

Influence des caractéristiques issues de la matrice de co-occurrences sur le diagnostic du cancer du sein

Atef BOUJELBEN*, Hedi TMAR*, Jameleddine M'NIF⁺, Mohamed ABID*

* CES-Ecole Nationale des Ingénieurs de Sfax , Tunisie

+Service de Radiologie et d'Imagerie Médicale, Hôpital Habib Bourguiba, Sfax, Tunisie

atef_boujelben@yahoo.fr,tmar@iuplo.univ-ubs.fr,jaameledine.mnif@rns.tn,

mohamed.abid@enis.rnu.tn

RESUME: Dans le cadre de l'analyse automatique des clichés mammographiques, la distinction entre microcalcification du sein, maligne ou bénigne, est un problème réputé complexe dont la faisabilité reste encore à prouver. Nous présentons dans cet article un descripteur de texture afin de classifier des clichés mammographiques. Nous testons l'influence de chaque caractéristique issue de la matrice de co-occurrences sur une analyse locale des régions d'intérêts "Regions Of Interest"(ROIs). En utilisant la base de test DDSM, la sensibilité peut atteindre 90% pour un classifieur type MLP, quant à la spécificité, elle peut atteindre 90% pour un classifieur type Bayes.

MOTS CLES: Analyse, Mammographie, Matrice de co-occurrences, Vecteur de caractéristiques

1 Introduction

Le cancer du sein est très fréquent; c'est la première cause de mortalité par cancer chez les femmes[1]. Récemment, le dépistage mammographique assisté par ordinateur a pris une grande importance[4] grâce à sa rapidité, sa consistance et sa capacité de fournir des solutions fiables pour la détection des lésions. Mais, bien que la mammographie assistée par ordinateur a été étudiée depuis ces deux dernières décennies, l'interprétation automatique des microcalcifications reste une tâche très difficile.

En effet, les problèmes d'interprétation des tissus issue du sein sont en relation avec la détection, identification des ROIs, analyse et classification de ces ROIs. Dans ce papier on se concentre à la phase d'analyse. En effet, l'analyse par texture peut se faire par des méthodes fréquentielles [5] [9] ou des méthodes statistiques [3] [7] [8] [10] [12]. Les méthodes statistiques, qui étudient les relations entre un pixel et ses voisins, semblent être les plus significatifs pour décrire un motif aléatoire. Pour cela dans la référence de Christoyianni et al. [3], où les régions sont définis par des radiologues, l'analyse statistique a été faite par deux types de statistiques dont l'une de premier ordre et l'autre du deuxième ordre qui est la matrice de co-occurrences. Dans ce travail, Christoyianni et al. font une comparaison de trois vecteurs caractéristiques (un vecteur pour chaque angle: 0,-45,45) de dix sept caractéristiques chacun. Par contre Lambrou et al. [7] se basent sur l'extraction de vingt caractéristiques de premier ordre, deuxième ordre (en particulier celle basée sur la matrice de co-occurrences) et de longueur de niveau de gris. D'autres travaux se basent sur les matrices de co-occurrences comme celui de Zwiglar et Denton [12] qui présentent une approche, basée sur la translation et la transportation de mesure, pour la sélection de sous ensemble de matrices de co-occurrences et celui de Mudigonda et al. [8] qui extraient un vecteur de vingt caractéristiques décomposé de cinq caractéristiques pour chaque angle (0, 45, 90, 135 et distance=1).

Dans ce travail, on propose de tester l'apport de chaque caractéristique issue de la matrice de co-occurrences sur la qualité de l'analyse tout en illustrant quatre angles. Puis on teste l'apport de la moyenne de chaque caractéristique (en 4 orientations) sur l'analyse en utilisant un vecteur d'ensemble. Le but général de ce travail est d'éliminer le calcul des caractéristiques qui n'influent pas sur la performance du diagnostic; et par conséquent augmenter le rapport temps sur qualité.

