

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A.MIRA-BEJAIA



Faculté de Technologie
Département de Génie Electrique
Laboratoire de Technologies Industrielles et de l'Information (LTII)

THÈSE

EN VUE DE L'OBTENTION DU DIPLOME DE DOCTORAT

Domaine : Sciences et Technologie Filière : Electronique
Spécialité : Electronique

Présentée par
OUKIL Samia

Thème

Détection de mélanomes-précoces par apprentissage automatique

Soutenue le : 05 mars 2022

Devant le Jury composé de :

Nom et Prénom

Grade

Mr BERRAH Smail	Professeur	Univ. de BEJAIA	Président
Mr KASMI Reda	MCA	Univ. de BOUIRA	Rapporteur
Mr MOKRANI Karim	Professeur	Univ. de BEJAIA	Co-Rapporteur
Mr BELAID Ahror	Professeur	Univ. de BEJAIA	Examineur
Mr BENSALIA Yassine	MCA	Univ. de BOUIRA	Examineur
Mr MEKHMOUKH Abdenour	MCA	Univ. de BEJAIA	Examineur

Année Universitaire : 2021/2022

Résumé

Le mélanome est connu comme une forme très agressive de cancer de la peau et l'une des tumeurs malignes dont la croissance est la plus rapide dans de nombreux pays. Le diagnostic assisté par ordinateur (DAO) est une technique populaire utilisée pour la détection du mélanome dans la littérature. Cependant, elle se heurte encore à de nombreuses complications, notamment lorsqu'il s'agit de distinguer les lésions malignes des lésions bénignes à un stade précoce. Cette thèse vise à discriminer le mélanome des lésions cutanées bénignes dans les images dermoscopiques afin d'améliorer le potentiel, l'efficacité et la précision d'un système d'aide à la décision médicale qui peut aider les dermatologues à diagnostiquer les lésions cutanées à un stade précoce. L'algorithme proposé utilise une segmentation automatique basée sur le k-means générant un masque assez précis pour chaque lésion. L'extraction de caractéristiques consiste en des attributs de couleur et de texture existants et nouveaux, mesurant comment la couleur et la texture varient à l'intérieur de la lésion. Afin de trouver les résultats optimaux, tous les attributs sont extraits des lésions sur cinq systèmes de couleurs différents (RGB, HSV, Lab, XYZ et YCbCr), et utilisés comme entrées pour trois classificateurs (KNN, SVM et ANN). L'ensemble d'images dermoscopiques PH2 de l'hôpital Pedro Hispano est utilisé pour évaluer la performance de l'algorithme proposé. Les résultats de notre algorithme sont comparés aux résultats d'articles publiés qui ont utilisé le même ensemble de données et il en ressort que la méthode proposée surpasse les résultats des travaux antérieurs en atteignant la sensibilité de 99,25%, la spécificité de 99,58% et la précision de 99,51%. Les résultats finaux montrent que les couleurs combinées à la texture sont des attributs puissants et pertinents pour la détection des mélanomes et montrent une amélioration par rapport à ce qui est montré dans la littérature en utilisant la même base de données.

Mots clés: Mélanome; segmentation; dermoscopie; super-pixel; K-means; Attributs; Classification.

Abstract

Melanoma is known as the most aggressive form of skin cancer and one of the fastest-growing malignant tumors worldwide. Several Computers Aid Diagnosis of melanoma are proposed, still, the algorithms encounter difficulties in the early stage of lesions. The paper aims to discriminate melanoma from a benign skin lesion in dermoscopic images. The proposed algorithm is based on the color and texture of skin lesions by introducing a novel technique of features extraction. The algorithm uses an automatic segmentation based on k-means generating a fairly accurate mask for each lesion. The feature extraction consists of the existing and novel color and texture attributes measuring how color and texture vary inside the lesion. In purpose to find the optimal results all the attributes are extracted from a lesion on five different system colors (RGB, HSV, Lab, XYZ, and YCbCr), and used as the inputs for three classifiers (KNN, SVM, and ANN). The PH2 set is used to assess the performance of the proposed algorithm. The results of our algorithm are compared to the results of published articles that used the same dataset and it shows that the proposed method outperforms the works of the state-of-art results by reaching the sensitivity of 99.25%, specificity of 99.58%, and accuracy of 99.51%. The final results show that the colors combined with texture are powerful and pertinent attributes for melanoma detection and show improvement over what is shown in the state-of-art using the same dataset.

Key words: Melanoma; segmentation; dermoscopy; super-pixel; K-means; features extraction; classification.

Remerciements

Ce manuscrit présente mon travail de thèse effectuée au sein du laboratoire LTII de l'université A.MIRA, Bejaia. Ce fut une expérience, des plus riches, durant laquelle j'ai eu à côtoyer plusieurs personnes qui méritent pleinement ma gratitude.

Avant tout, je remercie Dieu le tout-puissant de m'avoir accordé patience et courage pour la réalisation de cette thèse, résultat d'un travail de plusieurs années de recherches, qui n'aurait pu connaître le jour sans l'aide précieuse de mon directeur de thèse Monsieur KASMI Reda, à qui j'adresse mes vifs remerciements pour ses recommandations, ses conseils et son encadrement. Je tiens à lui exprimer ma haute reconnaissance pour sa confiance dans l'orientation de mes travaux, son soutien et sa disponibilité constante pendant toute la durée de ce travail.

À mon co-directeur de thèse, Professeur MOKRANI Karim, j'adresse mes sincères remerciements pour ses conseils tout au long de mon parcours.

Je souhaite, ensuite, remercier Professeure Maria Begonya pour son accueil au sein du Laboratoire eVIDA de l'Université Deusto, pour sa disponibilité, son écoute et ses conseils. Ainsi que tous les membres du laboratoire pour leur bon accueil et leur aide.

Mes remerciements particuliers vont à Monsieur ASRADJ Zahir pour ses encouragements et ses conseils durant ces longues années turbulentes.

Mes remerciements s'adressent également à Monsieur BERRAH Smail pour l'honneur qu'il m'accorde en présidant mon jury de thèse.

J'aimerais ensuite exprimer toute ma gratitude aux Messieurs BELAID Ahror, BENSAFIA Yassine et MEKHMOUKH Abdenour pour avoir accepté d'assister à mon jury en tant que examinateurs de ce travail.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Dédicaces

Je ne peux exprimer à travers ces lignes tous mes sentiments et ma profonde gratitude envers vous mes chers parents pour les sacrifices que vous avez consentis pour mon instruction et mon bien-être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance.

Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de vos innombrables sacrifices, bien que je ne vous en acquitte jamais assez. Que Dieu tout-puissant vous garde et vous procure santé, bonheur et longue vie pour que vous demeuriez flambeau illuminant mon chemin.

À mon cher frère Aissa et mes chères et adorables sœurs Fadila et Dyhia qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Puisse l'amour et la fraternité nous unissent à jamais. Je vous souhaite la réussite dans votre vie, avec tout le bonheur qu'il faut pour vous combler.

À mes chers petits-neveux et nièces Abd Raouf, Racha, Yani, Maya, Amir, Zinedine et Sofia. Aucune dédicace ne saurait exprimer tout l'amour que j'ai pour vous, votre joie et votre gaieté me combler de bonheur. Puisse Dieu vous garder, éclairer votre route et vous aider à réaliser à votre tour vos vœux les plus chers. Je vous aime énormément mes champions.

Merci du fond du cœur pour mon cher ami Nouh qui a toujours su m'écouter, m'encourager, m'aider et me soutenir durant la réalisation de cette thèse.

Merci à mes chères copines Aicha et Fatima que j'ai trouvées disponibles et compréhensives tout au long de ce parcours.

A tous ceux et celles qui m'ont accompagné et soutenu durant ce parcours, particulièrement mes chères amies Imane, Lilia, Meriem, Malika et Sarah.

Ces deux citations résonnent souvent en moi pour me donner du courage et de la volonté;

« Don't ever let someone tell you that you can't do something. Not even me. You got a dream, you gotta protect it. When people can't do something themselves, they're gonna tell you that you can't do it. You want something, go get it. Period.»

Will Smith

*« ÇA NE SE FERA PAS EN UN JOUR
MAIS UN JOUR ÇA SE FERA.»*

Franck Nicolas

Table des matières

Table des figures	viii
Liste des tableaux	x
Abréviations	xi
Introduction générale	1
Chapitre 1 État De l'Art Sur Les Méthodes de Segmentation Et De Classification D'images	3
1.1 Introduction	4
1.2 Le mélanome	4
1.3 Diagnostic assisté par ordinateur	6
1.4 Segmentation des lésions cutanées	7
1.5 Classification des lésions cutanées	16
1.6 Conclusion	19
Chapitre 2 Segmentation des images dermoscopiques	20
2.1 Introduction	21

2.2	Segmentation des lésions	21
2.3	Évaluation de l'algorithme de segmentation	27
2.4	Conclusion	30
Chapitre 3 Classification des lésions cutanées		31
3.1	Introduction	32
3.2	Extraction des caractéristiques	32
3.3	Sélection des caractéristiques	44
3.4	Classification	46
3.5	Résultats et discussions	50
3.6	Conclusion	54
Conclusions et perspectives		55
Bibliographie		60

Table des figures

1.1	Types des mélanomes (a) superficiel extensif, (b) nodulaire, (c) lentigo et (d) acral.	4
1.2	Dermoscope.	6
1.3	Schéma bloc d'un DAO.	7
1.4	Segmentation d'une lésion a) illustration de ROI, b) masque de la lésion.	8
1.5	Échantillon des lésions bénignes et malignes de la base des images dermoscopiques PH2.	16
2.1	Filtrage de l'image : (a) Image originale, (b) Élimination des poils avec filtre médiane 11x11.	22
2.2	Étape de prétraitement : Représentation de l'image originale sur les deux plans G ajusté et I_{new} en fonction de la valeur $\text{std}(G)$.	23
2.3	Segmentation des lésions : (a) Image originale, (b) Segmentation avec K-means.	24
2.4	Segmentation des coins par contours actifs : (a) initialisation du contour, (b) coins détectés, (c) superposition du contour sur l'image originale, (d) masque de lésion après suppression des coins.	26
2.5	Filtrage : (a) Masque de lésion avec objets supplémentaires, (b) Érosion, (c) Dilatation et masque final.	26

2.6	Segmentation des lésions : Superposition des contours automatiques et manuels.	29
3.1	Lésion sur-segmentée avec un algorithme de super-pixel.	33
3.2	Extraction des valeurs de couleurs R, G et B à partir des super-pixels.	34
3.3	Segmentation de lésion selon les couleurs dominantes.	36
3.4	Segmentation K-means : a) Une lésion avec trois classes b) Une lésion avec quatre classes.	37
3.5	Formation d'une matrice GLCM à partir de l'image d'entrée : représentée comme une matrice de niveaux d'intensité.	43
3.6	Mise à jour du relief pour une instance cible donnée lorsqu'elle est comparée à son plus proche voisin H et M.	46
3.7	Classification avec K-NN	47
3.8	Modèle de classificateur SVM	49
3.9	Architecture d'un neurone artificiel.	49
3.10	Architecture d'un réseau neuronal artificiel.	50

Liste des tableaux

2.1	Résultats et comparaison de la segmentation automatique proposée.	30
3.1	Les références des six couleurs dominantes sur les lésions.	34
3.2	Seuils des couleurs pour attribuer une couleur super-pixel donnée.	35
3.3	Classification des mélanomes et des lésions bénignes à l'aide de différents classificateurs et espaces couleurs sur l'ensemble de données PH2.	51
3.4	Comparaison des résultats de classification de notre méthode avec trois méthodes de la littérature.	52
3.5	Classification des mélanomes et des lésions bénignes à l'aide de différents classificateurs et espaces couleurs sur l'ensemble de données DermIS & DermQuest.	53

Abréviations

AA Apprentissage Automatique

ABCD Asymmetry, Border irregularity, Color variegation, and Diameter

ACC Accuracy

ANN Artificial Neural Networks

AUC Area Under the Curve

BNN Back-propagation Neural Network

CNN Convolutional Neural Network

DAO Diagnostic Assisté par Ordinateur

DC Coefficient de Dice

FCM Fuzzy C-Means

FCRN Fully Convolutional Residual Network

FN False Negatives

FP False Positives

GAC Geodesic Active Contour

GLCM Gray-level co-occurrence matrix

HSV Hue, Saturation et Value

ISIC International Skin Imaging Collaboration

J Indice de Jaccard

KNN	k-Nearest Neighbours
LS	Level Set
OMS	Organisation Mondiale de la Santé
RGB	Red, Green et blue
ROI	Region Of Interest
SEN	Sensibilité
SPE	Spécificité
SRM	Statistical Region Merging
std	standard deviation
SVM	Support Vector Machine
TDS	Total Dermoscopic Score
TN	True Negatives
TP	True Positives
UV	ultraviolet

Introduction générale

L'Apprentissage Automatique (AA) a connu une progression fulgurante ces dernières années, jouant un rôle essentiel en intégrant tous les domaines, notamment le domaine médical tel que le diagnostic assisté par ordinateur (DAO), analyse des images médicales, ...etc.

L'imagerie médicale a acquis une importance considérable dans le domaine médical tout au long de l'histoire. Elle a été utilisée pour diagnostiquer des maladies, planifier des traitements et évaluer les résultats. En outre, l'imagerie médicale est actuellement utilisée pour prévenir les maladies, généralement par le biais de programmes de dépistage. A ce titre, on s'intéresse à un problème de santé publique majeur : le cancer de la peau de type mélanome. En effet, depuis quelques années, l'incidence du mélanome ne cesse d'augmenter dans tous les pays du monde. D'après l'Organisation Mondiale de la Santé (OMS) 130 000 cas de mélanomes sont enregistrés dans le monde pour l'année 2020 [1].

Le mélanome est le cancer de la peau le plus dangereux, il peut se propager rapidement à d'autres parties du corps s'il n'est pas détecté et traité tôt. Le mélanome à son état primaire est difficile à détecter. En effet, à son état précoce, la lésion maligne présente peu de signes de malignité et peut être, facilement, confondu à une lésion bénigne.

Vu les difficultés d'identification par les méthodes manuelles classiques, telles que la règle ABCD (Asymmetry, Border Irregularity, Colour and Dermoscopic structure) [2], CASH [3] et seven-point checklist [4], qui sont des examens cliniques longs et ne constituent pas une solution efficace au vu du nombre important de cas suspects. Un dermatologue peut prendre beaucoup de temps pour n'examiner qu'un nombre limité de patients.

Il est donc nécessaire de développer un système automatique, rapide et efficace de discrimination des tumeurs malignes des tumeurs bénignes à un stade précoce.

Un tel système permet de détecter des caractéristiques peu discernables à l'œil nu ce qui augmente les chances de discrimination et permet un diagnostic sur une large population en un temps réduit.

L'objectif principal de cette thèse est donc de réaliser un système de diagnostic assisté par ordinateur pour une détection précoce du mélanome. Premièrement, nous avons proposé un algorithme de segmentation automatique basée sur le k-means pour générer un masque pour chaque lésion. Deuxièmement, en utilisant les masques générés par la segmentation, nous avons extrait des caractéristiques basées sur la couleur et la texture des lésions. Ensuite, un algorithme de sélection des attributs est utilisé afin de ne garder que les caractéristiques les plus pertinentes et d'éliminer les redondances. Enfin, pour la classification, nous avons testé différents espaces couleurs en les associant avec différents classificateurs afin de trouver la combinaison espace couleur/classificateur optimale. Finalement, nous avons testé et validé l'algorithme proposé sur un ensemble de données PH2 des images dermatoscopiques de l'hôpital Pedro Hispano [5]. Les résultats obtenus surpassent les résultats des travaux antérieurs qui ont utilisé la même base de données.

Cette thèse est organisée de la manière suivante :

Le premier chapitre est consacré à l'état de l'art des différentes approches de segmentation et de classification des travaux publiés.

Le deuxième chapitre présente l'approche de segmentation proposée.

Le troisième chapitre concerne la classification, après l'extraction et le choix des caractéristiques à partir de cinq espaces couleurs, trois classificateurs K-plus proche voisin (KNN), Machine à vecteurs de support (SVM) et le réseau neuronal artificiel (ANN) ont été utilisés.

Nous terminons par une conclusion générale et des perspectives sont ouvertes par ce travail de recherche.

Chapitre 1

Chapitre 1 État De l'Art Sur Les Méthodes de Segmentation Et De Classification D'images

1.1 Introduction

Ce chapitre est un aperçu de l'état de l'art des différents algorithmes de segmentation et de classification des lésions cutanées. La première partie est une étude des différentes approches de segmentation et leurs limitations lorsqu'elles sont appliquées avec les images dermoscopiques. La deuxième partie est consacrée aux approches de classification.

1.2 Le mélanome

Le cancer de la peau est une maladie cutanée caractérisée par la croissance anormale de cellules. Dans un premier temps, ces cellules se propagent uniquement dans leur environnement proche, il s'agit d'une tumeur bénigne. Ensuite, elles continuent à se propager et finissent par envahir d'autres parties du corps à travers le sang et la lymphe. Ces cellules tumorales, communément appelées métastases détériorent alors une ou plusieurs fonctions de l'organisme et peuvent conduire à la mort. On parle alors d'une tumeur maligne. Les cancers de la peau les plus fréquents sont : le carcinome basocellulaire, le carcinome épidermoïde et le mélanome.

Actuellement, le mélanome est l'un des cancers les plus dangereux, c'est un problème majeur de santé publique dans le monde. Une fois que la propagation des métastases commence, les traitements tels que la radiothérapie et la chimiothérapie sont inefficaces. L'immunothérapie, en revanche, augmente la durée de survie sans récurrence du mélanome. Les mélanomes sont regroupés en plusieurs sous-types dont les quatre principaux sont le mélanome superficiel extensif, nodulaire, lentigo et acral [6].



FIGURE 1.1 – Types des mélanomes (a) superficiel extensif, (b) nodulaire, (c) lentigo et (d) acral.

1.2.1 Facteurs de risques

Le mélanome provient de la transformation maligne des mélanocytes. Les mécanismes par lesquels le mélanome se développe sont encore mal connus, mais trois principaux facteurs de risque ont été identifiés :

- l'exposition aux ultraviolets UV,
- les prédispositions génétiques,
- les antécédent personnels de mélanome.

1.2.2 Développement du mélanome

Dans la plupart des cas, l'évolution du mélanome se divise en quatre stades bien définis.

- **Premier stade** : les cellules tumorales sont localisées uniquement dans l'épiderme et progressent horizontalement (mélanome in situ). A ce stade, une simple excision de la zone affectée suffit pour éliminer définitivement le cancer et éviter ainsi toute évolution.
- **Deuxième stade** : si aucun traitement n'est effectué, la tumeur se propage en profondeur, traverse la membrane basale et colonise le derme et l'hypoderme.
- **Troisième stade** : les cellules cancéreuses empruntent la circulation lymphatique pour envahir les ganglions proches de la lésion initiale.
- **Quatrième stade** : les métastases continuent de se développer et atteignent les organes profonds.

