# Predicting Malignancy from Mammography Findings and Surgical Biopsies

BIBM 2011 – November 13th 2011 – Atlanta, USA

CRACS
INESCPORTO LA

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

U.PORTO

Pedro Miguel Ferreira

Nuno A. Fonseca
Inês Dutra
Ryan Woods
Elizabeth Burnside

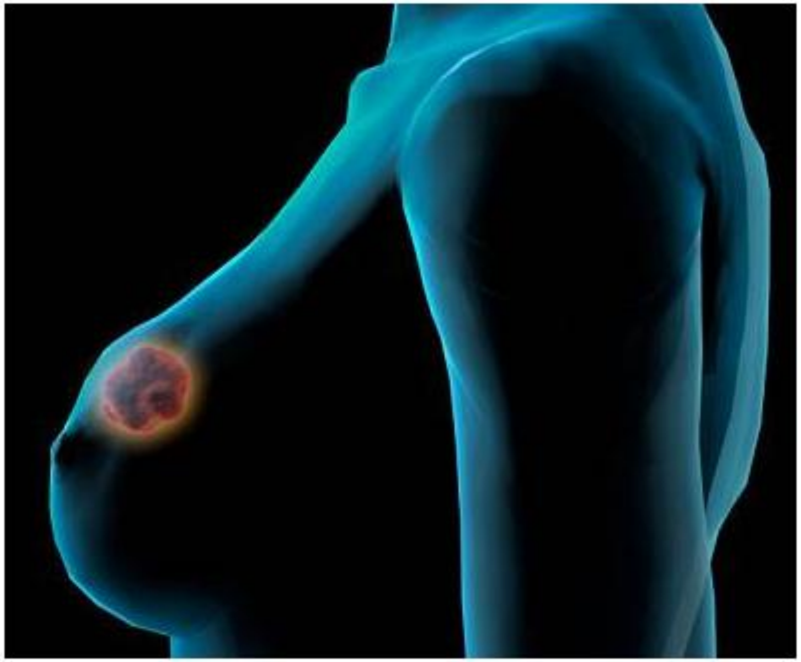
# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- *MammoClass* (online application)
- Conclusions and Future Work

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- *MammoClass* (online application)
- Conclusions and Future Work

# Breast Cancer



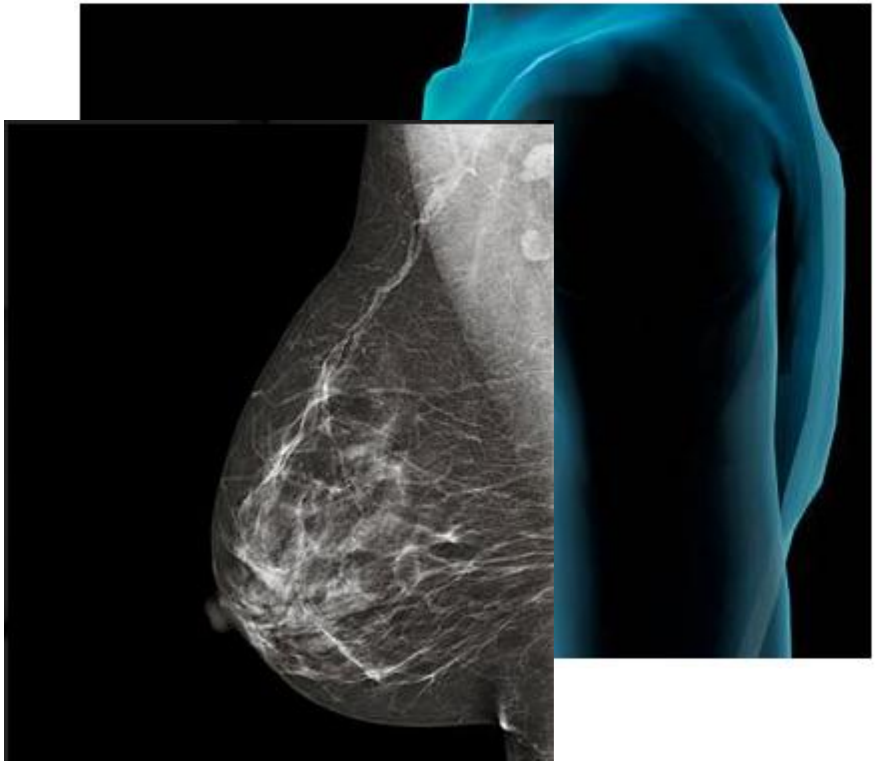
- USA:
  - 1 woman dies of breast cancer every 13 minutes
  - In 2011:
    - 230.480 invasive cancers
    - 39.520 (≈ 17%) expected to die

