

Une approche unifiée pour discriminer les tumeurs noires de la peau

Camille Serruys¹, Alain Giron¹, Romain Viard¹, Raoul Triller², Bernard Fertil¹

¹ INSERM U494 - CHU Pitié-Salpêtrière, 91 boulevard de l'hôpital,
75634 Paris cedex 13 - France
email: giron@imed.jussieu.fr

² Centre international de Dermatologie, 2 rue Barbès
92300 Levallois-Perret – France

RÉSUMÉ. Nous avons développé une méthodologie de traitement d'images de tumeurs noires fondée exclusivement sur l'utilisation de méthodes adaptatives à partir d'exemples documentés par les dermatologues.

Pour cela cinq dermatologues ont analysé indépendamment 600 images (cliniques et dermatoscopiques) en localisant les signes de malignité. Les réponses obtenues et le consensus dégagé ont servi de base pour les étapes de l'apprentissage supervisé. C'est ainsi qu'il a été possible de segmenter les lésions et de reconnaître certaines caractéristiques, intéressant directement le diagnostic, telles que la multiplicité de couleurs, la présence de réseau pigmenté ou l'irrégularité des bords de la tumeur, avec une même approche.

MOTS-CLÉS. Apprentissage, Classification, Segmentation, Mélanome

1. Introduction

Le mélanome malin est la plus grave des tumeurs cutanées. Si elle n'est pas retirée chirurgicalement à temps (exérèse) le pronostic de survie en est fortement altéré. L'exérèse des formes précoces (stade I) permet la guérison complète quand la tumeur n'est pas trop épaisse. La prévention et la détection précoce jouent donc un rôle essentiel, mais restent un enjeu difficile.

Actuellement, un dermatologue expérimenté peut identifier un mélanome avec 75% d'efficacité. Pour cela, il se fie en grande partie à l'observation visuelle de la tumeur en s'intéressant à des caractéristiques comme la coloration, la forme ou la texture. La principale difficulté est alors de pouvoir différencier les mélanomes de certaines tumeurs mélanocytaires moins proliférantes (Wolff 1998). Des techniques de microscopie de surface, tel la dermatoscopie, sont venues compléter les résultats de l'examen clinique et ont fait naître une nouvelle sémiologie, actuellement encore en évolution, et nécessitant un apprentissage conséquent par les dermatologues.

Bien que les dermatoscopes, ou les appareils photos utilisés pour l'examen clinique, soient maintenant associés à des outils informatiques d'acquisition et de stockage d'images, les outils de quantification et de traitement d'images sont généralement peu adaptés à la prise de décision. Dans ce contexte, l'objectif de ce projet est de déterminer un ensemble de "caractéristiques pertinentes" calculées à partir d'images cliniques et dermatoscopiques afin de discriminer au mieux les signes spécifiques des mélanomes.

Une approche unifiée pour discriminer les tumeurs noires de la peau

L'approche retenue est fondée sur l'utilisation de méthodes adaptatives supervisées, qui permettent au système d'acquérir sa compétence à partir d'exemples : ici des images répertoriées par les dermatologues et auxquelles sont associés les signes et les régions d'intérêt caractéristiques des lésions. Cette approche s'appuie sur trois étapes (Figure 1) :

- La constitution et la validation d'une base d'images destinée à l'apprentissage.
- L'extraction et le codage des informations destinées à la reconnaissance du facteur étudié qui se traduit par la recherche de paramètres pertinents et de leurs estimateurs.
- La classification des paramètres (opérée grâce à des modèles robustes et adaptatifs) afin de prédire la typologie de la lésion. Elle concerne le regroupement de l'ensemble des variables (correspondant au facteur étudié) et la classification de la lésion.

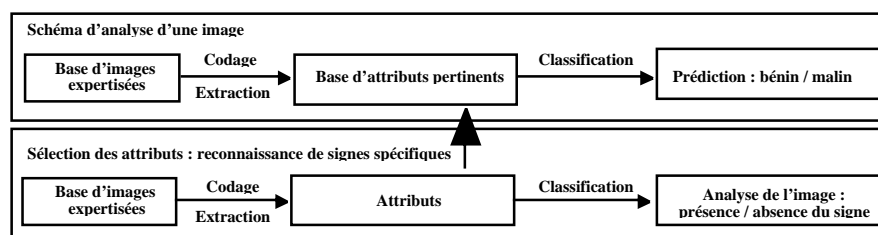


Figure 1 - Schéma général de classification des lésions à partir d'attributs.

