

Une méthode de partition des images de lésions mélanocytes cutanées

Valentina Cozza¹, Mario R. Guarracino², Rosanna Verde³

1. *Dip. di Matem. e Statistica, Univ. di Napoli ‘Federico II’, Via Cinthia, 80126 Napoli, Italie*

2. *ICAR-CNR, Via Pietro Castellino 111, 80131 Napoli, Italie*

3. *Dip. di Studi Europei e Mediterranei, Seconda Università di Napoli, Via del Setificio 15,
San Leucio, 81100 Caserta, Italie*

valentina.cozza@unina.it, mario.guarracino@na.icar.cnr.it, rosanna.verde@unina2.it

Mots clés : classification, images bio-médicales, descripteurs multi-valeurs

1 Introduction

Dans ce papier nous proposons de utiliser une méthode de classification de type dynamique ([7]; [5]), sur des images de lésions cutanées comme outil de diagnostique du mélanome malin. Le mélanome malin est une forme de cancer de la peau qui se développe à partir des mélanocytes et qui a été observé en forte augmentation ces dernières années (5000-6000 cas par an en France seulement). Seuls le diagnostic précoce et l’intervention chirurgical permettent d’envisager une guérison définitive. La différenciation entre le mélanome et les autres lésions pigmentées de la peau n’est pas triviale, même pour un dermatologue expérimenté. Cette problématique a naturellement attiré l’attention de nombreux chercheurs qui ont proposé différents systèmes pour la détection semi-automatique du mélanome (l’analyse discriminante, réseaux neuronales, *support vector machine*, ...) et pour l’individuation des caractéristiques qui sont les plus discriminantes. Les procédures conventionnelles pour reconnaître la nature maligne d’une tache pigmentée [2] sont :

- i. la règle ABCD pour analyser: l’Asymétrie (A), le Bord (B), la Couleur (C), la Structure Différentielle (D) ;
- ii. les méthodes d’analyse de patterns (globales et locaux) ;
- iii. la méthode de Menzies ;
- iv. la liste de contrôle de 7-points.

La première approche fournit une description des images des lésions par rapport à des mesures d’asymétrie calculées pour le bord, la couleur et les structures dermatoscopiques, à la présence des zones pigmentées qui s’interrompent brusquement au bord ou qui se dégradent lentement, au degré de la couleur des lésions et à l’existence d’un voile clair, aux mesures des différents composants de la structure (réseaux pigmentés, points, globules, zones sans structure et stries).

Les méthodes d’analyse de patterns se concentrent sur les caractéristiques de la structure de la lésion au niveau global et local. La méthode de Menzies focalise l’attentions sur les principaux indicateurs de négativité et positivité. La dernière méthode est basée sur une liste de contrôle à 7-points liés à des atypies dans le réseau pigmenté, à la présence du voile, à des zones vascularisés atypiques, à des stries, points, globules ou taches irrégulières, à une régression dans la structure de la lésion.

En général, dans le problème d’identification automatique ou semi-automatique de forme du mélanome malin, les caractéristiques qui sont prises en compte font référence aux

quatre composants de la règle ABCD: Asymétrie, Bord, Couleur, Structure Différentielle. Dans ce contexte nous considérons deux indicateurs pour l'asymétrie de la forme et de la couleur; deux indicateurs pour l'irrégularité et le degré de minceur du bord, des indicateurs de la distribution des trois couleurs (Rouge, Vert et Bleu), de l'Intensité, de la Luminosité et du Contraste, des indicateurs pour l'irrégularité de la Structure Différentielle (présence/absence, localisation et couleur de pigments, de points, de globules et de stries).

Du fait de la structure complexe des images à analyser et de l'incertitude souvent présente dans la mesure de la majeure partie des caractéristiques et attributs de cette structure (qualité de la photo, subjectivité dans l'assignation d'un point, ...) auxquelles s'accompagne la grande multiplicité de manifestations dermatoscopiques, nous proposons de décrire chaque image par des valeurs multiples des différentes mesures choisies comme descripteurs, selon le modèle des données symboliques [1]. À partir de ce type de description des lésions, nous pouvons utiliser une méthode de classification de type dynamique pour l'identification semi-automatique de deux typologie de lésions: mélanomes malins et lésions bénignes. L'objectif est de obtenir différentes classes de lésions malignes et bénignes (utilisant, pour les différencier, une information fournie par un expert) et de focaliser l'attention sur les classes des deux ensembles qui présentent une forte similarité. L'idée est de découvrir les attributs des différentes caractéristiques qui induisent à des erreurs d'attribution et sont les principales causes de diagnostics faussement positifs où incorrects.