Source: *U. S. Breast Cancer Statistics* – October 2011

- Portugal:
  - Per year:
    - 4500 new cases
    - 1500 deaths (33%)

Source: *Liga Portuguesa Contra o Cancro* – November 2011

# Breast Screening Programs

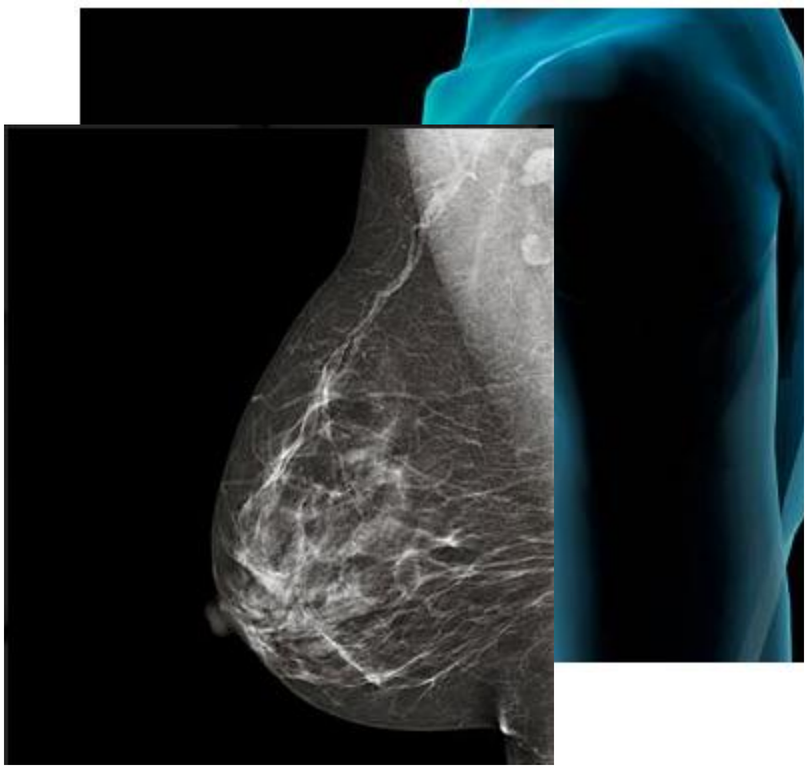


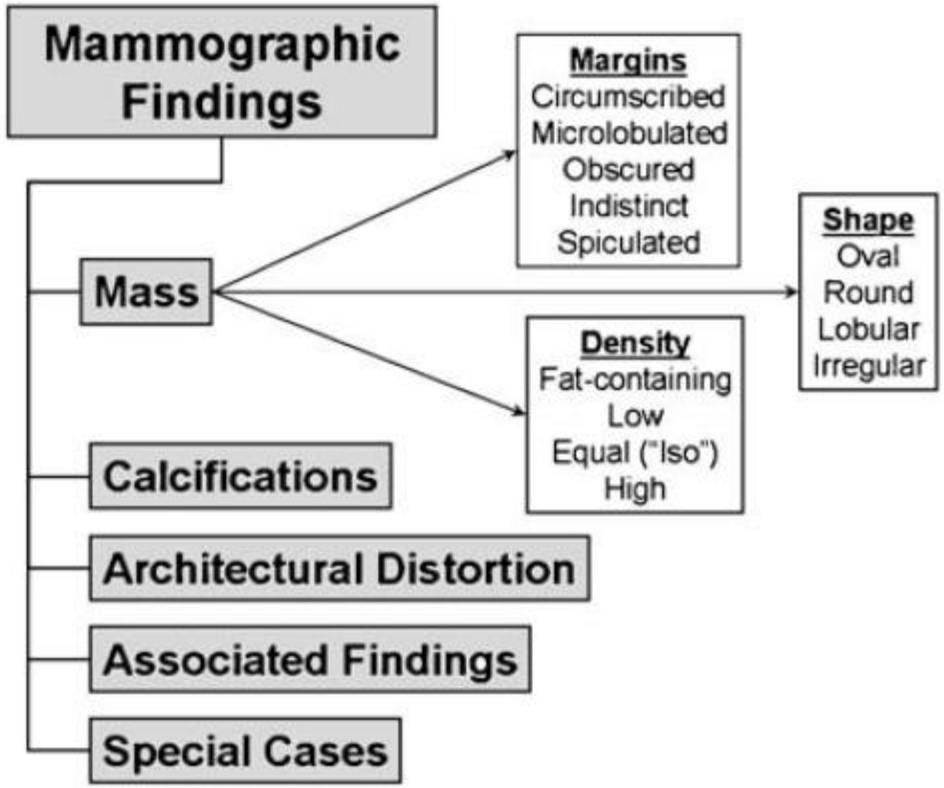
- Reduction of death rate in 30%

- **Mammography**:

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

# Mammography





# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- *MammoClass* (online application)
- Conclusions and Future Work

in *Studying the relevance of Breast Imaging Features* – 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 INESC PORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- *MammoClass* (online application)
- Conclusions and Future Work

# Data

- Source:
  - Ryan Woods (M.D.)
  - Elizabeth Burnside (M.D.)

- 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

| |
|---|
| 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 |

# Masses classification

## Prospective

- **Classification** of feature **mass density** for **180** cases **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 (348)**;
