Predicting Malignancy from Mammography Findings and Image Guided Core Biopsies

2nd Breast Cancer Workshop 2015 – April 7th 2015
Porto, Portugal

Pedro Ferreira

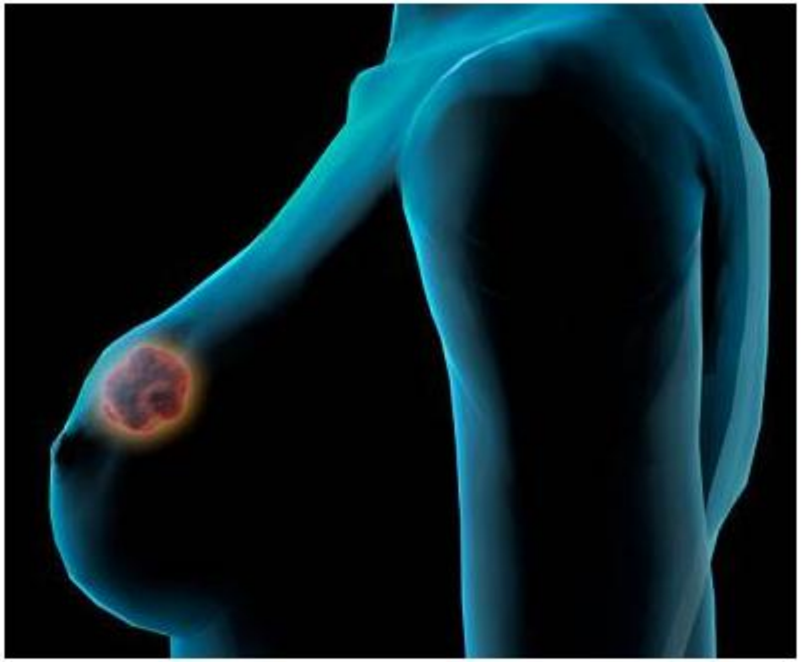
Nuno A. Fonseca
Inês Dutra
Ryan Woods
Elizabeth Burnside

CRACS INESCTEC
FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
U.PORTO
WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

Outline

- Breast Cancer
- Objectives
- Dataset
- Methodology
- Results and Analysis
- *MammoClass* (online tool)
- Conclusions

Breast Cancer



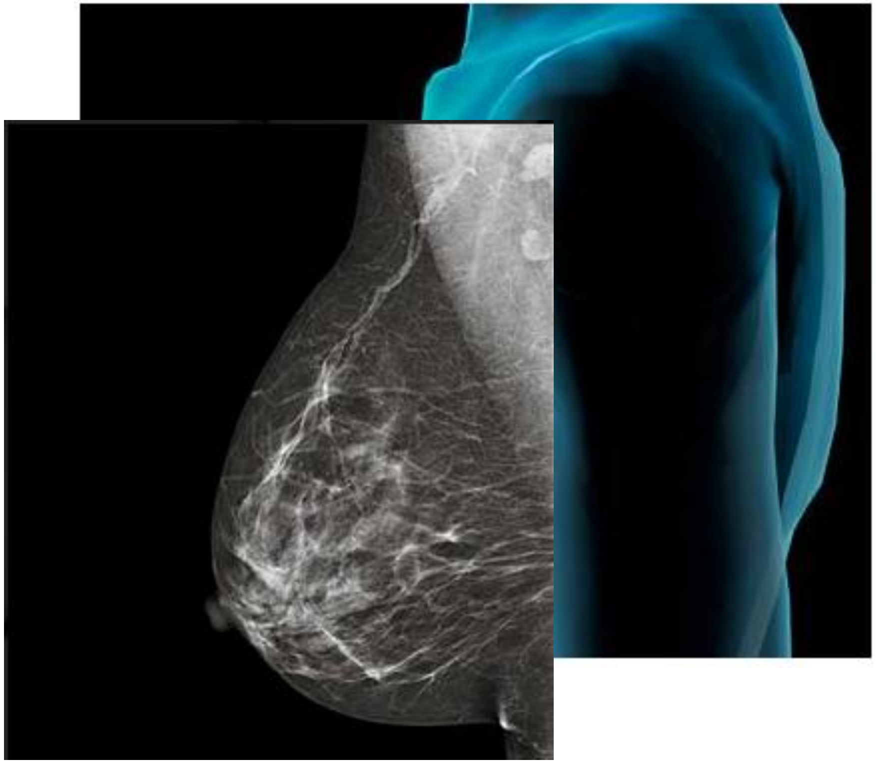
- USA:
  - About 1 in 8 women (≈ 12%) will develop invasive breast cancer during lifetime
  - In 2014:
    - 232.670 invasive cancers
    - 40.000 (≈ 17%) expected to die