Les études et les prototypes réalisés dans le passé se sont souvent fondés sur une modélisation des signes diagnostiques ou l'application de méthodes par apprentissage, type « réseaux de neurones » à certains de ces signes (Colot 1998, Schindewolf 1993, Sboner 1993, Binder 1998, Ercal 1994, Hintz-Madsen 1995, Kjoelen 1995).

Notre approche sera ici d'appliquer des techniques par apprentissage à l'ensemble des caractères susceptibles d'intervenir dans la discrimination des tumeurs. Cela concerne aussi bien des étapes de pré-traitement de l'image (segmentation), que la recherche de caractères généraux de la tumeur (asymétrie, couleur), de caractères spécifiques d'une zone (bords irréguliers), ou la détection de structures bien définies (pseudopodes, réseau pigmenté ...).

Dans ce cadre méthodologique, la constitution de la base d'apprentissage est bien évidemment fondamentale. Elle doit être représentative des caractères étudiés, doit rendre compte de ses différentes formes et être accompagnée d'un ensemble tout aussi représentatif de contre-exemples. Par ailleurs elle doit être validée en tenant compte de la variabilité des interprétations et parfois de l'absence de référence (gold-standard).

Pour répondre à cette attente, après une réunion de consensus, concernant les définitions et les signes à étudier, cinq dermatologues (3 en hôpital et 2 en cabinet privé), ont contribué à l'évaluation d'une base de 600 images extraites de la base initiale de 1800 images couleurs (600 * 400 pixels). Les données recueillies nous ont permis d'analyser la reproductibilité intra et inter-experts, et d'évaluer le degré de concordance entre médecins. La variabilité inter-dermatologue est, comme attendue, supérieure à la variabilité intra sauf pour les notions de multiplicité de couleurs et la recherche de réseau pigmenté, où elles sont du même ordre (Giron 2002). Pour certains signes, a été défini un consensus résultant d'un processus de

vote. Ce consensus, quantifié par un index compris entre 0 et 5 et traduisant le nombre de médecins ayant observé le caractère étudié, va servir de référence pour indiquer la présence d'un caractère dans la base d'apprentissage (Figure 2). On a ainsi pu montrer une plus grande concordance des diagnostics de mélanome avec les images dermatoscopiques. Parmi l'ensemble des caractères intervenant dans la discrimination des mélanomes, nous nous focaliserons sur quelques exemples dont la variété montre la généralité de la méthode.