2 Description symbolique des images

Une description symbolique [1] d'une image de lésion est définie à partir d'un ensemble de descripteurs de différente nature (qualitatifs, quantitatifs, catégoriques avec distribution associée) qui peuvent assumer des valeurs multiples (*multi-catégories, intervalles, modaux*). Nous choisissons un ensemble de descripteurs qui représentent des caractéristiques cliniques incluses dans la règle de l'ABCD. En particulier, l'asymétrie de la forme, de la structure et de la couleur par rapport aux deux axes principaux d'inertie. Car l'asymétrie de la structure réticulaire est plutôt une mesure liée à la régularité du réseau, elle sera prise en compte dans la description de la Structure Différentielle. L'asymétrie des différentes composantes est mesurée en termes des intervalles de valeurs $[min, max]$ sur les deux axes. Pour l'asymétrie de la forme d'une lésion m on considère le rapport entre les aires du quadrant droit A_{1d} et gauche A_{1g} , par rapport au premier axe et du quadrant supérieur A_{2s} et inférieur A_{2i} , par rapport au deuxième axe:

$$asy_F(m) = [min\{\frac{A_{1d}}{A_{1g}}, \frac{A_{2s}}{A_{2i}}\}, max\{\frac{A_{1d}}{A_{1g}}, \frac{A_{2s}}{A_{2i}}\}]. \quad (1)$$

Pour évaluer l'asymétrie de la couleur, nous considérons la partition de l'image en n cercles concentriques autour du centre des axes d'inertie. Le long des quatre directions des axes nous associons des histogrammes relatifs aux intensités de tons de gris. Soit I_{uv} l'intensité du ton de gris présent dans le cercle v dans le demi-plan $u = d, g, i, s$, une mesure de l'asymétrie de la couleur par rapport aux deux axes est donnée par:

$$K_r = \sum_{v=1,n} (I_{u_1,v} - I_{u_2,n-v+1})^2; \quad u_r = \{d, g\}, \{i, s\} \text{ et } r = 1, 2$$

De la même façon que pour l'asymétrie de la forme, le descripteur de l'asymétrie de la couleur prend en compte les intervalles des valeurs minimales et maximales de K_1 et K_2 .

Le Bord, ou contour, en littérature est mesuré par le diamètre plus grand, l'aire, l'irrégularité du bord, le rapport de minceur, l'index de circularité, la variance des distances des points sur le bord du barycentre. Considérant une partition de l'image de la lésion en 8 sections radiales, avec des angles à 45°, on calcule la mesure d'inertie pour chacune des α configurations obtenues en faisant pivoter de θ° (par exemple $\theta = 5^\circ$) la configuration autour au centre et en prenant les valeurs minimales et maximales.

L'épaisseur du Bord identifie une interruption brusque de la lésion. On considère donc une bande de dimension fixée le long du bord et par rapport aux 8 sections radiales, on calcule le gradient d'intensité des tons de gris, du centre vers le bord dans la direction orthogonale au bord. De manière plus simple, la description du bord peut être exprimée par les valeur minimales et maximales du gradient. Autrement on pourra considerer la distribution des gradients. Le descripteur sera, dans ce cas, de type modal.

Les caractéristiques de la Couleur se basent sur la distribution de la palette des trois couleurs (Rouge, Vert et Bleu) et sur les distributions des valeurs de l'Intensité, de la Luminosité et du Contraste. Nous utilisons des descripteurs de type *modal* qui représentent ces caractéristiques par la distribution correspondante de chaque image.

L'irrégularité de la Structure Différentielle est considérée, dans la littérature, comme un indicateur de haut risque de lésions malignes. Les descripteurs pris en considération sont les caractéristiques relatives aux aspects globaux: la présence d'un réseau pigmenté, de points, de globules et de stries ramifiées. Les points et les globules sont décrits par des attributs multi-catégoriques ordinaux comprenant leurs *absence* ou *présence*, leur forme *circulaire* ou *ovale*, leur couleur *noire*, *gris* ou *marron* et leur localisation *centrale* ou *périphérique*. Les stries sont décrites en fonction de leur *régularité* ou *irrégularité*. Pour tous les descripteurs considérés, des valeurs élevées correspondent à une caractérisation maligne ou à haut risque et des valeurs basses caractérisent de lésions bénignes.

