

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université

le xx mars 2023 par

Arthur Cartel FOAHOM GOUABOU

Apprentissage neuronal profond pour la détection du mélanome

Discipline

Informatique

École doctorale

184 - MATHÉMATIQUES ET INFORMATIQUE

Laboratoire/Partenaires de recherche

Laboratoire d'Informatique et Systèmes (LIS)

Invité

Rabah IGUERNAISSI
MCF, LIS, Aix-Marseille Univ.

Composition du jury

Serge MIGUET PR, LIRIS, Univ. Lumière Lyon	Rapporteur
Youssef CHAHIR MCF HDR, GREYC, Univ. de Caen	Rapporteur
Nadège THIRION PR, LIS, Univ. de Toulon	Examinatrice
Monique THONNAT DR, Centre INRIA, Univ. côte d'Azur	Examinatrice
Omar AIT AIDER MCF HDR, Institut Pascal, Univ. Clermont Auvergne	Examineur
Caroline GAUDY PU-PH, APHM, Inserm	Examinatrice
Abdellatif MOUDAFI PR, LIS, Aix-Marseille Univ.	Directeur de thèse
Jean-Luc DAMOISEAUX MCF, LIS, Aix-Marseille Univ.	Co-directeur de thèse

Affidavit

Je soussigné, Arthur Cartel Foahom Gouabou, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Abdellatif Moudafi et Jean-Luc Damoiseaux dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 13 janvier 2023



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Publications

Révisions en cours :

1. MONNIER Jilliana, **FOAHOM GOUABOU Arthur Cartel**, SERDI Meryem, COLLENNE Jules, IGUERNAISSI Rabah, GAUDY-MARQUESTE Caroline, DAMOISEAUX Jean-Luc, GROB Jean-Jacques and MERAD Djamel. *Automated detection of melanoma. Comparing an algorithm based on the characterization of disordered pattern of melanocytic lesions mimicking dermatologists practice, with a convolutional neural network approach*. Submitted In : *European Journal of Cancer* in November 2022.

Revue internationale

1. **FOAHOM GOUABOU Arthur Cartel**, COLLENNE Jules, MONNIER Jilliana, IGUERNAISSI Rabah, DAMOISEAUX Jean-Luc, MOUDAFI Abdellatif and MERAD Djamel. *Computer Aided Diagnosis of Melanoma using Deep Neural Networks and Game Theory: application on Dermoscopic Images of Skin Lesions*. In : *International Journal of Molecular Sciences*, 2022, vol. 23, no 22, p.13838.
2. **FOAHOM GOUABOU Arthur Cartel**, IGUERNAISSI Rabah, DAMOISEAUX Jean-Luc, MOUDAFI Abdellatif and MERAD Djamel. *End-to-End Decoupled Training for Long-tailed Classification applied on Skin Lesion Classification from Dermoscopic images*. In : *Electronics*, 2022, vol. 11, no 20, p.3275.
3. **FOAHOM GOUABOU Arthur Cartel**, DAMOISEAUX Jean-Luc, MONNIER Jilliana, IGUERNAISSI Rabah, MOUDAFI Abdellatif and MERAD Djamel. *Ensemble Method of Convolutional Neural Networks with Directed Acyclic Graph Using Dermoscopic Images: Melanoma Detection Application*. In : *Sensors*, 2021, vol. 21, no 12, p. 3999.

Conférences internationales

1. **FOAHOM GOUABOU Arthur Cartel**, IGUERNAISSI Rabah, DAMOISEAUX Jean-Luc, MOUDAFI Abdellatif and MERAD Djamel. *Rethinking decoupled training with bag of tricks for long-tailed recognition*. In : *2022 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2022. (Présentation orale + Poster).
2. MONNIER Jilliana, **FOAHOM GOUABOU Arthur Cartel**, SERDI Meryem, COLLENNE Jules, IGUERNAISSI Rabah, GAUDY-MARQUESTE Caroline, DAMOISEAUX Jean-Luc, GROB Jean-Jacques and MERAD Djamel. *Automated detection*

of melanoma. Comparing a Convolutional Neural Network (CNN) approach with an algorithm assessing disorder in the pattern of pigmented lesions, intended to mimic onco-dermatologists visual analysis. In : *European Society for Medical Oncology Congress (ESMO)*. Elsevier, 2022.

3. MONNIER Jilliana, **FOAHOM GOUABOU Arthur Cartel**, SERDI Meryem, COLLENNE Jules, SUFFIT Justine, IGUERNAISSI Rabah, GAUDY-MARQUESTE Caroline, RICHARD Marie-Aleth, DAMOISEAUX Jean-Luc, GROB Jean-Jacques and MERAD Djamel. *Automated detection of melanoma. Comparing an algorithm based on the characterization of pattern disorder of the lesion, mimicking dermatologist practice, with a convolutional neural network approach* . In : *European Academy of Dermatology and Venereology Congress (EADV)*, 2022.

Conférences et congrès nationaux

1. **FOAHOM GOUABOU Arthur Cartel**, IGUERNAISSI Rabah, DAMOISEAUX Jean-Luc, MOUDAFI Abdellatif and MERAD Djamel. *HMLoss: Une fonction de cout robuste au déséquilibre des classes*. In : *Actes du Colloque GRETSI*. Sept. 2022. (Poster).
2. MONNIER Jilliana, **FOAHOM GOUABOU Arthur Cartel**, SERDI Meryem, IGUERNAISSI Rabah, GAUDY-MARQUESTE Caroline, DAMOISEAUX Jean-Luc, GROB Jean-Jacques and MERAD Djamel. *Détection automatique du mélanome: comparaison d'un algorithme fondé sur la caractérisation de l'aspect désordonné de lésions mélanocytaires mimant la pratique des dermatologues, avec une approche par CNN (Convolutional Neural Network)*. In : *Annales de Dermatologie et de Vénérologie-FMC*, 2021, vol. 1, no 8, p. A135.
3. MONNIER Jilliana, **FOAHOM GOUABOU Arthur Cartel**, GAUDY-MARQUESTE Caroline, DAMOISEAUX Jean-Luc, GROB Jean-Jacques and MERAD Djamel. *Impact d'un artefact fréquent sur la détection automatique du mélanome à partir d'images dermoscopiques: approche deep learning combinée à l'algorithme Support Vector Machine*. In : *Annales de Dermatologie et de Vénérologie*. Elsevier Masson, 2020. p. A82.

Résumé

Le mélanome est un problème majeur de santé publique. Sa détection précoce est le moyen efficace pour réduire son taux de mortalité. Néanmoins, la détection manuelle des mélanomes est un processus chronophage, et le manque de dermatologues rend difficile un dépistage large et régulier de la population. Il est donc souhaitable d'investir des efforts de recherche dans le développement de systèmes automatiques de diagnostic pouvant aider les cliniciens dans cette tâche. Ces dernières années, l'apprentissage profond a permis de développer de tels systèmes atteignant des performances de diagnostic comparables à celles des dermatologues. Malgré ces résultats prometteurs, des problématiques freinent leur utilisation en contexte clinique réel. Premièrement, les lésions cutanées présentent un degré élevé de similarité, ce qui fait de leur classification automatique une tâche complexe. Deuxièmement, les jeux de données de lésions présentent une distribution asymétrique créant ainsi un biais lors de l'apprentissage des modèles. Troisièmement, le processus d'inférence des réseaux neuronaux est complexe et rend leur processus de décision difficile à comprendre pour les dermatologues.

L'objectif de ce travail est de contribuer à l'expansion des systèmes automatiques de diagnostic en contexte clinique réel. Pour cela, nous avons premièrement proposé une méthode ensembliste capable d'améliorer les performances de diagnostic du Mélanome. La méthode consiste à agréger plusieurs modèles binaires à l'aide d'un graphe acyclique dirigé. Ensuite, nous nous sommes intéressés au déséquilibre des classes et avons proposé une méthode d'apprentissage visant à rendre nos modèles robustes à ce problème. Elle consiste à initialiser l'apprentissage du modèle avec une fonction de coût L_f capable de guider le modèle vers une meilleure représentation des caractéristiques. Puis, une variante pondérée de L_f est utilisée pour affiner le réseau pré-entraîné et l'aider à mieux discriminer les classes minoritaires parmi les majoritaires. Finalement, préoccupé par la problématique d'explicabilité des réseaux de neurones, nous avons proposé une architecture, inspirée de la théorie des jeux, capable à la fois d'améliorer les performances de détection des systèmes de diagnostic existants et d'apporter plus de clarté dans leur processus de décision. L'architecture est constituée de sept réseaux de neurones convolutifs assimilés à des joueurs. La stratégie des joueurs consiste tout d'abord à regrouper les lésions selon trois niveaux de confiance : élevé, moyen et incertain. Par la suite, un sous-ensemble de joueurs a pour stratégie d'affiner le diagnostic des lésions difficiles ayant obtenues des prédictions avec des niveaux de confiance moyennes ou incertaines.

Mots clés : détection du mélanome, vision par ordinateur, apprentissage profond, méthode ensembliste, déséquilibre des classes, explicabilité

Abstract

Melanoma is a major public health problem. Early detection of melanoma is currently the most effective way to reduce its mortality rate. Nevertheless, the manual detection of melanomas is a difficult and time-consuming process; and the lack of dermatologists makes it difficult to carry out a wide and regular screening of the population. Therefore, it is desirable to invest research efforts in the development of computer aided diagnosis systems that can assist clinicians in this task. In recent years, deep learning techniques have made it possible to develop such systems able to reach the diagnostic performance of dermatologists. However, despite these promising results, some issues hinder their use in real clinical settings. First, skin lesions have a high degree of similarity, which makes their automatic classification a complex task. Second, skin lesion datasets have a skewed distribution which creates a bias when training the models. Third, the inference process of convolutional neural networks is complex and makes their decision process difficult to understand for non-experts.

The objective of this work is to contribute to the expansion of computer aided diagnosis systems in a real clinical context. To that end, we first propose an ensemble method capable of improving the diagnostic performance of Melanoma. The proposed method consists in aggregating several binary convolutional neural networks using a directed acyclic graph. Secondly, we address the class imbalance problem and propose a learning method called end-to-end decoupled training. This method aims at making convolutional neural networks robust to class imbalance. The method consists in initializing the training of the network with a loss function Lf able to guide the model towards a better representation of the features. Then, a weighted variant of Lf is used to refine the pre-trained network and help it to better discriminate minority classes from majority classes. Finally, concerned by the problem of explicability of convolutional neural networks, we propose a framework, inspired by game theory, capable of both improving the detection performance of existing diagnostic systems and bringing more clarity to their decision-making process. The proposed framework consists of a multi-class convolutional neural network and six binary convolutional neural networks assimilated to players. The strategy of the players consists first in clustering the lesions, using the introduced method of evaluating the confidence level of a prediction, according to three levels of confidence: high, medium, and uncertain. Subsequently, a subset of players has the strategy to refining the diagnosis of difficult lesions that have obtained predictions with medium or uncertain confidence levels.

Keywords: melanoma detection, computer vision, deep learning, ensemble method, class imbalance, explainability

Remerciements

Table des matières

Affidavit	2
Liste de publications et participations aux conférences	3
Résumé	5
Abstract	6
Remerciements	7
Table des matières	8
Table des figures	11
Liste des tableaux	13
Liste des acronymes	15
Glossaire	17
Introduction Générale	20
1. Le diagnostic du mélanome	24
1.1. Introduction	24
1.2. Le contexte du mélanome	25
1.2.1. Structure et fonctions de la peau	25
1.2.2. Le Mélanome	27
1.2.3. L'examen clinique et la prise en charge	29
1.3. L'apprentissage profond pour la détection du mélanome	32
1.3.1. Systèmes automatiques d'aide au diagnostic du mélanome basés sur un seul CNN	32
1.3.2. Systèmes automatiques d'aide au diagnostic du mélanome basés sur une méthode ensembliste de CNNs	35
1.3.3. Systèmes automatiques d'aide au diagnostic du Mélanome basés sur des CNNs combinés à des techniques traditionnelles d'apprentissage automatique	37
1.3.4. Défis clés du développement des systèmes automatiques d'aide au diagnostic des cancers de la peau	39
1.3.5. Métriques de performances couramment utilisées pour l'évaluation des systèmes automatiques d'aide au diagnostic	43

1.4. Conclusion	44
2. Construction d'un système de diagnostic automatique du mélanome	46
2.1. Introduction	47
2.2. État de l'art des méthodes ensemblistes de réseaux profonds	49
2.2.1. Apprentissage ensembliste de CNNs par bagging	49
2.2.2. Apprentissage ensembliste de CNNs par boosting	50
2.2.3. Apprentissage ensembliste de CNNs par stacking	50
2.2.4. Apprentissage ensembliste de CNNs appliqué à la détection du mélanome	51
2.3. Méthodes proposées	52
2.3.1. Base de données	53
2.3.2. Méthodologie	56
2.3.3. Entraînement des modèles	61
2.3.4. Définition des critères de performance utilisés	62
2.4. Résultats et discussions	63
2.4.1. Performance des modèles binaires	63
2.4.2. Impact du choix du nœud racine	64
2.4.3. Comparaison de l'approche DAG-CNN avec des CNNs multi- classes standards	65
2.4.4. Analyse des performances de la méthode DAG-CNN	66
2.4.5. Comparaison avec d'autres approches de fusion	76
2.5. Conclusion	77
3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond	79
3.1. Introduction	80
3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond	81
3.2.1. Approches du point de vue des données	82
3.2.2. Approches du point de vue de l'algorithme	83
3.2.3. Entraînement découplé	84
3.3. Définition du problème et analyses préliminaires	85
3.3.1. Le déséquilibre des classes	85
3.3.2. Effets du déséquilibre des classes sur l'apprentissage des modèles	86
3.4. Méthodes	91
3.4.1. Motivation théorique	91
3.4.2. Définition des fonctions de coûts	92
3.4.3. Description de la méthode d'entraînement découplé proposée .	94
3.4.4. Protocole expérimental	95
3.5. Résultats et discussions	103
3.5.1. Validation sur la base de données ISIC	103
3.5.2. Validation sur les jeux de données CIFAR et iNaturalist	106
3.6. Conclusion	107

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux	108
4.1. Introduction	108
4.2. Revue des méthodes d'amélioration de l'explicabilité	109
4.3. Méthodes proposées	111
4.3.1. Base de données	111
4.3.2. Description de la méthode	111
4.3.3. Entraînement des modèles	115
4.4. Résultats et discussions	115
4.4.1. Résultats des expérimentations	115
4.4.2. Discussions	119
4.5. Conclusion	120
5. Applications	121
5.1. Introduction	121
5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermatoscopiques sur la détection automatique du mélanome	122
5.2.1. Matériels et Méthodes	123
5.2.2. Résultats	125
5.3. Étude comparative des SAADs proposés combinés ou non avec la méthode d'entraînement découplé de bout en bout	128
5.3.1. Matériels et méthodes	128
5.3.2. Résultats et discussions	128
5.4. Conclusion	129
Conclusion générale et perspectives	130
5.5. Conclusion générale	130
5.6. Perspectives	133
5.6.1. Études approfondies sur l'impact des artefacts et des solutions	133
5.6.2. Études approfondies des échantillons aberrants	133
5.6.3. Mise en place d'une étude clinique de nos outils	134
5.6.4. Modélisation et intégration des concepts cliniques dans des SAADs afin d'améliorer leur explicabilité	134
Bibliographie	135
ANNEXES	153
A. Rappel théorique sur les CNNs	153
A.1. L'apprentissage profond	153
A.2. Réseaux de neurones convolutifs	157

Table des figures

1.1. Illustration de la structure de la peau	26
1.2. Illustration d'un dermoscope	29
1.3. Illustration de la méthode ABCDE	31
1.4. Illustration des principaux défis qui freinent l'utilisation des systèmes automatiques d'aide au diagnostic du mélanome.	42
2.1. Illustration de lésions cutanées	48
2.2. Illustration de la méthode DAG-CNN	53
2.3. Illustration de quelques exemples d'image avant et après avoir été pré-traitées	55
2.4. Illustration d'un bloc résiduel	57
2.5. Illustration de l'architecture ResNet-50 utilisée	57
2.6. Illustration des architectures VGG-16 et VGG-19 utilisées	58
2.7. Illustration d'un graphe acyclique dirigé	59
2.8. Courbes ROC obtenues avec la structure DAG-CNN	68
2.9. Illustration du processus de décision de notre méthode DAG-CNN	69
2.10. Exemples de prédictions correctes de Kératoses bénignes	70
2.11. Exemples de prédictions correctes de Mélanomes	71
2.12. Exemples de prédictions correctes de Naevus	72
2.13. Exemples de prédictions erronées de Kératoses bénignes	73
2.14. Exemples de prédictions erronées de Mélanome	74
2.15. Exemples de prédictions erronées de Naevus	75
3.1. Illustration de la distribution du jeu de données publiques ISIC-2018	81
3.2. Visualisation de la représentation des caractéristiques profondes apprises par le modèle <i>EfficientNetB3</i> entraîné à l'aide de la fonction de perte CE.	87
3.3. Visualisation de la représentation des caractéristiques profondes apprises par le modèle <i>efficientNetB3</i> entraîné à l'aide de la fonction de perte CS.	88
3.4. Distribution des probabilités de prédiction des échantillons de la classe de queue Dermatofibrome durant l'apprentissage du modèle <i>EfficientNetB3</i>	89
3.5. Distribution des probabilités de prédiction des échantillons de la classe de tête Naevus durant l'apprentissage du modèle <i>EfficientNetB3</i>	90
3.6. Illustration des distributions des contributions sur le coût total des deux fonctions introduites L_f et L_c comparées à celles des fonctions CE et FL.	92

3.7. Description brève du protocole expérimental de notre méthode.	96
3.8. Illustration des sept classes de lésions cutanées présentes dans la base de données ISIC-2018	98
3.9. Illustration des dix classes présentes dans le jeu de données CIFAR-10.	99
3.10. Illustration des 13 super-classes présentes dans le jeu de données iNaturalist-2018.	100
3.11. Tracé des courbes ROC des performances obtenues par le meilleur modèle de notre approche sur les sept types de lésions de la tâche.	104
3.12. Tracé des courbes d'apprentissage de EfficientNetB3 entraîné à l'aide de la perte d'entropie croisée pondérée et de notre schéma d'apprentissage.	106
4.1. Description du pipeline de notre système automatique d'aide au diagnostic basé sur la théorie des jeux.	112
4.2. Courbes ROC obtenues avec notre SAAD basé sur la théorie des jeux.	116
4.3. Illustration du processus de décision de notre SAAD basé sur la théorie des jeux.	118
5.1. Illustration d'artefacts présents sur les images dermoscopiques des lésions cutanées	123
5.2. Description de la méthodologie de génération artificielle de l'halo périphérique.	125
5.3. Illustration de l'impact du halo périphérique sur la performance des CNNs.	127
.1. Illustration de l'analogie entre un neurone biologique et un neurone artificiel	153
.2. Illustration du fonctionnement d'un neurone artificiel	154
.3. Illustration de l'entraînement d'un réseau de neurones artificiels	156

Liste des tableaux

1.1. Synthèse des méthodes développant un système automatique d'aide au diagnostic du mélanome basées sur un seul CNN.	34
1.2. Synthèse des méthodes utilisant un système automatique d'aide au diagnostic du Mélanome basées sur un assemblage de plusieurs CNNs.	36
1.3. Synthèse des méthodes utilisant un système automatique d'aide au diagnostic du Mélanome basées sur une combinaison de CNNs avec des techniques traditionnelles d'apprentissage automatique.	38
2.1. Répartition de la base de données.	54
2.2. Espace de recherche des hyperparamètres pour l'entraînement des modèles.	62
2.3. Performances des modèles binaires sur le jeu d'entraînement et de validation évaluées à l'aide d'une triple validation croisée.	64
2.4. Effets de la variation du nœud racine sur les performances de la structure DAG-CNN.	65
2.5. Validation-croisée des performances de chaque structure DAG-CNN avec la version ternaire du CNN utilisé.	66
2.6. Analyse statistique comparative des performances de la structure DAG-CNN avec des CNNs multi-classes sur le jeu de données de test.	66
2.7. Comparaison de notre approche avec des méthodes d'agrégation classiques et d'autres méthodes existantes de la littérature sur la même tâche de classification.	77
3.1. Répartition de la base de données ISIC-2018.	97
3.2. Structure détaillée de l'architecture EfficientNetB3 utilisée.	101
3.3. Performances en termes de score de BAcc obtenues par EfficientNetB3 entraîné avec diverses fonctions de perte et notre méthode d'entraînement découplé	103
3.4. Performances en termes de score de BAcc obtenues par le meilleur modèle de notre approche et d'autres travaux utilisant la même base de données ISIC-2018.	105
3.5. Performances durant la première étape de l'entraînement découplé de EfficientNetB3 entraîné à l'aide des fonctions de perte CE et L_f	105
3.6. Taux d'erreur obtenus sur les jeux de données CIFAR-10, CIFAR-100 et iNaturalist-2018.	107

4.1. Comparaison de notre approche avec d'autres SAADs existants pour la classification multi-classes des mélanomes, des naevus et des kératoses bénignes.	116
5.1. Répartition de la base de données.	124
5.2. Performances des modèles selon les prétraitements effectués sur les images dermoscopiques	126
5.3. Étude comparative des performances des systèmes de SAADs développés dans le cadre de cette thèse avec ou sans la méthode d'entraînement découplé.	129
.1. Récapitulatif des fonctions d'activation couramment utilisées en apprentissage profond.	159

Liste des acronymes

AUC

Aire sous la courbe caractéristique d'exploitation du récepteur –ou *Area under the receiver operating characteristic curve*–. 35–38, 44, 62, 66, 104, 115, 119, 125, 126

BAcc

Exactitude équilibrée –ou *Balanced accuracy*–. 33, 34, 36, 44, 62–66, 76, 77, 95, 101, 103, 105, 113, 115, 116, 119, 128, 131

CE

fonction de perte d'entropie croisée –ou *cross entropy loss function*–. 83, 86, 87, 91, 92, 94, 105

CNN

Réseau de neurone convolutif –ou *Convolutional neural network*–. 23, 25, 32–38, 40, 41, 45, 47, 49–54, 56, 57, 59, 60, 63–67, 76, 77, 80, 86, 95, 97, 103, 104, 107–113, 115, 119–122, 125, 128, 129, 131–133, 153, 157

CS

fonction de perte d'entropie croisée pondérée par l'inverse des fréquences de classe –ou *cost sensitive cross entropy loss function*–. 86, 87, 102, 105, 106, 115

DAG

Graphe acyclique dirigé –ou *directed acyclic graph*–. 47, 52, 58–60, 63–67, 76, 77, 121, 128, 129, 131

FC

Couche complètement connectée –ou *Fully connected layer*–. 33, 57, 58, 60, 86, 101, 111, 125, 157, 158

FL

fonction de perte focalisée –ou *focal loss function*–. 33, 83, 91, 94, 102

ReLU

Unité de rectification linéaire –ou *Rectified linear unit*–. 57, 58, 158, 159

ROC

Courbe caractéristique d'exploitation du récepteur –ou *Receiver Operating Characteristic*–. 44, 66, 104, 115

SAAD

Système automatique d'aide au diagnostic. 20, 22, 23, 25, 32, 33, 35, 37, 39–45, 47, 49, 50, 52–54, 67, 76, 77, 79, 80, 104, 107–112, 116, 119–122, 128–134, 153

SVM

Machine à vecteurs de support –ou *Support vector machine*–. 37, 38, 58, 64, 125

Glossaire

ABCDE

La règle ABCDE est une méthode de diagnostic clinique du mélanome qui repose sur cinq critères: l'asymétrie, la bordure irrégulière, la couleur inhomogène, le diamètre supérieur à 6 mm et l'évolutivité de la lésion. [30](#), [31](#), [37](#), [38](#)

Adaboost

Adaboost est un algorithme de classification basé sur des arbres décisionnels. [50](#)

Avg

Avg désigne la méthode d'agrégation basé sur le maximum des moyennes arithmétiques normalisées des prédictions des modèles de base. [51](#), [60](#), [76](#), [77](#)

bagging

Bagging ou bootstrap ou assemblage parallèle est une méthode d'ensemble qui consiste à entraîner différent modèle sur des ensembles sous-échantillonnés de la base de départ. [49-51](#)

boosting

boosting ou assemblage séquentielle est une méthode d'ensemble qui consiste à construire progressivement un ensemble en formant chaque nouvelle instance de modèle pour mettre l'accent sur des données d'entraînement que les modèles précédents ont mal classées. [49](#), [50](#)

Breslow

L'indice de Breslow mesure l'épaisseur maximale de la tumeur exprimée en millimètres. Il est l'un des principaux critères de classification du mélanome. [32](#), [63](#)

CIFAR

CIFAR est une large base de données publiques d'images. Selon la version, le jeu de données contient des images appartenant à 10 ou 100 catégories.. [97](#), [99](#), [101](#), [107](#)

Dropout

Le Dropout est une méthode de régularisation en apprentissage automatique. Elle consiste à supprimer de façon aléatoire des neurones dans un réseau neuronal. Elle permet de réduire considérablement le risque de surajustement durant l'entraînement des modèles. [158](#)

Gmean

Gmean désigne la méthode d'agrégation basé sur le maximum des moyennes géométriques normalisées des prédictions des modèles de base. [51](#), [60](#), [61](#), [76](#), [77](#)

halo

Le halo périphérique est un artefact fréquent présent sur des images dermoscopiques de lésions cutanées.. [23](#), [40](#), [45](#), [121–126](#), [129](#), [132](#), [133](#)

illuminant

En colorimétrie, un illuminant représente une source fictive définie par un tableau indiquant, par bandes de longueurs d'onde, l'intensité lumineuse relative que fournit la source dans le spectre visible. [56](#)

ImageNet

ImageNet est une base de données publiques d'images annotées produit par l'organisation du même nom, à destination des travaux de recherche en vision par ordinateur. [33](#), [56](#), [62](#)

iNaturalist

iNaturalist est une large base de données publiques d'images de plantes et d'animaux.. [97](#), [99](#), [100](#), [102](#), [107](#)

ISIC

L'archive ISIC est une plateforme open source contenant des images de lésions cutanées. Les images sont associées à des diagnostics de terrain et à d'autres métadonnées cliniques.. [22](#), [23](#), [33](#), [37](#), [40](#), [53](#), [54](#), [56](#), [80](#), [86](#), [97](#), [101](#), [102](#), [104](#), [107](#), [111](#), [123](#), [124](#), [128](#)

Max

Max désigne la méthode d'agrégation basé sur le maximum des prédictions des modèles de base. [51](#), [60](#), [61](#), [76](#), [77](#)

Max-Win

Max-win est une méthode d'agrégation basé sur vote majoritaire des modèles de base. [51](#), [52](#), [60](#), [61](#), [76](#), [77](#), [114](#)

phénotype

Elément de classification des individus selon la réaction de leur peau à une exposition solaire. [28](#)

Prod

Prod désigne la méthode d'agrégation basée sur le maximum des produits normalisés des prédictions des modèles de base. [51](#), [60](#), [61](#), [76](#), [77](#)

Retinopathy

Retinopathy est une large base de données publiques d'images rétiniennes à haute résolution prises dans diverses conditions d'imagerie. [33](#)

stacking

Le stacking est une méthode d'ensemble qui consiste à entraîner différents modèles de base sur les mêmes données et utiliser un autre modèle pour apprendre comment combiner au mieux les prédictions faites par l'ensemble des modèles de base. [49-51](#)

tenseur

Un tenseur est un tableau multidimensionnel contenant des données. [54](#), [157](#)

validation-croisée

La validation croisée -ou cross validation- est une méthode employée en apprentissage automatique pour estimer la fiabilité d'un modèle. [53](#), [62](#), [65](#), [76](#)

Introduction Générale

L'apprentissage profond (*deep learning* en anglais) est un domaine de l'apprentissage automatique qui consiste à utiliser des réseaux de neurones profonds comportant de nombreuses couches cachées, pour apprendre à effectuer des tâches complexes. L'apprentissage profond est largement inspiré de la structure et du fonctionnement du cerveau humain, en particulier de la façon dont le cerveau humain reconnaît et traite les stimuli visuels. Les premières recherches en apprentissage profond débutent dans les années 1950, mais c'est seulement dans les années 2010 que l'apprentissage profond a véritablement pris son essor grâce aux progrès réalisés dans les domaines de l'informatique et de la science des données. Depuis, l'apprentissage profond a été largement utilisé dans de nombreux domaines, notamment la reconnaissance de la parole et de l'image, la traduction automatique et la création de contenu génératif comme la musique ou les écrits.

Le diagnostic précoce du mélanome, une forme mortelle de cancer de la peau, est crucial pour augmenter les chances de guérison et de survie des patients. Traditionnellement, la détection des cancers de la peau est effectuée par des dermatologues qui examinent manuellement la peau des patients à la recherche de signes de cancer. Cependant, cette méthode peut être imparfaite, car elle dépend de la compétence et de l'expérience du dermatologue, et peut manquer de précision dans certains cas. L'apprentissage profond peut être utilisé pour développer des [Système automatique d'aide au diagnostic \(SAAD\)](#)s de mélanome à partir d'images de lésions de la peau. Ces systèmes peuvent être entraînés sur des bases de données d'images de lésions cutanées annotées, c'est-à-dire accompagnées d'étiquettes indiquant si une image contient ou non un type de cancer de la peau. Une fois entraînés, ces systèmes peuvent être utilisés pour analyser automatiquement de nouvelles images de la peau et détecter les signes de cancer de la peau de manière rapide et fiable. L'utilisation de l'apprentissage profond pour la détection du mélanome présente de nombreux avantages par rapport à la détection manuelle par un dermatologue. En particulier, elle permet de détecter les mélanomes de manière plus précise et plus rapide, ce qui peut contribuer à réduire le nombre de diagnostics erronés et à améliorer les chances de guérison des patients. De plus, l'utilisation de l'apprentissage profond peut également aider à réduire les coûts de santé en permettant de détecter les cancers de la peau plus tôt, ce qui peut éviter les coûts d'un traitement sur le long terme. Les travaux de Esteva *et al.* [53] ont été les précurseurs sur l'utilisation de l'apprentissage profond pour la détection des cancers de la peau. Dans cet article, les auteurs ont développé un système de détection de cancers de la peau basé sur l'apprentissage profond et l'ont entraîné sur une base de données d'images de lésions cutanées annotées par des dermatologues. Lorsqu'ils

ont testé leur système sur un ensemble de données indépendant, ils ont constaté que leur système pouvait classer avec une précision de 95% les images de la peau selon le type de cancer, ce qui est comparable aux performances des dermatologues. Cet article a ouvert la voie à de nombreuses autres recherches sur l'utilisation de l'apprentissage profond pour la détection des cancers de la peau et a contribué à l'émergence de nombreux systèmes de détection automatique de cancers de la peau basés sur l'apprentissage profond. A ce jour, à titre d'illustration, on recense près de 165 000 résultats sur le moteur de recherche google scholar répondant au terme de recherche « Apprentissage profond pour la détection des cancers cutanés ».

Malgré les résultats prometteurs apportés par l'apprentissage profond pour la détection des cancers de la peau en général et celui du mélanome en particulier, un certain nombre de problématiques demeurent :

- Un des principaux défis de la détection des cancers de la peau à l'aide de l'apprentissage profond est d'améliorer la précision de la détection. Il est important de minimiser le taux de faux positifs et de faux négatifs pour éviter d'une part de donner de faux diagnostics aux patients et d'autre part de manquer des cas de cancer dont la mortalité est très forte ;
- Pour que les modèles d'apprentissage profond soient efficaces pour la détection des cancers de la peau, ils doivent être entraînés sur des bases de données diversifiées représentant une grande variété de types de peau, de couleurs de peau et de types de cancers de la peau ;
- Les données d'images de la peau utilisées pour entraîner un modèle présentent en général un déséquilibre des données, d'où pour l'entraînement du modèle une sous-représentation d'une certaine classe (par exemple un type de cancer de la peau) en comparaison des autres classes, induisant un biais au niveau des modèles. Par exemple, si un modèle d'apprentissage profond a été entraîné sur une base de données qui a beaucoup plus d'images de cancers de type A que de cancers de type B, il pourrait être beaucoup plus précis pour détecter les cancers de type A que les cancers de type B. Cela a pour conséquence d'entraîner des erreurs de diagnostic pour les patients atteints de cancer de type B, et avoir de graves conséquences sur leur santé ;
- Le fonctionnement interne d'un réseau de neurones, c'est-à-dire comment il prend une décision à partir des données d'entrée, est généralement difficile à comprendre pour les dermatologues, qui parlent alors de « boîte noire ». N'étant pas en mesure de leur garantir l'exactitude et la fiabilité du système, la question de leurs responsabilités se pose ;
- La variabilité de la qualité des images est une autre problématique que rencontre le développement des systèmes automatiques profonds pour la détection des cancers cutanés. En effet, si la qualité des images dans le jeu de données d'entraînement est variable (par exemple certaines images sont très floues alors que d'autres sont nettes) ou encore que les images ont été prises avec différents appareils de capture (par exemple des scanners de peau de différentes marques) conduisant ainsi à des variations d'illumination ou de couleur, tout cela rendra

plus difficile l'apprentissage pour le modèle, diminuera sa fiabilité à détecter les cancers et affectera sa précision ;

- Une fois qu'un système de détection de cancers de la peau a été développé et testé en laboratoire, il est important de le mettre en œuvre en milieu clinique et de le valider auprès de patients réels. Cela peut poser des défis liés à la collecte des données en milieu clinique, à l'intégration du système dans les workflows cliniques existants et à la gestion sensible des données de patients.

La liste non exhaustive des problématiques ci-dessus fait de l'apprentissage profond pour la détection du mélanome un sujet de recherche pertinent et d'actualité. L'objectif de cette thèse est de contribuer à la mise en œuvre en milieu clinique des SAADs, ceci afin d'accompagner les dermatologues dans la détection des cancers cutanés. Pour atteindre cet objectif, nous avons développé de nouvelles architectures et méthodes d'apprentissage pour la tâche de diagnostic du mélanome, proposé des approches innovantes pour augmenter la robustesse des modèles neuronaux au problème de déséquilibre des classes, mais aussi l'explicabilité des SAADs. Nos méthodes ont été évaluées sur des données publiques réelles et ont été comparées aux approches de l'état de l'art.

Une des principales contributions de cette thèse est le développement d'un SAAD qui utilise une méthode de décomposition et ensembliste couplé à un graphe acyclique dirigé pour améliorer la performance de la classification multi-classes d'images dermoscopiques. Ce système a obtenu le meilleur score de précision équilibrée de 76,6% sur le jeu de données public ISIC-2018, surpassant ainsi les autres travaux connexes et des modèles multi-classes seuls. Ce travail a fait l'objet d'une publication pour un [article de revue](#) et est décrit dans le [Chapitre 2](#).

Outre l'amélioration de la précision des SAADs, nous nous sommes également penchés sur la résolution du problème du déséquilibre des classes dans les données disponibles. Dans le contexte de la détection des lésions cutanées, le déséquilibre des classes est encore exacerbé par l'existence de maladies rares et les conditions complexes au niveau du patient. La solution proposée dans cette thèse est une approche d'apprentissage découplé de bout en bout, qui utilise une nouvelle fonction de perte pour guider le réseau vers une meilleure représentation des caractéristiques. Une variante pondérée de cette fonction de perte a été aussi proposée pour améliorer la robustesse du réseau au déséquilibre des classes. Testée sur le jeu de données public ISIC-2018, cette approche a surpassé toutes les méthodes comparées avec une marge minimale de 2% avec un seul modèle. Ce travail a fait l'objet de trois publications, une dans une [revue](#), une autre dans une [conférence internationale](#) et une autre dans une [conférence nationale avec comité de lecture](#). Ce travail est décrit dans le [Chapitre 3](#).

La troisième contribution de cette thèse est la proposition d'une approche innovante pour augmenter l'explicabilité des SAADs et apporter plus de transparence à leur processus de décision. Pour résoudre ce problème, le système proposé dans cette thèse se base sur la théorie de la décision, en particulier la théorie des jeux, pour

assurer la transparence du processus de décision. Le système se compose d'un [Réseau de neurone convolutif](#) –ou *Convolutional neural network*– (CNN) multi-classes et de six CNNs binaires, assimilés à des joueurs. Les stratégies des joueurs sont choisies en fonction de leurs probabilités prédites, et la décision finale est prise par un processus de vote, la classe gagnante étant déterminée par l'utilité attendue la plus élevée. Cette approche a été testée sur le jeu de données public ISIC-2018 et a obtenu une précision équilibrée de 86,7%. Ces approches peuvent aider les dermatologues à comprendre le raisonnement qui sous-tend les décisions du SAAD et améliore la confiance et l'adoption de ces systèmes en milieu clinique. Ce travail a fait l'objet d'une publication dans une [revue](#) et est décrit dans le [Chapitre 4](#).

Enfin, nous avons mené une étude sur l'impact d'un des artefacts présents de manière récurrente sur les jeux de données d'images de lésion cutanées qu'est l' [halo](#) périphérique. Cette étude a permis d'établir pour la première fois les effets négatifs de cet artefact lors de l'apprentissage des modèles et a permis de poser les bases de résolution de ce problème. Cette étude a fait l'objet d'une communication dans un [congrès national](#) et est décrite dans le [Chapitre 5](#).

Une partie des travaux présentés dans cette thèse a été financée par l'Agence nationale de la recherche française à travers le projet [DIAMELEX](#) (ANR-20-CE45-0026).

Dans l'ensemble, le travail présenté dans cette thèse a fait progresser l'état de l'art dans le développement des SAADs pour le diagnostic précoce du mélanome et a abordé plusieurs défis qui ont entravé leur utilisation dans les milieux cliniques. Nos méthodes ont le potentiel d'améliorer l'efficacité et la précision du diagnostic du mélanome et peuvent contribuer au développement de SAADs plus transparents et fiables pour une utilisation dans le domaine médical.

1. Le diagnostic du mélanome

Sommaire

1.1. Introduction	24
1.2. Le contexte du mélanome	25
1.2.1. Structure et fonctions de la peau	25
1.2.2. Le Mélanome	27
1.2.2.1. Les différents types de cancers cutanés	27
1.2.2.2. Les facteurs de risque du mélanome	28
1.2.2.3. L'incidence du mélanome	28
1.2.3. L'examen clinique et la prise en charge	29
1.2.3.1. Le dermoscope	29
1.2.3.2. Les méthodes de diagnostic clinique	30
1.2.3.3. La prise en charge	30
1.3. L'apprentissage profond pour la détection du mélanome	32
1.3.1. Systèmes automatiques d'aide au diagnostic du mélanome basés sur un seul CNN	32
1.3.2. Systèmes automatiques d'aide au diagnostic du mélanome basés sur une méthode ensembliste de CNNs	35
1.3.3. Systèmes automatiques d'aide au diagnostic du Mélanome basés sur des CNNs combinés à des techniques traditionnelles d'apprentissage automatique	37
1.3.4. Défis clés du développement des systèmes automatiques d'aide au diagnostic des cancers de la peau	39
1.3.4.1. Les défis techniques	39
1.3.4.2. Les défis non-techniques	40
1.3.5. Métriques de performances couramment utilisées pour l'évaluation des systèmes automatiques d'aide au diagnostic	43
1.4. Conclusion	44

1.1. Introduction

Le mélanome représente un problème majeur de santé publique avec un taux d'incidence qui ne cesse d'augmenter. Lorsque le diagnostic intervient tôt dans l'évolution de la lésion, le taux de survie relative à 5 ans est de 88% pour les stades localisés. A contrario, la survie à 5 ans d'une personne atteinte d'un mélanome à un stade avancé, avec des métastases chute en deçà de 20% [59]. Détecter un mélanome à un stade

précoce, alors que relativement peu de signes de malignité sont visibles sur la lésion est donc très souhaitable pour assurer la survie du patient. Malgré l'utilisation d'outil innovant tel que le dermoscope et l'application des méthodes d'analyse, le diagnostic clinique des lésions cutanées reste très difficile, et la justesse du diagnostic est étroitement corrélée à l'expérience du dermatologue. Assister les dermatologues avec un SAAD pourrait permettre d'améliorer la reproductibilité du jugement clinique et d'aider les cliniciens inexpérimentés en leur donnant un second avis in-vivo. L'apprentissage profond ou - *Deep learning*- en général et les CNNs en particulier ont contribué à l'évolution considérable des performances de tels systèmes pour la détection de mélanome. Cependant, un certain nombre de questions restent ouvertes et freinent l'utilisation de tels systèmes en contexte clinique réel.

Dans ce chapitre introductif, nous présenterons tout d'abord les bases de la biologie de la peau humaine et ferons une description du mélanome et de son diagnostic clinique. Ensuite, nous effectuerons un état de l'art des SAADs existants pour le diagnostic d'une lésion cutanée, et exposerons une synthèse des défis clés entravant le développement de tels systèmes en contexte clinique réel. Finalement, nous ferons un rappel des fondements théoriques des CNNs.

1.2. Le contexte du mélanome

1.2.1. Structure et fonctions de la peau

La peau est un organe à part entière du corps humain et constitue la première ligne de défense du corps contre les bactéries et les virus. La peau assure plusieurs fonctions indispensables à la vie. Outre son rôle protecteur contre les agressions extérieures, elle aide à réguler la température corporelle, joue le rôle d'organe sensoriel grâce à ses terminaisons nerveuses mais également participe à la synthèse de la vitamine D essentiel au bon fonctionnement de l'organisme humain [124].

D'un point de vue chimique, la peau comprend en moyenne : 70% d'eau, 27,5% de protéines, 2% de matières grasses et 0,5% de sels minéraux et oligo-éléments [57]. La peau est un organe en perpétuelle évolution qui réagit de manière instantanée à toute sollicitation de son environnement. Elle est constituée de trois couches principales [124] : l'épiderme, le derme et l'hypoderme (Figure 1.1).

- L'**épiderme** est l'enveloppe externe de la peau. Il contient différents types de cellules, dont les mélanocytes. Il est lui-même composé de quatre étages : la couche cornée, la couche granuleuse, la couche épineuse et la couche basale. Dépourvu de vaisseaux sanguins, l'épiderme doit compter sur le derme pour se nourrir ;
- Le **derme** est une couche intermédiaire qui se situe sous l'épiderme. Il constitue le principal tissu de soutien de la peau et est responsable de sa solidité. Il est

1. Le diagnostic du mélanome – 1.2. Le contexte du mélanome

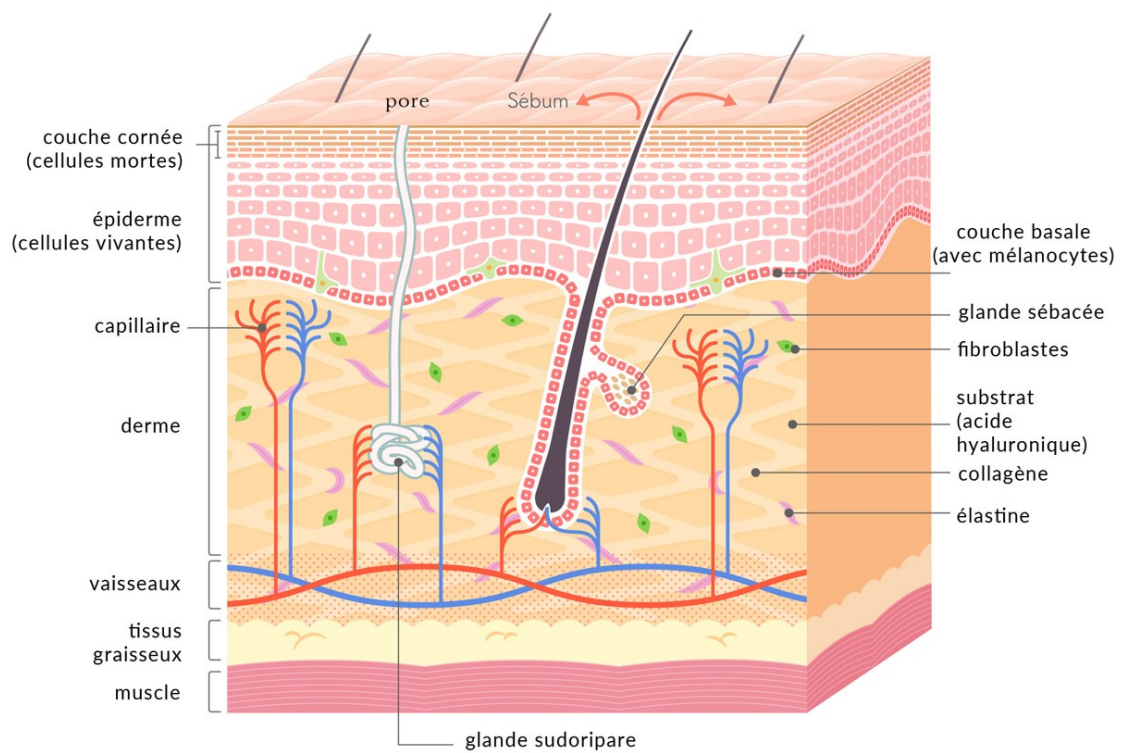


FIGURE 1.1. – Structure de la peau(Source :[122]).

1. Le diagnostic du mélanome – 1.2. Le contexte du mélanome

composé principalement de fibroblastes, cellules responsables de la production de fibres de collagène et de fibres élastiques, à l'origine de la souplesse et de l'élasticité de la peau. Le derme contient également des vaisseaux sanguins, des vaisseaux lymphatiques et des terminaisons nerveuses. Il est également le siège des follicules pileux (qui forment la base du poil), des glandes sudoripares (qui produisent la sueur) et sébacées (qui produisent le sébum). Tous ces constituants assurent l'hydratation et la nutrition de la peau et participent à la protection du corps contre les agressions ;

- L'**hypoderme** ou tissu graisseux est la couche la plus profonde de la peau et se situe sous le derme. Il est constitué de cellules graisseuses, les adipocytes. Cette couche de graisse stocke l'énergie, assure l'isolation pour conserver la chaleur du corps et protège les organes recouverts.

L'épaisseur de la peau varie en moyenne de 0,5 millimètre à 3 centimètres selon les parties du corps (cette épaisseur tient compte des 3 couches de la peau citées ci-dessus). Elle est la plus fine au niveau des paupières et la plus épaisse au niveau du dos, de l'abdomen ou encore au niveau des paumes des mains et des pieds.

La peau étant l'organe le plus en contact avec l'environnement extérieur, elle subit plusieurs types d'agression de différentes natures allant des agressions mécaniques (choc, frottements, pressions) aux agressions solaires (rayonnements ultra-violets) en passant par des agressions chimiques (produits toxiques) et microbiennes (bactéries, virus, champignons, ...). Parmi les types d'agressions précédemment citées, les rayons ultraviolets ont la particularité de pouvoir endommager l'ADN et de conduire à des effets néfastes comme des cancers de la peau [19].

1.2.2. Le Mélanome

Dans la [sous-section](#) précédente, nous avons présenté les bases de la biologie de la peau humaine ce qui nous a permis de situer le contexte biologique dans lequel s'inscrit notre thèse. Dans cette sous-section, nous nous intéresserons plus spécifiquement au mélanome et nous présenterons l'enjeu sociétal que représente cette maladie, ceci afin de mettre en avant l'intérêt d'un sujet de thèse portant sur la détection du mélanome. Le mélanome n'étant pas le seul cancer cutané existant, nous présenterons tout d'abord les différents types de cancer cutané qui existent. Par la suite, nous nous pencherons sur les facteurs de risque contribuant à l'apparition du mélanome. Enfin, nous ferons une brève analyse épidémiologique du mélanome.

1.2.2.1. Les différents types de cancers cutanés

Les cancers cutanés peuvent se diviser en deux catégories principales : les cancers non mélanocytaires et les mélanomes.

Le mélanome cutané : Le mélanome cutané est une tumeur qui se développe à partir des cellules mélanocytaires. Elle a pour effet de conduire à une multiplication

1. Le diagnostic du mélanome – 1.2. Le contexte du mélanome

de façon anarchique et incontrôlée de mélanocytes formant ainsi une lésion maligne au niveau de la peau [105]. Le mélanome est le cancer cutané le moins fréquent (5% des cancers cutanés) mais le plus agressif représentant la première cause de mortalité relative aux cancers cutanés [66]. Il est considéré comme une épidémie en forte augmentation notamment chez la population caucasienne [48]. Il se développe en majorité (80% des cas) sur une peau ne présentant aucune tache ou lésion et dans 20% des cas à partir d'un grain de beauté (naevus) préexistant [105].

Les cancers non-mélanocytaires : Les cancers non-mélanocytaires représentent la grande majorité des cancers cutanés. Les deux principaux types de cancers non-mélanocytaires (représentant environ 90% des cancers cutanés) sont le carcinome basocellulaire et le Carcinome épidermoïde [49]. Le carcinome basocellulaire est le cancer le plus répandu et se développe à partir des cellules de la couche basale de l'épiderme. Il est dû à des expositions fréquentes et répétées au soleil et à des coups de soleil importants lors de l'enfance. Le carcinome basocellulaire présente très peu de forme métastatique conduisant à un taux de mortalité faible. En l'absence de prise en charge, il peut se développer en profondeur et nécessitera alors une exérèse chirurgicale [105]. Le Carcinome épidermoïde quant à lui se développe à partir de la couche épineuse de l'épiderme. Il a l'aspect d'une croûte et peut atteindre toutes les parties du corps, y compris la muqueuse buccale et génitale mais est plus fréquent sur les zones exposées au soleil [105]. Le Carcinome épidermoïde a un pronostic moins assuré que le carcinome basocellulaire car il peut métastaser et nécessitera en cas de mauvaise prise en charge, une large exérèse.

1.2.2.2. Les facteurs de risque du mélanome

L'exposition fréquente et prolongée au soleil notamment aux rayonnements ultraviolets constitue le principal facteur de risque associé au mélanome [145]. Ceci est dû au fait que les rayonnements ultraviolets génèrent plusieurs altérations comme la production d'espèces réactives de l'oxygène, des modifications de la réponse immunitaire, la production de facteurs de croissance et des mutations génétiques [48, 98]. D'autres facteurs moins prépondérants existent et sont liés à l'hôte comme : un nombre important de nævi mélanocytaires (grains de beauté), un **phénotype** clair (yeux, cheveux, peau), antécédents personnels ou familiaux de mélanome, des taches de rousseur, et certaines maladies génétiques rares (Albinisme, Syndrome de Gorlin, Xeroderma pigmentosum) [105, 49].

1.2.2.3. L'incidence du mélanome

L'Institut national du cancer [135] estime à 15 513 le nombre de nouveaux cas de mélanome de la peau en France métropolitaine en 2018. L'incidence du mélanome n'a cessé d'augmenter ces dernières décennies, et a même été multiplié par 5 entre 1990 et 2018 chez l'homme et par près de 3 chez la femme [41]. De plus, le mélanome est le cancer qui augmente le plus chez l'homme parmi les tumeurs solides entre 2010

1. Le diagnostic du mélanome – 1.2. Le contexte du mélanome

et 2018 [135]. L'âge médian au diagnostic et au décès étant respectivement de 66 et 71 ans chez l'homme et de 60 et 73 ans chez la femme [135].

