouge) [51].

En imagerie la couleur est caractérisée par plusieurs outils comme : l’histogramme des couleurs et les moments de couleurs. La règle ABCDE [39], la règle de vérification en 7 points [49], et l’indice de Menzies [50], sont aussi utilisées, pour extraire les caractéristiques de couleur dans les images de lésions cutanées.

En dermoscopie la couleur permet de déterminer la profondeur de la lésion [52]. En effet depuis le dermoscope, la mélanine apparaît noire lorsqu’elle est située dans la couche cornée, brune lorsqu’elle est située dans l’épiderme moyen ou inférieur, et bleu-grise lorsqu’elle est présente dans le derme.

### 2.1.2.3 Les caractéristiques de texture

Selon Haralick dans [53] la texture peut être décrite statistiquement, en rendant compte de la manière dont les niveaux de gris de l’image s’organisent les uns par rapport aux autres. Il propose des mesures des caractéristiques texturales dont le contraste, la corrélation, l’entropie et la variance.

La texture peut aussi être décrite par une analyse spectrale de l’image. L’analyse spectrale consiste à appliquer des transformées comme la transformée de Fourier discrète ou un ensemble de filtres de Gabor disposés à plusieurs échelles et orientations pour décrire la texture de l’image.

En dermoscopie, les structures comme le réseau pigmentaire, les regroupements de points, de globules et de stries permettent de décrire la texture des lésions. Le réseau pigmentaire est constitué de lignes brunes qui se croisent et forment un motif réticulaire en forme de grille [54]. Le motif formé peut-être régulier (typique) ou irrégulier (atypique). Les points [54] sont de petites structures rondes et bien circonscrites (moins de 0,1 mm). Ces points

sont dits réguliers s'ils sont groupés au centre de la lésion, ou situés sur les lignes du réseau pigmentaire. Les globules [55] sont des structures rondes à ovales bien délimitées, d'une taille supérieure à 0,1 mm. Lorsque les points et les globules sont atypiques (différences de taille, de forme, de couleur ou de répartition), le diagnostic de mélanome est privilégié. Les stries [55] sont des projections pigmentées linéaires qui se localisent au bord des lésions et comprennent les stries radiales (stries linéaires) et les pseudopodes (projections bulbeuses). Certaines lésions présentent des zones sans de structures. Ces zones représentent plus de 10% de la lésion et sont caractérisées par une couleur (noir, brun, bleu, ou rouge) qui prédomine sur toutes les autres couleurs. Les zones sans structure peuvent être hypo-pigmentées, hyper-pigmentées ou normalement pigmentées. Les lésions qui ne présentent qu'un motif sans structure sont difficiles à diagnostiquer par dermoscopie. La Figure 2.3 présente quelques exemples de structures dermoscopiques.

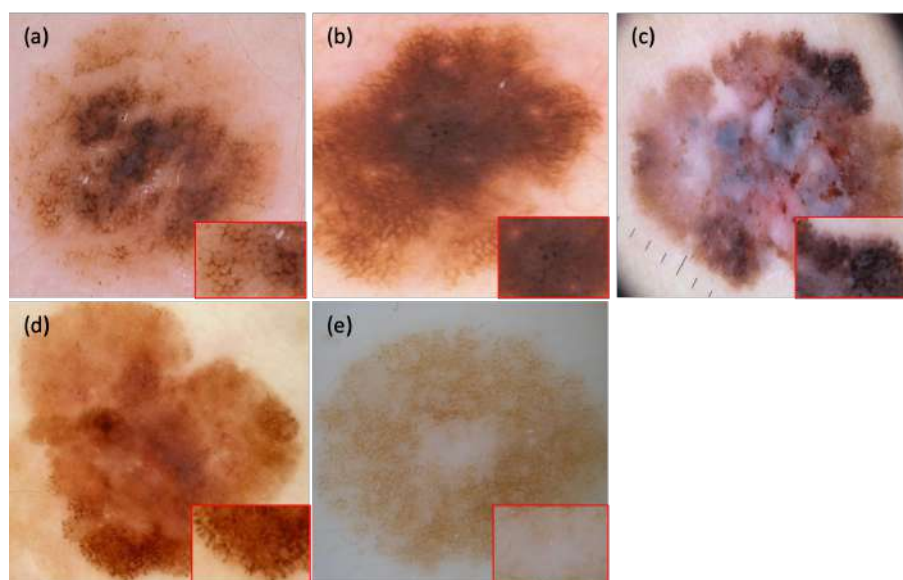


FIGURE 2.3 – (a) Exemple de réseau pigmentaire atypique sur une image dermoscopique de nevus, (b) Exemple de points irréguliers sur une image dermoscopique de nevus, (c) Exemple de globules irréguliers sur une image dermoscopique de mélanome, (d) Exemple de stries sur une image dermoscopique de mélanome, (e) Exemple de zone sans structure sur une image dermoscopique de nevus [4, 5, 6].

### 2.1.3 Classification

La classification consiste à catégoriser et à attribuer des étiquettes à une image à partir de caractéristiques extraites de l'image. Les outils de classification d'images sont principalement divisés en deux catégories : les outils de classification d'images supervisées et non supervisées.

Les outils de classification supervisés comme les k voisins les plus proches (kNN) [56], les

machines à vecteurs de support (SVM)[48, 57, 58], les réseaux de neurones artificiels [59], sont les plus utilisés pour la classification de lésions cutanées.

### **2.1.3.1 Voisins les plus proches**

Les ‘k plus proches voisins’ (K-Nearest Neighbors K-NN) est une méthode d’apprentissage supervisée qui est utilisée pour la classification et la régression [56]. Pour classer un exemple inconnu, l’algorithme du k-NN trouve en premier les voisins les plus proches parmi les données d’apprentissage. Après avoir localisé les k voisins les plus proches, il exécute une règle de vote majoritaire pour trouver la classe qui est apparue le plus souvent. La classe qui est apparue le plus souvent dans le voisinage correspond à la classe finale de la donnée.

Une particularité de ce classifieur est que toutes les observations de l’ensemble d’apprentissage restent en mémoire pendant la classification ce qui ralentit le processus. La performance du classifieur K-NN est lié au paramètre K qu’il convient de définir en fonction des observations. Des travaux scientifiques comme [56, 60], proposent des méthodes de classification de lésions cutanées basées sur les K-NN.

### **2.1.3.2 Machines à vecteurs de support**

Les machines à vecteurs de support (Support Vector Machine SVM) sont un ensemble de techniques d’apprentissage supervisé qui reposent sur deux notions : la marge maximale et la fonction noyau.

Le principe des SVM tel que défini dans [61] consiste à résoudre un problème de classification ou de discrimination en le ramenant à un problème linéaire de recherche d’un hyperplan optimal. Les données dans ce hyperplan sont séparées en plusieurs classes dont la frontière est la plus éloignée possible des points de données (ou marge maximale). Pour rendre les données linéairement séparables, les machines à vecteurs de support introduisent une fonction noyau (kernel) dans le produit scalaire induisant implicitement une transformation non linéaire des données vers un espace intermédiaire (feature space) de plus grande dimension. La fonction noyau définit un espace auto-reproduisant et isométrique par la transformation non linéaire de l’espace initial et dans lequel est résolu le problème linéaire.

Des travaux scientifiques comme [48, 57, 58], proposent des algorithmes de classification de lésions cutanés basés sur SVM.

### **2.1.3.3 Réseaux de neurones artificiels**

Le réseau de neurones artificiels est composé de neurones artificiels qui sont inspirés des neurones biologiques. Chaque neurone artificiel a des entrées et produit une sortie unique qui peut être envoyée à plusieurs autres neurones. A chaque neurone artificiel est

associé un poids de connexion (ou poids synaptique) nommé  $w$  (weight). La structure du neurone artificiel est schématisée dans la Figure 2.4.

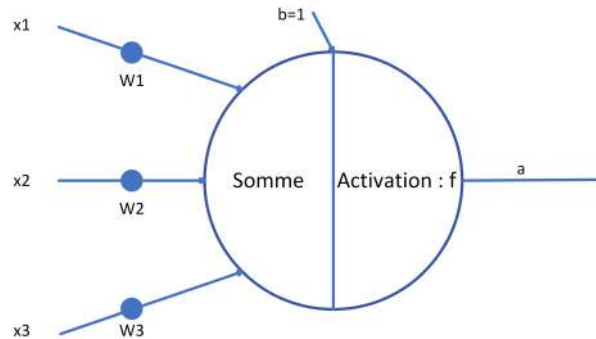


FIGURE 2.4 – Structure d'un neurone artificiel.

La sortie d'un neurone tel que défini par l'Équation 2.1 est obtenue en deux phases. La première phase consiste à calculer la somme pondérée des entrées à laquelle est ajouté un biais  $b$ . Ce biais peut être considéré comme un neurone externe supplémentaire qui envoie systématiquement le signal 1 de poids  $b$  au neurone. La deuxième phase consiste à activer le neurone en appliquant une fonction d'activation (encore appelée fonction de transfert). L'activation d'un neurone artificiel est liée au potentiel d'action du neurone biologique et représente le seuil à partir duquel un neurone émet un signal aux neurones suivants (s'il a été suffisamment stimulé).

$$a = f_{activation} \sum_{i=1}^n w_i x_i + b \quad (2.1)$$

Le réseau de neurones artificiels s'organise en plusieurs couches à savoir une couche d'entrée, une couche de sortie et zéro à plusieurs couches cachées [62]. Les neurones d'une couche sont entièrement et uniquement connectés aux neurones des couches immédiatement précédentes et immédiatement suivantes. La couche d'entrée reçoit les données externes (les caractéristiques extraites d'une image par exemple). La couche de sortie produit le résultat final. Entre ces deux couches se trouvent zéro ou plusieurs couches cachées. La Figure 2.5 illustre les interconnexions entre les différentes couches d'un réseau de neurones artificiels.

Avec :  $L$  le nombre de couches du réseau de neurones ;  $w_{ij}^l$  le poids qui relie le  $i^{eme}$  neurone de la  $l^{ieme}$  couche au  $j^{ieme}$  neurone de la  $l - 1^{ieme}$  couche.

### Fonction d'activation

Les fonctions d'activations comme la sigmoïde [63], la tangente hyperbolique (Tanh) [63], la fonction Unité Linéaire Rectifiée (ou ReLU pour Rectified Linear Unit) [63] sont les plus courantes et leurs tracés sont représentés sur la Figure 2.6.

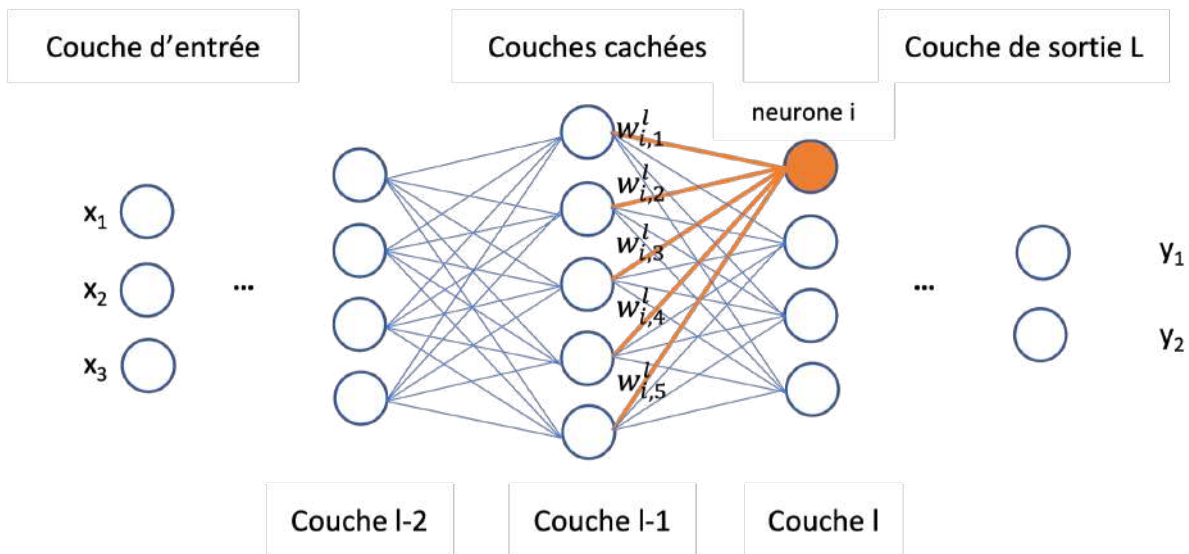


FIGURE 2.5 – Architecture d'un réseau de neurones artificiels.

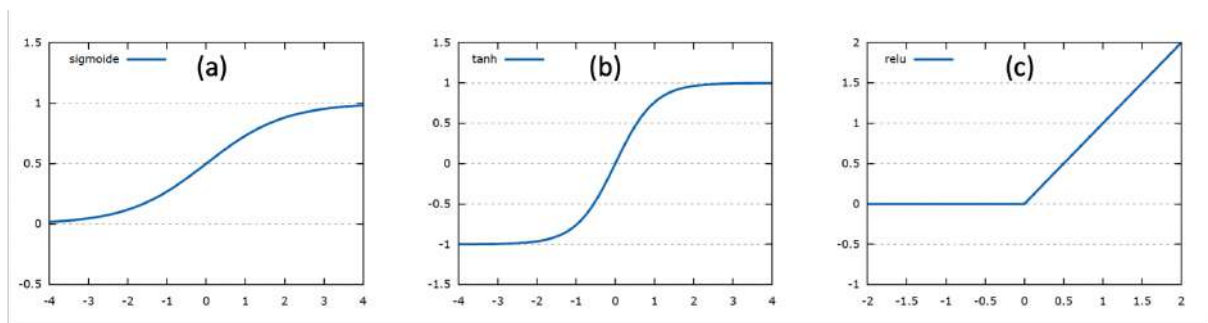


FIGURE 2.6 – Fonctions d'activation les plus connues. (a) Sigmoïde, (b) Tangente hyperbolique ( $\tanh$ ), et (c) ReLU (Rectified Linear Unit).

La fonction sigmoïde définie par l'Équation 2.2 transforme toutes ses entrées en des valeurs comprises entre 0 et 1. Les entrées plus grandes que 1 sont transformées en la valeur 1, de même les valeurs plus petites que 0 sont transformées en 0. La fonction tangente hyperbolique ( $\tanh$ ) définie par l'Équation 2.3 est une fonction d'activation non linéaire qui produit des valeurs comprises entre  $-1$  et  $1$ .

$$f(z) = \frac{1}{1 + e^{2z}} \quad (2.2)$$

$$f(z) = \frac{1 - e^{2z}}{1 + e^{2z}} \quad (2.3)$$

Le principal inconvénient de ces fonctions est la saturation. Une fois les neurones saturés (activé ou désactivé), les poids synaptiques de ces neurones ne sont plus mis à jour pendant la phase d'apprentissage pour améliorer les performances du réseau de neurones. L'erreur est rétro propagée dans les autres couches du réseau et utilisée pour la mise à jour



des poids. Ce problème est connu sous le nom du gradient évanescent (vanishing gradient) [64]. Une solution pour corriger le problème du gradient évanescent est l'utilisation de la fonction d'activation linéaire rectifiée (ou ReLU pour Rectified Linear Unit) [65]. La fonction ReLU définie par l'Équation 2.4 consiste à remplacer toutes les valeurs négatives reçues en entrées par des zéros. L'avantage d'utiliser la fonction ReLU est que tous les neurones ne sont pas activés en même temps. Il existe aussi d'autres versions améliorées de la fonction ReLU.

$$f(z) = \begin{cases} 0 & \text{si } z < 0 \\ 1 & \text{si } z \geq 0 \end{cases} \quad (2.4)$$

Une autre fonction d'activation généralement appliquée à la couche de sortie du réseau de neurones est le softmax. La fonction softmax définie par l'Équation 2.5, transforme un vecteur de  $K$  valeurs réelles en un vecteur de  $K$  valeurs réelles comprise entre 0 et 1, et dont la somme est égale à 1.

$$f_{(z)_i} = \frac{e^{z_i}}{1 + e^{2z}} \quad (2.5)$$

avec  $i = 1, \dots, K$

### Apprentissage des réseaux de neurones artificiels

L'apprentissage d'un réseau de neurones consiste à introduire de nombreuses données dans le réseau de neurones et à utiliser des algorithmes complexes, pour ajuster les poids des connexions synaptiques jusqu'à ce que la prédiction approche le niveau attendu [62]. Le processus d'apprentissage déroule de manière itérative les algorithmes de propagation en avant, de rétropropagation du gradient de l'erreur, et de mise à jour des poids jusqu'à minimiser l'erreur entre la prédiction et la valeur cible.

- Propagation en avant : L'information se propage dans une seule direction, de la couche d'entrée à la couche de sortie. Au cours de la propagation du signal, chaque neurone est activé suivant l'équation 3.1. De manière plus générale, la valeur d'activation  $a_i^l$  du  $i^{\text{ème}}$  neurone de la  $l^{\text{ème}}$  couche du réseau de neurones est définie par les équations 2.6 et 2.7 :

$$z_i^l = \sum_{j=1}^n w_{ij}^l a_{ij}^{l-1} + b^l \quad (2.6)$$

$$a_i^l = f_{\text{activation}}(z_i^l) \quad (2.7)$$

- Rétropropagation du gradient de l'erreur L'algorithme de rétropropagation est un algorithme d'optimisation qui permet de corriger les erreurs de prédiction en mettant en évidence les poids synaptiques qui contribuent à engendrer ces erreurs dans le réseau. Cet algorithme se déroule en deux étapes.

La première étape consiste à calculer l'erreur de prédiction à partir des sorties du réseau et des valeurs cibles. Il existe plusieurs fonctions d'erreur adaptées à différents problèmes à résoudre.

La deuxième étape consiste à calculer en fonction des paramètres du réseau, le gradient de l'erreur pour chaque neurone de la couche de sortie vers la couche d'entrée. Le gradient est calculé en utilisant le théorème de dérivation des fonctions composées (ou la règle de dérivation en chaîne).

En appliquant, la règle de dérivation en chaîne, le gradient de l'erreur pour un neurone de la couche de sortie s'écrit suivant l'Équation 2.8 :

$$\frac{\delta C}{\delta w_{ij}^L} = \frac{\delta z_i^L}{\delta w_{ij}^L} + \frac{\delta a_i^L}{\delta z_i^L} + \frac{\delta C}{\delta a_i^L} \quad (2.8)$$

avec :

$C$  la fonction d'erreur ;

$\frac{\delta C}{\delta a_i^L}$  la variation de la fonction d'erreur en fonction de la sortie du réseau  $a_i^L$  ;

$\frac{\delta a_i^L}{\delta z_i^L}$  la variation de la fonction d'activation en fonction de la somme pondérée  $z_i^L$  ;

$\frac{\delta z_i^L}{\delta w_{ij}^L}$  la variation de la fonction de somme en fonction du poids  $w_{ij}^L$ .

- Mise à jour des poids La mise à jour de poids du réseau de neurones est un processus itératif qui utilise les résultats de l'algorithme de rétropropagation et l'algorithme de descente de gradient pour déterminer les poids qui minimisent l'erreur de prédiction. La descente de gradient est un algorithme d'optimisation itératif du premier ordre utilisé pour trouver un minimum ou maximum local d'une fonction différentiable et convexe. Appliqué aux réseaux de neurones artificiel, l'algorithme de descente de gradient permet de trouver un minimum local de la fonction de coût. Les Équations 2.9 et 2.10 décrivent la mise à jour des poids :

$$w_{ij}^l(t+1) = w_{ij}^l(t) - \alpha \frac{\delta C}{\delta w_{ij}^l(t)} \quad (2.9)$$

$$b_{ij}^l(t+1) = b_{ij}^l(t) - \alpha \frac{\delta C}{\delta b_{ij}^l(t)} \quad (2.10)$$

Avec  $\alpha$ , le taux d'apprentissage compris entre 0 et 1.

Ainsi, les paramètres sont mis à jour en fonction de leur contribution sur l'erreur de prédiction.

La descente de gradient s'effectue de trois manières différentes : descente de gradient globale (batch gradient), et stochastique (stochastic gradient) par lot (mini-batch gradient).

Dans la descente de gradient globale, le réseau prend en compte tout l'ensemble d'apprentissage pour effectuer une mise à jour des poids. En effet, l'algorithme

calcul la moyenne des gradients de tous les exemples d'apprentissage, puis met à jour les paramètres en utilisant le gradient moyen. La descente de gradient stochastique consiste à mettre à jour les poids du réseau de neurones après observation de chaque donnée de l'ensemble d'apprentissage. Enfin, le processus de la descente de gradient par lot est le même que celui de la descente de gradient globale à la différence que la mise à jour des paramètres se fait sur de plus petit ensemble de données. La taille du petit ensemble de données est un hyperparamètre défini manuellement selon la taille de l'ensemble de données.

## 2.2 Classification basée sur l'apprentissage profond

Les méthodes de classification de lésions cutanées basées sur l'apprentissage profond (deep learning) utilisent plus particulièrement les réseaux de neurones à convolutions (Convolutional Neural Network CNN) mis en place par LeCun [66]. Ces réseaux sont utilisés avec beaucoup de succès en traitement d'images médicales.

Les réseaux de neurones à convolution sont composés de couches de convolution (convolution layer), de couches de mise en commun (pooling) qui sont responsables de l'extraction de caractéristiques et de couches entièrement connectées responsables de la classification. La Figure 2.7 présente l'architecture d'un CNN traditionnel.

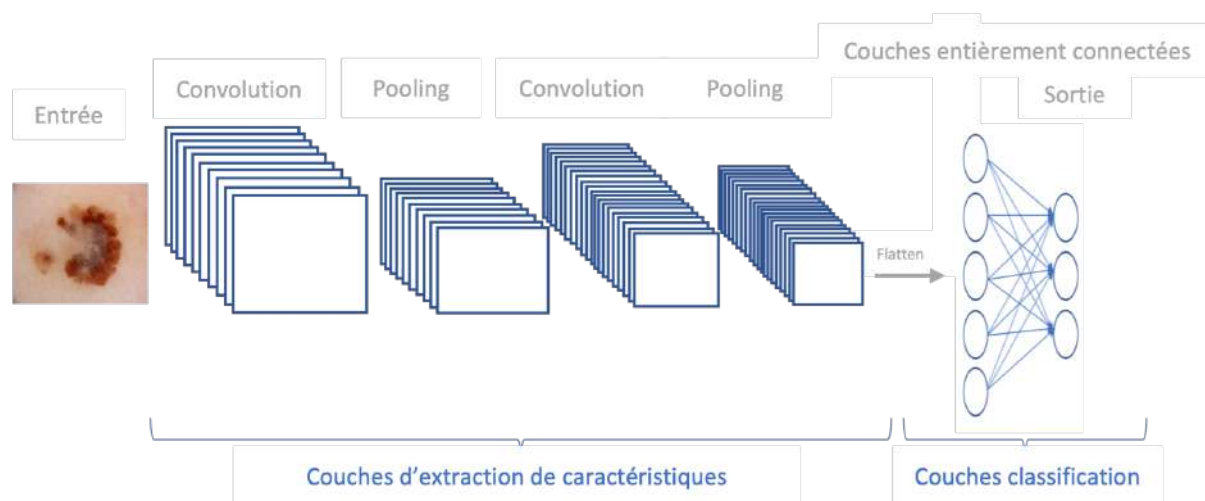


FIGURE 2.7 – Architecture traditionnelle d'un CNN.

- Couche de convolution La couche de convolution est une partie fondamentale des réseaux de neurones à convolution. Elle permet de détecter la présence de caractéristiques (features) dans les images d'entrée à travers des produits de convolution. L'opération de convolution consiste à faire glisser un noyau de convolution (filtre), sur la donnée d'entrée (image ou carte de caractéristiques) et à calculer la convolution entre le noyau et chaque partie de la donnée d'entrée balayée. Les filtres des

couches de convolution sont mis à jour pendant la phase d'apprentissage afin d'extraire les caractéristiques locales les plus utiles pour la tâche. Le nombre, la taille, et le pas de déplacement des filtres sont des hyperparamètres définis manuellement.

- **Couche de mise en commun** La couche de mise en commun est appliquée entre deux couches de convolution et permet de réduire la taille des cartes de caractéristiques générées par la couche de convolution. L'opération de mise en commun consiste à réduire un ensemble de pixels dans une fenêtre (région) en une valeur en sélectionnant le maximum (max-pooling), le minimum (min-pooling) ou en calculant la valeur moyenne des pixels de la fenêtre (average pooling) [62]. Le pooling favorise une invariance de base aux rotations et aux translations, et réduit les paramètres du réseau de neurones et donc le temps de calcul [67].
- **Couches entièrement connectées** L'architecture des couches entièrement connectées est similaire à celle des réseaux de neurones artificiels (ANNs : Artificial Neural Networks). Chaque nœud d'une couche entièrement connectée est directement connecté à chaque nœud de la couche précédente et de la couche suivante, comme le montre la partie droite de la Figure 2.7. La couche entièrement connectée transforme (aplaties "flatten") les dernières caractéristiques extraites en un seul vecteur qui sert d'entrée à la première couche entièrement connectée. La couche de sortie donne la prédiction la classe de l'image.

Pour les tâches de classification automatique de lésions cutanées, les réseaux de neurones à convolution sont utilisés de trois différentes manières : l'apprentissage à partir de zéro (training from scratch)[59], l'apprentissage par transfert (transfer learning) [68], l'extraction automatique de caractéristiques (features extraction) [57, 69].

### 2.2.1 Apprentissage à partir de zéro

Le processus d'apprentissage à partir de zéro du réseau de neurones à convolution est similaire à celui des réseaux de neurones artificiels (voir Section 2.1.3.3) et repose sur la descente des gradients pour minimiser l'erreur entre la prédiction et la donnée réelle. Cette forme d'apprentissage est coûteuse en termes de calcul et nécessite des données volumineuses.

### 2.2.2 Apprentissage par transfert

L'apprentissage par transfert de connaissance est une technique utilisée pour accélérer l'entraînement du réseau de neurones. Cette technique consiste à transférer les connaissances acquises par un réseau lors de la résolution d'un problème pour résoudre un autre problème [15]. L'apprentissage par transfert est peu coûteux et nécessite moins de données

d'apprentissage. Néanmoins, l'apprentissage par transfert n'est recommandé que lorsque le problème résolu généralise suffisamment le problème à résoudre.

### **2.2.3 Extraction automatique de caractéristiques avec les réseaux de neurones**

L'extraction automatique de caractéristiques consiste à extraire des caractéristiques d'images sans aucune intervention humaine au moyen de réseaux de neurones à convolution pré-entraînés sur des problèmes similaires au problème à résoudre. Comme présenté dans la section précédente, les réseaux à convolution sont composés de couches responsables de l'extraction de caractéristiques et de couches responsables de la classification [62]. Le processus d'extraction consiste à remplacer les couches responsables de la classification par un classifieur autre que les réseaux de neurones. Ainsi, les couches d'extraction de caractéristiques servent à extraire les caractéristiques sur l'image cible.

## **2.3 Évaluation des outils d'aide au diagnostic**

### **2.3.1 Méthodes de validation**

Une approche courante de validation des réseaux de neurones est la méthode de validation non croisée encore appelée "apprentissage-test". Cette méthode consiste à diviser la base de données en deux sous-ensembles respectivement d'apprentissage et de test. L'ensemble d'apprentissage sert à entraîner le réseau de neurones et l'ensemble de test sert à tester le modèle entraîné. Cette méthode est facile à mettre en œuvre et rapide à exécuter. L'inconvénient majeur de cette approche est qu'elle est sensible au découpage parce que seulement une partie de l'ensemble de données sert à l'apprentissage du modèle, de même pour l'évaluation du modèle. Ce qui pourrait favoriser un problème de sous-apprentissage et de sur-apprentissage.

Pour une meilleure généralisation du modèle une autre approche est proposée : la validation croisée. La validation croisée consiste à subdiviser la base de données en plusieurs plis (folds), servant à tour de rôle d'ensemble de test, et le reste de ces données est réservé à l'apprentissage. La subdivision de la base de données est faite aléatoirement (k-fold simple), ou suivant un principe (k-fold stratifié). La stratification est un processus de réorganisation des données pour s'assurer que chaque pli est approximativement représentatif de toutes les classes de la base de données.