  Source: *U. S. Breast Cancer Stats.* – accessed April 2015

- Portugal:
  - About 1 in 11 women (≈ 9%) will develop invasive breast cancer during lifetime
  - Per year:
    - 5600 new cases
    - 1500 deaths (27%)

  Source: *Laço Association* – accessed April 2015

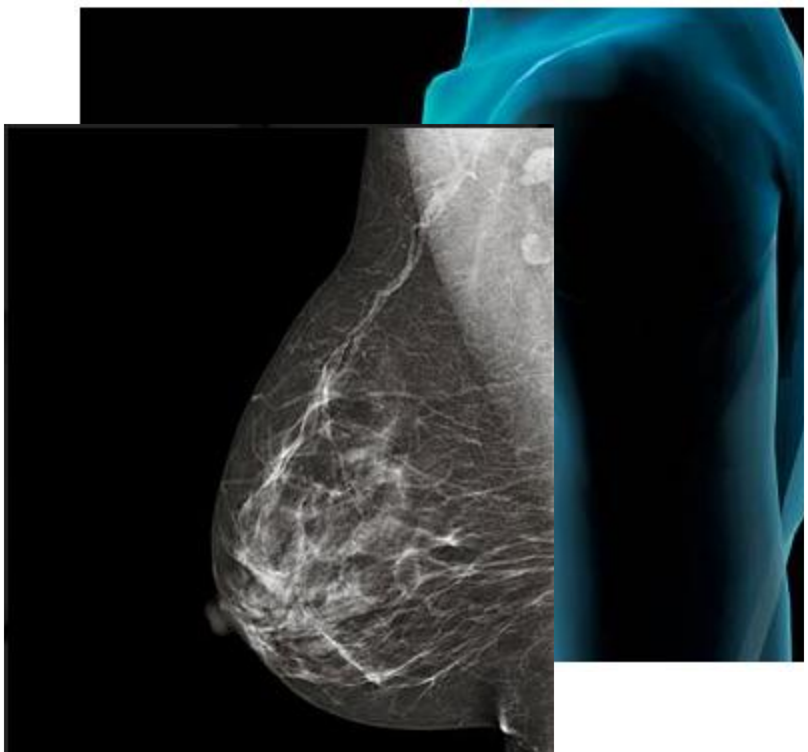
Breast Screening Programs

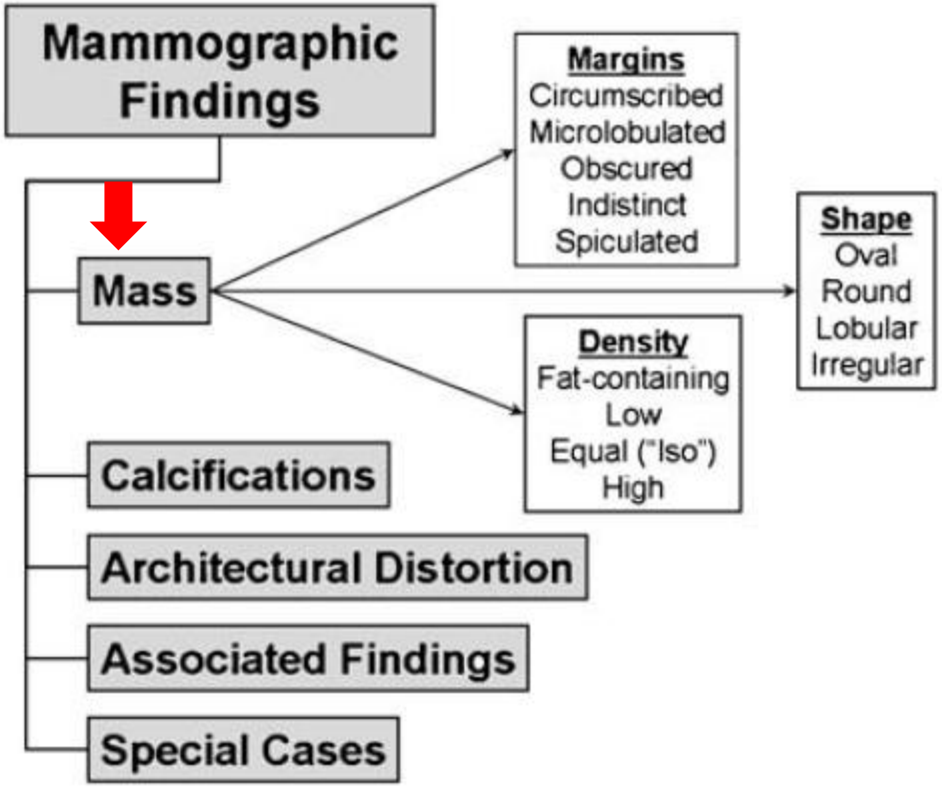


- Reduction of death rate in 30%
- **Mammography**:

  The cheapest and most eficient method to detect cancer in a preclinical stage

Mamography - BI-RADS® Descriptors





Outline

- Breast Cancer
- Objectives
- Dataset
- Methodology
- Results and Analysis
- *MammoClass* (online tool)
- Conclusions

- in P. Ferreira, et al., "**Studying the relevance of Breast Imaging Features**", in Proc. International Conference on Health Informatics (HEALTHINF), 2011.

Objectives

- Build classifiers capable of predicting **mass density** and **malignancy** from a reduced set of mammography findings

- Reduce the number of unnecessary biopsies

CRACS INESCTEC

Outline

- Breast Cancer
- Objectives
- Dataset
- Methodology
- Results and Analysis
- *MammoClass* (online tool)
- Conclusions

Dataset

- Source:

- 348 cases
- Each case refers to a breast nodule **retrospectively** classified according to BI-RADS® system
- From mammographies results
- Collected between October 2005 and December 2007

Attributes

13 attributes

age_at_mammo
CLOCKFACE_LOCATION_OR_REGION
MASS_SHAPE
MASS_MARGINS
SIDE
DEPTH
MASS_MARGINS_worst
QUADRANT_LOCATION_def
SIZE
OVERALL_BREAST_COMPOSITION
Density_num
retro_density
outcome_num

118 (33.9%)
malignant (+)

230 (66.1%)
benign (-)

Masses classification

Prospective