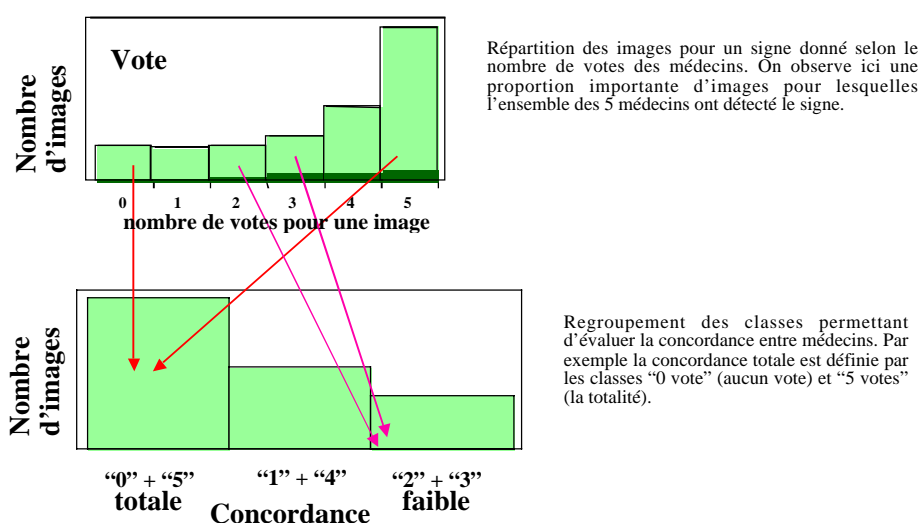


Figure 2 – Description du processus de vote des dermatologues

2. Segmentation automatique

Segmenter une lésion pour un dermatologue est une tâche aisée. Pourtant, mettre en place une méthode de segmentation automatique "universelle", nécessaire à l'extraction des informations de couleur, de forme, ou de texture, est extrêmement difficile. L'addition de bruit, la perte d'information spatiale et de profondeur, font que la segmentation n'est possible que si la tumeur est caractérisée par des couleurs et/ou des textures qui la différencient efficacement du fond. Cette problématique, résolue plus ou moins bien avec d'autres méthodes (clustering, détection de contours, agrégation ...) (Zhang 2000), constitue un champ d'application original de notre approche.

Partant du principe que la segmentation de la tumeur peut s'effectuer localement à partir d'une vision partielle de l'image, on utilisera des imagerie dont la taille sera ajustée au problème à résoudre. La méthodologie proposée (base d'imagerie expertisée, choix d'attributs pertinent, classification supervisée) s'applique ainsi à plusieurs niveaux de résolution, obtenus à partir de la modification de la taille des imagerie. La segmentation initiale, à un niveau très « grossier », cherche à localiser la lésion. Le niveau d'analyse évolue ensuite en fonction de la fenêtre d'observation. La localisation du bord devient alors

Une approche unifiée pour discriminer les tumeurs noires de la peau

plus précise et la contrainte imposée sur l'espace d'analyse permet d'éliminer de nombreux artefacts (Figure 3).

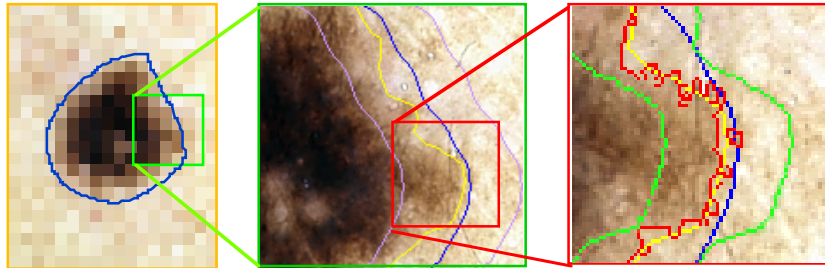


Figure 3 - segmentation (niveau 1, 2, 4)

Au premier niveau, sont traitées des imagerie de taille 4×4 sur des images où la résolution est divisée par 16. Cela correspond à un champ de taille 64×64 sur les images originales. En ce qui concerne les niveaux suivants, la résolution initiale est conservée, seule la taille des imagerie varie. On extrait des imagerie de taille 16×16 , 4×4 et 1×1 (un pixel). Une centaine d'images a servi à l'apprentissage (100 000 imagerie) et 80 au test (80 000 imagerie). La procédure d'analyse ne s'effectue plus sur l'ensemble de l'image mais sur une zone qui dépend de l'analyse précédente (contour prédit auquel s'ajoute une bande d'incertitude, définie par la largeur de l'imagerie au niveau précédent multipliée par un facteur dépendant de la taille). Cette façon de procéder permet au système de ne pas être perturbé par des artefacts une fois que la lésion a été « grossièrement » localisée au niveau 1.

L'étape supervisée de classification est réalisée par une régression logistique (spécifique de chaque niveau) appliquée à la cible (imagerie intérieure/extérieure à la tumeur) en intégrant les moyennes (des imagerie) obtenues aux résolutions plus faibles ainsi que d'autres variables (valeur du pixel central, intensité moyenne extérieure). Le taux de bonnes réponses (obtenu avec un seuil de 0,5) est de 75 %.

Cette approche, appliquée à un ensemble de lésions représentatif de la variabilité rencontrée, donne des résultats satisfaisants. Elle présente l'avantage de ne pas nécessiter de conditions initiales particulières lors de l'acquisition de l'image (centrage de la lésion, éclairage correct, ...) et s'affranchit en outre, de la plupart des artefacts présents dans l'image (poils, ombres, ...). L'approche multi-échelle permet de sélectionner des contours à différents niveaux de précision selon l'objectif envisagé : localiser la lésion, traiter les informations de contours (Serruys 2002).