### **2.3.2 Mesure de performance**

La mesure de performance des modèles est généralement basée sur la comparaison des prédictions du modèle avec les valeurs des données réelles. Pour un modèle idéal, les prédictions et les valeurs des données réelles devraient être égales mais en pratique,

ce n'est jamais le cas. La mesure de performance consiste à quantifier cet écart observé en pratique en utilisant des métriques. Le choix de la métrique dépend entièrement du type de modèle (régression, classification). Les métriques sont utilisées pendant la phase d'apprentissage et de test, et elles permettent de quantifier respectivement la qualité de l'ajustement et la qualité de prédiction.

Pour définir la qualité de l'ajustement d'un modèle de classification, il est possible d'utiliser des mesures basées sur l'erreur quadratique moyenne ou le maximum de vraisemblance [70]. Pour définir la qualité de la prédiction d'un modèle de classification, il est possible d'utiliser des mesures basées sur la matrice de confusion. La matrice de confusion est une matrice  $N * N$ , (où  $N$  est le nombre de classes à prédire) qui représente les différentes combinaisons de valeurs réelles et valeurs prédites. Le Tableau 2.1 montre une matrice de confusion d'un modèle de classification binaire.

TABLE 2.1 – Matrice de confusion.

Matrice de confusion		Réelle	
		Positif	Négatif
Prédiction	Positif	Vrai positif (VP)	Faux positif (FP)
	Négatif	Faux négatif (FN)	Vrai négatif (VN)

Les quatre cas possibles pour une matrice de confusion de taille  $2 \times 2$  sont :

- Vrai négatif (True Negative) : La prédiction est négative et coïncide avec la réalité ;
- Vrai positif (True Positive) : La prédiction est positive et coïncide avec la réalité ;
- Faux positif (False positive) : La prédiction est positive mais ne coïncide pas avec la réalité ;
- Faux négatif (False negative) : La prédiction est négative mais ne coïncide pas avec la réalité.

A partir de la matrice de confusion, on peut définir d'autres métriques telles que :

- Exactitude  $E$  (Accuracy), définie par l'Équation 2.11, est la proportion du nombre total de prédictions correctes.

$$E = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.11)$$

- Valeur prédictive positive  $VPP$  ou précision (Positive Predictive Value or Precision), définie par l'Équation 2.12, est la proportion de cas positifs qui ont été correctement identifiés.

$$VPP = \frac{VP}{VP + FP} \quad (2.12)$$

- Valeur prédictive négative VPN (Negative Predictive Value), définie par l'Équation 2.13 , est la proportion de cas négatifs qui ont été correctement identifiés.

$$VPN = \frac{VN}{VN + FN} \quad (2.13)$$

- Sensibilité ou Rappel (Sensitivity or Recall), définie par l'Équation 2.14 , est la proportion de cas positifs réels qui sont correctement identifiés.

$$Rappel = \frac{VP}{VP + FN} \quad (2.14)$$

- Spécificité S (Specificity), définie par l'équation 2.15 , est la proportion de cas négatifs réels qui sont correctement identifiés.

$$S = \frac{VN}{VN + FP} \quad (2.15)$$

Le poids de ces métriques dépend du problème que résous le modèle. Par exemple, pour la classification de lésion cutanées, il est primordial que le modèle identifie un mélanome tout en étant capable d'identifier correctement les lésions bénignes autant que possible. L'objectif principal est donc de cibler la sensibilité la plus élevée tout en optimisant la spécificité, ce qui augmente la précision globale du modèle.

Il existe une autre famille de métriques définie à partir des métriques précédentes : le score F1 (F1-score) et la courbe ROC (Receiver Operating Characteristic).

- Le score F1 est calculé à partir de la précision et du rappel du test. La formule du score F1 est définie par l'Équation 2.16 :

$$F_1 = \frac{VP}{VP + \frac{1}{2}(FP + FN)} \quad (2.16)$$

La valeur du score  $F_1$  est comprise entre 0 et 1.0. Un score de 1.0 indique une précision et un rappel parfaits, et score de 0, indique que la précision ou le rappel est nul. Cette métrique est utile si un équilibre entre la précision et le rappel est requise.

- ROC-AUC : La courbe ROC (receiver operating characteristic curve) est un tracé de la relation existante entre la sensibilité et la spécificité d'un test pour toutes les valeurs seuils possibles. L'ordonnée représente la sensibilité et l'abscisse correspond à la quantité (1 - spécificité). L'aire sous la courbe ROC (AUC Area Under the ROC Curve) est une mesure globale de la capacité d'un test à déterminer si une condition

TABLE 2.2 – Étude comparative des méthodes de classification.

Méthodes	Avantages	Inconvénients
<b>Approche traditionnelle</b>	<ul style="list-style-type: none"> <li>- Plus rapides à entraîner et à utiliser ;</li> <li>- Fonctionne efficacement avec des ensembles de données de petite à moyenne taille.</li> </ul>	<ul style="list-style-type: none"> <li>- Limités dans leur capacité à capturer des motifs complexes et des relations non linéaires dans les données ;</li> <li>- Nécessitent généralement une ingénierie de fonction pour extraire des caractéristiques pertinentes.</li> </ul>
<b>CNN à partir de zéro</b>	<ul style="list-style-type: none"> <li>- Capacité à apprendre des caractéristiques complexes à partir des données brutes ;</li> <li>- Adapté aux données d'images lorsque l'architecture est correctement conçue.</li> </ul>	<ul style="list-style-type: none"> <li>- Requiert un grand volume de données d'entraînement pour éviter le surapprentissage ;</li> <li>- Nécessiter une configuration et un ajustement minutieux des hyperparamètres.</li> </ul>
<b>Apprentissage par transfert</b>	<ul style="list-style-type: none"> <li>- Économise du temps et des ressources d'entraînement, car les modèles pré-entraînés ont déjà appris des caractéristiques générales à partir de grandes quantités de données ;</li> <li>- Efficace même avec un petit ensemble de données d'entraînement.</li> </ul>	<ul style="list-style-type: none"> <li>- Nécessite une base de données d'entraînement compatible avec les données cibles ;</li> <li>- Peut y avoir des problèmes de transfert de connaissance si les données cibles sont très différentes des données d'entraînement d'origine.</li> </ul>

spécifique est présente ou non. Une AUC de 0,5 représente un test sans capacité de discrimination (c'est-à-dire aléatoire), tandis qu'une AUC de 1,0 représente un test avec une discrimination parfaite.

## 2.4 Comparaison des méthodes de classification

Le Tableau 2.2 présente les avantages et inconvénients de différentes approches de classification d'images : approche Traditionnelle, CNN à partir de zéro, et apprentissage par transfert.

Le choix de l'une ou l'autre de ces approches dépend des ressources disponibles : expertise pour la sélection manuelle des caractéristiques, ressources matérielles pour l'im-



plémentation et l'entraînement et de la taille de la base de donnée. L'extraction manuelle des caractéristiques à l'aide de l'approche traditionnelle nécessite des connaissances dans le domaine pour sélectionner les caractéristiques pertinentes pour la tâche, et convient aux petits ensembles de données. Par contre, les approches d'apprentissage à base zéro et de transfert intègrent la sélection et la classification automatiques des caractéristiques et conviennent aux grandes bases de données. La spécificité de l'apprentissage par transfert est que, pour qu'il améliore les performances du modèle, la première tâche doit permettre de généraliser la seconde. Pour la classification d'images de carcinomes basocellulaires, de carcinomes épidermoïdes, de naevi et de mélanomes, par exemple, la première tâche peut être une classification binaire (cancers malins ou bénins). L'apprentissage par transfert est sujet à des limitations telles que le sur-apprentissage et peut ne pas être généralisable à toutes les tâches.

Dans la section 2.5, nous nous intéressons à l'apprentissage zéro, et en particulier à la manière dont il peut être adapté à de petits ensembles de données.

## 2.5 Méthode proposée : GCN

### 2.5.1 GCN

Les CNN sont de plus en plus indispensables à la classification d'images, mais leur utilisation pour les images médicales est difficile car ces modèles nécessitent de très grands ensembles d'apprentissage (des dizaines de milliers à des millions d'images)[13]. Pour contourner cette difficulté, l'apprentissage par transfert est proposé. Selon Brinker dans [16], l'apprentissage par transfert n'est efficace que si le modèle a été pré-entraîné sur des données qui présentent les mêmes similitudes que la nouvelle base de données d'apprentissage.

La combinaison du CNN avec une banque de filtres de Gabor est une nouvelle méthode de classification d'images par apprentissage profond. Nommée GCNN (Gabor CNN) ou GCN (Gabor Convolutional Network) cette méthode consiste à initialiser les filtres du CNN avec une banque de filtres de Gabor. Les travaux de [71, 72] ont proposés des architectures de GCNN pour la classification sur les jeux de données MNIST (Modified National Institute of Standards and Technology database), CIFAR10 et CIFAR100 (Canadian Institute for Advanced Research, 10 ou 100 classes) et . En faisant une étude comparative avec un CNN, ils ont observé une amélioration de la précision du modèle, une réduction du temps d'apprentissage et une optimisation de l'utilisation de la mémoire. De même les travaux de [73, 74] ont conclu que l'utilisation d'architectures de type GCN leur ont permis de renforcer l'apprentissage des informations de texture pour la reconnaissance d'objets. Nos travaux ont aussi contribué à améliorer la classification de lésion cutanée en proposant une architecture basée sur les GCN. Les sections suivantes présentent le modèle proposé et les résultats obtenus.

## 2.5.2 Méthode

La méthode proposée comprend trois étapes principales : la génération d'une banque de filtres de Gabor, la construction du CNN, et l'initialisation des paramètres. Les différentes étapes de la méthode proposée sont illustrées dans la Figure 2.8.

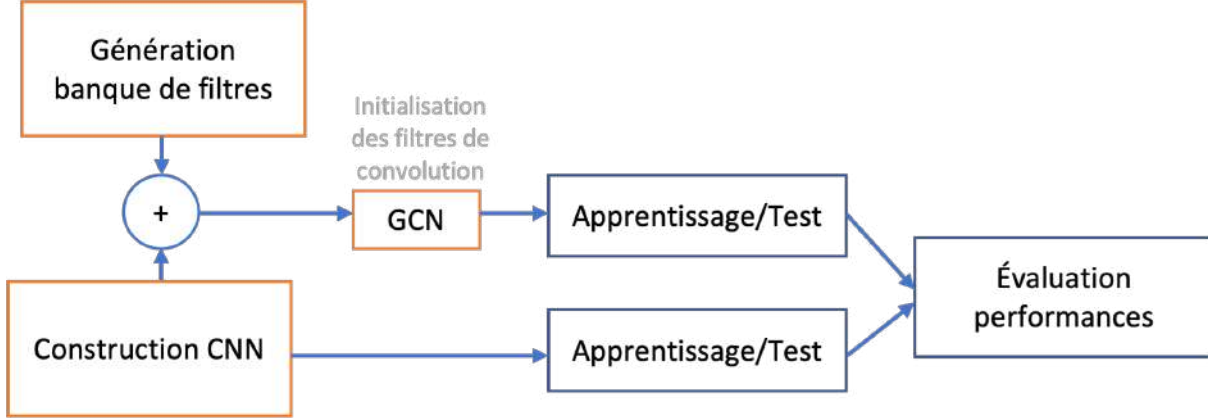


FIGURE 2.8 – Méthode GCN proposée.

### 2.5.2.1 Génération des filtres de Gabor

Le filtre de Gabor est un filtre linéaire utilisé dans les applications de traitement d'images pour la détection des bords, l'analyse de la texture, ou encore l'extraction de caractéristiques. La banque de filtres de Gabor est un ensemble de filtres linéaires de Gabor construits sur différentes orientations. Les banques de filtres peuvent permettre d'approximer les caractéristiques de certaines cellules du cortex visuel de certains mammifères (comme l'homme) [75].

Le filtre de Gabor 2D est un signal sinusoïdal de fréquence et d'orientation particulières, modulé par une fonction gaussienne (aussi appelée ondelette de Gabor). Le filtre de Gabor 2D est défini par l'Équation 2.17 [72] :

$$g(x, y; \sigma, \theta, \lambda, \gamma, \Psi) = \exp\left(\frac{-x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{x'}{\lambda} + \Psi\right)\right) \quad (2.17)$$

avec :

$$x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta$$

$g(x, y)$  : la réponse à valeur complexe du filtre de Gabor en un point spécifique de l'image avec les coordonnées  $(x, y)$ .

$\sigma$  : le sigma/écart-type de l'enveloppe gaussienne.

$\theta$  : l'orientation de la normale aux bandes parallèles de la fonction de Gabor.

$\lambda$  : la longueur d'onde de la composante sinusoïdale.

$\gamma$  : le rapport d'aspect spatial et spécifie l'ellipticité du support de la fonction de Gabor.

$\Psi$  : le décalage de phase de la fonction sinusoïdale.

Ces cinq paramètres contrôlent la forme et la taille des filtres de Gabor, et déterminent les caractéristiques spécifiques que le filtre peut extraire de l'image.

### 2.5.2.2 Construction du modèle

Nous avons proposé un réseau de neurones à convolution entraîné à zéro qui se compose de cinq blocs de convolution et d'une couche entièrement connectée. Les neurones de la couche de sortie du réseau sont activés par la fonction d'activation sigmoïde et les autres neurones sont activés par une fonction d'activation ReLU. Chaque couche de convolution est formée de filtre 2D dont la taille et la profondeur varie en fonction du niveau de la couche. Chaque couche de convolution est suivie d'une couche de mise en commun. La particularité de ce réseau réside dans le choix des hyperparamètres comme la profondeur des filtres, le nombre de couches de convolution et de couches cachées. La Figure 2.9 présente avec détail l'architecture du réseau CNN proposée.

### 2.5.2.3 Initialisation des filtres

Il existe différentes approches d'initialisation des paramètres de la couche convolutionnelle. Une première stratégie consiste à définir toutes les valeurs à une constante (zéro par exemple). Une autre consiste à échantillonner les valeurs des paramètres dans un intervalle, ou dans une limite maximale. Il existe également une approche d'initialisation des poids dans laquelle les poids sont initialement sélectionnés à partir de distributions aléatoires (normale ou gaussienne), mais sont ensuite mis à l'échelle. La mise à échelle est faite de manière que la variance des couches d'entrée et de sortie soit maintenue ou que la variance de la couche de sortie soit maintenue à une valeur souhaitée ou que la variance des gradients soit maintenue pendant la phase d'apprentissage. L'initialisation de Glorot/Xavier [76] et l'initialisation de He [77] sont deux méthodes populaires inspirées de cette approche.

L'initialisation de Glorot initialise les biais à zéro et les poids à partir d'une distribution aléatoire uniforme dans un intervalle. Les poids sont ensuite mis à l'échelle par un facteur d'échelle tel que la variance d'entrée soit égale à la variance de sortie. Cette méthode d'initialisation est mieux adaptée à la fonction sigmoid. L'initialisation de He modifie le facteur d'échelle des poids donné par Glorot [76] pour prendre en compte les non-linéarités du redresseur. Cette méthode d'initialisation est mieux adaptée aux fonctions d'activation ReLU.

Dans le cadre du modèle proposé, la troisième étape consiste à initialiser la première couche de convolution du modèle avec une banque de filtres de Gabor. Cette première couche permet d'extraire des caractéristiques robustes et discriminantes pour les couches suivantes. Les autres couches sont initialisées avec la méthode He.

## 2.5.3 Résultats expérimentaux et discussions

### 2.5.3.1 Base de données

La base des archives d'images de l'International Skin Imaging Collaboration (ISIC 2019) [4, 5, 6] est utilisée pour les expérimentations. L'archive ISIC 2019 contient 33 569 images dermoscopiques organisées en 9 classes : kératose actinique, carcinome basocellulaire, kératose bénigne, dermato-fibrome, nævus mélanocytaire, mélanome, carcinome spinocellulaire, lésion vasculaire. Cette base est réorganisée en deux classes : mélanome avec 4.523 images et non-mélanome avec 20.785 images. Les images dont les contours ne sont pas visibles ont été supprimées.

Les expérimentations ont été faites en premier sur une partie de la base de données et ensuite sur l'intégralité de la base. La méthode de validation utilisée est l'apprentissage-test avec 80% des images pour l'apprentissage et 20% pour le test. Pour l'apprentissage, les modèles GCN et CNN sont compilés avec l'optimiseur ADAM et le taux d'apprentissage est initialisé à 0,01. La métrique de performance utilisée est l'exactitude. Elle mesure la proportion de prédictions correctes faites par un modèle par rapport à l'ensemble des prédictions.

### 2.5.3.2 Résultats et discussions

La Figure 2.10 détaille les résultats de la précision des modèles GCN et CNN proposés durant la phase d'apprentissage et de test. Ces mesures ont été prises à différentes époques et l'apprentissage a été fait sur une partie de la base de données. De même, la Figure 2.11 détaille les performances des modèles GCN et CNN entraînés sur cette fois sur l'intégralité de la base de données.

En observant dans un premier temps les mesures du CNN (courbes en bleu) dans la Figure 2.10, on remarque dès les premières étapes et jusqu'à la fin, un écart entre les performances du modèle en apprentissage et celles du modèle en test. Cet écart est aussi observable sur les mesures du CNN (courbe en bleu) dans la Figure 2.11 mais est moins important que celui observé dans la Figure 2.10. Par exemple, à la dernière époque, l'écart sur la Figure 2.10 est de 0,0633 contre 0,0169 pour celui de la Figure 2.11. Cette observation nous permet de confirmer l'importance d'avoir une base de données large pour limiter le sur-apprentissage du modèle [78] et favoriser la généralisation du modèle.

Dans un deuxième temps, en observant les mesures du GCN (courbe en vert) dans la Figure 2.10 et Figure 2.11, on remarque toujours un écart entre les performances du modèle en apprentissage et celle du modèle en test, mais cet écart est moins important que

l'apprentissage soit réalisé sur une partie de la base de données ou sur la totalité de la base comparé aux écarts avec les CNN. L'utilisation des filtres de Gabor à la première couche de convolution permet d'extraire des caractéristiques robustes et discriminantes pour les couches suivantes. Ce qui explique la faible sensibilité des GCN à la taille de la base de données.

Enfin en observant les mesures de test des deux modèles GCN et CNN (courbes en trait plein) sur la Figure 2.10 et la Figure 2.11, le GCN présente de meilleures performances dès les premières étapes de l'apprentissage. À la fin, les deux modèles convergent presque avec une meilleure précision du GCN qui donne 96,39% contre 94,02% pour le CNN.

## Conclusion

Ce chapitre présente les outils d'aides au diagnostic de lésions cutanées, les méthodes d'évaluation de ces outils, et enfin notre contribution pour améliorer la précision de classification. La méthode proposée combine les filtres de Gabor au réseau CNN. Le CNN est une méthode puissante d'extraction de caractéristiques, mais il nécessite un grand nombre d'échantillons d'apprentissage pour éviter le surapprentissage. Pour réduire ce risque, le filtrage de Gabor a été utilisé pour extraire efficacement les informations spatiales, de bord et de texture. Le modèle est entraîné et évalué sur la base d'images dermoscopiques ISIC 2019. Les résultats obtenus par la méthode proposée surpassent les résultats du modèle CNN, même dans le cas où le nombre d'échantillons d'entraînement est limité.

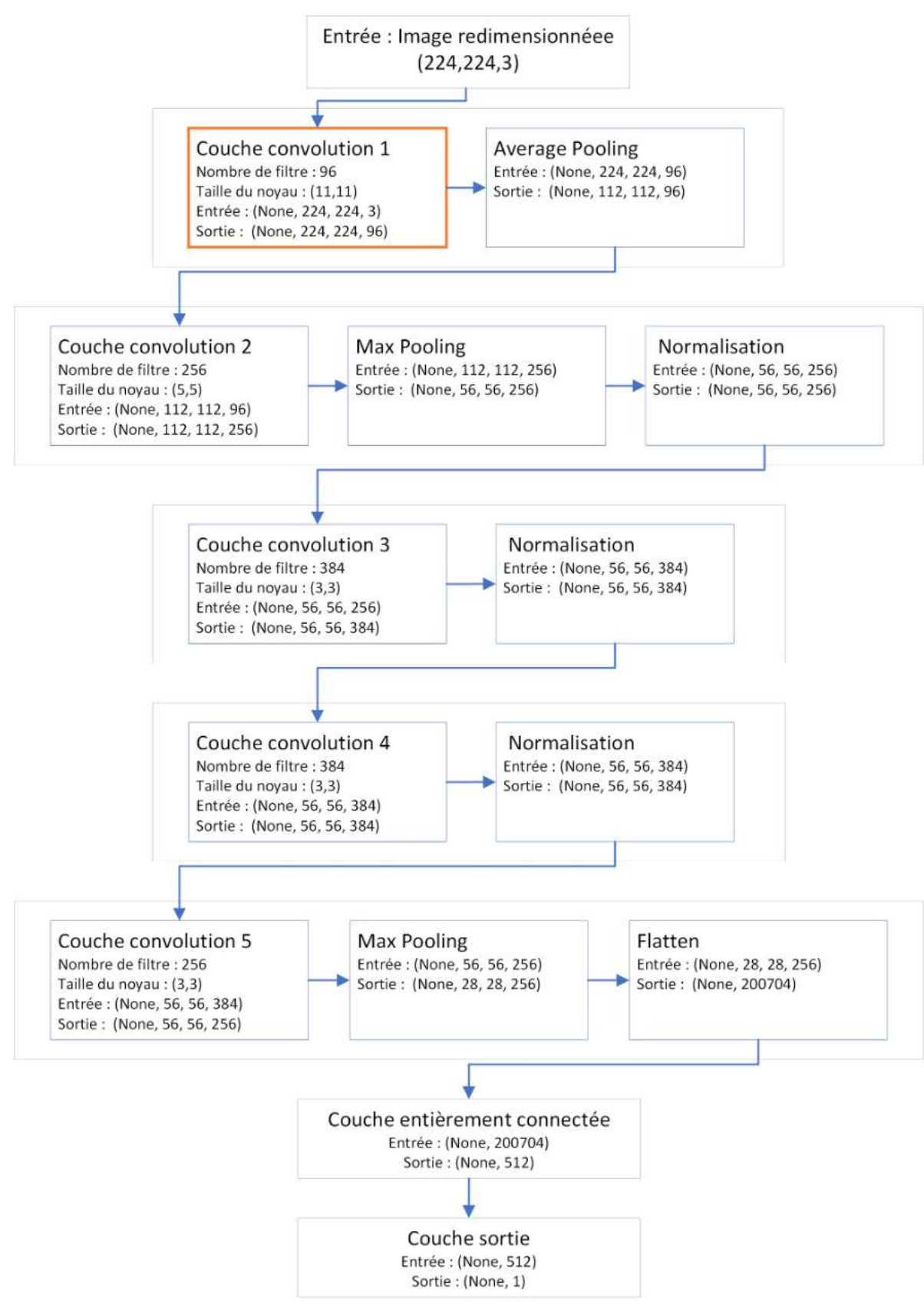


FIGURE 2.9 – Architecture du réseau proposé.

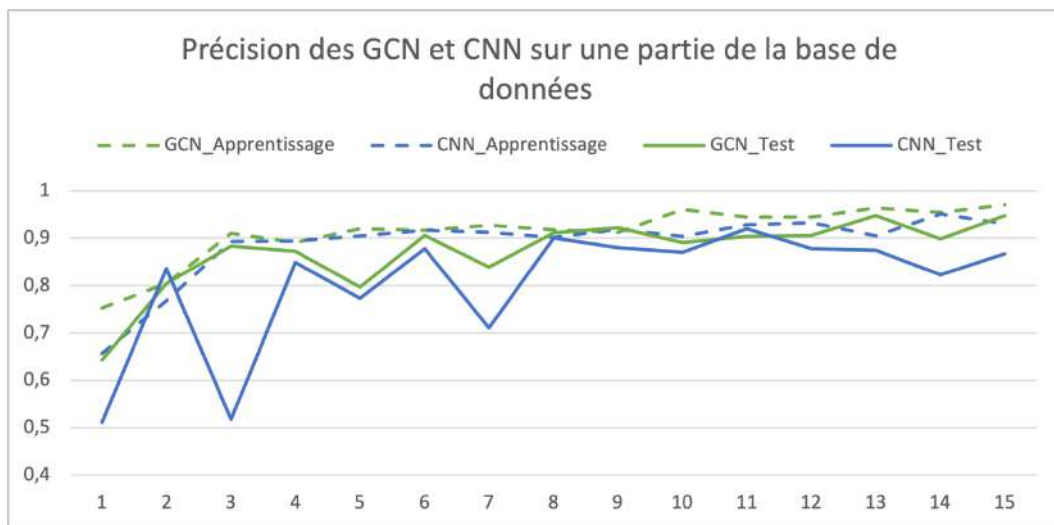


FIGURE 2.10 – Précisions des GCN et CNN sur une partie de la base de données.

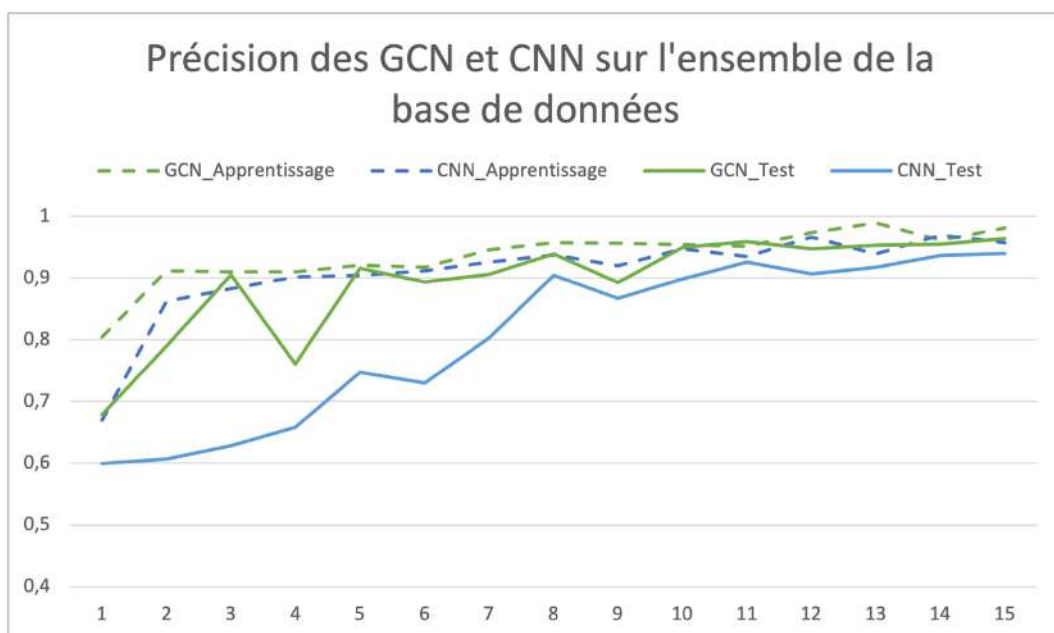


FIGURE 2.11 – Précisions des GCN et CNN sur l'ensemble de la base de données.





## CHAPITRE 3

# AUGMENTATION DE LA BASE D'IMAGES DERMOSCOPIQUES

### Sommaire

---

<b>3.1</b>	<b>Introduction</b>	<b>45</b>
<b>3.2</b>	<b>Augmentation basée sur des approches traditionnelles</b>	<b>46</b>
3.2.1	Transformations géométriques	46
3.2.2	Recadrage	47
3.2.3	Occlusion	47
3.2.4	Transformation de couleur	48
3.2.5	Injection de bruit	48
3.2.6	Filtrage	48
<b>3.3</b>	<b>Augmentation basée sur les réseaux générateurs antagonistes</b>	<b>48</b>
3.3.1	Généralités sur les réseaux générateurs antagonistes	48
3.3.2	Augmentation d'images médicales avec les GAN	49
<b>3.4</b>	<b>Augmentation basée sur des approches de fusion d'images</b>	<b>50</b>
3.4.1	Fusion d'images dans le domaine spatial	50
3.4.2	Fusion d'images dans le domaine transformé	51
<b>3.5</b>	<b>Comparaison des méthodes d'augmentation d'images</b>	<b>54</b>
<b>3.6</b>	<b>Méthode proposée : NSST PCNN</b>	<b>55</b>
3.6.1	PCNN simplifié et adaptif	56
3.6.2	Algorithme	58
3.6.3	Resultats expérimentaux et discussions	59

---

### 3.1 Introduction

Depuis l'introduction du premier réseau de neurones à convolution LeNet en 1998 par Yann LeCun et al. [79], différents types d'architectures de CNN sont proposés. Les architectures les plus connues sont : AlexNet [80], GoogLeNet [81], ResNet [82], et DenseNet

[83]. Les progrès réalisés dans l'architecture des réseaux de neurones, la croissance en puissance de calcul des ordinateurs, et l'accès aux grandes données ont favorisé l'application de la vision par ordinateur à de nombreuses tâches.