Ces données nous permettent de conclure aisément que le mélanome cutané constitue donc un enjeu majeur de santé publique.

1.2.3. L'examen clinique et la prise en charge

La première étape lors du diagnostic d'une lésion maligne par un dermatologue est l'examen visuel des régions suspectes. Un bon diagnostic est important du fait de la grande similarité entre certains types de lésions cutanées.

Dans cette sous-section, nous présenterons le dermoscope qui est l'outil couramment utilisé par les dermatologues durant leur examen clinique. Puis nous nous intéresserons aux méthodes employées par le dermatologue pour poser un diagnostic clinique sur une lésion cutanée.

1.2.3.1. Le dermoscope



FIGURE 1.2. – Photographie d'un dermoscope portable à lumière non polarisée et polarisée utilisé par les dermatologues lors de leur examen clinique (Source : [61]).

Le dermoscope (également appelé dermatoscope ou microscope à épiluminescence) est un outil largement utilisé par les dermatologues pour l'examen clinique

1. Le diagnostic du mélanome – 1.2. Le contexte du mélanome

des lésions cutanées. La figure 1.2 illustre l'utilisation d'un dermoscope par un dermatologue. Il s'agit d'un outil dermatologique non invasif permettant de visualiser les motifs et les structures des lésions cutanées avec une haute résolution. Il intègre une lentille de grossissement et différents schémas d'éclairage, tels que la lumière non polarisée et la lumière polarisée. La lumière polarisée permet de minimiser la réflexion de la lumière sur la surface de la peau et de mettre en évidence des motifs détaillés et la vascularisation de la lésion cutanée. Le dermoscope aide les dermatologues à reconnaître des caractéristiques spécifiques pour le diagnostic précoce du mélanome qui ne sont parfois pas visibles à l'œil nu [61]. La précision du diagnostic est fortement liée à l'expérience du dermatologue [74]. Sans support technique, la précision du diagnostic du mélanome par les dermatologues se situe entre 65%-80% [11], tandis qu'à l'aide d'un dermoscope, celle-ci se situe dans la plage 75%-84% [54, 6].

1.2.3.2. Les méthodes de diagnostic clinique

Le diagnostic clinique d'un mélanome repose sur l'analyse sémiologique à l'œil nu ou à l'aide d'un dermoscope d'une lésion cutanée.

Une des méthodes très utilisées d'analyse sémiologique à l'œil nu repose sur la règle ABCDE [47]. La règle ABCDE définit cinq critères déterminant pour le diagnostic du mélanome, l'asymétrie(A), la bordure(B) irrégulière, la couleur(C) inhomogène, le diamètre(D) supérieur à 6 mm et l'évolutivité(E) récente. La présence d'un ou plusieurs de ces critères permet d'établir un score de classification (maligne ou non) de la lésion [59]. La figure 1.3 illustre les cinq critères de la règle ABCDE.

Certaines formes cliniques ne répondent pas strictement aux critères de la règle ABCDE et rendent le diagnostic plus difficile [105]. Le dermatologue se retrouve ainsi dans l'obligation d'utiliser des approches plus globales comme l'identification du signe du « vilain petit canard » ou l'analyse des patrons à l'aide d'un dermoscope [49, 65].

Le concept de « vilain petit canard » décrit qu'une lésion mélanocytaire ne peut pas être évaluée séparément des autres naevus d'un individu donné [72]. Cette approche consiste donc pour le Dermatologue à inspecter l'ensemble des naevus d'un patient et à identifier celui ou ceux qui sont manifestement différents. La sensibilité de cette approche a été évaluée à 85% [49]. L'analyse des patrons à l'aide d'un dermoscope [12] consiste à reconnaître les patrons les plus courants associés à des lésions bénignes ou dans le cas contraire, rechercher des signes qui feraient suspecter une malignité [49].

1.2.3.3. La prise en charge

Le taux de survie relative à 5 ans d'un patient atteint d'un mélanome est de 91%. Cependant, ce taux présente de grandes variations en fonction du stade d'évolution de la lésion cutanée. Pour des stades localisés, ce taux se situe entre 40% et 98%,

DÉTECTION PRÉCOCE DU MÉLANOME (source INCa)
Soyez vigilant à tout changement :

- nouvelle tache brune qui apparaît
- grain de beauté différent des autres ou qui change d'aspect : gardez à l'esprit la règle « ABCDE »

ASYMÉTRIE
Grain de beauté de forme ni ronde ni ovale et dont les couleurs et les reliefs ne sont pas régulièrement répartis autour de son centre.

BORDS IRRÉGULIERS
Bords déchiquetés mal délimités.

COULEUR NON HOMOGENE
Présence désordonnée de plusieurs couleurs (noir, bleu, marron, rouge ou blanc).

DIAMÈTRE EN AUGMENTATION
Une lésion susceptible d'être un mélanome a généralement une taille supérieure à 6 mm.

ÉVOLUTION
Changement rapide de taille, de forme, de couleur ou d'épaisseur.

- La présence d'un ou de plusieurs de ces signes ne signifie pas que vous avez un mélanome, mais justifie de demander un avis médical.
- Il est important de consulter un dermatologue en cas de symptômes cutanés ou de facteurs de risque de cancers de la peau.

FIGURE 1.3. – Illustration de la règle [ABCDE](#) pour le diagnostic du mélanome (Source : [135]).

tandis que pour des personnes atteintes d'un mélanome à un stade avancé avec des métastases, leur taux de survie varie entre 9% et 19% [59]. La prise en charge du mélanome dépend ainsi du stade de ce dernier. Pour des stades localisés, la prise en charge est essentiellement chirurgicale et repose sur une excision chirurgicale avec des marges prédéfinies selon l'indice de Breslow [49]. Pour un cas de mélanome métastatique, la chirurgie n'est pas suffisante et la prise en charge s'accompagne d'un traitement médicamenteux [48]. Le traitement médicamenteux repose principalement sur l'immunothérapie notamment les thérapies anti-PD-1 et anti-CTLA4 [126].

1.3. L'apprentissage profond pour la détection du mélanome

Dans la littérature, plusieurs méthodes sont présentées afin de développer des SAADs du mélanome. Ces méthodes peuvent être regroupées en deux grandes familles en fonction de leur approche du problème [131, 137, 45, 84] : La première reposant sur les approches classiques des classifieurs automatiques, et la seconde sur les approches dites profondes à base de réseaux de neurones. Dans le cadre de notre thèse, nous centrerons notre étude sur les approches dites profondes. S'agissant de ces dernières, elles peuvent être subdivisées en trois types [131] : celles basées sur un seul CNN, celles combinant plusieurs CNNs par une méthode de fusion, et celles combinant des CNNs avec des techniques classiques d'apprentissage automatique.

Dans cette section, nous explorerons les approches d'apprentissage profond développées pour construire des SAADs du mélanome. Nous présenterons ensuite les défis clés du développement des SAADs du mélanome. Enfin nous aborderons les différentes métriques utilisées par ces systèmes.

1.3.1. Systèmes automatiques d'aide au diagnostic du mélanome basés sur un seul CNN

De nombreux travaux présentent des SAADs du mélanome conçus à l'aide d'un seul CNN. La plupart d'entre eux consistent à étudier différentes architectures de CNNs afin de comparer leurs performances et de sélectionner le CNN qui aura obtenu les meilleurs résultats.

Le premier travail utilisant un CNN pour la détection des lésions cutanées a été celui réalisé par Esteva *et al.* [53]. Ils ont entraîné un CNN à l'aide d'une grande base de données comprenant 129 450 images cliniques appartenant à 2 032 types de lésions cutanées. Les performances du CNN ainsi entraîné ont été comparées avec celles obtenues par 21 dermatologues certifiés pour effectuer une classification binaire, sur des images cliniques prouvées par biopsie. Les deux problématiques de diagnostic clinique évaluées par leurs travaux étaient : la classification des cancers non mélano-

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

cytaires contre des lésions cutanées de type kératose séborrhéique; et la classification du mélanome par rapport à des naevus. Leurs résultats ont montré que les SAADs basés sur un CNN atteignent des performances similaires à celles des experts, démontrant un niveau de compétence comparable à celui des dermatologues. Similairement, Lucius *et al.* [101] dans leurs travaux ont évalué les performances de huit CNNs dans la catégorisation de sept lésions cutanées courantes. Ils ont observé que le CNN le moins précis a donné de meilleurs résultats que les médecins généralistes et qu'un CNN peut améliorer la précision du diagnostic des médecins généralistes dans un contexte clinique. Les travaux de Lopez *et al.* [139] comparent différentes techniques d'apprentissage d'un CNN sur des lésions cutanées : (1) apprentissage à partir de zéro; (2) apprentissage par transfert à partir d'un modèle pré-entraîné sur un jeu de données plus important (ImageNet) en n'entraînant que les couches de type Couche complètement connectée –ou Fully connected layer– (FC); (3) apprentissage par transfert en entraînant une portion des couches profondes du CNN. La dernière technique d'apprentissage est celle qui a conduit aux meilleures performances, démontrant ainsi la robustesse de l'apprentissage par transfert sur des images dermoscopiques. Menegola *et al.* [112] ont effectué une étude plus approfondie des techniques d'apprentissage par transfert appliquées à la détection du mélanome. Ils ont pour cela comparé la performance sur des images dermoscopiques des modèles pré-entraînés sur différents jeux de données plus importants : Retinopathy et ImageNet. Curieusement, les résultats obtenus ont démontré que l'apprentissage par transfert à partir d'une tâche connexe (dans leur cas, à partir de Retinopathy, une tâche de classification médicale) donnait de moins bons résultats que l'apprentissage par transfert à partir d'une tâche moins connexe (à partir de ImageNet, une tâche de classification non-médicale).

D'autres travaux intègrent des innovations à la simple utilisation d'un réseau pré-entraîné. Kawahara *et al.* [90] par exemple, contrairement à l'utilisation traditionnelle des CNNs généralement entraînés sur une image à une seule résolution, ont proposé une nouvelle architecture de CNN conçue pour apprendre à partir d'informations provenant de différentes résolutions de l'image. Leur architecture est composée de plusieurs voies, et analyse simultanément différentes résolutions d'images et les interactions entre ces différentes résolutions en utilisant le même champ de vision. Dans le même contexte, Zhang *et al.* [176] ont proposé une nouvelle architecture de CNN intégrant un module d'attention. Le réseau proposé exploite la capacité intrinsèque d'auto-attention d'un CNN et génère des cartes d'attention dans les couches inférieures pour améliorer les performances de classification. Yao *et al.* [168] ont combiné trois techniques d'apprentissage profond afin de concevoir un SAAD robuste pour la classification multi-classes de lésions cutanées. Les techniques utilisées étaient : la fonction de coût fonction de perte focalisée –ou focal loss function– (FL) [97], la technique de pondération par la fréquence effective des classes [38] et la stratégie d'augmentation synthétique RandAugment [37]. Leur méthode a atteint une Exactitude équilibrée –ou Balanced accuracy– (BAcc) de 0,86 sur le jeu de données ISIC-2018.

Une synthèse des méthodes de construction de SAADs basés sur un CNN est pré-

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

sentée dans le Tableau 1.1.

TABLEAU 1.1. – Synthèse des méthodes développant un système automatique d'aide au diagnostic du mélanome basées sur un seul CNN.

Travaux	Contribution de l'étude	Classification	Performance
Esteva <i>et al.</i> [53]	Comparer la performance d'un CNN avec celles des dermatologues.	Multi-classes	AUC : 0.94
Lucius <i>et al.</i> [101]	Évaluer si les CNNs peuvent aider les médecins non-dermatologues à améliorer leur précision dans le diagnostic de sept lésions cutanées courantes.	Multi-classes	Acc : 83.88%
Lopez <i>et al.</i> [139]	Résoudre le problème de classification binaire, à partir d'images dermoscopiques, d'un mélanome et d'une lésion bénigne.	Binaire	Sens : 78.66%
Menegola <i>et al.</i> [112]	Étudier en profondeur l'impact de l'apprentissage par transfert pour la détection du mélanome.	Binaire	Acc : 84.5%
Kawahara <i>et al.</i> [90]	Améliorer les performances d'un CNN grâce aux informations additionnelles apportées par l'interactions de différentes résolutions d'une même image.	Multi-classes	Acc : 79.5%
Zhang <i>et al.</i> [176]	Améliorer la précision de classification des lésions cutanées en intégrant un module d'attention dans un CNN.	Multi-classes	-
Yao <i>et al.</i> [168]	Améliorer les performances d'un CNN en combinant plusieurs techniques d'apprentissage profond.	Multi-classes	BAcc : 85.8%

1.3.2. Systèmes automatiques d'aide au diagnostic du mélanome basés sur une méthode ensembliste de CNNs

Une autre technique efficace de construction de SAADs pour la détection du mélanome consiste à assembler un ensemble fini de CNNs. Harangi *et al.* [78], par exemple, ont fusionné les sorties de quatre CNNs en appliquant une stratégie de fusion pondérée ceci pour classifier trois types de lésions cutanées. Leur méthode a obtenu une Aire sous la courbe caractéristique d'exploitation du récepteur –ou *Area under the receiver operating characteristic curve*– (AUC) de 0,89, ce qui est supérieur aux performances de chaque CNN pris individuellement. Khatib *et al.* [92] ont également proposé un SAAD pour la détection de mélanome qui effectue une fusion des décisions de quatre CNNs. Dans le même ordre d'idée, Raza *et al.* [136] ont proposé une méthode ensembliste pour la classification des mélanomes combinant différents modèles pré-entraînés. Kausar *et al.* [89] ont utilisé le vote majoritaire et le vote majoritaire pondéré comme méthode d'agrégation pour assembler des modèles pour la classification des lésions cutanées. Yap *et al.* [169] ont proposé une méthode qui prend en compte plusieurs modalités d'image, y compris les méta-données du patient, pour améliorer les résultats de la classification. Le réseau CNN qu'ils ont utilisé a été appliqué différemment sur des images dermoscopiques et macroscopiques, et leurs caractéristiques ont été fusionnées pour effectuer la classification finale. Leur modèle multimodale a surpassé le modèle de base entraîné uniquement sur des images macroscopiques atteignant une AUC de 0,87. Gessert *et al.* [67] ont quant à eux assemblé plusieurs CNNs pré-entraînés pour effectuer une classification multi-classes des lésions cutanées. Ils ont tout d'abord entraîné des modèles sur des images de différentes résolutions obtenues par plusieurs stratégies de rognage. Ensuite, ils ont créé un grand ensemble comprenant des sous-ensembles optimaux de tous les modèles entraînés en se basant sur leurs performances sur la base de validation. Gong *et al.* [70] ont proposé et évalué trois stratégies d'assemblage de CNNs via des systèmes de vote pouvant améliorer la performance de classification des images dermoscopiques. Ichim *et al.* [85] ont proposé un système basé sur une décision prise à partir de six réseaux de neurones connectés à deux niveaux opérationnels. Le premier niveau contient cinq CNNs, tandis que dans le deuxième niveau, il y a un perceptron qui décide si la décision finale est un mélanome ou non.

Une synthèse des méthodes ensembliste de CNNs pour la construction de SAADs est présentée dans le Tableau 1.2.

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

TABLEAU 1.2. – Synthèse des méthodes utilisant un système automatique d'aide au diagnostic du Mélanome basées sur un assemblage de plusieurs CNNs.

Travaux	Contribution de l'étude	Classification	Performance
Harangi <i>et al.</i> [78]	Améliorer la précision individuelle des CNNs grâce à une méthode d'assemblage.	Multi-classes	AUC : 0.891
El-khatib <i>et al.</i> [92]	Améliorer la précision de détection des lésions cutanées ayant l'apparence d'un mélanome en combinant quatre CNNs.	Multi-classes	Acc : 88.33%
Raza <i>et al.</i> [136]	Améliorer la précision mélanome acral lentigineux grâce à une méthode d'assemblage.	Binaire	Acc : 97.93%
Kausar <i>et al.</i> [89]	Améliorer la précision de détection des lésions cutanées à partir d'un ensemble de modèles hétérogènes plus performants en utilisant le vote majoritaire et le vote majoritaire pondéré comme méthode de fusion.	Multi-classes	Acc : 92.0%
Yap <i>et al.</i> [169]	Améliorer les performances du diagnostic automatisé des lésions cutanées en combinant plusieurs modalités d'imagerie avec les métadonnées du patient.	Multi-classes	Acc : 72.9%
Gessert <i>et al.</i> [67]	Améliorer les performances du diagnostic automatisé des lésions cutanées en créant un grand ensemble à partir de différents modèles entraînés sur différentes résolutions d'image.	Multi-classes	BAcc : 74.2%
Gong <i>et al.</i> [70]	Évaluer trois stratégies d'assemblage de CNNs via des systèmes de vote pouvant améliorer la performance de classification des images dermoscopiques.	Multi-classes	AUC : 0.919%
Ichim <i>et al.</i> [85]	Améliorer les performances du diagnostic automatisé des lésions cutanées en combinant six réseaux de neurones connectés à deux niveaux opérationnels.	Binaire	Acc : 97.5%

1.3.3. Systèmes automatiques d'aide au diagnostic du Mélanome basés sur des CNNs combinés à des techniques traditionnelles d'apprentissage automatique

Certaines études conçoivent des SAADs en combinant des CNNs avec d'autres classifieurs. Dans ce contexte, Mahbod *et al.* [103] ont proposé une méthode qui combine les caractéristiques profondes extraites de différents CNNs pré-entraînés, à différents niveaux d'abstraction, avec un classifieur Machine à vecteurs de support –ou *Support vector machine*– (SVM) pour distinguer les mélanomes des lésions bénignes. De même, Hagerty *et al.* [75] ont présenté une approche qui combine des caractéristiques obtenues à partir de techniques conventionnelles de traitement d'images et des caractéristiques profondes extraites d'un CNN. En fusionnant les différentes caractéristiques, leur méthode a contribué à une amélioration de l'AUC de 7% par rapport au modèle CNN seul. Similairement, Almaraz-Damien *et al.* [9] ont proposé un SAAD basé sur la fusion de caractéristiques liées à la règle ABCDE et des caractéristiques profondes. Les caractéristiques profondes utilisées pour la fusion ont été obtenues par apprentissage par transfert sur des CNNs pré-entraînés. Yildirim-Yayilgan *et al.* [171] ont proposé un SAAD basé sur une combinaison de CNNs pré-entraînés et des caractéristiques structurelles ce qui leur a permis d'atteindre des gains de performance de 1 à 4%. Sekhar *et al.* [142] ont proposé un SAAD qui combine des caractéristiques de texture avec les CNNs. Les caractéristiques de texture sont extraites des lésions cutanées en appliquant une transformée en ondelettes dispersées et les caractéristiques ainsi extraites sont données comme informations supplémentaires à la couche entièrement connectée d'un CNN. Les auteurs ont utilisé des images dermoscopiques provenant du jeu de données ISIC-2016 pour entraîner et tester leur système et ont fait le constat que leur méthode permettait d'améliorer la précision de détection des mélanomes et des autres lésions cutanées atteignant une précision de classification de 98,13% pour l'identification du mélanome. Majtner *et al.* [104] ont proposé un SAAD pour le diagnostic de mélanomes basé sur des méthodes d'apprentissage profond combinées à des caractéristiques RSurf et des motifs binaires locaux. Les auteurs ont évalué leur approche sur le jeu de données ISIC-2016 et ont obtenu une précision, une sensibilité et une spécificité de classification élevées par rapport à d'autres classifieurs sur le même ensemble de données.

Une synthèse des méthodes de construction de SAADs basés sur des CNNs combinés à des techniques traditionnelles d'apprentissage automatique est présentée dans le Tableau 1.3.

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

TABLEAU 1.3. – Synthèse des méthodes utilisant un système automatique d'aide au diagnostic du Mélanome basées sur une combinaison de CNNs avec des techniques traditionnelles d'apprentissage automatique.

Travaux	Contribution de l'étude	Classification	Performance
Mahbod <i>et al.</i> [103]	Améliorer la détection des lésions cutanées en entraînant un SVM sur des caractéristiques extraites des CNNs pré-entraînés.	Binaire	AUC : 0.873
Hagerty <i>et al.</i> [75]	Améliorer la détection des lésions cutanées en combinant des caractéristiques de techniques conventionnelles de traitement d'images avec des caractéristiques extraites d'un CNN pré-entraîné.	Multi-classes	AUC : 0.94
Almaraz-Damien <i>et al.</i> [9]	Améliorer la détection des lésions cutanées en combinant des caractéristiques issues de la règle ABCDE avec des caractéristiques extraites d'un CNN pré-entraîné.	Binaire	Acc : 92.4%
Yildirim-Yayilgan <i>et al.</i> [171]	Améliorer la détection des lésions cutanées en entraînant un SVM sur des caractéristiques extraites des CNNs pré-entraînés combiné à des caractéristiques structurelles.	Binaire	AUC : 0.863
Sekhar <i>et al.</i> [142]	Améliorer la détection des lésions cutanées en ajoutant des caractéristiques de texture à la couche complètement connectée d'un CNN pré-entraîné.	Binaire	Acc : 98.1%
Majtner <i>et al.</i> [104]	Améliorer la détection des lésions cutanées en combinant deux SVMs. Un entraîné sur des caractéristiques extraite d'un CNN pré-entraîné et l'autre entraîné sur des caractéristiques extraites à l'aide des algorithmes RSurf et LBP.	Binaire	Acc : 80.5%

1.3.4. Défis clés du développement des systèmes automatiques d'aide au diagnostic des cancers de la peau

L'apprentissage profond a contribué à faire évoluer considérablement la recherche sur le développement des SAADs pour le diagnostic précoce des mélanomes. Certaines des méthodes existantes présentées précédemment atteignent des performances comparables voire surclassant parfois des dermatologues dans des contextes expérimentaux. Malgré ces résultats prometteurs obtenus, cela est loin d'être suffisant. De nombreuses questions nuancées subsistent face à l'application clinique de l'apprentissage profond aux soins de santé en général et au diagnostic des cancers de la peau en particulier [114].

Dans cette sous-section, nous explorerons de façon succincte les défis qui freinent l'utilisation des SAADs en contexte clinique réel. Ces défis recensés, découlant de notre étude bibliographique, peuvent être regroupés en deux catégories : les défis techniques et les défis non-techniques. La figure 1.4 illustre à travers une carte mémoire les différents défis recensés. Nous recommandons au lecteur les articles [71, 4, 173, 30, 133] pour une analyse plus détaillée des différents défis.

1.3.4.1. Les défis techniques

Nous considérons comme une question technique, tout ce qui, en termes de ressources et de méthodes, entrave la conception de SAADs. Elles peuvent être subdivisées en deux catégories : les questions dermatologiques ou algorithmiques.

- **Problèmes dermatologiques :** La qualité des données est une grosse problématique en dermatologie. La plupart des ensembles de données disponibles pour les images de lésions cutanées contiennent des artefacts visibles : encre de marqueur, marques de règle, poils, bulles de gel, halo noir périphérique, variation d'illumination, etc. De plus, le nombre d'images, les types de maladies, la taille des images, et les méthodes de prise de vue et de traitement de ces données diffèrent considérablement d'une base à une autre, ce qui rend laborieux l'analyse des différentes études faites sur le sujet et complique la possibilité de décrire quantitativement les différents modèles [4, 21]. En outre, il est difficile de collecter des images de certaines maladies rares. La plupart des bases de données ne contiennent que 2 à 14 types de lésions cutanées sur les plus de 3000 identifiées dans la littérature [96]. A cela, il faudrait ajouter que les images présentes dans ces bases de données ne sont pas représentatives de tous les types de peau, cultures, géographies et de la distribution des maladies [3, 123]. Une autre difficulté vient du fait qu'il y ait très peu de consensus chez les experts sur les critères cliniques de diagnostic des lésions cutanées, ce qui conduit à un diagnostic très subjectif [26]. Enfin, nous pouvons ajouter, malgré les travaux de Finnane *et al.* [58], que l'absence de protocole de standardisation des bases de données reste un réel obstacle à l'évolution des SAADs.

- **Problèmes algorithmiques :** L'implémentation d'un algorithme profond pour les modèles de classification automatique nécessite une énorme quantité de données. Bien que des efforts aient été faits dans le cadre du concours ISIC [31] pour construire des bases de données de plus en plus cohérentes pour les lésions cutanées, la distribution asymétrique des classes de ces jeux de données reste un problème important lors de l'entraînement d'un CNN [110]. Ceci a pour conséquence de rendre difficile l'apprentissage de modèles robustes au déséquilibre des classes. En outre, le progrès des CNNs dans la détection des cancers de la peau dépend d'une modélisation non linéaire de la problématique et d'une définition optimale mais laborieuse des hyperparamètres des modèles. Ces derniers éléments font assimiler les CNNs à des boîtes noires –ou *black box*– et rend leur processus de décision interne difficile à comprendre. L'amélioration de l'explicabilité des SAADs basés sur les CNNs pourrait permettre aux utilisateurs finaux (Dermatologues, Médecins, Patients) de ces systèmes de connaître clairement le comportement et les limites de ceux-ci, et de garantir la fiabilité et la sécurité des diagnostics.

1.3.4.2. Les défis non-techniques

Un défi certes non-technique, mais non moins important est la perception qu'a un patient de sa santé étudiée par une intelligence artificielle [39]. De même, en raison d'un effectif insuffisant de dermatologues et de ressources financières limitées, il est difficile de procéder à un large dépistage des lésions cutanées. Ceci conduit donc à ce que la majorité des lésions cutanées bénignes avec peu de signes de malignité ne soient pas vérifiées par histologie [35]. L'utilisation de leur diagnostic comme étiquette pour entraîner des modèles peut donc naturellement accroître le scepticisme des dermatologues à l'égard de l'utilisation des SAADs[30]. En outre, les normes et la réglementation sur les algorithmes d'aide à la décision [133] sont aujourd'hui une grande variable à prendre en compte dans la mise en œuvre de tels systèmes.

Les travaux issus de cette thèse s'attendent à traiter un certain nombre de problématiques présentées ci-dessus. Nous avons pour cela successivement mené les travaux suivants :

1. Mise en place d'un protocole expérimental et étude des images de lésions cutanées présentant un halo périphérique afin de mettre en lumière l'impact de cet artefact sur la performance des CNNs. Ce travail est décrit dans le Chapitre 5.
2. Définition d'une nouvelle méthode d'assemblage des CNNs capable d'améliorer les performances de diagnostic du mélanome comparées aux approches précédentes. Ce travail est présenté dans le Chapitre 2.
3. Proposition d'une méthode permettant de développer des SAADs plus robustes en présence de jeux de données déséquilibrés. Ce travail est présenté dans le Chapitre 3.

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

4. Étude du problème de l'explicabilité des [SAADs](#) basés sur l'apprentissage profond et proposition d' une nouvelle architecture des [CNNs](#) capable non seulement d'améliorer les performances de détection, mais aussi d'apporter plus de clarté dans le processus de décision. Ce travail est décrit dans le [Chapitre 4](#).

1. Le diagnostic du mélanome – 1.3. L'apprentissage profond pour la détection du mélanome

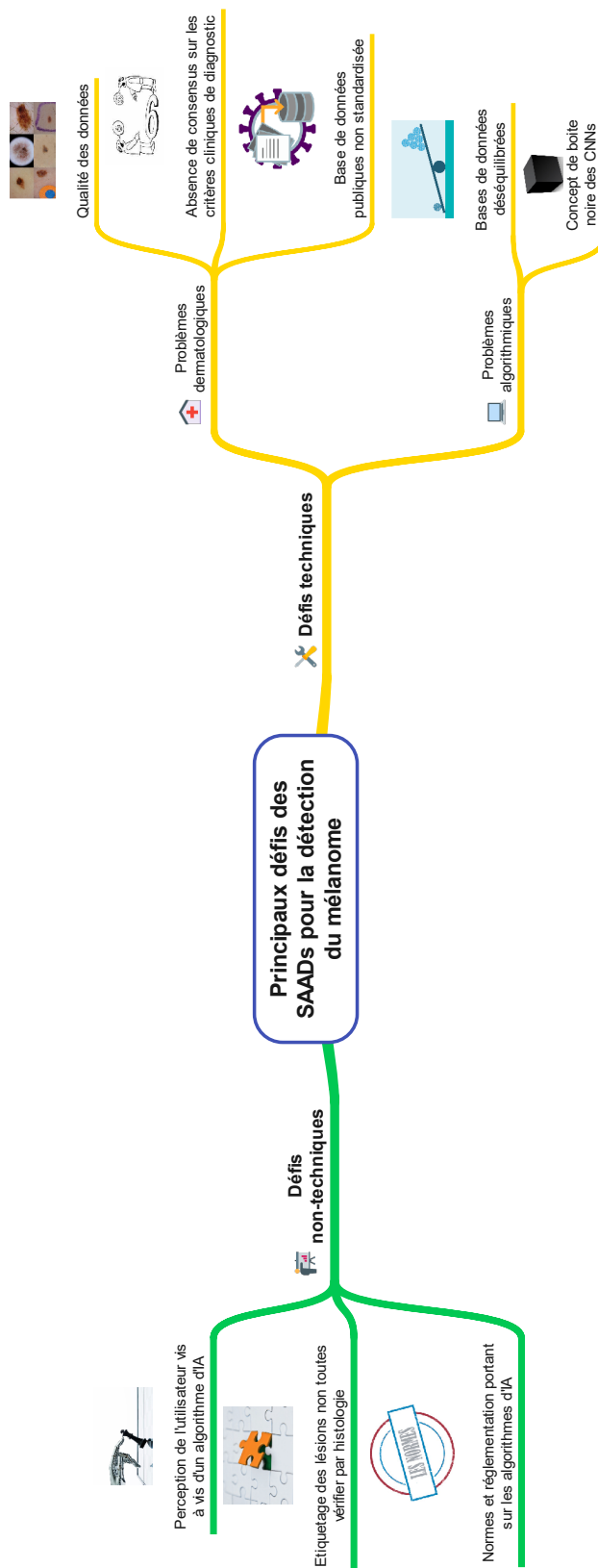


FIGURE 1.4. – Illustration des principaux défis qui freinent l'utilisation des SAADs en contexte clinique réel.

1.3.5. Métriques de performances couramment utilisées pour l'évaluation des systèmes automatiques d'aide au diagnostic

Plusieurs mesures sont utilisées dans la littérature pour évaluer la performance des SAADs. Nous présentons dans cette sous-section les plus utilisées dans des applications de détection de mélanome.

Tout d'abord, nous allons définir quatre valeurs prédictives clés pour une tâche de prédiction. Considérons pour simplification une tâche de classification binaire d'un patient malade ou sain. Les quatre valeurs prédictives clés d'une tâche de prédiction sont définies de la manière suivante :

- Vrai positif (VP) : il s'agit de l'ensemble des échantillons de patients malades que le modèle a prédit comme étant des échantillons appartenant à l'étiquette malade.
- Vrai négatif (VN) : il s'agit de l'ensemble des échantillons de patients sains que le modèle a prédit comme étant des échantillons appartenant à l'étiquette sain.
- Faux positif (FP) : il s'agit de l'ensemble des échantillons de patients sains pour lesquels le modèle a effectué une prédiction contraire. Le patient est sain, mais le modèle prédit qu'il ne l'est pas.
- Faux négatif (FN) : il s'agit de l'ensemble des échantillons de patients malades pour lesquels le modèle a effectué une prédiction contraire. Le patient est malade, mais le modèle prédit qu'il ne l'est pas.

Nous pouvons à présent définir les mesures¹ ci-dessous régulièrement utilisées dans la littérature relative à la détection de mélanome :

- La sensibilité ou taux de vrais positifs : elle évalue la capacité du modèle à identifier correctement les patients malades. Elle est calculée à l'aide de l'équation suivante :

$$Sensibilite = \frac{VP}{VP + FN} \quad (1.1)$$

- La spécificité ou taux de vrais négatifs : elle évalue la capacité du modèle à identifier correctement les patients sains. Elle est calculée à l'aide de l'équation suivante :

$$Specificite = \frac{VN}{VN + FP} \quad (1.2)$$

- L'exactitude -ou *accuracy*- : elle mesure le pourcentage de bonnes prédictions. Elle permet d'évaluer la capacité globale du modèle sur toutes les étiquettes. Elle est calculée à l'aide de l'équation suivante :

$$Exactitude = \frac{VP + VN}{VP + VN + FP + FN} \quad (1.3)$$

- La précision -ou *precision*- : elle mesure le taux de prédictions correctes parmi

1. Les formules des métriques présentées sont celles appliquées pour une tâche de classification binaire. Elles peuvent toutes être étendues à une tâche de classification multi-classes.

1. Le diagnostic du mélanome – 1.4. Conclusion

les prédictions positives. Similairement à la sensibilité, elle évalue la capacité du modèle à performer sur des échantillons de patients malades. Elle est calculée à l'aide de l'équation suivante :

$$Precision = \frac{VP}{VP + FP} \quad (1.4)$$

La plupart des bases de données du monde réel, et plus encore spécifiquement ceux touchant au domaine médical comme les lésions cutanées présentent une distribution asymétrique. Dans ce contexte, l'évaluation de la performance d'un modèle à l'aide de l'exactitude se retrouverait biaisée par la performance de la classe surreprésentée. Une métrique alternative et plus robuste dans ce cas de figure est le score de **BAcc**. **BAcc** est calculée à l'aide de l'équation suivante :

$$BAcc = \frac{Sensibilite + Specificite}{2} \quad (1.5)$$

Il existe généralement un compromis à effectuer entre la sensibilité et la spécificité. Par exemple, un dermatologue peut demander l'excision des naevus suspects à faible risque (faible spécificité), afin de réduire le risque de manquer des mélanomes (sensibilité élevée). Ce compromis peut être représenté graphiquement par une **Courbe caractéristique d'exploitation du récepteur –ou Receiver Operating Characteristic– (ROC)** [162]. La courbe **ROC** est un bon moyen de visualiser les performances d'un modèle afin de sélectionner le point de fonctionnement ou le seuil de décision approprié. L'**AUC** est une autre métrique souvent utilisée pour la comparaison des modèles. L'**AUC** est égale à la probabilité qu'un classifieur attribue un score élevé à un exemple positif choisi au hasard par rapport à un exemple négatif choisi au hasard. Il reflète le niveau de séparabilité entre les classes.

1.4. Conclusion

Dans ce chapitre, nous avons commencé par introduire le contexte autour du cancer du mélanome, nous avons par la suite fait un état de l'art des différentes approches proposées pour la construction des **SAADs** pour la détection du mélanome.

Dans la première partie, nous avons présenté les bases de la biologie de la peau humaine. Nous avons décrit les différents types de cancers et présenté les facteurs de risques et l'incidence du mélanome. Les méthodes courantes de diagnostic clinique et de prise en charge sont brièvement passées en revue. Cette partie a non seulement mis en avant le défi associé au bon diagnostic du mélanome, mais aussi le fait que le mélanome représente un enjeu majeur de santé publique. Ceci nous a permis de mettre en lumière la nécessité de développer des **SAADs** pouvant suppléer les dermatologues durant leur examen clinique.

Dans la deuxième partie, nous avons effectué une revue de la littérature des **SAADs** pour la détection du mélanome basés sur des techniques d'apprentissage profond.

1. Le diagnostic du mélanome – 1.4. Conclusion

Les approches existantes peuvent être regroupées en trois types : celles basées sur un seul CNN, celles qui combinent plusieurs CNNs par une méthode de fusion, et celles qui combinent des CNNs avec d'autres techniques traditionnelles d'apprentissage automatique. Certaines des approches existantes atteignent des performances comparables voire surclassant parfois celles des dermatologues dans des contextes expérimentaux. Malgré ces résultats prometteurs obtenus, cela est loin d'être suffisant car un certain nombre de défis demeurent. Nous avons également dans cette partie exposé une synthèse des défis clés entravant le développement des SAADs pour le diagnostic des cancers de la peau. Dans cette thèse, nous nous sommes tout particulièrement intéressés à traiter un certain nombre de ces problématiques notamment celle touchant à la qualité de la base de données à travers l'étude de l'halo périphérique, celle portant sur le problème de déséquilibre des classes et enfin celle concernant l'explicabilité des SAADs.

Le chapitre suivant présente une contribution de cette thèse qui se place dans la recherche d'une méthode pouvant améliorer la performance des SAADs existants pour la détection du mélanome.

2. Construction d'un système de diagnostic automatique du mélanome

Sommaire

2.1. Introduction	47
2.2. État de l'art des méthodes ensemblistes de réseaux profonds	49
2.2.1. Apprentissage ensembliste de CNNs par bagging	49
2.2.2. Apprentissage ensembliste de CNNs par boosting	50
2.2.3. Apprentissage ensembliste de CNNs par stacking	50
2.2.4. Apprentissage ensembliste de CNNs appliqué à la détection du mélanome	51
2.3. Méthodes proposées	52
2.3.1. Base de données	53
2.3.1.1. Préparation de la base de données	53
2.3.1.2. Prétraitement des images	54
2.3.2. Méthodologie	56
2.3.2.1. Les architectures des réseaux de neurones convolutifs utilisées	56
2.3.2.2. Agrégation de CNNs avec un graphe acyclique dirigé (DAG-CNN)	58
2.3.2.3. Définition des fonctions d'agrégation courantes	60
2.3.3. Entraînement des modèles	61
2.3.4. Définition des critères de performance utilisés	62
2.4. Résultats et discussions	63
2.4.1. Performance des modèles binaires	63
2.4.2. Impact du choix du nœud racine	64
2.4.3. Comparaison de l'approche DAG-CNN avec des CNNs multi-classes standards	65
2.4.4. Analyse des performances de la méthode DAG-CNN	66
2.4.5. Comparaison avec d'autres approches de fusion	76
2.5. Conclusion	77

2.1. Introduction

La détection précoce du mélanome est le moyen le plus efficace pour réduire son taux de mortalité. Les dermatologues accomplissent cette tâche avec l'aide d'un dermatoscope. Des SAADs développés sur des images dermoscopiques sont fortement souhaitables afin d' assister les dermatologues dans cette tâche.

L'approche conduisant à de très bonnes performances et actuellement utilisée par la communauté de vision par ordinateur afin de développer des SAADs sur des données médicales, repose sur des méthodes ensemblistes de CNNs [120]. Elle consiste en général à combiner plusieurs CNNs multi-classes afin d'obtenir des SAADs ayant les meilleures performances de prédiction. Cependant, la classification multi-classes des lésions cutanées demeure une tâche difficile d'une part parce que les lésions cutanées présentent un degré élevé de similitude. En effet, les mélanomes typiques (Figure 2.1 (a)) plus fréquents sont généralement plus facilement distinguables par rapport à d'autres lésions typiques de naevus (Figure 2.1 (b)) et kératose (Figure 2.1 (c)) et posent moins de problème de diagnostic aux dermatologues. A l'inverse, les naevus atypiques (Figure 2.1 (e)) ou kératoses atypiques (Figure 2.1 (f)) peuvent être confondues avec un mélanome atypique (Figure 2.1 (d)) car elles partagent plusieurs caractéristiques et ont donc un degré de similitude plus élevé. Une autre cause faisant de la classification multi-classes des lésions cutanées une tâche complexe provient du fait que les CNNs nécessitent une grande quantité de données étiquetées et une définition minutieuse des hyper-paramètres pour être performants. Décomposer un problème multi-classes en un problème binaire pourrait donc réduire cette complexité du problème initial et augmenter les performances globales des SAADs basés sur des modèles entraînés suivant cette hypothèse.

Dans ce chapitre, nous présenterons notre pipeline ensembliste de CNN ayant pour objectif d'améliorer la performance des SAADs existants pour la détection précoce du mélanome. Contrairement aux méthodes ensemblistes conventionnelles, la méthode proposée s'inspire des méthodes de décomposition et d'ensemble et utilise pour ce fait un *Graphe acyclique dirigé* –ou *directed acyclic graph*– (DAG) pour agréger des CNNs entraînés pour des tâches de classification binaire. Les principales contributions de ce travail sont :

- La décomposition de la tâche de classification multi-classes initiale des lésions cutanées en un problème binaire afin de réduire la complexité de la tâche et d'augmenter la performance globale de la classification ;
- L'utilisation d'un DAG comme méthode d'agrégation pour construire un SAAD basé sur une méthode ensembliste de CNNs.

Ce chapitre sera organisé comme suit : la Section 2.2 présentera une revue de littérature des méthodes d'apprentissage profond, dans la Section 2.3 la méthode proposée ainsi que le détail du protocole expérimental effectué seront présentés, nos résultats seront discutés dans la Section 2.4 et la Section 2.5 conclura le chapitre.

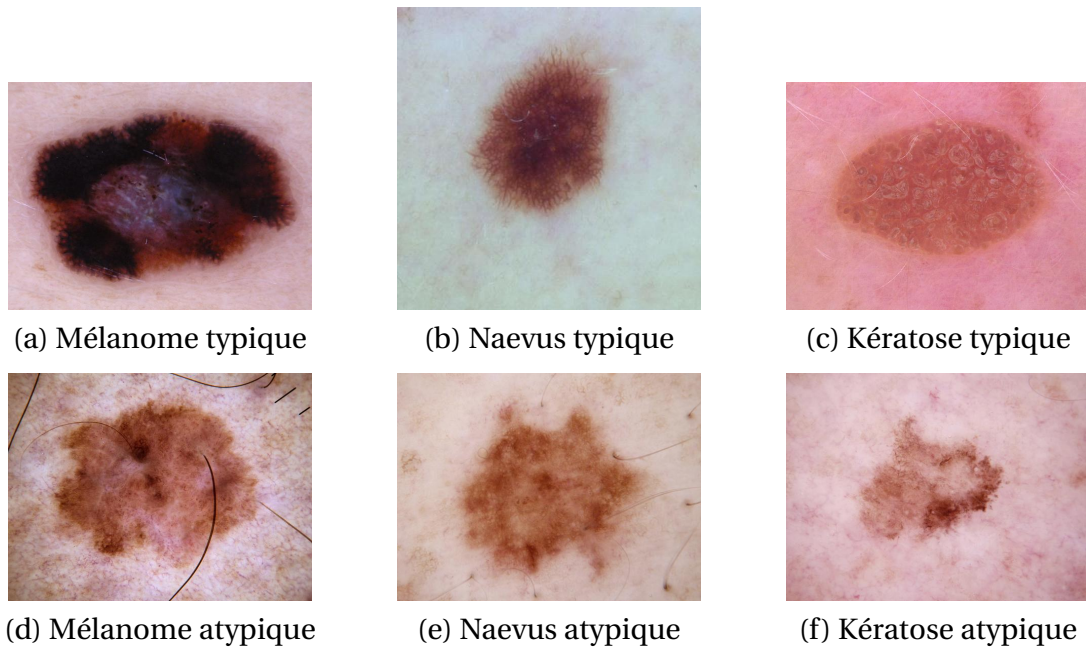


FIGURE 2.1. – Illustration de lésions cutanées tirées du jeu de données public ISIC [31, 157]. Les lésions typiques (a) mélanome, (b) naevus et (c) kératose ne posent pas de problème de diagnostic pour les dermatologues, alors que les lésions atypiques (d) mélanome, (e) naevus et (f) kératose sont beaucoup plus difficiles à différencier.

2.2. État de l'art des méthodes ensemblistes de réseaux profonds

Des études récentes ont montré que les SAADs les plus performants et les plus précis pour la classification d'images médicales sont fortement basés sur des méthodes ensemblistes de réseaux de neurones [120, 99, 117, 167].

Les méthodes ensemblistes (ou technique d'assemblage) sont des techniques d'apprentissage automatique qui combine plusieurs modèles de base afin de produire un modèle prédictif optimal ayant des performances meilleures que n'importe lequel des modèles de base pris individuellement [174]. Il existe trois grandes méthodes ensemblistes : la méthode d'assemblage parallèle (ou *bagging* ou *bootstrap*) [23], la méthode d'assemblage séquentielle (ou *boosting*) [62] et la méthode d'assemblage par empilage (*stacking*) [165].

Dans le domaine de l'apprentissage automatique, l'objectif des méthodes ensemblistes est de trouver une hypothèse appropriée qui maximise le score de précision de la prédiction. Cependant, il est très souvent difficile de trouver l'hypothèse optimale, c'est pourquoi la stratégie a évolué en la combinaison de plusieurs hypothèses en un classifieur supérieur plus proche d'une hypothèse optimale [120]. Dans le contexte de l'apprentissage profond, les hypothèses sont représentées par des modèles de CNNs adaptés. Ainsi, l'apprentissage ensembliste de réseaux profonds est défini comme la combinaison de plusieurs CNNs pour obtenir de meilleures performances de prédiction.

Dans cette section, nous allons passer en revue les méthodes ensemblistes de réseaux profonds développées pour des tâches de classification d'images.

2.2.1. Apprentissage ensembliste de CNNs par bagging

L'apprentissage ensembliste par *bagging* consiste à sous-échantillonner les données, en créant de nouvelles bases de données (dénommées ensembles de données *bootstrap*) pour entraîner chaque modèle. Pour déterminer le résultat de classification final, on procède par un vote des résultats de chaque modèle par une méthode d'agrégation (ou fusion). Les travaux de Hameed *et al.* [76] s'inscrivent dans ce contexte. Les auteurs de ce travail ont proposé une approche ensembliste de CNNs pour la classification d'images histopathologiques de cancer du sein de type non-carcinome et carcinome en utilisant un ensemble de données collectées à titre privé. Ils ont pour cela entraîné quatre CNNs pré-entraînés et fusionné les prédictions des modèles de base en prenant la moyenne des probabilités prédites. Similairement, Chen *et al.* [29] ont entraîné plusieurs CNNs et agrégé leur résultat à l'aide d'un vote majoritaire afin de construire un SAAD pour la classification d'images hyper-spectrales. Plus récemment, Yang *et al.* [167] ont proposé une méthode ensembliste en deux étapes dénommée formation

2. Construction d'un système de diagnostic automatique du mélanome – 2.2. État de l'art des méthodes ensemblistes de réseaux profonds

d'arbre profond – ou Deep Training Tree -. Leur approche consiste à former conjointement une série de modèles construits à partir des couches cachées d'un CNN de manière hiérarchique. Cette approche présente l'avantage de limiter le problème de saturation de gradients pour les couches cachées, et permet d'obtenir intrinsèquement des modèles de base sur des caractéristiques de niveau intermédiaire avec un coût de calcul réduit pour une solution ensembliste. Les meilleurs modèles de base sélectionnés sont finalement fusionnés à l'aide d'une méthode de fusion pondérée.

2.2.2. Apprentissage ensembliste de CNNs par boosting

L'apprentissage ensembliste par *boosting* consiste à construire progressivement un ensemble en formant chaque nouvelle instance de modèle pour mettre l'accent sur des données d'entraînement que les modèles précédents ont mal classées. Moghimi *et al.* [115] ont proposé un nouvel algorithme intégrant le *boosting* avec les CNNs (BoostCNN) basé sur la fonction de coût des moindres carrés. Les résultats qu'ils ont obtenus montrent que la méthode BoostCNN est capable d'atteindre des performances comparables à l'état de l'art sur plusieurs tâches de classification telles que la classification des oiseaux, des voitures et des avions. De même, Han *et al.* [77] ont proposé une approche combinant un CNN avec une méthode de *boosting* incrémentiel (IB-CNN). Leurs résultats expérimentaux ont démontré que leur approche apporte une amélioration significative par rapport au modèle de CNN de base pour la reconnaissance d'unités d'actions faciales. Dans certains cas, il a été démontré que les méthodes ensemblistes par *boosting* donnent une meilleure précision que celles par *bagging*. Cependant, l'un des principaux inconvénients de l'approche par *boosting* est qu'elle a tendance à sur-apprendre sur le jeu de données d'entraînement.

2.2.3. Apprentissage ensembliste de CNNs par stacking

L'apprentissage ensembliste de CNNs par *stacking* consiste à entraîner différents modèles de base sur les mêmes données et à utiliser un autre modèle pour apprendre comment combiner au mieux les prédictions faites par l'ensemble des modèles de base. Les travaux de Fang *et al.* [55] s'inscrivent dans ce contexte. Les auteurs ont proposé un nouveau SAAD qui regroupe trois CNNs pour la classification de la maladie d'Alzheimer. Pour cela, ils ont entraîné chacun des CNNs sur différentes modalités (Imagerie à résonance magnétique et tomographie par émission de positrons) d'images de sujets. Les vecteurs de probabilités de prédictions générés par les modèles de base sont par la suite utilisés pour l'entraînement d'un classifieur Adaboost afin de générer les prédictions finales. Une méthode ensembliste par *stacking* est également employée par Yazdizadeh *et al.* [170]. Leur approche est composée d'un ensemble de CNNs avec différentes valeurs d'hyperparamètres où les prédictions ont été agrégées à l'aide d'un classifieur de forêt aléatoire –ou *random forest classifier*–. Appliquée à une tâche de classification du mode de transport des données de voyage, leur approche a atteint un score d'exactitude de 91,8%. Muller *et al.* [120] ont analysé l'impact des différentes méthodes ensemblistes présentées dans les sous-sections précédentes

(Sous-section 2.2.2 et Sous-section 2.2.1) avec l'apprentissage ensembliste par *stacking* sur différents types d'images médicales. Leurs résultats ont révélé que l'apprentissage ensembliste par *stacking* a permis d'obtenir les meilleurs gains de performance.

2.2.4. Apprentissage ensembliste de CNNs appliqué à la détection du mélanome

Dans le contexte du diagnostic multi-classes des lésions cutanées en général et du mélanome en particulier, plusieurs méthodes ensemblistes de CNNs ont été proposées. De manière générale, tel que décrit dans la Sous-section 1.3.2, la plupart des études actuelles appliquant des méthodes ensemblistes de CNNs pour cette tâche utilisent la méthode ensembliste par *bagging*. Tout d'abord, plusieurs CNNs multi-classes sont entraînés pour la tâche de classification des lésions cutanées puis, leurs résultats sont fusionnés à l'aide d'une approche d'agrégation. Les méthodes d'agrégation les plus utilisées sont les suivantes :

- La stratégie du vote majoritaire ou **Max-Win** : La classe sélectionnée par la stratégie Max-Win est la classe qui reçoit le nombre maximum de prédiction de la part des modèles de base de l'ensemble.
- La stratégie du produit des probabilités (**Prod**) : Le produit des probabilités prédites par chaque CNN de base est calculé et la classe sélectionnée est déterminée par le maximum des produits normalisés.
- La stratégie des moyennes des probabilités (**Avg**) : La moyenne arithmétique des probabilités prédites par chaque CNN est calculée et la classe sélectionnée est déterminée par le maximum des moyennes normalisées.
- La stratégie de confiance maximale (**Max**) : La classe prédite par la stratégie de confiance maximale est la classe qui a reçu le score de prédiction maximum.
- Stratégie de la moyenne géométrique (**Gmean**) : La moyenne géométrique des probabilités prédites par chaque CNN est calculée et la classe sélectionnée est déterminée par le maximum des moyennes normalisées.