- **Classification** of feature **mass density just by one radiologist**:
  - low density;
  - iso-dense;
  - high density;
- **Brief** and superficial medical **report** (at the time of imaging);
- **Classification under stress**.

mass density

density_num

Retrospective

- **Classification** by a **group of experienced physicians** that **re-assess** all **exams**;
- **Review of mass density classification** made by radiologist (prospective study);
- **Classification without stress**;
- **Reference standard** for **mass density**.

mass density

retro_density

CRACS INESCTEC

Masses classification

348 cases (**retrospectively** classified)

180

(≈ 52%)

(**prospectively** classified)

168

(≈ 48%)

Outline

- Breast Cancer
- Objectives
- Dataset
- Methodology
- Results and Analysis
- *MammoClass* (online tool)
- Conclusions

Methodology

- WEKA
- Paired Corrected T-Tester
  - **Significance level: 0.05**

348

180 — 168

10 x strat. c. v. — **test**

Methodology – Experiments

<u>**10 x stratified. c. v.**</u>

180

- **E_1** – Predicting malignancy with *retro_density*
- **E_2** – Predicting malignancy with *density_num*
- **E_3** – Predicting malignancy without mass density

- **E_4** – Predicting *retro_density**
- **E_5** – Predicting *density_num**

* in all experiments the *low* and *iso* densities were merged into a single class

CRACS INESCTEC

Methodology – Algorithms applied

- ZeroR (baseline classifier)
- OneR
- DTNB
- PART

rules

- J48
- DecisionStump
- RandomForest
- SimpleCart
- NBTree

trees

- NaiveBayes
- BayesNet (TAN)

bayes

- SMO

functions

internal parameter variation

Results

348

180 168

10 x strat. c.v. **test**

CRACS INESCTEC

Results – Experiments

10 x stratified. c. v.