Néanmoins, la constitution de grandes bases de données d'images médicales est particulièrement délicate [84], à cause des règles de respect de la confidentialité des informations relatives au patient, de la difficulté à avoir accès aux spécialistes pour l'étiquetage des données, et du coût élevé des systèmes d'acquisition d'imagerie médicale. Ces obstacles ont conduit à la création de plusieurs méthodes d'augmentation des données, basées sur des techniques simples et complexes.

Ce chapitre est structuré comme suit. Les trois premières sections présentent trois différentes techniques d'augmentation de données. La Section 3.6 présente l'approche proposée, suivie des résultats expérimentaux et de la discussion. Ce chapitre se termine par la conclusion.

## 3.2 Augmentation basée sur des approches traditionnelles

L'augmentation des données à partir de techniques traditionnelles consiste à appliquer une transformation à une image qui fait correspondre les points de l'image à une position différente, ou à manipuler les valeurs d'intensité de l'image pour produire une autre image. Les transformations géométriques, le recadrage, l'occlusion, la transformation de couleur, l'injection de bruit et le filtrage sont quelques exemples de techniques d'augmentation d'images. D'après [84, 85] ces méthodes permettent aussi d'augmenter la diversité des bases de données et sont courantes parce qu'elles sont rapides, facilement reproductibles et fiables.

### 3.2.1 Transformations géométriques

Les transformations géométriques appliquées aux images peuvent être : la mise à l'échelle, la translation, la rotation, la réflexion, le cisaillement, ou parfois des transformations de perspective incluant l'inclinaison. Les transformations géométriques peuvent être exprimées mathématiquement par l'É 3.1 [14] :

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3.1)$$

Avec :

$(x, y)$  les coordonnées d'un pixel dans l'image d'origine.

$(x', y')$  les coordonnées d'un pixel de l'image transformée.

$A$  la matrice (3,3) dont les valeurs déterminent le type de transformation appliquée. Le 3.1 montre les différentes valeurs de  $A$  pour mettre en œuvre ces transformations.

TABLE 3.1 – Transformations basées sur l'Équation 3.1

Transformations	Matrice A	Équation
Mise à l'échelle	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x' &= c_x x \\ y' &= c_y y \end{aligned}$
Translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x' &= x + t_x \\ y' &= y + t_y \end{aligned}$
Rotation	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned}$
Cisaillement (vertical)	$\begin{bmatrix} 1 & s_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x' &= x + y s_v \\ y' &= y \end{aligned}$
Cisaillement (horizontal)	$\begin{bmatrix} 1 & 0 & 0 \\ s_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x' &= x \\ y' &= y + x s_h \end{aligned}$

### 3.2.2 Recadrage

Le recadrage est une technique qui consiste à échantillonner aléatoirement des blocs (ensemble de pixels voisins) dans une image et à les ajoutés à d'autres images de la base de données pour en augmenter la taille [86, 87]. Cette technique est en général utilisée lorsqu'il y a un déséquilibre entre les classes. En cas de déséquilibre de classes, les blocs sont sélectionnés sur des images de classes sous représentées pour équilibrer la base de données [88].

### 3.2.3 Occlusion

L'occlusion est une technique comparable au recadrage, à la différence que les blocs de l'image sont retirés pour générer une nouvelle image. Kompanek et al. dans [89] ont proposé une méthode basée sur l'occlusion pour l'augmentation d'images d'Imagerie par Résonance Magnétique (IRM).

### 3.2.4 Transformation de couleur

La transformation de couleur consiste à manipuler les pixels ou voxels de l'image. Cela se fait souvent en modifiant la luminosité ou le contraste de l'image [90, 91]. La correction gamma, l'égalisation d'histogramme, la balance des blancs, sont des méthodes qui permettent d'ajuster le contraste d'une image.

### 3.2.5 Injection de bruit

L'injection de bruit est une autre méthode d'augmentation d'images qui vise à simuler des images bruitées [86]. Les bruits généralement injectés sont : le bruit Gaussien, le bruit uniforme et le bruit sel et poivre.

L'injection de bruit gaussien ou uniforme consiste à modifier les intensités des images par l'échantillonnage aléatoire d'une distribution gaussienne ou d'une distribution uniforme. L'injection du bruit sel et poivre, consiste à mettre en noir ou en blanc de manière aléatoire des pixels de l'image.

### 3.2.6 Filtrage

Le filtrage d'une image est effectué avec l'opérateur de convolution. Pour ce faire, un noyau convolutif est déplacé dans l'image pour modifier les intensités de chaque pixel en fonction des valeurs des pixels environnants. Grâce à cette méthode, une image peut être rendue plus nette, plus floue ou plus lisse et produire ainsi une image augmentée. Certains noyaux sont capables de détecter et d'intensifier les bords des objets trouvés dans l'image et ont également été utilisés pour l'augmentation [88].

## 3.3 Augmentation basée sur les réseaux générateurs antagonistes

### 3.3.1 Généralités sur les réseaux générateurs antagonistes

Les réseaux générateurs antagonistes ou GAN (Generative Adversarial Network) ont une architecture basée sur les réseaux de neurones et sont conçus pour générer de données comme des images, du son, des vidéos et du texte. L'architecture d'un GAN telle que représentée dans la Figure 3.1 repose sur la mise en compétition de deux réseaux. Les deux réseaux sont appelés générateur et discriminateur.

Le discriminateur est un réseau de neurones à convolution composé d'une couche d'entrée, de couches cachées, d'une couche de sortie, et permet de séparer les données fausses des données réelles. Le discriminateur défini par la fonction  $D(x)$  reçoit en entrée une donnée  $x$  qui peut par exemple être une image, un signal sonore, ou un caractère, et retourne une valeur  $y$  comprise entre 0 et 1. La valeur de  $y$  est proche de 0 si le discriminateur estime

que la donnée  $x$  est fausse. Dans le cas contraire, le discriminateur renvoi une valeur  $y$  proche de 1.

Le générateur est un type de réseau de neurones à convolution inverse qui permet de générer de nouvelles données. Le générateur est défini initialement par une distribution  $P_Z$  sur un vecteur  $z$  qui sert d'entrée à la fonction génératrice  $G(z)$ . La distribution  $P_Z$  est généralement une distribution gaussienne ou une distribution uniforme et le vecteur d'entrée  $z$  est du bruit. Le rôle principal du générateur est d'apprendre la fonction  $G(z)$  qui transforme ce bruit non structuré en échantillons réalistes  $x'$  de même dimension que  $x$ . Le générateur définit donc implicitement une distribution de probabilité  $P_G$  comme la distribution des échantillons  $G(z)$ .

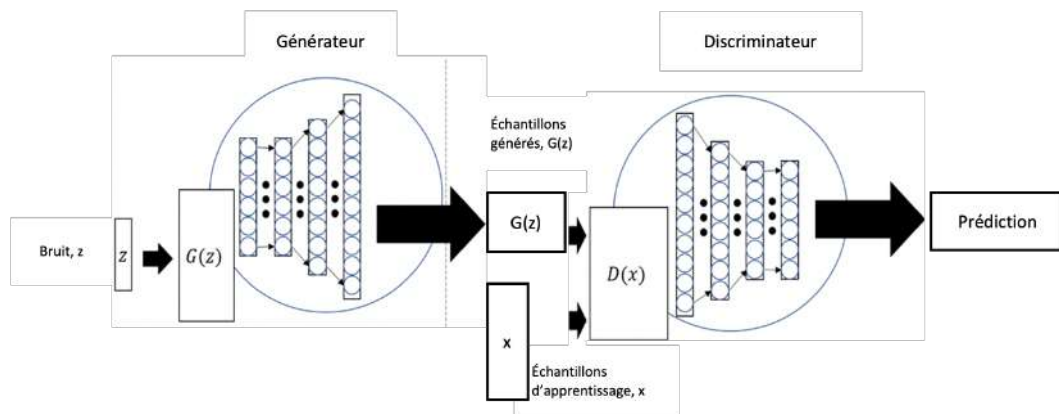


FIGURE 3.1 – Structure de base d'un GAN [7].

Le GAN, proposé à l'origine par Goodfellow et al. [7], est inspiré de la théorie des jeux (jeu minimax à 2 joueurs). Pendant la phase d'apprentissage, chaque modèle du GAN subit une fonction de coût. Chaque modèle actualise ses paramètres afin d'optimiser son propre coût. Le coût du discriminateur l'encourage à classer correctement les données comme réelles ou fausses, tandis que le coût du générateur l'encourage à générer des échantillons que le discriminateur classe incorrectement comme réels.

Le choix de la fonction de coût est déterminé par le problème à résoudre. Pour les GAN, les fonctions de coûts les plus connues sont l'entropie croisée (Cross-Entropy), la fonction de vraisemblance (Negative Log-Likelihood)[7], la divergence de Jensen-Shannon [7, 92], Wasserstein [92, 93], la divergence Kullback-Leiber [92]. La fonction de coût proposée Goodfellow et al. dans [94] est définie par l'Équation 3.2 :

$$L(G, D) = \sum_x \log(D(x)) + \sum_z \log[1 - D(G(z))] \quad (3.2)$$

### 3.3.2 Augmentation d'images médicales avec les GAN

Depuis leur introduction en 2014, différentes extensions de GAN sont proposées. Les modifications apportées concernent principalement l'architecture du générateur ou du

discriminateur, le type de bruit initial, les méthodes de normalisation et d'initialisation des paramètres et les fonctions de coûts. Les extensions de GAN les plus citées sont DCGAN (Deep Convolutional GAN)[95], les CycleGAN [96], les GAN à croissance progressive (Progressive Growing GAN) [97], StyleGAN [98].

Pour l'analyse d'images médicales, les GANs sont largement utilisés pour la reconstruction d'images [85, 97, 99], la segmentation [100], la classification [101, 102], la détection [103], le recalage et la synthèse d'images telles que l'image IRM du cerveau [104, 8], des images de lésions hépatiques [105, 106], et de lésions cutanées [107, 108, 109].

Les variantes de GANs basées sur l'architecture permettent de générer des images de très haute résolution [107], et des images de classes sous représentées dans les bases de données [106]. Les modèles GAN basés sur la variation de style ont permis à Zhiwei Qin et al. dans [107] de générer des images de lésions cutanées de haute résolution et diverses. Kora Venu et al. dans [106] ont proposé un modèle basé sur une architecture de GAN nommée Deep Convolutional Generative Adversarial Network (DCGAN) qui leur a permis de générer spécifiquement des images radiographiques normales parce que cette classe est sous-représentée dans leur base de donnée.

Un autre modèle pour générer des images réalistes est le modèle de traduction d'image à image (Image to Image translation). Cette technique consiste à générer une image à partir d'une autre image au lieu de générer l'image à partir du bruit (procédure habituelle avec les GANs). Alceu Bissoto et al. dans [109] ont proposé un modèle de traduction d'image à image nommé pix2pixHD. Le modèle synthétise de nouvelles images à partir de masques d'annotations.

### **3.4 Augmentation basée sur des approches de fusion d'images**

Les méthodes de fusion permettent d'intégrer plusieurs images sources pour en faire une image informative. Les images fusionnées peuvent provenir de différents systèmes d'acquisition ou du même système d'acquisition. Une fusion d'images efficace permet de préserver les caractéristiques pertinentes de toutes les images [37]. Les techniques de fusion d'images sont de plus en plus utilisées en vision par ordinateur, en surveillance, en imagerie médicale et en télédétection. La fusion se fait dans le domaine spatial ou dans un domaine fréquentiel [24]. La Figure 3.2 présente une classification des méthodes de fusion.

#### **3.4.1 Fusion d'images dans le domaine spatial**

Les méthodes de fusion dans le domaine spatial intègrent directement des blocs ou région de pixels des images à fusionner. Parmi ces méthodes, on peut citer : la méthode de filtrage passe-haut, la méthode de saturation de l'intensité de la teinte, la méthode de Brovey, la méthode d'analyse en composantes principales, la méthode moyenne, la méthode de sélection maximale, et la méthode de sélection minimale. Toutes fois, lorsqu'elles

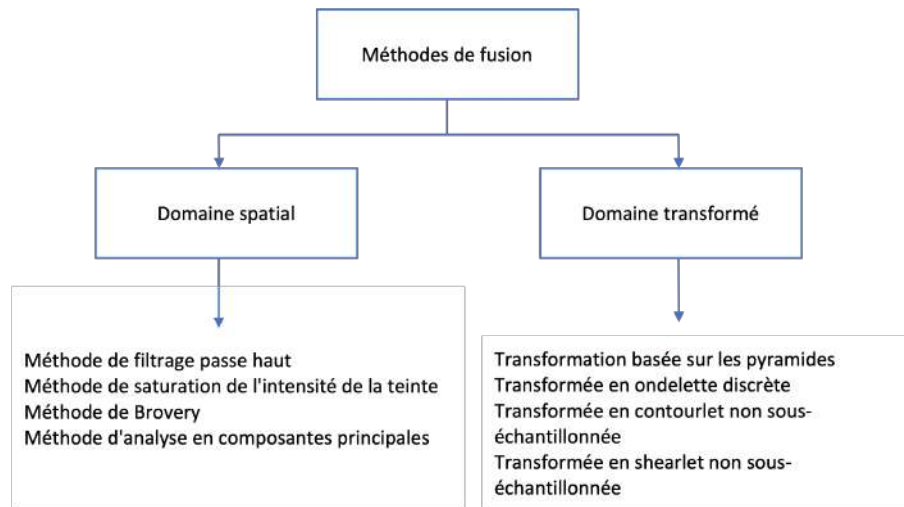


FIGURE 3.2 – Classification des méthodes de fusion.

sont appliquées directement sur des pixels, ces méthodes créent des distorsions spectrale et spatiale sur les images fusionnées. Une approche courante pour corriger ces distorsions est de mixer l'utilisation de techniques de fusion dans le domaine spatial et fréquentiel pour former de nouvelles méthodes.

### 3.4.2 Fusion d'images dans le domaine transformé

Les méthodes de fusion d'images médicales dans les domaines transformés sont généralement basées sur la transformation multi-échelle. Ces méthodes comprennent essentiellement trois étapes : la transformation des images sources à fusionner (décomposition), la fusion des coefficients obtenus après transformation, et la transformation inverse (reconstruction) des coefficients fusionnés pour produire l'image finale.

Les différentes méthodes de fusion d'images diffèrent par les méthodes de décomposition-reconstruction appliquées et la stratégie de fusion mise en place. La Figure 3.3 présente la procédure de fusion d'images dans le domaine fréquentiel.

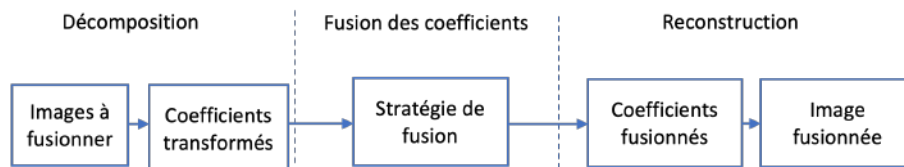


FIGURE 3.3 – Procédure de fusion d'images dans le domaine fréquentiel.

#### 3.4.2.1 Transformation d'image

Cette section présente quelques méthodes de transformations d'images utilisées pour la fusion d'images médicales.

### Transformation basée sur les pyramides

La pyramide est une représentation multi-résolution d'une image qui consiste à décomposer l'image à différentes résolutions, depuis l'image initiale (niveau inférieur) jusqu'à une image très grossière (niveau supérieur) [110]. Les images de la pyramide contiennent des informations descriptives concernant les contours, les gradients, le contraste de l'image décomposée. La Figure 3.4 présente un exemple d'une pyramide d'image.

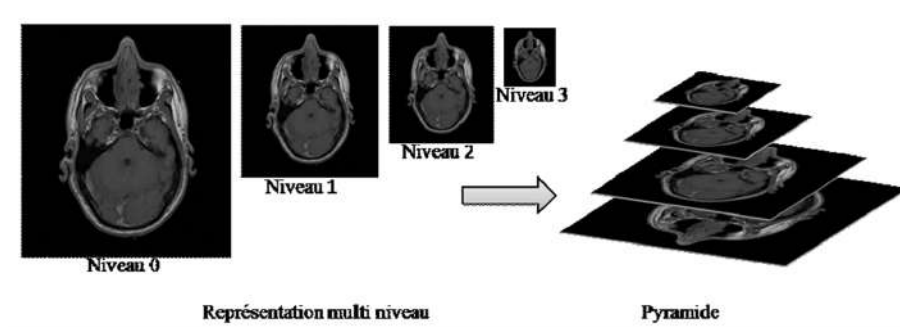


FIGURE 3.4 – Exemple pyramide d'image sur une coupe d'IRM. [8]

Il existe plusieurs types de construction de pyramide d'images. Les plus connues sont la pyramide de Gauss, et la pyramide Laplace. Dans la pyramide de Gauss, le niveau supérieur (basse résolution) est formé par une opération de filtrage sur l'image du niveau inférieur (haute résolution). Cette opération consiste à convoluer la matrice de l'image avec un filtre discret et effectuer une décimation à un facteur de 2 comme définit dans l'Équation 3.3 [111] :

$$G_l(i, j) = [G_{l-1} * w(i, j)]_{\downarrow 2} \quad (3.3)$$

Avec :

$G$  : Gaussienne de niveau  $l$ ;

$w$  : filtre passe-bas de taille  $5 \times 5$  [112].

La pyramide de Laplace est formée à partir d'une représentation pyramidale multi-résolution (la pyramide Gaussienne par exemple) [106]. Un niveau dans une pyramide de Laplace est formé par la différence entre ce niveau dans la pyramide de Gauss et la version étendue de son niveau supérieur dans la pyramide de Gauss.

### Transformée en ondelette

La transformée en ondelettes est une méthode fréquemment utilisée pour l'analyse multi résolution dans le traitement des images médicales. Shehanaz et al. dans [113] et Wang et al. [114] ont proposé des méthodes de fusion d'images médicales basées sur la transformée en ondelettes. La transformée en ondelettes décompose une image en plusieurs sous-bandes avec une combinaison de filtres passe-bas et passe-haut. Les sous-bandes  $LL(Low, Low)$  contiennent les coefficients approximatifs et les sous-bandes  $LH(Low, High)$ ,  $HL(High, Low)$



et  $HH$  (*High, High*) contiennent respectivement les coefficients verticaux, horizontaux et diagonaux détaillés [115]. La sous-bande  $LL$  à chaque niveau de décomposition peut encore être divisée en quatre parties.

La transformée en ondelettes discrètes est un système de sous-bandes hiérarchique, dans lequel toutes les sous-bandes sont espacées de manière logarithmique dans le domaine fréquentiel et représentent une structure de décomposition en bandes d'octave [114].

### **Transformée en contourlet non sous-échantillonnée**

La transformée de contourlet non sous-échantillonnée (NSCT : Nonsampled Contourlet Transform) est une méthode de décomposition multi-échelle, multidirectionnelle et à invariance de décalage basée sur la transformée de contourlet. Cette méthode décompose l'image en une sous-bande de basse fréquence et plusieurs sous-bandes de haute fréquence [116]. Le modèle pour le premier niveau de décomposition utilise une pyramide non sous-échantillonnée pour générer des composantes basse fréquence et haute fréquence. Pour les autres niveaux de décomposition, une banque de filtres directionnels non sous-échantillonnés est utilisée pour générer plusieurs composantes de sous-image [118].

Wang N. [114] et Li .L [117] dans leurs travaux ont proposé des méthodes de fusion d'images médicales basées sur la transformée en contourlet non sous-échantillonnée .

### **Transformée en shearlet non sous-échantillonnée**

La transformée en shearlet non sous-échantillonnée (NSST : Non Sampled Shearlet Transform) est une méthode de décomposition multi-échelle utilisée pour représenter efficacement les informations de haute et basse fréquence de l'image source [118]. Tout d'abord, l'image source est décomposée en bandes passe-bas et passe-haut à l'aide de la transformation pyramide laplacienne non sous-échantillonnée. Pour chaque niveau de décomposition, les bandes passe-haut sont soumises à des filtres de Shearlet invariants en translation et les bandes passe-bas sont à nouveau décomposées en bandes passe-bas et passe-haut pour le niveau suivant [119].

Subbiah Parvathy et al. dans [120] ont proposé une méthode basée sur des concepts d'apprentissage profond qui optimise le seuil des règles de fusion dans la transformée de Shearlet. La méthode proposée leur a permis de généraliser leur algorithme en l'utilisant sur différents types d'images.

### **Réseaux de neurones à convolution**

Les réseaux de neurones à convolution tels que présentés dans le Chapitre 1, sont composés de deux parties. Une première partie qui permet d'extraire des caractéristiques complexes des images et une deuxième partie qui permet de faire la classification.

Ainsi la première partie composée de couches convolutionnelles est utilisée pour réduire les images à fusionner en vecteur de caractéristiques (codage) [121, 122]. Les caractéristiques sont ensuite fusionnées et transformées en pixels (décodage).

Dans [122], Hui et al. ont proposé une méthode basée sur l'apprentissage profond appelée Densefuse. La méthode combine des couches convolutionnelles et un bloc dense layer

comme codeur pour extraire des caractéristiques profondes et une architecture spécifique de couches convolutionnelles comme décodeur pour reconstruire l'image fusionnée finale. Zhang et al. dans [121] ont proposé une méthode qui est aussi basée sur les réseaux de neurones à convolution.

### 3.4.2.2 Méthodes de fusion des coefficients

Il existe différentes méthodes de fusion et chacune de ces méthodes présente des effets de fusion différents.

#### Méthodes classiques

Les méthodes simples de fusion regroupent des fonctions comme minimum, maximum, moyenne et la combinaison linéaire. Ces méthodes dans la littérature sont utilisées pour fusionner les coefficients en basse fréquence [123].

#### Réseaux de neurones couplés par impulsion (PCNN)

Le PCNN (Pulse-Coupled Neural Networks) introduit par Johnson, J.L. et al. [124] est un réseau de neurones bidimensionnel (de la même taille que l'image à laquelle il est appliqué) dans lequel chaque neurone correspond à un pixel spécifique de l'image. Un neurone du PCNN fonctionne en de trois étapes : réception, modulation et activation (ou encore générateur d'impulsions) [125]. Chaque neurone du PCNN est activé ou désactivé en tenant compte des pixels du voisinage et l'intensité de pixel à la position du neurone. A la suite d'un certain nombre d'itération, la durée d'activation de chaque neurone permet d'apprécier les pixels qui contiennent le plus d'informations.

Les travaux de [114] et [126] ont proposé des méthodes de fusion d'images médicales basées sur ces réseaux.

## 3.5 Comparaison des méthodes d'augmentation d'images

Le Tableau 3.2 présente les avantages et inconvénients des méthodes d'augmentation d'images traditionnelles, les réseaux génératifs adversaires (GAN) et les méthodes de fusion d'images.

Les méthodes d'augmentation d'images traditionnelles sont simples à mettre en œuvre mais limitées en termes de diversité et de réalisme des données. Compte tenu de ce dernier inconvénient, nous nous sommes intéressés aux deux autres méthodes : GAN et fusion d'images. Les GAN peuvent générer des données très réalistes, comme le montre les travaux suivant, qui a généré des images dermoscopiques réalistes à haute résolution : [127, 128, 129]. Néanmoins pour générer des images, ces modèles nécessitent un entraînement complexe et peuvent être coûteux en ressources. Les méthodes de fusion d'images par contre ne nécessitent pas d'entraînement et sont utiles pour combiner des données à partir de plusieurs sources. Avec cette approche, la qualité de la fusion dépend de la qualité des images d'entrée et de la méthode de fusion utilisée.

TABLE 3.2 – Comparaison de méthodes d’augmentation d’images

Méthodes	Avantages	Inconvénients
<b>Méthodes traditionnelles</b>	<ul style="list-style-type: none"> <li>- Facilité d’utilisation et de mise en œuvre ;</li> <li>- Faible coût computationnel ;</li> <li>- Peut améliorer la robustesse des modèles envers les variations simples.</li> </ul>	<ul style="list-style-type: none"> <li>- Manquent de variété et de réalisme dans les données augmentées ;</li> <li>- Limités pour capturer des transformations complexes ou des caractéristiques fines.</li> </ul>
<b>GAN</b>	<ul style="list-style-type: none"> <li>- Générer des données hautement diversifiées et réalistes ;</li> <li>- S’adapte bien aux données complexes avec des structures intriquées ;</li> <li>- Possibilité de contrôler le réalisme des données générées.</li> </ul>	<ul style="list-style-type: none"> <li>- Requiert un entraînement intensif de modèles GAN, ce qui peut être coûteux en temps et en ressources ;</li> <li>- Problèmes de mode collapse où le GAN génère principalement un seul type d’image ou à des résultats imprévisibles.</li> </ul>
<b>Fusion d’images</b>	<ul style="list-style-type: none"> <li>- Utile pour intégrer des informations à partir de plusieurs sources ;</li> <li>- Permet de combiner des données multimodales.</li> </ul>	<ul style="list-style-type: none"> <li>- Qualité de la fusion dépend de la qualité des images d’entrée et de la méthode de fusion utilisée ;</li> <li>- Peu flexible pour créer des variations dans les données.</li> </ul>

Dans la section 3.6 qui présente la méthode proposée, nous nous intéressons à une approche d’augmentation par fusion, en sélectionnant les méthodes de fusion qui préservent le mieux les caractéristiques dermoscopiques présentes dans les images.

### 3.6 Méthode proposée : NSST PCNN

Cette section présente notre contribution pour augmenter les données d’images de cancers de la peau . Cette solution vient corriger la sous-représentation observée d’images de sujet à peau mixtes, foncées ou noires dans les ensembles de données de lésions cutanées. La méthode proposée, et illustrée dans la Figure 3.5, comprends deux parties. La première partie consiste en la transformation des images sources dans un domaine fréquentiel (décomposition), et la transformation inverse des coefficients fusionnés (reconstruction) en utilisant la transformée de Shearlet non sous échantillonnée (NSST voir 3.4.2.1) . La NSST grâce aux filtres de Shearlet directionnels et multi-échelles est particulièrement adaptées pour la représentation de structures orientées complexes comme celles qu’on retrouve dans les images dermoscopiques, comparée à d’autres méthodes comme la NSCT et la DWT (voir 3.4.2.1)

La deuxième partie de la partie proposée effectue la fusion des coefficients à l’aide

d'un réseau neuronal couplé par impulsions (PCNN). Cette stratégie de fusion permet d'extraire des caractéristiques saillantes et des régions d'intérêt dans les images. Comparé à d'autres méthodes de sélection basées sur les réseaux de neurones [130, 131], le PCNN s'adapte à des structures complexes et ne nécessite pas d'apprentissage. Le PCNN mis en place est une version actualisée simple, rapide et efficace.