Plutôt que de simplement agréger plusieurs CNNs multi-classes afin de construire un ensemble profond robuste, une approche innovante consisterait à décomposer le problème multi-classes initial en plusieurs tâches de classification moins complexes. C'est ce que révèle les travaux de Galar *et al.* [63]. Leurs auteurs démontrent que la frontière de décision d'un problème de classification multi-classes tend à être plus complexe qu'un problème de classification binaire et concluent qu'une tâche de classification multi-classes est généralement plus difficile qu'une tâche de classification binaire. Ces travaux précurseurs ont conduit des chercheurs à étudier les méthodes de décomposition et d'ensemble comme une alternative pour construire des classifieurs plus performants pour des tâches de classification multi-classes. L'idée derrière la méthode de décomposition et d'ensemble est de diviser le problème multi-classes en un ensemble de problèmes binaires, puis d'agréger les résultats. Les deux approches bien connues pour développer une stratégie de décomposition et d'ensemble sont les méthodes un contre tous –ou *one versus all*– (o-v-a) et un contre un –ou *one versus one*– (o-v-o) [63]. Pour une classification à N classes, chaque approche est décrite

comme suit :

- L'approche o-v-a consiste à construire un ensemble de N classifieurs binaires. Chaque classifieur est entraîné avec une classe pour laquelle une étiquette positive a été attribuée et toutes les autres classes ayant une étiquette négative. La décision finale correspond à la classe associée au classifieur ayant le score de prédiction le plus élevé.
- L'approche o-v-o consiste à construire tous les classifieurs binaires possibles à partir des N classes. Chaque classifieur est entraîné sur seulement deux classes parmi les N classes initiales. Ainsi, pour une tâche de classification multi-classes comprenant N classes, un total de $N(N - 1)/2$ classifieurs seront entraînés. Les prédictions de ces classifieurs sont par la suite fusionnées pour déterminer la décision finale.

La principale limite des approches o-v-o et o-v-a est qu'elles ont tendance à sur-apprendre comparativement à un classifieur multi-classes. De plus, l'algorithme **Max-Win** généralement associé à l'approche o-v-o ne dispose pas de limites sur l'erreur de généralisation [130]. Pour remédier à ces inconvénients, Platt *et al.* [130] ont proposé une méthode ensembliste de classifieurs binaires basée sur un arbre de décision appelé **DAG**. Les résultats obtenus ont démontré que le **DAG** offre non seulement de bonnes performances de généralisation, mais aussi que leur entraînement et évaluation ont un coût de calcul plus faible comparé à la plupart des méthodes ensemblistes.

Dans cette étude, nous nous sommes appuyés sur la méthode de décomposition et d'ensemble pour développer un **SAAD** pour la détection du mélanome. Pour cela, nous avons construit un nouvel ensemble de **CNNs** basé sur des modèles binaires et dont l'agrégation est effectuée à l'aide d'un **DAG**.

La section suivante décrit la méthodologie que nous avons employée pour construire ledit système.

2.3. Méthodes proposées

Dans cette section, nous faisons une description détaillée de notre approche pour construire un **SAAD** permettant de classifier les trois classes de lésion cutanée : mélanome, naevus et kératose. Le choix de ces trois classes a été effectué suivant notre collaboration avec des onco-dermatologues spécialisés dans la prise en charge du mélanome (dépistage et traitements).

Cette section est divisée en trois sous-sections. La première sous-section (**Sous-section 2.3.1**) est consacrée à la description du jeu de données, sa préparation et le prétraitement appliqué aux images. La deuxième sous-section (**Sous-section 2.3.2**) présente la théorie derrière la méthode **DAG**, elle définit la formulation des fonctions d'agrégation courantes et décrit l'architecture des **CNNs** utilisés. La troisième sous-section (**Sous-section 2.3.3**) décrit la méthodologie utilisée pour entraîner les modèles.

L'organigramme du SAAD proposé est illustré dans la figure 2.2.

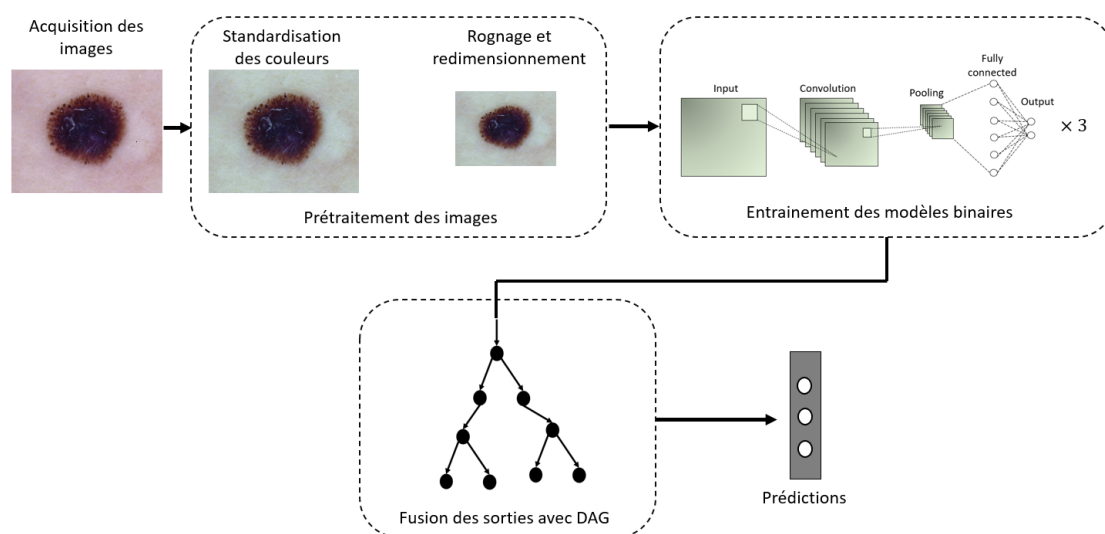


FIGURE 2.2. – Schéma illustratif du système automatique d'aide au diagnostic proposé. Les images des lésions cutanées sont d'abord prétraitées. Ensuite, trois CNNs sont entraînés pour une tâche de classification binaire pour différencier la lésion i d'une autre lésion j . Enfin, la sortie de chaque CNN est agrégée à l'aide d'un graphe acyclique dirigé (DAG) pour obtenir la prédiction finale.

2.3.1. Base de données

Nous avons utilisé la base de données publique ISIC-2018 [31, 157] pour construire notre SAAD pour la tâche de classification des classes mélanome, naevus et kératose.

2.3.1.1. Préparation de la base de données

La base de données ISIC-2018 comprend 1113 images de mélanomes, 6705 images de naevus et 1099 images de kératoses. La préparation de notre base de données a consisté tout d'abord à constituer l'ensemble des jeux de données pour l'entraînement et l'évaluation de notre système. Nous avons pour cela réparti aléatoirement l'ensemble de données original en deux jeux de données : un jeu de données d'entraînement (80%) et un jeu de données de test (20%). Afin de procéder à une validation-croisée de nos modèles, nous avons constitué trois ensembles de données à partir du jeu de données d'entraînement précédemment établi. Chacun des nouveaux ensembles ainsi constitués contient un jeu de données d'entraînement (80% du jeu d'entraînement précédemment établi) et un jeu de données de validation (20% du jeu d'entraînement précédemment établi). Pour réduire le biais induit par la distribution asymétrique des classes présentes dans les jeux de données d'entraînement, nous avons généré artificiellement de nouvelles images dans chacun des trois jeux de données d'entraînement constitué pour la validation-croisée. L'intérêt de générer

2. Construction d'un système de diagnostic automatique du mélanome – 2.3. Méthodes proposées

artificiellement des images pour entraîner des CNNs sur des lésions cutanées a été précédemment démontré par les travaux de Perez *et al.* [127]. Les techniques que nous avons utilisées pour générer les données artificielles étaient : le retournement horizontal, le retournement vertical, le décalage en largeur et en hauteur, et la rotation. Le tableau 2.1 décrit la distribution de l'ensemble de données et la proportion d'images générées utilisées pour l'entraînement des modèles.

TABLEAU 2.1. – Répartition de la base de données.

	kératose	mélanome	naevus
ISIC-2018	1099	1113	6705
Ratio	0.12	0.12	0.75
Jeu de données d'entraînement d'origine (80%)	769	779	4694
Jeu de données de test (20%)	220	223	1341
Jeu de données d'entraînement*	615	623	3755
Jeu de données artificielles généré à partir du jeu de données d'entraînement *	1385	1377	1245
Jeu de données d'entraînement finale *	2000	2000	5000
Ratio après augmentation artificielle des données*	0.22	0.22	0.55
Jeu de données de validation*	154	156	939

* Signifie que l'indication est valable pour chacun des trois ensembles constitués pour la validation-croisée.

2.3.1.2. Prétraitement des images

Nous avons veillé durant cette étude, à réduire au minimum les étapes de prétraitement afin de faciliter la reproductibilité de notre SAAD. Pour cela, nous avons appliqué des opérations de prétraitement standard couramment utilisées en apprentissage profond à savoir : la normalisation, le rognage et le redimensionnement des images. De plus, les images issues de la base de données ISIC-2018 ont été collectées à partir de plusieurs sources et acquises sous différentes modalités, ce qui fait qu'elles présentent une variabilité de luminosité et de couleur de fond. La figure 2.3 illustre cela. Afin de réduire ce biais susceptible d'affecter l'entraînement des modèles, nous avons également appliqué une méthode de standardisation des couleurs.

S'agissant de la normalisation des images effectuée, nous avons transformé les valeurs d'origine des pixels des images qui se situaient dans la plage [0, 255] pour les redéfinir dans la plage [0, 1]. Ceci est effectué en divisant chaque tenseur d'image par 255. La normalisation des images est une bonne pratique en apprentissage profond car elle permet d'accélérer la convergence des modèles.

2. Construction d'un système de diagnostic automatique du mélanome – 2.3.
Méthodes proposées

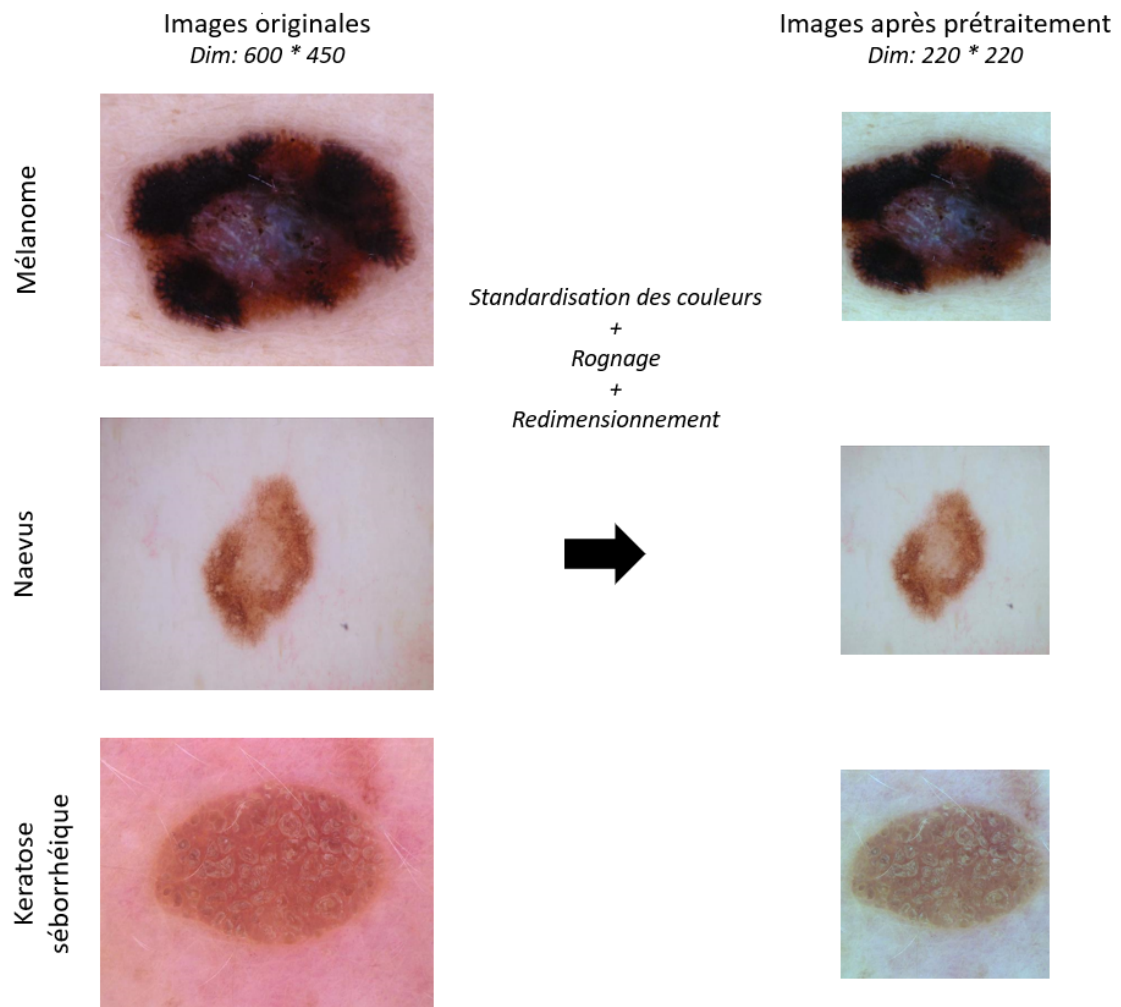


FIGURE 2.3. – Illustration de quelques exemples d'image avant et après avoir été prétraitées.

2. Construction d'un système de diagnostic automatique du mélanome – 2.3. Méthodes proposées

Toutes les images de la base de données ISIC-2018 ont une dimension de 600×450 pixels. Afin d'adapter ces images à la taille d'entrée requise des CNNs pré-entraînés utilisés, nous avons tout d'abord rogné les images afin d'obtenir des images de dimension 450×450 pixels pour préserver le ratio d'aspect. La dimension de 450×450 pixels est celle qui nous permettait d'obtenir la totalité de la lésion présente sur les images rognées. De plus, la fenêtre de rognage a été définie de sorte que son centre coïncide avec le centre de l'image, ceci afin de nous assurer de récupérer la totalité de la lésion. Ce choix est justifié par le fait que la quasi-totalité des lésions du jeu de données sont situées au centre de l'image. Ensuite, nous avons redimensionné nos images à la dimension d'entrée requise de 224×224 pixels de nos CNNs pré-entraînés grâce à la méthode d'interpolation bicubique.

Nous avons utilisé l'algorithme *-gray world-* proposé par Weijer *et al.* [163] pour effectuer la standardisation des couleurs des images. L'équation 2.1 décrit l'équation utilisée pour cette algorithme, où I_c représente une image couleur ayant les trois composantes rouge (R), verte (V) et bleue (B). e_c désigne l'illuminant de chaque composante, $c \in R, V, B$, et k est une constante choisie de telle sorte que l'illuminant ait une norme unitaire. Comme conseillé par Gijsenij *et al.* [68], nous avons modifié l'algorithme original en pré-segmentant l'image en plusieurs blocs d'images carrées de même dimension et en calculant la valeur moyenne de chaque pixel par bloc. Cette étape de standardisation des couleurs est particulièrement importante dans la mesure où Barata *et al.* [16] ont rapporté que cela apportait un gain substantiel dans la classification des lésions cutanées.

$$\frac{\int I_c(x) dx}{\int dx} = k e_c \quad (2.1)$$

2.3.2. Méthodologie

2.3.2.1. Les architectures des réseaux de neurones convolutifs utilisés

Plusieurs architectures de CNNs, pré-entraînés sur la base de données ImageNet contenant 14 millions d'images, sont open-source. Il est donc possible de réutiliser leurs poids et leurs biais, et ré-entraîner ces modèles afin de les adapter à des tâches de classification spécifiques; c'est ce qu'on appelle l'apprentissage par transfert. Nous avons testé notre approche sur trois architectures de CNNs bien connues et régulièrement utilisées avec succès dans la tâche de classification des lésions cutanées à savoir : les réseaux VGG16 et VGG19 [147] et les réseaux de neurones résiduelles –ou *residual neural networks*– (ResNet)-50 [80].

Description de l'architecture ResNet : un réseau très profond a pour conséquence de conduire au problème de saturation du gradient. L'objectif principal à la création du réseau ResNet était de réduire ce problème à l'aide d'un bloc résiduel (voir Figure 2.4). Les concepteurs du réseau ResNet [80] ont proposé plusieurs variantes de cette architecture. Dans notre étude, nous avons utilisé la version ResNet-50 qui comporte

2. Construction d'un système de diagnostic automatique du mélanome – 2.3. Méthodes proposées

50 couches convolutives avec des filtres de dimension 7×7 , 3×3 et 1×1 . Les couches convolutives sont regroupées en blocs résiduels pour construire l'ensemble de l'architecture. Chaque bloc résiduel se compose de quelques couches de convolution empilées, d'une couche de padding, d'une couche de normalisation par lot, et d'une couche de sous-échantillonnage de type échantillonnage maximum. La fonction d'activation utilisée étant la fonction **Unité de rectification linéaire –ou Rectified linear unit– (ReLU)**. Une couche de sous-échantillonnage de type échantillonnage moyen est placée à la fin des blocs résiduels pour compresser les cartes de caractéristiques en un vecteur, et une couche **FC** termine le réseau afin d'effectuer la classification. Dans le cadre de notre étude, nous avons modifié l'architecture ResNet-50 en remplaçant la couche de sortie **FC** par une nouvelle couche **FC** comprenant deux neurones afin de l'adapter à une tâche de classification binaire. La nouvelle architecture ainsi modifiée est illustrée sur la Figure 2.5. Pour plus de simplicité, nous désignerons par la suite cette architecture par *ResNet50₂*.

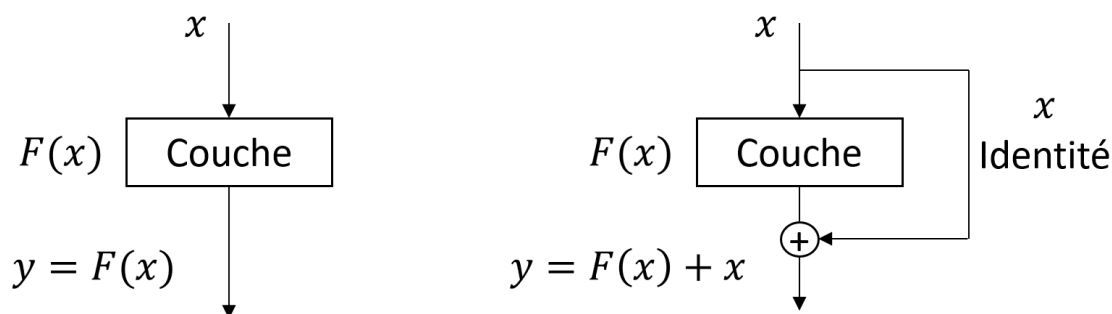


FIGURE 2.4. – Illustration de bloc standard (gauche) et d'un bloc résiduel (droite).

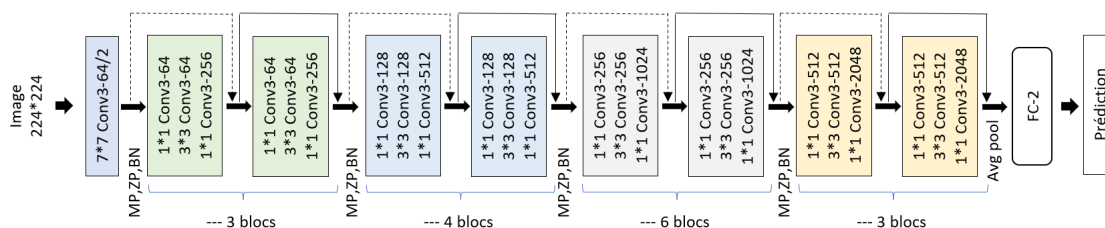


FIGURE 2.5. – illustration de l'architecture ResNet-50 utilisée. Nous avons remplacé la dernière couche entièrement connectée **FC** de l'architecture d'origine par une nouvelle couche **FC** constituée de deux neurones. MP, ZP, BN et Avg pool représentent respectivement la couche de sous-échantillonnage de type échantillonnage maximum, la couche de padding, la couche de normalisation par lot et la couche de sous-échantillonnage de type échantillonnage moyen.

Description de l'architecture VGGNet : VGGNet est une architecture de **CNN** bien documentée et couramment utilisée en vision par ordinateur. Plusieurs variantes de VGGNet ont été initialement proposées [147]. Les variantes se différencient par

2. Construction d'un système de diagnostic automatique du mélanome – 2.3. Méthodes proposées

la profondeur du réseau, allant de 16 à 19 couches. Dans le cadre de notre étude, nous avons utilisé les architectures VGG16 et VGG19, qui ont précédemment été utilisées avec succès sur des tâches de classification des lésions cutanées. VGG16 et VGG19 reçoivent tous deux une image d'entrée de dimension 224×224 pixels. Elles sont constituées de cinq blocs convolutifs. Chaque bloc convolutif se compose de deux, trois ou quatre couches convolutives avec un filtre de dimension 3×3 , des fonctions d'activation de types [ReLU](#) et une couche de sous-échantillonnage de type échantillonnage maximum. Les réseaux sont terminés par un bloc composé de trois couches [FCs](#). Dans notre étude, nous avons modifié les architectures VGG16 et VGG19 d'origine en supprimant leurs dernières couches [FCs](#) et en les remplaçant chacune par une nouvelle couche [FC](#) contenant deux neurones. Les nouvelles architectures VGG16 et VGG19 modifiées sont illustrées sur la figure 2.6. Nous désignerons par la suite ces architectures respectivement par $VGG16_2$ et $VGG19_2$.

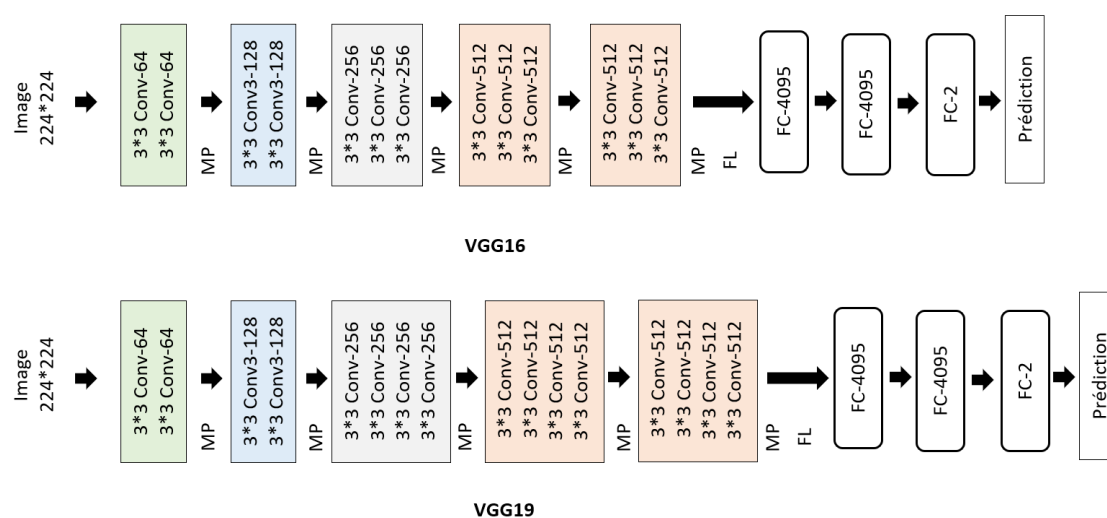


FIGURE 2.6. – illustration des architectures VGG-16 et VGG-19 utilisées. Nous avons remplacé les dernières couches entièrement connectées [FCs](#) des architectures d'origines par de nouvelles couches [FCs](#) constituées de deux neurones. MP, et FL représentent respectivement la couche de sous-échantillonnage de type échantillonnage maximum et l'opération de conversion en vecteur.

2.3.2.2. Agrégation de CNNs avec un graphe acyclique dirigé (DAG-CNN)

Un [DAG](#) est un graphe sans boucle dont les arêtes ont une orientation. L'algorithme [DAG](#) a été initialement introduit par Platt [130] pour étendre le classifieur [SVM](#) à des applications multi-classes. Le principe de base du [DAG](#) est de combiner un ensemble de classifieurs binaires afin de construire un classifieur multi-classes. Le principal avantage d'un [DAG](#) comme méthode d'agrégation est qu'il est plus rapide à construire et à évaluer comparé aux méthodes standards d'agrégation, ceci en conservant une précision comparable [130]. Dans notre approche, les classifieurs binaires utilisés

pour constituer le classifieur multi-classes à l'aide d'un DAG sont des CNNs.

Pour une formalisation adéquate, considérons un CNN binaire comme une fonction $S_{ij} : x \rightarrow R^2$, qui attribue deux prédictions de probabilités $p_i, p_j \in R$ à toute image d'entrée x . $p_i, p_j \in [0, 1]$ et $\sum p_i + p_j = 1$. Chaque modèle S_{ij} catégorise les images x dans une des classes c_i ou c_j . i, j désigne la nature de la lésion, dans notre cas i et $j \in \{\text{melanome}; \text{naevus}; \text{keratose}\}$ et i est différent de j . La figure 2.7 illustre un DAG pour une tâche de classification comportant trois classes $\{1;2;3\}$. Dans la figure 2.7, $\bar{1}$ indique que l'image x n'appartient pas à la classe 1.

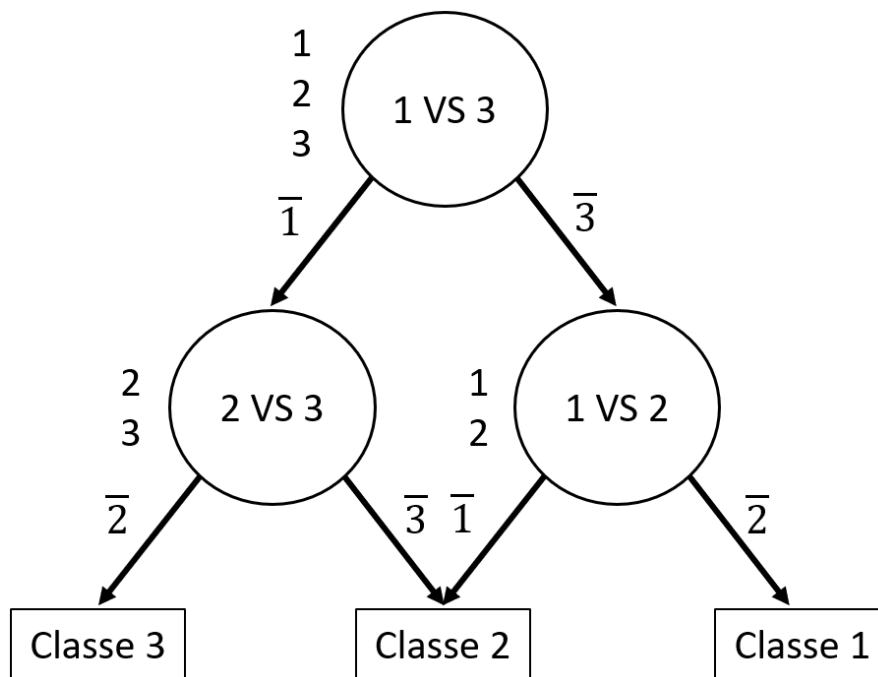


FIGURE 2.7. – Illustration d'un graphe acyclique dirigé pour une tâche de classification comportant trois classes.

Pour une tâche de classification multi-classes contenant N classes, il faudra $N(N - 1)/2$ classifieurs binaires pour construire un DAG-CNN. L'évaluation d'un DAG-CNN en fonction d'une entrée particulière x s'effectue de la manière suivante : tout d'abord le modèle S constituant le nœud racine est évalué. En fonction du résultat obtenu à la sortie du nœud, l'image x est orientée vers le nœud suivant situé à gauche ou à droite pour être de nouveau évaluée. Cette opération est effectuée successivement jusqu'à ce que le dernier nœud situé au bas du graphe soit atteint. Le chemin emprunté pour parvenir à la décision finale sur un DAG-CNN est appelé chemin d'évaluation. Pour un problème de classification contenant N classes, $N-1$ nœuds sont évalués au total pour produire une décision. L'algorithme 1 décrit les étapes suivies pour déterminer le chemin d'évaluation du DAG-CNN dans le cadre d'une tâche de classification comportant trois classes.

Algorithm 1 Evaluation d'un DAG-CNN pour une classification ternaire

Require: image x , trois CNNs binaires S_{ij} , une liste représentant les trois classes $class_list = [1, 2, 3]$

while $\text{len}(class_list) > 1$ **do**

choisir deux éléments i et j dans $class_list$

Générer la classe associée à x prédite par S_{ij}

if S_{ij} prédit x comme appartenant à la classe c_i **then**

retirer j de $class_list$

else

retirer i de $class_list$

end if

end while

x appartient à l'unique classe contenue dans $class_list$

2.3.2.3. Définition des fonctions d'agrégation courantes

Afin de comparer notre approche avec les approches existantes d'ensemble profond, il est nécessaire de redéfinir les fonctions d'agrégation couramment utilisées. Les fonctions d'agrégation sont des outils mathématiques capables de combiner plusieurs attributs en une seule sortie. Plus précisément, une fonction d'agrégation à n dimensions est une fonction monotone $f : [0, 1]^n \rightarrow [0, 1]$ qui satisfait la condition aux limites $f(0, \dots, 0) = 0$ et $f(1, \dots, 1) = 1$. Telle que décrit dans la [Sous-section 2.2.4](#), les fonctions d'agrégation couramment utilisées par les travaux existants sont : les fonctions [Avg](#), [Max-Win](#), [Gmean](#), [Prod](#) et [Max](#).

Pour comparer notre approche avec ces méthodes d'agrégation classiques, nous avons modifié les CNNs *ResNet50*₂, *VGG16*₂ et *VGG19*₂ précédemment décrits. Nous avons remplacé les dernières couches FCs de ces modèles qui comportaient deux neurones, par des couches FCs constituées de trois neurones afin de les adapter à une tâche de classification ternaire. Nous désignerons par la suite ces nouvelles architectures par *ResNet50*₃, *VGG16*₃ et *VGG19*₃.

Considérons un CNN ternaire j défini par la fonction $S_j : x \rightarrow R^3$, qui attribue des scores de probabilités $p_{i|j} \in R$ à toute image d'entrée x . $p_{i|j}$ désigne la probabilité prédite par le modèle j que l'image x appartienne à la classe i . $p_{i|j} \in [0, 1]$, $\sum_i p_{i|j} = 1$, avec $i \in \{\text{melanome}; \text{naevus}; \text{keratose}\}$ et $j \in \{\text{ResNet50}_3, \text{VGG16}_3, \text{VGG19}_3\}$. Soit p'_i la probabilité dérivée des scores $p_{i|j}$ qu'une image d'entrée x appartient à la classe i . Les différentes formulations de p'_i selon la méthode d'agrégation sont :

— méthode d'agrégation [Avg](#) :

$$p'_i = \frac{\sum_{j=1}^3 p_{i|j}}{3} \quad (2.2)$$

— méthode d'agrégation **Prod** :

$$p'_i = \frac{\prod_{j=1}^3 p_{i/j}}{\sum_{i=1}^3 \prod_{j=1}^3 p_{i/j}} \quad (2.3)$$

— méthode d'agrégation **Max** :

$$p'_i = \frac{\max_j p_{i/j}}{\sum_{i=1}^3 \max_j p_{i/j}} \quad (2.4)$$

— méthode d'agrégation **Gmean** :

$$p'_i = \frac{\sqrt[3]{\prod_{j=1}^3 p_{i/j}}}{\sum_{i=1}^3 \sqrt[3]{\prod_{j=1}^3 p_{i/j}}} \quad (2.5)$$

— méthode d'agrégation **Max-Win** :

$$p'_i = \frac{\sum_{j=1}^3 F(p_{i/j})}{3} \quad (2.6)$$

$$avec F(p_{i/j}) = \begin{cases} 1, & si p_{i/j} = \max_j(p_{i/j}) \\ 0, & sinon \end{cases} \quad (2.7)$$

2.3.3. Entraînement des modèles

Cette sous-section fait une description détaillée du protocole expérimental utilisé pour entraîner tous nos modèles. Toutes les expériences réalisées dans cette étude ont été menées à partir d'un ordinateur doté d'un processeur 3,2 GHz, de 16 Go de mémoire vive et d'une carte GPU Nvidia GeForce Rtx 2080. Nous avons utilisé le logiciel MATLAB 2020 pour préparer et standardiser les couleurs des images. La bibliothèque Keras avec Tensorflow comme backend a été utilisée pour entraîner nos modèles. Le code de cette étude est disponible en [open source](#)¹.

Nous avons utilisé l'algorithme d'optimisation Adam [93] pour mettre à jour les poids et les biais de nos modèles à chaque itération afin de minimiser la fonction de coût. La fonction de coût employée pour entraîner nos modèles est la fonction d'entropie croisée pondérée. Le terme général de la fonction d'entropie croisée est :

$$L_{CE} = -w_l \sum_{i=1}^N l_i \times \log(p_i) \quad (2.8)$$

Dans l'équation précédente, l_i représente l'étiquette de l'image i , p_i est la probabilité prédite par le modèle, w_l est le poids de la classe l et N est le nombre total

1. https://github.com/cartelgouabou/DAG_CNN_on_dermoscopic_images

d'images. Le poids de pondération w_l utilisé correspond à la fréquence normalisée inverse de chaque classe et est défini comme suit :

$$w_l = \frac{N}{n_l} \quad (2.9)$$

avec n_l représentant le nombre d'images appartenant à la classe l . Les poids des modèles utilisés ont été initialisés avec leurs poids initiaux issus de leurs pré-entraînement sur la base de données [ImageNet](#). Pour chaque modèle, nous avons évalué différents hyperparamètres et avons entraîné les modèles durant 150 époques. Plus précisément, les hyperparamètres que nous avons évalués étaient le taux d'apprentissage initial, variant de 0,01 à 0,0001, et le ratio des couches profondes à ré-entraîner, variant de 4% à 65%. De plus, nous avons employé une technique d'apprentissage dynamique en variant le taux d'apprentissage initial (l_r) suivant une décroissance polynomiale lorsque le taux d'erreur sur le jeu de données de validation ne diminuait pas après huit époques. Le calcul du taux d'apprentissage programmé sch_{l_r} s'effectue de la manière suivante :

$$sch_{l_r} = l_r \times \left(1 - \frac{epoque_courante}{Nombre_total_d_epoques}\right) \quad (2.10)$$

Le tableau 2.2 résume l'espace de recherche des hyperparamètres utilisés pour entraîner nos modèles. Durant l'entraînement, nous avons également utilisé une technique de régularisation consistant à arrêter l'entraînement aussitôt que l'exactitude sur le jeu de données d'entraînement était supérieure de plus de 10% à celle sur le jeu de données de validation. Cette technique de régularisation est employée afin d'éviter le sur-apprentissage des modèles. Les modèles ayant obtenus le meilleur score de [BAcc](#) sont ceux qui sont retenus pour générer les prédictions.

TABLEAU 2.2. – Espace de recherche des hyperparamètres pour l'entraînement des modèles.

Algorithme d'optimisation	Taille des Mini-batch	Nombre d'époques	ratio des couches profondes ré-entraînées	taux d'apprentissage
Adam	32	150	[4%, 65%]	[1e-2, 1e-4]

2.3.4. Définition des critères de performance utilisés

Les métriques utilisées dans cette étude pour évaluer notre approche sont l'[AUC](#) et la [BAcc](#). Les sensibilité et spécificité des modèles sont également calculées. Chacun des résultats présentés dans la section suivante a été obtenu en employant une triple [validation-croisée](#) sur le jeu de données d'entraînement et de validation. Chacun des résultats sera présenté par la suite sous forme de moyenne avec son écart-type associé.

Pour les études comparatives, des tests statistiques non paramétriques ont été effectués. Dans des cas de figure comparant deux modèles, nous avons utilisé un test *t* apparié (ou test de Student). Dans le cas où la comparaison de plus de deux modèles devait être faite, nous avons utilisé le test de Kruskal-Wallis [111], suivi d'un test de comparaison multiple utilisant le test de Dunn [50]. Les résultats ont été considérés comme statistiquement significatifs si $p\text{-valeur} < 0,05$. Les calculs statistiques ont été réalisés à l'aide du logiciel GraphPad Prism, version 5.03.

2.4. Résultats et discussions

Dans cette section, nous présentons les résultats obtenus par notre approche DAG-CNN. Premièrement, nous présentons les performances de chacun des classifieurs binaires. Deuxièmement, nous analysons l'effet du choix du nœud racine sur la performance de notre approche. Troisièmement, nous comparons notre méthode avec des modèles standard de CNNs entraînés pour une tâche de classification multi-classes. Nous analysons par la suite les performances de la meilleure structure DAG-CNN obtenue. Enfin, nous comparons notre approche à d'autres stratégies d'agrégation conventionnelles et d'autres travaux existants pour la même tâche de classification.

2.4.1. Performance des modèles binaires

Le tableau 2.3 montre les résultats obtenus sur le jeu de données d'entraînement avec les modèles *ResNet50*₂, *VGG16*₂ et *VGG19*₂ pour chaque tâche individuelle : *mélanome-versus-naevus*, *mélanome-versus-kératose* et *naevus-versus-kératose*. Le score de BAcc et la sensibilité pour chaque classe sont présentées. Dans l'ensemble, les classifieurs binaires ont obtenu de bonnes performances avec dans tous les cas de figures des scores de BAcc minimum de 80.2%.

Dans toutes les trois tâches de classification binaire, les modèles *VGG19*₂ ont obtenu les meilleures performances comparées aux modèles *ResNet50*₂ et *VGG16*₂. Ce résultat peut s'expliquer par le fait que l'architecture *VGG19* comporte plus de couches profondes par rapport aux architectures *VGG16* et *ResNet50*. Par conséquent, *VGG19* peut apprendre des caractéristiques plus discriminantes. Fait intéressant, parmi les trois tâches de classification binaire évaluées, la tâche *kératose-versus-naevus* était en général la plus facile à distinguer quelle que fût l'architecture. D'un point de vue dermatologique, la lésion kératose est très sombre et composée de motifs complètement différents des naevus, tels que des structures kératiniques, des kystes de corne ou un motif cérébriforme.

La tâche qui s'est révélée la plus difficile à distinguer fût la tâche *mélanome-versus-naevus*. Lorsque le mélanome est excisé à un stade précoce avec un indice de Breslow fin, la difficulté de différencier le mélanome des naevus est élevée même pour des experts dermatologues. De plus, certains mélanomes surviennent à partir des

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

naevus, ils peuvent donc partager les mêmes motifs et structures (motif réticulaire ou motif en pointillés). Toutefois, il est important de noter que les motifs des mélanomes sont en général plus irréguliers que ceux des naevus.

TABLEAU 2.3. – Performances des modèles binaires sur le jeu d'entraînement et de validation évaluées à l'aide d'une triple validation croisée.

Tâche	CNN	BAcc	S-MEL	S-NEV	S-KES
<i>MEL – vs – NEV</i>		80.6±0.3%	78.3±3.0%	84.0±3.5%	-
<i>MEL – vs – KES</i>	VGG19	85.8±1.6%	85.6±1.8%	-	84.5±2.2%
<i>NEV – vs – KES</i>		86.7±1.2%	-	88±0.8%	85.3±2.3%
<i>MEL – vs – NEV</i>		80.2±0.5%	75.0±1.0%	84±1.0%	-
<i>MEL – vs – KES</i>	VGG16	83.4±1.9 %	87±2.0%	-	79.5±0.5%
<i>NEV – vs – KES</i>		87.2±0.4%	-	83.5±1.5%	90.5±1.5%
<i>MEL – vs – NEV</i>		81.8±1.6%	82.3±0.4%	81.2±2.6%	-
<i>MEL – vs – KES</i>	ResNet50 ₂	81.3±0.5%	90±1.0%	-	71.5±1.9%
<i>NEV – vs – KES</i>		81.5±0.4%	-	87±1.3%	76±1.5%

S désigne la sensibilité.

2.4.2. Impact du choix du nœud racine

Le deuxième aspect que nous avons étudié était l'effet des variations du nœud racine sur la performance globale de notre approche. Pour cela nous avons analysé les performances en terme de BAcc de notre approche lorsque la tâche de classification placée sur le nœud racine variait. Les résultats de cette analyse sont présentés dans le tableau 2.4.

La première observation faite est que quelques soit la structure du DAG-CNN choisie, la performance globale de notre méthode d'ensemble profond est corrélée à la performance de chaque classifieur constituant l'ensemble. C'est ainsi que les DAG-CNNs basés sur l'architecture VGG19 ont obtenu de meilleurs résultats. En effet, les structures DAG-CNNs basées sur les modèles VGG19₂ ont atteint des BAcc entre 73,7% et 76,6%, contre 72,55% à 73,25% pour celles basées sur les modèles VGG16₂ et 70,1% à 71,1% pour celles basées sur les modèles ResNet50₂.

Les auteurs de [152, 1] ont observé que lors de l'agrégation de SVMs avec un DAG, la performance globale du modèle multi-classes résultant variait légèrement selon le classifieur SVM qui était placé sur le nœud racine. En général, Takahashi *et al.* [152] suggèrent de placer le classifieur avec la plus grande capacité de généralisation dans le nœud racine pour obtenir la structure optimale du DAG. La deuxième observation

TABLEAU 2.4. – Effets de la variation du nœud racine sur les performances de la structure DAG-CNN.

Nœud racine	CNN	BAcc
<i>MEL – vs – NEV</i>		73.7±1.2%
<i>MEL – vs – KES</i>	VGG19	74.2±0.9%
<i>NEV – vs – KES</i>		74.6±0.4
<i>MEL – vs – NEV</i>		72.6±1.6%
<i>MEL – vs – KES</i>	VGG16	72.9±1.7%
<i>NEV – vs – KES</i>		73.3±1.5%
<i>MEL – vs – NEV</i>		71.1±0.6%
<i>MEL – vs – KES</i>	ResNet50	70.1±1.0%
<i>NEV – vs – KES</i>		70.4±0.8%

Les meilleures performances sont marquées en gras.

fait est similaire à celle précédemment décrite. En effet, pour chacune des trois architectures testées, les structures DAG-CNNs avec les classifieurs binaires les plus performants placés sur le nœud racine sont celles qui ont obtenu les meilleures performances (confère [Sous-section 2.4.1](#)).

La meilleure structure DAG-CNN obtenue est celle basée sur l'architecture VGG19 et ayant un classifieur binaire conçu pour la tâche *kératose – versus – mélanome* comme nœud racine. La BAcc obtenue par cette structure est de $74,6 \pm 0,4\%$.

2.4.3. Comparaison de l'approche DAG-CNN avec des CNNs multi-classes standards

Dans cette sous-section nous évaluons notre hypothèse principale. Pour ce faire, nous comparons notre structure DAG-CNN avec les modèles *ResNet50*₃, *VGG16*₃ et *VGG19*₃ formés pour des tâches de classification comportant trois classes. Le tableau 2.5 présente les résultats de cette analyse obtenus par [validation-croisée](#). Quelle que soit l'architecture utilisée, les CNNs binaires agrégés à l'aide d'un DAG ont obtenu de meilleures performances que des CNNs conçus pour des tâches de classification multi-classes.

Afin d'effectuer une analyse plus approfondie, nous avons évalué chacun de ces modèles sur le jeu de données de test. Le tableau 2.6 présente les résultats de cette étude. Chaque étude comparative y est présentée avec les résultats des tests statistiques appariés *t* effectués. Nous avons constaté que les améliorations des performances en termes de BAcc obtenues avec la structure DAG-CNN étaient toutes statistiquement significatives. De plus, la structure DAG-CNN utilisant l'architecture VGG19 est celle

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

TABLEAU 2.5. – Validation-croisée des performances de chaque structure DAG-CNN avec la version ternaire du CNN utilisé.

Méthode	Architecture CNN	BAcc
DAG-CNN	VGG19	76.6±0.4%
VGG19 ₃	VGG19	76.5±1.5%
DAG-CNN	VGG16	73.3±1.5%
VGG16 ₃	VGG16	72.5±1.5%
DAG-CNN	ResNet50	71.1±0.6%
ResNet50 ₃	ResNet50	70.9±1.1%

Les meilleures performances sont marquées en gras.

ayant obtenu les meilleures performances, atteignant une BAcc de 76,6% sur le jeu de données de test. Ces résultats nous ont permis de conclure que la décomposition d'un problème multi-classes en un problème binaire réduit la complexité du problème initial et permet d'augmenter les performances globales de classification. Néanmoins, il est important de souligner, que les performances de la structure DAG-CNN sont étroitement liées à la performance de l'architecture de CNN choisie comme l'illustrent les résultats obtenus avec VGG16 et ResNet50 (voir Tableau 2.6). L'utilisation des architectures avec des performances potentiellement meilleures, comme EfficientNet[155] pourrait améliorer les performances de notre approche.

TABLEAU 2.6. – Analyse statistique comparative des performances de la structure DAG-CNN avec des CNNs multi-classes sur le jeu de données de test.

Architecture CNN	DAG-CNN	CNN multi-classes
VGG19	76.6%***	74.6%
VGG16	73.9%***	73.6%
ResNet50	70.4%***	62.4%

*** : p-valeur < 0.0001

2.4.4. Analyse des performances de la méthode DAG-CNN

La figure 2.8 montre les courbes ROCs obtenues par notre meilleure structure DAG-CNN sur le jeu de données de test pour chacune des trois lésions contenues dans notre tâche de classification. Notre approche a atteint des AUCs de 0,92, 0,87 et 0,88 respectivement pour les classes kératose, mélanome et naevus. Nous avons observé

que le mélanome restait la classe la plus difficile à classifier.

Par la suite, nous avons présenté notre SAAD à l'équipe d'onco-dermatologues avec qui nous travaillons en collaboration pour une analyse approfondie. Pour faciliter l'analyse du dermatologue, nous avons associé chaque prédiction fournie par un modèle CNN de base avec sa carte thermique correspondante, permettant ainsi de visualiser les régions contribuant à la prédiction. La génération des cartes thermiques a été implémentée à l'aide de l'algorithme Grad-CAM [143]. La figure 2.9 illustre la stratégie de décision de notre approche. Les flèches en vert représentent le chemin d'évaluation dans ce cas. Comme exemple illustratif, nous avons pris une image d'une lésion pigmentée suspecte. A première vue, cette lésion pouvait être une kératose bénigne ou bien un mélanome. Pour un choix entre la classe kératose et la classe naevus, le CNN situé au nœud racine a attribué une probabilité plus élevée de 0,8 pour l'appartenance à la classe kératose. En suivant notre pipeline, la lésion a donc été orientée vers le nœud correspondant à la tâche de classification *mélanome – versus – kératose*. La deuxième étape de notre structure DAG-CNN a permis d'établir la bonne prédiction finale en attribuant une probabilité de 1.0 pour l'appartenance de la lésion à la classe mélanome, qui était en réalité l'étiquette de vérité terrain. Le réseau réticulaire de la lésion en exemple est irrégulier et élargi sur la partie gauche. Sur cette partie et au milieu, on peut également observer une couleur de voile bleu blanc avec quelques points correspondant à une zone de régression, qui est associée au diagnostic de mélanome. Fait intéressant, la carte thermique révèle la partie gauche atypique de la lésion (zone marquée en rouge sur la carte thermique de la Figure 2.9) comme la principale zone de caractéristiques ayant contribué à la prédiction. Cette partie correspondant effectivement aux caractéristiques les plus suspectes pour le diagnostic du mélanome.

Nous nous sommes également intéressés à l'analyse des images des lésions bien classées ainsi que celles des lésions mal classées par notre architecture DAG-CNN pour chacune des trois classes contenues dans notre tâche de classification. Pour cela nous avons ressorti, pour chaque cas de figure d'images de lésions correctement et faussement prédites, neuf exemples de la classe kératose (voir figures 2.10 et 2.13), de la classe mélanome (voir figures 2.11 et 2.14) et de la classe naevus (voir figures 2.12 et 2.15). Nous avons observé en général que les images correctement classées avec une grande probabilité correspondaient à des lésions typiques de chacune des trois classes. A l'inverse, les images mal classées et avec des probabilités de prédictions également très grandes correspondaient à des lésions atypiques à chacune des trois classes. Ce résultat peut s'expliquer par le fait que les lésions atypiques sont naturellement moins fréquentes sur l'ensemble de la population examinée, ce qui a pour effet d'induire un biais lors de l'apprentissage des modèles neuronaux gourmands en données. D'un autre côté, l'observation effectuée au sujet des probabilités de prédictions générées par les modèles a révélé la problématique selon laquelle l'attribution d'un score de probabilité comme critère de décision ne constitue pas à elle seule un critère robuste pour l'aide à la décision.

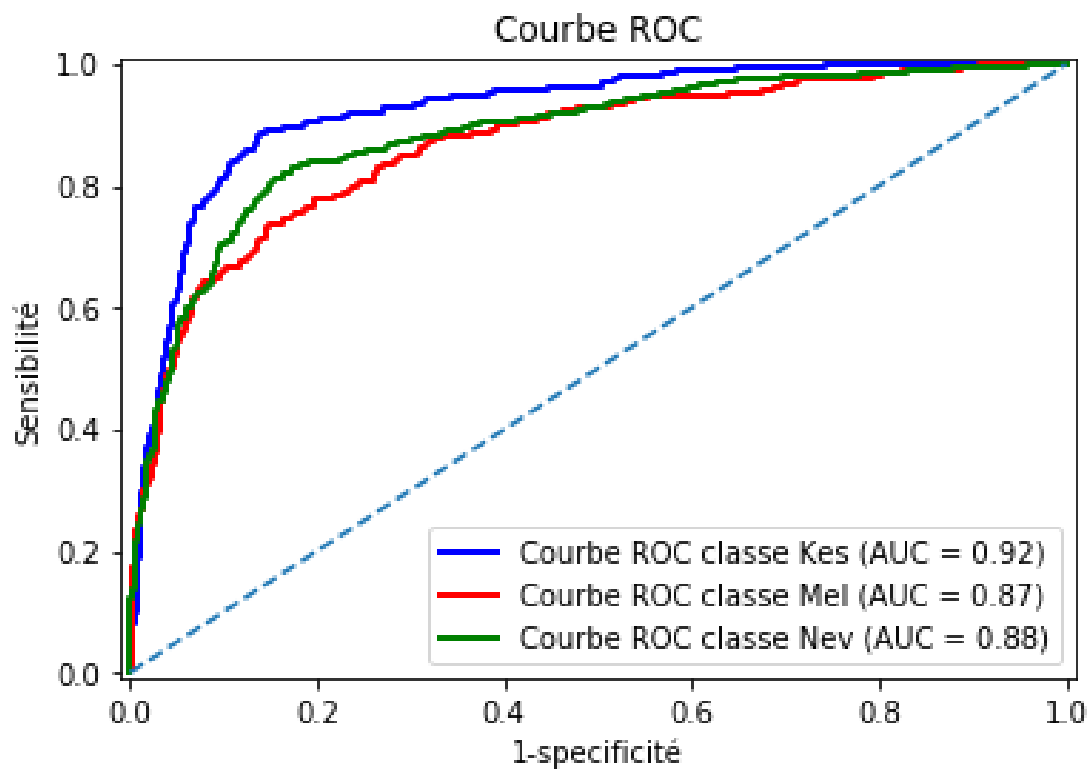


FIGURE 2.8. – Courbes ROC obtenues avec la meilleure structure DAG-CNN. L'AUC de chacune des trois classes mélanome (MEL), kératose séborrhéique (SEK) et naevus (NEV) est précisée.

