

Extraction et évaluation de caractéristiques adaptées pour la classification du Lentigo à partir d’images de Microscopie Confocale

Romain CENDRE¹, Alamin MANSOURI¹, Jean-Luc PERROT², Elisa CINOTTI³, Franck MARZANI¹

¹Laboratoire ImViA EA 7535, Université de Bourgogne, Dijon, France

²Service de Dermatologie-Oncologie-Allergologie, CHU de St Etienne, France

³U.O. Dermatologia, Dipartimento di Scienze Mediche, Università degli Studi di Siena, Italie

romain.cendre@gmail.com, alamin.mansouri@u-bourgogne.fr,
j.luc.perrot@chu-st-etienne.fr, elisacinotti@gmail.com, franck.marzani@u-bourgogne.fr

Résumé – La détection de cancer de la peau est l’un des défis de ces dernières décennies. Par ailleurs, diverses techniques d’imagerie ont pour objectif d’aider à la reconnaissance de ces pathologies malignes en contexte clinique. La Microscopie Confocale par Réflectance est un exemple de technique d’imagerie adaptée à la détection de maladie de la peau sur laquelle nous nous basons pour la détection de Lentigo. Les travaux présentés dans cet article portent sur la classification de ces images en trois catégories : sain, bénin et malin. Dans ce but, nous proposons et évaluons deux méthodes d’extraction de caractéristiques basées sur les descripteurs d’Haralick pour l’une et sur une extraction par Réseau Neuronal Convolutif pour l’autre. Ces deux méthodes sont évaluées et comparées à un travail suggérant une approche par ondelettes, étendu à ces mêmes catégories. D’après les résultats obtenus, l’approche par Réseau Neuronal Convolutif est la plus adaptée à cette problématique.

Abstract – Detection of skin cancer in time used to be a challenge over decades. In clinical circumstances, a couple of imaging techniques aims to improve early recognition of malignant pathologies. For instance, Reflectance Confocal Microscopy modality is a suitable technique for skin pathologies diseases in clinical context. This study intends to classify these images under three categories: healthy, benign and malignant Lentigo. We introduce and evaluate two extraction ways based on Haralick textures descriptors and a Deep features extraction. Then, we reproduce a work suggesting wavelets descriptors for this purpose, applied to these three categories. These methods are performed on different images scales and, according to results, Convolutional Neural Network is the most pertinent.

1 Introduction

De nos jours, les cancers de la peau sont parmi les pathologies malignes les plus récurrentes rencontrées cliniquement. Elles impactent socialement les personnes concernées en provoquant de forts désagréments, et peuvent conduire au décès des individus. De plus, elles occasionnent de fortes dépenses auprès des organismes de soin [1]. Pourtant, la plupart de ces répercussions pourraient être évitées grâce à des outils de détection adaptés et une prise en charge médicale adéquat.

Actuellement, la biopsie reste la technique de référence pour établir le diagnostic des pathologies en dermatologie. Cependant, cette technique reste dispendieuse et peu pratique dans un contexte clinique. En outre, de nombreuses modalités d’imagerie ont émergé afin de faciliter le processus de diagnostic et certaines d’entre elles sont désormais d’usage courant. La Dermatoscopie en est un exemple, à la fois peu onéreuse mais également très facile d’usage par le spécialiste, apportant une amélioration significative dans son diagnostic clinique par rapport à un simple examen à l’oeil nu [2]. Par ailleurs, les principaux travaux de recherche en détection automatique se focalisent sur cette modalité et obtiennent des résultats probants dans le cadre de lésions mélanocytiques [3]. Parmi ces travaux, nous retrouvons d’une part des processus de traitement établis

manuellement [4] et d’autre part, des méthodes basées sur des architectures de réseaux profonds cherchant à déterminer automatiquement les critères pertinents à partir des données d’apprentissage [5].