1.2.3 Détection et traitement du mélanomes

La détection précoce du mélanome constitue une étape importante du traitement des personnes atteintes. En effet, si la tumeur peut être retirée avant qu'elle n'ait franchi la jonction dermo-épidermique, l'individu est en phase de guérison. Généralement, le moyen de détection consiste en un examen à l'œil nu par un dermatologue. Ce dernier utilise couramment des dermoscopes (Figure [1.2](#)) pour une

meilleure visualisation. Plusieurs règles ont été développées afin de caractériser plus efficacement une lésion suspecte, la plus connue est la règle ABCD [2]. L'utilisation de cette dernière présente plusieurs limites. En effet, elle n'est pas applicable pour des lésions précoces.



FIGURE 1.2 – Dermoscope.

Lorsque le mélanome n'a pas encore formé de métastases, le seul traitement curatif est une exérèse chirurgicale. À ce jour, il n'existe pas de traitement pour le mélanome métastasé permettant une rémission complète. Il est donc nécessaire de développer un outil de détection automatique du mélanome précoce plus efficace.

1.3 Diagnostic assisté par ordinateur

Le diagnostic assisté par ordinateur (DAO) pour la détection du mélanome a été introduit afin d'améliorer et d'aider les dermatologues et d'autres cliniciens dans la prévention du cancer de peau. La construction d'un DAO repose principalement sur quatre étapes essentielles : le prétraitement de l'image, la segmentation, l'extraction des attributs et la classification. La figure 1.3 présente le schéma bloc d'un DAO.

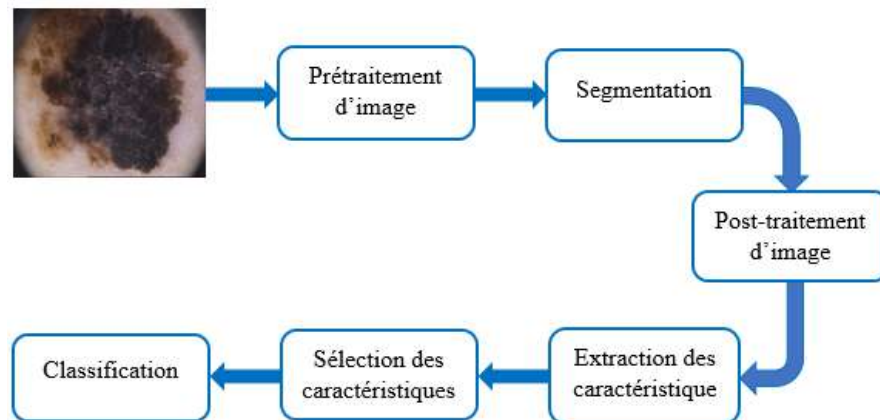


FIGURE 1.3 – Schéma bloc d'un DAO.

1. **Prétraitement** : regroupe l'ensemble des processus visant à améliorer la qualité de l'image acquise.
2. **Segmentation** : est une répartition de l'image en régions homogènes selon un critère déterminé : couleur, texture, niveau de gris, ...etc.
3. **Extraction d'attributs (caractéristiques)** : vise à extraire les informations pertinentes qui caractérisent chaque classe.
4. **Sélection des caractéristiques** : est utilisée pour sélectionner les caractéristiques les plus pertinentes et de réduire la dimension de l'espace des caractéristiques de manière à éliminer les caractéristiques redondantes.
5. **Classification d'images** : est une identification des classes en utilisant les attributs sélectionnés.

1.4 Segmentation des lésions cutanées

La segmentation des lésions cutanées joue un rôle crucial dans la détection automatique des mélanomes en délimitant la région d'intérêt (ROI) comme montré sur la figure 1.4, vu que dans cette région que toutes les caractéristiques sont extraites. L'étude comparative des performances des différentes techniques de segmentation présentées dans les articles [7, 8], démontrent la contribution de la segmentation dans l'amélioration de la fiabilité d'un système DAO. Pour cela, différentes techniques de segmentation automatique ont été proposées pour la délimitation de ROI [9, 10, 11].

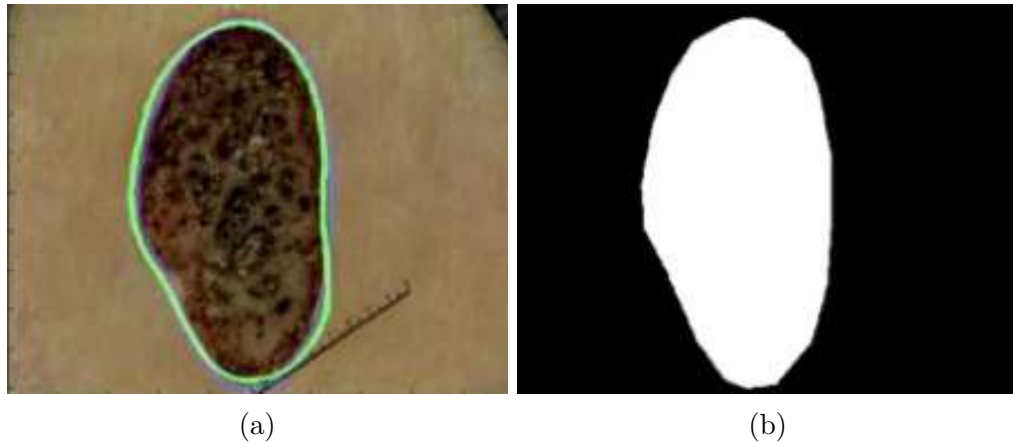


FIGURE 1.4 – Segmentation d’une lésion a) illustration de ROI, b) masque de la lésion.

1.4.1 Approches classiques pour la segmentation de lésions cutanées

Les approches classiques utilisées pour la segmentation des images dermoscopiques peuvent être classées en méthodes de seuillage, segmentation basée régions et la segmentation par la détection des contours [7].

a) Les méthodes de seuillage

Le principe des méthodes de seuillage est de déterminer une valeur seuil pour transformer une image en niveaux de gris à une image binaire [12].

A titre d’illustration, Celebi et al. [13] ont développé une méthode de segmentation automatique par seuil pour détecter les bords des lésions sur des images dermoscopiques. Des expériences sur un ensemble de 90 images montrent que la méthode proposée est robuste, rapide et précise par rapport à neuf méthodes de l’état de l’art. La moyenne et l’écart-type des erreurs XOR de la méthode proposée sont 8.31 et 4.06 respectivement.

Abbas et al. [14] proposent un algorithme qui se compose de quatre étapes principales : Tout d’abord, l’image dermoscopique RGB est transformée en espace couleur CIE $L^*a^*b^*$, le contraste de la lésion est ensuite amélioré, un algorithme de hill-climbing est ensuite utilisé pour détecter ROI et enfin, un seuillage adaptatif est appliqué pour déterminer la limite optimale de la lésion. Une étude comparative avec trois techniques de segmentation basée sur la couleur et la texture est menée pour montrer l’efficacité de la méthode.

Sur un ensemble de 100 images dermoscopiques, les mesures du taux de vrais positifs (TPR), du taux de faux positifs (FPR) et de la probabilité d'erreur (EP) ont atteint les valeurs de 94,25%, 3,56% et 4% respectivement.

Dans l'article de Nidhal et al. [15], une nouvelle méthode de segmentation des images dermoscopiques est introduite. La segmentation est principalement basée sur le seuillage de l'histogramme après l'étape du prétraitement qui consiste en une opération de filtrage en utilisant le filtre de Wiener afin d'éliminer les poils. La méthode proposée est évaluée en comparant les images segmentées automatiquement avec des masques manuelles. La précision de la méthode proposée était de 96,32%.

La méthode de segmentation des lésions cutanées proposée par Tajeddin et al. [16] est basée sur trois étapes principales. Après une phase de prétraitement, une technique de seuillage basée sur l'histogramme est utilisée pour obtenir un masque initial de la lésion. Le masque initial est utilisé pour échantillonner la couleur de la lésion et piloter un algorithme de propagation des contours. L'algorithme proposé a été testé sur la base de données ISIC de 900 images dermoscopiques et a obtenu, sur les deux métriques utilisées, coefficient de Dice et de Jaccard, un taux de 0,89 et 0,79, respectivement.

Maciel et al. [17] ont proposé une méthode de seuillage pour segmenter automatiquement les lésions cancéreuses de la peau. L'algorithme est basé sur une modification de la méthode de seuillage d'Otsu, analysant séparément des échantillons provenant de différentes zones de l'image. La méthode proposée a été testée et validée sur l'ensemble de données DermQuest. Les résultats expérimentaux montrent que la méthode peut potentiellement être utilisée pour segmenter les naevus atypiques et les mélanomes dans les lésions avec une peau de fond très hétérogène en atteignant une sensibilité de 94,60% et une spécificité de 99,51%.

Haidi et al. [18] ont proposé une méthode de segmentation des lésions basée sur la saillance de l'image combinée au seuil d'Otsu. La méthode comprend des étapes de filtrages et de segmentation. Dans l'étape de segmentation, selon la distribution de l'histogramme, une fonction d'optimisation est conçue pour ajuster la méthode traditionnelle du seuil d'Otsu afin d'obtenir des limites de lésions plus précises. Les résultats expérimentaux démontrent que la méthode est robuste et plus performante que les autres méthodes déjà existantes dans la littérature, avec une précision de 97,65%, 96,78% et 97,28% réalisée sur les trois ensembles de données EDRA, PH2 et ISIC 2016, respectivement.

b) Segmentation basée sur les régions

Repose sur l'idée que les valeurs des pixels d'une même région dans l'image varient très peu. Elle implique le regroupement de pixels en régions homogènes en utilisant la fusion de régions, la division de régions ou une combinaison des deux. Les images sont divisées en régions ou groupes de pixels similaires en fonction de leurs propriétés. Chaque pixel est comparé à ses voisins dans une région particulière et, selon certaines conditions.

Celebi et al. [19] ont présenté une approche non supervisée de la détection des frontières des lésions cutanées, basée sur l'algorithme de fusion de régions statistiques SRM. L'approche proposée comprend trois phases principales : le prétraitement, la segmentation et le post-traitement. La phase de prétraitement comprend le lissage de l'image et la suppression du cadre noir qui caractérise certaines images dans la base utilisée. La phase de segmentation est basée sur la fusion des régions. Enfin, la phase de post-traitement consiste à éliminer les régions appartenant à l'arrière-plan, à supprimer les régions isolées, à fusionner les régions restantes et enfin à élargir la frontière initiale par dilatation morphologique. La méthode est testée sur un ensemble de 90 images et est comparée à quatre méthodes automatiques publiées. La méthode SRM proposée obtient les meilleurs résultats avec un écart-type de l'erreur de frontière de 5.308%.

Une nouvelle approche itérative stochastique de fusion de régions est utilisée dans le travail de Wong et al. [20] pour segmenter les lésions, où la fusion stochastique de régions est initialisée d'abord au niveau du pixel, puis au niveau de la région jusqu'à convergence. Une fonction de vraisemblance de fusion de régions basée sur les statistiques régionales est introduite pour déterminer la fusion des régions de manière stochastique. Pour évaluer l'efficacité de la méthode proposée, une segmentation a été réalisée sur 60 images. Les résultats expérimentaux montrent que la méthode proposée permet d'obtenir une erreur de segmentation globale inférieure à 10% pour les lésions cutanées, ce qui est inférieur à celui obtenu par les méthodes existantes dans la littérature.

Récemment, Bama et al. [21] ont proposé une nouvelle approche de croissance de région basée sur les super-pixels pour la segmentation des lésions cutanées. Elle est basée sur les super-pixels du modèle de mélange gaussien qui segmentent l'image en régions homogènes. Le système proposé est testé sur l'ensemble de données PH2. Les résultats expérimentaux donnent une sensibilité de 79.98% et une spécificité de 97.32%.

c) Segmentation basée sur les contours

Consiste à détecter les contours d'un objet à l'aide de techniques d'évolution des courbes.

Cavalcanti et al. [22] ont proposé de segmenter les lésions cutanées avec les contours actifs. Premièrement les images sont prétraitées en créant une nouvelle représentation d'image (plan) où les images sont plus contrastées. Après seuillage de la lésion dans ce nouveau plan, la segmentation est affinée à l'aide d'un algorithme de contours actifs suivi d'un post-traitement basé sur les opérations morphologiques pour mieux déterminer les limites de la lésion. Les résultats expérimentaux sur un ensemble de données de 152 images donnent une sensibilité de 90.07% et une spécificité de 99.11%.

Dans l'article d'Abbas et al. [23], un algorithme de segmentation est développé sur la base des étapes suivantes, transformation de l'espace colorimétrique vers le modèle de couleur CIE-CAM02 orienté perception, étape de prétraitement pour corriger la réflexion spéculaire, amélioration du contraste à l'aide d'une transformation homomorphe filtre (HTF) et fonction sigmoïde non linéaire (NSF) et segmentation avec un modèle de contours actifs (ACM). Pour valider la technique proposée, des comparaisons avec trois autres algorithmes de segmentation ont été effectuées pour 210 images de la base de données EDRA. À partir de ces expériences, un taux de détection moyen réel de 91,01%, un taux de faux positifs de 6,35 et une probabilité d'erreur de 7,8 ont été obtenus.

Kasmi et al. [24] ont proposé une technique pour la segmentation, un algorithme de détection de contours biologiquement inspiré en utilisant le géodésique contour actif (GAC). L'algorithme utilise une initialisation automatique du contour à proximité de la limite réelle de la lésion. De plus, un nouveau plan d'image est trouvé en transformant l'image RGB originale en une image lissée qui permet au GAC de se déplacer sans se coller sur l'énergie locale minimale. Un post-traitement qui élimine les contours qui dépassent ou qui rentrent dans la lésion sont corrigés pour avoir un aspect d'un contour biologique et naturel. L'auteur a introduit les contours parallèles : plusieurs contours sont générés pour une lésion afin d'en choisir le plus précis. L'algorithme a été testé sur la base EDRA de 100 images, donnant une erreur de frontière médiane XOR de 6,7%. Sur une autre base de données de 1238 images caractérisées par des bords à faible contraste, une erreur médiane XOR de 23,9% est obtenue, comparée à des masques manuels des images segmentées par trois dermatologues.

La méthode de segmentation des lésions cutanées proposée par Farhan et al. [25] est basée sur les contours actifs. En prenant un contour initial de la

lésion suivi par le critère d'optimisation qui utilise un terme basé sur les régions (convergence des moyennes de la lésion et de la peau vers leurs maxima locaux) intégré au gradient de l'image (contours) pour effectuer la segmentation de l'image. La fonction objective consiste à maximiser la distance de la distribution des niveaux de gris entre la peau et la lésion pour ajuster une courbe aux limites de la lésion. La méthode proposée exploite les avantages des différentes méthodes de segmentation, notamment le seuillage, les méthodes basées sur les contours et les méthodes basées sur les régions. Elle est évaluée sur les deux ensembles de données dermoscopiques de l'International Skin Imaging Collaboration (ISIC) et PH2. La méthode de segmentation proposée atteint un indice dice de 80.17% et 86.54% respectivement.

Dans l'article d'Abhilash et al. [26], une segmentation adaptative de contours actifs est proposée. Une étape de prétraitement est utilisée pour éliminer les poils. L'image dermoscopique RGB est convertie en image en niveaux de gris, et l'opération de fermeture est effectuée sur l'image en niveaux de gris. La fermeture est une dilatation suivie d'une érosion. Enfin, un seuillage est effectué pour obtenir un masque. Les résultats obtenus par cette méthode ont fourni de meilleurs résultats de segmentation par rapport aux méthodes basées sur le seuillage d'Otsu et la croissance de région avec un taux d'erreur moyen de 0,13.

1.4.2 Segmentation basée clustering

La segmentation par clustering consiste à affecter à chaque pixel de l'image une classe qui définit les régions à délimiter sur l'image. Nous pouvons donc citer des algorithmes de classification de pixels non-supervisés comme C-moyennes floues [27] et les k-means proposés par MAC QUEEN [28].

a) Segmentation basée C-moyennes floues (FCM)

FCM est un algorithme de segmentation basé sur le regroupement de pixels similaires de manière itérative, les centres de regroupement étant ajustés à chaque itération [27]. En raison de sa nature itérative, le coût de calcul de l'algorithme est relativement élevé par rapport aux autres techniques de segmentation. C'est pourquoi un certain nombre d'approches, par exemple [29] et [30], ont été présentées pour permettre des accélérations significatives tout en maintenant de bonnes performances de segmentation.

Masood et al. [31] ont proposé une nouvelle méthode de segmentation qui combine FCM, le seuillage et la méthode level set (LS). Tout d'abord, l'image

est prétraitée pour éliminer les artefacts (poils, bulles d'air, et les marques de règle) en utilisant un filtre médian 7×7 . Ensuite, un seuillage basé sur le FCM est utilisé pour initialiser l'évolution du LS et réguler les paramètres de réglage. La méthode proposée a montré une bonne précision de segmentation des lésions cutanées avec un taux de détection vrai moyen de 92,6% et une erreur de faux positifs et de faux négatifs assez réduite, soit 4,66% et 7,34% respectivement.

Eltayef et al. [32] ont présenté une nouvelle méthode de segmentation des mélanomes à partir des images dermoscopiques. Pour détecter les poils et autres bruits dans les images, une étape de prétraitement est effectuée en appliquant le filtre Gabor. La région d'intérêt (ROI) de l'image est extraite en combinant la méthode FCM et Markov Random Field. La méthode a été évaluée sur un ensemble de données PH2 de 200 images dermoscopiques. Les résultats de la méthode proposée ont été comparés aux masques de la vérité terrain et atteignent une sensibilité de 93,2%, une spécificité de 98,0% et une précision de 94,0%.

Dans l'article de Jose Luis Garcia et al. [33], l'algorithme proposé de segmentation des lésions cutanées dans les images dermoscopiques est basé sur une classification floue des pixels et un seuillage de l'histogramme. La méthode a été testée sur deux bases de données publiques ISIC 2016 et ISIC 2017 contenant respectivement 379 et 600 images, et comparée avec les autres travaux de l'état de l'art, à l'aide des métriques comme l'exactitude, le coefficient de Dice, l'indice de Jaccard, la sensibilité et la spécificité : (0.934, 0.869, 0.791, 0.870 et 0.978) et (0.884, 0.760, 0.665, 0.869 et 0.923) respectivement.

b) Segmentation basée K-means

L'algorithme des k-means réalise un partitionnement automatique des pixels de l'image en k classes homogènes. Divers algorithmes ont été développés et appliqués pour la segmentation des lésions cutanées dans le passé. Cette partie présente l'application de certains de ces algorithmes et leurs performances sur la segmentation des ensembles de données d'images dermoscopiques.

Alvarez et al. [34], ont développé un algorithme de segmentation qui se compose des étapes suivantes : une phase du prétraitement de l'image qui consiste en une opération de fermeture morphologique réalisée indépendamment sur chacun des canaux de couleur de l'image RGB, en utilisant comme élément de sélection un disque de rayon égal à 5 pixels. Ensuite, un filtre médian 3×3 est appliqué à chaque canal pour éliminer les poils. Une segmentation de l'image en utilisant l'algorithme k-means, un calcul d'un ensemble de caractéristiques décrivant les propriétés de chaque région segmentée et enfin, un calcul d'un score final pour chaque région, représentant la probabilité de

correspondre à une segmentation de lésion appropriée. L'algorithme atteint un indice Jaccard d'environ 0,679.