3. Couleur

La pigmentation d'une lésion n'est pas uniforme et la présence de plusieurs couleurs est une des caractéristiques prédictives d'un mélanome malin (« variegated colouring »). Pour cela les dermatologues repèrent les plages de couleur homogène (fauve, marron, rouge ou noir ...), et déterminent le nombre de couleurs présentes dans la tumeur.

La cible a été définie ici par le consensus majoritaire (ensemble des votes ≥ 3 : au moins 3 dermatologues en accord) pour la question « Observez-vous plusieurs couleurs ? ». La région d'intérêt couvre l'ensemble de la lésion.

La forme des histogrammes, en RVB, des pixels à l'intérieur de la tumeur est différente selon les modalités « une » ou « plusieurs couleurs » (Figure 4). Nous avons donc sélectionné, comme variable discriminante, les pourcentages d'individus contenus dans chaque classe, chaque histogramme étant découpé en 3 classes (compromis raisonnable entre le nombre de variables à considérer, la résolution et le comptage par classe).

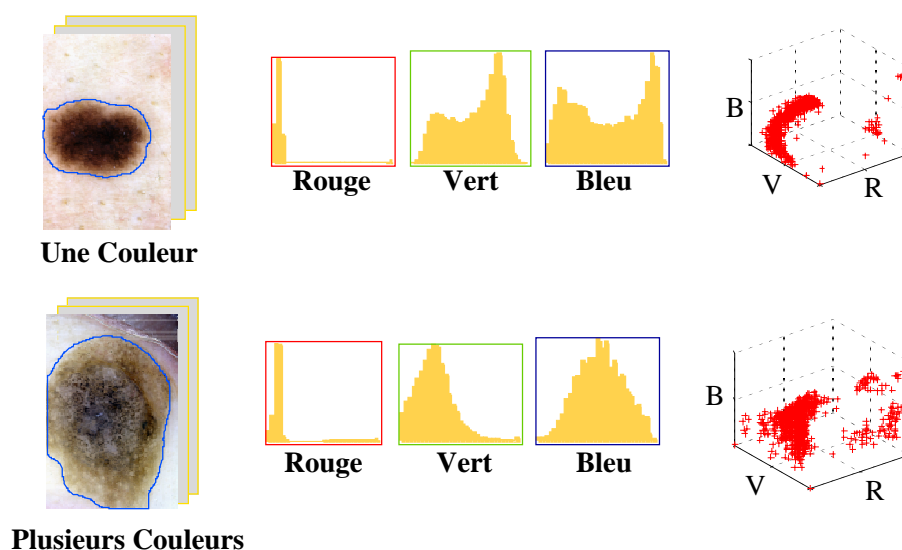


Figure 4 - Distribution des pixels de la lésion selon les 3 plans de couleurs RVB, pour une image classée « une couleur » et une image classée « plusieurs couleurs ».

Traitant ici des images (et non des imagerie que l'on peut aisément obtenir en grand nombre), nous avons appliqué une procédure de « leave-one-out » à deux méthodes de classification : une régression logistique et un réseau de neurones (perceptron multi-couche avec une couche cachée et technique de rétropropagation) pour valider les résultats. 70% des images (régression logistique) et 76 à 79 % (réseaux de neurones) sont bien classées à partir des variables issues des histogrammes RVB.

La répartition des prédictions du système suivant le vote des médecins montre que de façon générale, le classifieur code de façon satisfaisante l'information « plusieurs couleurs » (Figure 5). Le classifieur présente ainsi la même concordance avec le consensus général (Kappa = 0,56 : concordance = 78 %) que chacun des différents experts avec le consensus général (Kappa moyen = 0,53 : concordance moyenne = 77%)

Une approche unifiée pour discriminer les tumeurs noires de la peau

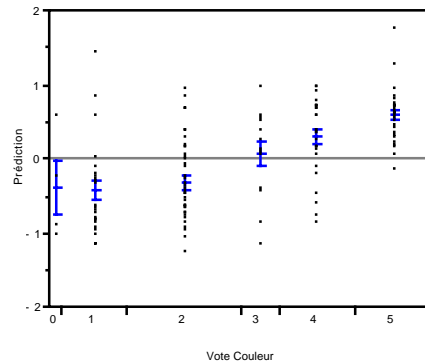


Figure 5 – Distribution des prédictions du système en fonction du nombre de votes des médecins en faveur de la multiplicité de couleurs.