180

Predicting malignancy with retro_density

Predicting retro_density

Exp.	Algorithm	CCI	K	F	AUROC
E1	SMO	**85.6**±7.3	**0.69**±0.16	**0.80**±0.11	**0.84**±0.08
E1	DTNB	81.6±8.2	0.60±0.18	0.74±0.13	0.88±0.07
E1	NaiveBayes	81.3±9.5	0.61±0.20	0.76±0.12	0.88±0.08
E1	J48	80.7±9.3	0.59±0.20	0.75±0.13	0.79±0.11
E2	SMO	**83.9**±7.7	**0.66**±0.17	**0.78**±0.11	**0.82**±0.08
E2	NaiveBayes	80.3±9.3	0.59±0.19	0.75±0.12	0.87±0.09
E2	DTNB	79.8±9.5	0.56±0.21	0.72±0.15	0.86±0.09
E2	J48	75.4±9.5	0.47±0.21	0.65±0.15	0.73±0.12
E3	SMO	**83.8**±7.7	**0.65**±0.17	**0.78**±0.11	**0.82**±0.09
E3	J48	76.3±9.9	0.49±0.22	0.67±0.15	0.76±0.13
E3	NaiveBayes	76.2±9.9	0.51±0.20	0.71±0.13	0.85±0.09
E3	DTNB	75.7±9.0	0.48±0.19	0.67±0.13	**0.81**±0.10
E4	SMO	**81.3**±8.2	**0.52**±0.21	**0.64**±0.17	**0.75**±0.11
E4	J48	74.4±8.8	0.32±0.24	0.47±0.21	0.67±0.15
E4	DTNB	73.5±10.0	0.34±0.24	0.51±0.19	**0.76**±0.12
E4	NaiveBayes	72.8±9.9	0.37±0.23	0.56±0.18	0.77±0.11
E5	NaiveBayes	**67.2**±12.1	**0.33**±0.25	**0.62**±0.15	**0.72**±0.14
E5	SMO	**66.8**±10.7	**0.31**±0.22	0.55±0.16	0.65±0.11
E5	J48	63.6±10.1	0.26±0.21	0.56±0.15	0.62±0.13
E5	DTNB	62.1±11.9	0.22±0.24	0.54±0.16	0.64±0.14

Results – Experiments

Predicting density

180

Results - Experiments

10 x stratified. c. v.

- E_4 – Predicting *retro_density*

SVM's

CCI: 81.3% (+/-8.2)

Sens: 0.57 (+/- 0.20)

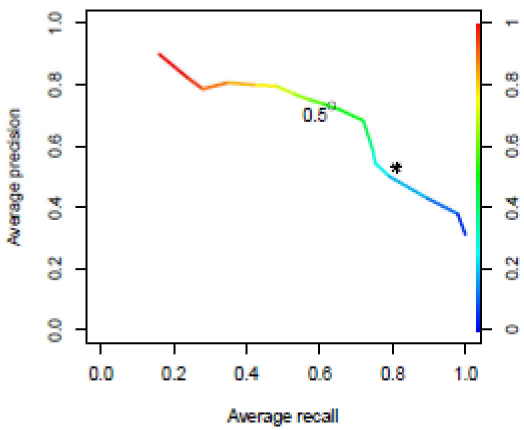
Spec: 0.92 (+/- 0.07)

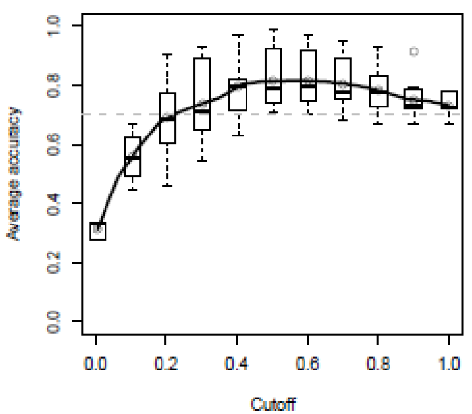
F: 0.64 (+/- 0.17)

Radiologist's accuracy = 70 %

Classifier ≈ 81 %

high density





168

Results - Experiments

TEST

- **E_6** – Predicting *retro_density*

(model E_4 applied)

SVM's

CCI: 84.5%

Sens: 0.57

Spec: 0.90

F: 0.55

180

SVM's

CCI: 81.3% (+/-8.2)