Ce papier est organisé comme suit : La section suivante présente l'étape d'extraction des caractéristiques. Les résultats expérimentaux et les interprétations seront résumés dans la section trois. On termine ce papier par une conclusion et quelques perspectives.

2 Extraction des caractéristiques

Le flot adopté illustre le parcours d'une image mammographique dès son acquisition jusqu'à la prise de décision. A partir de l'image d'entrée on localise les ROIs en utilisant une méthode de segmentation, après, on extrait les

caractéristiques en utilisant une méthode statistique. Ces caractéristiques serviront des entrées à des classificateurs pour enfin avoir une décision et/ou aide à la décision interprété dans la section 3.

Notre but revient à analyser localement les ROIs issues des images mammographiques. Pour cela avant de passer à l'étape d'extraction des vecteurs caractéristiques, on présente une méthode de segmentation qui permet de localiser les ROIs.

2.1 Identification des régions d'intérêts

L'étape d'identification des ROIs est décomposée en deux parties. En premier temps une étape de prétraitement de l'image scannée pour améliorer l'image bruitée : cette étape est faite par le filtre médian. En deuxième temps, après la sélection de la région douteuse par le radiologue, on procède par une méthode de croissance de région traduite afin d'isoler la région d'intérêt du reste du sous-image. L'étape d'analyse décrite dans ce qui suit se fait sur la région isolée.

2.2 Descripteur de texture

La méthode d'analyse par texture utilisée est celle des matrices de co-occurrences. En effet, on extrait à partir de la région à analyser (après identification des ROIs) une matrice de niveaux de gris. A partir de cette matrice on calcule les vecteurs caractéristiques. Le flot de l'analyse est représenté dans la figure 1.