On peut cependant noter que le consensus général entre médecins n'est pas très élevé. Comme la définition de la multiplicité des couleurs peut varier d'un médecin à l'autre et que par conséquent, il n'existe pas de consensus très fort entre les médecins, nous obtenons un classifieur qui, basé sur le consensus majoritaire, fournit une performance correcte (78%) qui pourrait cependant être améliorée. Cette amélioration passe par la construction d'un consensus plus important, soit en tenant compte seulement des images où le consensus est fort (votes 0 et 1 – votes 4 et 5), soit en raffinant l'échelle de couleurs par la prise en compte du nombre de couleurs détectées.

4. Asymétrie

L'asymétrie est une caractéristique essentielle dans la différenciation des tumeurs malignes et des lésions bénignes. Elle est généralement, évaluée par les dermatologues grâce à la comparaison des deux moitiés de la lésion selon l'axe principal de symétrie.

Pour rendre compte de ce critère, réalisé à l'étape de segmentation initiale), les axes de symétrie et le centre de gravité sont calculés à partir du masque de la lésion, en utilisant l'analyse en composantes principales (Figure 6).

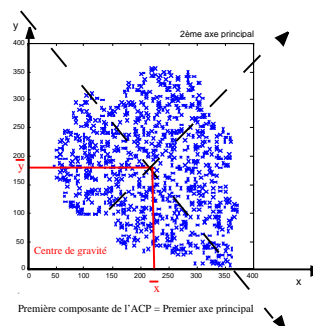


Figure 6 - Illustration des axes principaux.

L'asymétrie est alors codée par 3 indices :

- L'asymétrie centrale ("recouvrement" entre l'aire initiale et l'aire ayant subi une rotation de 180°),
- L'asymétrie axiale autour du 1^{er} axe d'inertie,
- L'asymétrie axiale autour du 2^{ème} axe d'inertie.

La cible est constituée par le consensus des médecins à la question concernant l'asymétrie de la tumeur. Par la régression logistique (avec procédure "leave-one-out"), nous obtenons un taux de bonnes classifications de 69 %.

On peut noter que les erreurs concernant les lésions symétriques déclarées "asymétriques" sont majoritairement commises sur des lésions de petite taille. On peut donc penser que l'indice de symétrie tient compte de la taille de la lésion alors que visuellement, on est moins sensible à l'asymétrie de forme lorsque la lésion est petite, les irrégularités étant globalement moins observables. L'évolution d'une lésion pouvant être caractérisée par l'apparition d'une extension qui entraîne une asymétrie de plus en plus visible, il est intéressant de constater que le critère que nous avons mis en place peut quantifier cette évolution.

5. Aspect général

C'est un critère de suspicion de malignité important mais difficile à formaliser. Aussi, les dermatologues ont parfois du mal à s'accorder sur ce critère. Il inclut, par exemple, aussi bien les irrégularités de forme que les irrégularités de bord.

Pour quantifier les irrégularités de bord, nous nous sommes basés sur les 4 niveaux de segmentation réalisés lors de la phase de segmentation de la tumeur. Au cours de cette tâche, le système a une vision de plus en plus précise des bords. Dans le cas d'un bord régulier (en terme de frontière), la différence entre le contour du niveau 1 et celui des niveaux suivants, sera tout à fait minime. Au contraire, lorsque le bord présente de multiples indentations, le système les voit apparaître au fur et à mesure que la résolution des images augmente (Fig 7).

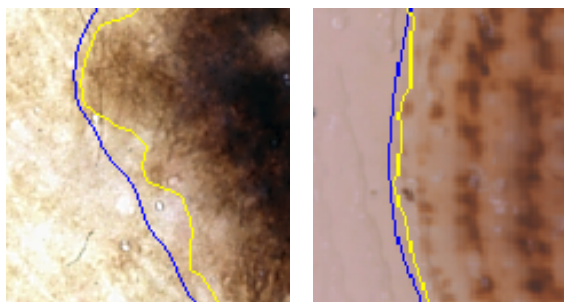


Figure 7 - Exemples (a) de bords avec indentations (b) de bords réguliers

Les attributs considérés sont ainsi les aires et les périmètres de la segmentation (niveau 1 à 4) normalisées entre 0 et 1. La cible est définie comme le consensus majoritaire des réponses des 5 médecins à la question : « les bords de la lésion sont-ils réguliers ? ». La

Une approche unifiée pour discriminer les tumeurs noires de la peau

régression logistique et le réseau de neurones (avec procédure “leave-one-out”) conduisent respectivement à 63 % et 52 % d’images bien classées.