Certaines modalités d’imagerie de plus en plus utilisées en contexte clinique, comme la Microscopie Confocale par Réflectance (MCR), semblent être délaissées par la littérature. Pourtant, cette modalité possède nombre d’avantages, dont sa robustesse face au type de peau [6] et des résultats prometteurs sur les lésions mélanocytiques et non mélanocytiques [7]. Quelques travaux portant sur la détection de lésions ont été menés, fondés sur une extraction de caractéristiques spatiales [8] ou encore fréquentielles [9].

Cet article s’inscrit dans une démarche de classification du Lentigo et de sa forme cancéreuse [10]. Un travail récent propose d’évaluer la pertinence d’une extraction sur base d’ondelettes, pour une classification de tissu sain et de Lentigo pathologique [9]. Nous utilisons cette méthode comme référence, et l’étendons à une classification selon trois classes : sain, bénin et malin. Par la suite, nous étudions la pertinence de deux méthodes d’extraction, la première sur la base des descripteurs d’Haralick et la seconde par extraction de caractéristiques à partir d’un Réseau Neuronal Convolutif (RNC).

Nous décomposons cet article en quatre sections distinctes.

Tout d'abord, la Section 2 décrit les méthodes proposées lors de ces expériences. Puis, en Section 3 nous nous intéressons aux données sur lesquelles sont appliquées ces méthodes. Enfin, en Section 4 nous présentons les résultats obtenus et en Section 5 nous concluons et élaborons des perspectives à ce travail.

2 Méthodes

Comme évoqué en Section 1, cette étude aborde trois méthodes d'extraction de caractéristiques dans le cadre de notre problématique. Nous décrirons en premier lieu le processus dans lequel ils s'inscrivent, avant de les décrire plus en détail.

2.1 Processus global

Premièrement, les caractéristiques extraites propres à chaque méthode sont normalisées afin de rendre la tâche de classification plus efficace mais également plus robuste [11]. Pour cela, chaque variable observée est centrée par sa moyenne μ , puis réduite par son écart-type σ selon les données mises à disposition pour l'entraînement.

Ensuite, ces caractéristiques sont intégrées dans un classifieur de type Support Vector Machine (SVM) avec une approche multi-classe de type "one vs all". Par ailleurs, le noyau sélectionné est de type linéaire afin d'évaluer de manière simple la capacité de séparation des diverses méthodes d'extraction. Le terme de régularisation est déterminé par validation croisée à partir d'une échelle logarithmique variant de 0,01 à 1000. L'ensemble de ce processus est visible en Figure 1.

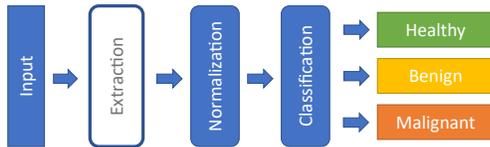


FIGURE 1 – Processus de classification.

2.2 Extracteurs de caractéristiques

La première méthode d'extraction de caractéristiques sur laquelle nous nous appuyons se base sur une décomposition en ondelettes de Daubechies [9]. Cette décomposition est réalisée à quatre échelles différentes. Pour chacune d'elles, nous récupérons un ensemble de coefficients horizontaux, verticaux et diagonaux. Afin de réduire la dimension des descripteurs, la distribution des coefficients est modélisée par une loi normale généralisée centrée dont la densité de probabilité f est décrite par l'Equation (1), avec Γ correspondant à la fonction gamma. Les auteurs de l'étude ne retiennent comme caractéristiques que les paramètres d'échelle α et de forme β . Dans un but de clarté, nous nommerons cette méthode "Ondelettes"

dans les prochains paragraphes. La décomposition par ondelettes est effectuée par la bibliothèque logicielle "PyWavelets" [12].

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

La seconde méthode que nous mettons en oeuvre se base sur treize des caractéristiques définies par Haralick, dont l'efficacité a été démontré sur des images texturées. Ces caractéristiques sont extraites selon les axes horizontal, vertical et les diagonales distinctes. Afin de simplifier la lecture, nous emploierons le terme "Haralick" pour nommer cette méthode dans la suite de cet article. Cette extraction se base sur la bibliothèque logicielle "Mahotas" [13].