Dans le travail d'Agarwal et al. [35], une méthode de segmentation des lésions cutanées basée sur la technique du clustering est utilisée. D'abord, l'image RGB est convertie en espace couleur $L^*a^*b^*$. Une combinaison stratégique est appliquée aux différents canaux de l'espace $L^*a^*b^*$ pour obtenir une nouvelle image en niveaux de gris dans laquelle les pixels de la lésion et de la peau sont plus différenciables. Ensuite, le clustering K-means est mis en œuvre pour segmenter la lésion de l'image prétraitée. Enfin, une étape de post-traitement utilise un filtre moyenneur et un seuillage basé sur la surface, pour supprimer les pixels bruités (les faux positifs) de l'image segmentée. La méthode proposée a été testée sur 50 images de la base de données public DermIS et DermQuest. Les valeurs du score de chevauchement et du coefficient de corrélation de la méthode proposée sont respectivement de 96,75% et 97,66%.

Jaisakthi et al. [36], proposent une méthode de segmentation automatique des lésions cutanées qui comprend deux étapes principales, à savoir le pré-traitement et la segmentation. Dans l'étape de pré-traitement, les bruits tels que l'illumination, les poils et les marques de règles sont supprimés à l'aide de techniques de filtrage. Dans la phase de segmentation, les lésions cutanées sont segmentées à l'aide de l'algorithme de segmentation GrabCut [37]. Ensuite, l'algorithme de regroupement des K-means est utilisé avec les caractéristiques de couleur apprises à partir des images d'entraînement pour améliorer les limites des segments. Pour évaluer la méthode proposée, les auteurs ont utilisé les deux ensembles de données ISIC 2017 et PH2. Ils ont obtenu des valeurs de coefficient de Dice de 0,8236 et 0,9139 pour ISIC 2017 et PH2, respectivement.

1.4.3 Segmentation basée sur les réseaux neuronaux convolutifs

Récemment, les réseaux neuronaux convolutifs ont été adoptés pour augmenter la précision de la segmentation des images de lésions cutanées.

Zhang et al. [38], ont développé une structure de réseau de neurones CNN pour segmenter les lésions cutanées. L'architecture du réseau comprend plusieurs couches de convolution, des couches softmax, plusieurs filtres et des fonctions d'activation. Une fonction d'activation non linéaire (telle que ReLU et ELU) est utili-

sée pour atténuer le problème de disparition du gradient et RMSprop/Adam sont incorporés pour optimiser l'algorithme de perte. La méthode a été testée sur l'ensemble de données ISIC 2017, et elle a atteint une précision de 0.91 qui est plus élevée que les architectures existantes.

Dans le travail de Yu et al. [39], un nouveau réseau résiduel entièrement convolutif (FCRN) de plus de 50 couches est proposé, pour la segmentation des images dermoscopiques. Des expériences sur la base de données ISIC 2016 ont montré l'efficacité de la méthode proposée. Le FCRN à 50 couches a donné de meilleurs résultats que les architectures VGG-16, GoogleNet, FCRN-38 et FCRN-100, avec une précision de segmentation de 94,9%.

Yuan et al. [40] utilisent un réseau convolutif profond à 19 couches pour la segmentation des lésions cutanées. L'augmentation des données d'image a été adoptée pour résoudre le problème des données limitées. De plus, les performances de segmentation ont été améliorées par l'utilisation de techniques telles que la normalisation par lots [41], l'optimisation Adam [42] et la fonction de perte basée sur l'indice Jacquard. Les résultats expérimentaux ont montré que la méthode proposée surpasse les autres algorithmes de l'état de l'art sur la base de données ISIC 2016, avec un indice de Jaccard de 0,861.

Yang et al. [43], ont utilisé l'architecture U-Net pour segmenter les lésions cutanées. Le modèle d'apprentissage profond multitâche proposé est entraîné et évalué sur l'ensemble d'images dermoscopiques ISIC 2017, qui comprend 2000 échantillons d'entraînement et 150 échantillons d'évaluation. Les résultats expérimentaux montrent que le modèle d'apprentissage profond multitâche proposé atteint des performances prometteuses sur la segmentation des lésions cutanées. La valeur moyenne de l'indice de Jaccard pour la segmentation des lésions est de 0,724.

Dans l'article de Lin et al. [44], une étude comparative de deux approches différentes de la segmentation des lésions cutanées est présentée. La première approche utilise des U-Nets et introduit une étape de prétraitement basée sur l'égalisation des histogrammes. La deuxième approche est une approche basée sur le clustering K-means. Les expériences ont été testées sur l'ensemble de données ISIC 2017 pour valider l'étude comparative. En comparant les deux algorithmes proposés, les U-Nets ont obtenu un indice de Jaccard significativement plus élevé (0.62) par rapport à l'approche de clustering (0.44).

1.5 Classification des lésions cutanées

Différencier les cas malins du bénins, comme le montre la figure 1.5 est une tâche très difficile, même pour des spécialistes expérimentés. L'utilisation de techniques informatiques et d'algorithmes peut être d'une grande aide.

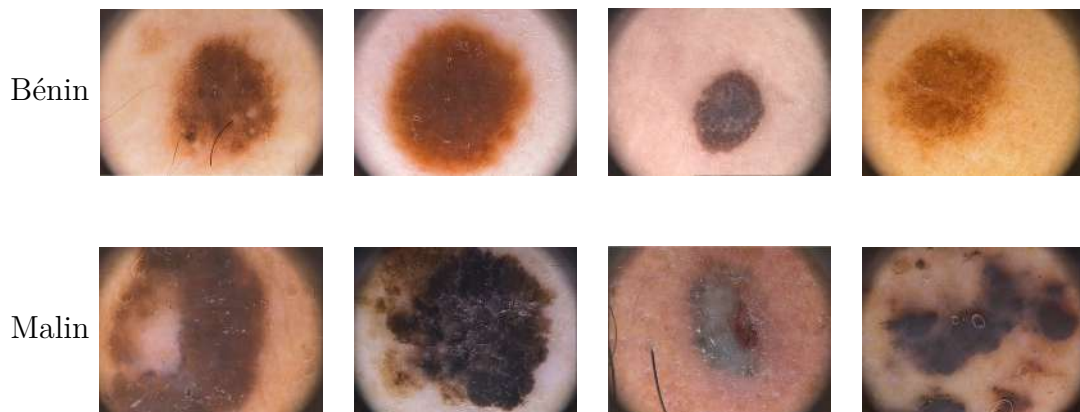


FIGURE 1.5 – Échantillon des lésions bénignes et malignes de la base des images dermoscopiques PH2.

Kasmi et al. [45], ont proposé d'automatiser la règle ABCD afin de discriminer les tumeurs malignes des tumeurs bénignes. Une étape de prétraitement est utilisée pour éliminer les artefacts (bulles et poils fins) à l'aide d'un filtre médian. Les poils épais sont détectés en utilisant des filtres de Gabor avec différentes fréquences et orientations. Ensuite, les lésions sont segmentées par la méthode des contours actifs géodésiques (GAC). Enfin, les valeurs de A, B, C et D sont estimées par différents algorithmes et le Total Dermoscopic Score (TDS) est calculé. Testé sur 200 images dermoscopiques, l'algorithme atteint un taux de sensibilité de 91,25%, une précision de 95,83% et une exactitude de 94,0%.

Jaworek et al. [46], ont utilisé la forme et la texture de la lésion pour détecter le mélanome à son stade précoce avec un classificateur SVM. La méthode proposée contient les étapes suivantes : l'étape du prétraitement qui comprend l'élimination du cadre noir, le lissage des bulles d'air et la suppression des poils noirs en utilisant un filtre gaussien et les opérations morphologiques. Ensuite, une segmentation des lésions basée sur l'algorithme de croissance de régions est utilisée. L'extraction des caractéristiques est basée sur les deux techniques, la règle ABCD pour la forme et

7-points checklist pour la texture. Ils ont utilisé un ensemble de 200 images et ont obtenu une sensibilité de 90%, une spécificité de 96% et une aire sous la courbe de 93,24%.

Dans l'article de Barata et al. [47], un ensemble de caractéristiques est utilisé. L'étude compare le poids des descripteurs de couleur (moments et histogrammes) et de texture (Gabor et Laws) pour le mélanome. Les résultats montrent que les descripteurs de couleur fonctionnent mieux que les descripteurs de texture et que de bons résultats de classification peuvent être obtenus en utilisant SVM, avec une sensibilité de 93% et une spécificité de 88% sur un ensemble de données de 176 images dermoscopiques. La fusion des descripteurs de couleur et de texture a également obtenu de bons résultats, avec une sensibilité de 96%, spécificité de 82% pour la combinaison des moments avec les descripteurs de texture Gabor et Laws.

Alfed et al. [48] combinent la texture (histogramme des gradients, histogramme des lignes) avec les angles des vecteurs de couleur et les moments de Zernike comme caractéristiques pour classer les mélanomes des lésions bénignes de l'ensemble de données PH2. En utilisant trois classificateurs : SVM, Adaboost et ANN, la méthode atteint une sensibilité de 99,41%, une spécificité de 98,18% et une précision de 98,79%.

Majumder et al. [49] ont utilisé un réseau neuronal à rétropropagation (BNN) basé sur les caractéristiques de la règle ABCD pour classer les mélanomes et les tumeurs bénignes dans l'ensemble de données PH2. Ils ont développé cinq caractéristiques pour classer les mélanomes malins et bénins. Ces caractéristiques sont le score d'asymétrie (AS), l'irrégularité des bords (B), la variation des couleurs (C), le diamètre (D1) et la différence entre les diamètres de Feret maximum et minimum de l'ellipse la mieux adaptée à la lésion (D2). La technique proposée atteint une précision de 98%, une sensibilité de 95% et une spécificité de 98,8%.

Hagerty et al. [50] mélangent les caractéristiques classiques et d'apprentissage profond après les avoir évaluées individuellement. D'une part, les trois caractéristiques classiques sont d'inspiration biologique et d'information clinique : réseau de pigments atypiques, distribution des couleurs et vaisseaux sanguins. Les informations cliniques comprennent les informations soumises au pathologiste - âge du patient, sexe, emplacement de la lésion, taille et antécédents du patient. Les caractéristiques d'apprentissage profond utilisent le transfert de connaissances par l'intermédiaire d'un réseau ResNet-50 qui est reconverti pour prédire la probabilité de classification du mélanome. Les scores de classification de chaque caractéristique individuelle, classique et d'apprentissage profond, sont ensuite assemblés à l'aide

de la régression logistique pour prédire une probabilité globale des mélanomes. Les résultats de classification des mélanomes effectuée sur un ensemble de données HAM10000 comprenant 9174 lésions, mesurés par l'aire sous la courbe (AUC) , montrent l'efficacité des caractéristiques de fusion en obtenant une précision de classification de 0,94.

Moura et al. [51] utilisent la règle ABCD et des caractéristiques de réseaux neuronaux convolutionnels pré-entraînés CNN : Vgg-m, Vgg-f, Vgg-verydeep-19 et CaffeNet. A partir du masque binaire, fourni par la base de données utilisée, la lésion cutanée est segmentée. Les images originales RGB sont converties dans l'espace couleur HSV . Ensuite, les caractéristiques sont extraites par les descripteurs de la composante générée par la somme des plans S et V, puis les descripteurs les plus pertinents sont choisis. Enfin, un classificateur Perceptron multicouche est utilisé pour distinguer le mélanome des lésions bénignes. La méthode proposée a été testée sur les deux bases de données PH2 et DermIS. La méthode a présenté de bons résultats avec un taux de précision de 94,9% et un indice de Kappa de 89,2%.

Hirano et al. [52] ont développé un système de diagnostic automatisé pour l'identification du mélanome en utilisant des données hyperspectrales (HSD) et GoogLeNet. Le GoogLeNet préformé, est utilisé pour alimenter le HSD, ils ont ajouté un nouveau réseau, appelé "Mini Network", qui effectue une réduction de taille juste avant la couche d'entrée de GoogLeNet. Les travaux ont utilisé 619 lésions pour l'entraînement et les tests. L'évaluation sur 5 validations croisées indique que sans augmentation des données, la sensibilité, la spécificité et la précision sont respectivement de 69.1%, 75.7% et 72.7%. Et avec l'augmentation des données, la sensibilité, la spécificité et la précision sont respectivement de 72.3%, 81.2% et 77.2%.

La méthode de classification du mélanome proposée par Jojoa Acosta et al. [53] se compose de deux étapes : tout d'abord, la région d'intérêt est délimitée à l'aide de la technique du masque et du réseau neuronal convolutif basé sur la région, puis un classificateur basé sur la structure ResNet152 est utilisé pour distinguer les lésions malignes. La méthode est évaluée sur l'ensemble de données ISIC 2017 et atteint une précision de 90,4%, une sensibilité de 82% et une spécificité de 92,5%.

Fekrache et al. [54] ont proposé un système automatisé qui utilise un algorithme de segmentation basé sur les colonies de fourmis, puis trois types de caractéristiques sont extraits, basées sur les propriétés géométriques de la lésion telles que décrites par la règle ABCD, les caractéristiques de textures calculant des caractéristiques basées sur l'histogramme du premier ordre, et la couleur relative.

112 caractéristiques sont extraites et 12 attributs pertinents sont sélectionnés à l'aide de l'algorithme Relief. Deux classificateurs sont utilisés, le KNN et ANN. Le système proposé est testé sur 172 images dermoscopiques et atteint une précision de 85.22% et 93.60% avec KNN et ANN respectivement.

1.6 Conclusion

Ce chapitre introductif a été consacré essentiellement à la présentation des différentes méthodes existantes dans la littérature relative aux processus de segmentation et de classification des images dermoscopiques. Comme mentionné le long de ce chapitre, la segmentation se présente sous différentes approches et est utilisée sous différentes manières.

Le chapitre suivant décrit la méthode de segmentation proposée, en l'occurrence la segmentation à base du K-means.

Chapitre 2

Chapitre 2 Segmentation des images dermoscopiques

2.1 Introduction

Ce chapitre est consacré à la présentation de la méthode de segmentation proposée et utilisée pour générer les masques des lésions.

2.2 Segmentation des lésions

L'algorithme est constitué de trois étapes principales. La première étape consiste à prétraiter l'image en utilisant un filtre médian pour éliminer les poils fins. La deuxième étape repose sur le choix du plan de représentation de l'image afin de mettre en évidence la lésion puis, la méthode k-means est appliquée pour segmenter facilement les images dermoscopiques. La troisième étape est un post-traitement des masques des images segmentées pour filtrer les objets détachés de la lésion, affiner le résultat et minimiser l'erreur de segmentation.

2.2.1 Prétraitement de l'image

L'étape de prétraitement a pour but de faciliter la segmentation en renforçant la similitude entre les pixels appartenant à une même région, ou en accentuant la dissimilarité entre pixels appartenant à des régions différentes ainsi que la suppression des artefacts.

Nous avons choisi deux plans d'images pour la segmentation, soit sur le plan G de l'espace couleur RGB après ajustement de son histogramme ou un nouveau plan I_{new} , proposé, qui combine les deux plans S (saturation) et V (valeur) de l'espace couleur HSV [55]. Le second plan I_{new} est trouvé comme indiqué dans l'équation 2.1

$$I_{new} = \begin{cases} (0.9S + 0.3V) & \text{si } (0.9S + 0.3V) < 1 \\ 1 & \text{sinon} \end{cases} \quad (2.1)$$

Ensuite, l'image représentée dans le nouveau plan G ajusté ou I_{new} est filtrée par un filtre médian 11x11 pour minimiser l'effet des poils tout en préservant les contours de la lésion (figure 2.1), sachant que les images de la base PH2, utilisées dans cette thèse, ne sont pas recouvertes de poils épais qui nécessite une méthode de suppression de poils.

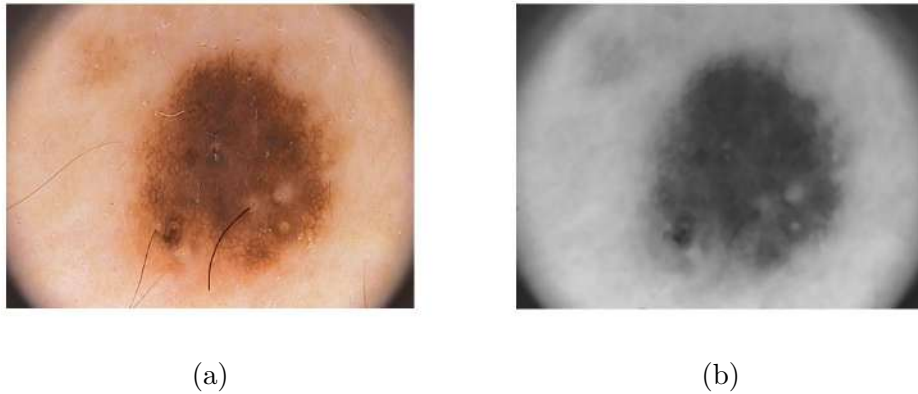


FIGURE 2.1 – Filtrage de l'image : (a) Image originale, (b) Élimination des poils avec filtre médiane 11x11.

2.2.2 Choix du plan de représentation de l'image

Le choix du plan convenable pour la segmentation est basé sur l'écart-type du plan vert ($\text{std}(G)$), selon l'équation 2.2. Nous avons constaté que si l'écart-type du plan vert G est supérieur à 0,25, la lésion est suffisamment mise en évidence pour être segmentée par les k-means (figure 2.2a sur le plan G ajusté), sinon, nous utilisons le plan trouvé par l'équation 2.1.

$$I = \begin{cases} I_{new} & \text{si } \text{std}(G) < 0,25 \\ G & \text{sinon} \end{cases} \quad (2.2)$$

I : L'entrée de K-means qui peut être un plan I_{new} ou G (plan vert). Ce choix dépend de l'écart type du plan vert ($\text{std}(G)$).

En raison de la faiblesse des limites des lésions (figure 2.2b), les méthodes de détection des contours ne peuvent pas bien délimiter toutes les lésions en utilisant les plans connus (R, G, B, H, S, V, \dots etc). Il est donc recommandé de modifier un plan ou de combiner des plans dans le but de mettre en évidence la lésion. Dans notre cas, l'intensité de la couleur S et la luminosité V sont combinées en réduisant les deux valeurs ($S \times 0.9$) et ($V \times 0.3$). Sachant que la lésion est plus colorée et moins claire que le fond, la combinaison des plans S et V réduits rend la lésion plus claire et le fond plus sombre (figure 2.2b sur le plan I_{new}). La figure 2.2 montre l'étape de prétraitement suivie.