- **Review of mass density classification** made by radiologist (prospective study);
- **Classification without stress**;
- **Reference standard** for **mass density**.

↓ mass density

***retro_density***

# Masses classification

348 cases (**retrospectively** classified)

180

(≈ 52%)

(**prospectively** classified)

168

(≈ 48%)

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- *MammoClass* (online application)
- Conclusions and Future Work

# Methodology

- WEKA


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

348

180 | 168

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

# 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**

# 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*

# Results

348

180 168

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

CRACS INESC PORTO LA

# Results – Experiments

**10 x stratified. c. v.**

180

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

# Results - Experiments

Predicting density

180

# Results - Experiments

**10 x stratified. c. v.**

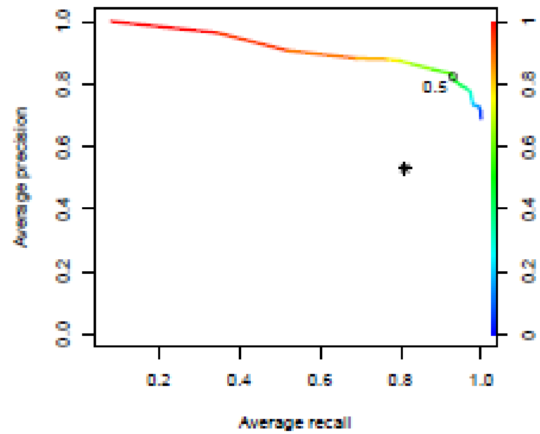
- **$E_4$** – Predicting *retro_density*

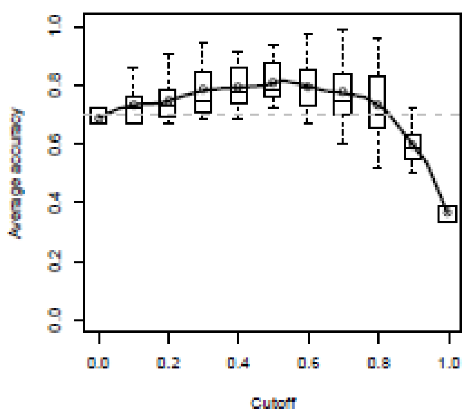
SVM's

CCI: 81.3% (+/-8.2)