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

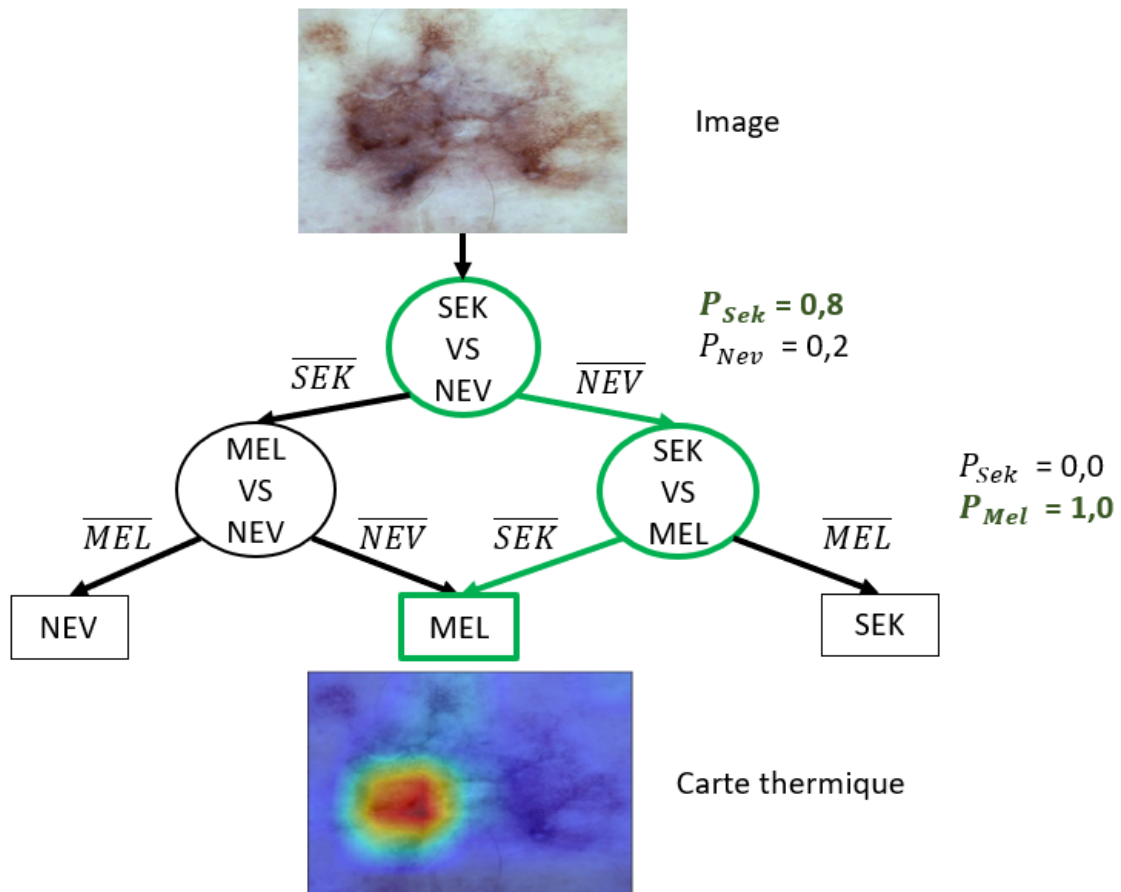


FIGURE 2.9. – Illustration du processus de décision de notre méthode DAG-CNN pour la détection du mélanome. La génération de cartes thermiques a été implémentée à l'aide de l'algorithme Grad-CAM [143]. MEL : mélanome, KES : kératose séborrhéique, NEV : naevus.

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions correctes de Kératoses bénignes

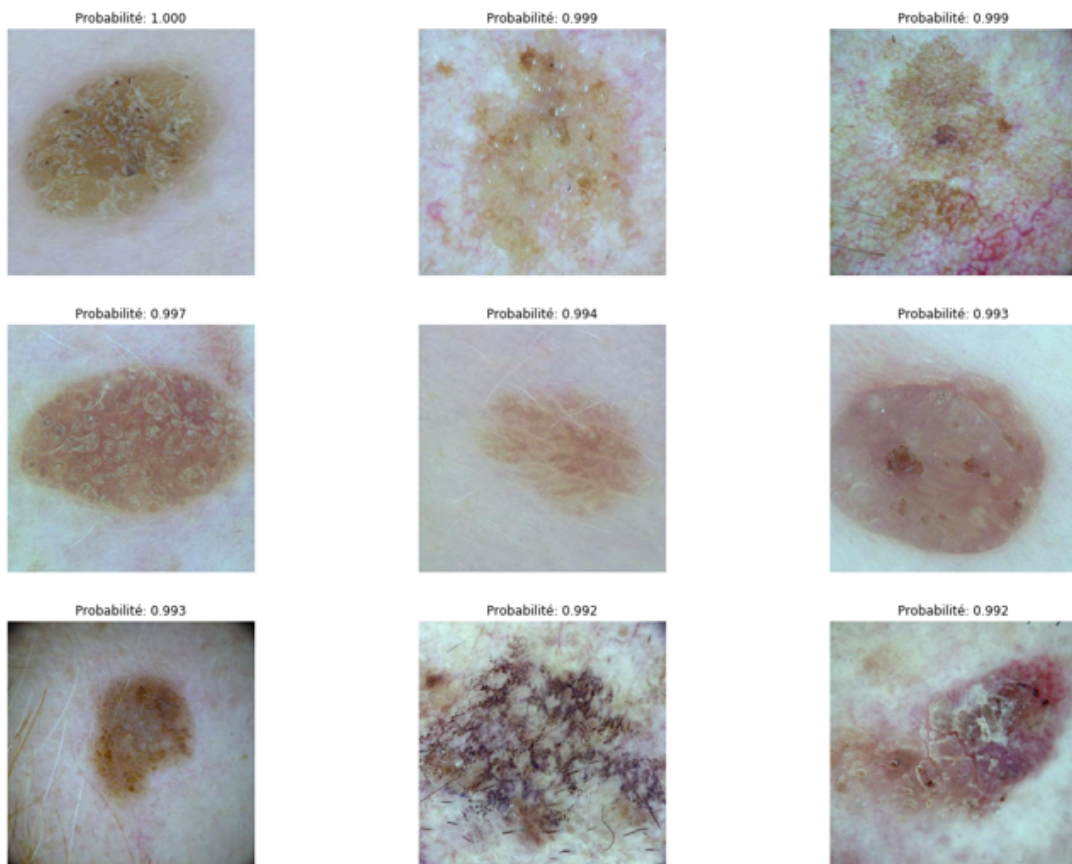


FIGURE 2.10. – Exemples de prédictions correctes de Kératoses bénignes avec leurs probabilités associées

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions correctes de Mélanomes

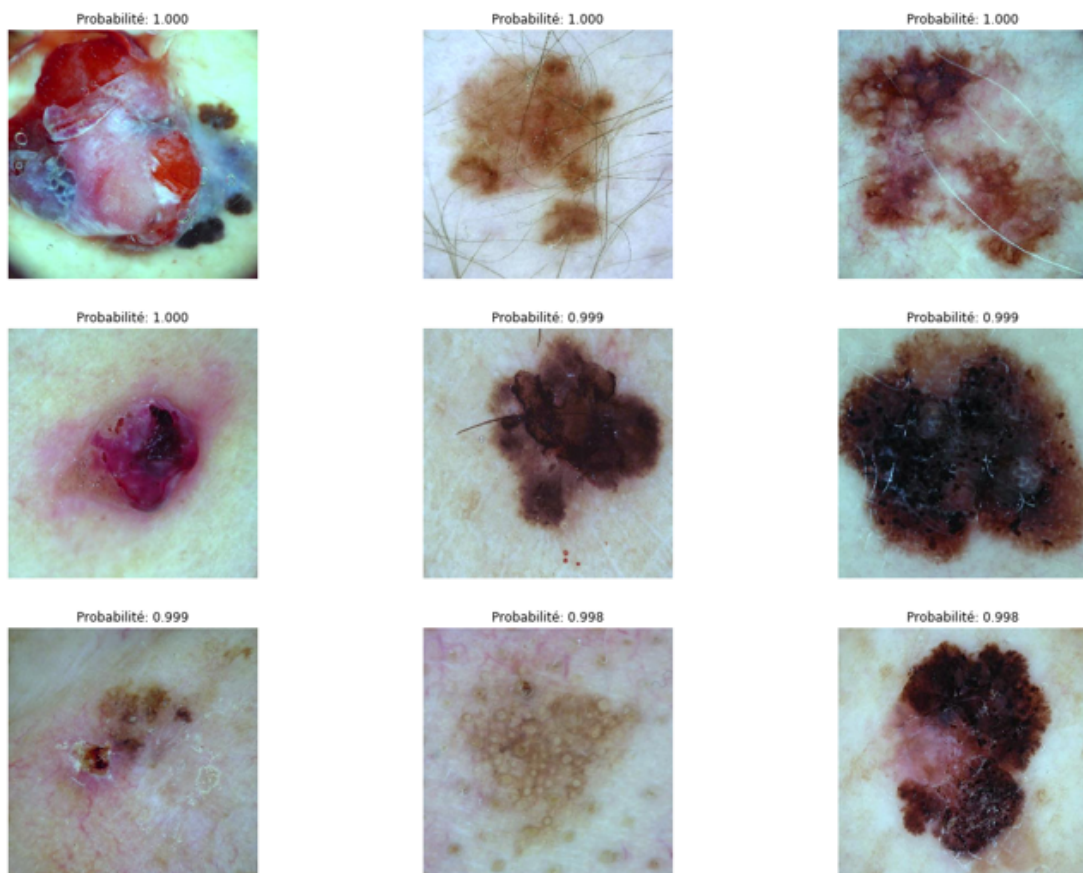


FIGURE 2.11. – Exemples de prédictions correctes de Mélanomes avec leurs probabilités associées

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions correctes de Naevus

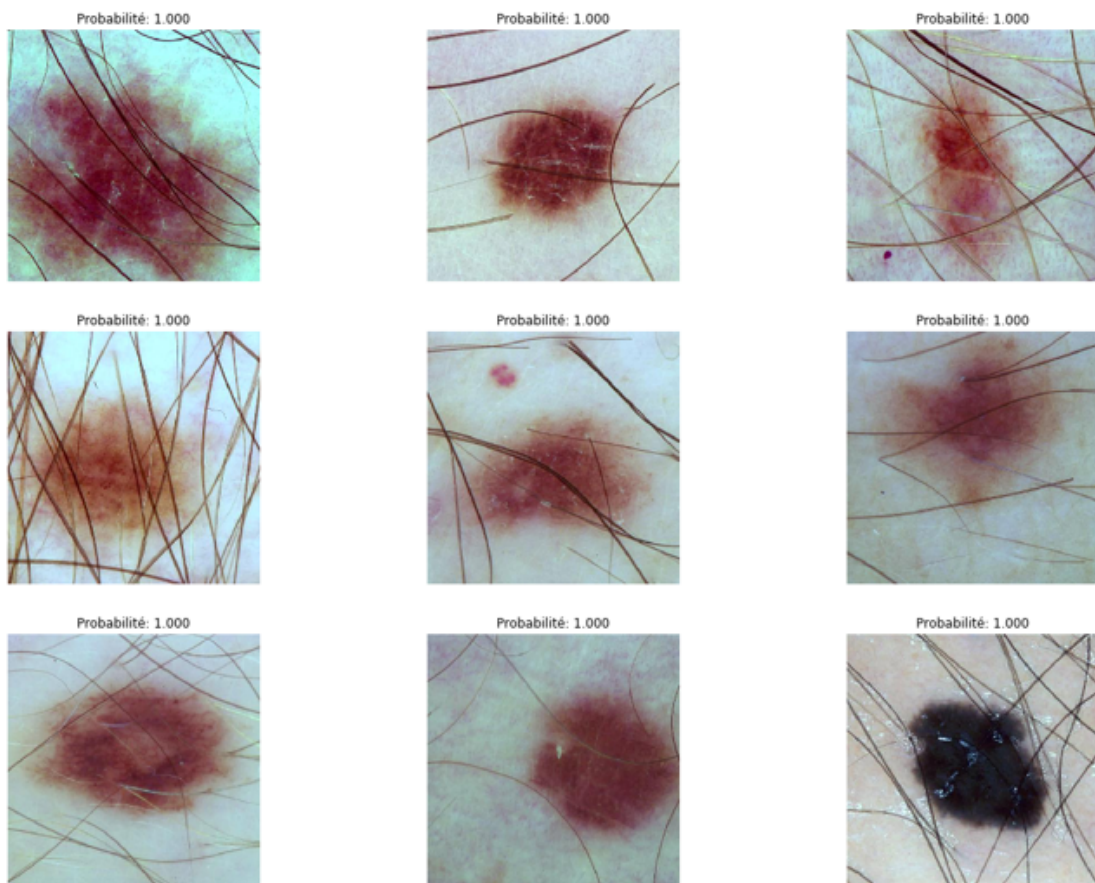


FIGURE 2.12. – Exemples de prédictions correctes de Naevus avec leurs probabilités associées

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions erronées de Kératoses bénignes

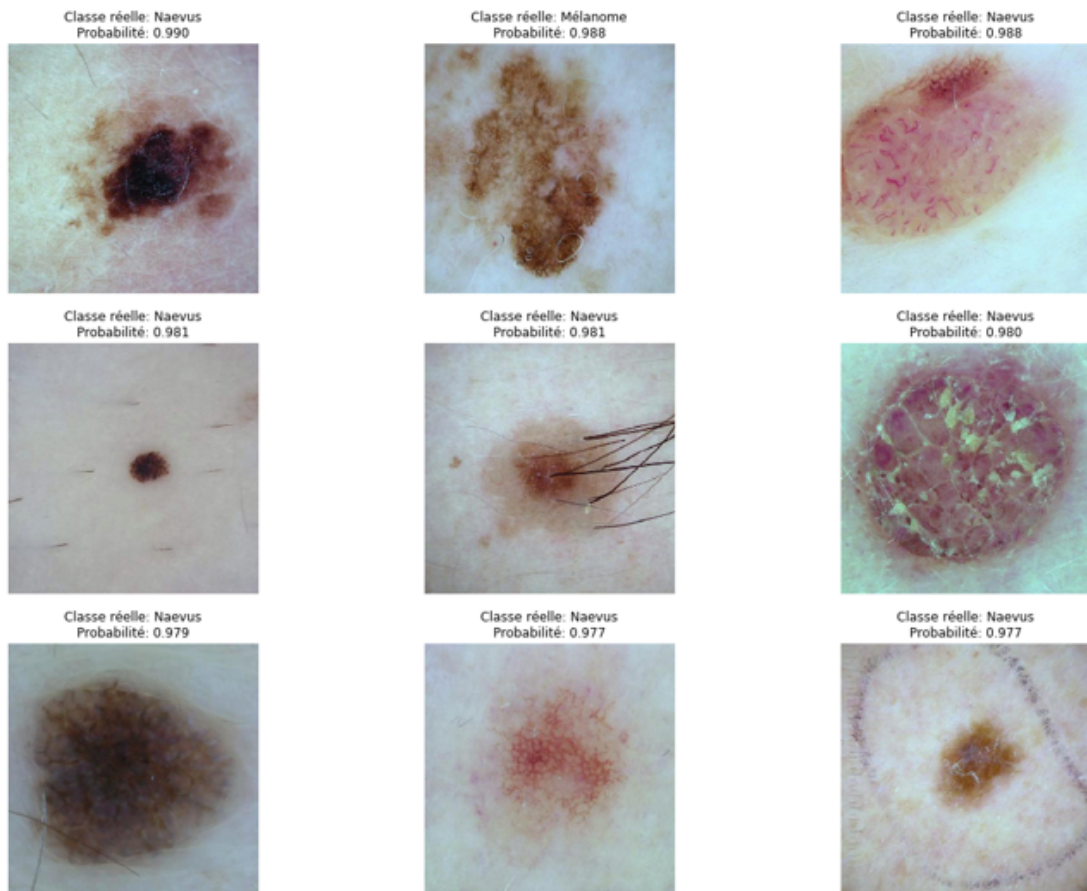


FIGURE 2.13. – Exemples de prédictions erronées de Kératoses bénignes avec les probabilités prédites associées et leurs classes réelles

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions erronées de Mélanomes

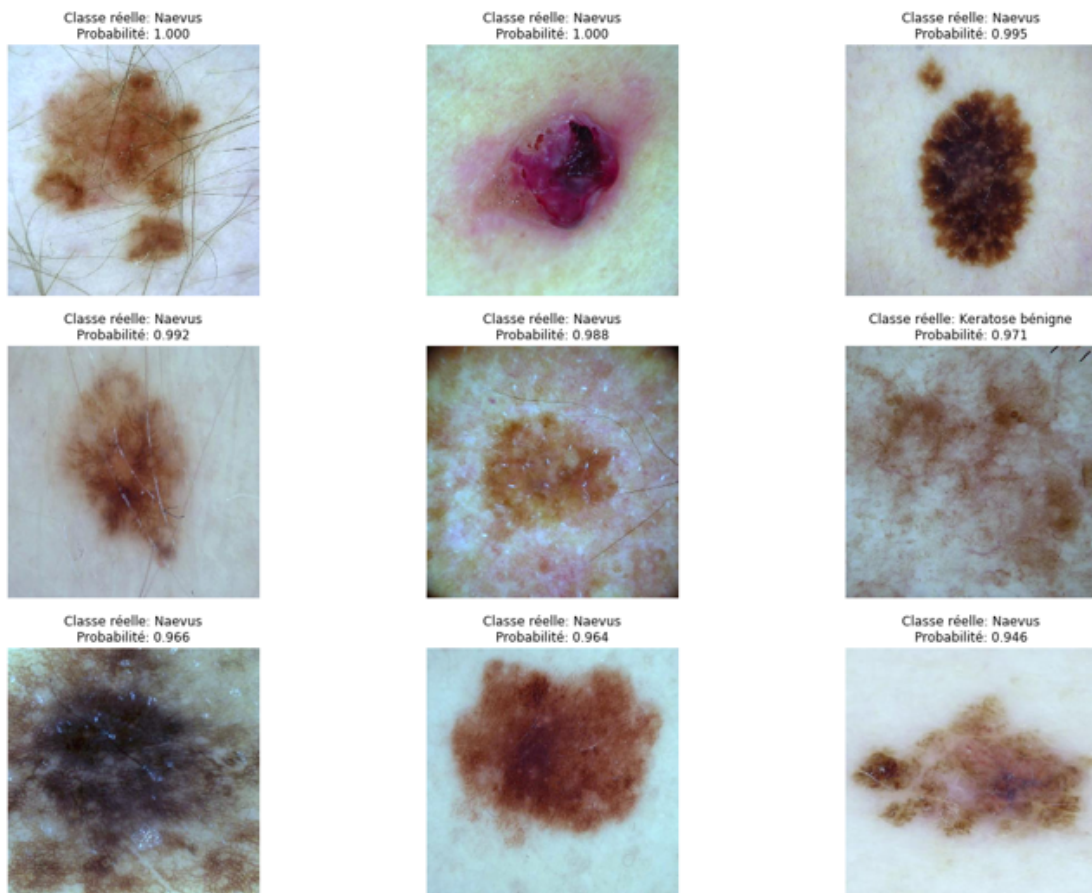


FIGURE 2.14. – Exemples de prédictions erronées de Mélanome avec les probabilités prédites associées et leurs classes réelles

2. Construction d'un système de diagnostic automatique du mélanome – 2.4.
Résultats et discussions

Exemples de prédictions erronées de Naevus

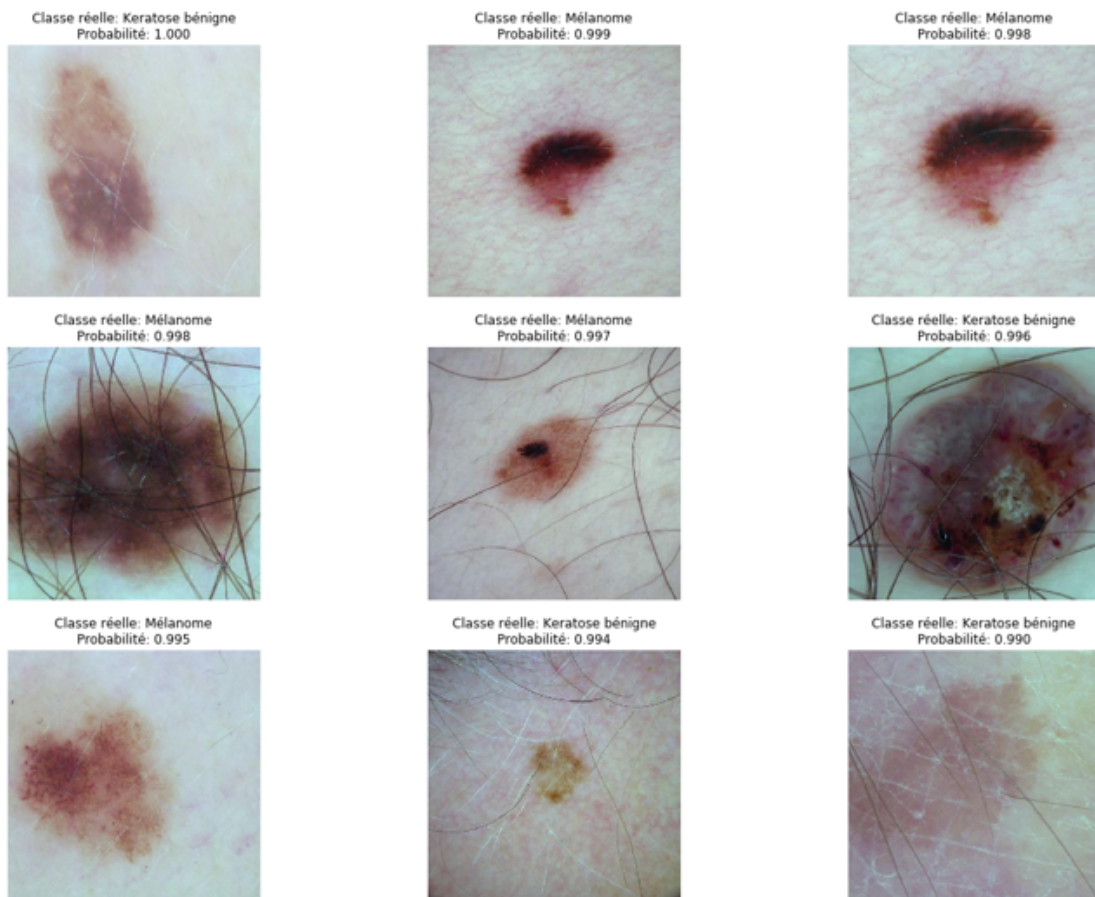


FIGURE 2.15. – Exemples de prédictions erronées de Naevus avec les probabilités prédites associées et leurs classes réelles

2.4.5. Comparaison avec d'autres approches de fusion

Nous avons comparé notre approche d'ensemble profond basée sur les DAGs avec les méthodes d'agrégation couramment utilisées (Avg, Max, Prod, Gmean et Max-Win). Pour cela, les meilleurs modèles obtenus avec *ResNet50*₃, *VGG16*₃ et *VGG19*₃ lors de la validation-croisée ont été fusionnés suivant ces méthodes d'agrégation. Les résultats présentés dans le tableau 2.7 résument pour chaque méthode les performances obtenues sur le jeu de données de test et leurs analyses statistiques associées.

La structure DAG-CNN a obtenu le meilleur score de BAcc (76,6%) comparée aux autres méthodes d'ensemble profond utilisant des CNNs multi-classes et des méthodes d'agrégation classiques. De plus, les scores de probabilité générés par notre approche se sont révélés statistiquement différents ($p < 0,05$) de ceux des autres méthodes d'agrégation classiques, ce qui confirme la robustesse de notre structure DAG-CNN et sa capacité à améliorer les performances d'un SAAD. Nous avons également constaté, parmi les méthodes d'agrégation classiques, que les méthodes Avg, Max et Gmean sont celles qui ont obtenues les meilleures performances, sans différence statistique entre leurs scores prédits. Ce résultat suggère que ce sont les meilleures approches d'agrégation à utiliser pour construire un SAAD sur des images dermoscopiques parmi les méthodes d'agrégations classiques. La méthode d'agrégation Prod est celle qui a obtenue la moins bonne performance. Ceci peut s'expliquer par le fait que le produit distribue plus significativement les fausses prédictions. Pour réduire cet effet dans l'application des méthodes d'ensemble profond, nous recommandons de ne fusionner que des classifieurs ayant des performances similaires.

Nous avons également comparé notre méthode avec des méthodes existantes sur la même tâche de classification à trois classes [15, 17]. En se basant sur le score de BAcc, notre approche a surclassé tous les travaux précédemment cités.

De plus, nous pensons que notre approche apporte une meilleure clarté sur le processus de décision car elle suit un flux de décision hiérarchique similaire au raisonnement en deux étapes[12] des dermatologues, contrairement aux approches conventionnelles qui se contentent d'agréger plusieurs CNNs sans offrir de transparence dans le processus de prise de décision.

2. Construction d'un système de diagnostic automatique du mélanome – 2.5. Conclusion

TABLEAU 2.7. – Comparaison de notre approche avec des méthodes d'agrégation classiques et d'autres méthodes existantes de la littérature sur la même tâche de classification.

Méthode d'agrégation	CNNs	BAcc	Test statistique	
			g.s	g.n.s
DAG	VGG19 ₂	76.6%	1	{2–4}
Avg	ResNet50 ₃ + VGG16 ₃ + VGG19 ₃	75.2%	2	{3–4}
Max	ResNet50 ₃ + VGG16 ₃ + VGG19 ₃	75.2%	2	{3–4}
Gmean	ResNet50 ₃ + VGG16 ₃ + VGG19 ₃	75.2%	2	{3–4}
Max-Win	ResNet50 ₃ + VGG16 ₃ + VGG19 ₃	74.6%	3	{4}
[15]	VGG16	74.3%	-	-
Prod	ResNet50 ₃ + VGG16 ₃ + VGG19 ₃	73.9%	4	{}
[17]	DenseNet-161	70%	-	-

$p < 0.05$

Le classement des groupes des méthodes avec un score de BAcc statistiquement similaire est indiqué par “g.s” (1 est le meilleur), tandis que “g.n.s” indique les groupes qui sont statistiquement moins bons. Le test statistique utilisé ici est le test de Kruskal-Wallis et le test de comparaison multiple de Dunn.

2.5. Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode d'ensemble profond permettant la construction d'un SAAD du mélanome. La méthode proposée a été évaluée sur une tâche de classification des images de lésions cutanées de mélanomes par rapport aux lésions cutanées de naevus et de kératoses séborrhéiques. Notre approche consiste en un ensemble de CNNs et s'inspire de la méthode de décomposition et d'ensemble. Pour ce faire, trois CNNs ont été entraînés pour une tâche de classification binaire de deux parmi les trois classes que comporte notre tâche de classification. Les sorties des trois modèles de bases ainsi constitués sont finalement agrégées à l'aide d'un DAG. Les résultats obtenus nous ont permis de démontrer que notre approche permettait d'obtenir de meilleurs résultats qu'un modèle CNN multi-classes standard. En outre, nous avons comparé notre méthode avec d'autres méthodes d'agrégation couramment utilisées dans la littérature : l'Avg, la Prod, de la Max, la Gmean et le Max-Win. La méthode proposée a été en mesure de surclasser toutes les méthodes d'agrégation précédemment citées. Ces résultats ont mis en évidence l'efficacité de notre approche DAG-CNN atteignant une BAcc sur la tâche de classification de 76.6%. Ceci nous a permis de corroborer notre hypothèse de départ selon laquelle la décomposition d'un problème multi-classes en un problème binaire réduit la complexité du problème multi-classes initial pour les CNNs et contribue à augmenter la performance des SAADs. Cependant, nous avons observé non seulement que les performances globales de notre méthode d'ensemble profond étaient

2. *Construction d'un système de diagnostic automatique du mélanome – 2.5.*
Conclusion

étroitement corrélées à la précision des modèles de base employés, mais aussi que la faible fréquence des images de lésions atypiques dans la base de données était susceptible de conduire à un biais affectant les performances des modèles de base entraînés sur ces données. Ces observations nous ont naturellement conduit à nous intéresser à la problématique du déséquilibre des classes en apprentissage profond. Le chapitre suivant présente l'étude que nous avons menée sur ce sujet ainsi que la contribution de cette thèse à la résolution de cette problématique.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond

Sommaire

3.1. Introduction	80
3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond	81
3.2.1. Approches du point de vue des données	82
3.2.2. Approches du point de vue de l'algorithme	83
3.2.3. Entraînement découplé	84
3.3. Définition du problème et analyses préliminaires	85
3.3.1. Le déséquilibre des classes	85
3.3.2. Effets du déséquilibre des classes sur l'apprentissage des modèles	86
3.4. Méthodes	91
3.4.1. Motivation théorique	91
3.4.2. Définition des fonctions de coûts	92
3.4.3. Description de la méthode d'entraînement découplé proposée	94
3.4.4. Protocole expérimental	95
3.4.4.1. Bases de données	97
3.4.4.2. Entraînement des modèles	100
3.4.4.3. Analyse comparative	102
3.5. Résultats et discussions	103
3.5.1. Validation sur la base de données ISIC	103
3.5.1.1. Comparaison de notre approche avec l'état de l'art des approches pour le déséquilibre des classes	103
3.5.1.2. Analyse des performances du meilleur modèle	104
3.5.1.3. Comparaison de notre approche avec d'autres SAADs pour la détection des lésions cutanées	104
3.5.1.4. Études ablatives	105
3.5.2. Validation sur les jeux de données CIFAR et iNaturalist	106
3.6. Conclusion	107

3.1. Introduction

L'essor de l'apprentissage profond en général et des CNNs en particulier a permis d'améliorer considérablement les performances des SAADs pour la tâche difficile de détection des lésions cutanées. Cependant, l'un des freins à l'utilisation de tels systèmes dans un contexte clinique réel vient du fait que les jeux de données d'entraînement disponibles présentent une distribution asymétrique qui affecte la performance des modèles entraînés. Cette distribution asymétrique est due au fait que certains types de lésions cutanées sont moins fréquents que d'autres, ce qui se répercute sur les bases de données constituées pour développer des SAADs. En effet, dans la quasi-majorité des jeux de données réels en général, et ceux des lésions cutanées en particulier, il existe des classes (classes de tête) qui contiennent un très grand nombre d'échantillons comparées à d'autres classes (classes de queue) qui sont sous-représentées. La difficulté d'entraîner un modèle sur un jeu de données déséquilibrées est due principalement à deux aspects :

- Premièrement, les méthodes d'apprentissage profond sont avides de données, mais les annotations des classes de queue peuvent être insuffisantes pour entraîner les modèles.
- Deuxièmement, les modèles entraînés sur ces jeux de données ont tendance à privilégier les classes de tête puisqu'elles constituent l'écrasante majorité des échantillons dans l'ensemble de données [154].

A titre d'exemple, le jeu de données public de lésions cutanées ISIC-2018 a un ratio de déséquilibre entre les classes de tête et les classes de queue supérieur à 50. La Figure 3.1 illustre la distribution de ce jeu de données. Dans la littérature, les approches basées sur l'entraînement découplé semblent être les plus performantes pour développer des modèles robustes au déséquilibre des classes [180, 87, 42].

Dans ce chapitre, nous présentons notre méthode d'apprentissage découplé ayant pour objectif de construire des modèles robustes au problème de déséquilibre des classes. Contrairement aux approches usuelles d'apprentissage découplé, notre méthode ne fige ni le classifieur ni l'extracteur de caractéristiques (ou *-features-*) durant l'apprentissage. Seule la définition de la fonction de perte de chaque étape est décisive, suivant un processus d'apprentissage de bout en bout. Pour ce fait, nous avons introduit dans cette étude deux nouvelles fonctions de perte dénommées L_f et L_c capables de répondre aux deux objectifs de l'entraînement découplé sans dissocier l'apprentissage de l'extracteur de caractéristiques et du classifieur du modèle.

Ce chapitre est organisé comme suit : dans la Section 3.2 nous effectuerons une étude bibliographique des approches existantes de résolution du problème de déséquilibre des classes. Dans la Section 3.3, nous présenterons une formulation théorique du problème du déséquilibre des classes et effectuerons des analyses préliminaires enfin d'étudier ses effets. Dans la Section 3.4, nous décrirons la méthodologie employée et ferons une présentation détaillée du protocole expérimental. Dans la Section 3.5 nous présenterons et analyserons les résultats obtenus et la Section 3.6 conclura ce

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond

chapitre.

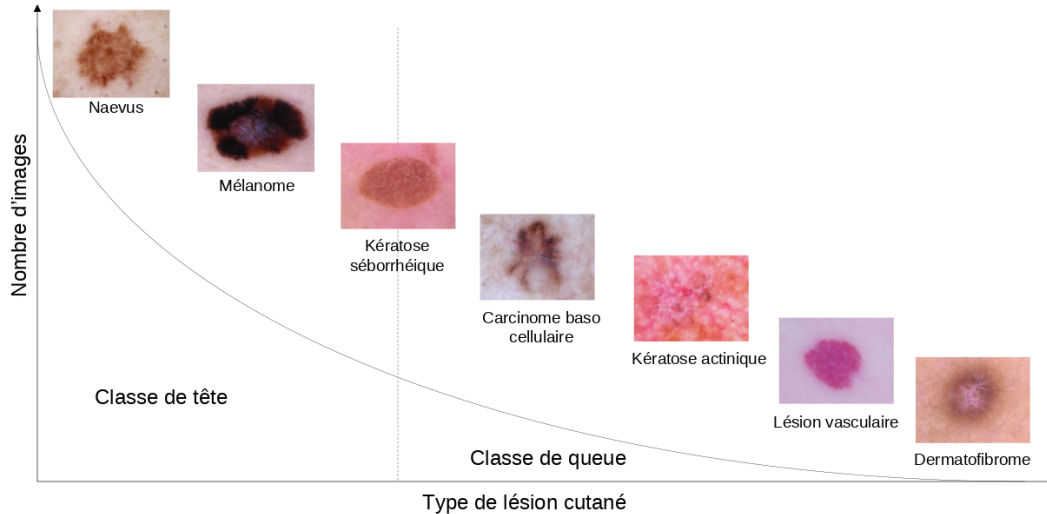


FIGURE 3.1. – Illustration de la distribution du jeu de données publiques ISIC-2018 [31, 157]. Le jeu de données présente une distribution asymétrique avec un ratio entre la classe de tête et la classe de queue approximativement égale à 50.

3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond

Diverses méthodes ont été proposées dans la littérature pour réduire le biais induit sur des modèles entraînés sur des jeux de données déséquilibrés. Les méthodes peuvent être regroupées en trois types d'approches [166, 178] :

- celles du point de vue des données qui opèrent sur le jeu de données d'apprentissage et modifient sa distribution;
- celles du point de vue des algorithmes qui maintiennent le jeu de données d'apprentissage inchangé et ajustent le processus d'apprentissage (ou d'inférence);
- celles basées sur l'entraînement découplé qui divisent le processus d'apprentissage en apprentissage de l'extracteur de caractéristiques et en apprentissage du classifieur.

Dans cette section, nous allons passer en revue chacune des trois approches de résolution du problème de déséquilibre des classes en apprentissage profond, en nous intéressant plus particulièrement à celles présentées dans la littérature pour résoudre cette problématique dans des tâches de vision par ordinateur.

3.2.1. Approches du point de vue des données

Les approches du point de vue des données agissent directement sur les données afin d'obtenir une distribution de classe équilibrée. Ces approches se basent principalement sur des techniques d'échantillonnage telles que le sous-échantillonnage, le sur-échantillonnage ou une combinaison des deux.

Le sur-échantillonnage consiste à générer de nouveaux échantillons de classes minoritaires à partir des données déséquilibrées disponibles. Le sur-échantillonnage aléatoire par exemple, est une stratégie de sur-échantillonnage qui consiste à répliquer de manière aléatoire des instances de la classe minoritaire. Une autre stratégie, appelée sur-échantillonnage ciblé, consiste à rééchantillonner uniquement les échantillons des classes minoritaires proches de la frontière de décision. Ces deux stratégies précédemment évoquées présentent des lacunes majeures. Le sur-échantillonnage aléatoire augmente le risque de sur-apprentissage du modèle ainsi que le coût de calcul. Le sur-échantillonnage ciblé, quant à lui, conduit à une région de décision plus spécifique de la classe minoritaire [27]. La technique de sur-échantillonnage par synthèse des minorités –ou *synthetic minority oversampling technique*– (SMOTE) [56] est un algorithme proposé pour résoudre ces problèmes. SMOTE tente de créer plus de diversités parmi les données des classes minoritaires en générant des échantillons synthétiques. Ces nouveaux échantillons de classes minoritaires sont obtenus en interpolant linéairement les échantillons existants des classes minoritaires. D'autres travaux proposent des méthodes pour améliorer intelligemment les techniques de sur-échantillonnage. Le *Mixup* [175], par exemple, est une méthode qui génère de nouvelles images synthétiques en faisant une combinaison convexe des images présentes dans le jeu de données. *Cutmix* [172] est une autre méthode qui mélange deux images en coupant un patch d'une image et en l'insérant dans une autre. Toujours dans le même ordre d'idée, la méthode *Cutout* [44] met à zéro certains pixels des échantillons existants pour créer de nouvelles images synthétiques. Une autre approche de sur-échantillonnage utilise les Réseaux antagonistes génératifs –ou *generative adversarial networks*– pour générer des échantillons réalistes à partir de classes minoritaires; cependant, non seulement leur entraînement est difficile, mais en plus, elles se généralisent mal sur d'autres ensembles de données [144, 7, 40]. Les méthodes de sur-échantillonnage récentes intègrent un algorithme d'augmentation automatique des données. La méthode *AutoAugment* [36] est la pionnière pour générer artificiellement de nouvelles images en se fondant sur un algorithme d'augmentation automatique. *RandAugment* [37] est une variante d'*AutoAugment* qui ne nécessite pas que la stratégie d'augmentation soit apprise à l'avance à partir de données étiquetées. Plus récemment encore, Muller *et al.* [121] ont proposé une politique d'automatisation de l'augmentation des données qui utilise le produit interne du gradient comme récompense pour optimiser la distribution et l'augmentation des données.

Le sous-échantillonnage est une autre technique courante pour résoudre le déséquilibre des classes. Elle consiste à supprimer les données de la classe majoritaire pour

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond

former un ensemble de données équilibrées. La principale limitation des méthodes de sous-échantillonnage est qu'elles peuvent supprimer des informations critiques requises par le modèle pour apprendre. Ainsi, plusieurs travaux proposent des méthodes pour choisir intelligemment les échantillons majoritaires afin de conserver des informations précieuses pour l'apprentissage. Mani *et al.* [106], par exemple, ont proposé plusieurs algorithmes qui suppriment les échantillons de la classe majoritaire en fonction de leur distance aux échantillons minoritaires prédits par un algorithme des K plus proches voisins –ou *K-nearest neighbors algorithm*– (K-NN).

3.2.2. Approches du point de vue de l'algorithme

Les approches du point de vue des algorithmes visent à ajuster l'apprentissage ou le processus de décision de manière à faciliter la tâche d'apprentissage spécifiquement pour les échantillons de classe minoritaire. Ces approches peuvent être regroupées en trois catégories : l'apprentissage sensible aux coûts, la perte de marge et le seuillage.

L'apprentissage sensible aux coûts consiste à modifier la fonction de coût pour rendre le classifieur plus sensible aux classes minoritaires [51]. La méthode la plus répandue de cette catégorie consiste à pondérer la perte par l'inverse de la fréquence des échantillons de chaque classe contenue dans le jeu de données [14]. Dans le même contexte, Cui *et al.* [38] ont proposé une fonction de coût pondérée par l'inverse des fréquences effectives des classes. De même, Ren *et al.* [138] ont proposé d'utiliser les fréquences d'étiquettes pour ajuster les prédictions du modèle durant l'apprentissage, de sorte que le biais dû au déséquilibre de classe puisse être atténué par des connaissances a priori. Lin *et al.* [97] ont proposé une version reformulée de la [fonction de perte d'entropie croisée –ou *cross entropy loss function*–](#) (CE) dénommée **FL** qui ajoute un facteur réduisant la contribution sur le coût total des échantillons correctement classés. Dans le même ordre d'idée, Tan *et al.* [153] ont proposé une fonction de perte qui réduit les contributions, sur le coût total, des échantillons de la classe majoritaire afin de permettre au modèle d'apprendre de meilleures caractéristiques discriminantes pour les échantillons de la classe minoritaire. Plus récemment, Park *et al.* [125] ont proposé une méthode intuitive qui utilise directement les fréquences d'étiquettes des échantillons d'apprentissage pour la pondération des pertes et ajuste cette pondération en fonction de l'influence de l'échantillon.

La perte de marge est une autre catégorie des approches du point de vue des algorithmes. Elle consiste à contraindre les classes minoritaires à avoir des marges plus importantes. Les travaux de Cao *et al.* [25], par exemple, proposent une fonction de perte de marge dénommée fonction de perte de marge liée à la distribution d'étiquettes –ou *label-distribution-aware margin loss function*– (LDAM). LDAM minimise une limite de généralisation basée sur la marge. De même, Menon *et al.* [113] ont proposé une modification de la fonction CE de manière à encourager une grande marge relative entre des paires d'étiquettes rares et dominantes. D'autres méthodes basées sur la perte de marge sont également proposées et étudiées dans [95, 83].

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.2. Revue des approches de résolution du problème de déséquilibre des classes en apprentissage profond

La dernière catégorie des approches du point de vue des algorithmes est le seuillage [183]. Dans des tâches de classification automatique, un seuil de 0,5 est généralement utilisé pour prédire la classe d'un échantillon à partir de la probabilité de prédiction générée par le classifieur. L'idée centrale derrière les méthodes de seuillage est d'ajuster le seuil de décision du classifieur pour obtenir de meilleures performances. L'article de revue de Buda *et al.* [24] fait un état de l'art des méthodes de seuillage appliquées en apprentissage profond.

3.2.3. Entraînement découplé

L'entraînement découplé est une autre approche existante permettant de résoudre le problème de déséquilibre des classes. Il consiste à disjoindre l'apprentissage d'un modèle en deux étapes : une étape d'apprentissage de représentation (première étape) et une étape d'apprentissage du classifieur (deuxième étape) [166, 178]. Cette approche semble atteindre de meilleures performances que les deux autres approches citées précédemment [178]. Kang *et al.* [87] sont les pionniers à avoir introduit l'entraînement découplé en apprentissage profond. Ils ont pour cela entraîné des modèles avec différentes stratégies d'échantillonnage pour apprendre de meilleures représentations des classes. Par la suite, ils ont étudié trois approches pour obtenir un classifieur avec des frontières de décision équilibrées, en prenant en compte les représentations précédemment apprises. Ces approches étaient : 1) réentraîner le classifieur linéaire paramétrique de manière à équilibrer les classes ; 2) utiliser un classifieur K-NN non paramétrique ; et 3) normaliser les poids du classifieur afin d'ajuster directement l'amplitude des poids pour qu'ils soient plus équilibrés. Leur approche a permis d'établir un nouvel état de l'art sur trois applications de vision par ordinateur qui présente une distribution asymétrique de leurs classes. Desai *et al.* [43] ont également appliqué l'entraînement découplé, ceci pour une tâche de génération de graphes scéniques. Les résultats qu'ils ont obtenus démontrent l'efficacité de l'entraînement découplé dans la gestion du déséquilibre de classe dans l'apprentissage des relations visuelles. D'autres études ont innové le processus d'entraînement découplé en améliorant l'étape d'entraînement du classifieur. Zhang *et al.* [177], par exemple, ont appliqué une couche supplémentaire pour calibrer le classifieur original en faisant correspondre la distribution des prédictions avec une distribution relativement équilibrée des classes. Dans le même contexte, Wang *et al.* [160] ont proposé une stratégie d'alignement de distribution unifiée pour la reconnaissance visuelle à partir d'un jeu de données déséquilibrées. Leur approche transfère les statistiques des classes sur-représentées pertinentes pour déduire la distribution des classes sous-représentées dans la deuxième étape.

Malgré le gain de performance apporté par l'entraînement découplé, un certain nombre de problématiques persistent et doivent être résolues. Premièrement, le choix de la bonne fonction de perte pour obtenir le meilleur extracteur de caractéristiques reste peu discuté. Deuxièmement, les méthodes de ré-échantillonnage (ou de pondé-

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.3. Définition du problème et analyses préliminaires

ration) adoptées durant la deuxième étape de l’entraînement découplé présentent encore certaines limites. En particulier, le biais induit par les classes sur-représentées [149]. Enfin et non des moindres, l’apprentissage découplé contredit l’expectation d’un entraînement de bout en bout attendu en apprentissage profond [166].

Dans cette étude, nous nous sommes inspirés de l’entraînement découplé afin de développer une méthode d’apprentissage des modèles robustes au déséquilibre des classes. La méthode proposée tente de résoudre chacune des trois problématiques de l’entraînement découplé mentionnées précédemment.

La section suivante présente une définition formelle du problème de déséquilibre des classes et analyse les effets de ce déséquilibre sur l’apprentissage des modèles

3.3. Définition du problème et analyses préliminaires

Dans la section précédente, nous avons fait un état de l’art des méthodes de résolution du problème de déséquilibre des classes en apprentissage profond. De cette étude bibliographique, il en est ressorti que l’ensemble des méthodes existantes peuvent être regroupées en trois types d’approches : les approches du point de vue des données, les approches du point de vue de l’algorithme, et les approches basées sur l’entraînement découplé. Parmi les trois approches précédemment citées, l’entraînement découplé est celle qui semble atteindre les meilleures performances.

Dans cette section, nous allons nous intéresser à la formalisation de la problématique du déséquilibre des classes. Nous allons également étudier les effets qu’induisent les jeux de données déséquilibrés sur l’apprentissage des modèles. Ces analyses préliminaires, étayées par des expérimentations compréhensibles, nous permettront d’introduire notre méthode d’entraînement découplé proposée pour résoudre le problème de déséquilibre des classes.

3.3.1. Le déséquilibre des classes

Nous considérons un jeu de données $D = (x_i, y_i)_{i=1}^N$ avec N échantillons d’apprentissage et C classes, x_i est un échantillon d’image de D et $y_i \in 1, 2, \dots, C$ est son étiquette de vérité terrain. On note D_k un sous-ensemble de D contenant tous les échantillons appartenant à la classe k . N_k représente le nombre d’échantillons de D_k . D est considéré comme un jeu de données déséquilibrées si nous avons $N_1 \geq N_2 \geq \dots \geq N_C$ et $N_1 \gg N_C$ après avoir ordonné N_k . Les classes pour lesquelles le nombre d’échantillons est proche de N_1 sont qualifiées de classes de tête. A contrario, celles pour lesquelles le nombre d’échantillons est proche de N_C sont qualifiées de classes de queue. Résoudre le problème de déséquilibre des classes revient donc à apprendre un modèle sur un jeu de données d’entraînement déséquilibrées qui se généralise bien sur un jeu de

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.3. Définition du problème et analyses préliminaires

données de test à la fois sur des classes de tête que sur des classes de queue.

Soit $M(x_i, w)$ un modèle CNN paramétré par w . Dans sa forme la plus générale, M contient deux composantes : un extracteur de caractéristiques $f(x_i) = x'_i$ et un classifieur $h(x'_i) = z_i$, où x'_i désigne les caractéristiques profondes extraites de l'image d'entrée x_i et z_i désigne la sortie du classifieur. La probabilité de prédiction p_i est généralement calculée à l'aide de la fonction *Softmax* appliquée à z_i . L'extracteur de caractéristiques est constitué de plusieurs couches empilées de convolution et de sous-échantillonnage, ainsi que de fonctions d'activation. Le classifieur est quant à lui construit à l'aide de couches FCs. L'extracteur de caractéristiques $f(x_i)$ est conçu pour apprendre les représentations hiérarchiques des caractéristiques de x_i , tandis que le classifieur a pour rôle d'interpréter les caractéristiques extraites x'_i et d'effectuer la tâche de classification.

Deux raisons peuvent expliquer pourquoi il est difficile d'entraîner $M(x_i, w)$ sur un jeu de données déséquilibrées. Tout d'abord, le nombre d'échantillons des classes de queue est petit, ce qui rend difficile l'entraînement de l'extracteur de caractéristiques $f(x_i)$ de manière à effectuer une bonne représentation des caractéristiques des classes de queue. Deuxièmement, la sur-représentation des classes de tête conduit le classifieur $h(x'_i)$ à avoir des probabilités de prédiction bien plus élevées pour les classes de tête par rapport aux classes de queue. L'objectif de l'étude présentée dans ce chapitre est de résoudre ces difficultés.

3.3.2. Effets du déséquilibre des classes sur l'apprentissage des modèles

Afin d'étudier les effets qu'induisent les jeux de données déséquilibrées sur l'apprentissage des modèles, nous avons mené deux expérimentations. Le jeu de données ISIC-2018 ainsi que le modèle *EfficientNetB3* ont été utilisés pour réaliser ces expérimentations.

La première expérience a pour objectif d'évaluer entre la fonction CE, couramment utilisée en vision par ordinateur, et sa version pondérée fonction de perte d'entropie croisée pondérée par l'inverse des fréquences de classe –ou *cost sensitive cross entropy loss function*– (CS) laquelle permet d'obtenir une meilleure représentation des caractéristiques profondes de la tâche. La deuxième expérience a pour objectif d'analyser comment un jeu de données déséquilibrées influence la distribution des probabilités de prédiction d'un modèle $M(x_i, w)$ durant son apprentissage.

La première expérience a consisté à entraîner *EfficientNetB3* avec les fonctions CE et CS pendant 50 époques. Pour la deuxième expérience, nous avons entraîné le réseau uniquement avec la fonction CE pendant 100 époques. Durant cette dernière expérience, nous avons également employé l'arrêt précoce de l'apprentissage comme

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.3. Définition du problème et analyses préliminaires

technique de régularisation pour éviter tout sur-apprentissage. Les résultats de ces expérimentations présentées ci-dessous ont été obtenus sur le jeu de données de validation. Le lecteur se référera à la [Sous-section 3.4.4](#) pour une description détaillée du protocole expérimental utilisé.