Enfin, la dernière méthode proposée se base sur les RNC, généralement appropriés à la classification d'images. Nous utiliserons à cette fin une architecture de type "InceptionV3" supposée pertinente dans ce cadre d'application [14]. Afin de tirer le meilleur parti de cette architecture, nous nous orientons vers de l'apprentissage par transfert. En effet, ce type de techniques est particulièrement adapté à des problématiques suffisamment proche des données initiales, et pour lesquelles la quantité de données disponible est restreinte. Pour cela, nous nous référons à une version de ce modèle entraînée sur la base d'images "ImageNet", dont nous retirons au préalable les couches responsables de la classification. Afin de conserver une cohérence entre le nombre de caractéristiques extraites et le nombre d'échantillons à notre disposition, nous ajoutons également une couche de "Global Pooling" pour réduire la dimension spatiale de chaque carte d'activation à un simple scalaire. Lors de ces prochaines lignes, nous utilisons le terme "RNC Max" (respectivement "RNC Moy") pour faire référence à un tel réseau comportant respectivement une couche de "Global Pooling" basé sur une fonction de maximum (respectivement moyenne). De plus, nous combinons aux précédents éléments une Analyse en Composantes Principales (ACP) afin de réduire considérablement le nombre de variables de manière non supervisée et analyser son incidence. Nous emploierons dans ce cas les termes "RNC Max + ACP" et "RNC Moy + ACP" dans la suite cet article. La manipulation des RNC est réalisée grâce à la bibliothèque logicielle "Keras" [15].

La Table 1 synthétise l'ensemble des méthodes d'extraction ainsi que le nombre associé de caractéristiques extraites.

Méthode	Nombre
Ondelettes	24
Haralick	52
RNC	2048
RNC+ ACP	20

TABLE 1 – Nombre de caractéristiques extraites par méthode.

3 Données

Les données de cette étude sont fournies par le service de dermatologie de l'hôpital de Saint-Étienne avec consentement du patient et proviennent d'une base contenant trois types d'images : Photographie clinique, Dermatoscopie et MCR. De plus, chaque patient est associé à une vérité terrain tirée de l'histopathologie [16].

Les images de la modalité MCR sont acquises au niveau de la jonction dermo-épidermique jugée pertinente pour le diagnostic du Lentigo bénin et malin par le spécialiste. Ces images natives peuvent contenir différents types de tissus, lésionnel ou non, et ont une résolution spatiale de 1000 pixels par 1000 pixels, avec une quantification sur un seul canal de 8 bits.

Un premier jeu de données est constitué à partir de la base d'images initiale et avec l'aide du spécialiste. Des annotations sont réalisées sur 68 patients, soit 2089 images MCR, selon trois critères : "Malin" pour les images comprenant au moins du tissu malin, "Bénin" pour les images composées d'une pathologie bénigne et enfin une étiquette "Sain" correspondant à l'absence de ces deux caractéristiques. Les ratios associés à ce jeu de données sont les suivants : 38% de données Maligne, 37% de données Saines et 25% de données Bénigne. Nous nommons ce jeu de données "Images Originales".

Un second jeu de données dérivé des images natives est également constitué, afin d'évaluer la capacité à différencier les types de tissus. A cette fin, une interface graphique permettant de réaliser des sous-images de 250 pixels par 250 pixels est mise à disposition du spécialiste. Ainsi, 200 sous-images incluant un type de tissu particulier sont extraites pour chacune des trois annotations mentionnées précédemment. Des exemples de ces sous-images extraites par le spécialiste sont visibles en Figure 2. Nous nommerons ce jeu de données "Sous-images".

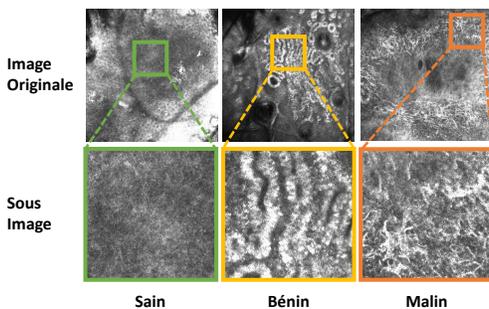


FIGURE 2 – Exemple de données originales et de sous-images extraites par le spécialiste.