Sens: 0.57 (+/- 0.20)

Spec: 0.92 (+/- 0.07)

F: 0.64 (+/- 0.17)

Results – Experiments

Predicting malignancy

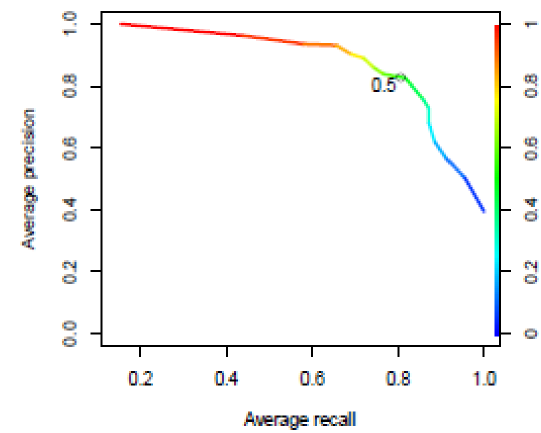
180

Results - Experiments

<u>10 x stratified. c. v.</u>

- **E_1** – Predicting malignancy with *retro_density*

malignant



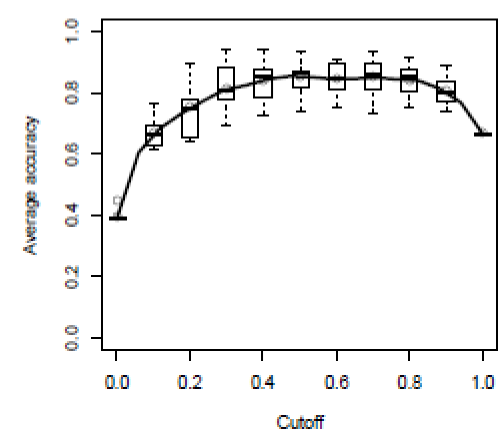
SVM's

CCI: 85.6% (+/-7.3)

Sens: 0.78 (+/- 0.15)

Spec: 0.91 (+/- 0.07)

F: 0.80 (+/- 0.11)



168

Results - Experiments

180

SVM's

CCI: 85.6% (+/-7.3)

Sens: 0.78 (+/- 0.15)

Spec: 0.91 (+/- 0.07)

F: 0.80 (+/- 0.11)

TEST

- **E_8** – Predicting malignancy with *retro_density*

(model E_1 applied)