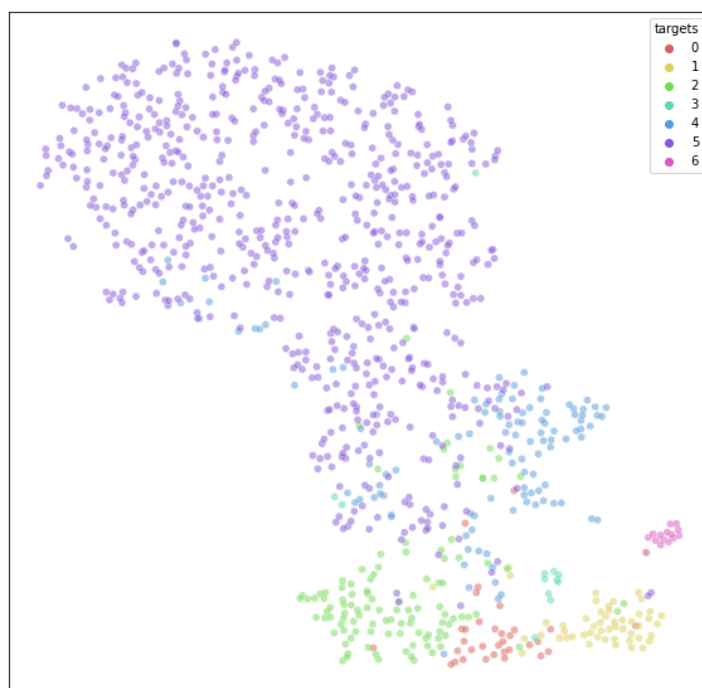


FIGURE 3.2. – Visualisation de la représentation des caractéristiques profondes apprises par le modèle *EfficientNetB3* entraîné à l'aide de la fonction de perte CE.

Les Figure 3.2 et 3.3 illustrent les représentations des caractéristiques profondes générées par un modèle *EfficientNetB3* entraîné respectivement avec les fonctions CE et CS. La visualisation en 2-dimensions des caractéristiques a été générée à l'aide de l'algorithme *t-Distributed stochastic neighbor embedding* (t-SNE). Comme le montre la Figure 3.3, la frontière de décision entre les classes pour le réseau entraîné avec la fonction CS n'est pas nette. Les points caractéristiques proches de la frontière de décision se chevauchent, ce qui conduit à de nombreux faux positifs. A l'inverse, dans la Figure 3.2, on observe que les caractéristiques générées par le modèle entraîné avec la fonction CE présente une frontière de décision plus nette dans l'espace des caractéristiques à 2-dimensions. Ces observations suggèrent que les fonctions de perte non pondérées permettent de construire des modèles avec des extracteurs de caractéristiques plus fidèles à la tâche de classification à réaliser.

Nous visualisons sur les figures 3.5 et 3.4 les distributions des probabilités durant l'apprentissage des échantillons appartenant respectivement à la classe de tête (naevus) et à la classe de queue (dermatofibrome). Comme première observation, nous

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.3. Définition du problème et analyses préliminaires

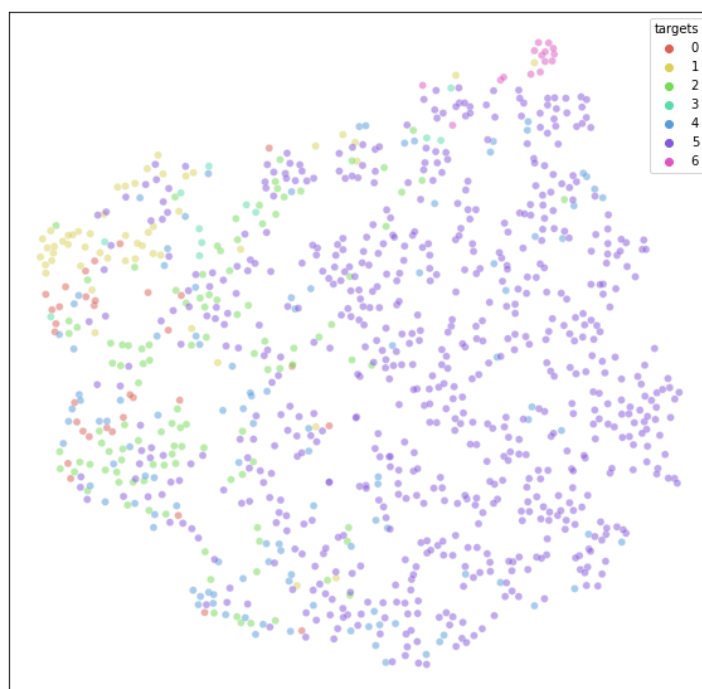


FIGURE 3.3. – Visualisation de la représentation des caractéristiques profondes apprises par le modèle efficientNetB3 entraîné à l’aide de la fonction de perte CS.

pouvons noter qu’à l’initialisation du modèle, toutes les probabilités ont des valeurs comprises dans l’intervalle $[0.1, 0.3]$. Ceci s’explique par le fait que les neurones du classifieur sont initialisés en considérant que toutes les classes sont équiprobables (dans notre cas nous avons sept classes), donnant ainsi des probabilités autour de $0,14$. Ensuite, on observe que l’apprentissage de la classe naevus semble aisé. En effet, durant l’apprentissage du modèle, les probabilités de prédiction des échantillons appartenant à la classe naevus croissent très rapidement avec une convergence approximativement atteinte dès l’époque 20. En revanche, l’apprentissage des échantillons appartenant à la classe dermatofibrome semble plus difficile. On observe en début d’apprentissage que le modèle a du mal à discriminer les échantillons de cette classe avec leurs probabilités de prédiction qui oscillent dans l’intervalle $[0.0, 0.3]$ jusqu’à l’époque 22 environ. A partir de l’époque 22, le modèle commence à mieux discriminer cette classe avec des probabilités de prédiction qui croissent progressivement avec une convergence qui commence à être atteinte autour de l’époque 52. De plus, malgré le début de convergence du modèle, ils demeurent des échantillons ayant des probabilités figées autour de l’intervalle $[0.0, 0.2]$. Ces échantillons peuvent être assimilés à des exemples très difficiles, voire aberrants, et pour lesquels le modèle peut s’en passer pour se concentrer sur des échantillons plus discriminants.

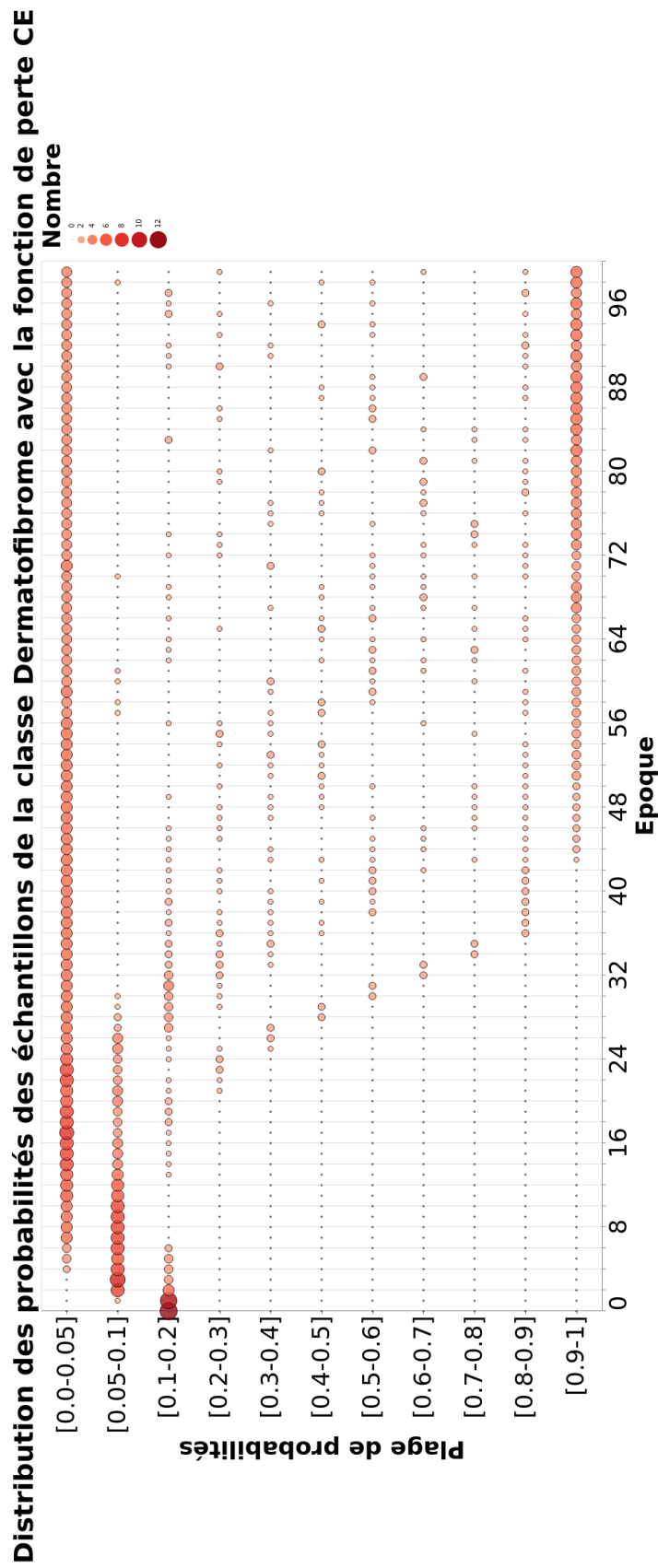


FIGURE 3.4. – Distribution des probabilités de prédiction des échantillons de la classe de queue Dermatofibrome durant l'apprentissage du modèle *EfficientNetB3*. Les échantillons présentés ici appartiennent au jeu de données de validation.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.3. Définition du problème et analyses préliminaires

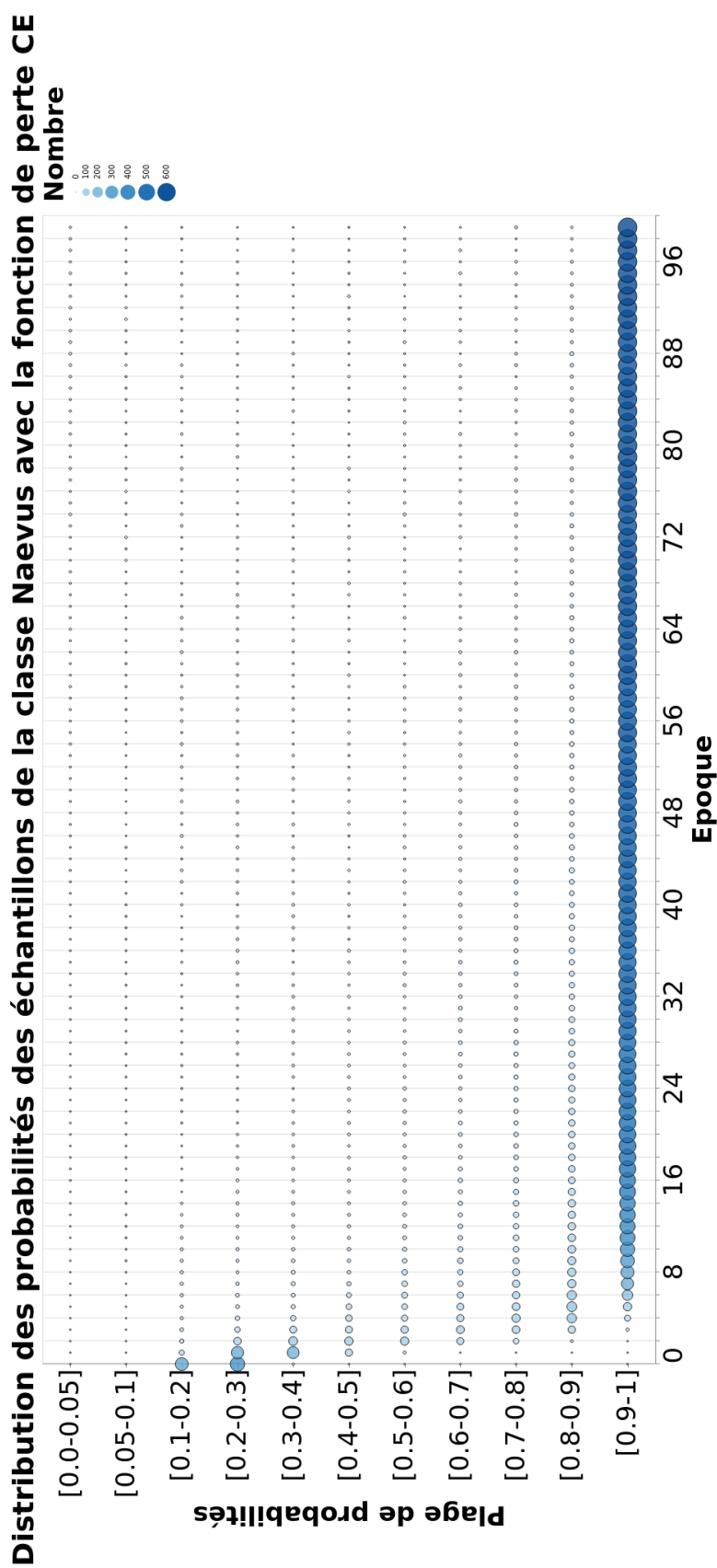


FIGURE 3.5. – Distribution des probabilités de prédiction des échantillons de la classe de tête Naevus durant l'apprentissage du modèle EfficientNetB3. Les échantillons présentés ici appartiennent au jeu de données de validation.

Dans la section suivante, nous allons décrire notre méthode permettant de résoudre le problème de déséquilibre de classe à la lumière des analyses préliminaires présentées ci-dessus. Le protocole expérimental employé sera également détaillé afin de faciliter sa reproduction par la communauté scientifique.

3.4. Méthodes

Dans cette section, nous faisons une description détaillée de notre méthode d'entraînement découplé proposée pour résoudre le problème de déséquilibre des classes.

3.4.1. Motivation théorique

Les travaux de Li *et al.* [95] assimilent le problème de déséquilibre des classes à la difficulté à classer une classe par rapport à une autre. Dans le même contexte, les travaux sur l'apprentissage profond présentés dans [46] démontrent que les échantillons difficiles possèdent plus d'informations discriminantes que des échantillons faciles. En effet, si un échantillon est bien classé, c'est un échantillon facile et le modèle en tire très peu d'information discriminante pour la tâche à effectuer. L'apprentissage des modèles devrait donc s'appesantir sur des échantillons difficiles, jugés plus informatifs quelle que soit la classe à laquelle ils appartiennent. En outre, Li *et al.* [95] révèlent également que la densité des échantillons ayant de larges gradients (échantillons très difficiles) est légèrement supérieure à la densité des échantillons ayant des gradients moyens. Ils ont observé que les échantillons avec de larges gradients affectent la stabilité du modèle car leurs gradients peuvent présenter un écart important par rapport aux autres échantillons. De plus, ces échantillons existent de manière stable même après que le modèle ait convergé. Ces observations leur ont permis de conclure que des échantillons avec des gradients larges sont assimilables à des valeurs aberrantes. Cette conclusion rejoint l'observation faite durant notre analyse préliminaire des effets induits par les jeux de données déséquilibrés sur l'apprentissage des modèles (voir [Sous-section 3.3.2](#)). Il est possible que ces valeurs aberrantes puissent correspondre à des données mal étiquetées.

Inspirés des approches basées sur l'entraînement découplé (voir [Sous-section 3.2.3](#)) et des observations issues de l'analyse préliminaire présentée dans la [Sous-section 3.3.2](#), nous avons développé une nouvelle méthode d'apprentissage permettant de construire des modèles robustes au déséquilibre des classes. La méthode proposée consiste en un entraînement en deux étapes basées sur deux nouvelles fonctions de perte désignées par L_f et L_c . La Figure 3.6 illustre la distribution de coût en fonction de la probabilité prédite des fonctions de perte [CE](#) (courbe bleue), [FL](#) (courbe rouge) et les fonctions L_f (courbe verte) et L_c (courbe cyan).

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

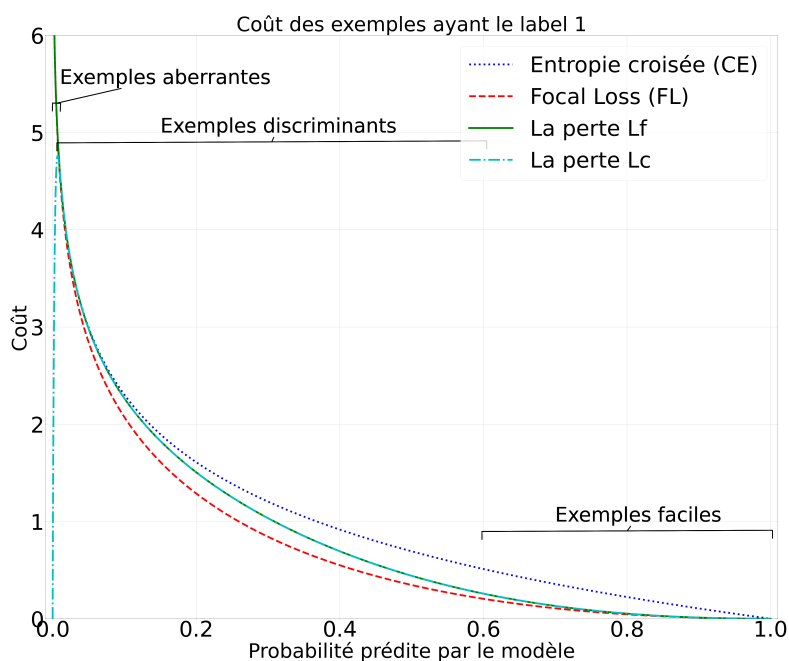


FIGURE 3.6. – Illustration des distributions des contributions sur le coût total des deux fonctions introduites L_f et L_c comparées à celles des fonctions CE et FL.

3.4.2. Définition des fonctions de coûts

Avant d'introduire les deux fonctions de perte L_f et L_c , nous devons revoir quelques notions notamment la formulation de la fonction CE, mais aussi la définition de la notion de difficulté d'un échantillon.

Soit un échantillon d'image x_i , et p_i sa probabilité générée par un modèle. p_i est obtenue en appliquant la fonction Softmax à la sortie z_i du modèle. y_i représente l'étiquette de vérité terrain de x_i . Nous désignerons par $p_c \in [0, 1]$ la probabilité générée par un modèle que x corresponde à son étiquette c .

Un échantillon facile x_i est un échantillon pour lequel le modèle a attribué une probabilité élevée ($p_c > 1 - \exp(-\eta)$, $\eta > 0$ avec η grand). A l'inverse, lorsque la probabilité attribuée par le modèle est faible (p_c autour de 0.5), l'échantillon est qualifié de difficile. Les échantillons très difficiles (ou exemples aberrants ou mal étiquetés) sont des échantillons pour lesquelles le modèle a attribué des probabilités très faibles ($p_c < \exp(-\eta)$, $\eta > 0$ avec η grand).

La fonction CE est définie par l'équation suivante :

$$CE(\mathbf{y}, \mathbf{p}) = - \sum_i y_i \log(p_i) \quad (3.1)$$

Nous introduisons à présent les fonctions L_f et L_c comme une meilleure alternative

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

aux fonctions de perte précédemment utilisées dans le cadre de l'entraînement découplé. La fonction de perte Lf est utilisée lors de la première phase d'apprentissage de l'entraînement découplé. Son objectif est de guider l'extracteur de caractéristiques à mieux représenter les caractéristiques de la tâche. La fonction de perte Lc est utilisée lors de la deuxième phase d'entraînement, elle a pour objectif de résoudre plus spécifiquement le problème de déséquilibre de classe. Les deux phases d'apprentissage se déroulent de bout en bout.

Comme décrit dans la [Sous-section 3.3.2](#), une bonne initialisation du réseau passe par la réduction de la contribution des échantillons faciles afin d'éviter que les classes de tête soient sur-représentées dans le coût total d'apprentissage. De plus, il est désirable de maintenir une contribution élevée pour les échantillons difficiles, afin d'aider le classifieur à mieux discriminer les classes de queue. En outre, les fonctions de coût non pondérées semblent les plus appropriées pour l'apprentissage de l'extracteur de caractéristiques. La fonction Lf répond aux trois critères ci-dessus mentionnés. Elle est définie par l'équation suivante :

$$Lf(p, y)(\mathbf{y}, \mathbf{p}) = - \sum_i w_i y_i \log(p_i) \quad (3.2)$$

le terme w_i dans l'équation précédente est calculé par :

$$\frac{\sin(\pi p_i)}{\pi p_i} \quad (3.3)$$

La fonction Lf est continue et dérivable. Le calcul des gradients permettant l'apprentissage du modèle s'effectue en dérivant Lf par rapport à l'entrée p_i à l'aide de l'équation suivante :

$$\nabla_{p_i} Lf = \frac{(\log(p_i) - 1) \sin(\pi p_i) - \pi p_i \log(p_i) \cos(\pi p_i)}{\pi p_i^2} \quad (3.4)$$

Une fois que le modèle a appris une bonne représentation des caractéristiques, nous devons guider l'apprentissage de son classifieur vers des échantillons discriminants pour le rendre robuste au déséquilibre de classe. Pour cela, nous voulons atténuer la contribution des gradients très importants les empêchant d'affecter la convergence du modèle. Ceci a pour effet d'aider le modèle à mieux se focaliser sur les échantillons difficiles (contenant plus d'informations). Poursuivant ce but, nous avons modifié la fonction Lf initiale en soustrayant le terme $\frac{\sin(\pi p_i)}{\pi p_i}$ de w_i par un autre sinus cardinal de fréquence plus élevée $\frac{\sin(\delta \pi p_i)}{\delta \pi p_i}$. Cette opération nous a permis d'obtenir une distribution de coût réduisant la contribution des échantillons ayant un large gradient. Cependant, la soustraction des deux sinus cardinaux induit également des oscillations susceptibles d'affecter la stabilité de l'apprentissage, c'est la raison pour laquelle nous avons ajouté un facteur d'amortissement à travers la fonction exponentielle pour lisser la courbe. Finalement, la nouvelle fonction de perte Lc qui en résulte est définie par :

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

$$Lc(\mathbf{y}, \mathbf{p}) = - \sum_i \left(\frac{\sin(\pi p_i)}{\pi p_i} - \frac{\sin(\delta \pi p_i)}{\delta \pi p_i} \exp^{-\delta p_i} \right) y_i \log(p_i) \quad (3.5)$$

Lc satisfait les propriétés mathématiques suivantes :

- Lc est continue et dérivable dans l'intervalle $[0, 1]$
- Lorsque le gradient d'un échantillon est très grand, correspondant à $p_c < \exp(-\eta)$, $\eta > 0$ avec η grand, la perte tend vers zéro et le modèle est moins affecté par les valeurs aberrantes.

$$\lim_{p_i \rightarrow 0} Lc(p, y) = 0 \quad (3.6)$$

- Lorsque le gradient d'un échantillon est très faible, correspondant à $p_c > 1 - \exp(-\eta)$, $\eta > 0$ avec η grand, la perte tend vers zéro, ce qui évite au modèle d'être submergé par des échantillons faciles.

$$\lim_{p_i \rightarrow 1} Lc(p, y) = 0 \quad (3.7)$$

Le paramètre δ permet de déplacer le maximum de la fonction Lc , modifiant ainsi le seuil pour lequel un échantillon peut être considéré comme étant aberrant. La variabilité d'un tel seuil est pertinente dans la mesure où la densité des larges gradients est fortement corrélée au jeu de données de la tâche de classification. Le choix de la valeur de δ se fait par recherche de grille.

En pratique, pour la deuxième étape, on utilisera une version pondérée de la fonction Lc . Le terme de pondération α ajouté peut être évalué par une technique de pondération au choix à l'instar de la technique classique de pondération par l'inverse des fréquences des classes [14]. La formulation de la fonction Lc pondérée est définie par l'équation suivante :

$$Lc(\mathbf{y}, \mathbf{p}) = - \sum_i \alpha \left(\frac{\sin(\pi p_i)}{\pi p_i} - \frac{\sin(\delta \pi p_i)}{\delta \pi p_i} \exp^{-\delta p_i} \right) y_i \log(p_i) \quad (3.8)$$

On peut voir sur la Figure 3.6 que la fonction Lf (courbe verte) réduit considérablement la contribution sur le coût total des échantillons faciles similairement à celle générée par la fonction FL (courbe rouge). En parallèle, la fonction Lf maintient une contribution élevée proche de celle générée par la fonction CE (courbe bleue) pour les échantillons difficiles. La fonction Lc (courbe cyan) suit la même distribution que Lf à la seule exception qu'elle réduit en plus la contribution des valeurs aberrantes.

3.4.3. Description de la méthode d'entraînement découplé proposée

La méthodologie globale suivie par notre méthode d'entraînement découplé est décrite par l'Algorithme 2. L'intuition derrière notre méthode est de poursuivre les objectifs de l'entraînement découplé sans dissocier l'entraînement de l'extracteur

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

des caractéristiques et du classifieur. Notre approche se déroule donc en deux étapes. Dans la première étape, un CNN est entraîné à l'aide de la fonction Lf . Cette étape vise à guider l'extracteur des caractéristiques afin d'apprendre une bonne représentation des caractéristiques pour la tâche donnée. La deuxième étape débute lorsque le nombre d'époques d'apprentissage atteint un seuil T . À ce stade, le modèle est ajusté avec la version pondérée de la fonction Lc . Cette étape vise à guider l'apprentissage du classifieur afin de mieux discriminer les classes de tête et de queue. Il est avantageux de définir T correspondant à l'époque à laquelle le modèle commence à converger vers un minimum local. Dans nos expérimentations, nous avons fait correspondre T à l'époque pour laquelle la performance du modèle en termes de BAcc sur le jeu de validation restait statique durant dix époques.

Algorithm 2 Méthodologie de l'entraînement découplé de bout en bout

Require: Base d'entraînement $D = (x_i, y_i)_{i=1}^N$, modèle $f(x, w)$, Nombre d'époque T , Epoque de transition T , Taille mini-batch m , Taille de la base N , facteur de réduction τ

Initialisation aléatoire de w

for $t=1$ to T **do**

$D_m \leftarrow \text{SampleMiniBatch}(D, m)$

$L(w) \leftarrow \frac{1}{m} Lf(f(x, y); y)$

$w_t \leftarrow w_{t-1} - \eta \nabla_w L(w_{t-1})$

end for

$\eta \leftarrow \eta / \tau$

for $t=T$ to T **do**

$D_m \leftarrow \text{SampleMiniBatch}(D, m)$

$L(w) \leftarrow \frac{1}{m} Lc(f(x, y); y)$

$w_t \leftarrow w_{t-1} - \eta \nabla_w L(w_{t-1})$

end for

3.4.4. Protocole expérimental

Cette sous-section fait une description détaillée du protocole expérimental utilisé pour préparer nos données et entraîner nos modèles. La Figure 3.7 donne un aperçu global du protocole expérimental suivi. Durant la première phase, nous entraînons un CNN avec la fonction de perte Lf pour guider le modèle à apprendre une meilleure représentation des caractéristiques profondes de la tâche. Ensuite, durant la deuxième phase, lorsque le nombre d'époques est supérieur à un seuil prédéfini T , nous ajustons l'entraînement du modèle à l'aide de la version pondérée de Lc pour effectuer la tâche de classification finale. Le protocole s'achève par une évaluation sur la base de test.

Les expériences présentées dans cette étude ont été réalisées à l'aide des ressources HPC/IA de l'Institut du Développement et des ressources en informatique scientifique (IDRIS) alloué par la société Grand Equipement National de Calcul Intensif (GENCI)

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

sous le matricule 2022-AD01101- 2626R1. La bibliothèque Pytorch a été utilisée pour entraîner nos modèles. Le code de cette étude est disponible en [open source](#)¹.

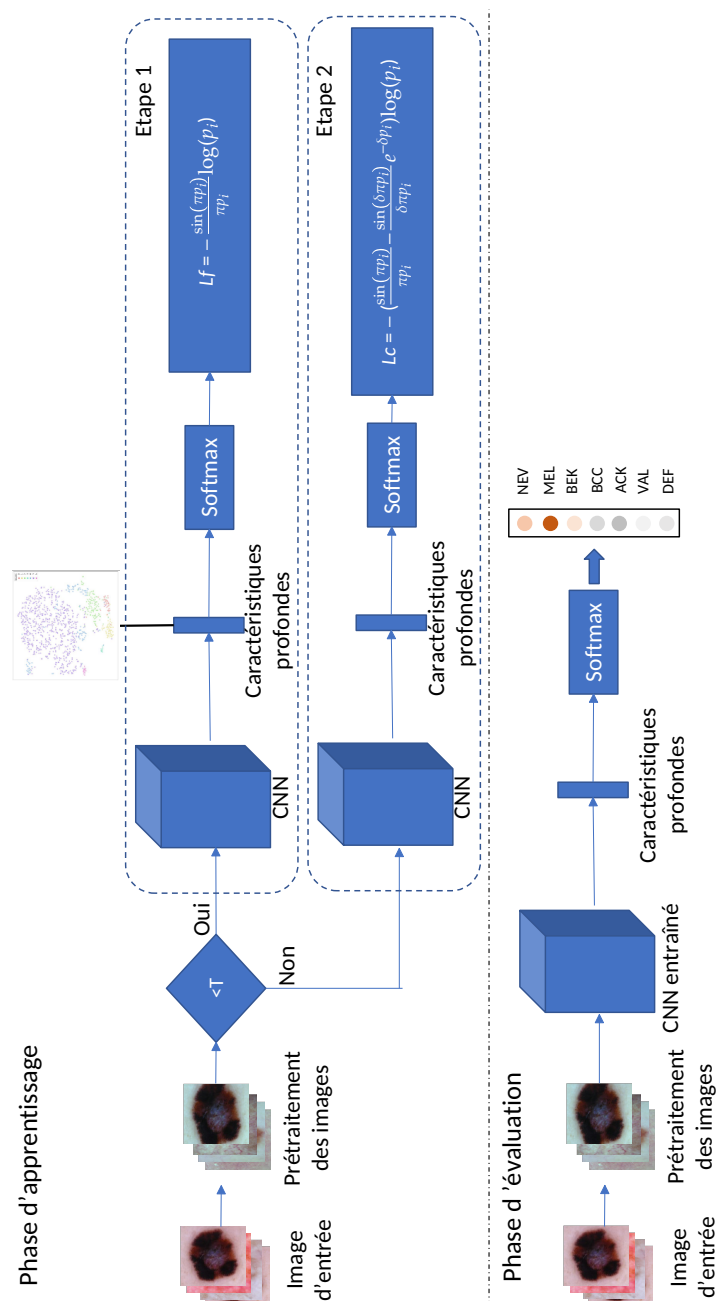


FIGURE 3.7. – Description brève du protocole expérimental de notre méthode.

1. https://github.com/cartelgouabou/End_to_end_decoupled_training_for_skin_lesion_classification

3.4.4.1. Bases de données

En plus de la base de données [ISIC-2018](#) des lésions cutanées, nous avons évalué notre méthode sur d'autres bases de données de référence en vision par ordinateur. Les autres jeux de données utilisés étaient les bases de données publiques [CIFAR \[94\]](#) et [iNaturalist-2018 \[159\]](#). Nous avons intégré ces jeux de données afin d'évaluer la capacité de notre méthode à s'étendre à d'autres tâches de classification plus générales.

La base de données ISIC-2018 : Le jeu de données [ISIC-2018](#) comprend 10 015 images dermoscopiques appartenant à sept types de lésions cutanées : mélanome (MEL), naevus (NEV), carcinome basocellulaire (CBC), kératose actinique (KAC), kératose séborrhéique (KES), dermatofibrome (DEF) et lésion vasculaire (VAL). Des échantillons d'images de chacune des sept catégories présentes dans ce jeu de données sont illustrés dans la [Figure 3.8](#). Comme le montre la [Table 3.1](#), ce jeu de données présente une distribution asymétrique. Pour une analyse approfondie de la problématique du déséquilibre, nous avons regroupé les différentes classes en trois catégories : les classes de tête (naevus, mélanome et kératose séborrhéique), les classes moyennes (carcinome basocellulaire et kératose actinique) et les classes de queue (dermatofibrome et lésion vasculaire). Nous avons utilisé 80% des images comme données d'entraînement, 10% des images comme données de validation et 10% des images comme données de test (voir [Table 3.1](#) pour la répartition) . S'agissant du prétraitement effectué sur les images, nous avons appliqué la même procédure décrite dans la [Sous-section 2.3.1](#), seule la taille de redimensionnement des images d'entrée a été redéfinie à 300×300 pixels afin de correspondre à la dimension d'entrée de la nouvelle architecture de [CNN](#) utilisée.

TABLEAU 3.1. – Répartition de la base de données ISIC-2018.

	NEV	MEL	KES	CBC	KAC	VAL	DEF	Total
ISIC-2018	6705	1113	1099	514	327	142	115	10 015
Ratio	0.67	0.11	0.11	0.06	0.03	0.01	0.01	1
Entraînement* (80%)	5364	890	879	411	262	114	92	8012
Validation* (10%)	671	111	110	51	33	14	12	1002
Test* (10%)	670	112	110	52	32	14	11	1001

* Répartition obtenue après avoir divisé la base de données initiale. Les abréviations NEV, MEL, KES, CBC, KAC, VAL et DEF correspondent respectivement aux classes naevus, mélanome, kératose séborrhéique, carcinome basocellulaire, kératose actinique, lésion vasculaire et dermatofibrome.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

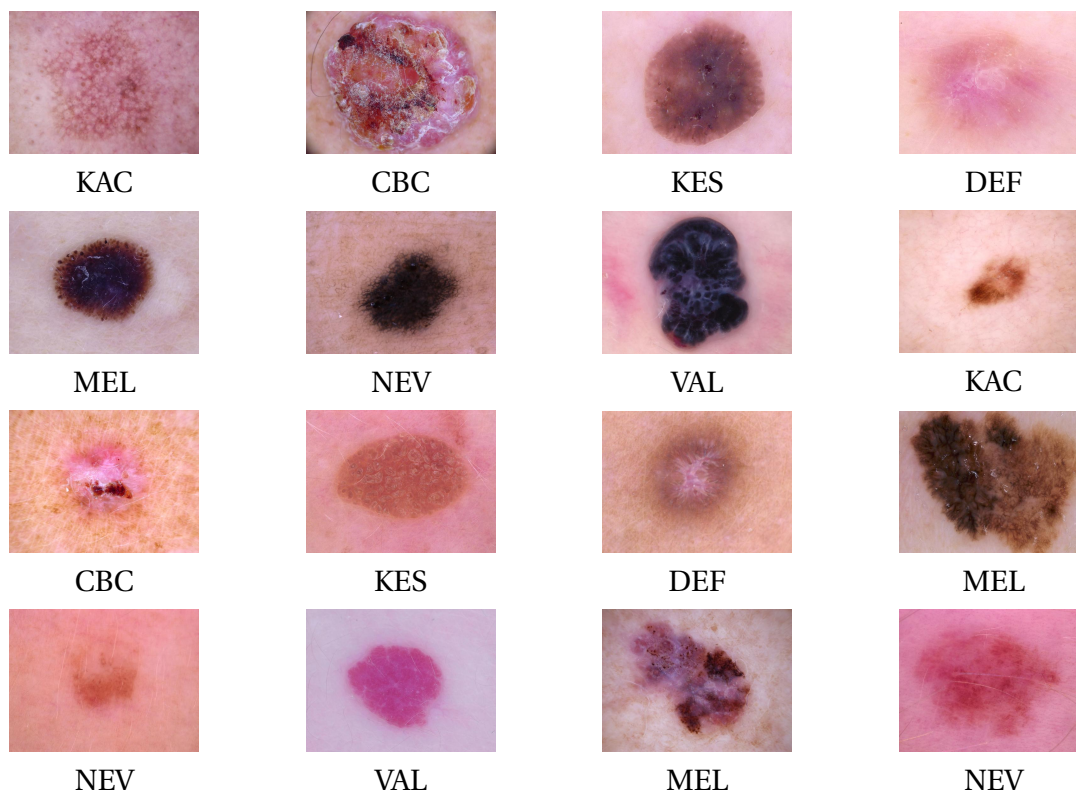


FIGURE 3.8. – Échantillons d'images représentant chacune des sept classes de lésions cutanées présentes dans le jeu de données ISIC-2018. Les abréviations NEV, MEL, KES, CBC, KAC, VAL et DEF correspondent respectivement aux classes naevus, mélanome, kératose séborrhéique, carcinome basocellulaire, kératose actinique, lésion vasculaire et dermatofibrome.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

Les bases de données CIFAR : Il existe deux versions du jeu de données publique **CIFAR**, **CIFAR-10** et **CIFAR-100**. Les jeux de données **CIFAR-10** et **CIFAR-100** contiennent tous les deux 50 000 images d'entraînement et 10 000 images de validation de taille 32×32 pixels. La base de données **CIFAR-10** contient 10 classes tandis que la base de données **CIFAR-100** compte 100 classes. La Figure 3.9 donne quelques échantillons d'images des classes contenues dans le jeu de données **CIFAR-10**. Leurs versions originales ne présentent pas de distribution asymétrique. Conformément aux travaux existants traitant le même sujet, nous avons artificiellement créé des versions déséquilibrées des jeux de données **CIFAR**. Ceci est effectué en réduisant la taille de chaque classe dans le jeu de données d'apprentissage tout en conservant intacte le jeu de données de test [38]. Le niveau de déséquilibre des données est défini par ρ , qui représente le ratio entre le nombre d'échantillon de la classe de tête et le nombre d'échantillon de la classe de queue. Pour nous assurer que notre méthode s'applique à une variété de contextes, nous avons évalué notre méthode sur deux valeurs de ρ : 100 et 50.

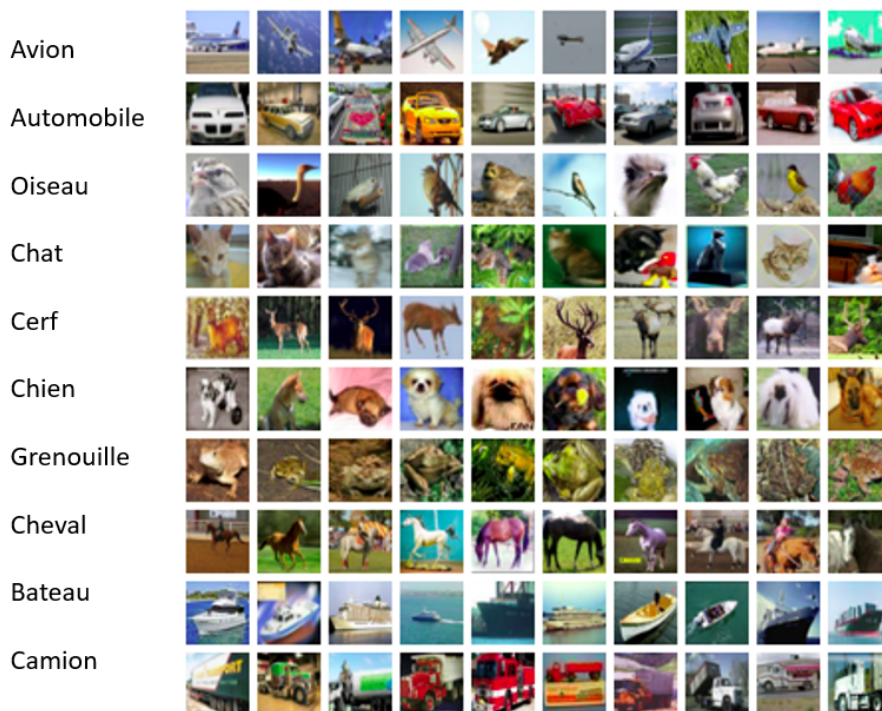


FIGURE 3.9. – Illustration des dix classes présentes dans le jeu de données **CIFAR-10**.

La base de données iNaturalist-2018 : **iNaturalist-2018** est un jeu de données publiques pour la tâche de classification et de détection des espèces. Il s'agit d'un jeu de données présentant naturellement une distribution asymétrique à longue queue. Le ratio ρ de déséquilibre de cette base est approximativement égal à 500. Le jeu de données **iNaturalist-2018** contient 437 513 images d'entraînement et 24 426 images de test appartenant à 8 142 classes qui peuvent être regroupées en 13 super-classes. La

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

Figure 3.10 donne quelques échantillons d'images des super-classes contenues dans le jeu de données iNaturalist-2018. Dans nos expériences, nous avons conservé cette répartition d'origine.



FIGURE 3.10. – Exemples d'images du jeu de données iNaturalist-2018. Chaque ligne affiche des images choisies au hasard dans chacune des 13 super-classes. (Source : [159])

3.4.4.2. Entraînement des modèles

Afin d'avoir une comparaison aussi fidèle que possible avec les méthodes existantes, nous avons pour chacun des jeux de données reproduit autant que possible les méthodes de préparation de base de données, de prétraitement des images, les architectures et les algorithmes d'apprentissage des travaux utilisant la base de données en question.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

TABLEAU 3.2. – Structure détaillée de l’architecture EfficientNetB3 utilisée.

Bloc N°	Nom de la couche	Résolution	Taille du filtre	Nombre de couches
1	Conv	300 × 300	3 × 3	1
2	MBCConv1	150 × 150	3 × 3	2
3	MBCConv6	150 × 150	5 × 5	3
4	MBCConv6	75 × 75	3 × 3	3
5	MBCConv6	38 × 38	3 × 3	5
6	MBCConv6	19 × 19	5 × 5	5
7	MBCConv6	10 × 10	5 × 5	6
8	MBCConv6	10 × 10	3 × 3	2
9	Conv	10 × 10	1 × 1	1
10	Global pooling	10 × 10		1
11	FC	10 × 10		1

Modèles entraînés sur le jeu de données ISIC : Nous avons utilisé le modèle pré-entraîné EfficientNetB3[155] pour mener toutes nos expérimentations. Seuls les neurones de la dernière couche FC ont été modifiés pour adapter le modèle à une tâche de classification comprenant sept classes. Le numéro de chaque bloc, le nom et la taille des couches de convolution par bloc correspondant, la taille de chaque filtre d’échantillonnage et le nombre de couches de l’architecture EfficientNetB3 utilisées sont décrits dans le Tableau 3.2. Nous avons utilisé l’algorithme d’optimisation Adam avec les paramètres suivants : $\beta_1 = 0,9$, $\beta_2 = 0,999$, $\epsilon = 10^{-7}$ et $amsgrad = False$. Tous les modèles ont été entraînés avec une taille de lot de 128 durant 100 époques. Similairement à la procédure suivie dans [129], nous avons utilisé le taux d’apprentissage cyclique proposé par [150] pour varier dans la plage de 0,001 à 0,00001 le taux d’apprentissage. Nous avons également appliqué la technique de régularisation d’arrêt précoce qui consiste à arrêter l’apprentissage du modèle tôt afin d’éviter le surapprentissage. Durant nos expérimentations, nous avons arrêté l’apprentissage du modèle dès que la performance en termes de BAcc sur le jeu de données de validation ne s’améliorait pas au bout de 20 époques et avons enregistré les paramètres ayant permis d’obtenir précédemment le score le plus élevé. La meilleure valeur obtenue par recherche sur grille de l’hyperparamètre δ était de 10^7 .

Modèles entraînés sur les jeux de données CIFAR : S’agissant de nos expériences sur les bases de données CIFAR, nous avons appliqué le prétraitement effectué dans [25]. Spécifiquement, les images d’entraînement et de test ont été normalisées dans l’intervalle [0, 1]. Durant l’apprentissage, quatre pixels d’une valeur de zéro ont été ajoutés à chaque côté de l’image, suivi d’un rognage aléatoire de taille 32 x 32 pixels. De plus, nous avons utilisé durant l’apprentissage des techniques de génération artificielle de données à l’instar de la rotation aléatoire, du décalage en largeur et en

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.4. Méthodes

hauteur, du retournement horizontal, mais aussi en utilisant la technique de génération artificielle *RandAugment* [37]. Nous avons utilisé l’architecture ResNet-32 pour réaliser toutes les expériences sur cette base de données et avons utilisé l’algorithme d’optimisation SGD avec un Momentum de 0,9 et une décroissance de poids de 2×10^{-4} . Les paramètres du réseau ont été initialisés en utilisant la méthode de Kaiming [81]. Tous les modèles ont été entraînés avec une taille de lot de 128 durant 200 époques. Similairement à la méthode suivie dans [38], nous avons linéairement augmenté le taux d’apprentissage durant les cinq premières époques, puis nous avons fixé notre taux d’apprentissage initial à 0,1 pour le nombre d’époques restant. Dans le cas où l’expérimentation suivait notre méthode d’apprentissage découplé, nous avons d’abord entraîné le modèle pendant 100 époques avec la fonction de coût L_f , puis nous avons ajusté l’entraînement avec la fonction de coût L_c pour les 100 époques restantes.

Modèles entraînés sur le jeu de données iNaturalist-2018 : Pour les expérimentations sur la base de données iNaturalist-2018, nous avons suivi la même stratégie de prétraitement que celle utilisée dans [25]. Spécifiquement, chaque image est redimensionnée à une taille de 256 x 256 pixels. Puis, un rognage aléatoire de taille 224 x 224 pixels est effectué sur l’image ou sur sa version ayant subi un retournement horizontal. Nous avons utilisé sur cette base de données l’architecture ResNet-50 pour réaliser toutes les expériences. Comme pour l’implémentation présentée précédemment sur ResNet-32, nous avons utilisé l’algorithme d’optimisation SGD avec un Momentum de 0,9 et une décroissance de poids de 2×10^{-4} . Nous avons entraîné les modèles durant 120 époques avec un taux d’apprentissage initial de 0,01. Le taux d’apprentissage est actualisé durant l’apprentissage en le multipliant par un facteur de 0,1 après 30 et 60 époques. Dans le cas où l’expérimentation suivait notre méthode d’apprentissage découplé, nous avons commencé par entraîner le modèle avec la fonction de coût L_f , puis à l’époque 60, nous avons poursuivi l’entraînement du modèle en utilisant la fonction L_c combinée à la technique de pondération par l’inverse des fréquences des classes. En plus de l’augmentation simple des données artificielles, nous avons également appliqué la technique de génération artificielle *RandAugment*.

3.4.4.3. Analyse comparative

Pour évaluer notre méthode, nous l’avons comparé à l’état de l’art des méthodes proposées pour traiter le problème de déséquilibre de classe en apprentissage profond. Les méthodes avec lesquelles nous avons comparés notre approche sont : la fonction CS [14], la fonction FL [97], la fonction de perte pondérée par le nombre effectif de classe –ou *class balanced loss*– [38], la fonction LDAM [25], la fonction de perte pondérée par l’influence des classes –ou *influence balanced loss*– [125], la méthode proposée par Zhang *et al.* [179] et la méthode d’entraînement découplé proposée dans [180]. Nous avons également comparé notre méthode avec d’autres travaux existants [109, 67, 168, 64, 15] appliqués à la tâche de classification multi-classes des lésions cutanées à partir de la base de données ISIC-2018.

3.5. Résultats et discussions

Cette section présente et discute les résultats obtenus avec notre méthode d’entraînement découplé. Étant donné que l’apprentissage d’un CNN suit un processus stochastique, nous avons pour certains résultats exécutés chacune de nos expériences à dix reprises avec des paramètres d’initialisation aléatoires. Ceci est effectué afin de nous assurer de la robustesse des résultats présentés. Chacun des résultats obtenus est présenté, par la suite, sous forme de moyenne avec son écart-type associé.

3.5.1. Validation sur la base de données ISIC

3.5.1.1. Comparaison de notre approche avec l’état de l’art des approches pour le déséquilibre des classes

TABLEAU 3.3. – Performances en termes de score de BAcc obtenues par EfficientNetB3 entraîné avec diverses fonctions de perte et notre méthode d’entraînement découplé

Méthodes	Classes de tête	Classes moyennes	Classes de queue	Global
CS [†] [14]	0.76 ± 0.02	0.83 ± 0.02	0.95 ± 0.03	0.84 ± 0.01
CBL [†] [38]	0.79 ± 0.01	0.86 ± 0.01	0.95 ± 0.04	0.85 ± 0.01
FL [†] [97]	0.80 ± 0.01	0.83 ± 0.05	0.88 ± 0.04	0.83 ± 0.01
LDAM [†] [25]	0.76 ± 0.03	0.78 ± 0.01	0.93 ± 0.04	0.82 ± 0.02
IBL [†] [125]	0.83 ± 0.01	0.81 ± 0.04	0.87 ± 0.01	0.82 ± 0.02
Notre méthode	0.81 ± 0.01	0.88 ± 0.02	0.98 ± 0.01	0.87 ± 0.01

Les résultats présentés sont ceux obtenus sur le jeu de données de test. Notre méthode a obtenu la meilleure performance. Le signe † indique que nous avons réimplémenté les méthodes comparées.

Le Tableau 3.3 synthétise les performances des modèles entraînés avec différentes fonctions de coût existantes et avec notre méthode d’apprentissage découplé. Les performances présentées dans ce tableau ont été obtenues sur le jeu de données de test. En analysant les performances obtenues selon que la classe appartienne au groupe des classes de tête, des classes moyennes, ou des classes de queue, nous faisons les observations suivantes : tout d’abord notre approche permet d’améliorer les performances des classes appartenant au groupe des classes moyennes (amélioration de 2%) et au groupe des classes de queue (amélioration de 3%) tout en maintenant une bonne performance pour le groupe des classes de tête ; de plus, notre méthode a obtenu la meilleure performance globale, atteignant un score moyen de BAcc de 87% avec une marge minimale de 2% par rapport aux autres méthodes. Ce résultat suggère que notre approche permet de construire un système robuste au déséquilibre des classes.

3.5.1.2. Analyse des performances du meilleur modèle

Pour approfondir l'analyse des performances de notre méthode, nous avons généré pour chaque type de lésions cutanées la courbe ROC des prédictions faites par le meilleur modèle obtenu à l'aide de notre méthode d'entraînement. La Figure 3.11 présente les courbes ROC obtenues avec leurs scores d'AUC associés. Notre modèle performe globalement bien avec un score d'AUC au moins supérieur à 95% sur toutes les classes. Fait intéressant, nous avons observé que les classes de queue ont toutes obtenu un score d'AUC de 100%, confirmant ainsi la conclusion précédente sur la robustesse de notre approche pour le déséquilibre des classes.