4 Résultats

Le protocole de validation employé est équivalent lors de chaque expérience. Pour ce faire, nous réalisons une validation

croisée imbriquée afin de permettre une sélection des hyperparamètres du modèle SVM, tout en assurant son évaluation objective [17]. Chacune des étapes de validation croisée est basée sur une stratégie de type "K-Fold" sur 5 échantillons. Nous utilisons la bibliothèque "Scikit Learn" pour réaliser ces démarches de validation, ainsi que pour la classification et l'évaluation des mesures [18].

Chaque expérience est évaluée à l'aide du F1-Score, dont la valeur tend vers 1 conjointement à la Précision et au Rappel, et permet une meilleure objectivité dans une situation où les données sont non balancées. De plus, un écart-type est calculé afin d'analyser la stabilité des modèles lors de la validation croisée imbriquée.

La Table 2 met en avant les différents F1-Scores obtenus lors de l'évaluation de chaque expérience, selon leur évaluation sur le jeu de données utilisant les images originales ou les sous-images.

Tout d'abord, nous remarquons une stabilité accrue de l'ensemble des méthodes évaluées et ce quel que soit l'échantillon utilisé lors de l'évaluation.

En revanche, nous obtenons des résultats mitigés lors de l'évaluation sur les images originales. La méthode basée sur la décomposition par ondelette semble peu concluante dans ces conditions. Nous retrouvons des performances acceptables avec l'extraction des caractéristiques d'Haralick ainsi qu'avec les "RNC Max", "RNC Max+ACP" et "RNC Moy+ACP". Par ailleurs, nous émettons une réserve concernant l'extraction "RNC Moy", et ce malgré le dispositif de validation mis en place, dans la mesure où le nombre de caractéristiques extraites avoisine le nombre d'observations à disposition.

Afin de vérifier la pertinence de ces méthodes à discriminer les différentes formes de tissu, nous avons réalisé leur évaluation sur la base de sous-images. Nous pouvons observer de très bons résultats sur l'ensemble des méthodes dans ces conditions d'évaluation. Cette fois-ci, la méthode basée sur la décomposition par ondelettes semble plus apte à caractériser ces sous-images que les descripteurs d'Haralick. En revanche, les méthodes basées sur RNC semblent exceller à cette tâche de classification. Néanmoins, nous devons émettre une réserve sur les résultats concernant le "RNC Max" et "RNC Moy" et ce malgré le processus de validation mis en place. En effet, le nombre de caractéristiques extraites est nettement supérieur au nombre d'observations (600 observations pour 2048 variables) pouvant conduire à des résultats trop optimistes.

Méthodes	Images Originales	Sous-images
Ondelettes	0.62 ± 0.01	0.87 ± 0.02
Haralick	0.71 ± 0.02	0.82 ± 0.01
RNC Max	0.76 ± 0.02	0.97 ± 0.02
RNC Max + ACP	0.75 ± 0.01	0.95 ± 0.02
RNC Moy	0.86 ± 0.02	0.99 ± 0.01
RNC Moy + ACP	0.76 ± 0.01	0.95 ± 0.01

TABLE 2 – Performances mesurées sur diverses méthodes à l'aide du F1-Score.

5 Conclusion et Perspectives

Dans cette étude nous avons évalué la pertinence de trois extracteurs de caractéristiques dans le cadre d'images MCR. Ces diverses méthodes, et plus particulièrement les RNC nous permettent de réaliser une classification robuste sur des sous-images dont la pathologie est isolée. Celle-ci est en revanche plus complexe sur les images originales et ce quelque soit la méthode utilisée.

L'une de nos hypothèses non formulée supposait une supériorité lors de la classification de la couche de "Global Pooling" basée sur une fonction de type moyenne par rapport à une fonction de type maximum. Cette hypothèse se fondait sur une plus grande tolérance au bruit de la moyenne. Or dans ce cadre, notre travail suggère une faible influence de ce paramètre sur les résultats.

