

Contexte

De nos jours, la mammographie est la méthode la plus efficace pour le dépistage du cancer du sein. Depuis le début des années 2000, un outil d'assurance qualité des rapports d'examen de mammographies a été adopté : BI-RADS (Breast Imaging-Reporting and Data System). Cependant, malgré cet effort de standardisation, près de 70% des biopsies demandées suite aux mammographies s'avèrent inutiles.

Objectif

L'objectif de ce projet est de proposer un processus d'aide à la décision pour accompagner le praticien dans son diagnostic du cancer du sein.

Ce processus d'aide à la décision sera conçu à partir de l'étude d'une base de données de patientes en mettant en application différentes méthodes de classification étudiées en cours.

Livrable

Un rapport comportant au minimum :

- Une description de la problématique
- Une description de la démarche d'analyse de données mise en œuvre
- Une description des méthodes de classification utilisées (a minima celles vues en cours : KNN, Naives Bayes, Arbre de décision)
- Une analyse comparative des résultats obtenus
- En perspective : proposition d'un processus d'aide à la décision

Dates

29 janvier 2014 : distribution du sujet

12 février 2014 : rendu du rapport au format pdf. (examenUE4_nom_prenom.pdf) à envoyer par mail à gerard.drays@mines-ales.fr et delahunay.elisabeth@gmail.com.

Données

Source : <http://archive.ics.uci.edu/ml/datasets/Mammographic+Mass>

1. Title: Mammographic Mass Data

2. Sources:

(a) Original owners of database:

Prof. Dr. Rüdiger Schulz-Wendtland
Institute of Radiology, Gynaecological Radiology, University
Erlangen-Nuremberg
Universitätsstraße 21-23
91054 Erlangen, Germany

(b) Donor of database:

Matthias Elter
Fraunhofer Institute for Integrated Circuits (IIS)
Image Processing and Medical Engineering Department (BMT)
Am Wolfsmantel 33
91058 Erlangen, Germany
matthias.elter@iis.fraunhofer.de
(49) 9131-7767327

(c) Date received: October 2007

3. Past Usage:

M. Elter, R. Schulz-Wendtland and T. Wittenberg (2007)
The prediction of breast cancer biopsy outcomes using two CAD
approaches that both emphasize an intelligible decision process.
Medical Physics 34(11), pp. 4164-4172

4. Relevant Information:

Mammography is the most effective method for breast cancer screening available today. However, the low positive predictive value of breast biopsy resulting from mammogram interpretation leads to approximately 70% unnecessary biopsies with benign outcomes. To reduce the high number of unnecessary breast biopsies, several computer-aided diagnosis (CAD) systems have been proposed in the last years. These systems help physicians in their decision to perform a breast biopsy on a suspicious

lesion seen in a mammogram or to perform a short term follow-up examination instead.

This data set can be used to predict the severity (benign or malignant) of a mammographic mass lesion from BI-RADS attributes and the patient's age.

It contains a BI-RADS assessment, the patient's age and three BI-RADS attributes

together with the ground truth (the severity field) for 516 benign and 445 malignant masses that have been identified on full field digital mammograms

collected at the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006.

Each instance has an associated BI-RADS assessment ranging from 1 (definitely benign)

to 5 (highly suggestive of malignancy) assigned in a double-review process by

physicians. Assuming that all cases with BI-RADS assessments greater or equal

UE4 - EXAMEN FOUILLE DE DONNEES 2014

a given value (varying from 1 to 5), are malignant and the other cases benign, sensitivities and associated specificities can be calculated. These can be an indication of how well a CAD system performs compared to the radiologists.

5. Number of Instances: 961

6. Number of Attributes: 6 (1 goal field, 1 non-predictive, 4 predictive attributes)

7. Attribute Information:

1. BI-RADS assessment: 0 to 6

* 0: Incomplete

* 1: Negative

* 2: Benign finding(s)

* 3: Probably benign

* 4: Suspicious abnormality

* 5: Highly suggestive of malignancy

* 6: Known biopsy - proven malignancy

2. Age: patient's age in years

3. Shape: mass shape: round oval lobular irregular

4. Margin: mass margin: circumscribed microlobulated obscured ill-defined spiculated

5. Density: mass density high iso low fat-containing

6. Severity: benign or malignant

8. Missing Attribute Values: Yes

- BI-RADS assessment: 2

- Age: 5

- Shape: 31

- Margin: 48

- Density: 76

- Severity: 0

9. Class Distribution: benign: 516; malignant: 445