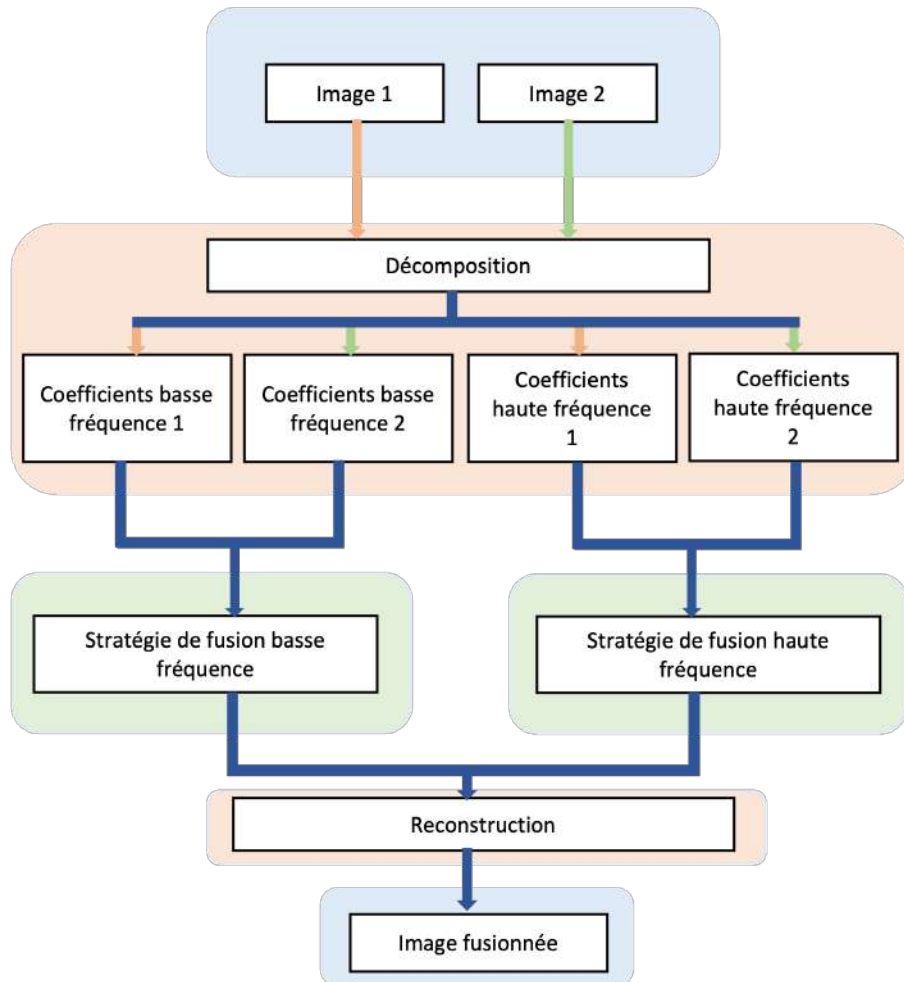


FIGURE 3.5 – Diagramme de la méthode proposée.

### 3.6.1 PCNN simplifié et adaptif

Comme présenté dans la section 3.4.2.2, le PCNN lorsqu'il est utilisé en fusion d'image est comparé à un réseau de neurones bidimensionnel. Chaque neurone de ce réseau, permet d'apprécier l'importance d'un pixel d'une image en tenant compte des pixels voisins. Le principe de fonctionnement d'un neurone comprend trois étapes : réception, modulation et activation (génération d'impulsions). La Figure 3.6 présente ces étapes d'activation d'un neurone PCNN. L'indice  $(i, j)$  indique la position du pixel dans l'image, et l'indice  $(k, l)$  fait référence aux pixels voisins autour d'un pixel, et  $n$  désigne l'itération en cours.

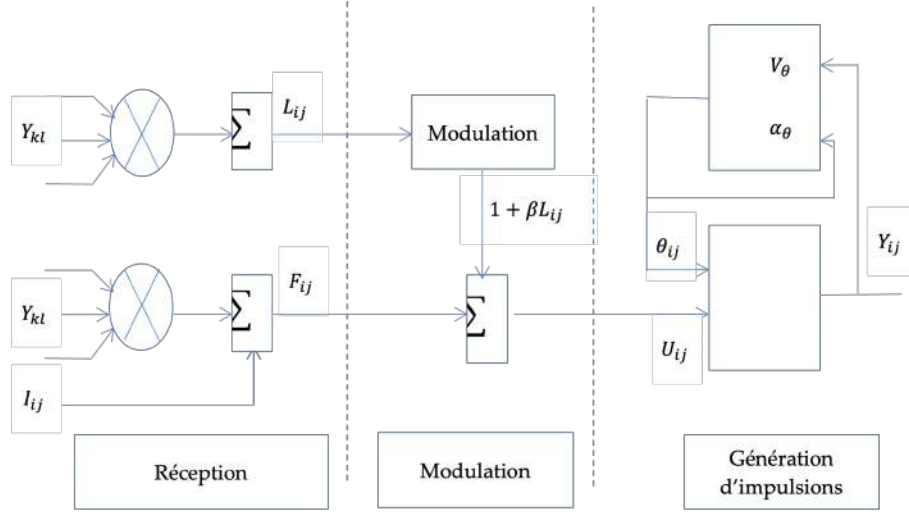


FIGURE 3.6 – Neurone à couplage d'impulsion de base.

Chaque neurone du PCNN est activé ou désactivé par un calcul itératif défini par les Équations 3.4 à 3.9.

$$F_{ij}(n) = \exp(-\alpha_F)F_{ij}(n-1) + V_F \sum_{kl} M_{ijkl}Y_{kl}(n-1) + I_{ij} \quad (3.4)$$

$$L_{ij}(n) = \exp(-\alpha_L)L_{ij}(n-1) + V_L \sum_{kl} W_{ijkl}Y_{ij}(n-1) \quad (3.5)$$

$$U_{ij}(n) = F_{ij}(n)(1 + \beta L_{ij}(n)) \quad (3.6)$$

$$Y_{ij}(n) = \begin{cases} 1 & \text{si } U_{ij}(n) \geq \theta_{ij}(n) \\ 0 & \text{sinon} \end{cases} \quad (3.7)$$

$$\theta_{ij}(n) = \exp(-\alpha_\theta)\theta_{ij}(n-1) + V_\theta Y_{ij}(n-1) \quad (3.8)$$

$$T_{ij}(n) = T_{ij}(n-1) + Y_{ij}(n) \quad (3.9)$$

L'étape de réception décrit par les Équations 3.4 et 3.5, est composée de deux modules  $F$  et  $L$ .

$L$  est le paramètre de liaison qui reçoit un stimulus local des neurones environnants. Le module d'alimentation  $F$  reçoit un stimulus externe de  $I$  le signal d'entrée (l'intensité du pixel). L'étape de modulation est définie mathématiquement par l'Équation 3.6.  $U$  appelé activation interne, combine les informations de l'entrée d'alimentation  $F$  et de l'entrée de liaison  $L$ . Avec l'entrée de liaison  $L$  modulée par un paramètre  $\beta$  appelé la force de liaison. La dernière étape, la génération d'impulsions est décrite par les Équations 3.7 et 3.8. Le module de sortie compare le paramètre d'activation interne  $U$  avec  $\theta$  le seuil dynamique. Si  $U$  est supérieur au seuil  $\theta$ , alors le neurone est activé et génère une impulsion, qui est caractérisée par  $Y = 1$ , sinon  $Y = 0$ . Le temps d'excitation de chaque neurone dénoté  $T$  est défini dans l'Équation 3.9.

Dans la méthode proposée, deux modifications ont été apportées au PCNN. La première modification a été apportée dans le but de simplifier le réseau et de réduire le temps de calcul et la deuxième a été apportée pour améliorer la performance du PCNN. Pour simplifier le processus d'activation des neurones, le module d'alimentation  $F$  reçoit un stimulus externe qui correspond à l'intensité  $I$  du pixel. Cette modification est définie dans l'Équation 3.10 qui par comparaison à l'Équation 3.4 est moins complexe et nécessite moins de paramètres.

$$F_{ij}(n) = I_{ij} \quad (3.10)$$

La deuxième modification est apportée au paramètre de liaison dénoté  $\beta$ . La force de liaison indique l'importance de la contribution des pixels environnants au pixel central. Ce paramètre généralement est fixé à une constante ce qui ne traduit pas la réalité. En effet, plus une région présente de caractéristiques, plus la force de liaison est importante. Pour corriger cela, nous avons proposé de déterminer  $\beta$  de manière dynamique, tel que défini dans l'Équation 3.11.

$$\beta_{(i,j)} = \frac{1}{1 + \exp(-G_{(k,l)})} \quad (3.11)$$

$\beta$  est défini comme une valeur locale et adaptive, calculée dans un voisinage de 3x3, en fonction de la magnitude du gradient  $G$ . La fonction sigmoïde permet de normaliser entre 0 et 1 la magnitude du gradient  $G$ . Cette modification va permettre d'améliorer la performance du PCNN en préservant au mieux les détails des images de sources dans l'image finale.

### 3.6.2 Algorithme

L'algorithme de fusion comprend quatre parties : la décomposition, la fusion basse fréquence, la fusion haute fréquence et la reconstruction.

Le processus de décomposition et reconstruction est réalisé avec la transformée de Shearlet non sous-échantillonnée, car les algorithmes basés sur la NSST sont invariants par rapport au décalage et peuvent éliminer efficacement les effets de bord. Une image décomposée avec NSST est transformée en coefficients basse fréquence et coefficients haute fréquence. Les coefficients de basses fréquences contiennent les informations de structure, de texture et de fond des images sources. Pour préserver ces informations, principalement celles de l'image de la lésion, la fusion est réalisée en utilisant la fonction d'intégration définie par l'Équation 3.12. Les paramètres  $\mu$  et  $\nu$  sont des coefficients pondérés fixés respectivement après expérimentations à 0,5 et à 0,1. Les coefficients des images dermoscopiques contribuent davantage que ceux de l'image de la peau foncée et saine, afin d'améliorer la préservation des informations de structure, de texture et de fond des images.

$$L(x, y) = \mu A_{Low} + \nu B_{Low} \quad (3.12)$$

Les coefficients de hautes fréquences contiennent des informations sur les détails des images et sont fusionnés en utilisant l’algorithme de fusion d’images basé sur les PCNN dont le pseudo-code est présenté dans le Tableau 3.3 . Le PCNN est un processus itératif qui génère à la fin deux matrices temps  $T_A$  et  $T_B$  pour les images sources  $A$  et  $B$ . La règle de sélection des coefficients est définie par l’Équation 3.13.

$$F_{img}(i, j) = \begin{cases} A_{img}(i, j) & \text{si } T_{ij,A} \geq T_{ij,B} \\ B_{img}(i, j) & \text{sinon} \end{cases} \quad (3.13)$$

Avec  $F_{img}$  le résultat de la fusion des coefficients obtenus des images sources  $A_{img}$  et  $B_{img}$ . Si  $T_{ij,A}$  est plus grand que  $T_{ij,B}$ , alors le pixel situé à la position  $(i, j)$  dans l’image décomposée de  $A$ , a des caractéristiques plus remarquables que le pixel correspondant au même endroit de l’image décomposée de  $B$ . Sinon c’est plus le pixel de l’image  $B_{img}$  qui est pris en compte dans l’image finale.

TABLE 3.3 – Algorithme pour la fusion d’images

---

Entrée : Images à fusionner.
Les images proviennent des bases de données source A et source B
Sortie : Image fusionnée

---

- 1 : Étape 1 : Décomposition de l’image avec NSST
- 2 : Sélectionner aléatoirement des images sources et décomposer chaque image source en cinq niveaux avec NSST.
- 3 : Étape 2 : Stratégie de fusion
- 4 : Fusion des coefficients basse fréquence
- 5 : Normaliser entre 0 et 1 les coefficients de basse fréquence ;
- 6 : Appliquer une combinaison linéaire pour fusionner les coefficients.
- 7 : Fusion des coefficients haute fréquence
- 8 : Normaliser entre 0 et 1 les coefficients de haute fréquence ;
- 9 : Initialiser le seuil dynamique  $\theta$  à 1 pour accélérer l’activation des neurones et les autres modules à 0 ;
- 10 : Stimuler le PCNN respectivement avec les coefficients normalisés ;
- 11 : Lancer le processus itératif de stimulation des neurones ;
- 12 : Sélectionner les coefficients de haute fréquence.
- 13 : Étape 3 : Reconstruction de l’image avec NSST inverse

---

### 3.6.3 Resultats expérimentaux et discussions

#### 3.6.3.1 Base de données

Le jeu de données comprend deux bases d’images. La première base appelées source A, contient des images dermoscopiques de mélanomes et de nævus qui proviennent des bases ISIC2019 et ISIC2020 Challenge [4, 5, 6]. La deuxième base appelée source B, comprends

des images couleur de peaux saines bronzées, foncées et très foncées. La Figure 3.7 présente certaines des images sources utilisées pour les expérimentations.

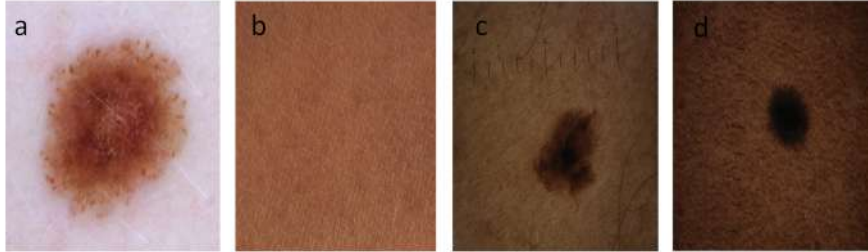


FIGURE 3.7 – (a) Source A : caractéristiques dermoscopiques d’un mélanome sur une peau très claire [4, 5, 6]. (b) Source B : image de peau foncée saine. (c) Caractéristiques dermoscopiques d’un nævus mélanocytaire sur une peau mate [4, 5, 6].(d) Caractéristiques dermoscopiques d’un nævus mélanocytaire sur peau foncée [4, 5, 6].

Pour organiser la base de données de la source A selon les différentes teintes de peaux, nous avons proposé une méthode.

La première étape consiste à délimiter les parties saines des parties affectées de chaque image de lésion de peau en faisant une segmentation.

La deuxième étape de cette méthode consiste à appliquer l’angle typologique individuel ATI (Individual Typology Angle ITA) sur les parties saines des images pour caractériser la teinte de la peau. L’ITA est un outil de classification objectif, qui à partir d’images de peau permet de classer les tons de peau en six catégories : très clair, clair, intermédiaire, bronzé, brun et foncé. Les pixels des régions saines sont convertis en espace CIELab pour obtenir la luminance  $L$  et  $b$  la quantité de jaune dans chaque pixel. La valeur moyenne de l’ITA est exprimée en degrés et est calculée à l’aide de l’Équation 3.14 [132].

$$ATI = \arctan\left(\frac{L - 50}{b}\right) \frac{180}{\pi} \quad (3.14)$$

Il existe une correspondance entre les six catégories peaux obtenu avec l’ITA et les six phénotypes cutanés de Fitzpatrick, présentés dans la section 1.1.3 du Chapitre 1. Le Tableau 3.4 présente la correspondance entre l’ATI, les catégories de peaux, et les six phénotypes cutanés de Fitzpatrick.

En appliquant la méthode basée sur l’ATI, nous observons que la base de données source A est organisée comme suit :

- Peau très claire : 1901 images ;
- Peau claire : 8956 images ;
- Peau intermédiaire : 471 images ;
- Peau bronzée : 51 images ;



TABLE 3.4 – Correspondance entre l’ATI, catégories de peaux et les six phénotypes cutanés de Fitzpatrick.

Plage ATI	Catégories de peaux	Six phénotypes cutanés de Fitzpatrick
$>55^\circ$	Très clair	Très clair (avec des taches de rousseur)
$> 41^\circ$ et $\leq 55^\circ$	Clair	Très clair
$> 28^\circ$ et $\leq 41^\circ$	Intermédiaire	Clair
$> 10^\circ$ et $\leq 28^\circ$	Bronzé	Mate
$> -30^\circ$ et $\leq 10^\circ$	Brun	Foncé
$\leq -30^\circ$	Foncé	Noir

— Peau brune : 13 images ;

— Peau non-défini : 46 images.

Les résultats montrent que les tons foncés sont sous-représentés avec 94,92% d’images de tons clairs (catégories très clair et clair) contre 4,68% d’images de tons foncés (catégories intermédiaire, bronzé, brun et foncé). Matthew Groh et al. dans [133] révèlent également le déséquilibre des types de peau dans la base de données Fitzpatrick 17k et dans les bases de données en général.

Après augmentation de la base de données et en appliquant l’ATI à la base, on obtient la répartition suivante :

— Peau très claire : 1901 images ;

— Peau claire : 8956 images ;

— Peau intermédiaire : 471 images ;

— Peau bronzée : 1000 images ;

— Peau brune : 1000 images ;

### 3.6.3.2 Configuration des expérimentations

La méthode proposée a été comparée à quatre autres méthodes de fusion d’image fréquemment utilisées dans la littérature. Ces quatre méthodes sont : la méthode de transfert de couleur (CT Color Transfer) [133], la méthode basée sur les ondelettes et la représentation clairsemée (DWT-SR Discrete Wavelet Transform and Sparse Representation)[134], la méthode basée sur les ondelettes et le transfert de couleur (DWT-CT Discrete Wavelet Transform and Color Transfer), la représentation clairsemée et le Laplacien modifié par la somme dans le domaine NSCT (NSCT-SR-SML)[123].

La qualité des images obtenues par ces cinq méthodes a été évaluée en utilisant des métriques et les réseaux de neurones à convolution.

Les métriques adoptées pour l’évaluation quantitative et objectives des images fusion-

nées sont :

- La performance de fusion basée sur le gradient  $Q_G$  [135] pour évaluer la quantité d'informations sur les bords qui est transférée des images sources à l'image fusionnée ;
- Les métriques  $Q_S$  [117],  $Q_C$  [135] et  $Q_Y$  [136] pour évaluer les similarités entre les cartes de saillance et les informations structurelles de l'image fusionnée et des images sources ;
- La métrique Chen-Blum QCB [136], qui est une métrique de fusion inspirée de la perception humaine pour évaluer la performance de visualisation humaine des images fusionnées.

Deux réseaux de neurones à convolution, appelés Model 1 et Model 2, inspirés du modèle présenté dans notre précédent travail [34] ont été entraînés. Le Model 1 a été entraîné avec la base de données source A et le Model 2 a été entraîné avec l'ensemble des images générées et l'ensemble des images de la base de données source A. Nous avons comparé l'exactitude (Accuracy), la précision (Precision), le rappel (Recall) et le score F1 (F1-score) de chaque modèle afin d'évaluer l'efficacité de la méthode d'augmentation des données proposée pour la classification des lésions cutanées pour les couleurs de peau sous-représentées dans la base de données.

Tous les algorithmes proposés et utilisés pour l'étude comparative ont été implémentés sur MATLAB R2020b. La décomposition NSST des images sources a été faite sur cinq niveaux. Le nombre d'itérations du PCNN a été fixé à 100 et la matrice  $W$  définie dans l'Équation 3.15 est obtenue en calculant la distance euclidienne entre le neurone central (à la position (2,2) et un neurone voisin (qui peut être à la position (1,1), (1,2), (1,3), (2,1), (2,3), (3,1), (3,2) et (3,3)).

$$W = \begin{bmatrix} 0,707 & 1 & 0,707 \\ 1 & 0 & 1 \\ 0,707 & 1 & 0,707 \end{bmatrix} \quad (3.15)$$

### 3.6.3.3 Résultats et discussions

#### Evaluation visuelle et qualitatives

La Figure 3.8 présente les résultats des cinq méthodes de fusions. Les images de la Figure 3.8(a1)-(a3) indiquent les caractéristiques dermoscopiques pour les peaux très claires et claires. Les images obtenues respectivement par les méthodes de fusion CT, DWT-SR, DWT-CT, NSCT-SR-SML et la méthode proposée NSST-PCNN sont présentées dans la Figure 3.8(b1)-(f3).

Le résultat obtenu par la méthode NSCT-SR-SML n'est pas naturel et ne présente pas

de caractéristiques communes aux caractéristiques dermoscopiques réelles. En observant dans la Figure 3.8 les images (b1)-(b3) générées avec DWT-SR et les images (f1)-(f3) générées en utilisant la méthode proposée, on constate que la méthode NSST-PCNN combine mieux les images sources et préserve distinctement les structures dermoscopiques comme les réseaux de pigments, les zones amorphes sans structure (tâches), les points et les globules entre les images dermoscopiques. Comparé aussi aux résultats de la CT et de la DWT-CT, la méthode proposée présente une meilleure combinaison des teintes de peaux fusionnées.

Visuellement, on remarque une variation du contraste et de la pigmentation des lésions dermoscopiques (Figure 3.8(a1)-(a3)) par rapport aux images fusionnées avec le PCNN-NSST (Figure 3.8(f1)-(f3)). Selon Lallas A. [125], la différence de contraste et de pigmentation observé selon la couleur de peau est naturelle car les lésions cutanées sur la peau foncées sont caractérisées par une hyperpigmentation centrale et un réseau périphérique brun foncé.

Bien que les résultats de l'évaluation visuelle montrent que les structures dermoscopiques sont préservées, il convient de noter que l'évaluation visuelle est une méthode subjective. Le Tableau 3.5 présente les résultats de cinq métriques appliquées aux différentes méthodes. Les métriques  $Q_G$ ,  $Q_S$ ,  $Q_C$ ,  $Q_Y$  évaluent la quantité d'informations sur les bords transférées, et la similitude entre les images fusionnées et les images sources. La méthode évaluée avec ces métriques est très performante si les valeurs tendent vers 1.

La méthode NSST-PCNN, comparée aux autres méthodes, et sur toutes les images générées, présente de meilleurs résultats pour les métriques  $Q_G$ ,  $Q_S$ , et  $Q_C$ . Cette performance est suivie par celles des méthodes DWT-CT et DWT-SR, qui présentent de meilleurs résultats pour les métriques  $Q_C$  et  $Q_Y$  respectivement. D'après les résultats obtenus, nous pouvons conclure que les images générées par le NSST-PCNN préservent mieux les détails et la similarité avec les images sources. De plus, la faible valeur de la métrique de performance de visualisation humaine  $Q_C B$ , confirme la différence visuelle de pigmentation observée entre les images réelles et les images générées par la NSST-PCNN. Nous pouvons donc avancer que les images fusionnées ne sont pas juste des copies d'images de lésions cutanées mais contiennent de vraies informations qui peuvent améliorer la performance des outils d'aide au diagnostic.

Le Tableau 3.6 montre le temps moyen pris par chaque algorithme pour générer une image. Les résultats indiquent que la méthode proposée a un temps d'exécution plus long que la plupart des méthodes. Cette valeur élevée peut s'expliquer par le choix de calculer de manière adaptive la force de liaison.

### **Évaluation avec réseaux de neurones**

Nous avons choisi d'apprécier l'impact de l'augmentation des données dans la classification

TABLE 3.5 – Résultats des métriques d'évaluation des différentes méthodes.

Images générées	Méthodes	QG	QS	QC	QY	QCB
b1	CT	0.7390	0.8080	0.7910	0.7820	0.5470
c1	DWT_SR	0.4990	0.6750	0.8330	0.5630	0.5530
d1	DWT_CT	0.6600	0.7700	0.8400	0.7170	0.5050
e1	NSCT_SR_SML	0.4440	0.5780	0.6410	0.4190	0.4450
<b>f1</b>	<b>NSST_PCNN</b>	<b>0.8060</b>	<b>0.8390</b>	<b>0.8810</b>	<b>0.8240</b>	<b>0.6190</b>
b2	CT	0.7730	0.9430	0.8900	0.9140	0.6310
c2	DWT_SR	0.6660	0.7560	0.8170	0.7900	0.6280
d2	DWT_CT	0.7540	0.7280	0.9320	0.9050	0.6130
e2	NSCT_SR_SML	0.4920	0.6200	0.8640	0.5050	0.4690
<b>f2</b>	<b>NSST_PCNN</b>	<b>0.8030</b>	<b>0.8800</b>	<b>0.9680</b>	<b>0.9750</b>	<b>0.6580</b>
b3	CT	0.4500	0.8380	0.7530	0.7610	0.4840
c3	DWT_SR	0.5930	0.8090	0.8020	0.7320	0.4590
d3	DWT_CT	0.6900	0.7930	0.7430	0.7570	0.5000
e3	NSCT_SR_SML	0.4880	0.7390	0.4280	0.5330	0.4800
<b>f3</b>	<b>NSST_PCNN</b>	<b>0.7690</b>	<b>0.8470</b>	<b>0.8090</b>	<b>0.8450</b>	<b>0.5740</b>
Toutes les images	CT	0.5760	0.8350	0.7640	0.7810	0.5260
	DWT_SR	0.5080	0.7190	0.7910	<b>0.8440</b>	0.5190
	DWT_CT	0.6360	0.7360	<b>0.8380</b>	0.6570	0.5120
	NSCT_SR_SML	0.3970	0.6180	0.5970	0.4480	0.4370
	<b>NSST_PCNN</b>	<b>0.7150</b>	<b>0.8280</b>	0.7700	0.7550	<b>0.5890</b>

TABLE 3.6 – Temps d'exécution.

Méthodes	CT	DWT_SR	DWT_CT	NSCT_SR_SML	NSST_PCNN
Times/s	<b>0.6685</b>	13.1643	6.9492	24.2362	23.9312

des lésions cutanées sur tous les types de peau en utilisant des réseaux de neurones à convolution. Les modèles utilisés sont inspirés du réseau de neurones présenté dans le chapitre 2.

80 % de l'ensemble de données a servi pour l'entraînement du modèle, et 20 % pour le test. L'exactitude (Accuracy), la précision (Precision), le rappel (Recall) et le score F1 (F1-score) du modèle 1 sont présentés dans le Tableau 3.7.

Le Modèle 1 a été renforcé avec les images générée du NSST-PCNN, puis testé avec les mêmes 20% de l'ensemble de données. Pour vérifier la généralisation du réseau neuronal, les modèles ont été testés uniquement avec des images réelles. L'exactitude (Accuracy), la précision (Precision), le rappel (Recall) et le score F1 (F1-score) du Modèle 2 sont présentés dans le Tableau 3.8.

La Figure 3.9 présente les résultats des modèles pour chacune des métriques. Le Modèle 2 comparé au Modèle 1 gagne en exactitude de 2,22%, 0,73%, 1,67%, 15,69% et 15,38%

TABLE 3.7 – Résultats des mesures d’évaluation du Modèle 1.

Métriques	Très claire	Claire	Intermédiaire	Bronzée	Foncée
Accuracy	0.9434	0.9480	0.9333	0.7255	0.7692
Precision	0.9091	0.9297	0.8900	0.7826	0.7500
Recall	0.9662	0.9557	0.9082	0.6667	1.0000
F1-score	0.9368	0.9425	0.8990	0.7200	0.8571

TABLE 3.8 – Résultats des mesures d’évaluation du Modèle 2.

Métriques	Très claire	Claire	Intermédiaire	Bronzée	Foncée
Accuracy	0.9518	0.9553	0.9500	0.8824	0.9231
Precision	0.9227	0.9365	0.9300	0.9130	0.9167
Recall	0.9713	0.9650	0.9208	0.8400	1.0000
F1-score	0.9464	0.9505	0.9254	0.8750	0.9565

respectivement pour les catégories de peaux très claire, claire, intermédiaire, bronzée et foncée. La précision du modèle a également augmenté de 1,36%, 0,68%, 4,00%, 13,04% et 16,67% respectivement pour les mêmes catégories citées plus haut. Aussi les résultats des autres métriques indiquent en général une meilleure performance du Modèle 2 et cette performance est plus remarquable dans les catégories de peaux foncées. L’augmentation des données a donc favorisé le renforcement et la généralisation du classifieur.

Comme pour les modèles basés sur le GAN proposés par Zhiwei Qin et al. dans [107] et Alceu Bissoto et al. dans [109], l’augmentation des données a amélioré le modèle, avec la particularité que l’augmentation des données proposée peut permettre de corriger le déséquilibre des teintes de la peau observé. L’introduction de nouvelles images inspirées d’images réelles mais contenant également d’autres caractéristiques, a permis de favoriser la généralisation du classifieur.

## Conclusion

Les méthodes de classification basées sur le deep Learning nécessitent beaucoup d’images et il est difficile d’avoir de larges bases de données d’images médicales. Pour pallier cette problématique, des méthodes d’augmentation de données simples à plus complexes sont proposées dans la littérature.

Dans ce chapitre nous avons présenté la revue de littérature sur le sujet et aussi proposé une méthode d’augmentation de données nommée NSST-PCNN pour corriger le déséquilibre observé par rapport à la représentation de tous les types de peaux dans la base de données d’image dermoscopiques.

La méthode NSST-PCNN est basée sur la décomposition d’images multi-échelles et la stratégie de fusion PCNN. Comparée aux méthodes existantes, les images de la méthode

proposée présentent une structure et des détails dermoscopiques plus informatifs. En particulier, la méthode proposée a l'avantage d'être adaptée aux images dermoscopiques. Les résultats des expériences prouvent également que cette méthode a amélioré les performances et la précision d'un classificateur de lésions cutanées basé sur un réseau de neurones à convolution, surtout pour les catégories de peaux sous-représentées dans la base données. En conclusion, ce travail propose une solution qui ne nécessite pas d'entraînement et améliore efficacement la classification d'images dermoscopiques.