SVM's

CCI: 81.0%
Sens: 0.57
Spec: 0.90
F: 0.63

with **real** values of **retro_density**

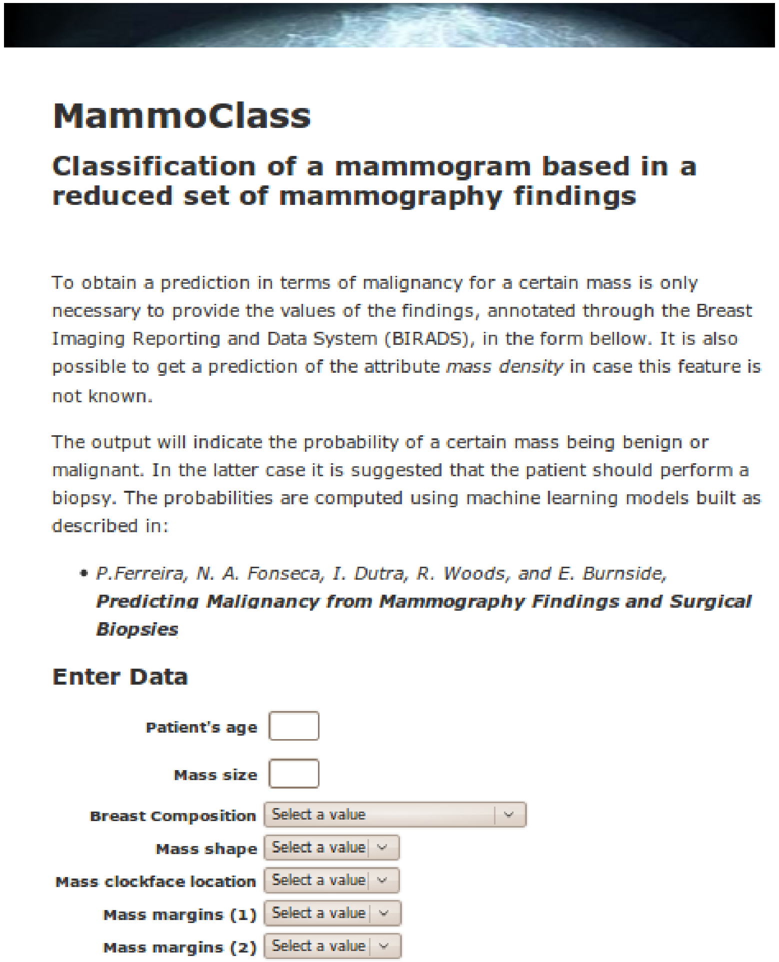
SVM's

CCI: 80.4%
Sens: 0.57
Spec: 0.89
F: 0.62

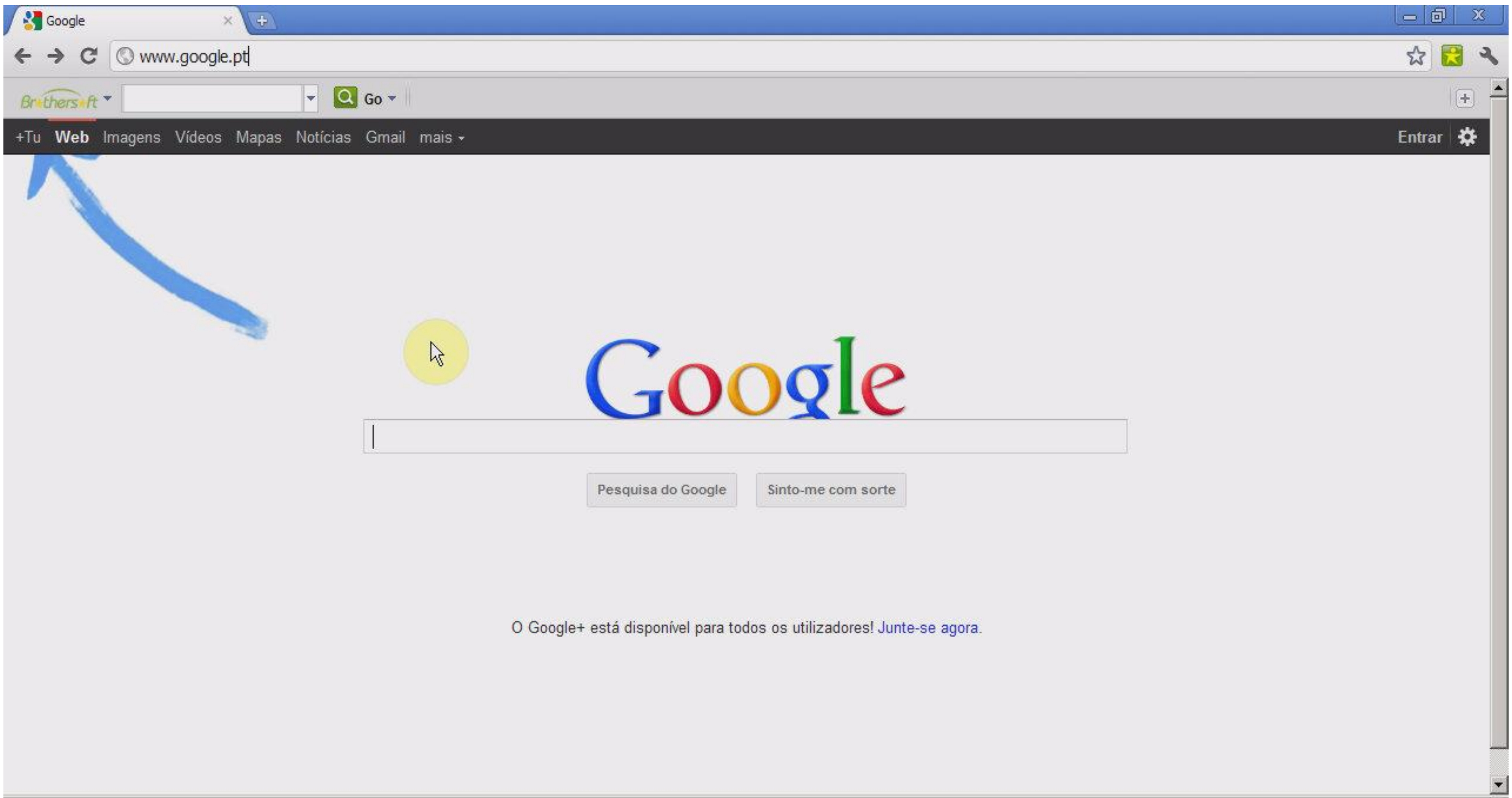
with **predicted** values of **retro_density** by classifier **E_6**