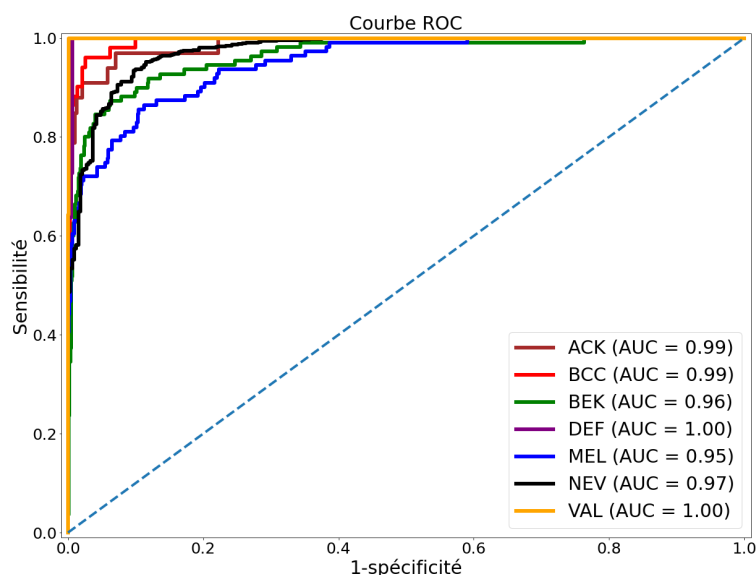


FIGURE 3.11. – Tracé des courbes ROC des performances obtenues par le meilleur modèle de notre approche sur les sept types de lésions de la tâche.

3.5.1.3. Comparaison de notre approche avec d'autres SAADs pour la détection des lésions cutanées

Nous avons également effectué une étude comparative des performances de notre méthode avec des SAADs existants développés sur la même base de données ISIC-2018 que nous avons utilisée. Cette étude comparative est présentée dans le Tableau 3.4. Les performances rapportées sont tirées des articles originaux. Notre méthode a obtenu la meilleure performance. De plus, malgré le fait que nous avons utilisé, dans le cadre de cette étude, une approche avec un seul modèle CNN, notre approche parvient à surclasser les performances de SAADs employant un ensemble de plusieurs CNN comme les travaux de Gessert *et al.* [67] et de Barata *et al.* [15].

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.5. Résultats et discussions

TABLEAU 3.4. – Performances en termes de score de BAcc obtenues par le meilleur modèle de notre approche et d’autres travaux utilisant la même base de données ISIC-2018.

Travaux	Approche SAAD utilisée	BAcc
Al-masni <i>et al.</i> [‡] [109]	Un CNN	0.81
Gessert <i>et al.</i> [‡] [67]	Ensemble de CNNs	0.76
Yao <i>et al.</i> [‡] [168]	Un CNN	0.86
Garg <i>et al.</i> [‡] [64]	Un CNN	0.74
Barata <i>et al.</i> [‡] [15]	Ensemble de CNNs	0.73
Notre méthode	Un CNN	0.88

Notre approche a obtenu la meilleure performance. Le signe [†] indique que le résultat est tiré de l’article original.

3.5.1.4. Études ablatives

Efficacité de la fonction de perte L_f pour la première phase de l’entraînement découplé : Nous avons comparé les performances durant la première phase de l’entraînement découplé de la fonction CE standard et notre fonction L_f proposée. Cette analyse a été réalisée afin d’évaluer laquelle des deux fonctions était la plus appropriée pour la première phase de l’entraînement découplé. Comme présenté dans le Tableau 3.5, la fonction de perte L_f a obtenu une meilleure performance avec un score de BAcc de 83% par rapport à la fonction CE qui a obtenu un score de 82%. Ce résultat démontre que la fonction de perte L_f est plus appropriée que la fonction CE pour la première étape de l’entraînement découplé.

TABLEAU 3.5. – Performances durant la première étape de l’entraînement découplé de EfficientNetB3 entraîné à l’aide des fonctions de perte CE et L_f .

Fonction de perte	BAcc
CE	0.82 ± 0.01
L_f	0.83 ± 0.00

Efficacité de notre méthode comparée à un schéma conventionnel d’apprentissage : Afin d’avoir une meilleure analyse de notre approche, nous avons tracé les courbes d’apprentissage sur les jeux de données d’entraînement et de validation du modèle EfficientNetB3 entraîné avec la fonction de coût CS (Figure 3.12a) et avec notre approche (Figure 3.12b). Nous pouvons observer sur la figure 8a que le modèle entraîné avec la fonction CS a atteint sa convergence autour de l’époque 40, par la suite ses performances ont commencé à stagner sans espoir d’amélioration. En revanche, pour le modèle entraîné avec notre approche (voir Figure 3.12b), nous observons que la deuxième phase d’entraînement a prolongé la convergence du modèle. En effet,

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.5. Résultats et discussions

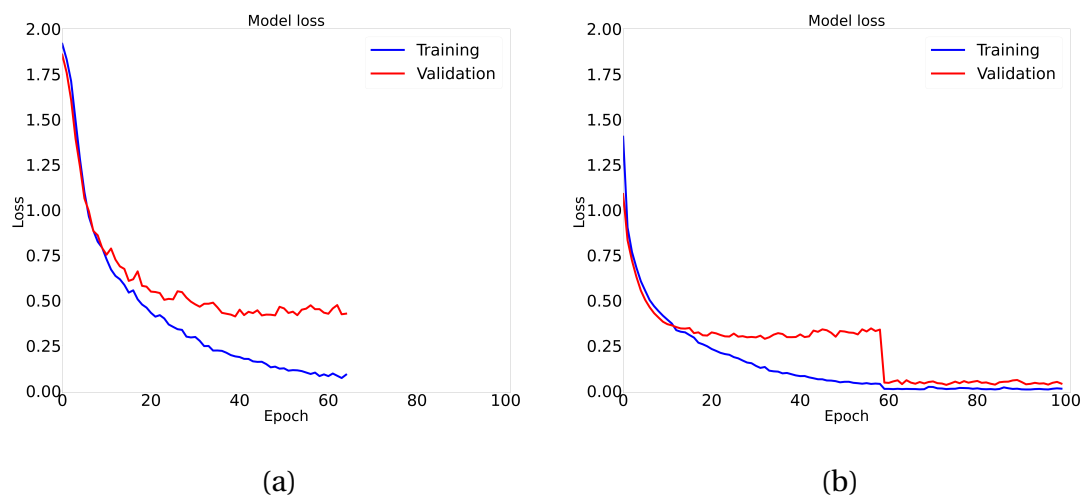


FIGURE 3.12. – Courbe d’apprentissage de EfficientNetB3 entraîné à l’aide de la perte d’entropie croisée CS (a) et de notre schéma d’apprentissage (b). Nous observons que la deuxième phase d’apprentissage de notre approche a prolongé la convergence du modèle avec un gain de performance significatif qui peut être observé à travers le gap visible autour de l’époque 60. Ceci met en évidence l’apport de notre approche par rapport à un schéma d’apprentissage classique.

lorsque le modèle a commencé à stagner autour de l’époque 40, le changement de fonction de coût de la fonction L_f à la fonction L_c nous a permis d’obtenir un gain de performance significatif, ce qui a permis d’améliorer les performances du modèle. Ceci s’observe à travers le gap visible sur la distribution des coûts autour de l’époque 60. Pour rappel, nous avons défini un délai de 10 époques avant de pouvoir changer automatiquement de fonction. Ce résultat met en évidence l’apport de notre approche par rapport à une procédure d’apprentissage classique.

3.5.2. Validation sur les jeux de données CIFAR et iNaturalist

Les performances globales en termes de score d’exactitude obtenues sur d’autres tâches de classification sont présentées dans le Tableau 3.6. Les performances de notre méthode ont été évaluées sur les mêmes jeux de données de test que les autres méthodes comparées. Les résultats indiquent que notre méthode surclasse l’état de l’art des approches sur le déséquilibre des classes dans tous les niveaux de déséquilibre évalués.

3. Contribution à la résolution du problème de déséquilibre des classes en apprentissage profond – 3.6. Conclusion

TABLEAU 3.6. – Taux d’erreur obtenus sur les jeux de données CIFAR-10, CIFAR-100 et iNaturalist-2018.

Base de données	CIFAR-10		CIFAR-100		iNaturalist-2018
Ratio de déséquilibre	100	50	100	50	500
Architecture	ResNet-32				ResNet-50
CE	28.01	22.91	61.20	56.54	42.03
FL [97]	29.24	23.17	62.05	56.98	41.70
CBL [‡] [38]	25.43	20.73	60.40	53.79	38.88
LDAM [‡] [25]	22.97	20.70	57.96	54.92	32.00
IBL [†] [125]	25.88	18.30	57.86	52.51	34.61
BAG [‡] [179]	19.97	16.41	52.17	48.31	29.13
Entraînement découplé [†] [180]	20.2	17.4	56.6	52.2	-
Notre méthode	18.22	14.92	52.65	47.86	27.75

Le symbole † indique que le résultat est copié de l’article original, tandis que le symbole ‡ signifie que le résultat est tiré de [179].

3.6. Conclusion

Dans ce chapitre, nous avons introduit une nouvelle méthode d’apprentissage de bout en bout pour résoudre le problème de déséquilibre des classes en apprentissage profond. Notre approche s’inspire de la méthode d’entraînement découplé afin de construire un schéma d’apprentissage conduisant à des modèles robustes. Plus spécifiquement, nous avons proposé deux nouvelles fonctions de perte L_f et L_c . La première fonction L_f guide l’extracteur des caractéristiques afin d’apprendre une bonne représentation des caractéristiques pour la tâche donnée. Tandis que la fonction L_c guide l’apprentissage du classifieur afin de mieux discriminer les classes de tête et de queue. Nous avons mené plusieurs études et effectué plusieurs expériences afin d’évaluer l’efficacité de notre méthode. L’approche proposée a été validée sur le jeu de données ISIC-2018, ainsi que sur trois autres jeux de données de référence en vision par ordinateur à savoir CIFAR-10, CIFAR-100 et iNaturalist-2018. Les résultats obtenus ont révélé que notre méthode permettait, avec un seul modèle CNN, de surclasser la plupart des SAADs existants pour la détection multi-classes des lésions cutanées. De plus, notre méthode a obtenu de meilleures performances que celles obtenues par les autres approches existantes pour résoudre le problème de déséquilibre des classes.

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux

Sommaire

4.1. Introduction	108
4.2. Revue des méthodes d'amélioration de l'explicabilité	109
4.3. Méthodes proposées	111
4.3.1. Base de données	111
4.3.2. Description de la méthode	111
4.3.2.1. Réseaux de neurones à convolutions	111
4.3.2.2. La théorie des jeux	112
4.3.2.3. Description de notre SAAD	112
4.3.3. Entraînement des modèles	115
4.4. Résultats et discussions	115
4.4.1. Résultats des expérimentations	115
4.4.2. Discussions	119
4.5. Conclusion	120

4.1. Introduction

Une autre des raisons précédemment évoquées justifiant le fait que des SAADs ne soient pas suffisamment répandus dans nos hôpitaux, au-delà de la crainte du dermatologue d'être induit en erreur par un faux négatif, est l'assimilation des CNNs à des boîtes noires (*black box*) rendant leur processus de décision difficile à comprendre pour un non-expert.

La théorie de la décision et en particulier la théorie des jeux est une solution potentielle au problème d'explicabilité des SAADs basés sur des CNNs car elle se concentre sur l'identification de la meilleure option de décision qui maximise l'utilité attendue du décideur.

Dans ce chapitre, nous présentons un nouveau pipeline basé sur une méthode ensembliste de CNNs pour le développement de SAAD pour le diagnostic du mélanome avec pour objectif non seulement d'améliorer les performances des systèmes existants, mais aussi d'apporter plus de transparence dans le processus de décision. Pour réaliser cette tâche, nous nous sommes inspirés de la façon dont les dermatologues diagnostiquent des lésions suspectes. En effet, ces derniers passent en revue de manière hiérarchique les lésions suspectes pour établir leurs hypothèses cliniques [12]. D'une part, certaines lésions pigmentées peuvent être facilement diagnostiquées comme les mélanomes typiques, les naevus et les kératoses bénignes typiques, le diagnostic est donc simple après un simple examen visuel. D'autre part, certaines lésions pigmentées atypiques peuvent être trompeuses et difficiles à diagnostiquer, les dermatologues doivent donc procéder en plusieurs étapes, en éliminant d'abord les diagnostics différentiels et en conservant les principales hypothèses basées sur une analyse de la lésion; pour ces lésions difficiles, la certitude du dermatologue sur le diagnostic est plus faible [52, 69] et un examen histologique est souvent demandé en supplément. En s'inspirant de cela, l'hypothèse formulée dans le cadre de ce travail est qu'un SAAD conçu de manière similaire à cette approche hiérarchique aurait de meilleures performances et améliorerait l'explicabilité du processus de décision pour les dermatologues. À cette fin, nous introduisons un nouveau schéma d'apprentissage basé sur une stratégie de vote hiérarchique prédéfinie de sept joueurs (CNNs). Les principales contributions de ce travail sont :

- La combinaison des CNNs et la théorie des jeux pour une tâche de classification. À notre connaissance, il s'agit de la première tentative d'intégration de la théorie des jeux dans un contexte de classification des lésions cutanées;
- L'introduction d'une méthode d'évaluation de la confiance d'une prédiction faite par un SAAD;

Ce chapitre sera organisé comme suit : la Section 4.2 présentera une revue de littérature des méthodes permettant d'améliorer l'explicabilité des algorithmes d'apprentissage profond, dans la Section 4.3 la définition de la théorie des jeux ainsi que la description de notre méthode sont effectuées, nos résultats seront discutés dans la Section 4.4 et la Section 4.5 conclura le chapitre.

4.2. Revue des méthodes d'amélioration de l'explicabilité

Aujourd'hui, l'un des principaux objectifs de la communauté de vision par ordinateur dans le domaine médical n'est plus seulement d'atteindre des performances élevées, mais aussi d'améliorer l'explicabilité des SAADs afin d'accroître leur utilisation dans un contexte clinique. Il existe principalement trois approches dans la littérature pour atteindre cet objectif : la visualisation de cartes de chaleur des caractéristiques, les systèmes de Recherche d'images basées sur le contenu –ou *content based image retrieval*- (CBIR) et l'incorporation des connaissances des dermatologues dans des SAADs.

Le principe de base de la visualisation des cartes de chaleur des caractéristiques est de mettre en évidence les régions d'une image qui contribuent à la décision d'un CNN. Les cartes d'activation de classe [182, 143] représentent la méthode la plus populaire pour réaliser cette tâche. Van *et al.* [118] ont été les premiers à utiliser cette méthode pour analyser les prédictions faites par un CNN sur des lésions cutanées. Ils ont montré que les CNNs étaient capables d'apprendre des caractéristiques similaires à celles utilisées par les dermatologues pour établir un diagnostic, à l'instar de l'irrégularité de la bordure et la couleur des lésions cutanées. Wang *et al.* [161] ont proposé un CNN multimodal pour le diagnostic des lésions cutanées qui utilise l'algorithme Grad-CAM [143] pour analyser la contribution de chaque caractéristique relativement aux méta-données du patient. Dans le même ordre d'idée, Alche *et al.* [5] ont proposé un CNN pour le diagnostic du cancer de la peau. L'approche proposée a été couplée avec l'algorithme Grad-CAM pour effectuer une analyse qualitative de leur modèle. Une autre méthode de visualisation des cartes de chaleur des caractéristiques repose sur l'algorithme Explication additive de Shapley –ou *SHapley Additive exPlanations*– (SHAP) [102]. L'algorithme SHAP permet d'analyser les prédictions au niveau du pixel et de s'assurer que le modèle se concentre sur les parties pertinentes des images pour la tâche à accomplir. SHAP a été utilisé par Mohammad *et al.* [116] pour améliorer l'explicabilité de son SAAD des lésions cutanées. Cependant, les approches de visualisation des cartes de chaleur des caractéristiques sont employées après l'entraînement du modèle, évitant ainsi tout impact sur les performances du modèle.

Le système CBIR est une alternative aux approches de visualisation des cartes de chaleur des caractéristiques. L'intuition derrière cette approche est d'aider à la prise de décision en récupérant et en affichant à partir d'une base de données des échantillons d'images visuellement pertinents vis à vis de l'échantillon examiné. Les travaux de Tschandl *et al.* [156], par exemple, comparent la précision diagnostic de ce système appliqué à des images dermoscopiques aux prédictions faites par un CNN. Ils ont constaté qu'avec l'aide simplement de 16 échantillons d'images récupérés, leur système atteignait la même performance qu'un modèle CNN conventionnel. La même démarche a été adoptée par Allegretti *et al.* [8] dans leurs travaux avec une performance atteinte par leur système supérieure à celles des dermatologues. Sadeghi *et al.* [141] ont analysé l'impact d'un système CBIR sur la routine quotidienne des dermatologues. Leur étude a révélé que ces systèmes aidaient les dermatologues novices à augmenter de manière significative leur précision de diagnostic de quatre types de lésions cutanées couramment observées. L'application d'un système CBIR afin d'améliorer l'explicabilité des SAADs des lésions cutanées a également été étudiée dans [18]. Toutefois, malgré la clarté sur le processus de décision qu'apportent ces systèmes, quelques inconvénients demeurent. En effet, l'intuition derrière ce type de système repose sur deux notions subjectives [32]. Tout d'abord, du point de vue sémantique, la corrélation entre la similarité des caractéristiques et celle des étiquettes de terrain n'est pas nécessairement vraie. De même, du point de vue de l'utilisateur, ce qui est considéré similaire d'un point de vue pathologique par un CNN ne correspond pas

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.3. Méthodes proposées

nécessairement aux mesures humaines de la similarité.

La dernière approche pour améliorer l'explicabilité des SAADs pour les lésions cutanées consiste à modéliser et incorporer les connaissances des dermatologues dans un SAAD. Cette dernière approche est celle suivie par les travaux de Kaymak *et al.* [91] et Barata *et al.* [15]. Les auteurs de ces travaux se sont inspirés de l'organisation taxonomique des lésions cutanées [12] afin de structurer leurs jeux de données pour développer des SAADs. Cependant, malgré le fait que cette approche apporte une rationalité à la décision, une étude récente de Alves *et al.* [10] démontre que les modèles formés à l'aide de cette approche sont moins performants qu'un simple modèle basé sur un CNN multi-classes.

4.3. Méthodes proposées

4.3.1. Base de données

Ce travail a été réalisé à l'aide du jeu de données public ISIC-2018. Nous avons appliqué notre approche à la même tâche de classification multi-classes mélanome, naevus et kératose séborrhéique effectuée dans le travail présenté au Chapitre 2. Nous avons également suivi la même répartition des jeux de données (entraînement, validation et test) ainsi que la même procédure de prétraitement décrite dans la Sous-section 2.3.1. Seule la taille de redimensionnement des images d'entrée a été redéfinie à 456×456 pixels afin de correspondre à la dimension d'entrée de la nouvelle architecture de CNN utilisée.

4.3.2. Description de la méthode

4.3.2.1. Réseaux de neurones à convolutions

Plusieurs architectures de CNNs sont open source, certaines d'entre elles ayant été précédemment entraînées sur le jeu de données ImageNet. Ainsi, il est possible de réutiliser les poids et les biais de ces architectures ainsi pré-entraînées et de les affiner pour les appliquer à d'autres tâches. C'est ce qu'on appelle l'apprentissage par transfert. Les réseaux EfficientNet [155] sont actuellement l'une des architectures les plus couramment utilisées en vision par ordinateur. Leurs auteurs [155] ont proposé huit versions de l'architecture ayant différentes échelles de résolution et de taille afin de les adapter selon la puissance de calcul disponible. Les huit versions proposées ont une appellation selon le niveau d'échelle variant de B0 à B7, la version B7 étant celle ayant la plus grande échelle. L'architecture EfficientNet a été régulièrement utilisée avec succès pour la tâche de classification des lésions cutanées [67, 88]. Dans ce travail, nous avons utilisé la version B5 d'EfficientNet comme architecture de base de nos modèles en raison des ressources disponibles. Nous avons à cet effet modifié le modèle original en remplaçant la dernière couche FC par une nouvelle couche FC comportant deux neurones pour effectuer une classification binaire ou trois neurones

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.3. Méthodes proposées

pour effectuer une classification ternaire. Les nouvelles couches ont été initialisées à l'aide de la méthode de Kaiming [81].

4.3.2.2. La théorie des jeux

La théorie des jeux est un cadre théorique permettant de modéliser les situations conflictuelles entre des acteurs concurrents et d'analyser le comportement des différents acteurs. La théorie des jeux a été développée à l'origine comme modèle mathématique dans le domaine de la recherche opérationnelle et a été appliquée à d'autres disciplines pour résoudre les problèmes de collaboration et de concurrence entre différents objectifs. Plus récemment, certains chercheurs ont étudié l'utilisation de la théorie des jeux pour l'apprentissage profond [79]. La théorie des jeux repose sur trois éléments principaux : les joueurs, les stratégies qu'ils utilisent, et les gains potentiels qu'ils reçoivent relativement à chacune des stratégies [73]. En général, le principe de la théorie des jeux peut se résumer comme le processus qui consiste à effectuer le choix qui maximise le gain de chaque joueur après qu'un des joueurs ait adopté une certaine stratégie, ceci en supposant que tous les joueurs sont rationnels.

4.3.2.3. Description de notre SAAD

Dans le cadre de cette étude, les modèles CNNs représentent les joueurs. Nous avons défini deux stratégies pour les joueurs. Dans la première nous regroupons les prédictions par niveau de confiance (élevé, moyen, incertain), et dans la seconde nous effectuons la prédiction finale. La valeur du gain dépend des probabilités de sortie des CNNs.

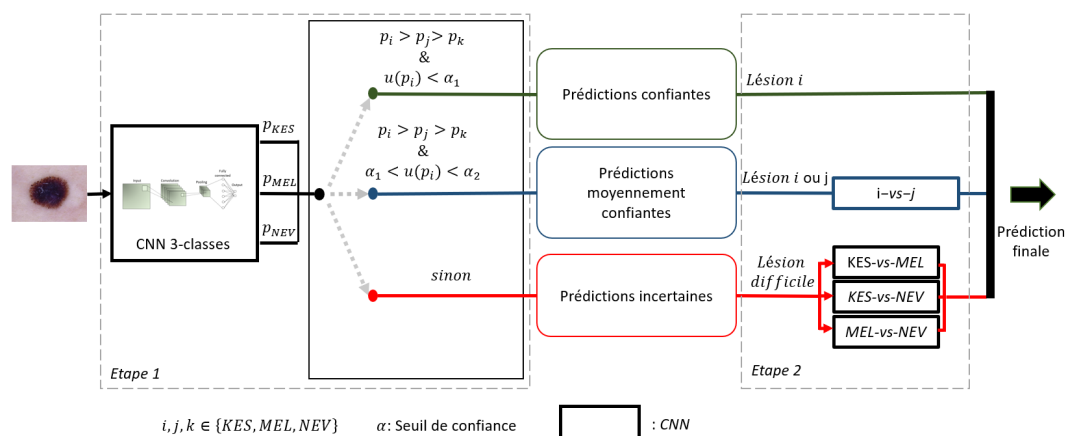


FIGURE 4.1. – Description du pipeline de notre système automatique d'aide au diagnostic basé sur la théorie des jeux.

Le pipeline global de notre approche est décrit par la Figure 4.1. Elle consiste en la combinaison de sept CNNs dont un a été entraîné à la tâche de classification multi-classes mélanome, naevus et kératose séborrhéique. Trois CNNs ont été entraînés pour classifier une classe de lésions cutanées par rapport à une autre classe à l'aide de

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.3. Méthodes proposées

la méthode un contre tous –ou *one versus all*– (o-v-a). Les trois autres CNNs ont été entraînés à distinguer deux types de lésions cutanées grâce à la méthode un contre un –ou *one versus one*– (o-v-o). Les CNNs entraînés avec la méthode o-v-a sont utilisés uniquement pour évaluer le niveau de confiance du CNN multi-classes et générer un score de confiance.

Soit un échantillon d'image x , et un modèle de CNN défini par la fonction $s : x \Rightarrow R^n$ qui génère un vecteur de probabilité \mathbf{p} de taille n . Les éléments du vecteur \mathbf{p} sont les probabilités p_i que x appartient à la classe i . $p_i \in [0, 1]$ et $\sum_n p_i = 1$. n représente le nombre de classes. Le classifieur s prend la décision de prédire x comme appartenant à la classe i à partir de l'équation ci-dessous :

$$class(x) = i, \text{ si } \max(\mathbf{p}) = p_i. \quad (4.1)$$

Ci-dessous, nous utiliserons les notations s_3 , $s_{i\bar{i}}$, et s_{ij} respectivement pour désigner le CNN ternaire, un CNN entraîné avec la méthode o-v-a et un CNN entraîné avec la méthode o-v-o.

s_3 génère trois probabilités p_{KES} , p_{MEL} et p_{NEV} que x correspond aux classes KES , MEL et NEV . La classe indique la nature de la lésion cutanée, KES pour la classe kératose séborrhéique, MEL pour la classe mélanome et NEV pour la classe naevus. Chaque modèle $s_{i\bar{i}}$ est spécialisé sur une classe i et génère deux probabilités p'_i et $p'_{\bar{i}}$ que x corresponde respectivement à la classe i ou pas. Chaque modèle s_{ij} génère également deux probabilités p_i et p_j que x corresponde à la classe i ou j .

Afin d'évaluer le niveau de confiance d'une probabilité de prédiction p_i , nous avons introduit la fonction u défini de la manière suivante :

$$u(p_i) = abs(p_i - p'_i) \quad (4.2)$$

La fonction u estime le niveau de confiance de la prédiction en calculant l'erreur absolu entre les prédictions effectuées par s_3 et $s_{i\bar{i}}$.

Durant la première étape de notre pipeline, s_3 et les trois classifieurs $s_{i\bar{i}}$ sont utilisés pour catégoriser chaque image d'entrée x selon le niveau de confiance de la prédiction : élevé, moyen et incertain. La répartition est effectuée à l'aide de l'équation 4.3. Dans l'équation 4.3, α_1 et α_2 sont obtenu par recherche sur grille relativement au score de BAcc obtenu sur le jeu de données de validation.

$$group(x) = \begin{cases} g1 : \text{elevation} & \text{pour } u(p_i) \leq \alpha_1 \\ g2 : \text{moyen} & \text{pour } \alpha_1 \leq u(p_i) \leq \alpha_2. \\ g3 : \text{incertain} & \text{sinon.} \end{cases} \quad (4.3)$$

Lorsque la prédiction est catégorisée comme appartenant au groupe $g1$ (élevé), la

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.3. Méthodes proposées

prédiction effectuée par s_3 est retenue. Lorsque la prédiction est catégorisée comme appartenant au groupe g_2 (moyen), le classifieur s_{ij} est utilisé pour affiner la prédiction de x . Dans le cas où la prédiction est catégorisée d'incertaine (groupe g_3), les trois classifieurs s_{BM} , s_{BN} et s_{MN} sont utilisés pour prédire la classe de x à l'aide de la stratégie **Max-Win**. L'Algorithme 3 décrit les étapes suivies par notre cadre pour prédire la classe d'une observation x .

Toutes les prédictions sont finalement normalisées à l'aide de l'équation 4.4 afin d'attribuer les probabilités de prédictions finales.

$$prob(i) = \frac{p_i}{p_i + p_j + p_k}; i, j, k = B, M, N \text{ and } i \neq j \neq k \quad (4.4)$$

Algorithm 3 Pseudo-code de notre pipeline.

Require: Image x , trois CNNs binaires s_{ij} , un CNN 3-classes s_3 , trois CNNs binaires $s_{i\bar{i}}$, une liste représentant les trois classes $class_list = [1, 2, 3]$, g_1 : groupe de prédictions avec un niveau de confiance certain, g_2 : groupe de prédictions avec un niveau de confiance moyen, g_3 : groupe de prédictions avec un niveau de confiance incertain

Ensure: Classe prédite pour l'échantillon x

Générer la probabilité de prédiction p_i de la classe associée à x avec s_3

Générer la probabilité p'_i d'appartenir à la classe i par le modèle $s_{i\bar{i}}$ spécialisé à cette classe

Estimez le niveau de confiance $u(p_i)$ de la prédiction faite par s_3 en utilisant l'équation $u(p_i) = abs(p_i - p'_i)$

if $u(p_i) < \alpha_1$ **then**

 catégorise la prédiction comme appartenant au groupe g_1

x appartient à la classe i

else if $(\alpha_1 < u(p_i) < \alpha_2) \ \& \ (p_i > p_j > p_k)$ **then**

 catégorise la prédiction comme appartenant au groupe g_2

 Générer les probabilités de prédiction p_i et p_j faites par le modèle s_{ij}

if $p_i > p_j$ **then**

x appartient à la classe i

else

x appartient à la classe J

end if

else

 catégorise la prédiction comme appartenant au groupe g_3

 Générer des probabilités de prédiction de tous les trois CNNs binaires s_{ij}

x appartient à la classe obtenue en appliquant la règle Max-Win sur les trois CNNs binaires s_{ij}

end if

4.3.3. Entraînement des modèles

Nous avons utilisé la bibliothèque Pytorch pour entraîner tous les modèles présentés dans cette étude. Le code de cette étude est disponible en [open source](#)¹ pour la communauté.

Durant l'apprentissage des modèles pré-entraînés, seules 75% des couches profondes ont été ré-entraînées afin de réduire le risque de sur-apprentissage. L'algorithme d'optimisation Adam et une taille de lot de 64 ont été utilisés pour mettre à jour les poids durant 100 époques. La fonction de perte CS a été adoptée pour l'évaluation du taux d'erreur durant l'apprentissage. Similairement à la procédure suivie dans [129], nous avons utilisé le taux d'apprentissage cyclique proposé dans [150] pour adapter le taux d'apprentissage durant l'apprentissage des modèles. Nous avons également appliqué l'arrêt précoce comme méthode de régularisation supplémentaire pour éviter le sur-apprentissage. Pour ce faire, l'apprentissage des modèles est stoppé lorsque le score de BAcc sur le jeu de données de validation ne s'améliore pas au bout de 15 époques consécutives.

4.4. Résultats et discussions

Cette sous-section présente les résultats et discussions des expériences réalisées. Étant donné que l'apprentissage des réseaux de neurones suit un processus stochastique, nous avons reproduit chacune de nos expériences à cinq reprises et nous avons rapporté par la suite les moyennes des résultats obtenus avec leur écart-type associé.

4.4.1. Résultats des expérimentations

Dans un premier temps, nous analysons les performances obtenues par notre approche dans la tâche de classification choisie. La Figure 4.2 présente les courbes ROC obtenues pour chacune des trois lésions cutanées de notre tâche de classification. Nous pouvons observer que notre méthode a obtenu de bonnes performances sur toutes les classes bien au-delà d'une prédiction hasardeuse ($AUC=0,5$), atteignant pour chaque lésion une AUC moyenne de 0,97, 0,93 et 0,96 respectivement pour les classes kératose séborrhéique, mélanome et naevus. Nous avons fait le constat que la classe mélanome reste la classe la plus difficile à détecter.

Dans un deuxième temps, nous avons comparé notre approche avec des travaux antérieurs basés sur des CNNs et développés pour la même tâche de classification des mélanomes, des naevus et des kératoses séborrhéiques. Le Tableau 4.1 présente les résultats de cette comparaison. Notre approche a surpassé toutes les méthodes comparées en atteignant une amélioration de l'AUC de 5% pour la classe mélanome,

1. <https://github.com/cartelgouabou/CAD-for-melanoma-detection-using-cnn-and-game-theory>

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.4. Résultats et discussions

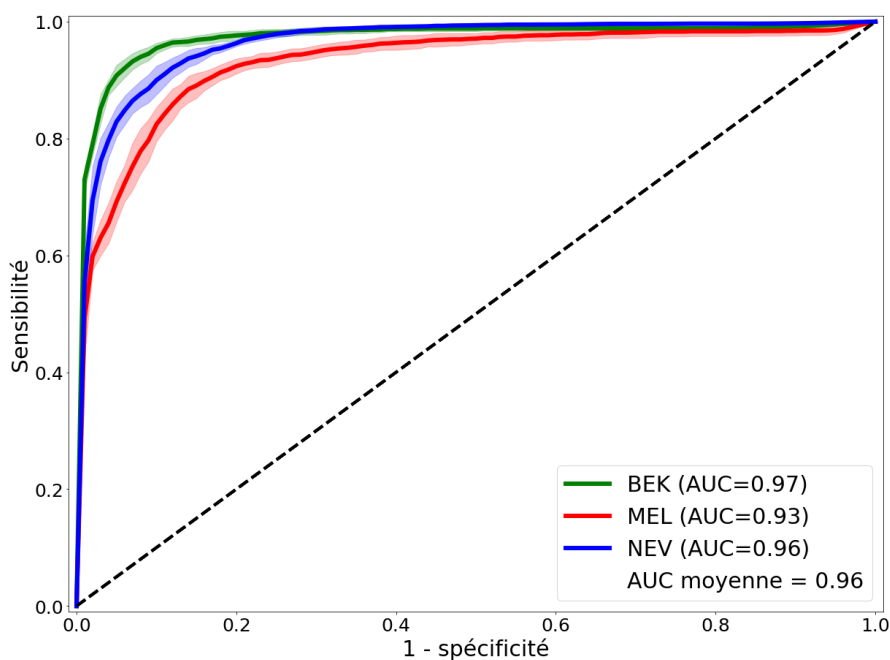


FIGURE 4.2. – Courbes ROC de notre approche. Les AUCs des courbes ROC obtenus sur le jeu de données de test pour chacune des lésions de la tâche sont précisées : mélanome (MEL), kératose bénigne (BEK) et nævus (NEV) sur l'ensemble de test.

4% pour la classe kératose séborrhéique, et 8% pour la classe nævus. De plus, notre approche a atteint un score de **Bacc** moyen de 0.86 conduisant à une amélioration de près de 9% des performances de notre précédente méthode de **SAAD** (DAG-CNN [61]) présentée dans le **Chapitre 2**.

TABLEAU 4.1. – Comparaison de notre approche avec d'autres SAADs existants pour la classification multi-classes des mélanomes, des nævus et des kératoses bénignes.

Travaux	BACC	AUC			
		Moyenne	MEL	BEK	NEV
Harangi <i>et al.</i> [78]	-	0.85	0.84	0.87	0.84
Bisla <i>et al.</i> [20]	-	0.92	0.88	-	-
Barata <i>et al.</i> [17]	0.70	0.88	-	-	-
Barata <i>et al.</i> [15]	0.74	0.92	0.80	0.92	0.85
DAG-CNN [61]	0.77 ± 0.00	-	0.87	0.93	0.88
Notre approche	0.86 ± 0.01	0.96 ± 0.00	0.93 ± 0.01	0.97 ± 0.01	0.96 ± 0.00

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.4. Résultats et discussions

Enfin, nous avons présenté notre approche à un dermatologue pour une évaluation en contexte clinique réel. La Figure. 4.3 décrit deux cas d'utilisation de notre pipeline. La Lésion 1 (voir Figure. 4.3(a)) est une image dermoscopique d'un mélanome typique avec une prédiction élevée (1.00) pour ce diagnostic et une faible prédiction pour les deux autres classes (une probabilité de 0.0 pour chacune des deux). La carte de chaleur révèle la zone de régression avec un voile bleu blanc correspondant à une région typique pour le diagnostic de mélanome, comme étant la zone ayant obtenu une forte activation des neurones pour la prédiction de la classe mélanome. Pour ce cas typique, aucune autre étape n'a été nécessaire et le modèle semble efficace. La Lésion 2 (voir figure. 4.3 (b)) représente une image dermoscopique d'une lésion pigmentée trompeuse. A première vue, cette lésion pourrait être une kératose bénigne ou un mélanome. La première étape a fourni des prédictions plus élevées pour le mélanome (0,53) et la kératose bénigne (0,46) et une prédiction plus faible pour le naevus (0,01) par le modèle s_3 avec un score de confiance de 54%. La deuxième étape utilisant un modèle entraîné spécifiquement pour distinguer la kératose bénigne et le mélanome a été efficace pour ce mélanome difficile à diagnostiquer avec une probabilité plus élevée (0,99) attribuée à la classe de vérité terrain qu'était le mélanome. Ces différentes étapes de notre pipeline mettent en lumière la capacité de notre approche à différencier et à classer des lésions pigmentées difficiles, ainsi que sa capacité à fournir un processus de décision plus transparent.

4. Amélioration de l'explicabilité des systèmes de diagnostic automatique du mélanome grâce à la théorie des jeux – 4.4. Résultats et discussions

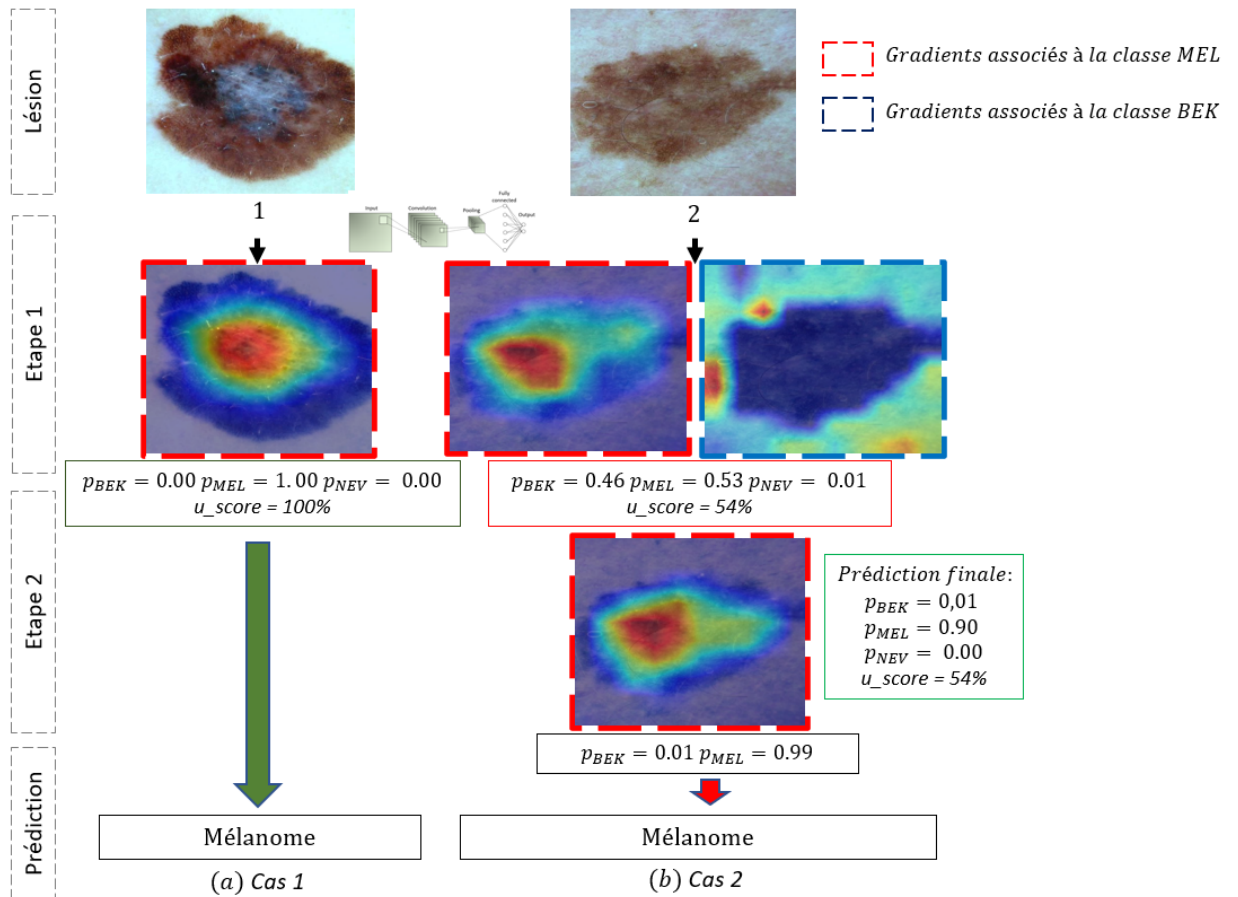


FIGURE 4.3. – Illustration du processus de décision de notre SAAD basé sur la théorie des jeux. La génération des cartes thermiques est implémentée avec Grad-CAM [143]

4.4.2. Discussions

Dans cette étude, nous avons présenté une méthode d'ensemble de CNNs avec pour but d'obtenir un SAAD du mélanome qui soit à la fois précis et explicable. Notre méthode repose sur un pipeline suivant une structure hiérarchique et combinant sept CNNs. Les scores de BAcc et de AUC ont été principalement utilisés pour analyser les performances de classification. La méthode proposée a atteint un score moyen de AUC de 96%, ce qui démontre la bonne performance de notre approche dans cette tâche. La performance de notre approche a également été comparée à celle des travaux précédents dans la même tâche et notre approche a obtenu de meilleure performance avec une marge minimale de 9% en termes de score de BAcc. Une analyse de cas d'utilisation de notre système a également été réalisée par un dermatologue pour évaluer l'efficacité notre approche en contexte clinique.

Dans cette étude, nous avons intégré la plupart des stratégies présentées dans la littérature afin d'accroître la précision des SAADs pour la classification des lésions cutanées. Similairement au travail de [132, 28, 60], nous avons utilisé l'architecture pré-entraînée EfficientNet pour construire tous nos modèles. Nous avons également employé des techniques d'augmentation artificielle des données comme dans [181, 108] afin de nous assurer de la robustesse de nos modèles. Nous avons enfin construit un pipeline combinant plusieurs modèles pour former un système ensembliste de CNNs, ce qui est une stratégie bien connue utilisée dans les travaux précédents [61, 89, 136, 108] et permettant d'obtenir de meilleure performance qu'un système basé sur un seul modèle. En outre, l'approche présentée dans cette étude intègre la visualisation des cartes thermiques, via l'algorithme Grad-CAM, ce qui est une méthode efficace permettant d'apporter plus d'applicabilité dans la prédiction effectuée par un modèle [100]. Ceci est particulièrement important pour notre tâche dans la mesure où les travaux de Van *et al.* [118] ont montré que la méthode Grad-CAM révèle des caractéristiques similaires à celles utilisées par les dermatologues pour établir leur diagnostic. De plus, pour avoir suivi un processus décisionnel hiérarchique similaire à la démarche du dermatologue pour établir un diagnostic, les résultats que nous avons obtenus, éclairés par l'analyse d'un dermatologue, nous ont permis de démontrer que notre approche est capable de fournir des informations pertinentes sur le processus de décision. Par ailleurs, la définition, pour la première fois à notre connaissance, d'une méthode d'évaluation du niveau de confiance d'une prédiction apporte plus de clarté à la prédiction faite par un SAAD et fournit aux dermatologues suffisamment d'éléments pour les aider à poser le meilleur diagnostic. En effet, comme indiqué dans [100], combiner différentes méthodes d'explication avec diverses modalités permet d'obtenir des explications plus complètes. De plus, fournir aux dermatologues suffisamment d'éléments pour étayer une prédiction renforce la confiance de ce dernier dans la décision du modèle et les aide à identifier les décisions potentiellement discutables [148].

Cette étude répond à une nécessité de construire des SAADs explicables devenue de

plus en plus pertinente compte tenu des normes éthiques et juridiques récentes [134, 13]. Néanmoins, notre étude présente un certain nombre de limites qui doivent être abordées. Tout d'abord, dans cette étude, nous ne nous sommes pas appesantis sur la résolution du problème de déséquilibre de classe qui est présent dans les tâches de vision par ordinateur et plus particulièrement dans la classification des lésions cutanées, ceci du fait que cette problématique a déjà fait l'objet d'autres travaux de notre part qui sont présentés dans le [Chapitre 2](#). Une perspective de ce travail serait donc de combiner la méthode proposée dans le [Chapitre 2](#) pour résoudre le problème de déséquilibre avec le pipeline proposé dans cette étude ce qui pourraient renforcer la robustesse de ce dernier. Deuxièmement, notre étude manque de suffisamment d'investigations sur l'utilisabilité et l'adoptabilité de notre application dans un scénario clinique réel, comme suggéré dans [86]. Bien que nous ayons intégré une analyse de cas d'utilisation de notre système par un dermatologue, il serait judicieux de mener une véritable étude clinique en intégrant un plus grand nombre de cliniciens. Néanmoins, les résultats obtenus montrent qu'il est clairement possible de construire des SAADs basés sur l'apprentissage profond qui soient explicables sans perdre en précision de diagnostic.

4.5. Conclusion

Dans cette étude, nous avons proposé un nouveau pipeline pour le développement de SAAD pour le diagnostic des lésions cutanées. L'objectif de l'approche proposée était non seulement d'augmenter la performance des systèmes existants, mais aussi d'apporter plus de transparence durant leur processus de décision. Pour ce faire, nous avons combiné sept modèles CNNs à travers la théorie des jeux et avons associé le système ainsi constitué à la visualisation des cartes de chaleurs. En outre, nous avons suivi un processus décisionnel hiérarchique inspiré de la démarche en deux étapes utilisée par les dermatologues pour poser un diagnostic sur des lésions suspectes. De plus, nous avons introduit une nouvelle méthode pour évaluer le niveau de confiance d'une prédiction générée par un système automatisé et l'avons intégrée dans notre pipeline. Les résultats montrent que notre approche peut effectivement améliorer la précision des SAADs par rapport à l'état de l'art. De plus, en menant une étude des cas d'utilisation de notre pipeline par un dermatologue, nous avons pu observer que le processus de décision de notre approche s'est révélé être plus intuitif et explicable, ce qui soutiendrait son utilisation dans des contextes cliniques réels. À ce stade, notre système a un grand potentiel d'utilisation dans des contextes cliniques réels en tant qu'outil de formation pour les dermatologues débutants.

5. Applications

Sommaire

5.1. Introduction	121
5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome	122
5.2.1. Matériels et Méthodes	123
5.2.2. Résultats	125
5.3. Étude comparative des SAADs proposés combinés ou non avec la méthode d'entraînement découplé de bout en bout	128
5.3.1. Matériels et méthodes	128
5.3.2. Résultats et discussions	128
5.4. Conclusion	129

5.1. Introduction

Dans les chapitres précédents, nous avons révélé que les CNNs représentent une des meilleures alternatives pour le diagnostic automatique des lésions cutanées à partir d'images. Pour ce faire, nous avons présenté et évalué trois contributions de cette thèse, la première étant un SAAD réalisé pour la tâche de diagnostic du mélanome qui repose sur l'agrégation de plusieurs CNNs binaires à l'aide d'un DAG. La deuxième contribution portant sur l'introduction d'un nouveau schéma d'apprentissage dénommé entraînement découplé de bout en bout dont l'objectif est de permettre la construction de modèles neuronaux robustes au déséquilibre des classes. La troisième contribution est le développement d'un SAAD intégrant la problématique d'explicabilité.

Les modèles proposés durant ces études ont été pour la plupart entraînés sur une base de données publique d'images dermoscopiques, images contenant un certain nombre d'artefacts comme le halo périphérique. Dans ce chapitre, nous nous intéresserons dans un premier temps à l'impact d'un halo périphérique sur la performance des modèles neuronaux, et évaluerons une technique de prétraitement à même de réduire l'influence de cet artefact lors du développement des SAADs pour la détection des mélanomes.

De plus, comme vu précédemment, les différentes méthodes que nous avons introduites peuvent être combinées afin d'améliorer encore plus leurs performances. Nous effectuerons donc une étude comparative combinant les SAADs proposés avec

la méthode MED de résolution de la problématique du déséquilibre des classes, ceci afin d'évaluer si oui ou non la combinaison de ces méthodes apporte un réel bénéfice de performance.

5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome

Les CNNs, composés de plusieurs couches séquentielles, sont entraînés "de bout en bout" à l'aide de grands jeux de données. Pour des applications de détection automatique des cancers de la peau, les jeux de données sont constitués à partir d'images cliniques ou dermoscopiques et de leurs diagnostics correspondants. Sans intervention humaine, les CNNs apprennent donc à détecter des caractéristiques représentatives au niveau des pixels et à les associer à un diagnostic spécifique. Cependant, ce processus autonome comporte le risque d'introduire des biais, conduisant à associer des artefacts spécifiques d'une image dermoscopique à un diagnostic spécifique. Une illustration des artefacts couramment rencontrés sur des images dermoscopiques est présentée sur la Figure 5.1. Les biais introduits par ces artefacts ont pour conséquence de conduire à un sur-apprentissage ou à une augmentation de l'erreur de généralisation [164]. De plus, les SAADs basés sur des CNNs sont de plus en plus utilisés par des cliniciens inexpérimentés pour un second avis diagnostic in-vivo [158]. Par conséquent, une application clinique plus large de SAADs entravés par des erreurs de diagnostic intrinsèques dû aux CNNs pourrait avoir des résultats préjudiciables sur le bien-être des patients.

Des travaux précédents ont étudié et tenté de réduire l'influence d'artefacts spécifiques aux images dermoscopiques sur la performance des CNNs. Les artefacts évalués étaient les traces de marqueur [164, 107] et les cheveux [151, 107]. La présente étude avec pour référence [119], à notre connaissance, est pionnière sur l'analyse de l'impact du halo périphérique sur la performance des algorithmes d'apprentissage profond. L'objectif de cette étude était d'analyser et quantifier l'influence d'un halo sur la performance diagnostique d'un SAAD basé sur un CNN. Nous référons aux lecteurs les travaux présentés dans [146, 128] pour des études plus récentes sur l'impact du halo.

5. Applications – 5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome

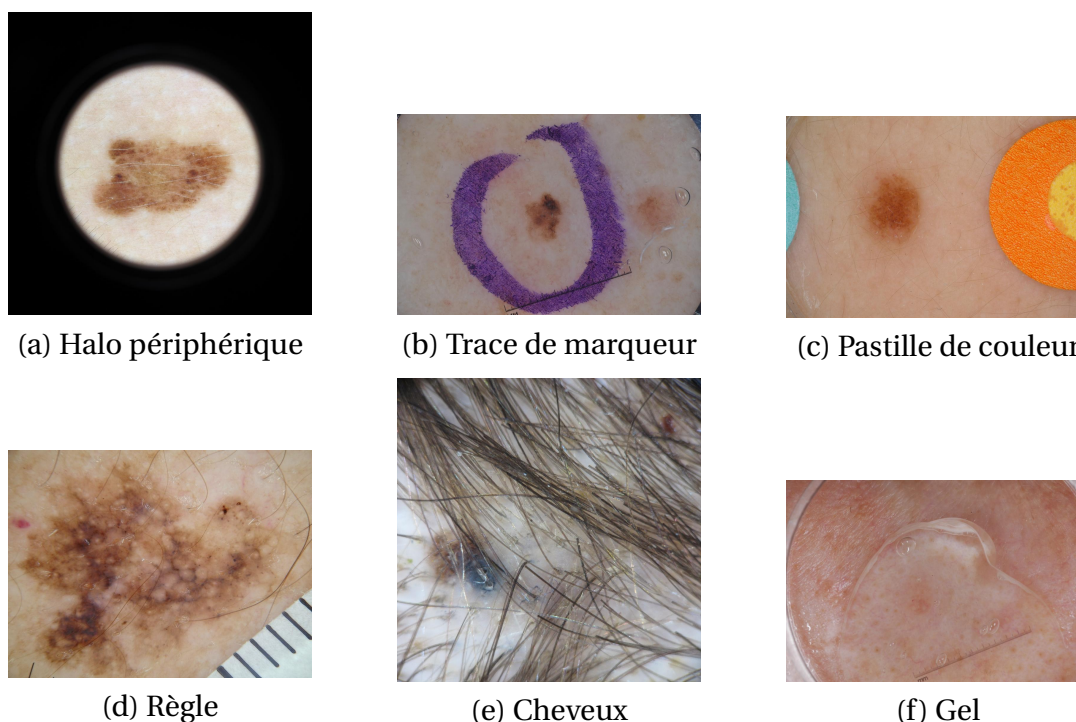


FIGURE 5.1. – Illustration d'artefacts fréquemment présents sur les images dermoscopiques des lésions cutanées.