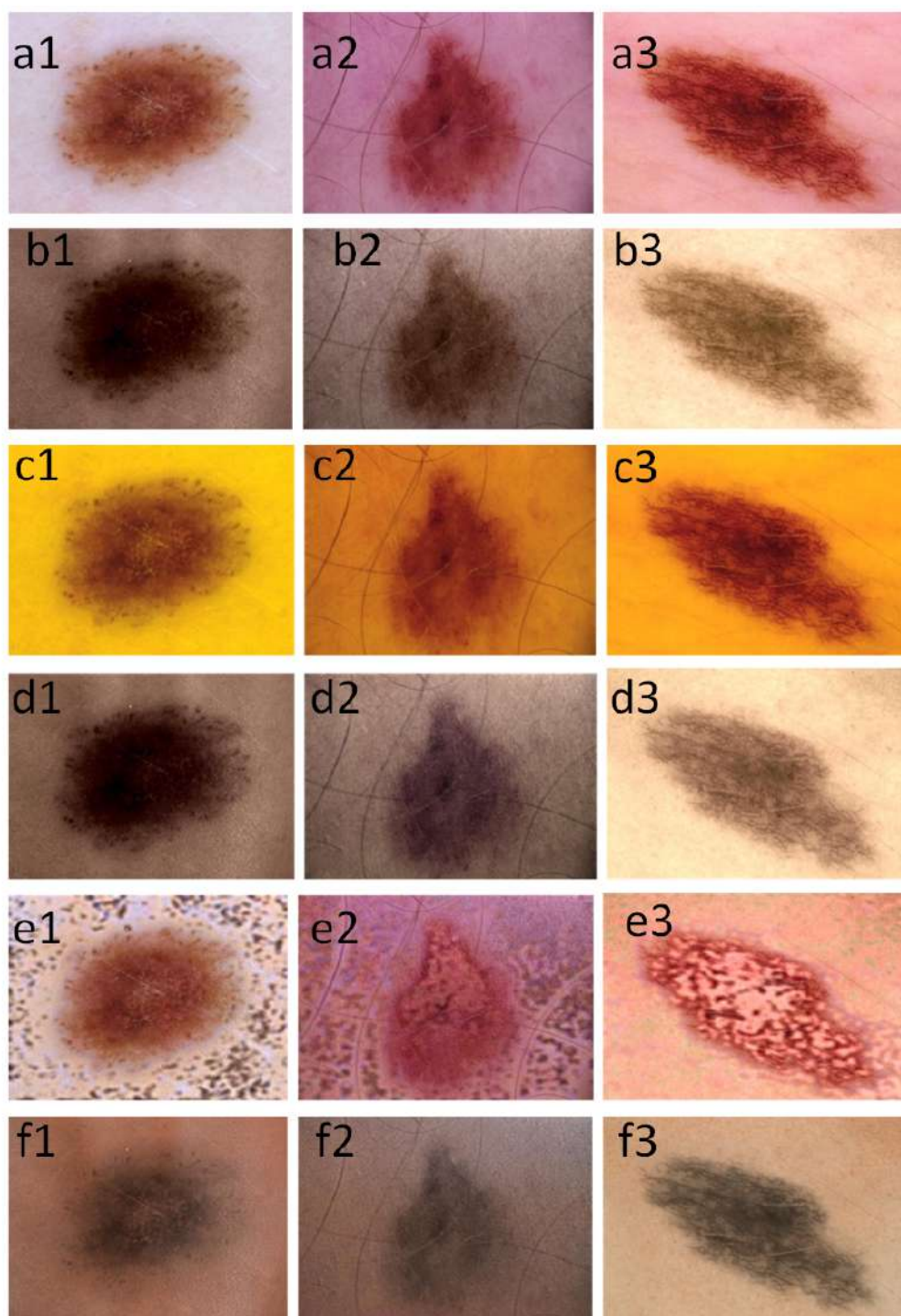


FIGURE 3.8 – (a1)–(a3) Images dermoscopiques de tons de peaux clairs provenant la base de données source A [9, 10, 11]. (b1-f3) Résultats obtenus en utilisant respectivement CT, DWT-SR, DWT-CT, NSCT-SR-SML, et la méthode proposée PCNN-NSST.

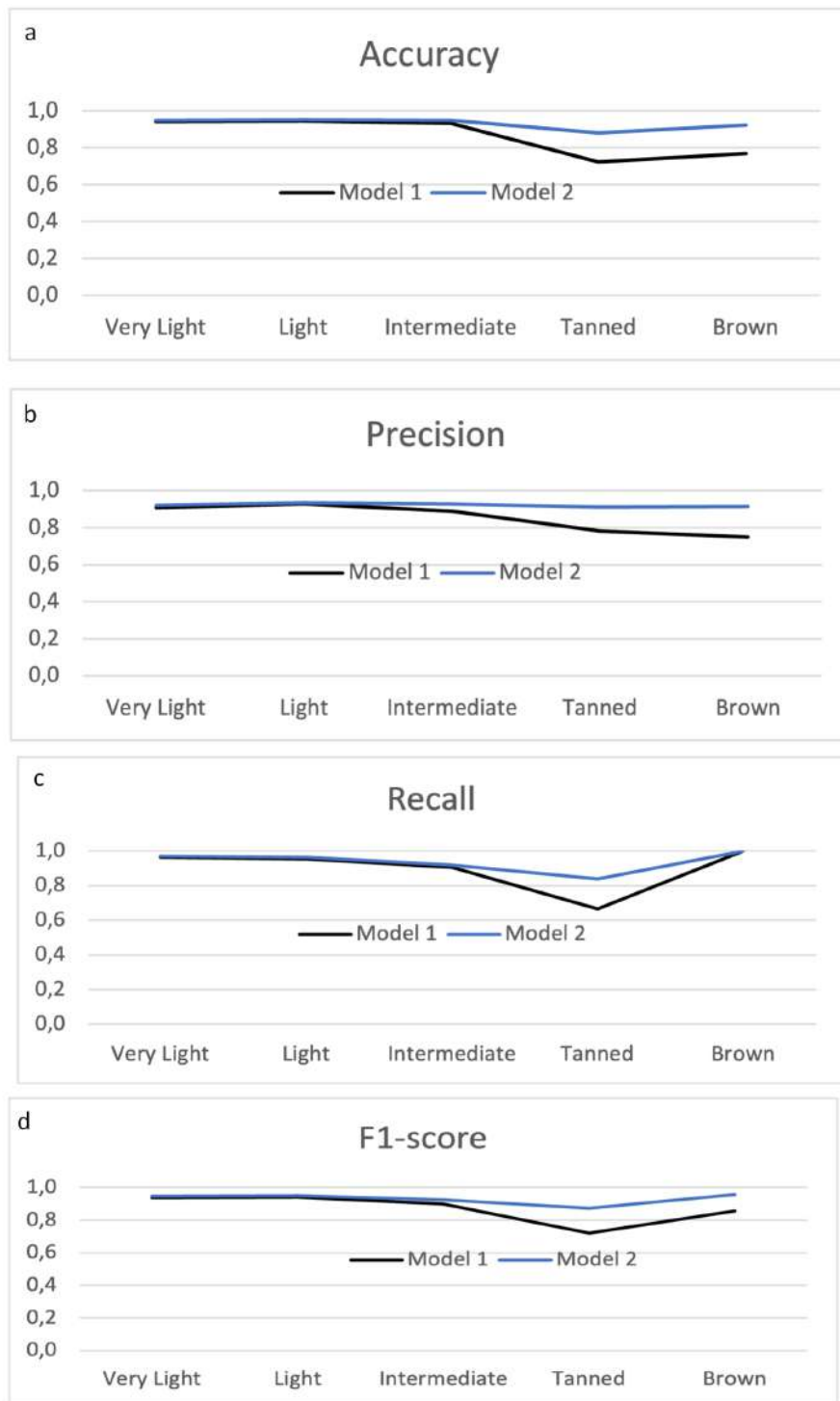


FIGURE 3.9 – Comparaison des résultats des métriques d'évaluation. (a) Accuracy, (b) Precision, (c) Recall et (d) F1-score.



## CHAPITRE 4

# DÉTECTION AUTOMATIQUE DE CARACTÉRISTIQUES DERMOSCOPIQUES

### Sommaire

---

<b>4.1</b>	<b>Détection basée sur un apprentissage traditionnel</b>	<b>70</b>
4.1.1	Histogramme des gradients orientés	71
4.1.2	Méthode de Viola-Jones	71
4.1.3	Modèles de pièces déformables	72
4.1.4	Correspondance des modèles	72
<b>4.2</b>	<b>Détections basées sur l'apprentissage profond</b>	<b>73</b>
4.2.1	Architecture des détecteurs d'objets basés sur le deep learning	73
4.2.2	Littérature sur la détection de caractéristiques dermoscopiques basée sur l'apprentissage profond	74
<b>4.3</b>	<b>Comparaison des méthodes de détection d'objets</b>	<b>75</b>
<b>4.4</b>	<b>Méthode proposée : Yolo-dermoscopic-features</b>	<b>76</b>
4.4.1	Yolov8	76
4.4.2	Algorithme	79
4.4.3	Résultats expérimentaux et discussions	81

---

### Introduction

Les algorithmes de détection d'objets permettent de déterminer l'emplacement des objets dans une image donnée, c'est à dire la localisation et la classification des objets. Ces algorithmes sont appliqués dans des systèmes temps réels tels que la détection de visages, la détection de piétons, la détection de véhicules, la détection de panneaux de signalisation, la vidéo surveillance, ainsi que dans des systèmes de santé, pour détecter des tumeurs, des lésions ou des noyaux dans des images de microscopie.

Sur des images dermoscopiques, les caractéristiques détectées peuvent être réticulaire (réseau pigmentaire "pigment network", réseau négatif "negatif network"), globulaire, ou étoilé

(stries ou "streaks"). La détection visuelle de ces caractéristiques est une tâche complexe en raison de leur très petite taille dans la lésion. D'où l'intérêt d'appliquer des outils de traitement d'images pour améliorer la précision de détection. Fournir ces informations aux spécialistes peut les aider à comprendre pleinement l'image de la lésion et à améliorer leurs précisions de diagnostics.

Les progrès récents de ces algorithmes ont été rendus possibles avec l'avènement des méthodologies d'apprentissage profond (deep Learning). Mais avant que l'apprentissage profond ne soit utilisé dans la détection d'objets, des méthodes traditionnelles basées sur l'extraction manuelle de caractéristiques étaient utilisées. La Figure 4.1 répertorie quelques méthodes de ces deux familles de méthodes de détection d'objets.

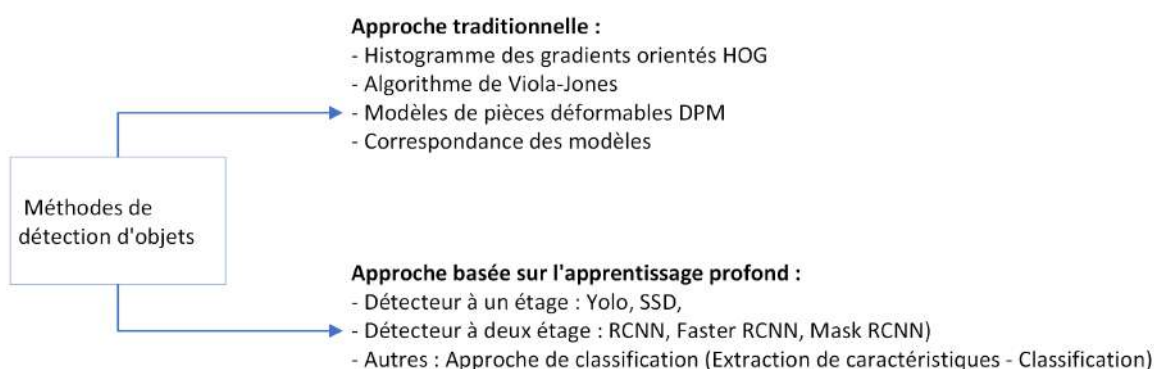


FIGURE 4.1 – Méthodes de détection d'objets.

Dans la suite de ce chapitre, nous présentons les deux familles de méthodes, la méthode proposée pour améliorer la détection automatique de caractéristiques dermoscopiques, les résultats expérimentaux et de la discussion. Ce chapitre se termine par une conclusion.

## 4.1 Détection basée sur un apprentissage traditionnel

Avant que l'apprentissage profond ne soit utilisé dans la détection d'objets, il existait plusieurs méthodes de détection d'objets basées sur des techniques traditionnelles de vision par ordinateur et des algorithmes d'apprentissage automatique (machine Learning). Ces méthodes de détection d'objets sans apprentissage profond comprennent : Histogramme des gradients orientés (HOG Histogram of Oriented Gradients) [137], algorithme de Viola-Jones [138], modèles de pièces déformables (DPM Deformable Part Models) [139], et correspondance des modèles (Template Matching) [140].

### 4.1.1 Histogramme des gradients orientés

L’histogramme des gradients orientés (HOG : Histogram of Oriented Gradients) [137] est un descripteur de caractéristiques utilisé en vision artificielle et en traitement d’images pour la détection et la reconnaissance d’objets. HOG permet de caractériser les formes et les structures des objets par la distribution des orientations de gradient dans une image. L’algorithme du descripteur HOG pour capturer les informations de gradient local dans une image, suit plusieurs étapes. L’image est d’abord divisée en petites régions connectées appelées cellules et, pour les pixels de chaque cellule, un histogramme des directions de gradient est compilé. Ensuite pour tenir compte des variations locales d’éclairage et de contraste, les valeurs de l’histogramme sont normalisées dans chaque bloc, qui est un groupe de cellules. Cette normalisation peut être réalisée à l’aide de méthodes telles que la norme L1 ou la norme L2. Enfin, le descripteur HOG final est la concaténation des histogrammes normalisés de toutes les cellules de l’image. Cela crée une représentation compacte qui capture les informations de gradient de l’image entière.

Le descripteur HOG résultant peut être utilisé comme vecteur de caractéristiques pour entraîner des algorithmes d’apprentissage automatique, tels que les machines à vecteurs de support (SVM) [141], pour des tâches de détection et de reconnaissance d’objets.

### 4.1.2 Méthode de Viola-Jones

La méthode Viola-Jones [138] est un algorithme classique de détection d’objets utilisé en particulier pour la détection de visages, mais aussi pour d’autres tâches de détection d’objets. Cette méthode consiste à diviser l’image en sous-régions rectangulaires appelées fenêtres de détection, puis à calculer les caractéristiques Haar qui contiennent les informations de variations d’intensité dans une image.

Ces caractéristiques sont ensuite appliquées à une série de classifieurs en cascade pour décider si elles contiennent ou non l’objet recherché. Chaque classifieur est formé à l’aide d’un algorithme d’apprentissage supervisé, Adaboost (Adaptive Booster) par exemple, qui sélectionne les caractéristiques les plus discriminantes pour la tâche de détection.

Barata et al. dans [142] se sont intéressés à la détection automatique de réseaux pigmentaires en proposant une méthode proche de l’algorithme de Viola Jones. La première étape consiste à appliquer des filtres directionnels pour mieux détecter les lignes du réseau pigmentaire. Cette étape est suivie de l’extraction de caractéristiques et de classification avec l’algorithme AdaBoost pour détecter les grandes structures liées qui représentent la zone du réseau pigmentaire.

### 4.1.3 Modèles de pièces déformables

Les modèles de pièces déformables (DPM Deformable Part Models) [139] sont une méthode de détection d'objets qui est particulièrement adaptée à la détection d'objets complexes avec une structure articulée. Ils capturent à la fois la forme globale de l'objet et les relations spatiales entre ses parties constitutives. La DPM représente les objets comme une collection de parties, où chaque partie est associée à un modèle d'apparence spécifique. Les parties sont généralement définies comme des taches ou des régions locales de l'image qui sont caractéristiques de la classe d'objet détectée.

Le DPM intègre un modèle de déformation pour tenir compte des relations spatiales et des variations géométriques entre les différentes parties. Ce modèle définit la disposition spatiale attendue des pièces les unes par rapport aux autres, ce qui permet une certaine souplesse dans leur position et leur orientation. Les caractéristiques HOG (ou des filtres de Gabor) sont ensuite calculées à la fois pour la région globale de l'objet et pour chaque région individuelle de la pièce.

Cette méthode nécessite un apprentissage supervisé avec un ensemble d'images d'entraînement annotées qui contiennent des annotations de boîtes de délimitation pour les instances d'objets. Le processus de formation implique l'apprentissage des modèles d'apparence de l'objet et de ses parties, ainsi que l'estimation des paramètres de déformation qui définissent leurs relations spatiales. Au cours de la phase de détection, le DPM balaie l'image à l'aide d'une fenêtre coulissante à la recherche d'instances d'objets potentiels. À chaque déplacement de la fenêtre, les caractéristiques sont calculées et un score de détection est généré. Les scores de détection sont combinés pour déterminer la présence et l'emplacement de l'objet dans l'image.

### 4.1.4 Correspondance des modèles

La mise en correspondance de modèles [140] est une méthode simple et intuitive qui consiste à comparer une image modèle à l'image d'entrée en faisant glisser le modèle sur l'image et en mesurant la similarité. Les mesures de similarité courantes comprennent la corrélation croisée, la somme des différences quadratiques (SSD Sum of Squared Differences) et la corrélation croisée normalisée (NCC Normalized Cross-Correlation). Cette méthode est efficace pour détecter des objets ayant des apparences distinctes, mais est sensible aux variations d'échelle, à la rotation et aux conditions d'éclairage.

Benam et al. dans [143] ont introduit un système qui permet de rechercher un motif (modèle) classé comme réseau pigmentaire dans des images qui présente la caractéristique. Le principe consiste à faire une correspondance entre le motif recherché et les zones de l'image qui présentent le motif, en utilisant l'algorithme de détection de points d'intérêt SURF (Speeded-Up Robust Features)[143].

## 4.2 Détections basées sur l'apprentissage profond

Les détecteurs d'objets basés sur le deep learning peuvent être divisés en plusieurs types mais plus principalement en types en fonction du nombre d'étapes impliquées dans le processus de détection : les détecteurs à une étape (one stage detector) et les détecteurs à deux étapes (two stage detector).

Les détecteurs à une étape utilisent un seul réseau neuronal pour prédire à la fois les classes d'objets et les boîtes englobantes. YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), et EfficientDet sont des exemples de détecteurs en une étape. Quant aux détecteurs à deux étapes, ils utilisent une approche en deux étapes pour détecter des objets dans une image. Dans un premier temps, ces algorithmes génèrent un ensemble de régions d'intérêt candidates, appelées propositions de régions (region candidates). Dans la deuxième étape, ces régions sont classées et les coordonnées de la boîte englobante des objets situés à l'intérieur des régions sont affinées. Le R-CNN rapide (Region-based Convolutional Neural Networks), Faster R-CNN, et Mask R-CNN est sont des exemples de détecteurs en deux étapes.

### 4.2.1 Architecture des détecteurs d'objets basés sur le deep learning

Les détecteurs d'objets basés sur le deep learning du point de vue des composants, intègrent généralement deux parties, un réseau dorsal (ou épine dorsale) basé sur les réseaux convolutionnels et utilisé pour l'extraction des caractéristiques de l'image, et un réseau de tête de détection utilisée pour prédire la classe et la boîte de délimitation de l'objet. En outre, les détecteurs d'objets développés ces dernières années insèrent souvent des couches entre l'épine dorsale et la tête, nommée réseau de cou (ou cou du détecteur).

- Le réseau dorsal est un module d'extraction de caractéristiques basé sur une architecture d'apprentissage en profondeur. Tous les modèles de l'épine dorsale sont fondamentalement des modèles de classification tels que le réseau VGGNet (Visual Geometry Group à crée VGGNet) [144], le bloc de transformation résiduelle (ResNet) [145], le réseau convolutionnel dense (DenseNet) [83] et l'approche du réseau partiel à étapes croisées pour Darknet (CSPDarknet Cross Stage Partial Network) [146] plutôt que des réseaux convolutionnels construits à zéro.
- Le "cou" est un ensemble de couches qui effectuent généralement des opérations d'agrégation spatiale et de fusion de caractéristiques sur la sortie de l'épine dorsale. L'objectif du cou est de combiner des informations provenant de différentes échelles et de différents niveaux d'abstraction afin d'améliorer la capacité du réseau à détecter des objets de tailles et de formes variées. Les blocs d'agrégation de chemins couramment utilisés dans le cou sont les suivants : le réseau pyramidal de caractéristiques (FPN Feature Pyramid Networks)[147], le réseau d'agrégation

tion de chemins (PAN Path Aggregation Network) [148]. Le point commun de ces méthodes est l'utilisation répétée de divers échantillonnages ascendants et descendants, de sommes de points ou de produits de points pour concevoir des stratégies d'agrégation.

- Le réseau de tête est responsable de la détection des objets et se compose de couches de convolutions, suivies d'une couche de prédiction. La couche de prédiction produit des boîtes de délimitation, des scores d'objectivité et des probabilités de classe. Parmi ces réseaux, il y a le réseau de proposition de régions (RPN Region Proposal Network) [149], Yolo [150], et R-CNN [149].

## 4.2.2 Littérature sur la détection de caractéristiques dermoscopiques basée sur l'apprentissage profond

Plusieurs travaux se sont intéressés à localisation de caractéristiques dermoscopiques en utilisant des détecteurs à deux étapes ou des approches en deux étapes proches de celle du détecteur à deux étapes. Jahanifar et al. [151] ont proposé un algorithme de segmentation à cinq attributs basés sur une architecture codeur-décodeur pour segmenter chaque attribut dermoscopique. Sorokin et al. [152] ont proposé une méthode de détection des attributs des lésions basée sur un modèle R-CNN de masque. Le modèle a été pré-entraîné sur la base de données COCO [153] et détecte cinq attributs dermoscopiques. Li et al. [154] ont proposé une méthode d'extraction des caractéristiques dermoscopiques appelée Lesion Feature Network (LFN). La méthode LFN consiste à subdiviser l'image en 996 zones de superpixels, et chaque zone de superpixels est ensuite classée dans l'une des cinq catégories suivantes : quatre types de caractéristiques dermoscopiques (réseau pigmentaire, réseau négatif, stries et kystes de type milia) et l'arrière-plan. Kawahara et al. [155] ont mis en œuvre une méthode basée sur un réseau de neurones à convolution pour détecter la présence de quatre critères dermoscopiques (réseau pigmentaire, réseau négatif, kystes de type milia, stries) dans une image dermoscopique. Leur approche a remporté la première place au concours ISIC- ISBI Part 2 : Dermoscopic Feature Classification Task challenge de 2017. Bissoto et al. [156] ont abordé la tâche de détection des attributs des lésions avec un réseau Inception-v4 finement réglé et pré-entraîné sur ImageNet [157]. Chaque image a été découpée en parcelles de dimensions 128x128 ou 299x299 et chaque parcelle a été soumise au modèle pour être classée en six classes : absence (réseau pigmentaire), réseau négatif, kystes de type milia, stries et globules. Z. Chen et al. [158] ont proposé un modèle U-Net multi-tâches pour détecter automatiquement les attributs dermoscopiques. La caractéristique du modèle proposé est que le chemin d'encodage U-Net a été remplacé par un modèle VGG16 pré-entraîné. Au lieu d'utiliser le modèle pré-entraîné avec des images des bases de données comme ImageNet ou COCO, qui sont très différentes des images médicales, Nguyen et al. [159] ont proposé un apprentissage par transfert agnostique (TATL) pour la détection des attributs des lésions cutanées. La

méthode proposée s’inspire du comportement des dermatologues et se compose de trois réseaux codeur-décodeur. Le premier réseau, Segment-Net, segmente les régions des lésions dans l’image. Ensuite, le deuxième réseau, Attribute-Agnostic Segmenter, identifie toute position anormale dans l’image segmentée. Enfin, un réseau Target-Segmenter est formé pour détecter un seul attribut. Pour les cinq caractéristiques dermoscopiques, cinq réseaux Target-Segmenter différents ont été utilisés.

Toutes ces méthodes présentées ont abordé la détection des caractéristiques dermoscopiques comme une tâche de classification [156] ou de segmentation [151], en travaillant soit sur l’image entière [152], [154], soit sur la lésion segmentée [155, 159], soit sur des superpixels [154, 155, 156]. Des méthodes de pré-traitement ont été appliquées par certains auteurs [152] et [154] pour agrandir la zone de la lésion en vue de la détection des caractéristiques, et pour supprimer les artefacts présents sur certaines images. À la fin des méthodes proposées, d’autres auteurs [151] et [156] ont appliqué un post-traitement pour restreindre les cartes de prédiction d’attributs à la zone de lésion et pour atténuer les faux positifs qui se produisent généralement dans les attributs abondants.

La plupart des travaux cités exploitent les bases de données fournies par le concours ISIC (ISIC 2017 ou ISIC 2018). Ces ensembles de données ont le potentiel de faire avancer la recherche sur le sujet. Cependant, toutes les structures dermoscopiques ne sont pas annotées dans ces bases de données et le nombre d’exemples pour la plupart des structures annotées est faible. En outre, pour les dermatologues, fournir des annotations locales est une tâche lourde, fastidieuse et chronophage [18].

Tout en tenant compte de ces difficultés, ce travail a pour objectif de proposer une méthode pour améliorer la détection automatique des caractéristiques dermoscopiques.

### 4.3 Comparaison des méthodes de détection d’objets

Le Tableau 4.1 présente les avantages et inconvénients des méthodes de détection d’objets classiques, R-CNN, Fast-RCNN et Yolo.

Le choix de l’une ou l’autre de ces approches dépend du type d’application (temps réel ou non), de la taille des caractéristiques recherchées, de la résolution des images et de la taille de la base de données. En raison de son architecture (détection d’objets en une seule fois), l’algorithme YOLO est plus rapide que les autres approches. Cette méthode est donc idéale pour les systèmes en temps réel.

Les caractéristiques dermoscopiques présentent de petites structures complexes (globules  $>0,1$  mm, points  $<0,1$  mm [55]). Compte tenu des propriétés des méthodes R-CNN et YOLO, ces deux algorithmes sont bien adaptés à la recherche de structures complexes telles que celles que l’on trouve dans les images dermoscopiques.

Dans la section 4.4, qui présente la méthode proposée, nous nous intéresserons à la

TABLE 4.1 – Comparaison de méthodes de détection d’objets

Méthodes	Avantages	Inconvénients
<b>Approche traditionnelle</b>	<ul style="list-style-type: none"> <li>- Précis lorsque les paramètres sont bien définis et adaptés à la tâche spécifique ;</li> <li>- Efficace pour détecter de petits objets dans des images à haute résolution.</li> </ul>	<ul style="list-style-type: none"> <li>- Généralement lent en raison du traitement séquentiel des régions proposées ;</li> <li>- Nécessite un prétraitement manuel pour extraire les régions d’intérêt ;</li> <li>- Le temps de traitement varie en fonction du nombre d’objets candidats.</li> </ul>
<b>R-CNN</b>	<ul style="list-style-type: none"> <li>- Extraction automatique des caractéristiques des régions proposées avec les CNN ;</li> <li>- Amélioration de la précision ;</li> <li>- Adaptable aux images de petite et grande résolution.</li> </ul>	<ul style="list-style-type: none"> <li>- Lenteur due à l’extraction indépendante des caractéristiques pour chaque région ;</li> <li>- Besoin de ressources importantes en termes de mémoire et de temps de calcul ;</li> <li>- Le processus en plusieurs étapes (extraction des RoI, CNN pour chaque RoI, classification) rend l’entraînement complexe.</li> </ul>
<b>YOLO</b>	<ul style="list-style-type: none"> <li>- Très rapide, car il prédit simultanément les boîtes de délimitation et les classes pour l’ensemble de l’image ;</li> <li>- Convient à la détection d’objets en temps réel ;</li> <li>- Adapté à la détection d’objets de petite taille et se chevauchant.</li> </ul>	<ul style="list-style-type: none"> <li>- Performance variable en fonction de la densité d’objets ;</li> <li>- Nécessité de données d’entraînement représentatives ;</li> <li>- Besoin de ressources importantes en termes de mémoire et de temps de calcul.</li> </ul>

méthode de détection YOLO, plus précisément à la version 8 de cette méthode, qui permet de détecter efficacement les petits objets quelle que soit la résolution de l’image.

## 4.4 Méthode proposée : Yolo-dermoscopic-features

### 4.4.1 Yolov8

Yolo (You Only Look Once) est un modèle de détection d’objets et de segmentation d’images développé par Joseph Redmon et Ali Farhadi [160] en 2015. Depuis ce temps, des mises à jour ont été apportées à l’architecture et aux fonctions de coût afin d’améliorer sa précision et ses performances. La Figure 4.2 présente la frise chronologique de différentes versions de YOLO.