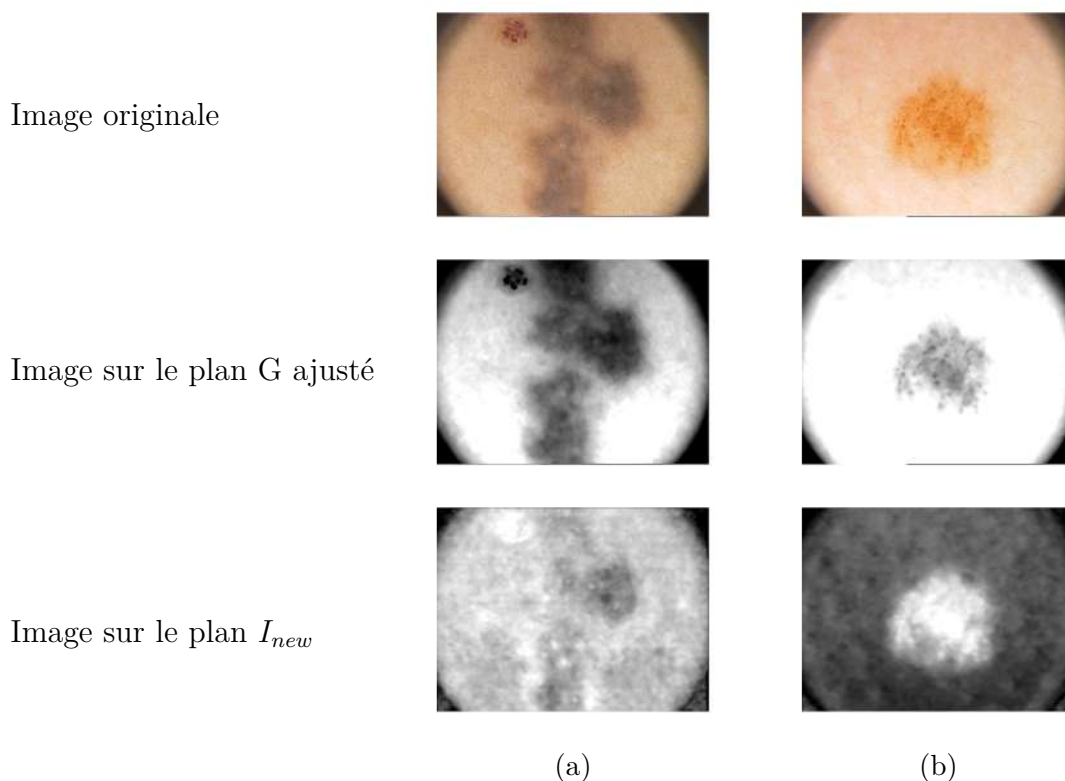


FIGURE 2.2 – Étape de prétraitement : Représentation de l'image originale sur les deux plans G ajusté et I_{new} en fonction de la valeur $\text{std}(G)$.

D'après la figure [2.2](#), nous constatons que les deux images originales (a) et (b) sont bien représentées dans deux plans différents. Pour cette raison, il était donc important d'avoir une combinaison de plans dans le but de mettre en évidence la lésion pour que le K-means segmente facilement et avec plus de précision les mélanomes.

2.2.3 Concept de base de la méthode K-means

Les méthodes du clustering impliquent le partitionnement d'un espace couleur en régions homogènes à l'aide d'algorithmes de regroupement non supervisés [\[56\]](#). Depuis le début des années 50 [\[57\]](#), le clustering a été abordé par de nombreux chercheurs et de nombreuses approches de clustering ont été explorées et étudiées. Le clustering K-means [\[28\]](#) est l'un des algorithmes d'apprentissage automatique non supervisé les plus simples et les plus populaires. La méthode K-means est considé-

rée comme l'une des méthodologies de base pour le regroupement et la classification de données dans K ensembles [58]. Elle fonctionne en divisant l'image en groupes de pixels qui ne se chevauchent pas, en fonction de leurs niveaux d'intensité. Le processus commence par la sélection des centroïdes des points de données, soit de manière aléatoire, soit selon un certain critère. Les pixels ou les points de données sont regroupés en fonction de leur distance minimale par rapport aux centroïdes sélectionnés. Après chaque itération, les valeurs moyennes des clusters formés sont trouvées et sont définies comme les centroïdes pour l'itération suivante. Le processus itératif se répète jusqu'à ce qu'il n'y ait plus de variation dans les centroïdes des clusters successifs ou en minimisant la fonction objective :

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \quad (2.3)$$

Où : $\|x_i^j - c_j\|^2$ est une mesure de distance choisie entre un point de données x_i^j et le centre de cluster c_j .

La figure 2.3 montre la segmentation des lésions à l'aide de K-means avec le nombre de classes $K=2$ afin de différencier les pixels de la lésion de ceux de la peau. La sortie de l'algorithme de clustering K-means génère un masque composée de pixels de la lésion et de pixels proches de ceux de la lésion, ces derniers doivent être filtrés dans l'étape du post-traitement.

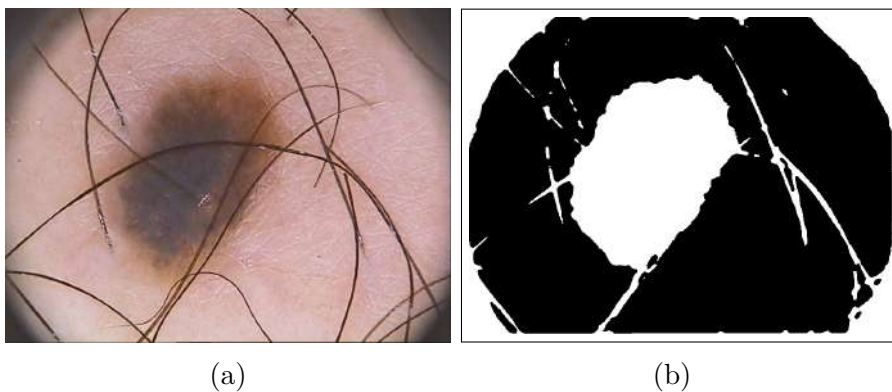


FIGURE 2.3 – Segmentation des lésions : (a) Image originale, (b) Segmentation avec K-means.

2.2.4 Post-traitement de l'image

Le post-traitement est essentiel pour filtrer tous les objets non-lésion détectés pour une segmentation précise des lésions et pour une extraction de caractéristiques efficace.

Nous avons appliqué un post-traitement à la segmentation effectuée. Ce dernier prend en considération les objets non lésionnels proviennent soit des poils, soit des coins sombres de l'image afin de les éliminer de la lésion. La suppression de ces objets consiste d'abord à détecter les coins sombres seuls, puis à les éliminer des masques générés par k-means.

Les coins (figure 2.4b) sont segmentés en utilisant les contours actifs [59, 24], qui sont basés sur le modèle d'évolution décrit dans l'équation 2.4.

Un contour initial C se déplace vers les bords et s'arrête par la fonction décroissante g comme exprimé dans l'équation 2.5. L'évolution de chaque point du contour $C(t)$ avec la vitesse V dans la direction normale est calculée comme le mouvement de la courbure dans la direction normale $K\vec{N}$. La fonction décroissante (équation 2.5) dépend du gradient de l'image.

Dans les contours, le gradient est élevé, ce qui fait que la fonction de décroissance tend vers zéro sur les bords. En multipliant la fonction décroissante par le mouvement de courbure, on arrête l'évolution des contours sur les bords. Et comme les contours de la lésion sont souvent peu nets, le deuxième terme de l'équation 2.4, projeté dans la direction normale, est une fonction d'arrêt supplémentaire.

$$\frac{\partial C(t)}{\partial t} = (g(I)k - \langle \nabla g(I), \vec{N} \rangle) \vec{N} \quad (2.4)$$

$$g(I) = \frac{1}{1 + (\nabla I)^2} \quad (2.5)$$

Avec,

C : Contour ;

∇I : Gradient du niveau de gris du plan I ;

k : Courbure ;

\vec{N} : Vecteur unitaire dans la direction normale ;

$\langle \nabla g(I), \vec{N} \rangle \vec{N}$: Le produit interne, projection du gradient de la fonction d'arrêt dans la direction normale.

La figure 2.4 montre les étapes de la détection des coins à l'aide de la méthode des contours actifs (GAC) ainsi que le masque des lésions après suppression de ces coins.

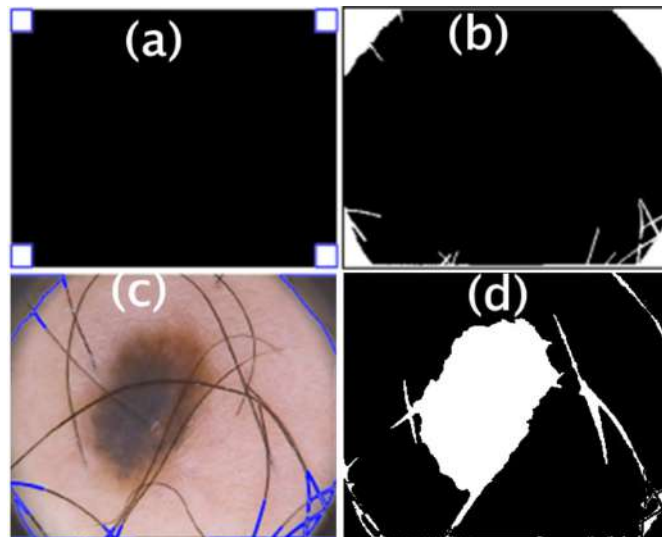


FIGURE 2.4 – Segmentation des coins par contours actifs : (a) initialisation du contour, (b) coins détectés, (c) superposition du contour sur l'image originale, (d) masque de lésion après suppression des coins.

Pour éliminer la zone de non-lésion, tous les objets dont la surface est inférieure à 700 pixels sont éliminés. Ensuite, pour éliminer les pixels isolés dans la région de la lésion par la segmentation, une étape supplémentaire d'ouverture morphologique est réalisée. Les objets sont érodés en utilisant l'opération morphologique d'érosion avec un élément structuré de disque de rayon égal à 10, suivie d'une dilatation en utilisant un élément structuré de disque de rayon égal à 15. Les étapes du filtrage sont présentées dans la figure ci-dessous (figure 2.5).

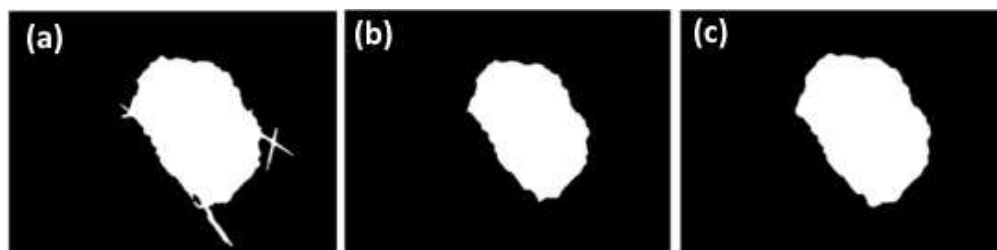


FIGURE 2.5 – Filtrage : (a) Masque de lésion avec objets supplémentaires, (b) Érosion, (c) Dilatation et masque final.

2.3 Évaluation de l'algorithme de segmentation

2.3.1 Base de données

Dans notre travail, nous avons utilisé l'ensemble de données PH2 qui est un ensemble public destiné à la recherche obtenue de l'hôpital Pedro Hispano [5, 60]. La base de données PH2 contient 200 images dermoscopiques d'une résolution de 768 x 560 pixels avec 80 nævus communs, 80 nævus atypiques et 40 mélanomes, et comprend également des masques manuels pour chaque lésion.

2.3.2 Mesures d'évaluation de l'algorithme

Les résultats de la segmentation sont évalués en utilisant plusieurs métriques : Sensibilité, spécificité, précision, indice de Jaccard et coefficient de Dice.

- **Sensibilité (SEN)** : c'est la proportion de positifs réels qui sont prédits positifs.

$$SEN = \frac{TP}{TP + FN} \quad (2.6)$$

- **Spécificité (SPE)** : il s'agit de la proportion de négatifs réels qui sont des négatifs prédits.

$$SPE = \frac{TN}{TN + FP} \quad (2.7)$$

- **Précision (ACC)** : elle est définie comme le pourcentage ou la proportion de prédictions correctes et peut être calculée en divisant le nombre de prédictions correctes par le nombre total de prédictions.

$$ACC = \frac{\text{Prédictions correctes}}{\text{Total des prédictions}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

- **Indice de Jaccard** : également connu sous le nom d'indice de similarité de Jaccard est une mesure de similarité pour les deux ensembles de données et peut être exprimé comme suit :

$$\text{Indice de Jaccard} = \frac{TP}{TP + FP + FN} \quad (2.9)$$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2.10)$$

Où : $J(A, B)$ est l'indice de Jaccard, avec : $0 < J(A, B) < 1$.

- **Coefficient de Dice (DC)** : l'indice de Jaccard est lié à l'indice de Dice selon :

$$DC = 2 \times J\left(\frac{1}{1 + J}\right) \quad (2.11)$$

$$DC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (2.12)$$

Où :

- **Vrais Positifs (TP)** : lorsque la prédiction d'une observation appartenant à une classe appartient effectivement à cette classe dans la réalité.
- **Vrais Négatifs (TN)** : lorsque la prédiction d'une observation n'appartenant pas à une classe n'appartient pas à cette classe en réalité.
- **Faux Positifs (FP)** : se produit lorsque la prédiction qu'une observation appartient à une classe alors qu'en réalité elle n'y appartient pas.
- **Faux Négatifs (FN)** : se produit lorsque la prédiction qu'une observation n'appartient pas à une classe alors qu'en réalité elle y appartient.

2.3.3 Résultats et discussion

La segmentation automatique génère un masque précis pour chaque lésion. La figure 2.6 montre quelques résultats de la méthode de segmentation proposée. Le contour manuel (ligne continue) et le contour automatique (ligne pointillée) sont superposés.

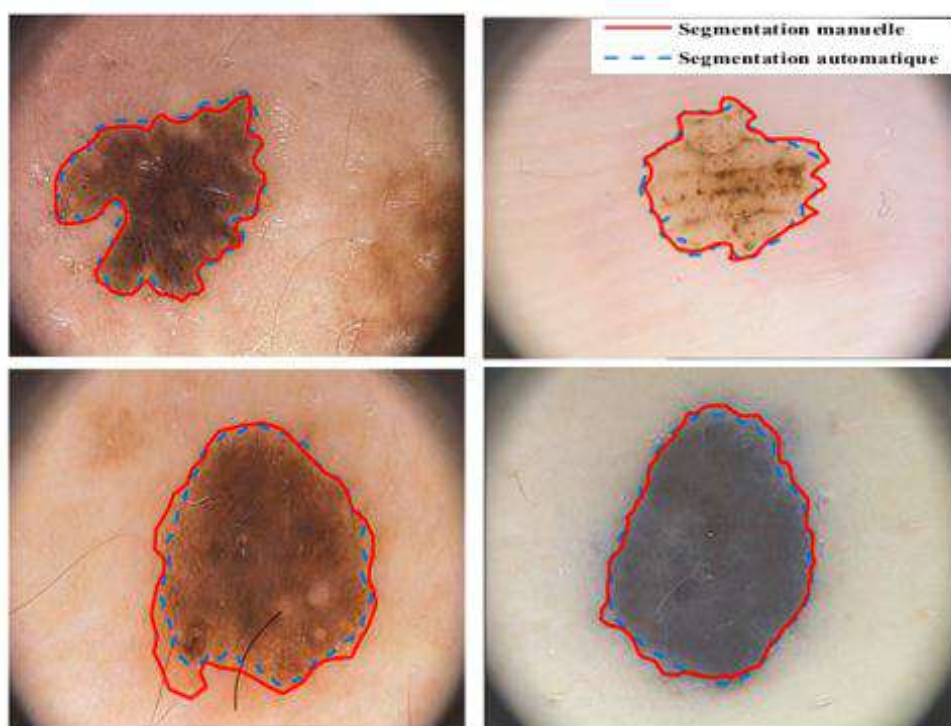


FIGURE 2.6 – Segmentation des lésions : Superposition des contours automatiques et manuels.

Les résultats de la segmentation automatique sont comparés aux résultats rapportés dans l'article publié de Jaisakthi et al. [36], où ils ont validé les résultats en utilisant la même base de données PH2. La table 2.1 montre la comparaison des résultats avec les cinq métriques : sensibilité (SEN), spécificité (SPE), précision (ACC), l'indice dice et Jaccard.

Métriques	Jaisakthi et al. [36]	Notre algorithme
ACC	0.9604	0.9606
SEN	0.8900	0.9236
SEP	0.9879	0.9870
Dice	0.9139	0.9167
Jaccard	0.8437	0.8462

TABLE 2.1 – Résultats et comparaison de la segmentation automatique proposée.

Les résultats de l'algorithme de segmentation proposé surpassent les résultats de Jaisakthi et al. [36] et présentent un meilleur équilibre entre sensibilité et spécificité. Il est important de noter qu'une sensibilité et une spécificité élevées indiquent la haute performance d'une méthode donnée [61].

Divers autres algorithmes basés sur l'approche K-means ont été développés et appliqués pour la segmentation des lésions cutanées dans le passé, en raison de sa simplicité et sa facilité à mettre en œuvre [35, 62]. Dans notre travail, une combinaison de plan de représentation d'image est utilisée pour mettre en évidence la lésion. Et cette combinaison permet de segmenter facilement la lésion et avec plus de précision, où $K = 2$ est le nombre de classes utilisé.

2.4 Conclusion

La segmentation joue un rôle essentiel pour la détection des mélanomes. Elle représente l'une des étapes les plus importantes de la chaîne de diagnostic assisté par ordinateur. Ce chapitre a été consacré à la présentation de l'approche proposée pour segmenter les lésions de l'ensemble de données PH2 en vue d'une classification, qui est présenté dans le chapitre suivant. L'algorithme de segmentation proposé à montrer son efficacité et sa précision à segmenter les lésions de la base PH2. Les résultats expérimentaux dépassent les résultats d'un travail publié en travaillant sur la même base de données.

L'algorithme de segmentation proposé génère un masque pour chaque lésion suffisamment proche de la segmentation manuelle, ils sont utilisés pour extraire les caractéristiques présentées dans le chapitre suivant.

Chapitre 3

Chapitre 3 Classification des lésions cutanées

3.1 Introduction

Afin de pouvoir procéder à la classification des lésions, nous avons proposé un algorithme basé sur les caractéristiques de couleur et de texture des lésions cutanées en introduisant de nouvelles techniques d'extraction de caractéristiques. Principalement, les attributs de couleur et de texture extraits mesurent la variation de la couleur et de la texture à l'intérieur de la lésion. Une fois les caractéristiques les plus pertinentes sont sélectionnées, nous avons procédé à la classification des lésions en cherchant une combinaison optimale plan d'espace couleur et classificateur.

Le présent chapitre reviendra en détail sur la mise en œuvre de la technique utilisée d'extraction des attributs afin de discriminer les lésions malignes des lésions bénignes.

3.2 Extraction des caractéristiques

Dans ce travail de recherche, trois caractéristiques différentes utilisant la caractéristique couleur, l'entropie et la matrice de Co-occurrence (GLCM), sont extraites de la lésion cutanée.

3.2.1 Les couleurs de références

Après inspection visuelle de toutes les lésions de l'ensemble de données PH2, nous concluons qu'il existe, subjectivement, six couleurs dominantes, que nous avons nommées : marron clair, marron, marron foncé, noir, blanc et bleu-gris, que nous devons quantifier.