Ce module met en évidence qu’une partie de l’information peut être extraite à partir des différents contours de la segmentation. La classification ne donne cependant pas de bons résultats car la cible fournie ne traduit pas directement l’irrégularité de la forme. Or cette information est cliniquement interprétée de façon globale. D’autres attributs, type dimension fractale, peuvent cependant être appliqués à ce schéma méthodologique.

6. Réseau pigmenté

Nous cherchons ici à reconnaître les zones de réseau pigmenté, signe dermatoscopique, rendant compte d’un maillage plus ou moins marqué de la tumeur permettant notamment de distinguer les lésions mélanocytaires, des tumeurs non-mélanocytaires. La problématique est ici typiquement une analyse de texture, qui doit être conduite à partir d’imagelettes de taille appropriée (Fertil 2001). Les bases d’apprentissage et de test sont ainsi constituées de 22000 (resp. 11 000) imagelettes de 256 pixels obtenues à partir de 28 (resp. 13) images « avec réseau » et 28 (resp. 14) images « sans réseau ».

L’étape de codage est effectuée ici par une projection des imagelettes sur les composantes de l’ACP (transformation de Karhunen-Loeve), obtenues à partir de l’analyse de la base d’apprentissage. Après une étude de robustesse, les 50 premières composantes ont été sélectionnées (Figure 8).

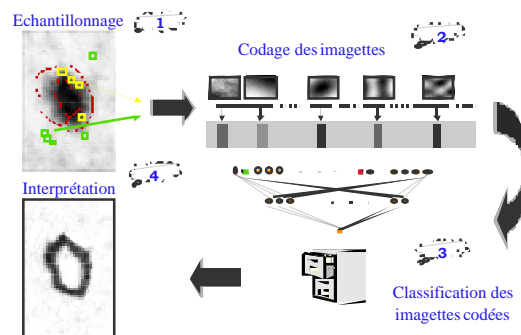


Figure 8 - Schéma général de la reconnaissance d'un signe.

Compte tenu de la complexité de la tâche, plusieurs méthodes de classification ont été utilisées à savoir : la régression logistique, les réseaux neuronaux (perceptron multicouches) et une méthode de classification linéaire sous contrainte de complexité (théorie de l'apprentissage statistique de Vapnik (Vapnik 1995)). Les trois classifieurs donnent des résultats comparables (de 63 à 66 % de bonne classification).

L'étude des erreurs montre que les bords ont tendance à être reconnus comme « réseau ». Cela s'explique par le fait que le réseau pigmenté se situe le plus souvent près des bords et

que le classifieur a du mal à différencier cette structure lorsque les bords sont peu contrastés ou lorsque ce dernier présente des structures autres, telles que les globules ou les pseudopodes.

La sensibilité du classifieur est globalement correcte alors que sa spécificité reste faible, le pouvoir discriminatoire du système n'est donc pas optimal. Il permet cependant, de localiser les zones de réseau pigmenté, nous pouvons ainsi imaginer que les erreurs de classement (les individus "non réseau pigmenté" classés en "réseau pigmenté") pourrait bénéficier d'une étape supplémentaire d'analyse par un deuxième classifieur plus spécialisé.

7. Conclusion

Nous avons proposé une approche qui permet, exclusivement par des techniques adaptatives, de tester et d'évaluer le pouvoir discriminatoire d'une large variété d'attributs pour la reconnaissance des mélanomes. Deux étapes essentielles ont ainsi été formalisées :

- La validation de la base d'images
- La détermination d'attributs pertinents associés à un signe.

La variabilité des lésions, de leurs formes, de leurs couleurs, de leurs textures, la variabilité des structures associées (réseau pigmenté, globules, ...), la présence d'artefacts (poils, ombres, réflexions, ...), et enfin la qualité des images elles-mêmes sont autant de facteurs qui rendent difficiles la tâche d'un classifieur automatique. L'utilisation de méthodes adaptatives a permis de prendre en compte l'ensemble des facteurs de variabilité dont on a pu mesurer au cours de ce travail toute l'importance.