Dans la dernière version de Yolo, les réseaux CSPDarknet, PAFPN et Yolo sont utilisés respectivement comme réseau de base, réseau de cou et réseau de tête, comme le montre la





FIGURE 4.2 – Frise chronologique de différentes versions de YOLO.

Figure 4.3. L’objectif en combinant ces réseaux est de rendre l’apprentissage et l’inférence beaucoup plus rapides, et d’améliorer la précision du modèle. Ce sont ces raisons qui ont motivées le choix de l’algorithme Yolo dans ce travail.

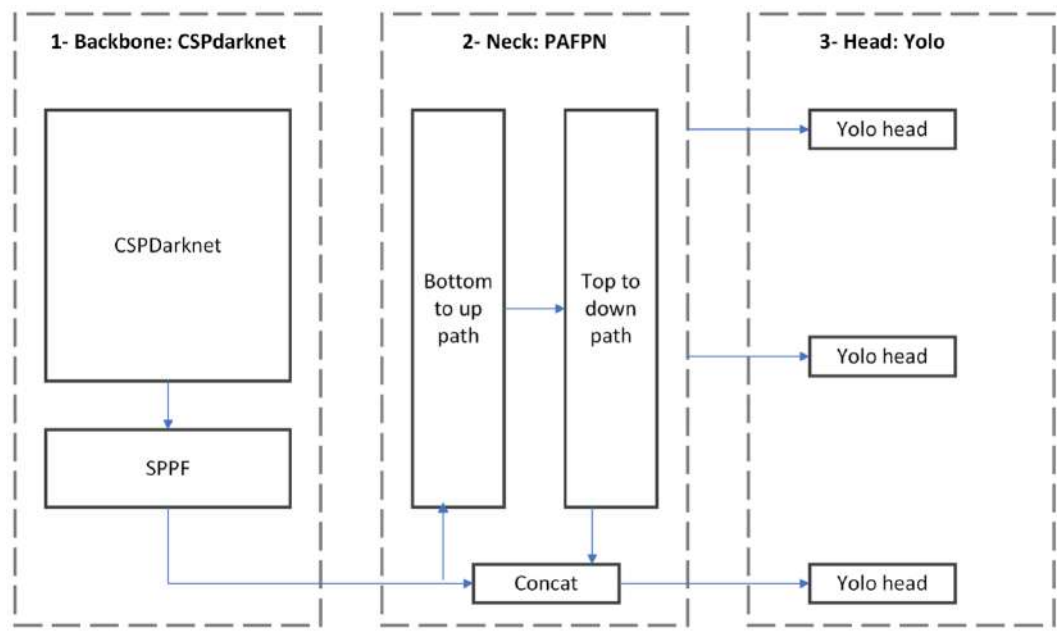


FIGURE 4.3 – Yolo Architecture [12].

La fonction de perte (fonction de perte totale) dans l’algorithme de détection d’objets YOLO est une combinaison de trois fonctions de perte : la perte de classification, la perte de localisation et la perte de confiance.

- perte de classification ( $cls_{loss}$ ) : mesure la précision de la prédiction de la classe pour chaque objet dans l’image. Elle est calculée en utilisant la fonction softmax et la perte de cross-entropy. Cette fonction est définie par l’Équation 4.1 :

$$ClassificationLoss = \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c \in classes}^B (p_i(c) - \hat{p}_i(c))^2 \quad (4.1)$$

avec :

$1_{ij}^{obj} = 1$  si un objet apparaît dans la cellule  $i$ , sinon 0.

$\hat{p}_i(c)$  représente la probabilité conditionnelle de la classe  $c$  dans la cellule  $i$ .

- perte de localisation (*segloss*) : mesure l'erreur de prédiction des coordonnées de la boîte englobante  $(x, y, w, h)$  d'un objet dans une image. Elle est calculée comme la différence entre les coordonnées prédites et réelles de la boîte englobante et est généralement pondérée par le score de confiance de l'objet. Cette fonction est définie par l'Équation 4.2 :

$$\begin{aligned} CoordinateLoss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \\ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i}) + (\sqrt{h_i} - \sqrt{\hat{h}_i})] \end{aligned} \quad (4.2)$$

avec :

$1_{ij}^{obj} = 1$  si un objet est présent dans la  $j_{eme}$  boîte de délimitation (ayant l'indice d'intégrité le plus élevé) et dans la cellule  $i$ , sinon 0. Ce paramètre est utilisé pour moduler la perte en fonction de la présence d'un objet dans une cellule.

$\lambda_{coord}$  augmente le poids de la perte dans les coordonnées de la boîte englobante. Par défaut,  $\lambda_{coord} = 5$ .

- perte de confiance (box-loss) : mesure la probabilité que la boîte englobante prédite contienne un objet. Cette fonction de perte permet d'entraîner le modèle à faire des prédictions plus précises sur l'emplacement et la classe des objets dans une image. Elle est calculée à l'aide de la perte d'entropie croisée binaire et pénalise le modèle lorsqu'il prédit des scores d'objectivité faibles pour les vrais positifs et des scores d'objectivité élevés pour les faux positifs. Cette fonction est définie par l'Équation 4.3 :

$$ConfidenceLoss = \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(C_i - \hat{C}_i)^2] + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} [(C_i - \hat{C}_i)^2] \quad (4.3)$$

avec :

$\hat{C}_i$  le score de confiance de la case  $j$  dans la cellule  $i$ .

$1_{ij}^{obj} = 1$  si un objet est présent dans la  $j_{eme}$  boîte de délimitation et dans la cellule  $i$ , sinon 0. Ce paramètre est utilisé pour moduler la perte en fonction de la présence

d'un objet dans une cellule.

$1_{ij}^{noobj}$  est le complément de  $1_{ij}^{obj}$ .

$\lambda_{noobj}$  pondère la perte en cas de détection de l'arrière-plan (absence d'objet). Par défaut  $\lambda_{noobj} = 0,5$ .

La fonction de perte totale Yolo est calculée comme la somme de ces trois pertes, avec des poids appliqués à chaque terme de perte pour équilibrer leurs contributions à la perte globale. Les poids sont généralement fixés de telle sorte que la perte de localisation ait un poids plus élevé que les deux autres pertes, car une localisation précise est essentielle pour la détection d'objets. Cette fonction est définie par l'Équation 4.4 et détaillée dans l'Équation 4.5 :

$$Loss_{yolo} = ClassificationLoss + CoordinateLoss + ConfidenceLoss \quad (4.4)$$

$$Loss_{yolo} = \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c \in classes}^B (p_i(c) - \hat{p}_i(c))^2 + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(C_i - \hat{C}_i)^2] + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{noobj} [(C_i - \hat{C}_i)^2] \\ + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i}) + (\sqrt{h_i} - \sqrt{\hat{h}_i})] \quad (4.5)$$

Au cours de la phase d'apprentissage, la fonction de perte est minimisée à l'aide de la descente stochastique de gradient (SGD Stochastic Gradient Descent) ou d'un algorithme d'optimisation similaire pour mettre à jour les poids du réseau de neurones. En minimisant la perte, le réseau est entraîné à faire des prédictions plus précises des boîtes de délimitation des objets, des scores de confiance et des probabilités de classe dans les images. La sortie de l'algorithme de détection Yolo est donc un tenseur qui contient les boîtes de délimitation prédites, les scores de confiance et les probabilités de classe pour chaque zone de l'image.

## 4.4.2 Algorithme

Le pipeline de l'approche proposée : Yolo-dermoscopic-feature, comme toutes les approches basées sur l'apprentissage profond, se compose de trois phases : la phase de construction ou d'organisation du jeu de données, la phase d'apprentissage du modèle, et la phase de prédiction. La Figure 4.4 présente les différentes étapes de l'algorithme Yolo-dermoscopic-feature. Les modules verts de la Figure 4.4 représentent les parties sur lesquelles des modifications ont été apportées.

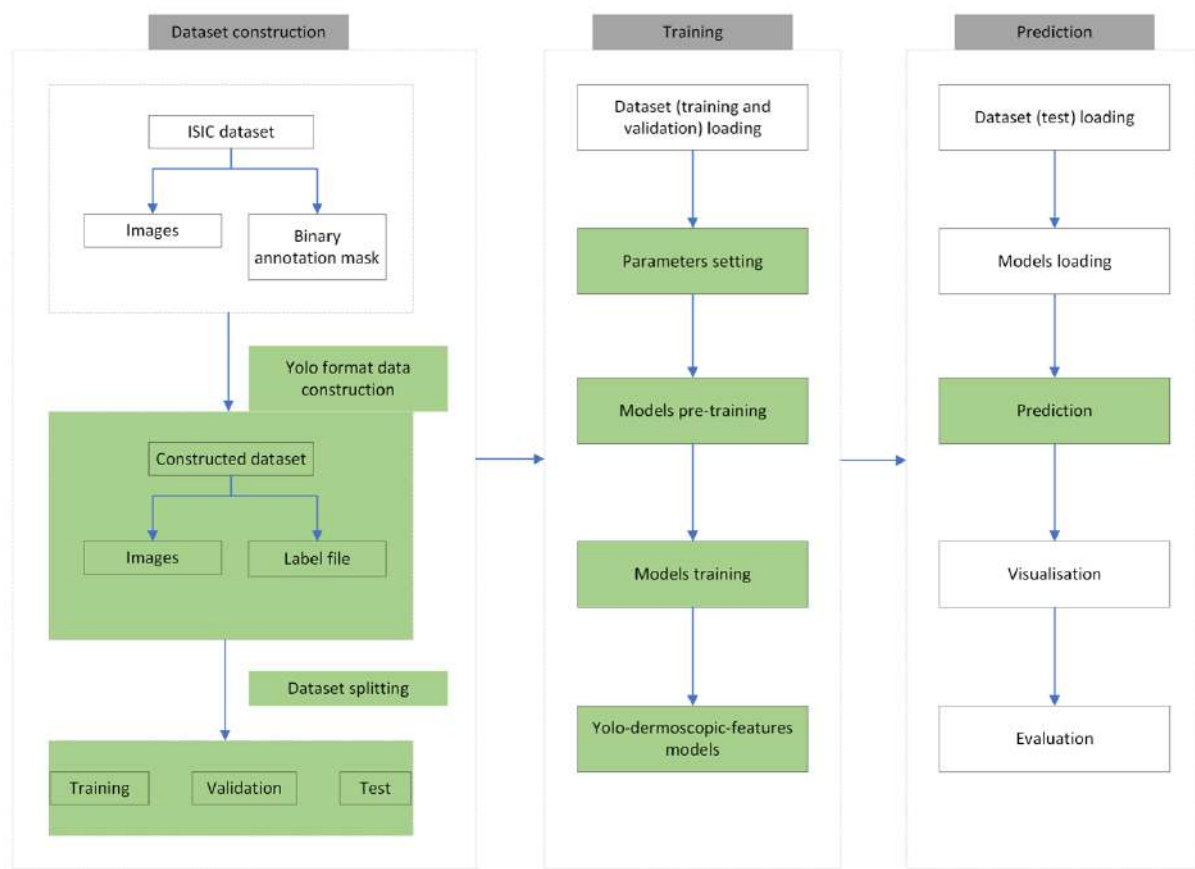


FIGURE 4.4 – Pipeline de l’approche Yolo-dermoscopic-feature. Les parties en vert dans la figure représentent les parties sur lesquelles des modifications ont été apportées.

La phase de construction de l’ensemble de données consiste à générer à partir masques binaires le fichier d’annotation des caractéristiques. Les masques binaires sont fournis avec la base d’image. Le format de fichier utilisé dans ce travail est le JavaScript Object Notation (JSON). JSON est un format d’échange de données léger basé sur un sous-ensemble de la norme ECMA-262 du langage de programmation JavaScript [159]. Le Tableau 5.1 résume les différentes étapes pour obtenir le fichier.

Le fichier généré renseigne entre autres sur, les types de caractéristiques dermoscopiques présents sur l’image (label), les coordonnées des points qui délimitent la caractéristique dermoscopique (points), et sur les dimensions de l’image (imageHeight et imageWidth). La Figure 4.5 présente la structure hiérarchique du fichier json formaté.

La deuxième partie de l’algorithme Yolo-dermoscopic-feature consiste à pré-entraîner et à entraîner les modèles. Le modèle utilisé est la dernière version de Yolo. Dans ce travail, au lieu d’utiliser directement un modèle pré-entraîné sur une base de données comme ImageNet, le modèle est d’abord pré-entraîné à segmenter la zone des lésions sur des images dermoscopiques. Ensuite, le modèle est utilisé pour la localisation des caractéristiques dermoscopiques. Afin de corriger le déséquilibre de caractéristiques observé dans la base

TABLE 4.2 – Algorithme de génération de fichiers d’annotations à partir des d’images binaires

---

Entrée : Masque binaire
Sortie : Fichiers d’annotation

---

- 1 : Initialisation :
- 2 :  $N_i$  : nombre d’images binaires
- 3 :  $N_c$  : nombre de contours
- 4 : Pour chaque image  $i=(1, N_i)$
- 5 :     Trouver le nombre de contours  $N_c$  dans le masque
- 6 :     Créer et initialiser le format du fichier json
- 7 :     Pour chaque image  $j= (1, N_c)$
- 8 :         Obtenir la coordonnée du contour
- 9 :         Écrire la coordonnée dans le fichier json
- 10 :     Fin Pour
- 11 : Sauvegarder le fichier json
- 12 : Fin Pour

---

de données, cinq modèles sont entraînés pour les cinq caractéristiques dermoscopiques (globules, réseau pigmentaire, réseau négatif, kystes de type milia et stries). En d’autres termes, un modèle par caractéristique.

### 4.4.3 Résultats expérimentaux et discussions

#### 4.4.3.1 Base de données

La base de données utilisée provient du défi ISIC 2018 et contient 2594 images de lésions cutanées et 12970 masques correspondant aux caractéristiques dermoscopiques annotées (5 pour chaque image). Les masques ont été obtenus à partir d’annotations d’experts, avec une validation croisée de plusieurs évaluateurs. Les cinq caractéristiques dermoscopiques annotées sont les globules, le réseau pigmentaire, le réseau négatif, les stries et les kystes de type milia. La résolution des images dermoscopiques de l’ensemble de données varie entre 640x480 et 1944x2592. Parmi les images, 603 images présentent des globules, 1523 images présentent un réseau pigmentaire, 191 images présentent un réseau négatif, 100 images présentent des stries et 682 images présentent des kystes de type milia.

La Figure 4.6 présente les structures de ces caractéristiques, des images dermoscopiques de la base de données et les masques binaires correspondants aux images. L’absence de pixels blancs dans un masque binaire traduit l’absence de la caractéristique dans l’image dermoscopique.

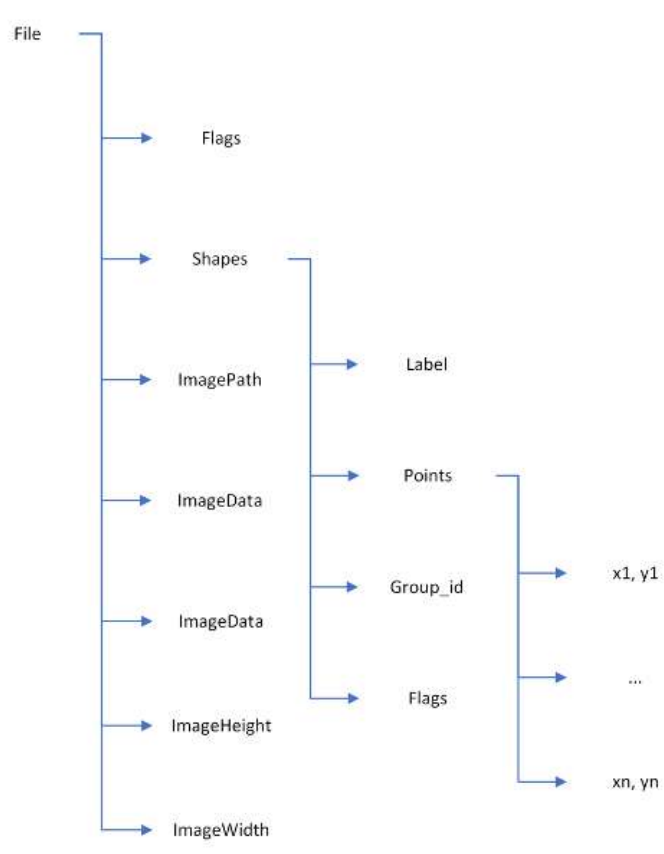


FIGURE 4.5 – Structure hiérarchique du fichier json généré.

#### 4.4.3.2 Configurations pour les expérimentations

Les expériences sont réalisées sur une machine de 40 Go de RAM et un GPU NVIDIA A100 Tensor Core. 70 % de l'ensemble de données est utilisés pour l'apprentissage, 20 % pour la validation et 10 % pour le test. Le nombre d'époques est fixé à 100.

Pour évaluer la qualité d'ajustement des paramètres pendant la phase d'apprentissage, nous avons observé l'évolution de trois fonctions de pertes de l'algorithme Yolo. Il s'agit de la perte de confiance (box-loss), de la perte de localisation (seg-loss) et de la perte de classification (cls-loss). La perte confiance indique dans quelle mesure le modèle peut localiser le centre d'un objet et dans quelle mesure la boîte englobante prédite couvre un objet. Les pertes de localisation et de classification indiquent la capacité du modèle à segmenter et à prédire la classe correcte d'un objet donné (voir Section 4.4).

Nous avons choisi d'évaluer la qualité de prédiction sur trois points : évaluation visuelle, évaluation qualitative et étude comparative.

- Évaluation visuelle : Elle consiste à comparer visuellement les masques de segmen-

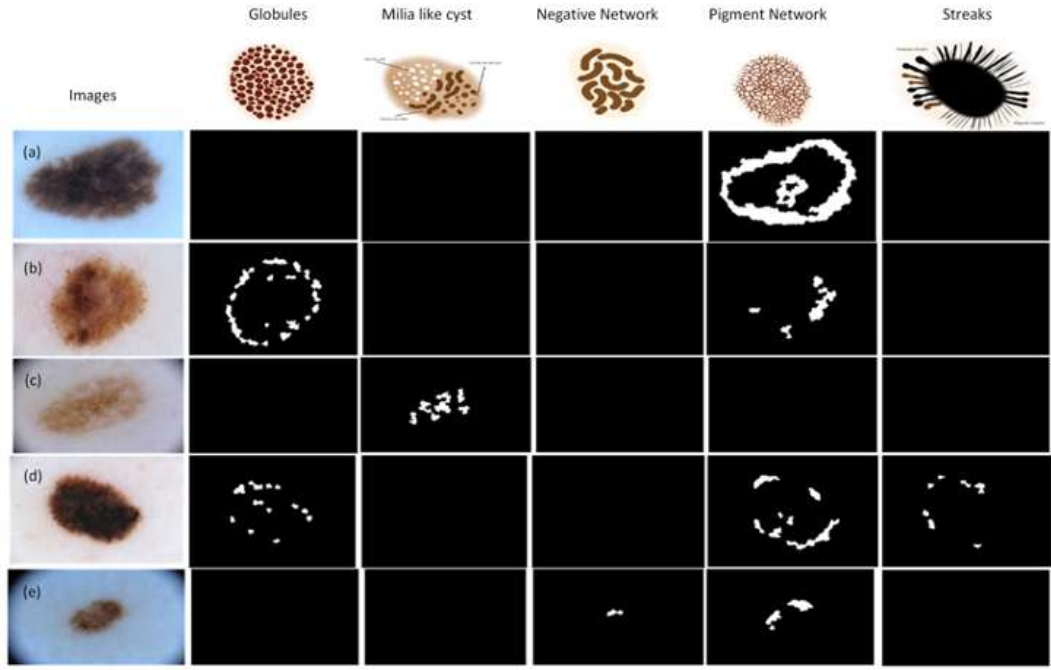


FIGURE 4.6 – Structures des caractéristiques et images de l’ensemble de données. (a) Image dermoscopique d’un naevus mélanocytaire, (b) Image dermoscopique d’un mélanome, (c) Image dermoscopique d’un naevus, (d) Image dermoscopique d’un mélanome, (e) Image dermoscopique d’un naevus.

tations prédits sur quelques exemples, aux masques de segmentations obtenus des annotations par des spécialistes.

- Évaluation qualitative : Elle permet d’apprécier objectivement les similitudes entre les prédictions et les données réelles en utilisant des métriques. A cette fin, cinq métriques ont été sélectionnés : le coefficient de similarité de Dice (DSC Dice Similarity Coefficient) [161], le coefficient de similarité de Jaccard (JSC Jaccard Similarity Coefficient) [162], la précision [18], le rappel [18] et la précision moyenne (mAP) [161]. Les métriques : DSC et JSC sont des mesures statistiques utilisées pour évaluer la similarité de deux échantillons. Toutes ces six métriques sont calculées à partir de la matrice de confusion (voir section 3.3.2) et définie par les équations 4.6 à 4.10.

$$DSC = \frac{2VP}{2VP + FP + FN} \quad (4.6)$$

$$JSC = \frac{VP}{VP + FP + FN} \quad (4.7)$$

$$Precision = \frac{VP}{VP + FP} \quad (4.8)$$

$$Recall = \frac{VP}{VP + FN} \quad (4.9)$$

$$mAP = \frac{Precision + Recall}{2} \quad (4.10)$$

Avec  $VP$  les vrais positifs,  $FP$  les faux positifs et  $FN$  les faux négatifs (voir Section 2.3).

- Étude comparative : Elle consiste dans un premier temps à comparer les algorithmes de deux méthodes de la littérature à l’algorithme de Yolo-dermoscopic-feature. Ensuite elle compare les résultats des métriques DSC, JSC et Recall de ces méthodes aux résultats obtenus avec Yolo-dermoscopic-feature.

#### 4.4.3.3 Résultats et discussions

Cette sous-section présente les résultats expérimentaux obtenus pour chacune des cinq caractéristiques dermoscopiques au cours de la phase d’apprentissage et de prédiction.

##### Évaluation pendant la phase d’apprentissage

L’étape de formation ou d’apprentissage consiste à mettre à jour les paramètres d’un réseau neuronal afin d’obtenir la combinaison de paramètres permettant la meilleure prédiction possible. L’apprentissage est supervisé lorsque le réseau neuronal est entraîné à l’aide d’une image et d’étiquettes [34]. Après chaque époque, les fonctions de perte et les métriques sont calculées pour observer l’état d’apprentissage du modèle.

La Figure 4.7 présente les diagrammes de perte des modèles pendant la formation et la validation. Les fonctions de perte convergent toutes vers zéro à partir de l’époque 80. En outre, il existe une similitude entre les paires de courbes d’entraînement et de validation. Cela signifie que le modèle généralise bien et s’adapte aux données observées de la base d’apprentissage ainsi qu’aux données non observées de la base de validation [78]. Pour éviter que la précision des modèles cesse de s’améliorer après un certain temps, voire se dégrade en raison de l’apprentissage par le bruit (arrêt précoce [78]), le nombre d’époques est limité à 100.

La Figure 4.8 montre les mesures de performance pendant l’apprentissage. Les modèles commencent à s’améliorer en termes de précision, de rappel et de précision moyenne après environ 50 époques. Une stabilisation est observée après environ 80 époques et les mesures convergent vers 1 qui est la valeur de performance la plus élevée. Ces courbes permettent d’observer que le processus d’apprentissage est progressif et que le modèle s’améliore réellement à chaque époque. La convergence de tous les modèles vers 1 est cohérente avec la convergence des courbes de perte (voir Figure 4.7) vers 0 et a motivé le choix de limiter



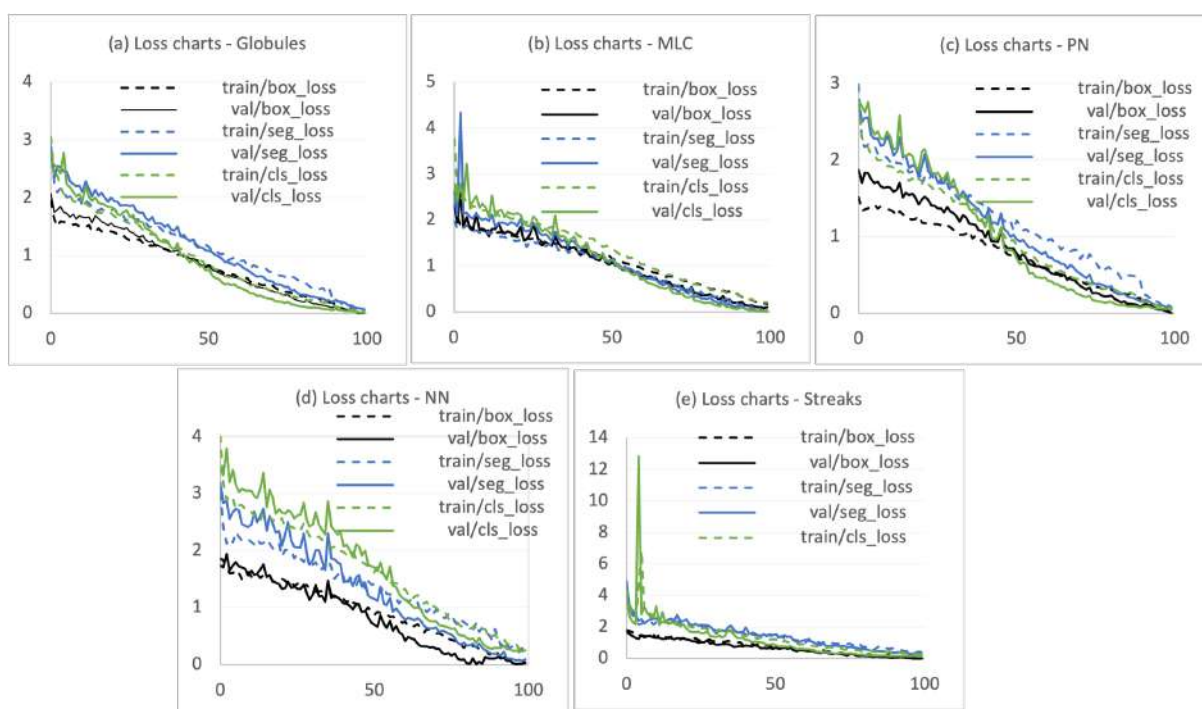


FIGURE 4.7 – Graphiques de perte des modèles pendant l’entraînement et la validation. (a) (b) (c) (d) et (e) représentent respectivement les courbes du modèle de détection de globules, du modèle de détection des kystes de type milia, du modèle de détection du réseau pigmentaire, du modèle de détection du réseau négatif et du modèle de détection des stries.

le nombre d’époques à 100 pour éviter l’overfitting [4, 78].

### Évaluation des prédictions

La phase de prédiction suit la phase d’apprentissage. Les prédictions ont été effectuées sur de nouvelles images qui n’ont pas été utilisées pour former les modèles. Les cinq modèles entraînés sont d’abord chargés et ensuite, sur la base de l’ensemble de test, chaque modèle effectue des prédictions. Les exemples des Figure 5.9 - Figure 5.13 montrent les résultats de prédiction des cinq caractéristiques. Pour chaque prédiction, le modèle renvoie les boîtes de délimitation détectées, le masque de segmentation et les scores de classe et de confiance prédits. De plus, l’algorithme Yolo-dermoscopic-features permet de générer un fichier d’annotation basé sur les prédictions. Ce label peut ensuite être utilisé pour annoter des images non annotées et ainsi augmenter le nombre d’images annotées.