Pour mesurer les moyennes des six couleurs dominantes dans les lésions, les images sont sur-segmentées en utilisant la technique des super-pixels pour obtenir 900 bulles, comme le montre la figure 3.1. La sur-segmentation super-pixel est basée sur un algorithme de regroupement itératif linéaire [63], attribuer la même valeur aux pixels ayant des valeurs proches dans un espace réduit qu'on appelle super-pixel (bulle), la figure 3.1 montre les contours des super-pixel de la lésion.

Super-pixel est couramment défini comme une région atomique perceptive obtenue par agrégation de pixels voisins sur la base de critères de similarité spa-

tiale et d'apparence. La mesure de similarité correspond à la quantification de la similarité entre deux super-pixels.

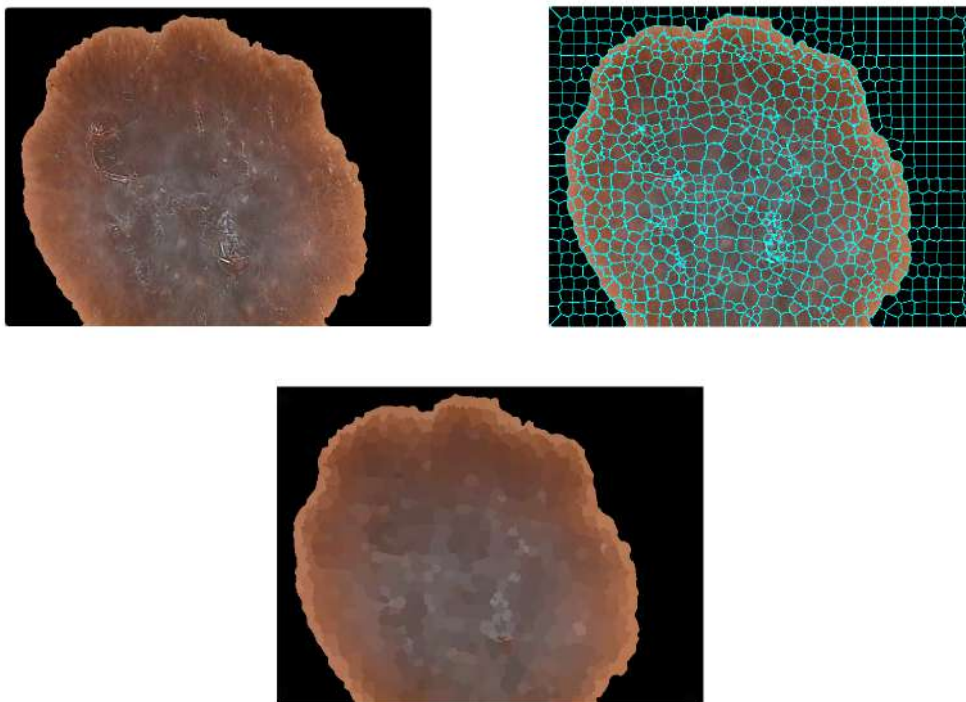


FIGURE 3.1 – Lésion sur-segmentée avec un algorithme de super-pixel.

À partir de tous les super-pixels qui apparaissent visuellement comme étant de la même couleur, nous avons extrait et enregistré les valeurs des couleurs R, G et B comme le montre la figure 3.2. Les couleurs de références, qui sont les valeurs qui représentent les couleurs dominantes (marron clair, marron, marron foncé, noir, blanc et bleu-gris), sont calculées comme les moyennes des valeurs extraites des super-pixels selon l'équation 3.1.

$$R_{ref_M} = \frac{1}{N} \sum_1^N R_i \quad (3.1)$$

Où :

R_{ref_M} : Référence du canal rouge pour la couleur marron (une procédure similaire est utilisée pour les canaux G et B.) ;

N : Nombre de bulles utilisées.



FIGURE 3.2 – Extraction des valeurs de couleurs R, G et B à partir des super-pixels.

L'opération est répétée pour 100 images. Même procédure est effectuée pour les six couleurs dominantes. Les couleurs de références mesurées sont présentées dans le tableau 3.1







Couleurs	Apparence	Les couleurs de référence [R_ref, G_ref, B_ref]
Marron clair		[200, 155, 130]
Marron		[160, 100, 67]
Marron foncé		[126, 67, 48]
Blanc		[230, 230, 230]
Noir		[31, 26, 26]
Bleu-gris		[75, 112, 137]

TABLE 3.1 – Les références des six couleurs dominantes sur les lésions.

3.2 EXTRACTION DES CARACTÉRISTIQUES

Une fois que nous avons eu les références des six couleurs fréquentes, nous avons mesuré les distances couleurs de chaque super-pixel d'une lésion donnée, par rapport aux six références couleurs mentionnées dans le tableau 3.1 en utilisant la distance euclidienne (équation 3.2).

$$Dist(i) = \sqrt{(R_i - R_{ref})^2 + (G_i - G_{ref})^2 + (B_i - B_{ref})^2} \quad (3.2)$$

Où :

Dist(i) : Distance d'un pixel donné i à l'une des couleurs de référence (tableau 3.1) ;

R_i, G_i, B_i : Canal rouge, vert et bleu d'un pixel i.

Un super-pixel est attribué à une couleur donnée si sa distance normalisée est inférieure à un seuil empirique T, variable pour chaque couleur. Les seuils sont indiqués dans le tableau 3.2.

Couleurs	Marron clair	Marron	Marron foncé	Blanc	Noir	Bleu-gris
Seuils	0.12	0.25	0.2	0.25	0.25	0.5

TABLE 3.2 – Seuils des couleurs pour attribuer une couleur super-pixel donnée.

Une couleur donnée est considérée comme présente uniquement si elle occupe une surface supérieure à 5% de la surface de la lésion.

La figure 3.3 montre l'image originale et les emplacements des couleurs détectées.

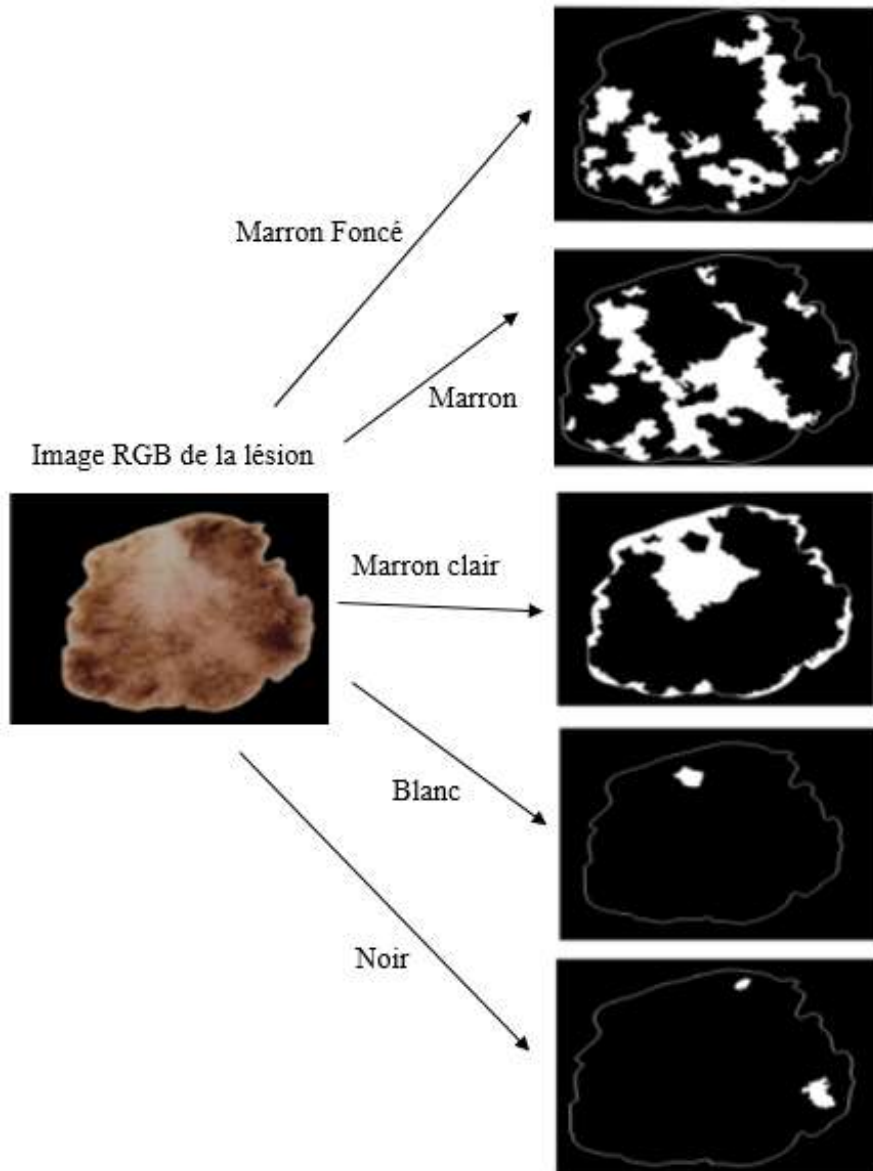


FIGURE 3.3 – Segmentation de lésion selon les couleurs dominantes.

Le nombre de couleurs détectées dans une lésion donnée est utilisé comme le nombre de cluster K du K -means. La lésion est segmentée en K classes.

3.2 EXTRACTION DES CARACTÉRISTIQUES

La figure 3.4 montre deux exemples où la lésion est segmentée en K classes, le nombre de couleurs dominantes, trouvées par K-means.

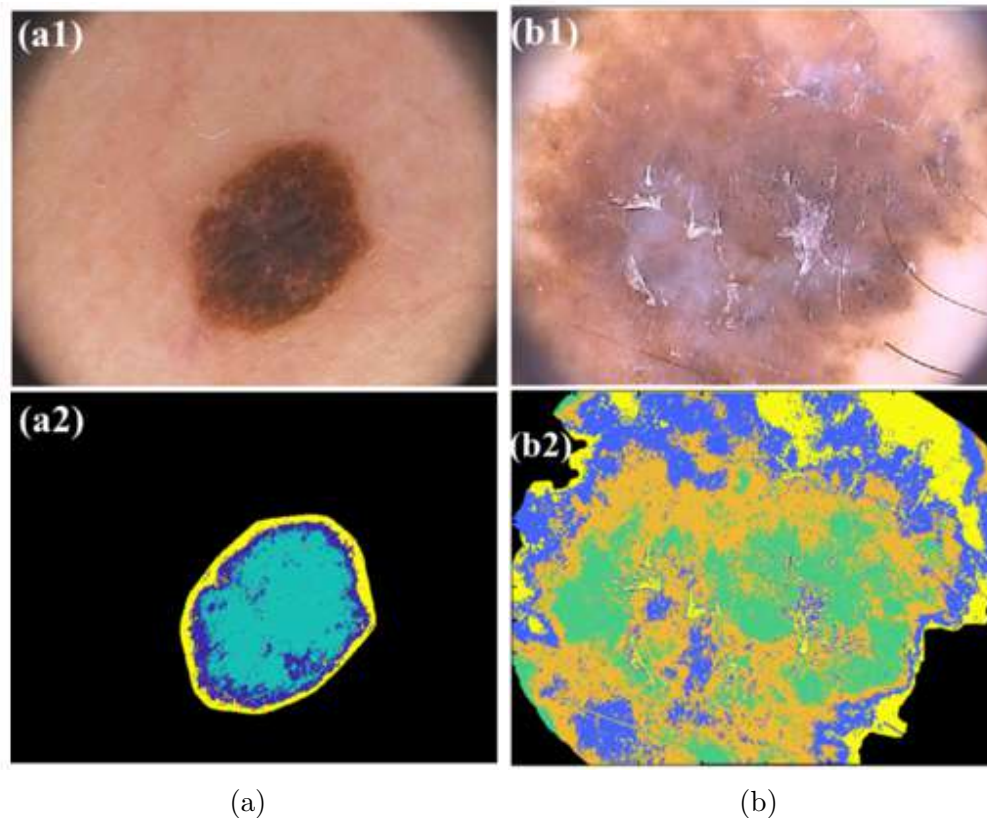


FIGURE 3.4 – Segmentation K-means : a) Une lésion avec trois classes b) Une lésion avec quatre classes.

Dans ce travail, nous présentons les équations des caractéristiques que nous avons extraites des plans R, G et B de l'espace couleur RGB. Les mêmes caractéristiques sont extraites des plans H, S, V de l'espace couleur HSV, des plans L, a, b de l'espace couleur Lab, des plans X, Y, Z de l'espace couleur XYZ, des plans Y, Cb, Cr de l'espace couleur YCbCr. La lésion est segmentée en K classes (nombre de couleurs dominantes trouvées pour chaque lésion) avec k-means.

3.2.2 Extraction de la caractéristique couleur

La présence de lésions multicolores et non homogènes est le signe le plus important de malignité [64, 65]. Les travaux de [66, 67, 68] montrent que la couleur est une caractéristique pertinente et efficace pour différencier les mélanomes malins et bénins, qui semblent très similaires au stade initial.

Les équations 3.3 à 3.27 utilisent les K classes trouvées par k-means pour mesurer, principalement, le degré d'inhomogénéité à l'intérieur de la lésion.

a) La moyenne

La moyenne des plans rouge, vert et bleu (avgR, avgG, avgB) est calculée pour chaque classe n (équation 3.3) :

$$avgR(n) = \frac{1}{I \times J} \sum_{i=1}^I \sum_{j=1}^J R(i_n, j_n) \quad (3.3)$$

Avec :

avgR(n) : Moyenne de la classe n (n = 1 à K) pour le plan R ;

R : Plan rouge ;

(i_n, j_n) : Coordonnées des pixels de la classe n.

b) Couleurs relatives

Les rapports des couleurs relatives C₁, C₂ et C₃ en considérant la moyenne maximale des K classes (équations 3.4 - 3.6).

$$C_1 = \max_{n=1...K} \left(\frac{avgR(n)}{R_{peau}} \right) \quad (3.4)$$

$$C_2 = \max_{n=1...K} \left(\frac{avgG(n)}{G_{peau}} \right) \quad (3.5)$$

$$C_3 = \max_{n=1...K} \left(\frac{avgB(n)}{B_{peau}} \right) \quad (3.6)$$

c) Couleurs relatives normalisées [69]

Les rapports relatifs normalisés des plans R, G et B sont C₄, C₅ et C₆ en considérant les classes avec la couleur moyenne maximale.

$$C_4 = \frac{C_1}{C_1 + C_2 + C_3} \quad (3.7)$$

$$C_5 = \frac{C_2}{C_1 + C_2 + C_3} \quad (3.8)$$

$$C_6 = \frac{C_3}{C_1 + C_2 + C_3} \quad (3.9)$$

d) Les différences relatives [69]

Les différences relatives des plans R, G et B sont C_7 , C_8 et C_9 respectivement en considérant les classes de couleur moyenne maximale.

$$C_7 = C_1 - R_{peau} \quad (3.10)$$

$$C_8 = C_2 - G_{peau} \quad (3.11)$$

$$C_9 = C_3 - B_{peau} \quad (3.12)$$

Où : R_{peau} , G_{peau} et B_{peau} sont les moyennes des plans de couleur rouge, vert et bleu respectivement de la peau entourant la lésion.

e) Différences relatives normalisées [69]

Les différences relatives normalisées des plans R, G et B sont C_{10} , C_{11} et C_{12} respectivement en sélectionnant les classes avec la couleur moyenne maximale.

$$C_{10} = \frac{C_7}{(C_7 + C_8 + C_9)} \quad (3.13)$$

$$C_{11} = \frac{C_8}{(C_7 + C_8 + C_9)} \quad (3.14)$$

$$C_{12} = \frac{C_9}{(C_7 + C_8 + C_9)} \quad (3.15)$$

f) Variation de luminance entre classes

Nous calculons les distances euclidiennes entre les classes pour chaque canal : C_{13} , C_{14} et C_{15} . La luminance varie à l'intérieur de la lésion (entre les classes trouvées par k-means). Si la distance euclidienne est grande, cela signifie qu'il y a une grande variation de luminance à l'intérieur de la lésion. Sachant que

la lésion du mélanome n'est pas uniforme, plus la distance entre les classes est grande, plus la probabilité que la lésion soit un mélanome est grande. La variation de luminance est mesurée pour les trois plans (Rouge : équation 3.16. Vert : équation 3.17. Bleu : équation 3.18).

$$C_{13} = \sqrt{\sum_{i=n}^{K-1} (avgR(n) - avgR(n+1))^2} \quad (3.16)$$

$$C_{14} = \sqrt{\sum_{i=n}^{K-1} (avgG(n) - avgG(n+1))^2} \quad (3.17)$$

$$C_{15} = \sqrt{\sum_{i=n}^{K-1} (avgB(n) - avgB(n+1))^2} \quad (3.18)$$

g) La variance

Plus la variance est grande plus la lésion est probablement un mélanome. C_{16} , C_{17} et C_{18} sont les variances entre les classes dans les différents canaux de couleur.

$$C_{16} = \frac{\sum_{n=1}^K (avgR(n) - \overline{avgR})^2}{K} \quad (3.19)$$

$$C_{17} = \frac{\sum_{n=1}^K (avgG(n) - \overline{avgG})^2}{K} \quad (3.20)$$

$$C_{18} = \frac{\sum_{n=1}^K (avgB(n) - \overline{avgB})^2}{K} \quad (3.21)$$

h) L'écart-type

Plus l'écart-type entre les classes est grand, plus il est probable qu'il s'agisse d'un mélanome. C_{19} , C_{20} et C_{21} sont les écart-types entre les classes dans les différents canaux de couleur. Ceci est fait pour les trois plans (Rouge : équation 3.22. Vert : équation 3.23. Bleu : équation 3.24).

$$C_{19} = \sqrt{C_{16}} \quad (3.22)$$

$$C_{20} = \sqrt{C_{17}} \quad (3.23)$$

$$C_{21} = \sqrt{C_{18}} \quad (3.24)$$

i) La variance de la couleur normalisée

Les équations C_{22} , C_{23} et C_{24} mesurent les variances de la couleur normalisée entre les classes, c'est à dire les différences entre les couleurs (R, G, B) de toutes les classes et les couleurs (R, G, B) de la peau (différences relatives des classes K).

$$C_{22} = \frac{\sum_{n=1}^K \left(avgRD(n) - \overline{avgRD} \right)^2}{K} \quad (3.25)$$

$$C_{23} = \frac{\sum_{n=1}^K \left(avgGD(n) - \overline{avgGD} \right)^2}{K} \quad (3.26)$$

$$C_{24} = \frac{\sum_{n=1}^K \left(avgBD(n) - \overline{avgBD} \right)^2}{K} \quad (3.27)$$

Avec :

Équations [3.28](#), [3.29](#) et [3.30](#) représentent la variance des différences relatives des classes.

$$avgRD(n) = |avgR(n) - R_{peau}| \quad (3.28)$$

$$avgGD(n) = |avgG(n) - G_{peau}| \quad (3.29)$$

$$avgBD(n) = |avgB(n) - B_{peau}| \quad (3.30)$$

Où :

$avgR(n)$, $avgG(n)$, $avgB(n)$: Moyennes de la classe n dans les plans rouge, vert et bleu respectivement ;

n : Nième classe (n=1 à K) ;

R_{peau} , G_{peau} et B_{peau} : Couleur de la peau qui entoure la lésion dans les plans rouge, vert et bleu.