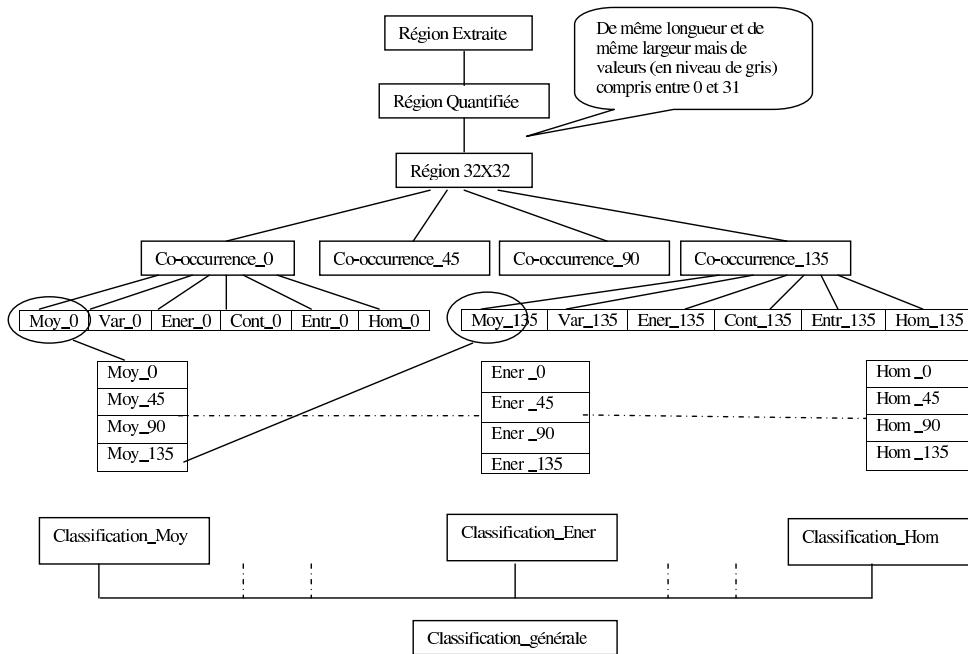


FIG. 1 – *Les étapes d'analyses adoptées*

- Après l'étape de segmentation, illustrée dans la section précédente, on obtient une région d'intérêt: une matrice en niveau de gris.
- Cette matrice est quantifiée; c'est-à-dire, on obtient une autre matrice de même longueur et de même largeur mais de valeur en niveau de gris entre 0 et 31 (les valeurs étaient entre 0 et le nombre maximum en niveau de gris). On obtient donc une région de taille 32X32 (cette valeur est prise en tenant compte de la taille de la tumeur). Le but de cette étape est la réduction du calcul des matrices de co-occurrences.
- On calcule ensuite les matrices de co-occurrences de cette région (co-occurrence_0, co-occurrence_45, co-occurrence_90 et co-occurrence_135 dont chacune a une taille 32X32).
- Pour chaque matrice, on calcule les six caractéristiques : moyenne, variance, énergie, contraste, entropie et homogénéité qui sont représentés respectivement.

$$Moyenne = \sum_i \sum_j p(i,j) \quad (1)$$

Où $p(i, j)$ correspond aux éléments de la matrice de co-occurrences .

$$Variance = \sum_i \sum_j (i - moy)^2 p(i,j) \quad (2)$$

$$Energie = \sum_i \sum_j p(i,j)^2 \quad (3)$$

$$Contraste = \sum_i \sum_j (i - j)^2 p(i,j) \quad (4)$$

$$Entropie = - \sum_i \sum_j p(i,j) \log p(i,j) \quad (5)$$

$$Homogeneite = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i,j) \quad (6)$$

- Les vecteurs caractéristiques obtenues sont répartis de la façon suivante : six vecteurs de quatre caractéristiques chacun (une même caractéristique pour les quatre angles) et un vecteur d'ensemble qui est constitué de six valeur (chaque valeur est la moyenne d'une caractéristique en quatre orientations).

Les vecteurs caractéristiques générés suite à cette phase d'analyse vont servir à une phase de décision qu'on va détailler dans le section suivante.

3 Résultats Expérimentaux

Les résultats expérimentaux, qui consistent à évaluer la spécificité et la sensibilité, sont obtenues en développant un outil en Java qui contient les phases segmentation, analyse et classification. Mais avant de passer à l'étape de décision et/ou aide à la décision, on va décrire la base de test utilisée.

3.1 Base d'images de test

Pour évaluer la performance de la méthode d'analyse basée sur l'extraction des vecteurs caractéristiques, on a utilisé une base de données universelle DDSM¹. Cette base contient 2620 cas décomposés en 15 volumes malades, 16 bénignes et 12 normales. Dans la pratique, on a travaillé sur une partie qui contient 80 ROIs pris au hasard; c'est à dire de volumes différents. La base adoptée contient 35 ROIs malades, 25 ROIs bénignes et 20 ROIs normaux issues de la base DDSM.

3.2 Classification expérimentale

Pour classifier les vecteurs issus de chaque caractéristique de la matrice de co-occurrences et du vecteur total, on a utilisé trois classificateurs Kpp_v (K plus proches voisins), Bayes et MLP². Les résultats sont donnés dans le tableau 1.

classifieur	sensibilité			spécificité		
	Kpp_v	Bayes	MLP	Kpp_v	Bayes	MLP
Moyenne	40%	25 %	90 %	65 %	90 %	30 %
Variance	40%	25%	90 %	60 %	90 %	25%
Energie	50%	30%	90 %	65 %	75 %	25 %
Contraste	40%	75%	90 %	65 %	25 %	25 %
Entropie	50%	75%	90 %	65 %	25 %	25 %
Homogénéité	60%	30%	75 %	65 %	25 %	30 %
Vecteur d'ensemble	60%	30%	75 %	75 %	30 %	50 %

TAB. 1 – *Résultats des tests des vecteurs caractéristiques*

Les meilleures résultats de cette étude sont obtenues avec le classifieur Kpp_v en utilisant l'homogénéité et le vecteur d'ensemble. Tandis que pour le Bayes et le MLP, on voit une alternance des résultats des méthodes de classifications;