5.2.1. Matériels et Méthodes

Préparation de la base de données : La base de données publique ISIC-2019 [157, 34, 33] a été utilisée pour entraîner les modèles. Dans cette étude, nous nous sommes intéressés à la tâche de classification binaire mélanome versus naevus. A cet effet, nous avons conservé uniquement les images dermoscopiques appartenant à ces deux classes contenues dans la base de données initiale, la base résultante contenant 4522 images dermoscopiques de mélanome et 12875 images dermoscopiques de naevus. Pour se conformer à l'objet de notre étude, nous avons exclu de la base de données toutes les images dermoscopiques présentant des artefacts autres que le halo périphérique notamment : les images dermoscopiques floues, les images dermoscopiques contenant une règle, des images dermoscopiques contenant des bulles, des traces de gel, des traces de marqueur ou des cheveux (voir Figure 5.1). La base de données ainsi constituée a été répartie en un jeu de données d'entraînement (80%) et un jeu de données de test (20%) (voir Table 5.1).

Afin d'étudier plus spécifiquement l'impact du halo, nous avons constitué trois groupes de jeux de données à partir du jeu de données d'entraînement et six groupes de jeux de données à partir du jeu de données de test. Pour ce qui est des groupes constitués à partir du jeu de données d'entraînement, nous avons défini les groupes suivants :

5. Applications – 5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome

TABLEAU 5.1. – Répartition de la base de données.

	mélanome	naevus
ISIC-2019	4 522	12 875
Jeu de données d'entraînement* (80%)	1 259	5 716
Jeu de données de test* (20%)	271	1 283

* Indique que les images dermoscopiques présentant des artefacts autre que le halo périphérique ont été enlevées.

- HO-Tr : Ce groupe contient toutes les images dermoscopiques du jeu de données d'entraînement constitué;
 - SH-Tr : Ce groupe contient toutes les images dermoscopiques du jeu de données HO-Tr à l'exception de celles présentant un halo;
 - RH-Tr : Ce groupe contient toutes les images dermoscopiques du jeu de données HO-Tr où les images dermoscopiques présentant un halo ont été rognées;
- Pour ce qui est des groupes constitués à partir du jeu de données de test, nous avons défini les groupes suivants :
- HO-Te : Ce groupe contient toutes les images dermoscopiques du jeu de données de test constitué;
 - SH-Te : Ce groupe contient toutes les images dermoscopiques du jeu de données HO-Te à l'exception de celles présentant un halo;
 - HA5-Te : Ce groupe contient toutes les images dermoscopiques du groupe SH-Te avec ajout d'un halo artificiel à 5%;
 - HA15-Te : Ce groupe contient toutes les images dermoscopiques du groupe SH-Te avec ajout d'un halo artificiel à 15%;
 - HA45-Te : Ce groupe contient toutes les images dermoscopiques du groupe SH-Te avec ajout d'un halo artificiel à 45%;
 - RH-Te : Ce groupe contient toutes les images dermoscopiques du jeu de données HO-Te où les images dermoscopiques présentant un halo ont été rognées;

Méthodes de prétraitement : Dans cette étude, quatre méthodes de prétraitement ont été employées : la standardisation des couleurs, la normalisation et le rognage des images dermoscopiques, et l'ajout de masque d'**halo** périphérique.

En ce qui concerne les méthodes de standardisation des couleurs et de normalisation des images, nous avons utilisé la même procédure décrite dans la [Sous-section 2.3.1](#). Ces prétraitements ont été appliqués à toutes les images dermoscopiques indépendamment du scénario évalué. Pour ce qui est du rognage, nous avons appliqué ce prétraitement uniquement sur les images dermoscopiques des groupes des jeux de données concernées (RH-Tr et RH-Te). Enfin, les masques d'**halo** périphérique ont été ajoutés uniquement sur les images dermoscopiques des groupes des jeux de données concernées (HA1-Te, HA2-Te, HA3-Te). Les masques d'**halo** ont été générés en mettant simplement à zéro les pixels des images. L'évaluation des zones où appliquer le masque en fonction de la dimension voulue est décrite par la Figure 5.2. Dans la

5. Applications – 5.2. Analyse de l’impact d’un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome

Figure 5.2, a représente le ratio d’un halo artificiel voulu, A_{image} représente l’aire d’une image x de dimension $L \times l$, A_{Lesi} représente l’aire du cercle décrit par le halo artificiel et R son rayon. De façon numérique, le masque du halo artificiel est appliqué sur l’image en mettant à zéro tous les pixels de l’image x situés au-delà du rayon R .

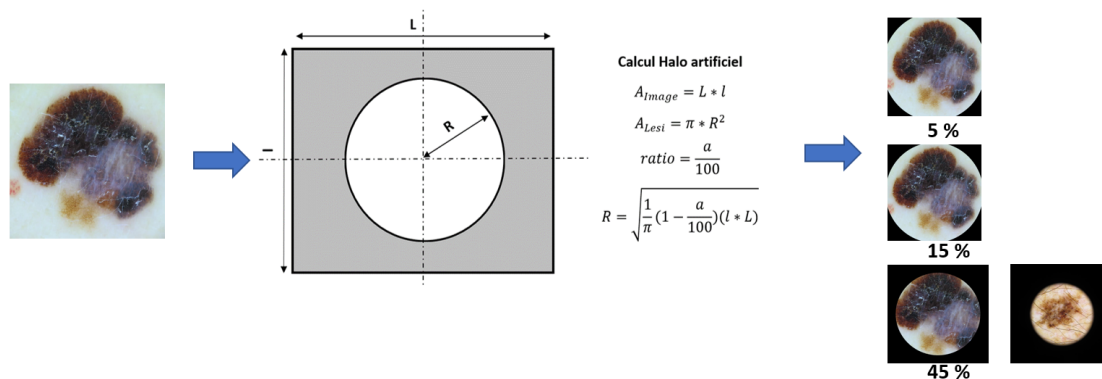


FIGURE 5.2. – Description de la méthodologie employée pour évaluer les zones de l’image où appliquer le masque d’halo périphérique.

Entraînement des modèles : L’architecture pré-entraînée ResNet-50 a été utilisée durant nos expérimentations. La dernière couche de l’architecture initiale a été modifiée afin de l’adapter à une tâche de classification binaire. Tout d’abord, nous avons ré-entraîné les modèles sur les groupes des jeux de données d’entraînement prédéfinis. Ensuite, nous avons extrait les caractéristiques de l’avant dernière couche FC des modèles pour entraîner un algorithme SVM afin d’obtenir la prédiction finale. La génération des cartes de chaleur a été réalisée à l’aide de l’algorithme Grad-CAM [143]. Pour mémoire, une carte de chaleur permet de visualiser les régions de l’image contribuant à la prédiction.

5.2.2. Résultats

Les métriques de performance analysées durant cette étude sont l’AUC, la sensibilité et la spécificité. Le Tableau 5.2 présente les performances obtenues par les différents modèles selon les prétraitements effectués aux images dermoscopiques. La première partie du Tableau 5.2 démontre l’impact du halo. On observe que le modèle a obtenu un score d’AUC, de sensibilité et de spécificité respectivement de 0.91, 73% et 89% sur des images dermoscopiques dépourvues d’halo. Cependant, les ajouts de masque de halo artificiel de dimension 5%, 15% et 45% de la surface de l’image dermoscopique ont conduit respectivement à une diminution importante et graduelle du score de sensibilité : 62%, 57%, 27%, sans perte du score de spécificité : 90%, 91%, 97%. L’impact du halo sur la performance des CNNs est encore plus perceptible en observant la Figure 5.3. En tête de la Figure 5.3, nous observons une image dermoscopique sans halo appartenant à la classe mélanome. Cette image dermoscopique a obtenu une probabilité de prédiction de 1.00 pour le diagnostic du mélanome (vrai positif). La

5. Applications – 5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome

carte de chaleur associée à cette prédiction révèle une concentration des caractéristiques profondes ayant contribué à la prédiction sur la lésion. Cette même image dermoscopique avec ajout d'un halo artificiel de dimension correspondant à 45% de la superficie de l'image dermoscopique a eu une probabilité de diagnostic de prédiction de 0.03 pour le diagnostic du mélanome, l'orientant donc vers un diagnostic incorrect de naevus (faux négatif). La carte de chaleur associée à cette prédiction révèle la zone de l'image contenant le halo comme étant celle ayant majoritairement contribué à cette prédiction.

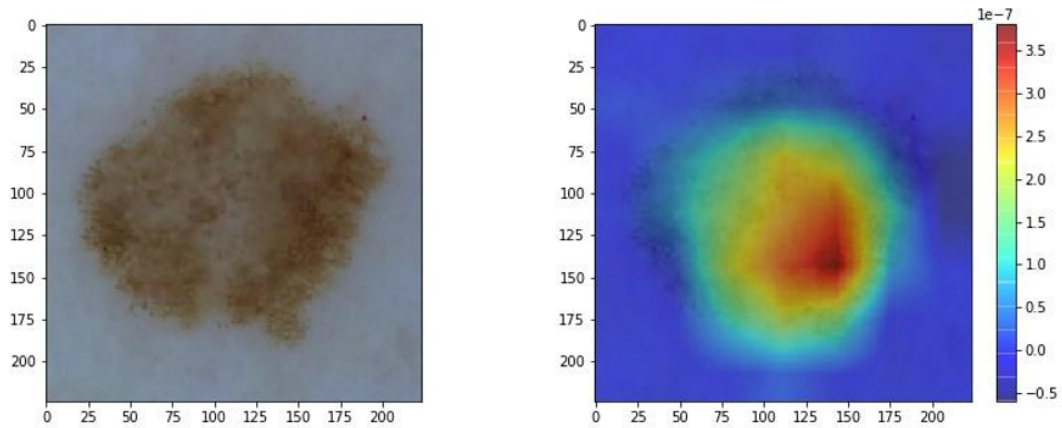
Lorsqu'on s'intéresse à la deuxième partie des résultats présentés dans le Tableau 5.2, on observe que le score de sensibilité du modèle, entraîné avec des images d'un halo naturellement présent dans la base de données, est de 67%. Ce score augmente considérablement jusqu'à atteindre 90% tout en conservant un score de spécificité comparable lorsque l'on applique un rognage sur les images dermoscopiques.

TABLEAU 5.2. – Performances des modèles selon les prétraitements effectués sur les images dermoscopiques .

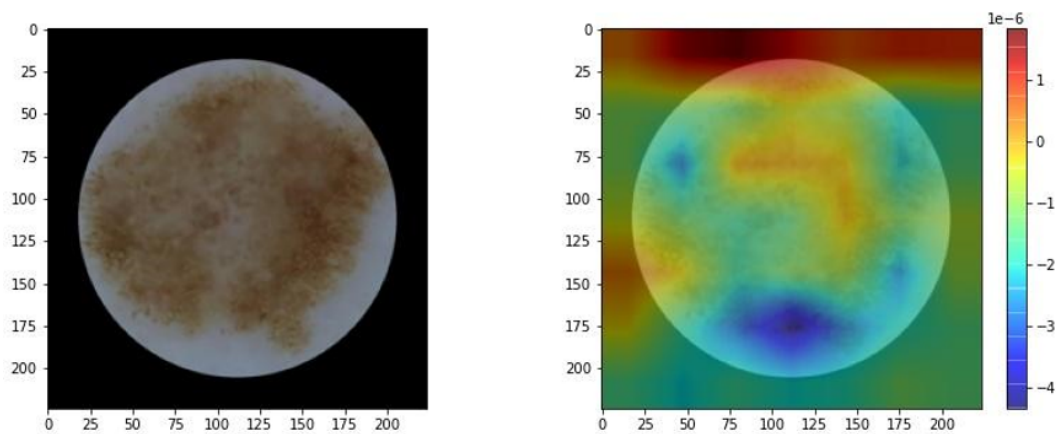
Jeu d'entraînement	Jeu de test	Prétraitement effectué	AUC	sensibilité	spécificité
SH-Tr	SH-Te	Images dépourvues d'halo noir périphérique	0.91	73%	89%
SH-Tr	HA5-Te	ID avec halo artificiel à 5%	0.87	62%	90%
SH-Tr	HA15-Te	ID avec halo artificiel à 15%	0.86	57%	91%
SH-Tr	HA45-Te	ID avec halo à 45%	0.82	27%	97%
HO-Tr	HO-Te	ID avec halo noir périphérique originel	0.73	67%	65%
RH-Tr	RH-Te	ID rognées automatiquement en périphérie	0.85	90%	58%

* Répartition obtenue après avoir divisé la base de données initiale. ID est l'abréviation d'image dermoscopique

5. Applications – 5.2. Analyse de l'impact d'un artefact récurrent présent sur des images dermoscopiques sur la détection automatique du mélanome



(a) Prédiction (classes=MEL)= 1.00



(b) Prédiction (classes=MEL)= 0.03

FIGURE 5.3. – Illustration de l'impact d'un halo périphérique sur la performance des CNNs. Au-dessus nous avons une image dermoscopique (ID) d'un mélanome qui a été correctement prédit comme tel. La carte de chaleur associée à cette prédiction révèle une concentration des caractéristiques profondes ayant contribué à la prédiction sur la lésion. En dessous, nous avons la même ID sur laquelle un masque d'un halo périphérique à 45% a été appliqué. L'ajout du masque a conduit à une prédiction erronée de l'ID et la carte de chaleur révèle la zone de l'image contenant le halo comme étant celle ayant majoritairement contribué à cette prédiction.

5.3. Étude comparative des SAADs proposés combinés ou non avec la méthode d’entraînement découplé de bout en bout

5.3.1. Matériels et méthodes

Préparation de la base de données : Similairement aux expérimentations effectuées dans les [Chapitre 2](#) et [Chapitre 4](#), nous avons utilisé la base de données publique ISIC-2018 [157, 33] pour réaliser cette étude comparative. Nous avons également suivi les méthodes de standardisation des couleurs et de normalisation des images, ainsi que la même répartition des données d’entraînement, de validation et de test que celles décrites dans la [Sous-section 2.3.1](#).

Entraînement des modèles : Comme dans l’étude présenté au [Chapitre 4](#), nous avons choisi l’architecture pré-entraînée EfficientNetB5 pour construire nos modèles. La dernière couche de l’architecture initiale a été modifiée afin de l’adapter selon les cas soit à une tâche de classification binaire, soit à une tâche de classification ternaire. L’entraînement des modèles a été réalisé en suivant la méthode d’entraînement découplé de bout en bout (MED) décrite dans la [Sous-section 3.4.3](#).

5.3.2. Résultats et discussions

Le [Tableau 5.3](#) reporte les performances précédemment obtenues avec la méthode [DAG-CNN](#) ainsi qu’avec la méthode de développement de [SAAD](#) à partir de la théorie des jeux. Le [Tableau 5.3](#) présente également les performances obtenues par chacune des deux méthodes précédemment évoquées combinées avec la méthode MED. On observe globalement que le fait d’entraîner les modèles suivant la méthode MED permet d’améliorer la performance des deux [SAADs](#) précédemment proposés. Plus spécifiquement, nous observons que la méthode [DAG-CNN](#) est celle qui a obtenu le gain de performance le plus significatif avec une augmentation de performance en terme de [BAcc](#) de 11%. Nous pensons que ce gain significatif n’est pas seulement dû à l’utilisation de la méthode MED, mais aussi au fait d’avoir remplacé l’architecture VGG19 par EfficientNetB5 qui semble plus adaptée pour la tâche de classification des cancers de la peau. En ce qui concerne la méthode de [SAAD](#) basée sur la théorie des jeux, nous observons également un gain de performance de 2% lorsque les modèles utilisés ont été entraînés à l’aide de la méthode MED. Dans ce dernier cas, étant donné que nous avons réutilisé la même architecture de [CNN](#), nous pouvons conclure que ce gain de performance est principalement dû à la méthode d’entraînement MED employée.

TABLEAU 5.3. – Étude comparative des performances des systèmes de SAADs développés dans le cadre de cette thèse avec ou sans la méthode d’entraînement découplé.

Méthode	Architecture	BACC
DAG-CNN*	VGG19	0.77
DAG-CNN + MED	EfficientNetB5	0.88
SAAD avec la théorie des jeux*	EfficientNetB5	0.87
SAAD avec la théorie des jeux + MED	EfficientNetB5	0.89

*résultats précédemment obtenues. MED est une abréviation de méthode d’entraînement découplé de bout en bout

5.4. Conclusion

Dans ce chapitre, nous avons tout d’abord présenté notre étude sur l’impact du **halo** périphérique sur la performance des **CNNs**, puis nous avons fait une étude comparative pour évaluer l’apport de la méthode MED introduite au [Chapitre 3](#) sur les deux systèmes de **SAADs** introduites aux [Chapitre 2](#) et [Chapitre 4](#).

Dans la première partie, nous avons établi pour la première fois que le **halo** fréquemment présent sur des images dermoscopiques impacte la performance des algorithmes d’apprentissage profond. Nous avons démontré que l’artefact **halo** interfère sur la prédiction diagnostic initialement correcte des mélanomes effectuée par un modèle et augmente le nombre de faux négatifs. Les résultats de nos expérimentations suggèrent qu’avec un simple rognage de l’image dermoscopique, il est possible de réduire le biais introduit par cet artefact. La conclusion de cette étude est qu’il est primordial de limiter la présence de cet artefact lors de la prise de photographies dermoscopiques ou de l’éliminer par traitement d’images avant leur utilisation pour l’entraînement d’un algorithme d’apprentissage profond.

Dans la deuxième partie, nous avons effectué une analyse comparative des performances obtenues par un **SAAD DAG-CNN** ou un **SAAD** basé sur la théorie des jeux dont les modèles neuronaux ont été entraînés ou pas à l’aide de la méthode MED. Globalement, nous avons pu observer dans les deux types de **SAAD**, que la méthode MED permettait d’obtenir un gain de performance minimal de 2%, permettant ainsi de conforter l’efficacité de cette méthode d’apprentissage.

Conclusion générale et perspectives

Dans ce qui suit, nous résumerons les principales contributions des recherches menées dans cette thèse et nous concluons en présentant des défis restants, les orientations futures et des idées qui mériteront d'être étudiées.

5.5. Conclusion générale

Le mélanome se développe en général à partir de l'épiderme, qui est une couche supérieure de la peau, ce qui conduit à ce que son diagnostic par le dermatologue puisse se faire par simple examen visuel des régions suspectes. Par ailleurs, cette caractéristique intrinsèque aux cancers de la peau fait en sorte que des images des lésions cutanées peuvent, à elles seules, être exploitées par des SAADs pour pouvoir poser un diagnostic préliminaire de manière in-vivo. L'avènement des techniques d'apprentissage profond a permis à la communauté de vision par ordinateur de développer de tels systèmes qui soient de plus en plus performants pour réaliser cette tâche. Toutefois, la quasi-majorité des recherches effectuées sur ce sujet demeurent dans un cadre expérimental, ceci dû à un certain nombre de problématiques de recherche restées ouvertes jusqu'à ce jour. Dans cette thèse, nous avons examiné de manière exhaustive les principales problématiques freinant l'utilisation en contexte clinique des SAADs pour diagnostiquer les lésions cutanées en général, et le cancer du mélanome en particulier. Parmi ces problématiques, nous nous sommes tout d'abord intéressés à l'amélioration de la précision des SAADs existants. En tentant de développer un SAAD performant capable de surpasser les systèmes existants, nous avons été confrontés à des défis majeurs portant sur la qualité des données disponibles et leur distribution asymétrique susceptible de créer un biais lors du développement de notre système. En outre, nous nous sommes également intéressés à la problématique visant à apporter plus de transparence au processus de décision des SAADs, un choix conforté au regard des exigences des normes éthiques et juridiques récentes vis-à-vis de l'IA et de leur application dans le domaine médical. Nous avons à cet effet construit de nouvelles architectures pour la tâche de diagnostic du mélanome permettant d'améliorer leurs performances, développé de nouvelle méthode d'apprentissage plus robustes face au déséquilibre des classes, et proposé des approches innovantes apportant plus de clarté dans le processus de décision des SAADs. Les méthodes proposées ont été développées et testées sur des données publiques réelles, ont été comparées avec

l'état de l'art des approches existantes, et obtenu d'excellents résultats. Un résumé des contributions de cette thèse et des principaux résultats est présenté ci-dessous.

Tout d'abord, au [Chapitre 2](#), nous avons proposé un [SAAD](#) pour classer les images dermoscopiques en trois classes de diagnostic : mélanome, naevus et kératose séborrhéique. De nos recherches bibliographiques, l'assemblage de différents modèles se révélait être la stratégie la plus performante pour améliorer les performances de ces systèmes. Nous avons donc à cet effet introduit un nouveau schéma d'assemblage de [CNNs](#) inspiré des méthodes de décomposition et d'ensemble pour améliorer les performances de notre [SAAD](#). Contrairement aux méthodes ensemblistes conventionnelles pour la tâche de détection du mélanome, nous avons utilisé un [DAG](#) pour agréger des [CNNs](#) binaires pour la tâche de détection du mélanome. Evalué sur le jeu de données public ISIC 2018, notre [SAAD](#) a atteint le meilleur score de [BAcc](#) de 76,6% comparé à des modèles multi-classes seuls, des ensembles de [CNNs](#) multi-classes réalisés avec des méthodes d'agrégation classiques et d'autres travaux connexes. Les résultats obtenus nous ont permis de démontrer que l'utilisation d'un [DAG](#) est une approche pertinente pour améliorer les performances d'un [SAAD](#) pour la tâche de classification multi-classes des images dermoscopiques. Toutefois, constat a été fait que la base de données que nous avons utilisée présentait une distribution asymétrique des classes. Autrement dit, dans cette base de données, mais aussi dans la quasi-majorité des bases de données publiques existantes pour cette tâche, il existe des types de lésions cutanées, à l'instar du Naevus, qui sont représentés avec un très grand effectif par rapport à d'autres types de lésions moins fréquentes dans la vie courante à l'instar des kératoses bénignes. Cette problématique connue sous le nom de déséquilibre des classes affecte l'apprentissage des modèles neuronaux entraînés sur ces données et engendre un biais favorisant la classe majoritaire au détriment de la classe sous-représentée.

Dans le [Chapitre 3](#), nous avons conduit deux expériences pouvant nous permettre d'effectuer une analyse approfondie de la manière selon laquelle une base de données déséquilibrée influençait l'apprentissage d'un modèle. De cette analyse, nous avons pu observer durant l'apprentissage, que le taux d'erreur commis par le modèle sur les échantillons de la classe sur-représentée diminuait beaucoup plus rapidement que celui sur la classe sous-représentée. En outre, fait intrigant, après la convergence du modèle, il existait des échantillons particulièrement de la classe sous-représentée dont le modèle non seulement n'avait pas été capable d'améliorer leur prédiction, mais qui en plus étaient restés quasiment figés sur leurs valeurs prédictives de départ. Ce fait nous a amené à déduire que ce type d'échantillons pourrait être assimilable à des données aberrantes ou des données mal étiquetées. Nous avons donc posé comme hypothèse qu'une stratégie permettant d'améliorer la robustesse des modèles vis-à-vis du déséquilibre des classes, pourrait être de simplement pondérer durant l'apprentissage les échantillons selon leur niveau de difficulté basé sur leurs valeurs prédictives. La deuxième hypothèse que nous avons posée portait sur les caractéristiques des fonctions de coût à utiliser dans le cadre d'un entraînement découplé.

Les hypothèses obtenues à l'issue de cette étude préliminaire nous ont permis de développer une nouvelle méthode d'apprentissage des modèles neuronaux robuste au déséquilibre des classes. Plus précisément, la méthode proposée initialise l'apprentissage d'un réseau à l'aide d'une nouvelle fonction de perte L_f que nous avons introduite et qui est capable de guider le modèle vers une meilleure représentation des caractéristiques. Ensuite, l'apprentissage du réseau ainsi pré-entraîné est affiné à l'aide d'une variante pondérée de la fonction L_f précédemment introduite, ceci avec pour objectif d'orienter le modèle à spécifiquement discriminer les différentes classes de manière plus robuste vis-à-vis du déséquilibre des classes. Nous avons évalué notre modèle sur la tâche de classification des lésions cutanées mais aussi sur deux autres tâches non connexes présentant cette même problématique de déséquilibre des classes. Les résultats ont révélé que notre méthode d'apprentissage permettait d'améliorer les performances comparativement à l'état de l'art et ceci avec une marge minimale de 2%. La même amélioration a été observée sur les autres tâches de classification. La méthode proposée a la capacité de servir de base solide pour renforcer la robustesse des modèles à d'autres domaines d'application de l'apprentissage profond rencontrant également cette problématique du déséquilibre des classes à l'instar de la segmentation d'images, la détection de fraude et le traitement du langage naturel.

Dans le [Chapitre 4](#), nous nous sommes intéressés à la nécessité d'améliorer l'explicabilité des SAADs, condition indispensable pour étendre leur utilisation en contexte clinique réel. L'aspect le plus difficile de ce travail était de proposer une méthode à même d'améliorer l'explicabilité du SAAD sans toutefois en réduire la performance. Poursuivant cet objectif, nous avons introduit un nouveau schéma d'apprentissage suivant un flux hiérarchique pour la détection du mélanome à partir d'images dermoscopiques en s'inspirant d'une démarche d'examen clinique du dermatologue. L'approche proposée est constituée de sept modèles CNNs agrégés à l'aide de la théorie des jeux. L'approche intègre également la visualisation des cartes thermiques à partir de l'algorithme Grad-CAM. Une des principales contributions de ce travail a été l'introduction d'une méthode pour évaluer le niveau de confiance d'une prédiction générée par un SAAD. Les résultats obtenus montrent que notre approche peut effectivement améliorer la précision des SAADs par rapport à l'état de l'art. De plus, lors de l'utilisation de notre SAAD par un dermatologue collaborant avec nous au sein de notre laboratoire, nous avons pu observer que le processus de décision de notre approche s'est avéré plus intuitif et explicable, ce qui soutiendrait son utilisation dans des contextes cliniques réels. D'ores et déjà, le système présenté a un grand potentiel d'utilisation dans des contextes cliniques réels en tant qu'outil de formation pour les dermatologues débutants.

Enfin, dans le [Chapitre 5](#), nous avons dans un premier temps présenté une étude succincte de l'impact du halo sur la performance des modèles. En effet, le halo est un artefact très fréquent sur les images dermoscopiques des bases de données publiques et affecte la qualité de ces données pour leur utilisation dans le cadre d'un apprentissage automatique. L'objectif de l'étude effectuée était donc d'analyser et

de quantifier l'influence du **halo** sur la performance diagnostic d'un **SAAD** basé sur un **CNN**. Les résultats de cette étude nous ont permis de démontrer que l'artefact **halo** interférait sur la prédiction diagnostic initialement correcte des mélanomes effectuée par un modèle et augmentait le nombre de faux négatifs. Cette observation nous a permis de conclure qu'il était primordial de limiter la présence de cet artefact lors de la prise de photographies dermoscopiques ou de l'éliminer par traitement d'images avant son utilisation pour l'entraînement d'un algorithme d'apprentissage profond. Un traitement d'image qui peut être appliqué par exemple serait simplement de rogner les images contenant cet artefact. Dans ce chapitre, nous avons également présenté les résultats de la combinaison de notre méthode d'apprentissage, présentée au **Chapitre 3** et permettant d'améliorer la robustesse des modèles à la problématique du déséquilibre des classes, avec chacun des deux **SAADs** présentés dans les **Chapitre 2** et **Chapitre 4**. Les résultats de cette dernière expérimentation nous ont permis de conforter l'efficacité de notre méthode d'apprentissage et ont permis d'améliorer avec une marge minimum de 2% les performances de nos deux schémas de **SAADs**.

5.6. Perspectives

5.6.1. Études approfondies sur l'impact des artefacts et des solutions

Dans le chapitre **Chapitre 5**, nous avons démontré que l'**halo** périphérique impactait négativement la performance des **SAADs**, basés sur des algorithmes d'apprentissage profond, pour la détection des mélanomes. La solution préliminaire que nous avons proposée pour résoudre cette problématique était d'appliquer un rognage automatique des images afin de réduire au minimum la surface de l'image présentant cet artefact. A notre connaissance, au moment de la rédaction de ce manuscrit, seuls deux autres travaux avaient été menés afin d'étudier cet artefact, dont un seul parmi les deux proposait une tentative de résolution de ce problème en appliquant deux méthodes consistant à peindre l'image. Récemment, une architecture en apprentissage profond dénommée réseau auto-codeur s'est révélée être très performante pour des tâches de débruitage d'images, et nous envisageons son utilisation pour supprimer l'**halo** périphérique des images dermoscopiques.

5.6.2. Études approfondies des échantillons aberrants

Dans le chapitre **Chapitre 3**, nous avons proposé une méthode d'entraînement pouvant augmenter la robustesse des modèles neuronaux vis à vis du déséquilibre des classes. Un des aspects de la méthode proposée est qu'elle réduit l'influence des valeurs aberrantes (pour rappel les échantillons dont la prédiction est très éloignée de leur étiquette réelle). Une étude approfondie de ces échantillons nous permettrait de mieux comprendre leur nature, leurs caractéristiques et pourrait nous permettre de développer une méthode plus spécifique à la résolution de cette problématique.

5.6.3. Mise en place d'une étude clinique de nos outils

Bien que l'objectif principal de notre thèse soit de contribuer à l'expansion des SAADs pour la détection des mélanomes en contexte clinique, tous les outils que nous avons développés ont été principalement réalisés dans un contexte expérimental. Nous avons pu dans bien des cas échanger avec une équipe de dermatologues de l'hôpital de la Timone à Marseille, et ainsi pu faire valider nos outils et intégrer leurs recommandations lors de nos développements. Il n'en demeure pas moins vrai qu'une étude clinique plus approfondie en intégrant un effectif plus large de cliniciens, permettrait d'encore mieux valider nos outils dans un contexte clinique réel.

5.6.4. Modélisation et intégration des concepts cliniques dans des SAADs afin d'améliorer leur explicabilité

En dermatologie, il existe un certain nombre de concepts cliniques à la disposition du praticien lui permettant de poser un diagnostic face à des lésions cutanées atypiques. Parmi ces concepts cliniques, nous pouvons citer ceux du « vilain petit canard » et de l'analyse des patrons définis au [Chapitre 1](#). Modéliser ces concepts cliniques et développer des méthodes d'intégration et/ou de combinaison de ces modèles avec des approches d'apprentissage profond, pourrait non seulement apporter plus d'explicabilité dans les SAADs, mais permettrait aussi d'améliorer leurs performances.

Bibliographie

- [1] Shigeo ABE. « Analysis of Multiclass Support Vector Machines ». In : *International Conference on Computational Intelligence for Modelling Control and Automation* (jan. 2003).
- [2] David H. ACKLEY, Geoffrey E. HINTON et Terrence J. SEJNOWSKI. « A learning algorithm for boltzmann machines ». en. In : *Cognitive Science* 9.1 (jan. 1985), p. 147-169. ISSN : 0364-0213. DOI : [10.1016/S0364-0213\(85\)80012-4](https://doi.org/10.1016/S0364-0213(85)80012-4).
- [3] Adewole S ADAMSON et Avery SMITH. « Machine learning and health care disparities in dermatology ». In : *JAMA dermatology* 154.11 (2018). Publisher : American Medical Association, p. 1247-1248. DOI : [10.1001/jamadermatol.2018.2348](https://doi.org/10.1001/jamadermatol.2018.2348).
- [4] Adekanmi ADEGUN et Serestina VIRIRI. « Deep learning techniques for skin lesion analysis and melanoma cancer detection : a survey of state-of-the-art ». en. In : *Artif Intell Rev* 54.2 (fév. 2021), p. 811-841. ISSN : 1573-7462. DOI : [10.1007/s10462-020-09865-y](https://doi.org/10.1007/s10462-020-09865-y).
- [5] Miguel Nehmad ALCHE, Daniel ACEVEDO et Marta MEJAIL. « EfficientARL : improving skin cancer diagnoses by combining lightweight attention on Efficient-Net ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 3354-3360. DOI : [10.1109/ICCVW54120.2021.00374](https://doi.org/10.1109/ICCVW54120.2021.00374).
- [6] Abder-Rahman A. ALI et Thomas M. DESERNO. « A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data ». In : *Medical Imaging 2012 : Image Perception, Observer Performance, and Technology Assessment*. T. 8318. SPIE, fév. 2012, p. 421-431. DOI : [10.1117/12.912389](https://doi.org/10.1117/12.912389).
- [7] Adamu ALI-GOMBE et Eyad ELYAN. « MFC-GAN : Class-imbalanced dataset classification using Multiple Fake Class Generative Adversarial Network ». en. In : *Neurocomputing* 361 (oct. 2019), p. 212-221. ISSN : 0925-2312. DOI : [10.1016/j.neucom.2019.06.043](https://doi.org/10.1016/j.neucom.2019.06.043).
- [8] Stefano ALLEGRETTI et al. « Supporting skin lesion diagnosis with content-based image retrieval ». In : *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021, p. 8053-8060. DOI : [10.1109/ICPR48806.2021.9412419](https://doi.org/10.1109/ICPR48806.2021.9412419).

- [9] Jose-Agustin ALMARAZ-DAMIAN et al. « Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures ». en. In : *Entropy* 22.4 (avr. 2020). Number : 4 Publisher : Multidisciplinary Digital Publishing Institute, p. 484. ISSN : 1099-4300. DOI : [10.3390/e22040484](https://doi.org/10.3390/e22040484).
- [10] Beatriz ALVES, Catarina BARATA et Jorge S MARQUES. « Diagnosis of Skin Cancer Using Hierarchical Neural Networks and Metadata ». In : *Iberian Conference on Pattern Recognition and Image Analysis*. Springer. 2022, p. 69-80. DOI : [10.1007/978-3-031-04881-4_6](https://doi.org/10.1007/978-3-031-04881-4_6).
- [11] Giuseppe ARGENZIANO et H Peter SOYER. « Dermoscopy of pigmented skin lesions – a valuable tool for early ». en. In : *The Lancet Oncology* 2.7 (juill. 2001), p. 443-449. ISSN : 1470-2045. DOI : [10.1016/S1470-2045\(00\)00422-8](https://doi.org/10.1016/S1470-2045(00)00422-8).
- [12] Giuseppe ARGENZIANO et al. « Dermoscopy of pigmented skin lesions : results of a consensus meeting via the Internet ». In : *Journal of the American Academy of Dermatology* 48.5 (2003), p. 679-693. DOI : [10.1067/mj.2003.281](https://doi.org/10.1067/mj.2003.281).
- [13] Alejandro Barredo ARRIETA et al. « Explainable Artificial Intelligence (XAI) : Concepts, taxonomies, opportunities and challenges toward responsible AI ». In : *Information fusion* 58 (2020), p. 82-115. DOI : [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012).
- [14] Yuri Sousa AURELIO et al. « Learning from imbalanced data sets with weighted cross-entropy function ». In : *Neural processing letters* 50.2 (2019), p. 1937-1949. DOI : [10.1007/s11063-018-09977-1](https://doi.org/10.1007/s11063-018-09977-1).
- [15] Catarina BARATA, M Emre CELEBI et Jorge S MARQUES. « Explainable skin lesion diagnosis using taxonomies ». In : *Pattern Recognition* 110 (2021), p. 107413. DOI : [10.1016/j.patcog.2020.107413](https://doi.org/10.1016/j.patcog.2020.107413).
- [16] Catarina BARATA, M Emre CELEBI et Jorge S MARQUES. « Improving Dermoscopy Image Classification Using Color Constancy ». In : *IEEE Journal of Biomedical and Health Informatics* 19.3 (mai 2015). Conference Name : IEEE Journal of Biomedical and Health Informatics, p. 1146-1152. ISSN : 2168-2208. DOI : [10.1109/JBHI.2014.2336473](https://doi.org/10.1109/JBHI.2014.2336473).
- [17] Catarina BARATA et Jorge S MARQUES. « Deep learning for skin cancer diagnosis with hierarchical architectures ». In : *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. ISSN : 1945-8452. IEEE. 2019, p. 841-845. DOI : [10.1109/ISBI.2019.8759561](https://doi.org/10.1109/ISBI.2019.8759561).
- [18] Catarina BARATA et Carlos SANTIAGO. « Improving the explainability of skin cancer diagnosis using CBIR ». In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2021, p. 550-559. DOI : [10.1007/978-3-030-87199-4_52](https://doi.org/10.1007/978-3-030-87199-4_52).
- [19] Julia BENEDETTI. *Structure et fonction de la peau*. URL : <https://www.msmanuals.com/fr/accueil/troubles-cutan%C3%A9s/biologie-de-la-peau/structure-et-fonction-de-la-peau> (visité le 11/07/2022).

- [20] Devansh BISLA et al. « Towards automated melanoma detection with deep learning : Data purification and augmentation ». In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2019. DOI : <https://doi.org/10.48550/arXiv.1902.06061>.
- [21] Alceu BISSOTO et al. « (De) Constructing Bias on Skin Lesion Datasets ». In : *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2019, p. 2766-2774. DOI : [10.1109/CVPRW.2019.00335](https://doi.org/10.1109/CVPRW.2019.00335).
- [22] Léon BOTTOU. « Stochastic Gradient Descent Tricks ». In : *Neural Networks : Tricks of the Trade : Second Edition*. Sous la dir. de Grégoire MONTAVON, Geneviève B. ORR et Klaus-Robert MÜLLER. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 2012, p. 421-436. ISBN : 978-3-642-35289-8. DOI : [10.1007/978-3-642-35289-8_25](https://doi.org/10.1007/978-3-642-35289-8_25).
- [23] Leo BREIMAN. « Bagging predictors ». en. In : *Mach Learn* 24.2 (août 1996), p. 123-140. ISSN : 1573-0565. DOI : [10.1007/BF00058655](https://doi.org/10.1007/BF00058655).
- [24] Mateusz BUDA, Atsuto MAKI et Maciej A. MAZUROWSKI. « A systematic study of the class imbalance problem in convolutional neural networks ». en. In : *Neural Networks* 106 (oct. 2018), p. 249-259. ISSN : 0893-6080. DOI : [10.1016/j.neunet.2018.07.011](https://doi.org/10.1016/j.neunet.2018.07.011).
- [25] Kaidi CAO et al. « Learning imbalanced datasets with label-distribution-aware margin loss ». In : *Advances in neural information processing systems* 32 (2019). DOI : [10.48550/arXiv.1906.07413](https://doi.org/10.48550/arXiv.1906.07413).
- [26] Cristina CARRERA et al. « Validity and reliability of dermoscopic criteria used to differentiate nevi from melanoma : a web-based international dermoscopy society study ». In : *JAMA dermatology* 152.7 (2016). Publisher : American Medical Association, p. 798-806. DOI : [10.1001/jamadermatol.2016.0624](https://doi.org/10.1001/jamadermatol.2016.0624).
- [27] Nitesh V CHAWLA. « Data Mining for Imbalanced Datasets : An Overview ». In : *Data Mining and Knowledge Discovery Handbook* (2010), p. 875. DOI : [10.1007/978-0-387-09823-4_45](https://doi.org/10.1007/978-0-387-09823-4_45).
- [28] Yixin CHEN, Yifan ZHU et Yanfeng CHANG. « CycleGAN Based Data Augmentation For Melanoma images Classification ». In : *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition*. 2020, p. 115-119. DOI : [10.1145/3430199.3430217](https://doi.org/10.1145/3430199.3430217).
- [29] Yushi CHEN et al. « Deep Learning Ensemble for Hyperspectral Image Classification ». In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.6 (juin 2019). Conference Name : IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, p. 1882-1897. ISSN : 2151-1535. DOI : [10.1109/JSTARS.2019.2915259](https://doi.org/10.1109/JSTARS.2019.2915259).
- [30] Travers CHING et al. « Opportunities and obstacles for deep learning in biology and medicine ». In : *Journal of The Royal Society Interface* 15.141 (2018). Publisher : The Royal Society, p. 20170387. DOI : [10.1098/rsif.2017.0387](https://doi.org/10.1098/rsif.2017.0387).

- [31] Noel CODELLA et al. « Skin lesion analysis toward melanoma detection 2018 : A challenge hosted by the international skin imaging collaboration (isic) ». In : *arXiv preprint arXiv :1902.03368* (2019). DOI : [10.48550/arXiv.1902.03368](https://doi.org/10.48550/arXiv.1902.03368).
- [32] Noel CF CODELLA et al. « Collaborative Human-AI (CHAI) : Evidence-based interpretable melanoma classification in dermoscopic images ». In : *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer, 2018, p. 97-105. DOI : [10.1007/978-3-030-02628-8_11](https://doi.org/10.1007/978-3-030-02628-8_11).
- [33] Noel CF CODELLA 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)*. IEEE, 2018, p. 168-172.
- [34] Marc COMBALIA et al. « Bcn20000 : Dermoscopic lesions in the wild ». In : *arXiv preprint arXiv :1908.02288* (2019).
- [35] Michael D CORBO, Ronald VENDER et Judy WISMER. « Comparison of dermatologists' and nondermatologists' diagnostic accuracy for malignant melanoma ». In : *Journal of cutaneous medicine and surgery* 16.4 (2012). Publisher : SAGE Publications Sage CA : Los Angeles, CA, p. 272-280. DOI : [10.1177/120347541201600410](https://doi.org/10.1177/120347541201600410).
- [36] Ekin D CUBUK et al. « Autoaugment : Learning augmentation strategies from data ». In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, p. 113-123. DOI : [10.1109/CVPR.2019.00020](https://doi.org/10.1109/CVPR.2019.00020).
- [37] Ekin D CUBUK et al. « Randaugment : Practical automated data augmentation with a reduced search space ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, p. 702-703. DOI : [10.48550/arXiv.1909.13719](https://doi.org/10.48550/arXiv.1909.13719).
- [38] Yin CUI et al. « Class-balanced loss based on effective number of samples ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, p. 9268-9277. DOI : [10.48550/arXiv.1901.05555](https://doi.org/10.48550/arXiv.1901.05555).
- [39] Clara CURIEL-LEWANDROWSKI et al. « Artificial intelligence approach in melanoma ». In : *Melanoma* (2019), p. 1-31. DOI : [10.1007/978-1-4614-7147-9_43](https://doi.org/10.1007/978-1-4614-7147-9_43).
- [40] Kumari DEEPSHIKHA et Anugunj NAMAN. « Removing Class Imbalance using Polarity-GAN : An Uncertainty Sampling Approach ». In : *arXiv preprint arXiv :2012.04937 [cs]* (déc. 2020). DOI : [10.48550/arXiv.2012.04937](https://doi.org/10.48550/arXiv.2012.04937).
- [41] Gautier DEFOSSEZ et al. *Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018*. 2019. ISBN : 979-10-289-0559-0. URL : https://www.e-cancer.fr/content/download/266450/3759432/file/Synthese_Estimations%20nationales%20incidence%20et%20mortalite%20par%20cancer_juillet_2019.pdf (visité le 11/07/2022).

- [42] Alakh DESAI et al. « Learning of visual relations : The devil is in the tails ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 15404-15413. DOI : [10.1109/ICCV48922.2021.01512](https://doi.org/10.1109/ICCV48922.2021.01512).
- [43] Alakh DESAI et al. « Learning of visual relations : The devil is in the tails ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 15404-15413. DOI : [10.48550/arXiv.2108.09668](https://doi.org/10.48550/arXiv.2108.09668).
- [44] Terrance DEVRIES et Graham W. TAYLOR. « Improved Regularization of Convolutional Neural Networks with Cutout ». In : *arXiv preprint arXiv :1708.04552* (2017). DOI : [10.48550/arXiv.1708.04552](https://doi.org/10.48550/arXiv.1708.04552).
- [45] Mehwish DILDAR et al. « Skin Cancer Detection : A Review Using Deep Learning Techniques ». en. In : *International Journal of Environmental Research and Public Health* 18.10 (jan. 2021). Number : 10 Publisher : Multidisciplinary Digital Publishing Institute, p. 5479. ISSN : 1660-4601. DOI : [10.3390/ijerph18105479](https://doi.org/10.3390/ijerph18105479).
- [46] Qi DONG, Shaogang GONG et Xiatian ZHU. « Class rectification hard mining for imbalanced deep learning ». In : *Proceedings of the IEEE International Conference on Computer Vision*. 2017, p. 1851-1860. DOI : [10.48550/arXiv.1712.03162](https://doi.org/10.48550/arXiv.1712.03162).
- [47] Ana F DUARTE et al. « Clinical ABCDE rule for early melanoma detection ». In : *European Journal of Dermatology* 31.6 (2021), p. 771-778. DOI : [10.1684/ejd.2021.4171](https://doi.org/10.1684/ejd.2021.4171).
- [48] Patricia DUBOT. « Etude du rôle de la glucosylcéramidase dans le mélanome cutané ». fr. Thèse de doct. Université Paul Sabatier - Toulouse III, déc. 2021. URL : <https://tel.archives-ouvertes.fr/tel-03663094> (visité le 11/07/2022).
- [49] Shireen DUMONT. « Dépistage du cancer de la peau en Suisse : étude transversale des tendances (1997-2012) des inégalités socioéconomiques ». fre. Thèse de doct. University of Geneva, 2021. DOI : [10.13097/archive-ouverte/unige:151525](https://doi.org/10.13097/archive-ouverte/unige:151525).
- [50] Olive Jean DUNN. « Multiple comparisons using rank sums ». In : *Technometrics* 6.3 (1964), p. 241-252. DOI : [10.1080/00401706.1964.10490181](https://doi.org/10.1080/00401706.1964.10490181).
- [51] Charles ELKAN. « The foundations of cost-sensitive learning ». In : 17.1 (2001), p. 973-978.
- [52] Dirk M ELSTON. « Management of atypical pigmented lesions ». In : *Journal of the American Academy of Dermatology* 70.1 (2014), p. 142-145. DOI : [10.1016/j.jaad.2013.08.003](https://doi.org/10.1016/j.jaad.2013.08.003).
- [53] Andre ESTEVA et al. « Dermatologist-level classification of skin cancer with deep neural networks ». In : *Nature* 542.7639 (fév. 2017), p. 115-118. ISSN : 1476-4687. DOI : [10.1038/nature21056](https://doi.org/10.1038/nature21056).

- [54] G. FABBROCINI et al. « Teledermatology : from prevention to diagnosis of non-melanoma and melanoma skin cancer ». In : *Int. J. Telemedicine Appl.* 2011 (jan. 2011), 1 :1-1 :5. ISSN : 1687-6415. DOI : [10.1155/2011/125762](https://doi.org/10.1155/2011/125762).
- [55] Xusheng FANG, Zhenbing LIU et Mingchang XU. « Ensemble of deep convolutional neural networks based multi-modality images for Alzheimer's disease diagnosis ». en. In : *IET Image Processing* 14.2 (2020), p. 318-326. ISSN : 1751-9667. DOI : [10.1049/iet-ipr.2019.0617](https://doi.org/10.1049/iet-ipr.2019.0617).
- [56] Alberto FERNANDEZ et al. « SMOTE for Learning from Imbalanced Data : Progress and Challenges, Marking the 15-year Anniversary ». en. In : *Journal of Artificial Intelligence Research* 61 (avr. 2018), p. 863-905. ISSN : 1076-9757. DOI : [10.1613/jair.1.11192](https://doi.org/10.1613/jair.1.11192).
- [57] Le FIGARO. *Structures et rôles de la peau*. URL : <https://sante.lefigaro.fr/mieux-etre/beaute/structures-roles-peau/quoi-peau-est-elle-composee> (visité le 11/07/2022).
- [58] Anna FINNANE et al. « Proposed technical guidelines for the acquisition of clinical images of skin-related conditions ». In : *JAMA dermatology* 153.5 (2017). Publisher : American Medical Association, p. 453-457. DOI : [10.1001/jamadermatol.2016.6214](https://doi.org/10.1001/jamadermatol.2016.6214).
- [59] Arthur Cartel FOAHOM GOUABOU, Frédéric HEIM et Jean-Luc DAMOISEAUX. *Apprentissage profond pour l'aide au diagnostic du mélanome à partir d'exemple*. Research Report. Université de Haute-Alsace, août 2019. URL : <https://hal-amu.archives-ouvertes.fr/hal-02515203> (visité le 11/07/2022).
- [60] Arthur Cartel FOAHOM GOUABOU et al. « End-to-End Decoupled Training : A Robust Deep Learning Method for Long-Tailed Classification of Dermoscopic Images for Skin Lesion Classification ». In : *Electronics* 11.20 (2022), p. 3275. DOI : [10.3390/electronics11203275](https://doi.org/10.3390/electronics11203275).
- [61] Arthur Cartel FOAHOM GOUABOU et al. « Ensemble Method of Convolutional Neural Networks with Directed Acyclic Graph Using Dermoscopic Images : Melanoma Detection Application ». In : *Sensors* 21.12 (2021), p. 3999. DOI : [10.3390/s21123999](https://doi.org/10.3390/s21123999).
- [62] Y. FREUND. « Boosting a Weak Learning Algorithm by Majority ». en. In : *Information and Computation* 121.2 (sept. 1995), p. 256-285. ISSN : 0890-5401. DOI : [10.1006/inco.1995.1136](https://doi.org/10.1006/inco.1995.1136).
- [63] Mikel GALAR et al. « An overview of ensemble methods for binary classifiers in multi-class problems : Experimental study on one-vs-one and one-vs-all schemes ». en. In : *Pattern Recognition* 44.8 (août 2011), p. 1761-1776. ISSN : 0031-3203. DOI : [10.1016/j.patcog.2011.01.017](https://doi.org/10.1016/j.patcog.2011.01.017).

