



Center for Machine Learning and Intelligent Systems

[About](#) [Citation Policy](#) [Donate a Data Set](#)
[Contact](#)

☒ Repository ☐ Web



[View ALL Data Sets](#)

Mammographic Mass Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.

Data Set Characteristics:	Multivariate	Number of Instances:	961	Area:	Life
Attribute Characteristics:	Integer	Number of Attributes:	6	Date Donated	2007-10-29
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	4365

Source:

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](mailto:matthias.elter@iis.fraunhofer.de)
 (49) 9131-7767327

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

Data Set 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 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.

Class Distribution: benign: 516; malignant: 445

Attribute Information:

6 Attributes in total (1 goal field, 1 non-predictive, 4 predictive attributes)

1. BI-RADS assessment: 1 to 5 (ordinal, non-predictive!)
2. Age: patient's age in years (integer)
3. Shape: mass shape: round=1 oval=2 lobular=3 irregular=4 (nominal)
4. Margin: mass margin: circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5 (nominal)
5. Density: mass density high=1 iso=2 low=3 fat-containing=4 (ordinal)
6. Severity: benign=0 or malignant=1 (binominal, goal field!)

Missing Attribute Values:

- BI-RADS assessment: 2
- Age: 5
- Shape: 31
- Margin: 48
- Density: 76
- Severity: 0

Relevant Papers:

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

Papers That Cite This Data Set¹:



N/A

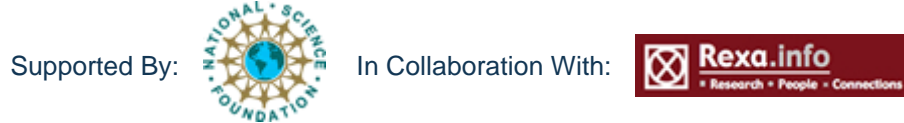
Citation Request:

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

[1] Papers were automatically harvested and associated with this data set, in collaboration with Rexa.info



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)