### Évaluation visuelle

En observant les prédictions, nous remarquons quelques points forts de l’algorithme Yolo-dermoscopic-features :

- Les masques de segmentation pour les prédictions effectuées (Figure 4.9-Figure

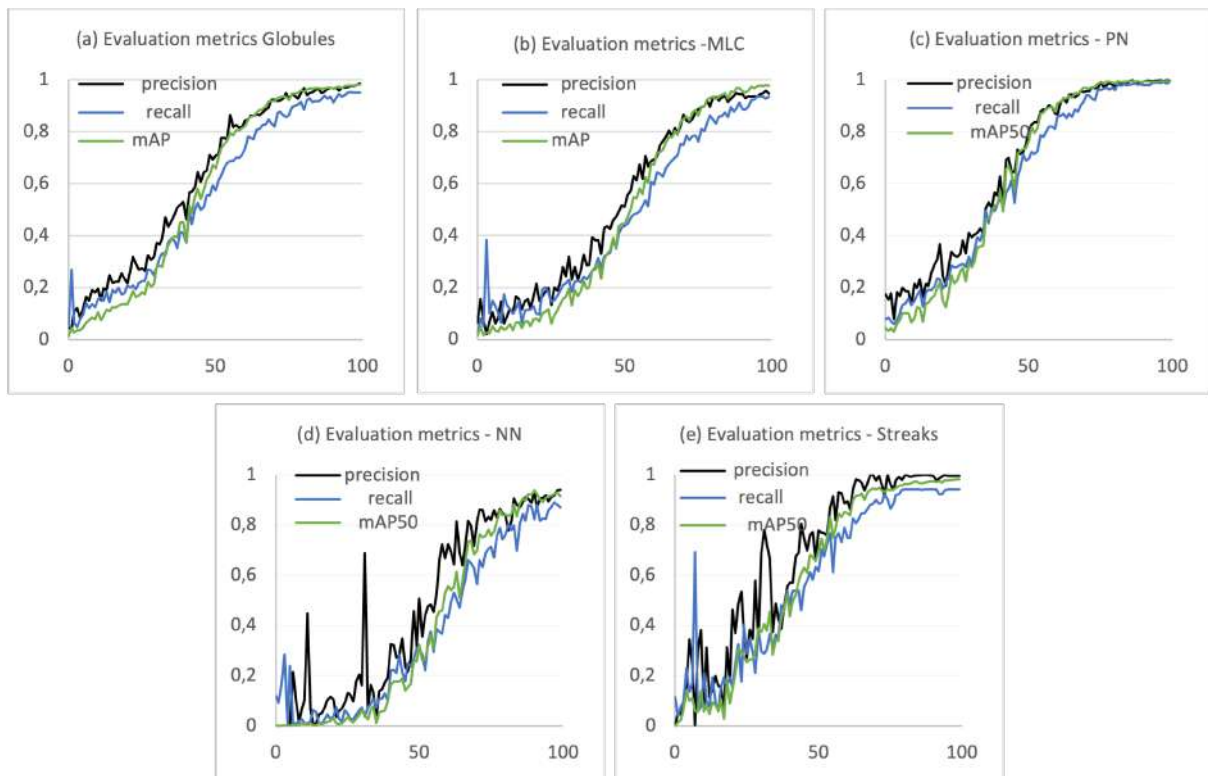


FIGURE 4.8 – Mesures de performance des modèles pendant la phase d’apprentissage. (a) (b) (c) (d) et (e) représentent respectivement les mesures de performance du modèle de détection de globules, du modèle de détection des kystes de type milia, du modèle de détection du réseau pigmentaire, du modèle de détection du réseau négatif et du modèle de détection des stries.

4.13(c)) présentent de grandes similitudes avec les masques de segmentations fournis par des spécialistes (Figure 4.9-Figure 4.13 (b)).

- Le modèle détecte bien la présence des structures dermoscopiques dans différentes régions de l’image.
- Dans certaines images, les caractéristiques sont situées sur les bords. Dans ces cas, l’algorithme détecte avec précision les zones affectées tout en délimitant les parties centrales qui ne présentent pas la caractéristique.
- La méthode permet de faire des prédictions sur des images de petits et grands diamètres.

### Évaluation qualitative (avec les métriques)

Le Tableau 4.3 montre les performances des modèles sur la base de test. Les statistiques présentées correspondent à la moyenne obtenue sur toutes les images de l’ensemble

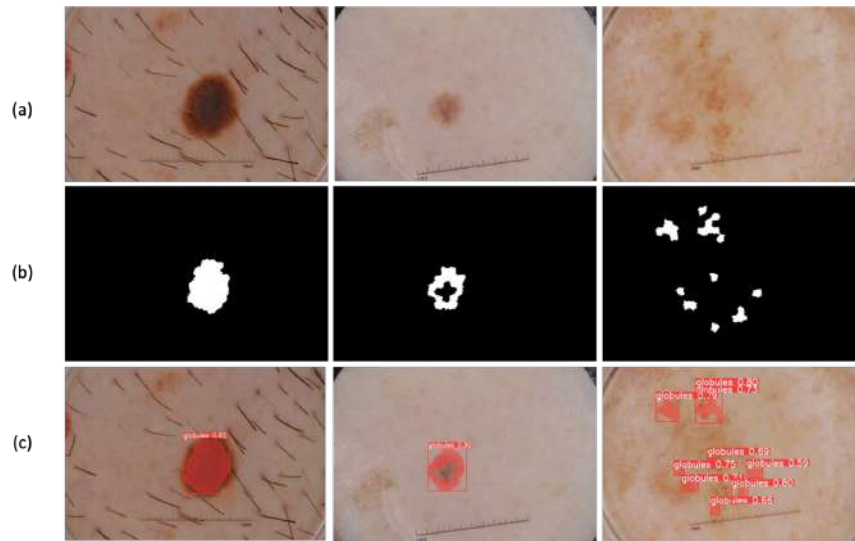


FIGURE 4.9 – (a) Images dermoscopiques, (b) ) Masques d’annotation des globules par les experts (c) Prédications des positions des globules.

de test. Les valeurs calculées pour chacune des métriques DSC, JSC, précision, rappel et précision moyenne convergent vers le score de performance le plus élevé 1. En se basant sur les résultats obtenus à l’aide des métriques, on peut objectivement conclure que les prédictions faites par les modèles ont de fortes similitudes avec les annotations faites par les experts. Cependant, même si les résultats sont intéressants pour toutes les caractéristiques, nous remarquons une meilleure performance pour les structures dermoscopiques bien représentées dans la base de données.

TABLE 4.3 – Performances de l’algorithme Yolo-dermoscopic-features.

Metrics	Globules	Milia like cyst	Pigment network	Negative network	Streaks	All features
DSC	0,984	0,979	1	0,947	0,969	0,9758
JSC	0,970	0,960	1	0,900	0,940	0,954
Precision	0,985	0,946	0,995	0,940	0,996	0,9724
Recall	0,950	0,935	0,992	0,871	0,942	0,938
Average precision	0,980	0,976	0,994	0,915	0,981	0,9692

### Étude comparative

L’étude comparative est la dernière étude réalisée et consiste à comparer les résultats de l’algorithme Yolo-dermoscopic-features avec deux autres méthodes de l’état de l’art. La première méthode sélectionnée est proposée par Kawahara et al [155] pour détecter la présence de quatre critères dermoscopiques (réseau pigmentaire, réseau négatif, kystes de type milia, stries) dans une image dermoscopique. Leur approche a remporté la première

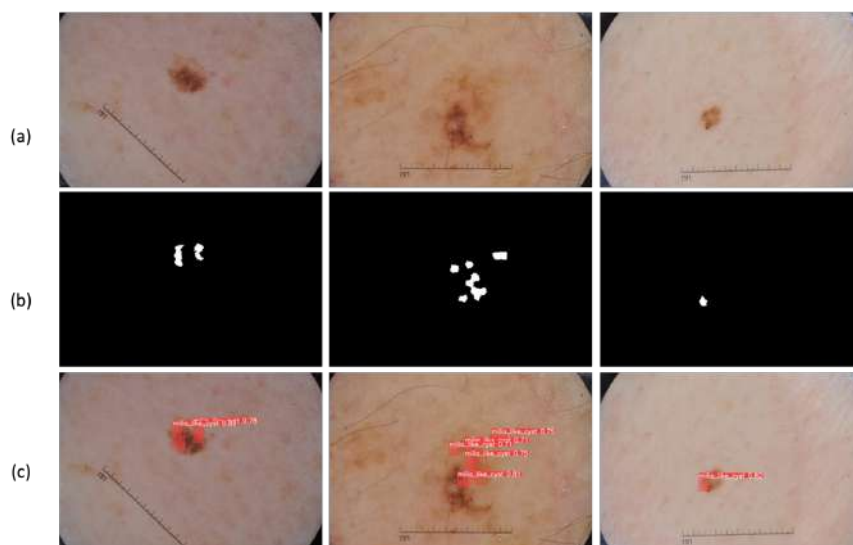


FIGURE 4.10 – (a) Images dermoscopiques, (b) ) Masques d’annotation des MLC par les experts (c) Prédications des positions des MLC.

place lors de l’édition 2017 de l’ISIC- ISBI Part 2 : Dermoscopic Feature Classification Task challenge. La deuxième méthode retenue pour la comparaison est proposée par Li et al [154] pour classer quatre types de caractéristiques dermoscopiques (réseau pigmentaire, réseau négatif, stries et kystes de type milia) et l’arrière-plan.

Le Tableau 4.4 présente une comparaison des différentes méthodes d’extraction des caractéristiques dermoscopiques. Nous remarquons que toutes les méthodes sont basées sur l’apprentissage profond, utilisent des bases de données ISIC, et permettent de détecter plusieurs caractéristiques dermoscopiques. La méthode Yolo-dermoscopic-features est celle qui extrait le plus de caractéristiques (5 caractéristiques contre 4 pour les autres). Nous avons estimé la complexité des trois méthodes comparées en fonction de l’approche d’extraction des caractéristiques. Kawahara et [155] proposent une approche globale pour l’extraction des caractéristiques, tout comme la méthode Yolo-dermoscopic-features proposée. Néanmoins, pour améliorer la précision du diagnostic, nous avons choisi d’entraîner plusieurs modèles, ce qui augmente le nombre de paramètres à entraîner et, par conséquent, le temps d’apprentissage et de prédiction. Li et al. dans [154] ont choisi une approche locale pour l’extraction des caractéristiques, ce qui signifie qu’ils ont d’abord subdivisé l’image en zones de superpixels et qu’ils ont ensuite appliqué leur méthode à chaque superpixel. Cette approche rend l’algorithme plus lourd et augmente le temps d’apprentissage. Compte tenu de cette analyse, nous pouvons conclure que la méthode la plus rapide et la moins complexe serait celle de Kawahara.

Le Tableau 4.5 présente les résultats des comparaisons des trois méthodes sur les métrique DSC, JSC et recall. Pour deux de ces métriques : JSC et le recall, la méthode proposée donne des résultats plus intéressants.

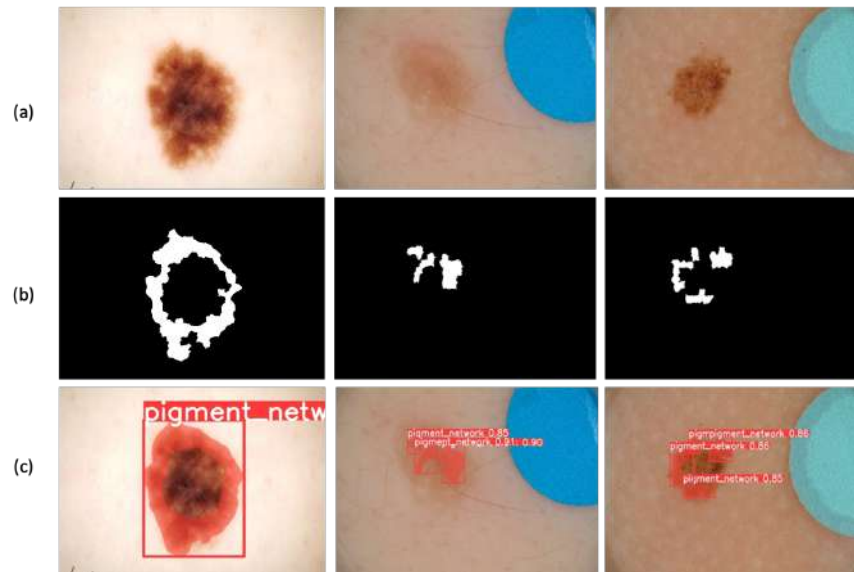


FIGURE 4.11 – (a) Images dermoscopiques, (b) ) Masques d’annotation des PN par les experts (c) Prédications des positions des PN.

## Conclusion

Dans ce chapitre, une méthode de détection de caractéristiques dermoscopiques basée sur un algorithme de détection d’objets en une seule prise Yolo est proposée. Cette solution permet d’avoir une compréhension complète des images dermoscopiques et de donner au médecin des informations plus pertinentes afin d’augmenter leur précision de diagnostic. L’objectif de ce travail est de proposer une méthode plus efficace que celles existantes dans la littérature. Pour atteindre ce but, nous avons proposé un algorithme appelé Yolo-dermoscopic-features défini en trois points : i) construction de la base de données, ii) pré-entraînement d’un modèle pour la détection des lésions cutanées, (iii) entraînement de cinq modèles pour la détection de cinq caractéristiques dermoscopiques. Les résultats montrent une amélioration de la prédiction avec les caractéristiques dermoscopiques de Yolo par rapport aux autres méthodes. Un autre avantage de l’algorithme proposé est qu’il peut également générer un fichier d’annotation basé sur les prédictions. Cette étiquette peut ensuite être utilisée pour annoter des images non annotées et ainsi augmenter le nombre d’images annotées.

Une des limites de ce travail est le nombre de modèles utilisé (autant de modèles que de caractéristiques dermoscopiques). Ce choix a été fait en raison de la très petite taille des éléments dermoscopiques, de la complexité de certaines structures et du fait qu’une lésion peut présenter la même caractéristique à plusieurs endroits. Bien que ce choix fournisse des prédictions intéressantes, une perspective serait d’utiliser un seul modèle pour localiser toutes les caractéristiques dermoscopiques. Les travaux futurs exploreront des techniques avancées pour atteindre cet objectif.

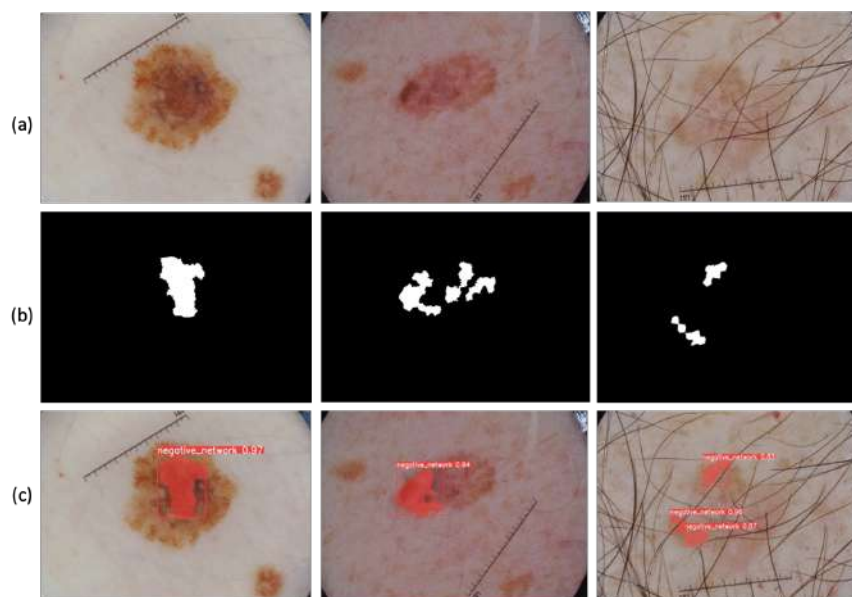


FIGURE 4.12 – (a) Images dermoscopiques, (b) ) Masques d’annotation des NN par les experts (c) Prédications des positions des NN.

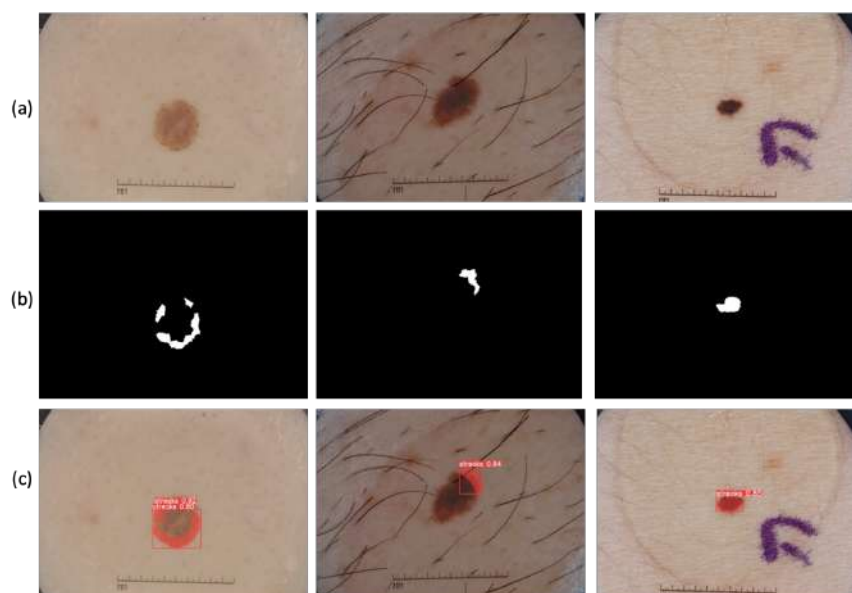


FIGURE 4.13 – (a) Images dermoscopiques, (b) Masques d’annotation des streaks par les experts (c) Prédications des positions des streaks.

TABLE 4.4 – Comparaison de différentes méthodes de détection de caractéristiques dermoscopiques.

<b>Méthode</b>	<b>Techniques</b>	<b>Caractéristique Dermoscopique</b>	<b>Base de données</b>	<b>Estimation de la complexité</b>
<b>Kawahara</b>	CNN et carte interpolée	pigment network, negative network, milia like cysts, streaks	ISIC	Faible
LFN	CNN	pigment network, negative network, streaks, milia-like cyst	ISIC	Élevé
YDL	Yolo	globules, pigment network, negative network, streaks, milia like cyst	ISIC	Élevé

TABLE 4.5 – Performances de différentes méthodes de détection de caractéristiques dermoscopiques.

<b>Méthode</b>	<b>JSC</b>	<b>DSC</b>	<b>Recall</b>
Kawahara J.	0,895	-	0,534
LFN	0,833	0,710	0,693
Yolo-dermoscopic-features	0,954	0,976	0,938





## CONCLUSION GÉNÉRALE ET PERSPECTIVES

### Conclusion générale

Le travail présenté dans ce manuscrit est un ensemble d'approches de solution en vue d'améliorer la précision de diagnostic de cancer de la peau. En tenant compte de l'incidence de plus en plus croissante de ces lésions, de la malignité des mélanomes, et de la difficulté de diagnostic, ce travail plus spécifiquement propose des outils d'aide à la détection précoce de lésions mélanocytaires et non mélanocytaires. Toutes les images traitées dans ce manuscrit proviennent d'une modalité d'imagerie médicale : la dermoscopie. Les outils proposés sont basés sur des méthodes d'apprentissage profond et concernent trois tâches : i) la classification binaire de lésions mélanocytaires et non mélanocytaires, ii) l'augmentation de la base de données, iii) la détection automatique de caractéristiques dermoscopiques.

Le Chapitre 1 présente la peau, sa structure, ses fonctions, les pathologies de la peau, les stratégies de diagnostic et quelques modalités d'imagerie de la peau.

Le Chapitre 2 présente les outils d'aides au diagnostic de lésions cutanées, les méthodes d'évaluation de ces outils, et la méthode proposée pour améliorer la précision de classification. La méthode proposée combine les filtres de Gabor au réseau CNN pour extraire efficacement les informations spatiales, de bord et de texture de lésions et les classées. La stratégie de diagnostic pendant une consultation consiste à distinguer les lésions malignes des lésions bénignes et dans un deuxième temps à identifier le type de lésion maligne et bénigne. La méthode proposée prend en compte la première étape et permet d'atteindre une précision de classification de 96,39%.

Le Chapitre 3 présente une revue de littérature sur les méthodes d'augmentation d'images, les métriques d'évaluation des méthodes d'augmentation, et l'approche proposée pour augmenter la base de données et faciliter une meilleure généralisation de l'outils d'aide au diagnostic. Cette méthode vient répondre au besoin d'avoir de larges bases de

données pour l'entraînement du modèle. La méthode proposée NSST PCNN est basée sur la décomposition d'images multi-échelles et la stratégie de fusion PCNN. Les peaux de teinte foncée sont généralement peu représentées dans les bases de données [133]. Nous sommes aussi arrivés à cette même conclusion en appliquant l'angle typologique individuel pour dénombrer les différents phénotypes de peau sur les images de la base de données utilisée. Avec la méthode d'augmentation proposée, nous arrivons à renforcer la base de données et à améliorer les précisions de prédictions et la généralisation des modèles, quel que soit le type de peau.

Le Chapitre 4 présente la revue de littérature sur les méthodes de détection automatique de caractéristiques dermoscopiques, et propose une méthode Yolo-dermoscopic-features basée sur un modèle Yolo et qui permet la détection automatique de cinq caractéristiques dermoscopiques (qui sont globules, pigment network, negative network, streaks, milia-like cyst). L'objectif est de segmenter et délimiter les régions où ces caractéristiques sont présentes. Les résultats permettent d'observer une forte similitude entre les prédictions faites par des experts et les prédictions de la méthode proposée. Les résultats de ces approches de solutions, bien qu'intéressants, peuvent être affinés et ne remplacent pas l'avis des spécialistes : ces méthodes sont des outils d'aide au diagnostic. La section Perspectives propose différentes perspectives dans la continuité des travaux de cette thèse.

## Perspectives

L'objectif de ce travail est d'améliorer la précision du diagnostic des lésions cutanées à l'aide d'outils d'intelligence artificielle. Pour atteindre cet objectif, nous nous sommes concentrés sur l'optimisation de l'architecture des réseaux de neurones à convolution, l'augmentation de la base de données et la détection automatique des caractéristiques.

Les résultats des différentes approches proposées dans nos travaux ouvrent de nouvelles perspectives de recherche, toujours dans le but de renforcer les systèmes d'aide au diagnostic des lésions cutanées. Ces perspectives concernent la base de données et l'architecture des modèles.

Dans le prolongement des travaux du Chapitre 4, la base d'images dermoscopique peut davantage être renforcée. Ce renforcement peut se faire par association d'autres images dermoscopiques à la base de données existante, par association d'images dermoscopiques à d'autres modalités d'imagerie des lésions de la peau, et par association de métadonnées cliniques aux bases d'images dermoscopiques. Les métadonnées cliniques qui proviennent d'informations collectées par les spécialistes (dermatologues et médecins) pendant l'examen clinique, elles renseignent entre autres sur la position de la lésion (main, dos, visage, ongles, pieds...), le sexe du patient et l'âge du patient. Certaines lésions sont fréquentes à des positions spécifiques par exemple. Le principe de l'extension de la base de données

est de collecter le maximum d'information possible afin de décrire au mieux la lésion et d'augmenter la quantité de caractéristiques mesurables par le système. En plus des images dermoscopiques utilisées dans nos travaux, il existe d'autres modalités comme la microscopie confocale par réflectance (MCR), la spectroscopie par réflectance, et l'imagerie multispectrale qui sont aussi utilisés pour examiner les lésions de la peau (voir section 2.3.3) . Les travaux de Cendre et al. [1] par exemple se sont intéressés à l'analyse d'images multimodales pour le diagnostic de lentigo malin. Pour renforcer la base d'image, nous prévoyons une collaboration avec d'autres chercheurs et spécialistes, disposant d'autres bases de données.

Pour augmenter la base d'images, il est également possible de synthétiser les images dermoscopiques avec des modèles génératifs plus précisément en faisant une traduction d'image à image avec les réseaux générateurs antagoniste. Le principe de cette approche est de générer une nouvelle image à partir d'un masque sémantique existant. Le travail de Bissoto et al. [110] qui porte sur la génération d'images dermoscopiques de haute résolution est l'une des seules publications sur le sujet. La difficulté de cette approche réside dans la création de masques sémantiques. Nous proposerons d'utiliser les fichiers d'annotation générés par l'algorithme Yolo-dermoscopic-features présenté au Chapitre 5 pour créer les masques sémantiques.

Une fois la base de données renforcée nous envisageons utiliser des architectures de réseaux de convolutions profonds (deep CNN) qui sont adaptées à de très grands ensembles de données. L'exploitation de différentes idées novatrices dans la conception des architectures des CNN est un domaine de recherche prometteur. Ces innovations impliquent l'augmentation de la taille et nombre du réseau, et l'optimisation des hyperparamètres. L'apprentissage d'ensemble qui consiste à combiner des architectures multiples et diverses est l'un des domaines de recherche prospectifs des CNN [162]. L'objectif de cette approche est d'aider le modèle à améliorer la généralisation et la robustesse sur diverses catégories d'images. Un réseau profond possède un grand nombre d'hyperparamètres tels que la fonction d'activation, la taille du noyau, le nombre de neurones par couche et la disposition des couches. La sélection des hyperparamètres et le temps d'évaluation d'un réseau profond font de l'ajustement des paramètres une tâche complexe. La définition des hyperparamètres est une tâche fastidieuse et intuitive, qui ne peut être définie par une formulation explicite. À cet égard, les algorithmes génétiques [163] peuvent également être utilisés pour optimiser automatiquement les hyperparamètres en orientant la recherche sur la base des résultats précédents ou de manière aléatoire.



## LISTES DES PUBLICATIONS

### **Journal avec comité de lecture :**

- Chabi Adjobo, E. ; Sanda Mahama, A.T. ; Gouton, P. ; Tossa, J. (2023) Automatic Localization of Five Relevant Dermoscopic Structures Based on YOLOv8 for Diagnosis Improvement. *J. Imaging*, 9, 148.
- Chabi Adjobo, E., Sanda Mahama, A. T., Gouton, P., and Tossa, J. (2022). Towards accurate skin lesion classification across all skin categories using a penn fusion-based data augmentation approach. *Computers*, 11(3), 44.

### **Conférences :**

- Adjobo, E. C., Mahama, A. T. S., Gouton, P., et Tossa, J. (2019, November). Proposition of convolutional neural network based system for skin cancer detection. In *2019 15th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)* (pp. 35-39). IEEE.

### **Participations aux évènements scientifiques :**

- Adjobo, E. C., Mahama, A. T. S., Gouton, P., et Tossa, J. (2020, Janvier, 06). Amélioration du diagnostic de lésions cutanées avec les réseaux conventionnels. Séminaire d'informatique Institut de Mathématiques et Sciences Appliquées, Dangbo.



- [1] Romain Cendre. *Classification par méthodes d'apprentissage supervisé et faiblement supervisé d'images multimodales pour l'aide au diagnostic du lentigo malin en dermatologie*. phdthesis, Université Bourgogne Franche-Comté, January 2021.
- [2] Thomas B. Fitzpatrick. The Validity and Practicality of Sun-Reactive Skin Types I Through VI. *Archives of Dermatology*, 124(6) :869–871, June 1988.
- [3] Philippe Bahadoran , Sarah Norrenberg , and Jean-Luc Lévy (first). Microscopie confocale in vivo : quel avenir en dermatologie ?, October 2016.
- [4] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018 : A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv :1902.03368*, 2019.
- [5] Marc Combalia, Noel CF Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Cristina Carrera, Alicia Barreiro, Allan C Halpern, Susana Puig, et al. Bcn20000 : Dermoscopic lesions in the wild. *arXiv preprint arXiv :1908.02288*, 2019.
- [6] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1) :180161, August 2018. Number : 1 Publisher : Nature Publishing Group.
- [7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11) :139–144, 2020.
- [8] Francesco Calimeri, Aldo Marzullo, Claudio Stamile, and Giorgio Terracina. Biomedical data augmentation using generative adversarial neural networks. In *Artificial Neural Networks and Machine Learning–ICANN 2017 : 26th International*

*Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26*, pages 626–634. Springer, 2017.

- [9] Rafael C. Gonzalez, Richard E. Woods, and Barry R. Masters. Digital Image Processing, Third Edition. *Journal of Biomedical Optics*, 14(2) :029901, 2009.
- [10] Qaisar Abbas, Irene Fondón Garcia, M. Emre Celebi, Waqar Ahmad, and Qaisar Mushtaq. A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images. *Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI)*, 19(1) :e490–497, February 2013.
- [11] Azadeh Noori Hoshyar, Adel Al-Jumaily, and Afsaneh Noori Hoshyar. The Beneficial Techniques in Preprocessing Step of Skin Cancer Detection System Comparing. *Procedia Computer Science*, 42 :25–31, January 2014.
- [12] Zoe Apalla, Aimilios Lallas, Elena Sotiriou, Elizabeth Lazaridou, and Demetrios Ioannides. Epidemiological trends in skin cancer. *Dermatology Practical & Conceptual*, 7(2) :1–6, April 2017.
- [13] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Noguees, Jianhua Yao, Daniel Mollura, and Ronald M. Summers. Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE transactions on medical imaging*, 35(5) :1285–1298, May 2016.
- [14] Rafael C. Gonzalez and Richard E. Wood. *Digital Image Processing*. Number 0-201-18075-8 in 9353062985. Pearson India, 4th edition edition, 2018.
- [15] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kallou, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection : A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018.
- [16] Titus Josef Brinker, Achim Hekler, Jochen Sven Utikal, Niels Grabe, Dirk Schandendorf, Joachim Klode, Carola Berking, Theresa Steeb, Alexander H. Enk, and Christof von Kalle. Skin Cancer Classification Using Convolutional Neural Networks : Systematic Review. *Journal of Medical Internet Research*, 20(10) :e11936, October 2018.
- [17] Matthew Groh, Caleb Harris, Luis Soenksen, Felix Lau, Rachel Han, Aerin Kim, Arash Koochek, and Omar Badri. Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1820–1828, 2021.