L'évaluation des résultats donne des taux de bonne classification qui dépendent des caractéristiques étudiées mais aussi du degré de consensus entre les médecins qui peut être dégagé pour les caractériser. Cette méthode permet, grâce à ce consensus de s'affranchir, en partie, des variations inter-individuelles et de conclure dans des cas où il n'y a pas de référence absolue.

Bibliographie

- Andreassi, L., R. Perotti, et al. (1999). "Digital Dermoscopy Analysis for the Differentiation of Atypical Nevi and Early Melanoma." *Archives of Dermatology* 135: 1459-1465.
- Binder, M., H. Kittler, et al. (1998). "Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network." *Melanoma Research* 8(3): 261-6.
- Colot, O., D. De Brucq, et al. (1998). an image processing and image analysis method for melanoma detection. 9e forum des jeunes chercheurs, brest
- Ercal, F., A. Chawla, et al. (1994). "Neural network diagnosis of malignant melanoma from color images." *IEEE Trans Biomed Eng* 41(9): 837-45
- Ercal, F., H. C. Lee, et al. (1994). Skin Cancer Diagnosis Using Hierarchical Neural Networks and Fuzzy Systems. *Artificial Neural Networks in Engineering conference*.
- Fertil B., Brahmi D., Cassoux N., Serruys C., Triller R., Le Hoang P. and Giron A. A versatile learning by sample strategy to perform Statistical Pattern classification in biomedical images. *SPIE Electronic Imaging*, 4305, 65-72, (2001).

Une approche unifiée pour discriminer les tumeurs noires de la peau

- Giron A., Serruys C., Triller R., Aitken G., Fertil B.: Analysis of image-based decisions from macroscopic and dermatoscopic examination of small pigmented tumors. 20th World Congress of Dermatology, (2002) Paris.
- Hintz-Madsen, M., L. K. Hansen, et al. (1995). Design and Evaluation of Neural Classifiers Application to Skin Lesion Classification. IEEE Workshop on Neural Networks for Signal Processing (NNSP'95).
- Kjoelen, A., M. J. Thompson, et al. (1995). "Performance of AI Methods in Detecting Melanoma." IEEE Transactions on Biomedical Engineering 41(9): 837-845.
- Lefevre, E., O. Colot, et al. (2000). Knowledge modeling methods in the framework of evidence theory. an experimental comparison for melanoma detection. Proceedings of SMC'2000, IEEE International Conference on Systems, Man, and Cybernetics, Nashville, Tennessee, USA
- Schindewolf, T., W. Stolz, et al. (1993). "Classification of melanocytic lesions with color and texture analysis using digital image processing." Anal Quant Cytol Histol 15(1):1-11.
- Sboner, A., E. Blanzieri, et al. (2001). A Knowledge Based System for Early Melanoma Diagnosis Support. Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP'2001), London, UK.
- Serruys C., Giron A., Triller R., Fertil B.: Steps toward the automatic detection of melanoma : feature recognition. 20th World Congress of Dermatology, (2002), Paris.
- Vapnik, V. N. The Nature of Statistical Learning Theory, Springer-Verlag. (1995).
- Wolff, I. H., J. Smolle, et al. "Sensitivity in the clinical diagnosis of malignant melanoma." Melanoma Research 8(5): 425-9. (1998).
- Zhang, Z., W. V. Stoecker, et al. "Border detection on digitized skin tumor images." IEEE Trans Med Imaging 19(11): 1128-43(2000).

Summary

An adaptive and supervised approach has been applied to images of melanoma in order to segment lesions and detect features of malignancy. Dermatologists were asked to label and provide information on regions of interest and consensus between dermatologists has been used as targets for the supervised classification (neural network, logistic regression ..). Tasks as different as segmentation of tumor, study of global features (coloring, symmetry) or detection of local structures (pigmented networks) have been evaluated.

Remerciements

Tous nos remerciements à la Ligue Nationale Contre le Cancer, à la Société Française de Dermatologie et à la société KXEN pour leurs contributions à ce projet.