MammoClass

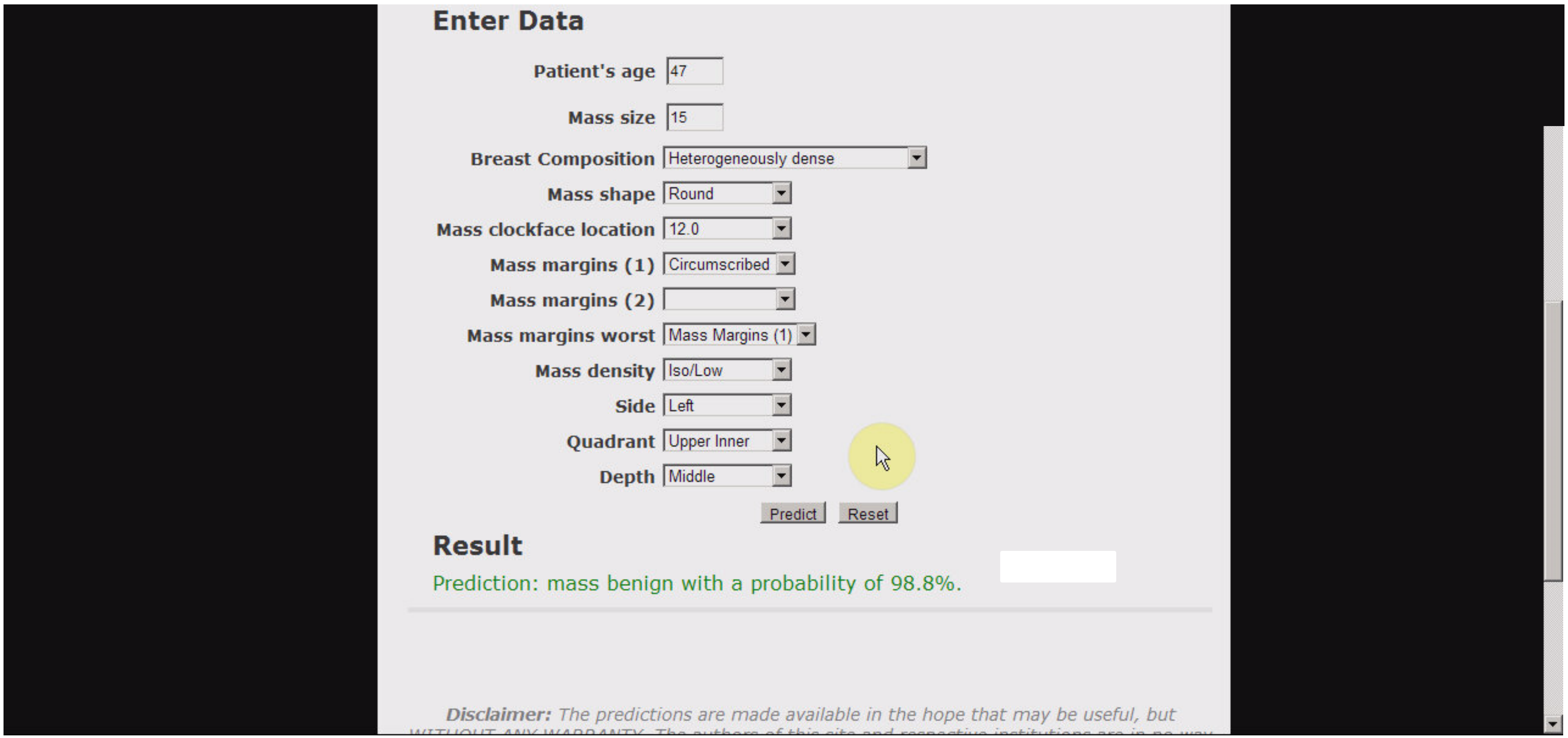
- Online tool freely available at:
  - http://cracs.fc.up.pt/mammoclass/



MammoClass


Google
www.google.pt
Go
+Tu Web Imagens Vídeos Mapas Notícias Gmail mais
Entrar
Google
Pesquisa do Google
Sinto-me com sorte
O Google+ está disponível para todos os utilizadores! Junte-se agora.