K: 0.52 (+/- 0.21)

F: 0.64 (+/- 0.17)




**Radiologist's accuracy = 70 %**

**Our classifier ≈ 81 %**

168

# Results - Experiments

**TEST**

- **$E_6$** – Predicting *retro_density*

(model $E_4$ applied)

SVM's

CCI: 84.5%

K: 0.46

F: 0.91

180

SVM's

CCI: 81.3% (+/-8.2)

K: 0.52 (+/- 0.21)

F: 0.64 (+/- 0.17)

# Results – Experiments

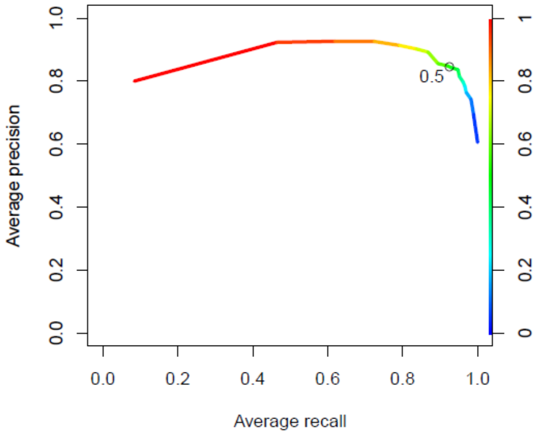
## Predicting malignancy

180

# Results - Experiments

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

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

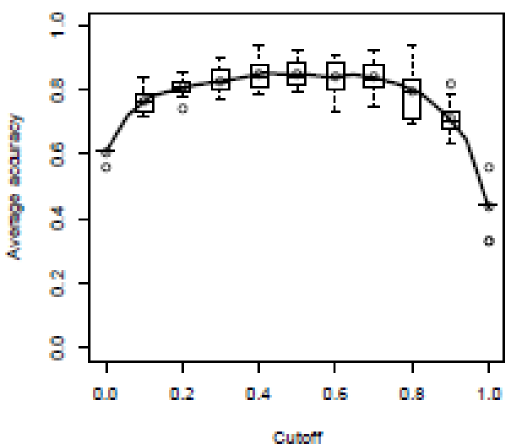



SVM's

CCI: 85.6% (+/-7.3)

K: 0.69 (+/- 0.16)

F: 0.80 (+/- 0.11)



168

# Results - Experiments

180 SVM's

CCI: 85.6% (+/-7.3)

K: 0.69 (+/- 0.16)

F: 0.80 (+/- 0.11)

**TEST**

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

(model $E_1$ applied)