3 Algorithme de Classification Dynamique

Nous proposons, pour la partition de l'ensemble E des images des lésions en k classes, un algorithme de type Nuées Dynamiques ([7], [4]) généralisé au cas de données multi-valuers. Dans son schéma classique, cet algorithme recherche une partition P^* de E en k classes non vides et un vecteur L^* de k prototypes $(g_1, \dots, g_i, \dots, g_k)$ qui représente, au mieux, par rapport à un critère Δ , les k classes $(C_1, \dots, C_i, \dots, C_k)$ de la partition P^* :

$$\Delta(P^*, L^*) = \text{Min}\{\Delta(P, L) \mid P \in P_k, L \in \Lambda_k\}.$$

avec : P_k l'ensemble des partitions de E en k classes non vides et Λ_k l'espace de représentation des prototypes. Ce critère exprime l'adéquation entre la partition P et le vecteur L des k prototypes. Il est défini comme la somme sur toutes les classes C_i et sur tous les objets m de C_i des mesures de proximités $\delta(x_m, G_i)$ entre chaque vecteur x_m de description (intervalles, distributions, multi-catégoriques) de m et le vecteur de description G_i du prototype g_i de la classe C_i . Les mesures $\delta(\cdot)$ sont supposées différents par rapport au type de descripteur. La principale mesure de proximité dans le cas de descripteurs à intervalles a été proposée par [3]. De Carvalho [6] a proposé un mesure à deux composantes pour les descripteurs multi-catégoriques. Une mesure L_2 entre distributions a été introduite par [5]. La mesure globale de dissimilarité entre m et g_i est obtenue par une combinaison linéaire des mesures de proximité choisies en fonction de différents descripteurs. L'algorithme procède alternativement par une étape de représentation suivie d'une étape d'allocation. Au fin de tenir séparées les classes de lésions bénignes de quelles malignes, un contrant supervisé est introduit dans la procédure.

4 Les caractéristiques discriminantes

A partir des classes appartenant aux deux groupes (lésions bénignes et malignes), issues dans l'algorithme de classification automatique, nous calculons les dissimilarités entre les prototypes respectifs: g_i et $g_{i'}$, avec $i \neq i'$) Nous sommes intéressés par l'évaluation des dissimilarités entre les prototypes de classes de lésions bénignes et de lésions malignes. Les valeurs de la mesure de dissimilarité considérée sont calculées au niveau global par rapport à tous les descripteurs pris en comptes dans notre analyse, avec un système de pondération initialement arbitrer. Afin de tenir en compte le pouvoir discriminant des différents descripteurs dans la procédure semi-automatiques, nous proposons d'introduire un système de pondérations sur les quatre caractéristiques principales A, B, C, D et sur les leurs composants:

$$\delta(g_i, g_{i'}) = \sum_{t=1}^4 \mu_j \sum_{j \in W_t} \nu_j(G_i^j, G_{i'}^j) \quad (\text{avec } \sum_{j \in W_t} \mu_j = 1) \quad (2)$$

où : W_t est l'ensemble des descripteurs appartenant aux caractéristiques A, B, C et D. Les points sont donc recherchés de façon optimale, par la maximisation d'un critère de meilleur séparation de classes.

5 Conclusions et perspectives

Notre contribution a été l'introduction d'un schéma de classification non supervisé dans le contexte de la détection d'images médicales. Sur le modèle utilisé pour l'analyse des données complexes nous avons proposé un ensemble de descripteurs qui font référence aux caractéristiques classiques dans ce type d'analyse, selon la règle ABCD, avec comme différence de leur associer des mesures opportunes et d'assumer des descripteurs à valeurs multiples. Une proposition de sélection des caractéristiques les plus discriminantes de classes de lésions bénignes et malignes est également suggéré. Le présent travail est encore en cours de développement et les résultats seront comparés avec ceux obtenus par des méthodes classiques (par ex. *support vector machine*).

6 Références bibliographiques

- [1] H.-H. Bock, E. Diday, *Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data*. Springer, 2000.
- [2] I. Maglogiannis, D. I. Kosmopoulos, "Computational vision systems for the detection of malignant melanoma", *Oncology reports* n. 15 , 2006, p. 1027-1032.
- [3] M. Chavent, "Analyse des Données Symboliques. Une méthode divisive de classification", *Thèse de l'Université de PARIS-IX Dauphine, 1997*.
- [4] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier, H. Ralambondrainy, *Classification Automatique des Données*. Bordas, 1989, Paris.
- [5] M. Chavent, F. A. T. De Carvalho, Y. Lechevallier, R. Verde, "Trois nouvelles méthodes de classification automatique de données symboliques de type intervalle", *Revue de Statistique Appliquées* n. 4 , 2003, p. 5-29.
- [6] F.A.T. de Carvalho, "Extension based proximities between Boolean symbolic objects", in *Data Science, Classification and Related Methods*, Hayashi, C. et al.(eds.), Springer-Verlag, 1998, Tokyo, 370-378.
- [7] E. Diday, "Le méthode des Nuées dynamique", *Revue de Statistique Appliquée* vol. 19, n. 2, 1971, p. 19-34.