1. Digital Database for Screening Mammography

2. Multi Layer Perceptron

c'est à dire qui a la meilleure spécificité n'a pas la meilleure sensibilité et vice versa. Ce qui explique l'utilisation d'un seul critère dans la littérature. Par exemple, Vachier et al. [11] et Joseph et al. [6] illustrent la sensibilité qui peut aboutir à 73,5% pour le premier en utilisant une base de données locale et 90% pour le deuxième en adoptant une base de données DDSM. Par contre Mudigonda et al. [8] illustrent la spécificité qui vaut 82,1% en combinant la base de données MIAS³ et une base locale. De même pour l'outil de ImageChecker Mammographic[2], il donne la main à l'utilisateur pour régler un seuil de détection donné. Ce réglage dépend des objectifs de lecture selon que la sensibilité ou la spécificité soit privilégiée.

4 Conclusion

Dans ce papier, on s'est intéressé à l'optimisation de la phase d'analyse. Pour aboutir à cette phase on a procédé par la méthode des matrices de co-occurrences. Cette étape est utilisée pour initialiser un processus décisionnel destiné à révéler la présence ou l'absence des microcalcifications sur les clichés numérisés et ceci par l'intervention d'un radiologue. Les résultats trouvés varient entre 90% et 25%. Cependant, la phase d'analyse peut comporter d'autres caractéristiques comme le Skewness et le Kurtosis qui peuvent mieux converger les résultats. En plus la phase de détection automatique des microcalcifications reste une tâche importante pour améliorer la qualité de détection surtout en ciblant les variétés des types de microcalcifications.

Références

- [1] www.gyneweb.fr/Sources/senologie/depistsein.htm.
- [2] www.r2tech.com/mammography/home/index.php.
- [3] I. Christoyianni, E. Dermatas, and G. Kokkinakis. Neural classification of abnormal tissue in digital mammography using statistical features of the texture. In *Proceedings of The 6th IEEE International Conference on Electronics Circuits and Systems*, pages 117–120, Université de patras, 1999.
- [4] H. Chuong, D. Hailey, R. Warburton, J. H. MacGregor, E. D. Pisano, and J. Joyce. La mammographie numérique comparativement à la mammographie sur film avec écran : évaluation technique, clinique et économique. Technical report, Office canadien de coordination de l'évaluation des technologies de la santé, octobre 2002.
- [5] C. B. R. Ferreira and D. L. Borges. Automated mammogram classification using a multiresolution pattern recognition approach. In *Proceedings of the XIV Brazilian Symposium on Computer Graphics and Image Processing "SIBGRAPI 01"* IEEE, 2001.
- [6] Y. Joseph, M. Gavrielides, M. K. Markey, and L. Jonathan. Computer-aided classification of breast microcalcification clusters: Merging of features from image processing and radiologists. *Medical Imaging, Proceedings of the SPIE*, 5032:882–889, 2003.
- [7] T. Lambrou, A. D. Linney, R. D. Speller, and A. Todd-Pokropek. Statistical classification of digital mammograms using features from the spatial and wavelet domains. *Medical Image Understanding and Analysis*, 3, july 2002.
- [8] R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels. Gradient and texture analysis for the classification of mammographic masses. *IEEE Transactions On Medical imaging*, 19:459–464, octobre 2000.
- [9] R. Rangayyan, R. Ferrari, J. Desautels, and A. Frère. Optimal segmentation of mammographic images. In *Proceedings of the XIII Brazilian Symposium on Computer Graphics and Image Processing SIBGRAPI.00*, 2000.
- [10] N. Székely, N. Toth, and B. Pataki. A hybrid system for detecting masses in mammographic images. In *Instrumentation and Measurement Technology Conference Como Italy*, IMTC, Mai 2004.
- [11] C. Vachier, S. Muller, and S. Bothorel. Analyse morphologique des clichés mammographiques en vue de la détection des opacités du sein. In *Mathematics and Image Analysis*, Paris, septembre 2000.
- [12] R. Zwigelaar and E. .R.Denton. Optimal segmentation of mammographic images. In *International Workshop on Digital Mammography*, Norwich, UK, 2004.

3. Mammographic Image Analysis Society