3.2.3 Extraction de la caractéristique texture

La texture de l'image représente la distribution spatiale des niveaux d'intensité des pixels dans une image. Les caractéristiques texturales sont fréquemment utilisées pour identifier les lésions malignes car elles permettent de mesurer l'irrégularité des lésions cutanées [70].

L'entropie, la matrice de cooccurrence sont les principales mesures utilisées dans ce travail.

a) Distance d'entropie

Pour mesurer l'inhomogénéité de la lésion, nous avons mesuré les distances euclidiennes entre les entropies des différentes classes de la même lésion, pour chaque canal, R, G, et B notés respectivement T_1 , T_2 et T_3 . Plus la distance d'entropie entre les classes de la même lésion est grande, plus la probabilité que la lésion soit un mélanome est grande.

$$T_1 = \sqrt{\sum_{i=n}^{K-1} (EntropieR(n) - EntropieG(n+1))^2} \quad (3.31)$$

$$T_2 = \sqrt{\sum_{i=n}^{K-1} (EntropieG(n) - EntropieG(n+1))^2} \quad (3.32)$$

$$T_3 = \sqrt{\sum_{i=n}^{K-1} (EntropieB(n) - EntropieB(n+1))^2} \quad (3.33)$$

Où :

EntropieR(n), EntropieG(n), EntropieB(n) : Entropie de la nⁱème classe dans le plan rouge, vert et bleu respectivement. Les trois caractéristiques sont mesurées à partir du premier plan pour chaque espace couleur (R, H, X et Y).

b) Matrice de co-occurrence des niveaux de gris (GLCM)

La Matrice de cooccurrence des niveaux de gris (GLCM) est une technique d'extraction de caractéristiques texturales globales qui mesure la distribution statistique des intensités en combinaison à des positions spécifiques dans l'image. La Matrice de cooccurrence est une matrice carrée dont les dimensions correspondent au nombre de niveaux d'intensité G. L'élément de la

3.2 EXTRACTION DES CARACTÉRISTIQUES

matrice $P(i,j)$, spécifié par la position de la ligne et de la colonne avec i et j , donne la fréquence à laquelle les valeurs de pixel i et j sont apparues dans des directions globales spécifiées par l'angle θ , comme le montre la figure 3.5. Les composantes extraites en utilisant la GLCM sont le contraste, la corrélation, l'homogénéité et l'énergie [71].

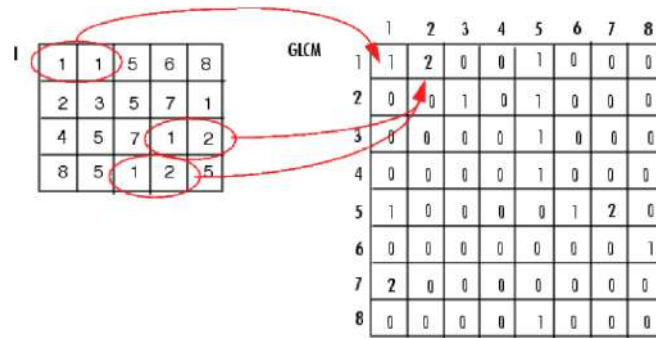


FIGURE 3.5 – Formation d'une matrice GLCM à partir de l'image d'entrée : représentée comme une matrice de niveaux d'intensité.

Nous avons utilisé la Matrice de cooccurrence en fixant l'intervalle $[0 \ 1]$ entre les pixels, pour mesurer trois propriétés : Contraste, Homogénéité, et Énergie. Ces trois propriétés sont mesurées à partir du premier plan pour chaque espace couleur (R, H, X et Y).

$$T_4 = \text{Contraste} = \sum_i^M \sum_j^M N |i - j|^2 G(i, j) \quad (3.34)$$

$$T_5 = \text{Homogénéité} = \sum_i^M \sum_j^M N \frac{G(i, j)}{1 + |i - j|} \quad (3.35)$$

$$T_6 = \text{Énergie} = \sum_i^M \sum_j^M N G(i, j)^2 \quad (3.36)$$

Avec :

G : Matrice de co-occurrence du plan H (teinte) ;

M, N : Dimension de la matrice de co-occurrence.

3.3 Sélection des caractéristiques

L'étape de sélection des caractéristiques est utilisée afin de sélectionner les caractéristiques les plus pertinentes et de réduire la dimension de l'espace des caractéristiques de manière à éliminer les caractéristiques non pertinentes et/ou redondantes. De plus, ces caractéristiques peuvent influencer la performance du processus de classification, c'est-à-dire le rendre plus lent [72].

Plusieurs avantages sont associés à l'application de schémas de sélection de caractéristiques, en réduisant le temps d'extraction des caractéristiques, diminuant la complexité de la classification, améliorant le taux de précision de la classification, diminuant le temps de formation et de test, et en simplifiant la compréhension et la visualisation des données.

Pour cela, avant l'étape de classification, nous avons sélectionné les caractéristiques les plus discriminantes par la méthode Relief [73] parmi toutes les caractéristiques de couleur et de texture extraites. L'algorithme Relief attribue un score à chaque caractéristique en fonction de son importance.

3.3.1 Algorithme Relief

Le concept de Relief [73, 74] présenté dans l'algorithme 1 est d'estimer la qualité des caractéristiques d'après la capacité de leurs valeurs à distinguer entre les instances qui sont proches les unes des autres. Pour cela, étant donnée une instance R_i sélectionnée aléatoirement, Relief recherche pour cette instance deux plus proches voisins, une de la même classe que R_i appelé hit H, est l'autre de la classe différente appelé miss M. Puis, met à jour l'estimation de la qualité $W[A]$ de chaque attribut A selon les valeurs de R_i , M et H.

Si les instances R_i et H ont des valeurs différentes pour l'attribut A, alors l'attribut A sépare entre deux instances de la même classe ce qui est indésirable, donc on diminue l'estimation de la qualité $W[A]$.

Si les instances R_i et M ont des valeurs différentes pour l'attribut A, alors l'attribut A sépare entre deux instances de classes différentes, ce qui est désirable, donc on augmente l'estimation de la qualité $W[A]$.

Le processus est répété m fois tel que m est un paramètre prédéfini par l'utilisateur. La dernière étape du cycle, met à jour le poids d'une caractéristique A dans

3.3 SÉLECTION DES CARACTÉRISTIQUES

W si la valeur de la caractéristique diffère entre l'instance cible et l'instance la plus proche H ou la plus proche M (voir Figure 3.6). Les caractéristiques qui ont une valeur différente entre R_i et M soutiennent l'hypothèse qu'elles sont pertinentes du résultat, l'estimation de la qualité $W[A]$ est donc augmentée. À l'inverse, les caractéristiques présentant des différences entre R_i et H fournissent des preuves du contraire, de sorte que l'estimation de la qualité $W[A]$ est diminuée [75].

Algorithm 1 Algorithme de sélection Relief

Entrée : pour chaque instance d'apprentissage, un vecteur de valeurs de caractéristiques et sa valeur de classe.

Sortie : le vecteur W des scores des caractéristiques qui estiment la qualité (poids) des caractéristiques.

Initialiser tous les poids des caractéristiques $W[A]=0$

pour $i = 1$ **jusqu'à** m **faire**

Sélectionner aléatoirement une instance R_i

Trouver le plus proche voisin de la même classe que R : H

Trouver le plus proche voisin de la classe différente R : M

pour $A = 1$ **jusqu'à** a **faire**

$W[A] = W[A] - \text{diff}(A, R_i, H)/m + \text{diff}(A, R_i, M)/m$

fin pour

fin pour

Retourner W

La fonction diff de l'algorithme 1 calcule la différence de valeur de la caractéristique A entre deux instances et, où $I_1 = R_i$ et I_2 est soit H soit M, lors de la mise à jour des poids. Pour les caractéristiques discrètes (par exemple catégoriques ou nominales), diff est défini comme suit :

$$\text{diff}(A, I_1, I_2) = \begin{cases} 0 & \text{si } \text{valeur}(A, I_1) = \text{valeur}(A, I_2) \\ 1 & \text{sinon} \end{cases} \quad (3.37)$$

Et pour les caractéristiques continues (par exemple, ordinales ou numériques), diff est défini comme suit :

$$\text{diff}(A, I_1, I_2) = \frac{|\text{valeur}(A, I_1) - \text{valeur}(A, I_2)|}{\max(A) - \min(A)} \quad (3.38)$$

Les valeurs maximale et minimale de A sont déterminées sur l'ensemble des instances. Cette normalisation garantit que les mises à jour des poids se situent

3.4 CLASSIFICATION

entre 0 et 1 pour les caractéristiques discrètes et continues. De plus, lors de la mise à jour de $W[A]$, la division de la sortie de diff par m garantit que tous les poids finaux seront normalisés dans l'intervalle $[-1, 1]$.

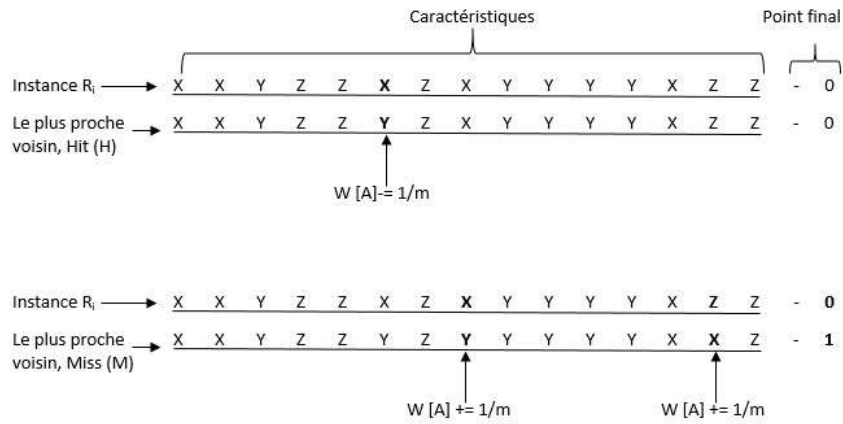


FIGURE 3.6 – Mise à jour du relief pour une instance cible donnée lorsqu'elle est comparée à son plus proche voisin H et M .

Dans cet exemple, les caractéristiques sont discrètes avec des valeurs possibles de X, Y ou Z, et le point final est binaire avec une valeur de 0 ou 1. Remarquez que lorsque la valeur d'une caractéristique est différente, le poids de la caractéristique correspondante augmente de $1/m$ pour le voisin le plus proche M et diminue de $1/m$ pour le voisin le plus proche H .

3.4 Classification

Après la segmentation automatique, l'extraction et la sélection des caractéristiques ainsi que trois classificateurs KNN, SVM et NN sont utilisés pour identifier le mélanome.

- **K-NN [76]** : l'algorithme des k-plus proches voisins est un algorithme de classification supervisé, qui se base sur l'ensemble de données pour effectuer une prédiction. En effet, pour une observation, qui ne fait pas parti de l'ensemble de données, qu'on souhaite prédire, l'algorithme va chercher les K instances de l'ensemble de données les plus proches de notre observation.

Ensuite pour ces K voisins, l'algorithme se basera sur leurs variables de sortie y pour calculer la valeur de la variable y de l'observation qu'on souhaite prédire. Par ailleurs, le mode des variables y des K plus proches observations est utilisé pour la prédiction.

De plus, K -NN calcule la distance entre deux observations. Plus deux points sont proches l'un de l'autre, plus ils sont similaires. Il existe plusieurs fonctions de calcul de distance, notamment, la distance euclidienne, la distance de Manhattan, la distance de Minkowski, celle de Jaccard, la distance de Hamming ...etc. La fonction de distance est choisie selon le type de données manipulées.

La classification dépend du nombre de voisins choisis. La figure 3.7 représente un ensemble de données d'entraînement avec deux classes, rouge et bleu. L'entrée est donc bidimensionnelle dans ce cas, la sortie est la couleur à classer et le point blanc est une nouvelle entrée.

Si le point blanc est une nouvelle entrée dont on veut prédire la classe, on va simplement regarder les k voisins les plus proches de ce point et regarder quelle classe constitue la majorité de ces points, afin d'en déduire la classe du nouveau point. Dans cet exemple, si on utilise le 5-NN et la distance euclidienne comme mesure de similarité, on peut prédire que la nouvelle donnée appartient à la classe rouge puisqu'elle a 3 rouges et 2 bleus dans son entourage.

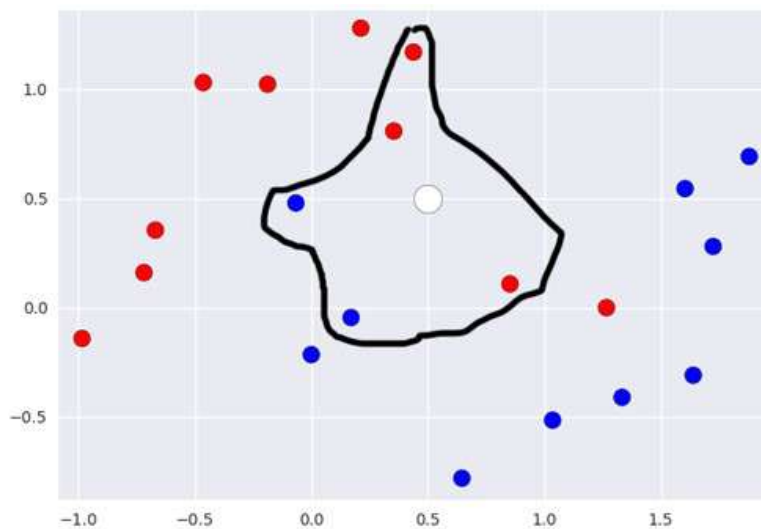


FIGURE 3.7 – Classification avec K-NN

Pour une classification binaire, la valeur de K doit être fixée à un nombre

impair pour éliminer la possibilité d'un tirage au sort entre les décisions. Les points de données d'apprentissage étant tracés dans l'espace des caractéristiques, les échantillons de test voisins sont classés comme mélanome ou bénin en fonction de la distance minimale par rapport à la majorité des K échantillons.

La méthode que nous proposons utilise le KNN avec un nombre de voisins égal à 3 et la distance euclidienne est utilisée [3.39](#).

$$\text{Distance Euclidienne : } d(X_c, X_r) = \sqrt{\sum_{i=1}^N (X_{ci} - X_{ri})^2} \quad (3.39)$$

La lésion qui sera classée comme bénigne ou maligne, sera classée sur la base du vote le plus important qu'il reçoit de son voisin le plus proche.

- **SVM** [77](#) : il s'agit d'un classificateur sélectif qui découle des premiers travaux théoriques de Vapnik et Chervonenkis en 1995, appliqués à partir de 2000. SVM est un algorithme d'apprentissage supervisé, et qui est spécialisé dans la résolution de problèmes de discrimination et de régression mathématiques. les SVMs ont pour but de séparer les données en classes à l'aide d'un hyperplan, de telle façon que la distance entre les différentes classes de données et l'hyperplan qui les sépare soit maximale. Cette distance est aussi appelée **marge** et les SVMs sont ainsi qualifiés de **séparateurs à vaste marge**, les **vecteurs de support** étant les données les plus proches de l'hyperplan. Un hyperplan est un plan qui isole des objets ayant des appartenances à des catégories totalement différentes.

Dans cette thèse, le nombre de catégories est de deux, les catégories bénigne et maligne se situent de part et d'autre de l'hyperplan. L'hyperplan peut être caractérisé comme suit :

$$w \cdot x + b = 0 \quad (3.40)$$

Où :

x : désigne le vecteur d'entrée à N dimensions ;

w : est le poids du vecteur défini par $w = w_1, w_2, w_3, \dots, w_n$;

b : est la fonction de biais du modèle.

La figure 3.8 montre l'hyperplan en pointillés et les classes binaires +1 et -1.

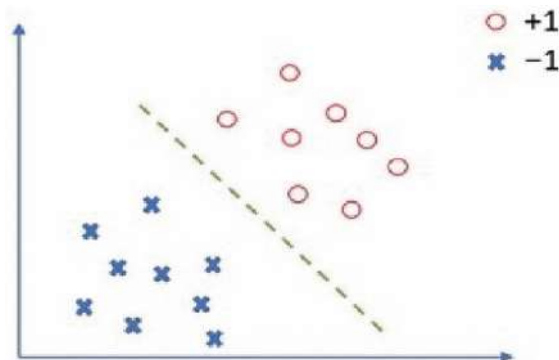


FIGURE 3.8 – Modèle de classificateur SVM.

Pour que le SVM puisse trouver l'hyperplan, il est nécessaire de lui donner des données d'entraînement. En l'occurrence, on donne au SVM un ensemble de points, dont on sait déjà si ce sont des ronds rouges ou des croix bleus, comme dans la Figure 3.8. A partir de ces données, le SVM va estimer l'emplacement le plus plausible de l'hyperplan.

- **ANN [78]** : inspirés des neurones du cerveau, les réseaux de neurones artificiels sont des fonctions mathématiques à plusieurs paramètres, ajustables. Le concept de réseau de neurones artificiels repose sur trois étapes principales : tout a d'abord, pour chaque neurone dans une couche, multiplier la valeur d'entrée par le poids. Ensuite, pour chaque couche, additionner toutes les pondérations des neurones et ajouter un biais. Enfin, appliquer la fonction d'activation sur cette valeur pour calculer une nouvelle sortie. La figure 3.9 montre l'architecture d'un neurone artificiel.

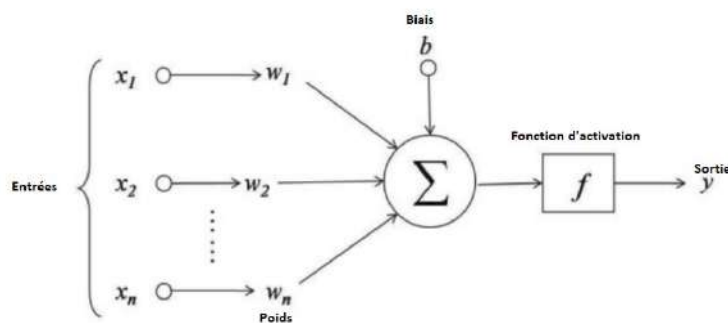


FIGURE 3.9 – Architecture d'un neurone artificiel.

Avec :

$$y = \sum (\text{poids} * \text{entrée}) + \text{biais};$$

x_1, x_2, \dots, x_n sont des variables d'entrée ;

w_1, w_2, \dots, w_n sont les poids des entrées respectives ;

b est le biais.

Chaque neurone effectue ainsi un calcul, et c'est la succession des couches de neurones qui permet d'obtenir des réseaux complexes. Durant la phase d'entraînement, le réseau va ajuster automatiquement les paramètres de chaque neurone, c'est-à-dire les valeurs des poids et du biais afin de minimiser l'erreur moyenne calculée sur l'ensemble des exemples entre la sortie attendue et celle observée (figure 3.10).