De futures investigations devraient permettre une meilleure classification de ces images. La première piste envisagée se base sur un découpage de l'image MCR et sur l'agrégation de caractéristiques à basse échelle. La seconde approche centrée sur les RNC, consistera à analyser l'impact du "fine-tuning" sur la capacité du modèle à classifier ces mêmes images. De plus nous souhaiterions par cette approche exploiter le potentiel des cartes d'activations générées par ces réseaux pour réaliser des segmentations par apprentissage faiblement supervisé.

Ce travail s'inscrit comme le premier maillon d'un ensemble de travaux et devrait, à plus long terme, nous permettre d'envisager une prise de décision au niveau du patient, voire également multi-modale par l'utilisation des autres modalités d'image.

Remerciements

Ces travaux ont été réalisés dans le cadre d'un financement proposé par le Conseil Régional de Bourgogne Franche Comté et d'un co-financement par le Fonds européen de développement régional.

Nous tenons à remercier le Dr. Perrot ainsi que le Dr. Cinotti pour leur mise à disposition de la base d'images, mais également pour leur précieux travail d'annotation.

Références

- [1] Aaron S. Farberg and Darrell S. Rigel. The Importance of Early Recognition of Skin Cancer, 2017.
- [2] Christoph Sinz, Philipp Tschandl, et al. Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. *Journal of the American Academy of Dermatology*, 77(6) :1100–1109, 2017.
- [3] Hitoshi Iyatomi, Kerri Ann Norton, et al. Classification of melanocytic skin lesions from non-melanocytic lesions. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, pages 5407–5410, 2010.
- [4] Sameena Pathan, K. Gopalakrishna Prabhu, and P. C. Sidalingaswamy. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review. *Biomedical Signal Processing and Control*, 39 :237–262, 2018.
- [5] Andre Esteva, Brett Kuprel, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639) :115–118, 2017.
- [6] Babar K. Rao, Ann M John, et al. Diagnostic Accuracy of Reflectance Confocal Microscopy Skin. *Archives of Pathology & Laboratory Medicine*, 143(March) :326–329, 2018.
- [7] Attiya Haroon, Shahram Shafi, and Babar K. Rao. Using Reflectance Confocal Microscopy in Skin Cancer Diagnosis. *Dermatologic Clinics*, 35(4) :457–464, 2017.
- [8] Marco Wiltgen, A. Gerger, et al. Automatic identification of diagnostic significant regions in confocal laser scanning microscopy of melanocytic skin tumors. *Methods of Information in Medicine*, 47(1) :14–25, 2008.
- [9] Abdelghafour Halimi, Hadj Batatia, et al. Wavelet-based statistical classification of skin images acquired with reflectance confocal microscopy. *Biomedical Optics Express*, 8(12) :5450–5467, 2017.
- [10] Frédérique-Anne Le Gal, Laurence Toutous-Trellu, et al. Lentigo maligna : un mélanome particulier. *Revue Médicale Suisse*, 7 :765–71, 2011.
- [11] Arnulf B A Graf and Silvio Borer. Normalization in Support Vector Machines. In Bernd Radig and Stefan Florczyk, editors, *Pattern Recognition*, pages 277–282, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- [12] G Lee, R Gommers, et al. PyWavelets - Wavelet Transforms in Python, 2006.
- [13] Luis Pedro Coelho. Mahotas : Open source software for scriptable computer vision. *arXiv preprint arXiv :1211.4907*, 2012.
- [14] Geert Litjens, Thijs Kooi, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(December 2012) :60–88, 2017.
- [15] François Chollet and Others. Keras, 2015.
- [16] E. Cinotti, B. Labeille, et al. Dermoscopy vs. reflectance confocal microscopy for the diagnosis of lentigo maligna. *Journal of the European Academy of Dermatology and Venereology*, 32(8) :1284–1291, 2018.
- [17] Gavin C. Cawley and Nicola L. Talbot. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research*, 11 :2079–2107, 2010.
- [18] Fabian Pedregosa, Gaël Varoquaux, et al. Scikit-learn : Machine Learning in Python. *Journal of machine learning research*, 12(Oct) :2825–2830, 2011.