- [18] Esther Chabi Adjobo, Amadou Tidjani Sanda Mahama, Pierre Gouton, and Joël Tossa. Towards accurate skin lesion classification across all skin categories using a pcnn fusion-based data augmentation approach. *Computers*, 11(3) :44, 2022.
- [19] Characteristics of publicly available skin cancer image datasets : a systematic review - The Lancet Digital Health.
- [20] Pr Benoit de WAZIERES. Collège National des Enseignants de Médecine Interne. *Algorithms and Interfaces*, page 23, 2012.
- [21] Deevya L Narayanan, Rao N Saladi, and Joshua L Fox. Ultraviolet radiation and skin cancer. *International journal of dermatology*, 49(9) :978–986, 2010.
- [22] Howard W. Rogers, Martin A. Weinstock, Steven R. Feldman, and Brett M. Coldiron. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the US Population, 2012. *JAMA Dermatology*, 151(10) :1081–1086, October 2015.
- [23] SPF. Encadré. L'épidémiologie du mélanome cutané en France et en Europe. Numéro thématique. Le bronzage artificiel : une menace bien réelle, mais évitable, pour la santé publique, 2012.
- [24] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5) :34–41, 2001.
- [25] L. Ban, S. Labbouz, D. Grindlay, J. M. Batchelor, and S. Ratib. Risk of skin cancer in people with vitiligo : a systematic review and meta-analysis. *British Journal of Dermatology*, 179(4) :971–972, 2018. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjd.16703>.
- [26] Koffi Kouamé Pacôme Gbandama, Almamy Diabaté, Kouamé Alexandre Kouassi, Yao Isidore Kouassi, Ange-Sylvain Allou, and Mamadou Kaloga. Squamous Cell Carcinoma Associated with Cosmetic Use of Bleaching Agents : About a Case in Ivory Coast. *Case Reports in Dermatology*, 11(3) :322–326, 2019. Publisher : Karger Publishers.
- [27] Romuald Jolivot. *Development of an imaging system dedicated to the acquisition analysis and multispectral characterisation of skin lesion*. PhD thesis, Université de Bourgogne, 2011.
- [28] I Stanganelli and MA Pizzichetta. Dermoscopy : overview, technical procedures and equipment, color. *Medscape*, 2018.
- [29] J.F. Nicolas, H. Michalaki, E. Peyron, P. Machado, E. Cozzani, and D. Schmitt. Pathologie acquise de la jonction dermo-épidermique. *médecine/sciences*, 9(4) :376, 1993.
- [30] B. Dréno. Anatomie et physiologie de la peau et de ses annexes. *Annales de Dermatologie et de Vénérologie*, 136 :S247–S251, October 2009.
- [31] Masson. *Comprendre la peau, annales de dermatologie et de vénéréologie*. Masson, Elsevier, 2005.

- [32] Oma N. Agbai, Kesha Buster, Miguel Sanchez, Claudia Hernandez, Roopal V. Kundu, Melvin Chiu, Wendy E. Roberts, Zoe D. Draelos, Reva Bhushan, Susan C. Taylor, and Henry W. Lim. Skin cancer and photoprotection in people of color : A review and recommendations for physicians and the public. *Journal of the American Academy of Dermatology*, 70(4) :748–762, April 2014. Publisher : Elsevier.
- [33] Sarah Girardeau-Hubert, Céline Deneuve, Hervé Pageon, Kahina Abed, Charlotte Tacheau, Nükhet Cavusoglu, Mark Donovan, Dominique Bernard, and Daniel Asse-lineau. Reconstructed Skin Models Revealed Unexpected Differences in Epidermal African and Caucasian Skin. *Scientific Reports*, 9(1) :7456, May 2019. Number : 1 Publisher : Nature Publishing Group.
- [34] Esther Chabi Adjobo, Amadou Tidjani Sanda Mahama, Pierre Gouton, and Joël Tossa. Proposition of Convolutional Neural Network Based System for Skin Cancer Detection. In *2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 35–39, November 2019.
- [35] M. Alexandra Kredlow, Suzanne L. Pineles, Sabra S. Inslicht, Marie-France Marin, Mohammed R. Milad, Michael W. Otto, and Scott P. Orr. Assessment of Skin Conductance in African American and Non-African American Participants in Studies of Conditioned Fear. *Psychophysiology*, 54(11) :1741–1754, November 2017.
- [36] A. Lallas, C. Reggiani, G. Argenziano, A. Kyrgidis, R. Bakos, N. C. M. S. Masiero, A. B. Scheibe, H. Cabo, F. Ozdemir, A. M. Sortino-Rachou, B. Gerceker Turk, E. Moscarella, C. Longo, and I. Zalaudek. Dermoscopic nevus patterns in skin of colour : a prospective, cross-sectional, morphological study in individuals with skin type V and VI. *Journal of the European Academy of Dermatology and Venereology : JEADV*, 28(11) :1469–1474, November 2014.
- [37] Harold A. Lancer. Lancer ethnicity scale (LES). *Lasers in Surgery and Medicine*, 22(1) :9–9, 1998. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-9101%281998%2922%3A1%3C9%3A%3AAID-LSM4%3E3.0.CO%3B2-T>.
- [38] A. J. Herzberg and S. M. Dinehart. Chronologic aging in black skin. *The American Journal of Dermatopathology*, 11(4) :319–328, August 1989.
- [39] Franz Nachbar, Wilhelm Stolz, Tanja Merkle, Armand B. Coggnetta, Thomas Vogt, Michael Landthaler, Peter Bilek, Otto Braun-Falco, and Gerd Plewig. The ABCD rule of dermatoscopy : High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, 30(4) :551–559, April 1994. Publisher : Elsevier.
- [40] Richard Ashton and Barbara Leppard. *Differential Diagnosis in Dermatology*. CRC Press, London, 5 edition, April 2021.
- [41] Silvana Díaz, Thomas Krohmer, Álvaro Moreira, Sebastián E Godoy, and Miguel

- Figuerola. An instrument for accurate and non-invasive screening of skin cancer based on multimodal imaging. *IEEE Access*, 7 :176646–176657, 2019.
- [42] 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, December 2004.
- [43] Yan Pan, Daniel S. Gareau, Alon Scope, Milind Rajadhyaksha, Nizar A. Mullani, and Ashfaq A. Marghoob. Polarized and Nonpolarized Dermoscopy : The Explanation for the Observed Differences. *Archives of Dermatology*, 144(6) :828–829, June 2008.
- [44] Griffin Lentsch, Erica G. Baugh, Bonnie Lee, Michelle Aszterbaum, Christopher B. Zachary, Kristen M. Kelly, and Mihaela Balu. Research Techniques Made Simple : Emerging Imaging Technologies for Noninvasive Optical Biopsy of Human Skin. *Journal of Investigative Dermatology*, 142(5) :1243–1252.e1, May 2022.
- [45] Anne Koenig, Blandine Roig, Jimmy Le Digabel, Gwendal Josse, and Jean-Marc Dinten. Accessing deep optical properties of skin using diffuse reflectance spectroscopy. In *Clinical and Biomedical Spectroscopy and Imaging IV*, volume 9537, pages 42–49. SPIE, July 2015.
- [46] David Huang, Eric A. Swanson, Charles P. Lin, Joel S. Schuman, William G. Stinson, Warren Chang, Michael R. Hee, Thomas Flotte, Kenton Gregory, Carmen A. Puliafito, and James G. Fujimoto. Optical Coherence Tomography. *Science (New York, N. Y.)*, 254(5035) :1178–1181, November 1991.
- [47] Jeffrey Glaister, Robert Amelard, Alexander Wong, and David A. Clausi. MSIM : Multistage Illumination Modeling of Dermatological Photographs for Illumination-Corrected Skin Lesion Analysis. *IEEE Transactions on Biomedical Engineering*, 60(7) :1873–1883, July 2013. Conference Name : IEEE Transactions on Biomedical Engineering.
- [48] K. Melbin and Y. Jacob Vetha Raj. Integration of modified ABCD features and support vector machine for skin lesion types classification. *Multimedia Tools and Applications*, 80(6) :8909–8929, March 2021.
- [49] Jeremy Kawahara, Sara Daneshvar, Giuseppe Argenziano, and Ghassan Hamarneh. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE Journal of Biomedical and Health Informatics*, 23(2) :538–546, March 2019. Conference Name : IEEE Journal of Biomedical and Health Informatics.
- [50] Scott W. Menzies, Kerry A. Crotty, Christian Ingwar, and William H. McCarthy. *An Atlas of Surface Microscopy of Pigmented Skin Lesions : Dermoscopy, Second Edition*. McGraw-Hill Book Company Australia, Sydney, austrln edition edition, October 2002.

- [51] Abundance of the benign melanocytic universe : Dermoscopic–histopathological correlation in nevi - Woltsche - 2017 - The Journal of Dermatology - Wiley Online Library.
- [52] Yellow and orange in cutaneous lesions : clinical and dermoscopic data - Bañuls - 2015 - Journal of the European Academy of Dermatology and Venereology - Wiley Online Library.
- [53] 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. Conference Name : IEEE Transactions on Systems, Man, and Cybernetics.
- [54] Harald Kittler, Ashfaq A. Marghoob, Giuseppe Argenziano, Cristina Carrera, Clara Curiel-Lewandrowski, Rainer Hofmann-Wellenhof, Josep Malvehy, Scott Menzies, Susana Puig, Harold Rabinovitz, Wilhelm Stolz, Toshiaki Saida, H. Peter Soyer, Eliot Siegel, William V. Stoecker, Alon Scope, Masaru Tanaka, Luc Thomas, Philipp Tschandl, Iris Zalaudek, and Allan Halpern. Standardization of terminology in dermoscopy/dermatoscopy : Results of the third consensus conference of the International Society of Dermoscopy. *Journal of the American Academy of Dermatology*, 74(6) :1093–1106, June 2016.
- [55] Oriol Yélamos, Ralph P. Braun, Konstantinos Liopyris, Zachary J. Wolner, Katrin Kerl, Pedram Gerami, and Ashfaq A. Marghoob. Dermoscopy and dermatopathology correlates of cutaneous neoplasms. *Journal of the American Academy of Dermatology*, 80(2) :341–363, February 2019.
- [56] Lucia Ballerini, Robert B Fisher, Ben Aldridge, and Jonathan Rees. A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. *Color medical image analysis*, pages 63–86, 2013.
- [57] Noel Codella, Junjie Cai, Mani Abedini, Rahil Garnavi, Alan Halpern, and John R. Smith. Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images. In Luping Zhou, Li Wang, Qian Wang, and Yinghuan Shi, editors, *Machine Learning in Medical Imaging*, Lecture Notes in Computer Science, pages 118–126, Cham, 2015. Springer International Publishing.
- [58] 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 : The Official Journal of the Computerized Medical Imaging Society*, 31(6) :362–373, September 2007.
- [59] Lei Bi, Jinman Kim, Euijoon Ahn, and Dagan Feng. Automatic Skin Lesion Analysis using Large-scale Dermoscopy Images and Deep Residual Networks, March 2017. arXiv :1703.04197 [cs].

- [60] Elif Kanca and Selen Ayas. Learning Hand-Crafted Features for K-NN based Skin Disease Classification. In *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–4, June 2022.
- [61] Mayank Arya Chandra and S. S. Bedi. Survey on SVM and their application in imageclassification. *International Journal of Information Technology*, 13(5) :1–11, October 2021.
- [62] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [63] Siddharth Sharma, Simone Sharma, and Anidhya Athaiya. ACTIVATION FUNCTIONS IN NEURAL NETWORKS. *International Journal of Engineering Applied Sciences and Technology*, 04(12) :310–316, May 2020.
- [64] Sepp Hochreiter. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02) :107–116, April 1998. Publisher : World Scientific Publishing Co.
- [65] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep Sparse Rectifier Neural Networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, June 2011. ISSN : 1938-7228.
- [66] Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, R. Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten Digit Recognition with a Back-Propagation Network. In *Advances in Neural Information Processing Systems*, volume 2. Morgan-Kaufmann, 1989.
- [67] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10) :1995, 1995.
- [68] 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, February 2017. Bandiera\_abtest : a Cg\_type : Nature Research Journals Number : 7639 Primary\_atype : Research Publisher : Nature Publishing Group Subject\_term : Diagnosis;Machine learning;Skin cancer Subject\_term\_id : diagnosis;machine-learning;skin-cancer.
- [69] Victor Pomponiu, Hossein Nejati, and N-M Cheung. Deepmole : Deep neural networks for skin mole lesion classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2623–2627. IEEE, 2016.
- [70] Przemyslaw Biecek and Tomasz Burzykowski. *Explanatory model analysis : explore, explain, and examine predictive models*. CRC Press, 2021.
- [71] Syed Shakib Sarwar, Priyadarshini Panda, and Kaushik Roy. Gabor filter assisted energy efficient fast learning Convolutional Neural Networks. In *2017 IEEE/ACM*

- International Symposium on Low Power Electronics and Design (ISLPED)*, pages 1–6, July 2017.
- [72] Gökhan Özbulak and Hazim Kemal Ekenel. Initialization of convolutional neural networks by Gabor filters. In *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, May 2018.
- [73] Hu Yao, Li Chuyi, Hu Dan, and Yu Weiyu. Gabor Feature Based Convolutional Neural Network for Object Recognition in Natural Scene. In *2016 3rd International Conference on Information Science and Control Engineering (ICISCE)*, pages 386–390, July 2016.
- [74] Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han, and Jianzhuang Liu. Gabor Convolutional Networks. *IEEE Transactions on Image Processing*, 27(9) :4357–4366, September 2018. Conference Name : IEEE Transactions on Image Processing.
- [75] Martial Mermillod, Patrik Vuilleumier, Carole Peyrin, David Alleysson, and Christian Marendaz. The importance of low spatial frequency information for recognising fearful facial expressions. *Connection Science*, 21(1) :75–83, March 2009. Publisher : Taylor & Francis \_eprint : <https://doi.org/10.1080/09540090802213974>.
- [76] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, March 2010. ISSN : 1938-7228.
- [77] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers : Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [78] Xue Ying. An Overview of Overfitting and its Solutions. *Journal of Physics : Conference Series*, 1168 :022022, February 2019. Publisher : IOP Publishing.
- [79] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, November 1998. Conference Name : Proceedings of the IEEE.
- [80] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.
- [81] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015. ISSN : 1063-6919.
- [82] 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)*, pages 770–778, June 2016. ISSN : 1063-6919.

- [83] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [84] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pages 117–122, May 2018.
- [85] Yan Wang, Biting Yu, Lei Wang, Chen Zu, David S. Lalush, Weili Lin, Xi Wu, Jiliu Zhou, Dinggang Shen, and Luping Zhou. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage*, 174 :550–562, July 2018.
- [86] Ikram Brahim, Dominique Fourer, Vincent Vigneron, and Hichem Maaref. Deep Learning Methods for MRI Brain Tumor Segmentation : a comparative study. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, November 2019. ISSN : 2154-512X.
- [87] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5) :545–563, 2021. \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1754-9485.13261>.
- [88] Clara M. Tam, Dong Zhang, Bo Chen, Terry Peters, and Shuo Li. Holistic multi-task regression network for multiapplication shape regression segmentation. *Medical Image Analysis*, 65 :101783, October 2020.
- [89] Matej Kompanek, Martin Tamajka, and Wanda Benesova. Volumetric Data Augmentation as an Effective Tool in MRI Classification Using 3D Convolutional Neural Network. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 115–119, June 2019. ISSN : 2157-8702.
- [90] Charles Poynton. *Digital video and HD : Algorithms and Interfaces*. Elsevier, 2012.
- [91] Clément Chadebec, Elina Thibeau-Sutre, Ninon Burgos, and Stéphanie Allassonnière. Data augmentation in high dimensional low sample size setting using a geometry-based variational autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [92] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.
- [93] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017.

- [94] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11) :139–144, 2020.
- [95] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv :1511.06434*, 2015.
- [96] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [97] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv :1710.10196*, 2017.
- [98] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.
- [99] Ohad Shitrit and Tammy Riklin Raviv. Accelerated magnetic resonance imaging by adversarial neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 30–38. Springer, 2017.
- [100] Dwarikanath Mahapatra and Behzad Bozorgtabar. Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution. *arXiv preprint arXiv :1710.04783*, 2017.
- [101] Ali Madani, Mehdi Moradi, Alexandros Karargyris, and Tanveer Syeda-Mahmood. Chest x-ray generation and data augmentation for cardiovascular abnormality classification. In *Medical imaging 2018 : Image processing*, volume 10574, pages 415–420. SPIE, 2018.
- [102] Chen-Yi Lu, Dan Jeric Arcega Rustia, and Ta-Te Lin. Generative adversarial network based image augmentation for insect pest classification enhancement. *IFAC-PapersOnLine*, 52(30) :1–5, 2019.
- [103] Maria JM Chuquicusma, Sarfaraz Hussein, Jeremy Burt, and Ulas Bagci. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 240–244. IEEE, 2018.
- [104] Camilo Bermudez, Andrew J Plassard, Larry T Davis, Allen T Newton, Susan M Resnick, and Bennett A Landman. Learning implicit brain mri manifolds with deep learning. In *Medical Imaging 2018 : Image Processing*, volume 10574, pages 408–414. SPIE, 2018.



- [105] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321 :321–331, 2018.
- [106] Sagar Kora Venu and Sridhar Ravula. Evaluation of deep convolutional generative adversarial networks for data augmentation of chest x-ray images. *Future Internet*, 13(1) :8, 2020.
- [107] Zhiwei Qin, Zhao Liu, Ping Zhu, and Yongbo Xue. A gan-based image synthesis method for skin lesion classification. *Computer Methods and Programs in Biomedicine*, 195 :105568, 2020.
- [108] Saisai Ding, Jian Zheng, Zhaobang Liu, Yanyan Zheng, Yanmei Chen, Xiaomin Xu, Jia Lu, and Jing Xie. High-resolution dermoscopy image synthesis with conditional generative adversarial networks. *Biomedical Signal Processing and Control*, 64, February 2021.
- [109] Alceu Bissoto, Fábio Perez, Eduardo Valle, and Sandra Avila. Skin lesion synthesis with generative adversarial networks. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis : First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*, pages 294–302. Springer, 2018.
- [110] Florence Laporterie. *Représentations hiérarchiques d’images avec des pyramides morphologiques : application à l’analyse et à la fusion spatio-temporelle de données en observation de la Terre*. PhD thesis, École nationale supérieure de l’aéronautique et de l’espace (Toulouse; 1972 . . . , 2002.
- [111] Walid Aribi, Ali Kalfallah, Noomène Elkadri, Leila Farhat, Wicem Siala, Jamel Daoud, and MS Bouhelel. Évaluation de techniques pyramidales de fusion multimodale (irm/tep) d’images cérébrales. In *5th International Conference : Sciences of Electronic, Technologies of Information and Telecommunications*, volume 19, pages 547–551, 2009.
- [112] Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. In *Readings in computer vision*, pages 671–679. Elsevier, 1987.
- [113] Shaik Shehanaz, Ebenezer Daniel, Sitaramanjaneya Reddy Guntur, and Sivaji Satrasupalli. Optimum weighted multimodal medical image fusion using particle swarm optimization. *Optik*, 231 :166413, 2021.
- [114] Nianyi Wang and Weilan Wang. An image fusion method based on wavelet and dual-channel pulse coupled neural network. In *2015 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pages 270–274. IEEE, 2015.

- [115] Huibin Yan and Zhongmin Li. Infrared and visual image fusion based on multi-scale feature decomposition. *Optik*, 203 :163900, 2020.
- [116] Arthur L Da Cunha, Jianping Zhou, and Minh N Do. The nonsubsampled contourlet transform : theory, design, and applications. *IEEE transactions on image processing*, 15(10) :3089–3101, 2006.
- [117] Liangliang Li, Hongbing Ma, Zhenhong Jia, and Yujuan Si. A novel multiscale transform decomposition based multi-focus image fusion framework. *Multimedia Tools and Applications*, 80 :12389–12409, 2021.
- [118] Manli Wang, Zijian Tian, Weifeng Gui, Xiangyang Zhang, and Wenqing Wang. Low-light image enhancement based on nonsubsampled shearlet transform. *IEEE Access*, 8 :63162–63174, 2020.
- [119] Bo Li, Hong Peng, Xiaohui Luo, Jun Wang, Xiaoxiao Song, Mario J Pérez-Jiménez, and Agustín Riscos-Núñez. Medical image fusion method based on coupled neural p systems in nonsubsampled shearlet transform domain. *International Journal of Neural Systems*, 31(01) :2050050, 2021.
- [120] Velmurugan Subbiah Parvathy, Sivakumar Pothiraj, and Jenyfal Sampson. A novel approach in multimodality medical image fusion using optimal shearlet and deep learning. *International Journal of Imaging Systems and Technology*, 30(4) :847–859, 2020.
- [121] Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang. Ifcnn : A general image fusion framework based on convolutional neural network. *Information Fusion*, 54 :99–118, 2020.
- [122] Hui Li and Xiao-Jun Wu. Densefuse : A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5) :2614–2623, 2018.
- [123] Yuanyuan Li, Yanjing Sun, Xinhua Huang, Guanqiu Qi, Mingyao Zheng, and Zhiqin Zhu. An image fusion method based on sparse representation and sum modified-laplacian in nsct domain. *Entropy*, 20(7) :522, 2018.
- [124] John L Johnson and Dieter Ritter. Observation of periodic waves in a pulse-coupled neural network. *Optics letters*, 18(15) :1253–1255, 1993.
- [125] Weiwei Kong, Longjun Zhang, and Yang Lei. Novel fusion method for visible light and infrared images based on nsst–sf–pcnn. *Infrared Physics & Technology*, 65 :103–112, 2014.
- [126] Liangliang Li and Hongbing Ma. Pulse coupled neural network-based multimodal medical image fusion via guided filtering and wseml in nsct domain. *Entropy*, 23(5) :591, 2021.
- [127] Christoph Baur, Shadi Albarqouni, and Nassir Navab. Melanogans : high resolution skin lesion synthesis with gans. *arXiv preprint arXiv :1804.04338*, 2018.

- [128] Saisai Ding, Jian Zheng, Zhaobang Liu, Yanyan Zheng, Yanmei Chen, Xiaomin Xu, Jia Lu, and Jing Xie. High-resolution dermoscopy image synthesis with conditional generative adversarial networks. *Biomedical Signal Processing and Control*, 64 :102224, 2021.
- [129] Christoph Baur, Shadi Albarqouni, and Nassir Navab. Generating highly realistic images of skin lesions with gans. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis : First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 5*, pages 260–267. Springer, 2018.
- [130] Zeyu Wang, Xiongfei Li, Haoran Duan, Yanchi Su, Xiaoli Zhang, and Xinjiang Guan. Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform. *Expert Systems with Applications*, 171 :114574, 2021.
- [131] B Rajalingam and R Priya. Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. *International Journal of ChemTech Research*, 11(06) :160–176, 2018.
- [132] Newton M Kinyanjui, Timothy Odonga, Celia Cintas, Noel CF Codella, Rameswar Panda, Prasanna Sattigeri, and Kush R Varshney. Fairness of classifiers across skin tones in dermatology. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020 : 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI*, pages 320–329. Springer, 2020.
- [133] Yang Xiao, Etienne Decencière, Santiago Velasco-Forero, H el ene Burdin, Thomas Bornschl ogl, Fran oise Bernerd, Emilie Warrick, and Th er ese Baldeweck. A new color augmentation method for deep learning segmentation of histological images. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 886–890. IEEE, 2019.
- [134] Sarmad Maqsood and Umer Javed. Multi-modal medical image fusion based on two-scale image decomposition and sparse representation. *Biomedical Signal Processing and Control*, 57 :101810, 2020.
- [135] Zheng Liu, Erik Blasch, Zhiyun Xue, Jiying Zhao, Robert Laganier, and Wei Wu. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision : a comparative study. *IEEE transactions on pattern analysis and machine intelligence*, 34(1) :94–109, 2011.
- [136] Guanqiu Qi, Gang Hu, Neal Mazur, Huahua Liang, and Matthew Haner. A novel multi-modality image simultaneous denoising and fusion method based on sparse representation. *Computers*, 10(10) :129, 2021.
- [137] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detec-

- tion. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.
- [138] Paul Viola and Michael J Jones. Robust real-time face detection. *International journal of computer vision*, 57 :137–154, 2004.
- [139] Ross Girshick, Forrest Iandola, Trevor Darrell, and Jitendra Malik. Deformable part models are convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 437–446, 2015.
- [140] Frédéric Jurie and Michel Dhome. A simple and efficient template matching algorithm. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 544–549. IEEE, 2001.
- [141] Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, Shavira Narrandes, Yang Wang, and Wayne Xu. Applications of support vector machine (svm) learning in cancer genomics. *Cancer genomics & proteomics*, 15(1) :41–51, 2018.
- [142] Catarina Barata, Jorge S Marques, and Jorge Rozeira. A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE transactions on biomedical engineering*, 59(10) :2744–2754, 2012.
- [143] Ardalan Benam, Mark S Drew, and M Stella Atkins. A cbir system for locating and retrieving pigment network in dermoscopy images using dermoscopy interest point detection. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 122–125. IEEE, 2017.
- [144] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.
- [145] Sasha Targ, Diogo Almeida, and Kevin Lyman. Resnet in resnet : Generalizing residual architectures. *arXiv preprint arXiv :1603.08029*, 2016.
- [146] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. Cspnet : A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 390–391, 2020.
- [147] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
- [148] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018.
- [149] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

- [150] Juan Terven and Diana Cordova-Esparza. A comprehensive review of yolo : From yolov1 to yolov8 and beyond. *arXiv preprint arXiv :2304.00501*, 2023.
- [151] Mostafa Jahanifar, Neda Zamani Tajeddin, Navid Alemi Koohbanani, Ali Gooya, and Nasir Rajpoot. Segmentation of skin lesions and their attributes using multi-scale convolutional neural networks and domain specific augmentations. *arXiv preprint arXiv :1809.10243*, 2018.
- [152] Andrey Sorokin. Lesion analysis and diagnosis with mask-rcnn. *arXiv preprint arXiv :1807.05979*, 2018.
- [153] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco : Common objects in context. In *Computer Vision–ECCV 2014 : 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.
- [154] Yuexiang Li and Linlin Shen. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 18(2) :556, 2018.
- [155] Jeremy Kawahara and Ghassan Hamarneh. Fully convolutional neural networks to detect clinical dermoscopic features. *IEEE journal of biomedical and health informatics*, 23(2) :578–585, 2018.
- [156] Alceu Bissoto, Fábio Perez, Vinícius Ribeiro, Michel Fornaciali, Sandra Avila, and Eduardo Valle. Deep-learning ensembles for skin-lesion segmentation, analysis, classification : Recod titans at isic challenge 2018. *arXiv preprint arXiv :1808.08480*, 2018.
- [157] 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*, pages 248–255. Ieee, 2009.
- [158] Eric Z Chen, Xu Dong, Xiaoxiao Li, Hongda Jiang, Ruichen Rong, and Junyan Wu. Lesion attributes segmentation for melanoma detection with multi-task u-net. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 485–488. IEEE, 2019.
- [159] Duy MH Nguyen, Thu T Nguyen, Huong Vu, Quang Pham, Manh-Duy Nguyen, Binh T Nguyen, and Daniel Sonntag. Tat1 : task agnostic transfer learning for skin attributes detection. *Medical Image Analysis*, 78 :102359, 2022.
- [160] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once : Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [161] Mehmet Şirin Gündüz and Gültekin Işık. A new yolo-based method for real-time crowd detection from video and performance analysis of yolo models. *Journal of Real-Time Image Processing*, 20(1) :5, 2023.

- [162] Amitava Halder and Debangshu Dey. Atrous convolution aided integrated framework for lung nodule segmentation and classification. *Biomedical Signal Processing and Control*, 82 :104527, 2023.