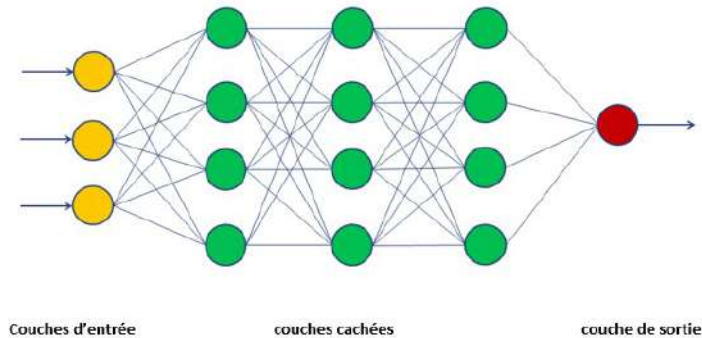


FIGURE 3.10 – Architecture d'un réseau neuronal artificiel.

3.5 Résultats et discussions

En vue de trouver la configuration optimale entre les classificateurs et les espaces couleur, nous avons extrait les mêmes caractéristiques des images dans cinq espaces couleur, RGB, HSV, Lab, XYZ et YCbCr. Ensuite, nous avons utilisé trois classificateurs : KNN, ANN et SVM pour chaque espace couleur. Les performances de l'algorithme proposé sont évaluées à l'aide de trois métriques, la sensibilité (SEN), la spécificité (SPE) et la précision (ACC), (équations 2.6 - 2.8).

L'évaluation est effectuée sur la base de données PH2 qui consiste à discriminer les mélanomes des lésions bénignes (naevus atypique et naevus commun). En mesurant la moyenne de 10 validations croisées pour chaque configuration (espace couleur/classificateur), les résultats sont reportés dans le tableau 3.3. Les

caractéristiques sont triées en fonction de leur pertinence.

Espaces couleurs	Caractéristiques	KNN			ANN			SVM		
		SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)
RGB	$T_1, T_3, T_2, T_4,$ T_5, C_{13}, C_{11}, C_2	99.23	98.26	98.46	98.25	99.45	99.2	99.25	99.58	99.51
HSV	$T_1, T_2, T_3, C_{22},$ $C_{13}, C_{16}, C_{19}, T_4$	95.73	99.11	98.44	92.25	99.45	98	94.92	99.55	98.62
LAB	$T_{12}, T_4, T_5, T_{16},$ C_4, C_7	94	98.81	97.85	95.25	98.02	97.45	95.71	96.19	96.09
XYZ	$C_{15}, T_4, C_{21}, C_{14},$ C_{18}, T_5, T_3, C_{21}	91.66	95.13	94.44	84	97.02	94.4	90.83	96.87	95.66
YCbCr	$T_2, T_3, T_4, T_1, T_5,$ $T_6, C_{15}, C_{21}, C_{13}$	99.21	98.81	98.98	98.5	99.82	99.55	98.8	99.2	99.12

TABLE 3.3 – Classification des mélanomes et des lésions bénignes à l’aide de différents classificateurs et espaces couleurs sur l’ensemble de données PH2.

Pour les deux classificateurs (KNN et ANN) le meilleur résultat est trouvé sur l’espace couleur YCbCr avec neuf attributs : six attributs de texture ($T_2, T_3, T_4, T_1, T_5, T_6$) et 3 attributs de couleur (C_{15}, C_{21}, C_{13}). Le KNN utilise un échantillon pour le test et les échantillons restants pour l’apprentissage, il atteint 99,21% de sensibilité, 98,81% de spécificité et 98,98% de précision. L’ANN utilise 60% de la base de données pour l’apprentissage et 40% pour le test, atteint 99,5% de sensibilité, 99,45% de spécificité et 99,55% de précision.

Avec le SVM, le meilleur résultat est trouvé sur l’espace couleur RGB avec huit attributs : cinq attributs de texture (T_1, T_3, T_2, T_4, T_5) et 3 attributs de couleur (C_{13}, C_{11}, C_2). En utilisant 60% de la base PH2 pour l’apprentissage et 40% pour le test, le résultat de la classification atteint 99,25% de sensibilité, 99,58% de spécificité et 99,51% de précision.

Les trois classificateurs donnent de bons résultats en utilisant une fusion de caractéristiques (couleur/texture). Nous constatons à part pour le XYZ, les caractéristiques de texture sont sélectionnées comme étant les plus pertinentes pour RGB, HSV, Lab et YCbCr.

Le tableau 3.4 montre une comparaison des résultats trouvés par notre algorithme et trois résultats publiés qui utilisent la même base de données PH2. Globalement, notre système surpasse toutes les autres méthodes.

Méthode et l'année de publication	SEN (%)	SPE (%)	ACC (%)
Barata [47], 2014	93	88	-
Alfed [48], 2017	99.4	98.18	98.79
Majumder [49], 2018	95	98.8	98
Notre Méthode	99.25	99.58	99.51

TABLE 3.4 – Comparaison des résultats de classification de notre méthode avec trois méthodes de la littérature.

La méthode d'Alfed [48] a donné de bons résultats en atteignant une sensibilité de 99,4% (notre : 99,25%) et une précision de 98,79% (notre : 99,51%). Notre algorithme présente le meilleur équilibre entre sensibilité et spécificité.

Plusieurs articles montrent que la présence de plusieurs couleurs dans une lésion constitue un important bio-marqueur de malignité [66, 67, 68, 79, 80]. Dans notre travail, la lésion est divisée en K classes, où K est le nombre de couleurs constituant la lésion. Et ces répartitions sont utilisées pour extraire des caractéristiques. Les caractéristiques les plus sélectionnées comme pertinentes, que ce soit la texture ou la couleur, mesurent le désordre ou la variation entre les classes. Ceci est cohérent avec la croissance incontrôlée des mélanomes.

Quelques chercheurs ont adopté une approche par palette de couleurs et ont donc formé un classificateur pour reconnaître les couleurs dans les lésions [81, 82]. Notre méthode est basée sur la détection de nombre de couleurs dominantes dans la lésion en utilisant la segmentation super-pixel. Et la présence de plusieurs couleurs dans une lésion représente probablement une lésion maligne.

La différence entre la valeur du pixel de la lésion et la valeur moyenne des pixels de la peau environnante donne une mesure de la différence de couleur relative, et est donc utilisée comme une caractéristique de couleur [83]. Nous avons exploité cette caractéristique avec une autre manière en divisant la lésion en sous régions en utilisant K -means, K est le nombre de couleurs dominantes détectées.

La différence de couleur relative entre les périphéries intérieure et extérieure de la lésion révèle des informations importantes sur la lésion. Celebi et al. [84]

ont obtenu un ensemble de 108 caractéristiques en calculant la moyenne et l'écart-type pour 6 espaces de couleur afin de déterminer la différence de couleur relative entre la périphérie interne, la périphérie externe et la lésion. Pour notre cas, cette caractéristique est calculée en considérant les classes de couleur moyenne maximale des plans R, G et B.

L'extraction des caractéristiques géométriques [45, 85], de texture [86, 87, 88], de couleurs [68] et d'une combinaison de celles-ci [89, 90] ont été utilisées pour identifier et classifier les cancers de la peau. Cependant, l'identification des caractéristiques les plus discriminantes pour identifier le mélanome à son stade initial reste une tâche difficile [91]. Pour cela, l'extraction des caractéristiques couleur et texture utilisées dans ce travail améliorent d'avantage l'identification et la classification du mélanome à l'état précoce.

Afin de pouvoir confirmer la sensibilité et l'efficacité de l'algorithme proposé ainsi que de voir ses limites, nous l'avons testé sur une autre base de données public appelée DermIS & DermQuest [92, 93]. Cette dernière contient 170 images dermoscopiques, composée de 100 lésions Bénéignes et 70 lésions malignes. Les résultats sont reportés dans le tableau suivant :

Espaces couleurs	Caractéristiques	KNN			ANN			SVM		
		SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)
RGB	$T_1, T_3, T_2, T_4,$ T_5, C_{13}, C_{11}, C_2	90	94	92.4	88.8	88	88.3	83.3	88	86
HSV	$T_1, T_2, T_3, C_{22},$ $C_{13}, C_{16}, C_{19}, T_4$	91.2	94.2	93	87.1	93.5	90.9	90.7	92	91.4
LAB	$T_{12}, T_4, T_5, T_{16},$ C_4, C_7	88.8	84	86	79.6	77.3	78.2	83.3	88	86
XYZ	$C_{15}, T_4, C_{21}, C_{14},$ C_{18}, T_5, T_3, C_{21}	88.4	89.6	87.4	77.1	86.6	82.7	85.1	85.3	85.2
YCbCr	$T_2, T_3, T_4, T_1, T_5,$ $T_6, C_{15}, C_{21}, C_{13}$	94.4	94.6	94.5	80	90.8	89.7	85.1	89.3	87.5

TABLE 3.5 – Classification des mélanomes et des lésions bénignes à l'aide de différents classificateurs et espaces couleurs sur l'ensemble de données DermIS & DermQuest.

Pour les deux classificateurs (SVM et ANN) le meilleur résultat est trouvé sur l'espace couleur HSV. Le SVM atteint 90.7 % de sensibilité, 92 % de spécificité et 91.4 % de précision. L'ANN atteint 87.1 % de sensibilité, 93.5 % de spécificité et 90.9 % de précision.

Avec le KNN, le meilleur résultat est trouvé sur l'espace couleur YCbCr. Il atteint 94.4 % de sensibilité, 94.6 % de spécificité et 94.5 % de précision

3.6 Conclusion

Le processus d'extraction de caractéristiques et de classification a été présenté le long de ce chapitre en vue d'une détection des mélanomes précoces. Premièrement, nous avons référencé les couleurs suspects en utilisant la segmentation par super-pixel.

Deuxièmement, la lésion est segmentée en classes par K-means en fonction du nombre de couleurs suspectes présentes. Ensuite, ces répartitions sont utilisées pour extraire de nombreuses caractéristiques de couleur et de texture, puis l'algorithme Relief sélectionne les caractéristiques les plus pertinentes.

Finalement, nous avons évalué l'algorithme proposé sur la base de données PH2, avec différentes configurations qui combinent cinq espaces de couleur (RGB, HSV, Lab, XYZ et YCrCb) et trois classificateurs (KNN, SVM et ANN). Les résultats montrent que notre algorithme est efficace et très sensible à l'identification des mélanomes en atteignant une sensibilité de 99,25%, une spécificité de 99,58% et une précision de 99, 51% avec le classificateur SVM et l'espace couleur RGB. L'algorithme montre également de bons résultats avec différents espaces de couleur et différents classificateurs, obtenant les meilleurs résultats en les comparant à trois articles qui ont utilisé la même base de données.

Conclusions et perspectives

Conclusion générale

La détection des mélanomes à un stade précoce est la meilleure façon de diminuer le taux de mortalité des personnes atteintes par cette maladie. En effet, ce type de cancer de peau se présente comme des lésions, qui apparaissent sur la peau sous différentes et complexes formes, couleurs et textures. C'est pourquoi les tumeurs malignes peuvent être confondues avec les tumeurs bénignes et rend le diagnostic très difficile pour les dermatologues. D'où l'intérêt d'utiliser des techniques d'apprentissage automatique pour améliorer la précision du diagnostic du mélanome et réduire son incidence élevée.

L'objectif fixé le long de cette thèse est la détection automatique des mélanomes précoces. Cette étude présente un algorithme de segmentation automatique et une classification basée sur des descripteurs de couleurs et de textures.

En premier lieu, nous avons présenté un état de l'art sur les méthodes de segmentation et de classification des images dermoscopiques, en vue de comprendre leurs concepts, leurs principes d'utilisation et leurs applications. Nous avons d'abord présenté une définition du mélanome et des systèmes du diagnostic assisté par ordinateur et ses différents éléments constitutifs pour la détection du mélanome. Ensuite, nous avons dressé un aperçu sur les approches de segmentation utilisées pour les images dermoscopiques. Suivie d'une présentation des méthodes de classification des mélanomes.

Nous avons proposé une approche de segmentation qui procède par les méthodes du clustering, appelées K-means. Nous avons commencé par un prétraitement pour faciliter la segmentation. Pour ce faire, une combinaison de plans est utilisée pour mettre en évidence la lésion. Une fois la lésion est segmentée avec les K-means, les lésions sont post-traitées dans le but d'éliminer les objets détachés et améliorer le rendu de la segmentation. Le post-traitement repose sur l'utilisation des contours actifs ainsi que les opérations morphologiques pour éliminer tous les objets non lésionnels. À l'issue des résultats obtenus, nous constatons que la segmentation proposée génère un masque assez précis pour chaque lésion.

Pour la classification, Les masques générés par la segmentation sont utilisés pour l'extraction des caractéristiques couleurs et textures. Afin de trouver la meilleure combinaison espace couleur/classificateur, les mêmes attributs de couleurs et de textures sont extraits de cinq espaces de couleur (RGB, HSV, Lab,XYZ et YCrCb) pour être testés avec trois classificateurs (KNN, SVM et ANN). L'évaluation de la classification est faite sur l'ensemble PH2, une base de données publique et référencée, qui est constituée de 200 images dermoscopiques dont 40 lésions de mélanomes et 160 lésions bénignes. Les résultats montrent l'efficacité de l'algorithme proposé dans l'identification des mélanomes en atteignant une sensibilité de 99,25%, une spécificité de 99,58% et une précision de 99,51% avec le classificateur SVM en utilisant l'espace couleur RGB. Il a obtenu également de bons résultats avec différents espaces de couleur et différents classificateurs. Enfin, L'expérience montre que notre algorithme est plus performant et obtenant les meilleurs résultats, notamment une meilleure balance entre la sensibilité et la spécificité, par rapport aux trois algorithmes déjà publiés et évalués avec la même base de données.

Perspectives

Les résultats de l'algorithme proposé montrent son efficacité à identifier le mélanome, cependant, des améliorations sont encore possibles. Pour la détection des mélanomes, en plus des caractéristiques proposées, des caractéristiques supplémentaires doivent être considérées pour prendre en compte les différents types de lésions. Les intervalles de couleurs de références sont trouvés manuellement, il est intéressant de développer un détecteur automatique de couleurs dominantes sur une lésion.

L'algorithme est évalué sur une base de données limitée, d'autres bases de données avec un nombre d'images plus importants est nécessaire pour voir les limitations de la méthode pour pouvoir l'améliorer d'avantage.

Les réseaux de neurones convolutifs, ont donné des résultats intéressants. Nous pouvons exploiter cette piste sur des ensembles de données plus importants des images dermoscopiques, le comparer avec des approches classiques de classification et/ou les fusionner afin de détecter les mélanomes avec une grande précision sur tous les types de peau.

Contributions

Dans cette section, les contributions de la thèse sont les suivantes :

Nous avons proposé une segmentation automatique et une classification des lésions cutanées en utilisant le K-means.

- Transformation d'image : deux plans d'images, le plan G de l'espace couleur RGB après ajustement de son histogramme ou un nouveau plan I_{new} proposé qui combine les deux plans S (saturation) et V (valeur) de l'espace couleur HSV de l'image originale, sont formulés. Puis le nouveau plan G ajusté ou I_{new} est filtré par un filtre médian 11x11 pour que le K-means segmente facilement et avec plus de précision les mélanomes. Elle présente de nombreux avantages : la plupart des artefacts tels que les poils fins sont filtrés et la lésion est mise en évidence.
- Extraction des caractéristiques : la lésion est divisée en K classes en utilisant K-means, où K est le nombre de couleurs dominantes et extraites de la lésion segmentée. Et ces répartitions sont utilisées pour extraire des caractéristiques basées sur la couleur et la texture. Les caractéristiques sélectionnées comme discriminantes mesurent la variation ou le désordre entre les classes au sein de la même lésion. Ceci correspond à la croissance incontrôlée des mélanomes.

Productions scientifiques

L'ensemble des travaux réalisés depuis le début de notre thèse a donné lieu à une publication scientifiques, trois conférences internationales et deux conférences nationales.

Publications

1. Oukil.S, Kasmi.R, Mokrani.K, García-Zapirain.B. Automatic Segmentation and Melanoma Detection Based On Color and Texture Features in Dermoscopic Image. *Skin Research and Technology*, 2021. doi :10.1111/srt.13111.

Conférences internationales

1. Samia Oukil, Reda Kasmi, and Karim Mokrani. U-net and k-means for dermoscopic skin lesion images : Segmentation and comparison. *Soft Computing and Electrical Engineering* ;Springer : Cham, Switzerland, 2, 2020.
2. S. OUKIL, R. KASMI, K. MOKRANI, A. MEKHMOUKH, I. CHERFA, S.S. GHENNAM. Colors Skin Lesions Detection For Melanoma Discrimination. *The Electrical Engineering International Conference EEIC'19*, Bejaia, Algérie, December 5, 2019.
3. I. CHERFA, A. MEKHMOUKH, K. MOKRANI, R. KASMI, S. OUKIL, S.S. GHENNAM. Medical Image Segmentation Using Adaptively Regularized Kernel-based Fuzzy C-Means (ARKFCM). *The Electrical Engineering International Conference EEIC'19*, Bejaia, Algérie, December 5, 2019.

Conférences nationales

1. S. OUKIL, R. KASMI, L. BENSADI, M.BENABDLHAK. Détection de mélanomes précoces par apprentissage automatique. *Première Journée Doctorale De La Faculté De Technologie*, Bejaia, Algérie, 15 Décembre 2021.
2. S. OUKIL, R. KASMI, L. BENSADI. Segmentation of Melanoma Skin Cancer from Dermoscopic Images Using Machine Learning Techniques. *National Conference of Mathematics and Applications CNMA 2021*, Mila, Algeria, December 11, 2021.

Bibliographie

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, pp. 209-249, 2021.
- [2] F. Nachbar, W. Stolz, T. Merkle, A. B. Coggnetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, pp. 551-559, 1994.
- [3] J. S. Henning, S. W. Dusza, S. Q. Wang, A. A. Marghoob, H. S. Rabinovitz, D. Polsky, and A. W. Kopf, "The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy," *Journal of the American Academy of Dermatology*, vol. 56, pp. 45-52, 2007.
- [4] G. Argenziano, C. Catricalà, M. Ardigo, P. Buccini, P. De Simone, L. Eibenschutz, A. Ferrari, G. Mariani, V. Silipo, and I. Sperduti, "Seven-point checklist of dermoscopy revisited," *British Journal of Dermatology*, vol. 164, pp. 785-790, 2011.
- [5] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2013, pp. 5437-5440.
- [6] A. Hartinger, "Détection du cancer de la peau par tomographie d'impédance électrique," *École Polytechnique de Montréal*, 2012.

