

Comparaison des mammographies par des méthodes d'apprentissage

Irina Diana Coman*, Djamel Abdelkader Zighed**

*Laboratoire ERIC, 5 avenue Pierre Mendès-France, 69676 Bron, France
irina.coman@gmail.com
<http://eric.univ-lyon2.fr>

**Laboratoire ERIC, 5 avenue Pierre Mendès-France, 69676 Bron, France
zighed@univ-lyon2.fr
<http://eric.univ-lyon2.fr>

1 Introduction et état de l'art

Les mammographies sont le moyen le plus répandu pour la détection du cancer du sein. Des études ont démontré qu'une lecture double des mammographies augmente la sensibilité du diagnostic jusqu'à 15% (Bird et al., 1992) et de plus, (Destounis et al., 2004), que les outils d'aide au diagnostic automatique du cancer du sein (ADACS) peuvent améliorer même les résultats d'une double lecture des mammographies. Pourtant, les techniques existantes d'ADACS ont une série d'inconvénients.

Les méthodes existantes d'ADACS peuvent être classées dans deux classes : celles qui essaient d'identifier des signes de cancer (Pluim et al., 2003) et celles qui essaient une classification des mammographies (Zaiane et al., 2002). Dans le premier cas, les inconvénients principaux sont le coût, dû au traitement d'images et à la classification des signes trouvés et le fait que les signes des phases de début de cancer sont plus subtiles que ceux recherchés par ces méthodes. Dans le deuxième cas, l'inconvénient principal est le taux de réussite plus faible (qui baisse parfois jusqu'à 56,25%) et sa forte variation selon le jeu de données considéré pour la validation.

2 Contribution

Par rapport aux autres techniques existantes dans le domaine d'ADACS, notre approche est basée sur les techniques des médecins et se propose d'utiliser moins le traitement d'images et plus des techniques d'apprentissage automatique afin d'obtenir une classification des clichés dans deux classes : symétriques et non symétriques.

Dans une première étape, nous comparons des zones des clichés pour obtenir une mesure quantitative de la similitude. Pour obtenir les zones, nous avons proposé trois méthodes issues de la pratique des médecins, plus précisément la méthode maillage (qui propose une segmentation tenant compte de la structure de symétrie du sein), la méthode fenêtre (qui représente un balayage vertical ou horizontal de l'image avec une zone de taille fixe) et la méthode rideau (qui propose un traitement progressif du cliché, soit en direction verticale soit horizontale).

Comparaison des mammographies

Le graphique des différences entre les zones des deux clichés, met en évidence, par des piques, les éventuelles asymétries. La hauteur des piques est une mesure de la taille des différences, tandis que la largeur des piques est une mesure de la localisation. Tenant compte du fait que les dissemblances naturelles sont normalement répandues sur une zone plus large, en temps que les asymétries dues au cancer sont plus localisées, nous avons choisi de prendre en compte la hauteur et la largeur des plus grands piques pour l'étape suivante de classification des clichés dans les deux classes : symétriques et non symétriques. Pour la classification nous avons utilisé les arbres de décision (Breiman et al., 1984), plus précisément des arbres C4.5(Quinlan, 1993).

Nous avons testé les trois méthodes sur un jeu de données de 202 couples de clichés, en utilisant 73% de données pour l'apprentissage et le reste de 27% pour la validation. La méthode rideau a eu un taux de réussite de 62% et la méthode maillage un taux de réussite de 68%. Les meilleurs résultats ont été obtenus par la méthode fenêtre, avec un taux de réussite de 70%.

Conclusions et perspectives Les résultats que nous avons obtenu, (un taux de réussite de 70%), sont des résultats préliminaires. Tenant compte du fait que l'approche proposée est nouvelle dans le domaine et aussi du fait que nous avons eu une base de données spécialement annotées seulement sur les aspects de cancer, nous considérons les résultats encourageants.

Nous envisageons de continuer les tests sur une base de données mieux annotée, spécialement sur les aspects d'asymétrie. Nous envisageons aussi d'effectuer plusieurs tests afin de trouver les meilleurs paramètres des trois méthodes, la meilleure représentation des clichés et les méthodes de classification les plus adéquates.

Références

- Bird, R., T. Wallace, et B. Yankaskas (1992). Analysis of cancers missed at screening mammography. *Radiology* 184, 613–617.
- Breiman, L., J. Friedman, R. Olshen, et C. Stone (1984). *Classification and Regression Trees*. Wadsworth.
- Destounis, S. V., P. DiNitto, W. Logan-Young, E. Bonaccio, M. L. Zuley, Kathleen, et M. Willison (2004). Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate?initial experience. *Radiology* 232, 578–584.
- Pluim, J. P. W., J. A. Maintz, et M. A. Viergever (2003). Mutual information based registration of medical images: A survey. *IEEE Transactions On Medical Imaging*.
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Zaiane, O. R., M.-L. Antonie, et A. Coman (2002). Mammography classification by an association rule-based classifier. *International Workshop on Multimedia Data Mining*.

Summary

This article proposes a new approach in the field of computer aided diagnosis of breast cancer, based on clinical practice and using bilateral comparative analysis and machine learning techniques in order to classify mammograms as symmetrical or asymmetrical.