MammoClass


Enter Data
Patient's age
47
Mass size
15
Breast Composition
Heterogeneously dense
Mass shape
Round
Mass clockface location
12.0
Mass margins (1)
Circumscribed
Mass margins (2)
Mass margins worst
Mass Margins (1)
Mass density
Iso/Low
Side
Left
Quadrant
Upper Inner
Depth
Middle
Predict
Reset
Result
Prediction: mass benign with a probability of 98.8%.
Disclaimer: The predictions are made available in the hope that may be useful, but

CRACS INESCTEC

Conclusions

a) We built **models** (integrated in *MammoClass*) that **predict malignancy and mass density** based on mammography findings;

b) Machine learning **classifiers** to **predict mass density** may **aid radiologists** during the prospective mass classification;

c) One of our classifiers can **predict malignancy even in the absence of mass density**, since we can **fill up** this **attribute** using our **mass density predictor**.

Thank you!

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

http://cracs.fc.up.pt/mammoclass

pedroferreira@dcc.fc.up.pt
nunofonseca@acm.org
ines@dcc.fc.up.pt
rwoods@gmail.com
eburnside@uwhealth.org

Appendices

State of the Art

- R. Woods, et al., **“Validation of Results from Knowledge Discovery: Mass Density as a Predictor of Breast Cancer”**, in J Digit Imaging, pp. 418-419, 2009.
- R. Woods, et al., **“The Mammographic Density of a Mass is a Significant Predictor of Breast Cancer”**, in Radiology, 2010.
- P. Ferreira, et al., **“Studying the relevance of Breast Imaging Features”**, in Proc. of the International Conference on Health Informatics (HEALTHINF), 2011.

- **Conclusions:**

  mass density can be an **important** attribute when predicting **malignancy**

State of the Art

- 4 datasets
- Main target of study:
  - **Breast Cancer**

State of the Art

- W. H. Wolberg & O. L. Mangasarian, **“Multisurface method of pattern separation for medical diagnosis applied to breast cytology”**, in Proc. of the National Academy of Sciences, 87, pp. 9193-9196, 1990.

- Y. Wu, et al, **“Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer”**, in Radiology 187:81-87, 1993.

- H. A. Abbass, **“An evolutionary artificial neural networks approach for breast cancer diagnosis”**, in Articial Intelligence in Medicine 25:265, 2002.

- W. N. Street, et al, **“An Inductive Learning Approach to Prognostic Prediction”**, in ICML, p. 522, 1995.

- T. Ayer, et al, **“Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration”**, in Cancer 116(14):3310-3321, 2010

Data distribution

- 348

348	*retro_density*		Total
outcome_num	***high***	***iso***	
malignant	59 (70.2%)	59 (22.3%)	**118 (33.9%)**
benign	25 (29.8%)	205 (77.7%)	**230 (66.1%)**
Total	**84 (24.1%)**	**264 (75.9%)**	

Data distribution

- 180

180	*retro_density*		Total
outcome_num	***high***	***iso***	
malignant	42 (75.0%)	29 (23.4%)	**71 (39.4%)**
benign	14 (25.0%)	95 (76.6%)	**109 (60.6%)**
Total	**56 (31.1%)**	**124 (68.9%)**	

180	*density_num*		Total
outcome_num	***high***	***iso***	
malignant	51 (63.0%)	20 (20.2%)	**71 (39.4%)**
benign	30 (37.0%)	79 (79.8%)	**109 (60.6%)**
Total	**81 (45.0%)**	**99 (55.0%)**	

Data distribution

- 168

168	*retro_density*		Total
outcome_num	*high*	*iso*	
malignant	17 (60.7%)	30 (21.4%)	**47 (28.0%)**
benign	11 (39.3%)	110 (78.6%)	**121 (72.0%)**
Total	**28 (16.7%)**	**140 (83.3%)**	