BIBLIOGRAPHIE

- [7] A. A. Adeyinka and S. Viriri, "Skin lesion images segmentation: A survey of the state-of-the-art," in *International conference on mining intelligence and knowledge exploration*, 2018, pp. 321-330.
- [8] R. B. Oliveira, E. Mercedes Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: a review," *Computer methods and programs in biomedicine*, vol. 131, pp. 127-141, 2016.
- [9] M. A. Al-Masni, M. A. Al-Antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer methods and programs in biomedicine*, vol. 162, pp. 221-231, 2018.
- [10] F. F. X. Vasconcelos, A. G. Medeiros, S. A. Peixoto, and P. P. Reboucas Filho, "Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour," *Cognitive Systems Research*, vol. 55, pp. 44-59, 2019.
- [11] J. Ramya, H. Vijaylakshmi, and H. M. Saifuddin, "Segmentation of skin lesion images using discrete wavelet transform," *Biomedical Signal Processing and Control*, vol. 69, p. 102839, 2021.
- [12] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, pp. 146-165, 2004.
- [13] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Research and Technology*, vol. 19, pp. e252-e258, 2013.
- [14] Q. Abbas, I. F. Garcia, M. Emre Celebi, W. Ahmad, and Q. Mushtaq, "A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images," *Skin Research and Technology*, vol. 19, pp. e490-e497, 2013.
- [15] N. K. El Abbadi and A. H. Miry, "Automatic segmentation of skin lesions using histogram thresholding," 2014.
- [16] N. Z. Tajeddin and B. M. Asl, "A general algorithm for automatic lesion segmentation in dermoscopy images," in *2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME)*, 2016, pp. 134-139.
- [17] M. Zortea, E. Flores, and J. Scharcanski, "A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images," *Pattern Recognition*, vol. 64, pp. 92-104, 2017.

BIBLIOGRAPHIE

- [18] H. Fan, F. Xie, Y. Li, Z. Jiang, and J. Liu, "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold," *Computers in biology and medicine*, vol. 85, pp. 75-85, 2017.
- [19] M. Emre Celebi, H. A. Kingravi, H. Iyatomi, Y. Alp Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, and H. S. Rabinovitz, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, pp. 347-353, 2008.
- [20] A. Wong, J. Scharcanski, and P. Fieguth, "Automatic skin lesion segmentation via iterative stochastic region merging," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 929-936, 2011.
- [21] S. Bama, R. Velumani, N. Prakash, G. Hemalakshmi, and A. Mohanarathinam, "Automatic segmentation of melanoma using superpixel region growing technique," *Materials Today: Proceedings*, vol. 45, pp. 1726-1732, 2021.
- [22] P. G. Cavalcanti and J. Scharcanski, "A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images," *Computer methods and programs in biomedicine*, vol. 112, pp. 684-693, 2013.
- [23] Q. Abbas, I. Fondón, A. Sarmiento, and M. Emre Celebi, "An improved segmentation method for non-melanoma skin lesions using active contour model," in *International Conference Image Analysis and Recognition*, 2014, pp. 193-200.
- [24] R. Kasmi, K. Mokrani, R. Rader, J. Cole, and W. Stoecker, "Biologically inspired skin lesion segmentation using a geodesic active contour technique," *Skin Research and Technology*, vol. 22, pp. 208-222, 2016.
- [25] F. Riaz, S. Naeem, R. Nawaz, and M. Coimbra, "Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images," *IEEE journal of biomedical and health informatics*, vol. 23, pp. 489-500, 2018.
- [26] K. Abhilash, P. Kumar, Z. Ravindran, and A. N. Rahman, "Adaptive Active Contour Segmentation for Melanoma Diagnosis from Dermoscopy Images," in *Sixth International Conference on Intelligent Computing and Applications*, 2021, pp. 137-144.
- [27] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1-8, 1980.
- [28] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, pp. 281-297.

BIBLIOGRAPHIE

- [29] T. W. Cheng, D. B. Goldgof, and L. O. Hall, "Fast fuzzy clustering," *Fuzzy sets and systems*, vol. 93, pp. 49-56, 1998.
- [30] L. Szilagyi, Z. Benyo, S. M. Szilágyi, and H. Adam, "MR brain image segmentation using an enhanced fuzzy c-means algorithm," in *Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439)*, 2003, pp. 724-726.
- [31] A. Masood and A. A. Al-Jumaily, "Fuzzy C mean thresholding based level set for automated segmentation of skin lesions," *Journal of signal and information processing*, vol. 4, p. 66, 2013.
- [32] K. Eltayef, Y. Li, and X. Liu, "Detection of melanoma skin cancer in dermoscopy images," in *Journal of Physics: Conference Series*, 2017, p. 012034.
- [33] J. L. Garcia-Arroyo and B. Garcia-Zapirain, "Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding," *Computer methods and programs in biomedicine*, vol. 168, pp. 11-19, 2019.
- [34] D. Alvarez and M. Iglesias, "k-Means clustering and ensemble of regressions: an algorithm for the ISIC 2017 skin lesion segmentation challenge," *arXiv preprint arXiv:1702.07333*, 2017.
- [35] A. Agarwal, A. Issac, M. K. Dutta, K. Riha, and V. Uher, "Automated skin lesion segmentation using K-means clustering from digital dermoscopic images," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, 2017, pp. 743-748.
- [36] S. M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms," *IET Computer Vision*, vol. 12, pp. 1088-1095, 2018.
- [37] C. Rother, V. Kolmogorov, and A. Blake, "' GrabCut" interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)*, vol. 23, pp. 309-314, 2004.
- [38] X. Zhang, "Melanoma segmentation based on deep learning," *Computer assisted surgery*, vol. 22, pp. 267-277, 2017.
- [39] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on medical imaging*, vol. 36, pp. 994-1004, 2016.

BIBLIOGRAPHIE

- [40] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE transactions on medical imaging*, vol. 36, pp. 1876-1886, 2017.
- [41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015, pp. 448-456.
- [42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [43] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, "A novel multi-task deep learning model for skin lesion segmentation and classification," *arXiv preprint arXiv:1703.01025*, 2017.
- [44] B. S. Lin, K. Michael, S. Kalra, and H. R. Tizhoosh, "Skin lesion segmentation: U-nets versus clustering," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1-7.
- [45] R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule," *IET Image Processing*, vol. 10, pp. 448-455, 2016.
- [46] J. Jaworek-Korjakowska, "Computer-aided diagnosis of micro-malignant melanoma lesions applying support vector machines," *BioMed research international*, vol. 2016, 2016.
- [47] C. Barata, M. Ruela, T. Mendonça, and J. S. Marques, "A bag-of-features approach for the classification of melanomas in dermoscopy images: The role of color and texture descriptors," in *Computer vision techniques for the diagnosis of skin cancer*, ed: Springer, 2014, pp. 49-69.
- [48] N. Alfed and F. Khelifi, "Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images," *Expert Systems with Applications*, vol. 90, pp. 101-110, 2017.
- [49] S. Majumder and M. A. Ullah, "Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, 2018, pp. 185-188.
- [50] J. R. Hagerty, R. J. Stanley, H. A. Almubarak, N. Lama, R. Kasmi, P. Guo, R. J. Drugge, H. S. Rabinovitz, M. Oliviero, and W. V. Stoecker, "Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images," *IEEE journal of biomedical and health informatics*, vol. 23, pp. 1385-1391, 2019.

BIBLIOGRAPHIE

- [51] N. Moura, R. Veras, K. Aires, V. Machado, R. Silva, F. Araújo, and M. Claro, "ABCD rule and pre-trained CNNs for melanoma diagnosis," *Multimedia Tools and Applications*, vol. 78, pp. 6869-6888, 2019.
- [52] G. Hirano, M. Nemoto, Y. Kimura, Y. Kiyohara, H. Koga, N. Yamazaki, G. Christensen, C. Ingvar, K. Nielsen, and A. Nakamura, "Automatic diagnosis of melanoma using hyperspectral data and GoogLeNet," *Skin Research and Technology*, vol. 26, pp. 891-897, 2020.
- [53] M. F. Jojoa Acosta, L. Y. Caballero Tovar, M. B. Garcia-Zapirain, and W. S. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images," *BMC Medical Imaging*, vol. 21, pp. 1-11, 2021.
- [54] F. Dalila, A. Zohra, K. Reda, and C. Hocine, "Segmentation and classification of melanoma and benign skin lesions," *Optik*, vol. 140, pp. 749-761, 2017.
- [55] A. R. Smith, "Color gamut transform pairs," *ACM Siggraph Computer Graphics*, vol. 12, pp. 12-19, 1978.
- [56] M. E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, "A state-of-the-art survey on lesion border detection in dermoscopy images," *Dermoscopy image analysis*, vol. 10, pp. 97-129, 2015.
- [57] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, pp. 651-666, 2010.
- [58] M. B. Bonab, S. Z. M. Hashim, A. K. Z. Alsaedi, and U. R. a. Hashim, "Modified K-means combined with artificial bee colony algorithm and differential evolution for color image segmentation," in *Computational intelligence in information systems*, ed: Springer, 2015, pp. 221-231.
- [59] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International journal of computer vision*, vol. 22, pp. 61-79, 1997.
- [60] M. E. Celebi, T. Mendonca, and J. S. Marques, *Dermoscopy image analysis* vol. 10: CRC Press, 2015.
- [61] D. A. Okuboyejo and O. O. Olugbara, "A review of prevalent methods for automatic skin lesion diagnosis," *The Open Dermatology Journal*, vol. 12, 2018.
- [62] M. Anas, K. Gupta, and S. Ahmad, "Skin cancer classification using K-means clustering," *International Journal of Technical Research and Applications*, vol. 5, pp. 62-65, 2017.

BIBLIOGRAPHIE

- [63] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, pp. 2274-2282, 2012.
- [64] T. Sreelatha, M. Subramanyam, and M. Prasad, "Shape and color feature based melanoma diagnosis using dermoscopic images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 5371-5380, 2021.
- [65] A. Maiti, H. Shekhargiri, B. Chatterjee, V. Rajinikanth, F. Shi, and N. Dey, "Classification of melanoma through fused color features and deep neural networks," *Information Technology and Intelligent Transportation Systems*, vol. 323, p. 86, 2020.
- [66] R. J. Stanley, W. V. Stoecker, and R. H. Moss, "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Research and Technology*, vol. 13, pp. 62-72, 2007.
- [67] M. Lingala, R. J. Stanley, R. K. Rader, J. Hagerty, H. S. Rabinovitz, M. Oliviero, I. Choudhry, and W. V. Stoecker, "Fuzzy logic color detection: Blue areas in melanoma dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 38, pp. 403-410, 2014.
- [68] S. Pathan, V. Aggarwal, K. G. Prabhu, and P. Siddalingaswamy, "Melanoma detection in dermoscopic images using color features," *Biomedical and Pharmacology Journal*, vol. 12, pp. 107-115, 2019.
- [69] M. E. Celebi, H. A. Kingravi, Y. A. Aslandogan, and W. V. Stoecker, "Detection of blue-white veil areas in dermoscopy images using machine learning techniques," in *Medical Imaging 2006: Image Processing*, 2006, pp. 1861-1868.
- [70] R. B. Oliveira, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for pigmented skin lesion classification in images: review and future trends," *Neural Computing and Applications*, vol. 29, pp. 613-636, 2018.
- [71] F. Albrechtsen, "Statistical texture measures computed from gray level cooccurrence matrices," *Image processing laboratory, department of informatics, university of oslo*, vol. 5, 2008.
- [72] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, pp. 1157-1182, 2003.
- [73] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine learning proceedings 1992*, ed: Elsevier, 1992, pp. 249-256.
- [74] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Aai*, 1992, pp. 129-134.

BIBLIOGRAPHIE

- [75] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *Journal of biomedical informatics*, vol. 85, pp. 189-203, 2018.
- [76] B. V. Dasarathy, "Nearest neighbor (NN) norms: NN pattern classification techniques," *IEEE Computer Society Tutorial*, 1991.
- [77] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.
- [78] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115-133, 1943.
- [79] M. E. Celebi and A. Zornberg, "Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification," *IEEE systems journal*, vol. 8, pp. 980-984, 2014.
- [80] C. Barata, M. E. Celebi, and J. S. Marques, "Development of a clinically oriented system for melanoma diagnosis," *Pattern Recognition*, vol. 69, pp. 270-285, 2017.
- [81] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A multiple expert-based melanoma recognition system for dermoscopic images of pigmented skin lesions," in *2008 8th IEEE International Conference on BioInformatics and BioEngineering*, 2008, pp. 1-6.
- [82] S. Seidenari, G. Pellacani, and C. Grana, "Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment," *British Journal of Dermatology*, vol. 149, pp. 523-529, 2003.
- [83] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic detection of blue-white veil and related structures in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 670-677, 2008.
- [84] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 362-373, 2007.
- [85] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. McCarthy, I. Osman, A. W. Kopf, and D. Polsky, "Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria," *Jama*, vol. 292, pp. 2771-2776, 2004.
- [86] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using GLCM approach," *International journal of scientific and research publications*, vol. 3, pp. 1-5, 2013.

BIBLIOGRAPHIE

- [87] R. Amelard, J. Glaister, A. Wong, and D. A. Clausi, "High-level intuitive features (HLIFs) for intuitive skin lesion description," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 820-831, 2014.
- [88] D. O. T. Bruno, M. Z. Do Nascimento, R. P. Ramos, V. R. Batista, L. A. Neves, and A. S. Martins, "LBP operators on curvelet coefficients as an algorithm to describe texture in breast cancer tissues," *Expert Systems with Applications*, vol. 55, pp. 329-340, 2016.
- [89] E. Almansour and M. A. Jaffar, "Classification of Dermoscopic skin cancer images using color and hybrid texture features," *IJCSNS Int J Comput Sci Netw Secur*, vol. 16, pp. 135-9, 2016.
- [90] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE transactions on medical imaging*, vol. 36, pp. 849-858, 2016.
- [91] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, pp. 35-45, 2009.
- [92] "Dermatology Information System," <http://www.dermis.net>, Page consultée le 16 août 2017, 2012.
- [93] "DermQuest," <http://www.dermquest.com>, Page consultée le 16 août 2017, 2012.

ملخص

يُعرف سرطان الجلد بأنه شكل شديد العدوانية من سرطان الجلد وواحد من أسرع الأورام الخبيثة نموًا في العديد من البلدان. التشخيص بمساعدة الكمبيوتر (DAO) هو أسلوب شائع يستخدم في الكشف عن سرطان الجلد. ومع ذلك، فإنه لا يزال يواجه العديد من المضاعفات، خاصة عندما يتعلق الأمر بالتمييز في مرحلة مبكرة عن الآفات الخبيثة من الحميدة. تهدف هذه الرسالة إلى تمييز الورم الميلانيني عن الآفات الجلدية الحميدة في صور التنظير الجلدي من أجل تحسين إمكانيات وكفاءة ودقة نظام دعم القرار الطبي الذي يمكن أن يساعد أطباء الأمراض الجلدية في تشخيص الآفات الجلدية في مرحلة مبكرة. تستخدم الخوارزمية المقترحة تجزئة تلقائية تعتمد على الوسائل K-means قناع دقيق إلى حد ما لكل آفة. يتكون استخراج الميزة من سمات الألوان والملمس الموجودة والجديدة، والتي تقيس كيفية اختلاف اللون والملمس داخل الآفة. من أجل العثور على أفضل النتائج، يتم استخراج جميع السمات من الآفات الموجودة على خمسة أنظمة ألوان مختلفة (RGB, HSV, Lab, XYZ, YCbCr)، وتستخدم كمداخل لثلاثة مصنفات (KNN, SVM, ANN). تُستخدم مجموعة صور تنظير الجلد بمستشفى بيدرو هيسبانو PH2 لتقييم أداء الخوارزمية المقترحة. تمت مقارنة نتائج الخوارزمية الخاصة بنا بنتائج الأبحاث المنشورة التي استخدمت نفس مجموعة البيانات ووجد أن الطريقة المقترحة تتفوق على نتائج الأعمال السابقة من خلال تحقيق حساسية 99.25% وخصوصية 99.58% وصحة 99.51%. تظهر النتائج النهائية أن الألوان مجتمعة مع الملمس هي سمات قوية وذات صلة لاكتشاف سرطان الجلد وتظهر تحسناً على ما هو معروض في الأدبيات باستخدام نفس قاعدة البيانات.

الكلمات الرئيسية: سرطان الجلد؛ تجزئة. تنظير الجلد. سوبر بكسل، K-means. صفات.

Résumé

Le mélanome est connu comme une forme très agressive de cancer de la peau et l'une des tumeurs malignes dont la croissance est la plus rapide dans de nombreux pays. Le diagnostic assisté par ordinateur (CAD) est une technique populaire utilisée pour la détection du mélanome dans la littérature. Cependant, elle se heurte encore à de nombreuses complications, notamment lorsqu'il s'agit de distinguer les lésions malignes des lésions bénignes à un stade précoce. Cette thèse vise à discriminer le mélanome des lésions cutanées bénignes dans les images dermoscopiques afin d'améliorer le potentiel, l'efficacité et la précision d'un système d'aide à la décision médicale qui peut aider les dermatologues à diagnostiquer les lésions cutanées à un stade précoce. L'algorithme proposé utilise une segmentation automatique basée sur le k-means générant un masque assez précis pour chaque lésion. L'extraction de caractéristiques consiste en des attributs de couleur et de texture existants et nouveaux, mesurant comment la couleur et la texture varient à l'intérieur de la lésion. Afin de trouver les résultats optimaux, tous les attributs sont extraits des lésions sur cinq systèmes de couleurs différents (RGB, HSV, Lab, XYZ et YCbCr), et utilisés comme entrées pour trois classificateurs (KNN, SVM et ANN). L'ensemble d'images dermoscopiques PH2 de l'hôpital Pedro Hispano est utilisé pour évaluer la performance de l'algorithme proposé. Les résultats de notre algorithme sont comparés aux résultats d'articles publiés qui ont utilisé le même ensemble de données et il en ressort que la méthode proposée surpasse les résultats des travaux antérieurs en atteignant la sensibilité de 99,25%, la spécificité de 99,58% et la précision de 99,51%. Les résultats finaux montrent que les couleurs combinées à la texture sont des attributs puissants et pertinents pour la détection des mélanomes et montrent une amélioration par rapport à ce qui est montré dans la littérature en utilisant la même base de données.

Mots clés : Mélanome ; segmentation ; dermoscopie ; super-pixel ; K-means ; Attributs.

Abstract

Melanoma is known as the most aggressive form of skin cancer and one of the fastest growing malignant tumors worldwide. Several Computers Aid Diagnosis of melanoma are proposed, still the algorithms encounter difficulties on the early stage of lesions. The paper aims to discriminate melanoma from benign skin lesion in dermoscopic images. The proposed algorithm is based on the color and texture of skin lesions by introducing a novel technique of features extraction. The algorithm uses an automatic segmentation based on k-means generating a fairly accurate mask for each lesion. The feature extraction consists on the existing and a novel color and texture attributes measuring how color and texture vary inside lesion. In purpose to find the optimal results all the attributes are extracted from lesion on five different system colors (RGB, HSV, Lab, XYZ and YCbCr), and used as the inputs for three classifiers (KNN, SVM and ANN). The PH2 set is used to assess the performance of the proposed algorithm. The results of our algorithm are compared to the results of published articles that used the same dataset and it shows that the proposed method outperforms the works of the state-of-the-art results by reaching the sensitivity of 99.25%, specificity of 99.58% and accuracy of 99.51%. The final results show that the colors combined with texture are powerful and pertinent attributes for melanoma detection and shows improvement over what is shown in the state-of-the-art using the same dataset.

Key words : Melanoma ; segmentation ; dermoscopy ; super-pixel ; K-means ; features extraction ; classification.