- [64] Rishu GARG, Saumil MAHESHWARI et Anupam SHUKLA. « Decision Support System for Detection and Classification of Skin Cancer Using CNN ». en. In : *Innovations in Computational Intelligence and Computer Vision*. Sous la dir. de Manoj Kumar SHARMA et al. Advances in Intelligent Systems and Computing. Singapore : Springer, 2021, p. 578-586. ISBN : 9789811560675. DOI : [10.1007/978-981-15-6067-5_65](https://doi.org/10.1007/978-981-15-6067-5_65).
- [65] Caroline GAUDY-MARQUESTE et al. « Ugly Duckling Sign as a Major Factor of Efficiency in Melanoma Detection ». In : *JAMA Dermatology* 153.4 (avr. 2017), p. 279-284. ISSN : 2168-6068. DOI : [10.1001/jamadermatol.2016.5500](https://doi.org/10.1001/jamadermatol.2016.5500).
- [66] Jeffrey E. GERSHENWALD et Gery P. GUY Jr. « Stemming the Rising Incidence of Melanoma : Calling Prevention to Action ». In : *JNCI : Journal of the National Cancer Institute* 108.1 (jan. 2016), djv381. ISSN : 0027-8874. DOI : [10.1093/jnci/djv381](https://doi.org/10.1093/jnci/djv381).
- [67] Nils GESSERT et al. « Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data ». en. In : *MethodsX* 7 (jan. 2020), p. 100864. ISSN : 2215-0161. DOI : [10.1016/j.mex.2020.100864](https://doi.org/10.1016/j.mex.2020.100864).
- [68] Arjan GIJSENIJ, Theo GEVERS et Joost van de WEIJER. « Computational Color Constancy : Survey and Experiments ». In : *IEEE Transactions on Image Processing* 20.9 (sept. 2011). Conference Name : IEEE Transactions on Image Processing, p. 2475-2489. ISSN : 1941-0042. DOI : [10.1109/TIP.2011.2118224](https://doi.org/10.1109/TIP.2011.2118224).
- [69] Yana GONCHAROVA et al. « Dermoscopic features of facial pigmented skin lesions ». In : *International Scholarly Research Notices* 2013 (2013). DOI : [10.1155/2013/546813](https://doi.org/10.1155/2013/546813).
- [70] An GONG, Xinjie YAO et Wei LIN. « Classification for Dermoscopy Images Using Convolutional Neural Networks Based on the Ensemble of Individual Advantage and Group Decision ». In : *IEEE Access* 8 (2020). Conference Name : IEEE Access, p. 155337-155351. ISSN : 2169-3536. DOI : [10.1109/ACCESS.2020.3019210](https://doi.org/10.1109/ACCESS.2020.3019210).
- [71] Manu GOYAL et al. « Artificial intelligence-based image classification methods for diagnosis of skin cancer : Challenges and opportunities ». en. In : *Computers in Biology and Medicine* 127 (déc. 2020), p. 104065. ISSN : 0010-4825. DOI : [10.1016/j.combiomed.2020.104065](https://doi.org/10.1016/j.combiomed.2020.104065).
- [72] J. J. GROB et J. J. BONERANDI. « The 'Ugly Duckling' Sign : Identification of the Common Characteristics of Nevi in an Individual as a Basis for Melanoma Screening ». In : *Archives of Dermatology* 134.1 (jan. 1998), p. 103-104. ISSN : 0003-987X.
- [73] Deepti GUPTA et al. « Game Theory Based Privacy Preserving Approach for Collaborative Deep Learning in IoT ». In : *Deep Learning for Security and Privacy Preservation in IoT*. Springer, 2021, p. 127-149. DOI : [10.1007/978-981-16-6186-0_8](https://doi.org/10.1007/978-981-16-6186-0_8).

- [74] Holger A HAENSSLE et al. « Man against machine : diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists ». In : *Annals of Oncology* 29.8 (2018). Publisher : Oxford University Press, p. 1836-1842. DOI : [10.1093/annonc/mdy166](https://doi.org/10.1093/annonc/mdy166).
- [75] Jason R. HAGERTY et al. « Deep Learning and Handcrafted Method Fusion : Higher Diagnostic Accuracy for Melanoma Dermoscopy Images ». In : *IEEE Journal of Biomedical and Health Informatics* 23.4 (juill. 2019). Conference Name : IEEE Journal of Biomedical and Health Informatics, p. 1385-1391. ISSN : 2168-2208. DOI : [10.1109/JBHI.2019.2891049](https://doi.org/10.1109/JBHI.2019.2891049).
- [76] Zabit HAMEED et al. « Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models ». en. In : *Sensors* 20.16 (jan. 2020). Number : 16 Publisher : Multidisciplinary Digital Publishing Institute, p. 4373. ISSN : 1424-8220. DOI : [10.3390/s20164373](https://doi.org/10.3390/s20164373).
- [77] Shizhong HAN et al. « Incremental boosting convolutional neural network for facial action unit recognition ». In : *Advances in neural information processing systems* 29 (2016). DOI : [10.48550/arXiv.1707.05395](https://doi.org/10.48550/arXiv.1707.05395).
- [78] Balazs HARANGI. « Skin lesion classification with ensembles of deep convolutional neural networks ». en. In : *Journal of Biomedical Informatics* 86 (oct. 2018), p. 25-32. ISSN : 1532-0464. DOI : [10.1016/j.jbi.2018.08.006](https://doi.org/10.1016/j.jbi.2018.08.006).
- [79] Tanmoy HAZRA et Kushal ANJARIA. « Applications of game theory in deep learning : a survey ». In : *Multimedia Tools and Applications* 81.6 (2022), p. 8963-8994. DOI : [10.1007/s11042-022-12153-2](https://doi.org/10.1007/s11042-022-12153-2).
- [80] Kaiming HE et al. « Deep residual learning for image recognition ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 770-778. DOI : [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [81] Kaiming HE et al. « Delving deep into rectifiers : Surpassing human-level performance on imagenet classification ». In : *Proceedings of the IEEE international conference on computer vision*. 2015, p. 1026-1034. DOI : [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- [82] Geoffrey E HINTON et al. « Learning distributed representations of concepts ». In : *Proceedings of the eighth annual conference of the cognitive science society*. T. 1. Amherst, MA, 1986, p. 12.
- [83] Youngkyu HONG et al. « Disentangling label distribution for long-tailed visual recognition ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, p. 6626-6636. DOI : [10.48550/arXiv.2012.00321](https://doi.org/10.48550/arXiv.2012.00321).
- [84] Azadeh Noori HOSHYAR, Adel AL-JUMAILY et Riza SULAIMAN. « Review on automatic early skin cancer detection ». In : *2011 International Conference on Computer Science and Service System (CSSS)*. Juin 2011, p. 4036-4039. DOI : [10.1109/CSSS.2011.5974581](https://doi.org/10.1109/CSSS.2011.5974581).

- [85] Loretta ICHIM et Dan POPESCU. « Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks ». In : *IEEE Access* 8 (2020). Conference Name : IEEE Access, p. 179189-179202. ISSN : 2169-3536. DOI : [10.1109/ACCESS.2020.3028248](https://doi.org/10.1109/ACCESS.2020.3028248).
- [86] Tharindu KALUARACHCHI, Andrew REIS et Suranga NANAYAKKARA. « A review of recent deep learning approaches in human-centered machine learning ». In : *Sensors* 21.7 (2021), p. 2514. DOI : [10.3390/s21072514](https://doi.org/10.3390/s21072514).
- [87] Bingyi KANG et al. « Decoupling Representation and Classifier for Long-Tailed Recognition ». In : *International Conference on Learning Representations*. 2019. DOI : [10.48550/arXiv.1910.09217](https://doi.org/10.48550/arXiv.1910.09217).
- [88] Sagar KARKI, Pradnya KULKARNI et Andrew STRANIERI. « Melanoma classification using EfficientNets and Ensemble of models with different input resolution ». In : *2021 Australasian Computer Science Week Multiconference*. 2021, p. 1-5. DOI : [10.1145/3437378.3437396](https://doi.org/10.1145/3437378.3437396).
- [89] Nabeela KAUSAR et al. « Multiclass Skin Cancer Classification Using Ensemble of Fine-Tuned Deep Learning Models ». In : *Applied Sciences* 11.22 (2021), p. 10593. DOI : [10.3390/app112210593](https://doi.org/10.3390/app112210593).
- [90] Jeremy KAWAHARA et Ghassan HAMARNEH. « Multi-resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers ». en. In : *Machine Learning in Medical Imaging*. Sous la dir. de Li WANG et al. Lecture Notes in Computer Science. Cham : Springer International Publishing, 2016, p. 164-171. ISBN : 978-3-319-47157-0. DOI : [10.1007/978-3-319-47157-0_20](https://doi.org/10.1007/978-3-319-47157-0_20).
- [91] Sertan KAYMAK, Parvaneh ESMAILI et Ali SERENER. « Deep learning for two-step classification of malignant pigmented skin lesions ». In : *2018 14th Symposium on Neural Networks and Applications (NEUREL)*. IEEE. 2018, p. 1-6. DOI : [10.1109/NEUREL.2018.8587019](https://doi.org/10.1109/NEUREL.2018.8587019).
- [92] Hassan EL-KHATIB, Dan POPESCU et Loretta ICHIM. « Deep Learning–Based Methods for Automatic Diagnosis of Skin Lesions ». en. In : *Sensors* 20.6 (jan. 2020). Number : 6 Publisher : Multidisciplinary Digital Publishing Institute, p. 1753. ISSN : 1424-8220. DOI : [10.3390/s20061753](https://doi.org/10.3390/s20061753).
- [93] Diederik P. KINGMA et Jimmy BA. « Adam : A Method for Stochastic Optimization ». In : *arXiv preprint arXiv :1412.6980 [cs]* (jan. 2017). DOI : [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- [94] Alex KRIZHEVSKY, Geoffrey HINTON et al. « Learning multiple layers of features from tiny images ». In : (2009).
- [95] Buyu LI, Yu LIU et Xiaogang WANG. « Gradient harmonized single-stage detector ». In : *Proceedings of the AAAI conference on artificial intelligence*. T. 33. 01. 2019, p. 8577-8584. DOI : [10.1609/aaai.v33i01.33018577](https://doi.org/10.1609/aaai.v33i01.33018577).
- [96] Henry W LIM et al. « The burden of skin disease in the United States ». In : *Journal of the American Academy of Dermatology* 76.5 (2017). Publisher : Elsevier, p. 958-972. DOI : [10.1016/j.jaad.2016.12.043](https://doi.org/10.1016/j.jaad.2016.12.043).

- [97] Tsung-Yi LIN et al. « Focal loss for dense object detection ». In : *Proceedings of the IEEE international conference on computer vision*. 2017, p. 2980-2988. DOI : [10.1109/ICCV.2017.324](https://doi.org/10.1109/ICCV.2017.324).
- [98] Jennifer A. LO et David E. FISHER. « The melanoma revolution : From UV carcinogenesis to a new era in therapeutics ». In : *Science* 346.6212 (nov. 2014). Publisher : American Association for the Advancement of Science, p. 945-949. DOI : [10.1126/science.1253735](https://doi.org/10.1126/science.1253735).
- [99] Robert LOGAN et al. « Deep Convolutional Neural Networks With Ensemble Learning and Generative Adversarial Networks for Alzheimer’s Disease Image Data Classification ». In : *Frontiers in Aging Neuroscience* 13 (2021). ISSN : 1663-4365. DOI : [10.3389/fnagi.2021.720226](https://doi.org/10.3389/fnagi.2021.720226).
- [100] Adriano LUCIERI, Andreas DENGEL et Sheraz AHMED. « Deep Learning Based Decision Support for Medicine—A Case Study on Skin Cancer Diagnosis ». In : *arXiv preprint arXiv :2103.05112* (2021). DOI : [10.48550/arXiv.2103.05112](https://doi.org/10.48550/arXiv.2103.05112).
- [101] Maximiliano LUCIUS et al. « Deep Neural Frameworks Improve the Accuracy of General Practitioners in the Classification of Pigmented Skin Lesions ». In : *Diagnostics* 10.11 (2020). Number : 11 Publisher : Multidisciplinary Digital Publishing Institute, p. 969. ISSN : 2075-4418. DOI : [10.3390/diagnostics10110969](https://doi.org/10.3390/diagnostics10110969).
- [102] Scott M LUNDBERG et Su-In LEE. « A unified approach to interpreting model predictions ». In : *Advances in neural information processing systems* 30 (2017). DOI : [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874).
- [103] Amirreza MAHBOD et al. « Fusing fine-tuned deep features for skin lesion classification ». en. In : *Computerized Medical Imaging and Graphics* 71 (jan. 2019), p. 19-29. ISSN : 0895-6111. DOI : [10.1016/j.compmedimag.2018.10.007](https://doi.org/10.1016/j.compmedimag.2018.10.007).
- [104] Tomas MAJTNER, Sule YILDIRIM-YAYILGAN et Jon Yngve HARDEBERG. « Combining deep learning and hand-crafted features for skin lesion classification ». In : *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE. 2016, p. 1-6. DOI : [10.1109/IPTA.2016.7821017](https://doi.org/10.1109/IPTA.2016.7821017).
- [105] Estelle MALDONADO. « Le mélanome cutané : stratégie thérapeutique et rôle du pharmacien d’officine ». In : (2020). URL : <https://dumas.ccsd.cnrs.fr/dumas-03101487> (visité le 11/07/2022).
- [106] Inderjeet MANI et I ZHANG. « kNN approach to unbalanced data distributions : a case study involving information extraction ». In : *Proceedings of workshop on learning from imbalanced datasets*. T. 126. ICML. 2003, p. 1-7.
- [107] Roman C MARON et al. « Reducing the impact of confounding factors on skin cancer classification via image segmentation : technical model study ». In : *Journal of Medical Internet Research* 23.3 (2021), e21695. DOI : [10.2196/21695](https://doi.org/10.2196/21695).
- [108] Roman C MARON et al. « Robustness of convolutional neural networks in recognition of pigmented skin lesions ». In : *European journal of cancer* 145 (2021), p. 81-91. DOI : [10.1016/j.ejca.2020.11.020](https://doi.org/10.1016/j.ejca.2020.11.020).

- [109] Mohammed A AL-MASNI, Dong-Hyun KIM et Tae-Seong KIM. « Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification ». In : *Computer methods and programs in biomedicine* 190 (2020), p. 105351. DOI : [10.1016/j.cmpb.2020.105351](https://doi.org/10.1016/j.cmpb.2020.105351).
- [110] Maciej A MAZUROWSKI et al. « Training neural network classifiers for medical decision making : The effects of imbalanced datasets on classification performance ». In : *Neural networks* 21.2-3 (2008). Publisher : Elsevier, p. 427-436. DOI : [10.1016/j.neunet.2007.12.031](https://doi.org/10.1016/j.neunet.2007.12.031).
- [111] Patrick E MCKIGHT et Julius NAJAB. « Kruskal-wallis test ». In : *The corsini encyclopedia of psychology* (2010), p. 1-1. DOI : [10.1002/9780470479216.corpsy0491](https://doi.org/10.1002/9780470479216.corpsy0491).
- [112] Afonso MENEGOLA et al. « Knowledge transfer for melanoma screening with deep learning ». In : *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. ISSN : 1945-8452. Avr. 2017, p. 297-300. DOI : [10.1109/ISBI.2017.7950523](https://doi.org/10.1109/ISBI.2017.7950523).
- [113] Aditya Krishna MENON et al. « Long-tail learning via logit adjustment ». In : *arXiv preprint arXiv :2007.07314* (2020). DOI : [10.48550/arXiv.2007.07314](https://doi.org/10.48550/arXiv.2007.07314).
- [114] Riccardo MIOTTO et al. « Deep learning for healthcare : review, opportunities and challenges ». In : *Briefings in Bioinformatics* 19 (nov. 2018), p. 1236-1246. ISSN : 1477-4054. DOI : [10.1093/bib/bbx044](https://doi.org/10.1093/bib/bbx044).
- [115] Mohammad MOGHIMI et al. « Boosted convolutional neural networks. » In : *BMVC*. T. 5. 2016, p. 6. DOI : [10.5244/C.30.24](https://doi.org/10.5244/C.30.24).
- [116] Shorf MOHAMMAD. « An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection ». In : *Multimedia Systems* (avr. 2021). DOI : [10.1007/s00530-021-00787-5](https://doi.org/10.1007/s00530-021-00787-5).
- [117] Mohanad MOHAMMED et al. « A stacking ensemble deep learning approach to cancer type classification based on TCGA data ». en. In : *Sci Rep* 11.1 (déc. 2021), p. 15626. ISSN : 2045-2322. DOI : [10.1038/s41598-021-95128-x](https://doi.org/10.1038/s41598-021-95128-x).
- [118] Pieter Van MOLLE et al. « Visualizing convolutional neural networks to improve decision support for skin lesion classification ». In : *Understanding and interpreting machine learning in medical image computing applications*. Springer, 2018, p. 115-123. DOI : [10.1007/978-3-030-02628-8_13](https://doi.org/10.1007/978-3-030-02628-8_13).
- [119] J MONNIER et al. « Impact d'un artefact fréquent sur la détection automatique du mélanome à partir d'images dermoscopiques : approche deep learning combinée à l'algorithme Support Vector Machine ». In : *Annales de Dermatologie et de Vénérologie*. T. 147. 12. Elsevier. 2020, A82. DOI : [10.1016/j.annder.2020.09.022](https://doi.org/10.1016/j.annder.2020.09.022).
- [120] Dominik MÜLLER, Iñaki SOTO-REY et Frank KRAMER. *An Analysis on Ensemble Learning optimized Medical Image Classification with Deep Convolutional Neural Networks*. arXiv :2201.11440 [cs]. Avr. 2022. DOI : [10.48550/arXiv.2201.11440](https://doi.org/10.48550/arXiv.2201.11440).

- [121] Samuel G MÜLLER et Frank HUTTER. « Trivialaugment : Tuning-free yet state-of-the-art data augmentation ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 774-782. DOI : [10.48550/arXiv.2103.10158](https://doi.org/10.48550/arXiv.2103.10158).
- [122] NATSTORY-COSMETIQUES. *LE COLLAGÈNE*. URL : <https://natstory-cosmetiques.fr/le-collagene/> (visité le 11/07/2022).
- [123] Cristian NAVARRETE-DECHENT et al. « Automated dermatological diagnosis : hype or reality? » In : *The Journal of investigative dermatology* 138.10 (2018), p. 2277. DOI : [10.1016/j.jid.2018.04.040](https://doi.org/10.1016/j.jid.2018.04.040).
- [124] *Panorama des cancers en France - Ref : PANOKFR2022*. URL : https://www.e-cancer.fr/content/download/336357/4805630/file/01_brochure_Panorama-2022-V2-@.pdf (visité le 11/07/2022).
- [125] Seulki PARK et al. « Influence-balanced loss for imbalanced visual classification ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 735-744. DOI : [10.48550/arXiv.2110.02444](https://doi.org/10.48550/arXiv.2110.02444).
- [126] Meredith S. PELSTER et Rodabe N. AMARIA. « Neoadjuvant Immunotherapy for Locally Advanced Melanoma ». en. In : *Curr. Treat. Options in Oncol.* 21.2 (fév. 2020), p. 10. ISSN : 1534-6277. DOI : [10.1007/s11864-020-0700-z](https://doi.org/10.1007/s11864-020-0700-z).
- [127] Fábio PEREZ et al. « Data augmentation for skin lesion analysis ». In : *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, p. 303-311. DOI : [10.1007/978-3-030-01201-4_33](https://doi.org/10.1007/978-3-030-01201-4_33).
- [128] Samuel William PEWTON et Moi Hoon YAP. « Dark Corner on Skin Lesion Image Dataset : Does It Matter? » In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, p. 4831-4839.
- [129] Tri-Cong PHAM et al. « AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function ». In : *Scientific Reports* 11.1 (2021), p. 1-13. DOI : [10.1038/s41598-021-96707-8](https://doi.org/10.1038/s41598-021-96707-8).
- [130] John C PLATT et al. « Large margin DAGs for multiclass classification ». In : *Advances in neural information processing systems*. 2000, p. 547-553.
- [131] Dan POPESCU et al. « New Trends in Melanoma Detection Using Neural Networks : A Systematic Review ». In : *Sensors* 22.2 (2022), p. 496. DOI : [10.3390/s22020496](https://doi.org/10.3390/s22020496).
- [132] Tryan Aditya PUTRA, Syahidah Izza RUFAlDA et Jenq-Shiou LEU. « Enhanced skin condition prediction through machine learning using dynamic training and testing augmentation ». In : *IEEE Access* 8 (2020), p. 40536-40546. DOI : [10.1109/ACCESS.2020.2976045](https://doi.org/10.1109/ACCESS.2020.2976045).

- [133] Andreas RAUBER, Roberto TRASARTI et Fosca GIANNOTTI. « Transparency in algorithmic decision making ». In : *ERCIM News* 116 (2019), p. 10-11. URL : https://www.researchgate.net/profile/Michele-Loi-2/publication/338622244_How_to_include_ethics_in_machine_learning_research/links/5eac67d345851592d6b1f7d8/How-to-include-ethics-in-machine-learning-research.pdf#page=10 (visité le 02/08/2022).
- [134] Andreas RAUBER, Roberto TRASARTI et Fosca GIANNOTTI. « Transparency in algorithmic decision making ». In : *ERCIM News* 116 (2019), p. 10-11.
- [135] *Rayonnements ultraviolets et risques de cancer - Ref : FRUV21*. URL : <https://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Rayonnements-ultraviolets-et-risques-de-cancer> (visité le 11/07/2022).
- [136] Rehan RAZA et al. « Melanoma classification from dermoscopy images using ensemble of convolutional neural networks ». In : *Mathematics* 10.1 (2021), p. 26. DOI : [10.3390/math10010026](https://doi.org/10.3390/math10010026).
- [137] Navid RAZMJOOY et al. « Computer-aided Diagnosis of Skin Cancer : A Review ». In : *Current Medical Imaging Reviews* 16.7 (août 2020), p. 781-793. DOI : [10.2174/1573405616666200129095242](https://doi.org/10.2174/1573405616666200129095242).
- [138] Jiawei REN et al. « Balanced meta-softmax for long-tailed visual recognition ». In : *Advances in neural information processing systems* 33 (2020), p. 4175-4186. DOI : [10.48550/arXiv.2007.10740](https://doi.org/10.48550/arXiv.2007.10740).
- [139] Adria ROMERO LOPEZ et al. « Skin lesion classification from dermoscopic images using deep learning techniques ». In : *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*. Fév. 2017, p. 49-54. DOI : [10.2316/P.2017.852-053](https://doi.org/10.2316/P.2017.852-053).
- [140] David E. RUMELHART, Geoffrey E. HINTON et Ronald J. WILLIAMS. « Learning representations by back-propagating errors ». en. In : *Nature* 323.6088 (oct. 1986). Number : 6088 Publisher : Nature Publishing Group, p. 533-536. ISSN : 1476-4687. DOI : [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [141] Mahya SADEGHI, Parmit K CHILANA et M Stella ATKINS. « How users perceive content-based image retrieval for identifying skin images ». In : *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer, 2018, p. 141-148. DOI : [10.1007/978-3-030-02628-8_16](https://doi.org/10.1007/978-3-030-02628-8_16).
- [142] Kotra Sankar Raja SEKHAR et al. « Dermoscopic image classification using CNN with Handcrafted features ». In : *Journal of king Saud University-science* 33.6 (2021), p. 101550. DOI : [10.1016/j.jksus.2021.101550](https://doi.org/10.1016/j.jksus.2021.101550).
- [143] Ramprasaath R SELVARAJU et al. « Grad-cam : Visual explanations from deep networks via gradient-based localization ». In : *Proceedings of the IEEE international conference on computer vision*. 2017, p. 618-626. DOI : [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74).

- [144] Pourya SHAMSOLMOALI et al. « Imbalanced data learning by minority class augmentation using capsule adversarial networks ». en. In : *Neurocomputing* 459 (oct. 2021), p. 481-493. ISSN : 0925-2312. DOI : [10.1016/j.neucom.2020.01.119](https://doi.org/10.1016/j.neucom.2020.01.119).
- [145] Joshua M. SHARFSTEIN. « A Spotlight on Sunscreen Regulation ». en. In : *N Engl J Med* 373.2 (juill. 2015), p. 101-103. ISSN : 0028-4793, 1533-4406. DOI : [10.1056/NEJMp1504912](https://doi.org/10.1056/NEJMp1504912).
- [146] Katharina SIES et al. « Dark corner artefact and diagnostic performance of a market-approved neural network for skin cancer classification ». In : *JDDG : Journal der Deutschen Dermatologischen Gesellschaft* 19.6 (2021), p. 842-850. DOI : [10.1111/ddg.14384](https://doi.org/10.1111/ddg.14384).
- [147] Karen SIMONYAN et Andrew ZISSERMAN. « Very Deep Convolutional Networks for Large-Scale Image Recognition ». In : (2015). Sous la dir. d'Yoshua BENGIO et Yann LECUN. DOI : [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- [148] Amitojdeep SINGH, Sourya SENGUPTA et Vasudevan LAKSHMINARAYANAN. « Explainable deep learning models in medical image analysis ». In : *Journal of Imaging* 6.6 (2020), p. 52. DOI : [10.3390/jimaging6060052](https://doi.org/10.3390/jimaging6060052).
- [149] Saptarshi SINHA, Hiroki OHASHI et Katsuyuki NAKAMURA. « Class-wise difficulty-balanced loss for solving class-imbalance ». In : *Proceedings of the Asian conference on computer vision*. 2020. DOI : [10.48550/arXiv.2010.01824](https://doi.org/10.48550/arXiv.2010.01824).
- [150] Leslie N SMITH. « Cyclical learning rates for training neural networks ». In : *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2017, p. 464-472. DOI : [10.1109/WACV.2017.58](https://doi.org/10.1109/WACV.2017.58).
- [151] Alina SULTANA et al. « Removal of artifacts from dermatoscopic images ». In : *2014 10th International Conference on Communications (COMM)*. IEEE. 2014, p. 1-4. DOI : [10.1109/ICComm.2014.6866757](https://doi.org/10.1109/ICComm.2014.6866757).
- [152] Fumitake TAKAHASHI et Shigeo ABE. « Optimizing directed acyclic graph support vector machines ». In : *Artificial Neural Networks in Pattern Recognition (ANNPR)* (2003), p. 166-173.
- [153] Jingru TAN et al. « Equalization loss for long-tailed object recognition ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, p. 11662-11671. DOI : [10.48550/arXiv.2003.05176](https://doi.org/10.48550/arXiv.2003.05176).
- [154] Jingru TAN et al. « Equalization loss v2 : A new gradient balance approach for long-tailed object detection ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, p. 1685-1694. DOI : [10.48550/arXiv.2012.08548](https://doi.org/10.48550/arXiv.2012.08548).
- [155] Mingxing TAN et Quoc LE. « Efficientnet : Rethinking model scaling for convolutional neural networks ». In : *International Conference on Machine Learning*. PMLR, 2019, p. 6105-6114. DOI : [10.48550/arXiv.1905.11946](https://doi.org/10.48550/arXiv.1905.11946).

- [156] P. TSCHANDL et al. « Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features ». en. In : *British Journal of Dermatology* 181.1 (2019), p. 155-165. ISSN : 1365-2133. DOI : <https://doi.org/10.1111/bjd.17189>. (Visité le 05/01/2021).
- [157] Philipp TSCHANDL, Cliff ROSENDAHL et Harald KITTLER. « The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions ». en. In : *Sci Data* 5.1 (août 2018), p. 180161. ISSN : 2052-4463. DOI : [10.1038/sdata.2018.161](https://doi.org/10.1038/sdata.2018.161).
- [158] Philipp TSCHANDL et al. « Human–computer collaboration for skin cancer recognition ». In : *Nature Medicine* 26.8 (2020), p. 1229-1234. DOI : [10.1038/s41591-020-0942-0](https://doi.org/10.1038/s41591-020-0942-0).
- [159] Grant VAN HORN et al. « The inaturalist species classification and detection dataset ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 8769-8778. DOI : [10.48550/arXiv.1707.06642](https://doi.org/10.48550/arXiv.1707.06642).
- [160] Chaozheng WANG et al. « Label-Aware Distribution Calibration for Long-tailed Classification ». In : *arxiv preprint arXiv :2111.04901 [cs]* (nov. 2021). DOI : [10.48550/arXiv.2111.04901](https://doi.org/10.48550/arXiv.2111.04901).
- [161] Sutong WANG et al. « Interpretability-based multimodal convolutional neural networks for skin lesion diagnosis ». In : *IEEE Transactions on Cybernetics* (2021). DOI : [10.1109/TCYB.2021.3069920](https://doi.org/10.1109/TCYB.2021.3069920).
- [162] Yanal WAZAEFI. « Automatic diagnosis of melanoma from dermoscopic images of melanocytic tumors : Analytical and comparative approaches ». These de doctorat. Aix-Marseille, déc. 2013. URL : <https://www.theses.fr/2013AIXM4106> (visité le 02/08/2022).
- [163] J. van de WEIJER, T. GEVERS et A. GIJSENIJ. « Edge-Based Color Constancy ». In : *IEEE Transactions on Image Processing* 16.9 (sept. 2007). Conference Name : IEEE Transactions on Image Processing, p. 2207-2214. ISSN : 1941-0042. DOI : [10.1109/TIP.2007.901808](https://doi.org/10.1109/TIP.2007.901808).
- [164] Julia K WINKLER et al. « Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition ». In : *JAMA dermatology* 155.10 (2019), p. 1135-1141. DOI : [10.1001/jamadermatol.2019.1735](https://doi.org/10.1001/jamadermatol.2019.1735).
- [165] David H. WOLPERT. « Stacked generalization ». en. In : *Neural Networks* 5.2 (jan. 1992), p. 241-259. ISSN : 0893-6080. DOI : [10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- [166] Lu YANG et al. « A Survey on Long-Tailed Visual Recognition ». en. In : *Int J Comput Vis* 130.7 (juill. 2022), p. 1837-1872. ISSN : 1573-1405. DOI : [10.1007/s11263-022-01622-8](https://doi.org/10.1007/s11263-022-01622-8).
- [167] Yun YANG et al. « Two-Stage Selective Ensemble of CNN via Deep Tree Training for Medical Image Classification ». In : *IEEE Trans. Cybern.* (2021), p. 1-14. ISSN : 2168-2267, 2168-2275. DOI : [10.1109/TCYB.2021.3061147](https://doi.org/10.1109/TCYB.2021.3061147).

- [168] Peng YAO et al. « Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification ». In : *IEEE Transactions on Medical Imaging* 41.5 (mai 2022). Conference Name : IEEE Transactions on Medical Imaging, p. 1242-1254. ISSN : 1558-254X. DOI : [10.1109/TMI.2021.3136682](https://doi.org/10.1109/TMI.2021.3136682).
- [169] Jordan YAP, William YOLLAND et Philipp TSCHANDL. « Multimodal skin lesion classification using deep learning ». en. In : *Experimental Dermatology* 27.11 (2018), p. 1261-1267. ISSN : 1600-0625. DOI : [10.1111/exd.13777](https://doi.org/10.1111/exd.13777).
- [170] Ali YAZDIZADEH, Zachary PATTERSON et Bilal FAROOQ. « Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey ». In : *IEEE Transactions on Intelligent Transportation Systems* 21.6 (juin 2020). Conference Name : IEEE Transactions on Intelligent Transportation Systems, p. 2232-2239. ISSN : 1558-0016. DOI : [10.1109/TITS.2019.2918923](https://doi.org/10.1109/TITS.2019.2918923).
- [171] Sule YILDIRIM-YAYILGAN et al. « Pre-trained CNN based deep features with hand-crafted features and patient data for skin lesion classification ». In : *International Conference on Intelligent Technologies and Applications*. Springer, 2021, p. 151-162. DOI : [10.1007/978-3-030-71711-7_13](https://doi.org/10.1007/978-3-030-71711-7_13).
- [172] Sangdoon YUN et al. « Cutmix : Regularization strategy to train strong classifiers with localizable features ». In : *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, p. 6023-6032. DOI : [10.48550/arXiv.1905.04899](https://doi.org/10.48550/arXiv.1905.04899).
- [173] Bin ZHANG et al. « Opportunities and Challenges : Classification of Skin Disease Based on Deep Learning ». In : *Chinese Journal of Mechanical Engineering* 34.1 (nov. 2021), p. 112. ISSN : 2192-8258. DOI : [10.1186/s10033-021-00629-5](https://doi.org/10.1186/s10033-021-00629-5).
- [174] Cha ZHANG et Yunqian MA, éd. *Ensemble Machine Learning*. en. Springer New York, NY. ISBN : 978-1-4419-9326-7. DOI : [10.1007/978-1-4419-9326-7](https://doi.org/10.1007/978-1-4419-9326-7). URL : <https://link.springer.com/book/10.1007/978-1-4419-9326-7>.
- [175] Hongyi ZHANG et al. « mixup : Beyond Empirical Risk Minimization ». In : *International Conference on Learning Representations*. 2018. DOI : [10.48550/arXiv.1710.09412](https://doi.org/10.48550/arXiv.1710.09412).
- [176] Jianpeng ZHANG et al. « Attention residual learning for skin lesion classification ». In : *IEEE transactions on medical imaging* 38.9 (2019). Publisher : IEEE, p. 2092-2103. DOI : [10.1109/TMI.2019.2893944](https://doi.org/10.1109/TMI.2019.2893944).
- [177] Songyang ZHANG et al. « Distribution alignment : A unified framework for long-tail visual recognition ». In : *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, p. 2361-2370. DOI : [10.48550/arXiv.2103.16370](https://doi.org/10.48550/arXiv.2103.16370).
- [178] Yifan ZHANG et al. *Deep Long-Tailed Learning : A Survey*. arXiv :2110.04596 [cs]. Oct. 2021. DOI : [10.48550/arXiv.2110.04596](https://doi.org/10.48550/arXiv.2110.04596).

- [179] Yongshun ZHANG et al. « Bag of tricks for long-tailed visual recognition with deep convolutional neural networks ». In : *Proceedings of the AAAI conference on artificial intelligence*. T. 35. 4. 2021, p. 3447-3455. URL : <https://ojs.aaai.org/index.php/AAAI/article/view/16458>.
- [180] Zhiwei ZHANG. « You Only Need End-to-End Training for Long-Tailed Recognition ». In : *arXiv preprint arXiv :2112.05958 [cs]* (déc. 2021). DOI : [10.48550/arXiv.2112.05958](https://doi.org/10.48550/arXiv.2112.05958).
- [181] Chen ZHAO et al. « Dermoscopy image classification based on StyleGAN and DenseNet201 ». In : *Ieee Access* 9 (2021), p. 8659-8679. DOI : [10.1109/ACCESS.2021.3049600](https://doi.org/10.1109/ACCESS.2021.3049600).
- [182] Bolei ZHOU et al. « Learning deep features for discriminative localization ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 2921-2929. DOI : [10.48550/arXiv.1512.04150](https://doi.org/10.48550/arXiv.1512.04150).
- [183] Zhi-Hua ZHOU et Xu-Ying LIU. « Training cost-sensitive neural networks with methods addressing the class imbalance problem ». In : *IEEE Transactions on knowledge and data engineering* 18.1 (2005), p. 63-77. DOI : [10.1109/TKDE.2006.17](https://doi.org/10.1109/TKDE.2006.17).

ANNEXES

A. Rappel théorique sur les CNNs

Dans la [section](#) précédente, nous avons fait un état de l’art des méthodes de conception de [SAADs](#) pour la détection du mélanome basées sur des techniques d’apprentissage profond. Les réseaux de neurones couramment utilisés pour cette tâche sont les [CNNs](#). Nous avons utilisé ces derniers pour développer tous les modèles présentés dans cette thèse.

Dans cette section, nous allons faire un tour d’horizon de l’intuition derrière les [CNNs](#). Nous allons notamment expliquer ce que c’est que l’apprentissage profond, mais aussi décrire le fonctionnement d’un [CNN](#).

A.1. L’apprentissage profond

L’origine de l’apprentissage profond date du début du XX^{ème} siècle notamment avec les travaux de Geoffrey Hinton [82, 140] considéré comme le père fondateur de l’apprentissage profond. L’idée derrière l’apprentissage profond est d’observer et répliquer la manière dont fonctionne le cerveau. Le cerveau étant constitué de milliards de neurones connectés entre eux, l’unité de base du calcul en apprentissage profond est donc naturellement le neurone artificiel.

Dans cette sous-section, nous allons présenter la théorie de base qui sous-tend l’apprentissage profond. Nous allons premièrement décrire l’analogie qui existe entre le neurone biologique et le neurone artificiel. Ensuite, nous allons présenter le processus d’apprentissage d’un réseau de neurones artificiels.

A.1.1. Principe de base

L’analogie entre un neurone biologique et un neurone artificiel peut se faire de la manière suivante :

- Les entrées d’un neurone correspondent aux dendrites ;
- La sortie correspond à l’axone ;
- Le noyau du corps cellulaire correspond à la fonction d’activation.

La Figure .1 illustre l’analogie qui existe entre un neurone biologique et un neurone artificiel.

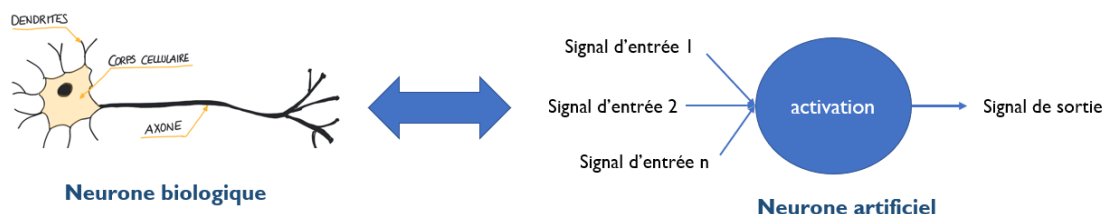


FIGURE .1. – Illustration de l’analogie entre un neurone biologique et un neurone artificiel.

Un neurone artificiel reçoit, de la part d'autres neurones auxquels il est connecté, des valeurs d'entrée (ou variables) ayant des valeurs numériques x_1, x_2, \dots, x_n et calcule une valeur de sortie y . Chaque entrée x_i a un poids associé w_i , qui est attribué en fonction de son importance relative par rapport aux autres entrées. Un biais b est ajouté à la somme pondérée des entrées x_i afin d'ajuster la sortie du neurone. Une fonction d'activation f est finalement appliquée sur cette somme. Le calcul de la sortie y du neurone s'effectue selon la formule suivante :

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (.1)$$

La Figure .2 illustre le fonctionnement d'un neurone artificiel.

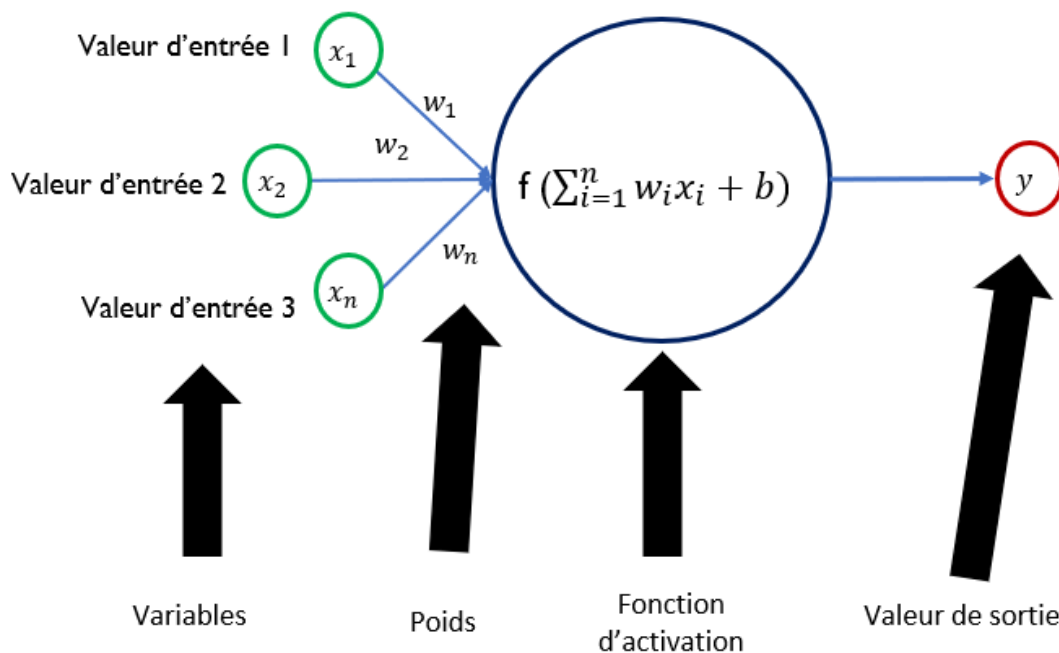


FIGURE .2. – Illustration du fonctionnement d'un neurone artificiel.

Un réseau de neurones artificiels se construit finalement en connectant différents neurones artificiels entre eux similairement aux réseaux de neurones du cerveau.

A.1.2. Apprentissage d'un réseau de neurones

L'apprentissage automatique consiste à entraîner des algorithmes sur des données connues afin qu'ils puissent faire des prédictions sur des données inconnues. Ils existent pour cela deux catégories d'apprentissage : l'apprentissage supervisé et l'apprentissage non supervisé.

L'apprentissage supervisé consiste à entraîner l'algorithme sur un jeu de données où à chaque échantillon est associée une étiquette ou classe ou catégorie. L'algorithme en-

trainé suivant cette approche se base donc sur la correction des erreurs de prédiction entre la classe prédite et l'étiquette d'origine pour mettre à jour ses poids. A contrario, l'apprentissage non supervisé consiste à entraîner l'algorithme sur des données non étiquetées. Le but de l'apprentissage est donc pour l'algorithme d'identifier dans les données des caractéristiques permettant de mettre en évidence des groupes ou des classes.

Dans le cadre de nos travaux de thèse, nous employons exclusivement pour l'entraînement de nos modèles des techniques d'apprentissage supervisées. Soit un ensemble d'apprentissage noté S , contenant n observations et défini par :

$$S = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (.2)$$

L'apprentissage supervisé d'un réseau de neurones artificiels consiste à présenter au réseau l'ensemble d'apprentissage S afin qu'il apprenne les paramètres W et b qui généralisent le mieux l'ensemble. Ce processus d'apprentissage se décline en trois étapes chronologiques principales : la propagation des informations, le calcul du taux d'erreur et la rétropropagation de l'erreur.

- **La propagation des informations** : Durant cette première phase, chacune des variables d'entrée d'entraînement x_i est passée individuellement ou par lot à travers le réseau de telle sorte que tous les neurones appliquent leur transformation $f(x)$ aux informations qu'ils reçoivent des neurones de la couche précédente et les envoient aux neurones de la couche suivante. Lorsque les données ont traversé toutes les couches, et que tous les neurones ont effectué leurs calculs, la couche finale sera atteinte et générera un vecteur de prédictions \mathbf{P} pour ces exemples d'entrée.
- **Calcul du taux d'erreur** : La deuxième étape consiste à estimer l'erreur (ou perte) du réseau, ceci à l'aide d'une fonction différentiable de coût L (ou fonction objective, ou fonction de perte). L'objectif de la fonction L est de calculer l'erreur entre les sorties attendues \mathbf{Y} des données de l'ensemble d'apprentissage S et les données prédites par le réseau de neurones \mathbf{P} . Idéalement, nous voulons que notre coût soit nul, c'est-à-dire sans divergence entre la valeur estimée et la valeur attendue. Par conséquent, au fur et à mesure de l'apprentissage du modèle, les poids des interconnexions des neurones seront progressivement ajustés jusqu'à l'obtention de bonnes prédictions ceci grâce à la rétropropagation de l'erreur.
- **Rétropropagation de l'erreur** : Une fois le coût calculé, cette information est propagée en arrière d'où son nom de rétropropagation. En partant de la couche de sortie, cette information est récursivement appliquée tout le long du réseau jusqu'à la couche d'entrée pour calculer les dérivées partielles sur tous les poids. Les dérivées partielles sont utilisées par un algorithme d'optimisation de descente de gradients [2] pour mettre à jour les poids du réseau. Ceci se fait globalement

à l'aide de l'équation suivante :

$$\mathbf{w}^t = \mathbf{w}^{t-1} - \eta \nabla L(\mathbf{w}^{t-1}) \quad (.3)$$

où η représente le taux d'apprentissage contrôlant l'amplitude de la mise à jour des poids, et ∇L représente le gradient de la fonction de coût L par rapport à \mathbf{w}^{t-1} . Il faut noter que d'autres algorithmes d'optimisation existent dont les plus utilisés sont la Descente de gradient stochastique –ou *Stochastic gradient descent*– (SGD) [22] et Adam [93].

La Figure .3 illustre les phases d'apprentissage d'un réseau de neurones artificiels.

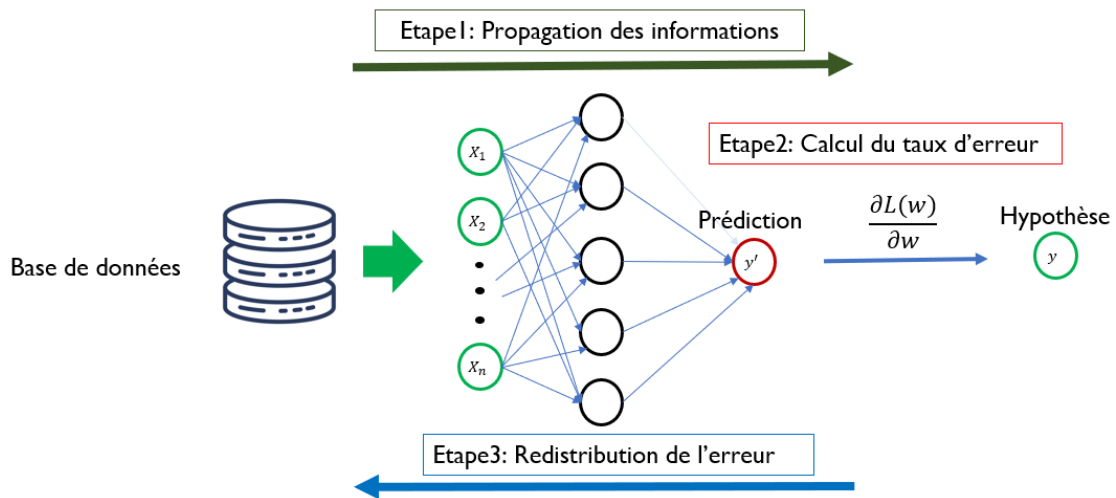


FIGURE .3. – Illustration de l'entraînement d'un réseau de neurones artificiels.

A.2. Réseaux de neurones convolutifs

Les CNNs représentent un type spécifique de réseaux de neurones artificiels. Ils sont particulièrement utilisés pour le traitement d'image, de vidéo et la reconnaissance/détection d'objet mais ils peuvent aussi s'appliquer à d'autres types de données. Un CNN est constitué de plusieurs couches empilées successivement. Les principales couches étant : la couche de convolution, la couche d'échantillonnage et la couche FC. Entre chaque couche d'un CNN, des fonctions d'activations sont appliquées afin d'introduire de la non-linéarité.

Dans cette sous-section, nous allons décrire le fonctionnement des principales couches d'un CNN mais aussi présenter les fonctions d'activation couramment utilisées en apprentissage profond ainsi que leurs propriétés.

A.2.1. Les différentes couches d'un CNN

Un CNN est un algorithme d'apprentissage profond qui peut prendre une image¹ en entrée, attribuer de l'importance à divers aspects de l'image et être capable de les différencier les uns des autres.

- **Couches de convolution** : la couche de convolution est la composante clé d'un CNN. Elle fonctionne comme un extracteur de caractéristiques. Pour cela, elle applique des opérations de convolution sur sa variable d'entrée. Théoriquement parlant, la convolution est un terme mathématique utilisé pour décrire une procédure permettant de combiner deux fonctions afin de produire une troisième. Dans le cas d'un CNN, il s'agit de la combinaison entre une image et un filtre (ou matrice ou *kernel*) afin de produire une carte de caractéristiques. Ces filtres sont en réalité des **tenseurs** de poids qui se mettent à jour via l'algorithme de rétropropagation.
- **Couche d'échantillonnage** (ou *pooling*) : La couche d'échantillonnage est généralement insérée entre deux couches de convolution. Elle a pour objectif de réduire la taille spatiale des cartes de caractéristiques reçues en entrée, tout en conservant les caractéristiques essentielles. Plusieurs techniques d'échantillonnage existent à l'instar de l'échantillonnage maximum -ou *max pooling*- et l'échantillonnage moyen -ou *average pooling*-. La couche d'échantillonnage génère à sa sortie le même nombre de cartes de caractéristiques qu'en entrée mais considérablement compressées.
- **Couche FC** : la notion de " complètement connecté " définit que chaque neurone de la couche précédente est connecté à chaque neurone de la couche suivante. L'objectif de cette couche est d'utiliser les caractéristiques de haut niveau issues des couches précédemment présentées pour classer les données d'entrée. Les

1. Il existe des réseaux de neurones convolutifs pour des variables d'entrée de dimension 1 ou 3, mais dans le cadre de notre thèse, seuls les réseaux de neurones utilisés pour des variables d'entrée de dimension 2 sont utilisés.

couches **FC** sont généralement suivies d'une couche de **Dropout** afin de régulariser les sorties des neurones et de réduire le sur-apprentissage. Une fonction Softmax ou Sigmoid est en pratique appliquée à la dernière couche **FC** du réseau afin de générer un vecteur de probabilités des prédictions.

A.2.2. Les fonctions d'activation

Les fonctions d'activation f permettent de transmettre ou non une valeur de sortie selon le signal d'entrée reçu. En apprentissage profond, il s'agit généralement de fonctions non-linéaires du fait que la plupart des problèmes complexes ont un comportement non linéaire. Plusieurs fonctions d'activation sont présentées dans la littérature. Les plus utilisées en pratique sont : la fonction Sigmoide, la fonction Tangente hyperbolique (TanH), la fonction **ReLU** et la fonction Softmax.

- La fonction sigmoïde : prend une entrée réelle et la réduit dans l'intervalle $[0, 1]$. Historiquement, elle a été fréquemment utilisée, mais elle présente le principal inconvénient de saturer à 0 ou 1 (le gradient est faible et proche de zéro). Ceci n'est pas souhaitable car pendant la rétropropagation, cela pourrait faire disparaître le gradient, rendant l'apprentissage difficile. Ce phénomène est connu sous le nom de problème de saturation du gradient (ou vanishing gradient). Dans ce contexte, l'initialisation des poids joue un rôle fondamental. Par exemple, des poids trop importants pourraient conduire la plupart des neurones à la saturation.
- La fonction TanH : prend une entrée de valeur réelle et la réduit à l'intervalle $[-1, 1]$. Comme la fonction Sigmoide, elle souffre également du problème de saturation du gradient.
- La fonction **ReLU** : prend une entrée de valeur réelle et la réduit à l'intervalle $[0, +\infty[$. Cette fonction d'activation est la plus utilisée en apprentissage profond en raison de ses propriétés : il s'agit d'une fonction non linéaire, différentiable, et ne provoquant pas de saturation de gradients.
- La fonction Softmax : elle prend en entrée un vecteur réel et le transforme en un vecteur de probabilités. Elle est généralement utilisée dans la couche finale d'un modèle de classification ayant plusieurs sorties (Classification multi-classes).

Le Tableau .1 récapitule les fonctions d'activation précédemment présentées avec leurs formulations.

TABLEAU .1. – Récapitulatif des fonctions d’activation couramment utilisées en apprentissage profond.

Fonction d’activation	Formulation	Intervalle de sortie
Sigmoïde	$\frac{1}{1+e^{-z}}$	$[0, 1]$
TanH	$\frac{1-e^{-2z}}{1+e^{-2z}}$	$[-1, 1]$
ReLU	$\max(0, z)$	$[0, +\infty[$
Softmax	$\frac{e^{-z_i}}{\sum e^{-z_i}}$	$[0, 1]$

z_i représente les valeurs obtenues à la sortie des neurones.