SVM's

CCI: 81.0%

K: 0.50

F: 0.87

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

SVM's

CCI: 78.0%

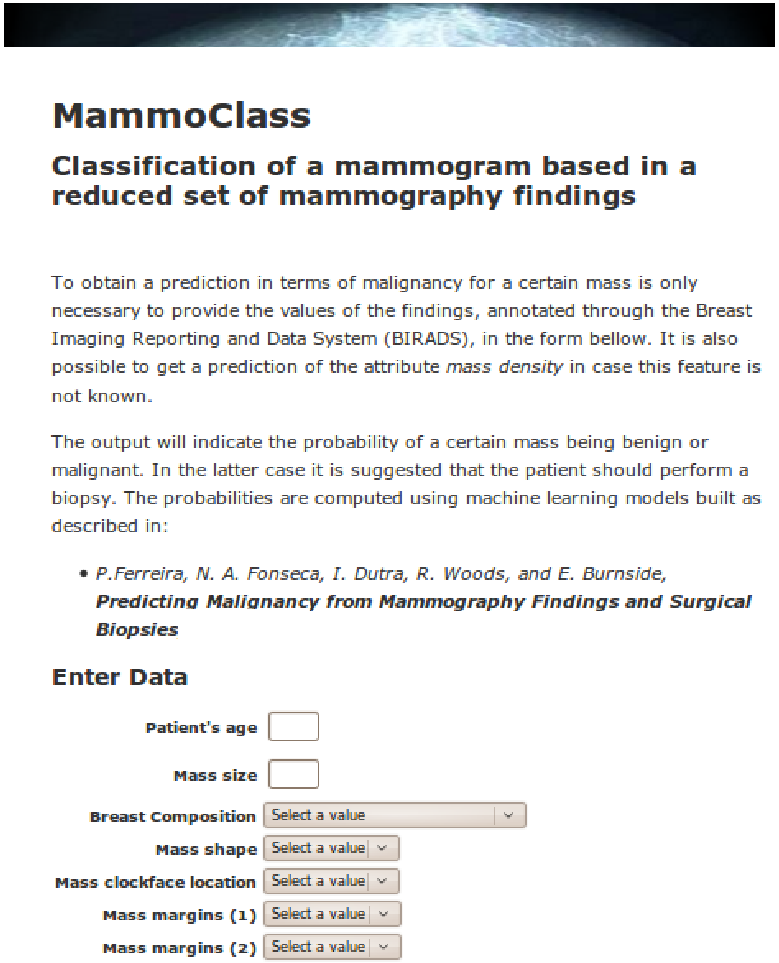
K: 0.45

F: 0.85

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

# *MammoClass*

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



## MammoClass

### Classification of a mammogram based in a reduced set of mammography findings

To obtain a prediction in terms of malignancy for a certain mass is only necessary to provide the values of the findings, annotated through the Breast Imaging Reporting and Data System (BIRADS), in the form bellow. It is also possible to get a prediction of the attribute *mass density* in case this feature is not known.

The output will indicate the probability of a certain mass being benign or malignant. In the latter case it is suggested that the patient should perform a biopsy. The probabilities are computed using machine learning models built as described in:

- *P.Ferreira, N. A. Fonseca, I. Dutra, R. Woods, and E. Burnside,* ***Predicting Malignancy from Mammography Findings and Surgical Biopsies***

### Enter Data

Patient's age

Mass size

Breast Composition Select a value

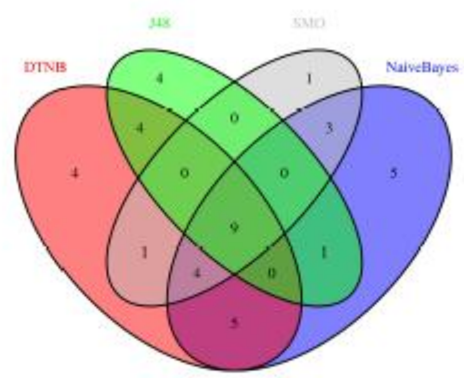
Mass shape Select a value

Mass clockface location Select a value

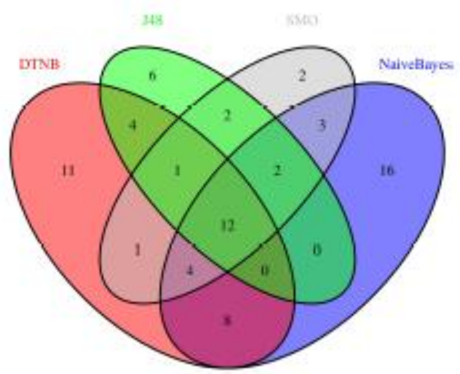
Mass margins (1) Select a value

Mass margins (2) Select a value

# Misclassified Instances




$E_1$




$E_4$

CRACS INESCPORTO LA

# Conclusions and Future Work

a) **Automatic classification** of a mammography can reach **equal or better results** than the classification performed by a **radiologist**;

b) Machine learning **classifiers** can **predict mass density** with **higher quality** than the one obtained by radiologists

   a) our classifier can **predict malignancy** in the absence of mass density, since we can **fill up** this **attribute** using our **mass density predictor.**

CRACS INESCPORTO LA

# Conclusions and Future Work

a) Apply other machine learning techniques based on statistical relational learning;

b) Investigate how other features can affect malignancy or are related to the other attributes;

c) Study why the **parameter variation** on **WEKA algorithms** has a strong **impact** on the **performance** of **classifiers**;

d) Investigate with the radiologist why some **instances** are **consistently misclassified** by all algorithms.

# Thank you!

Pedro Miguel Ferreira

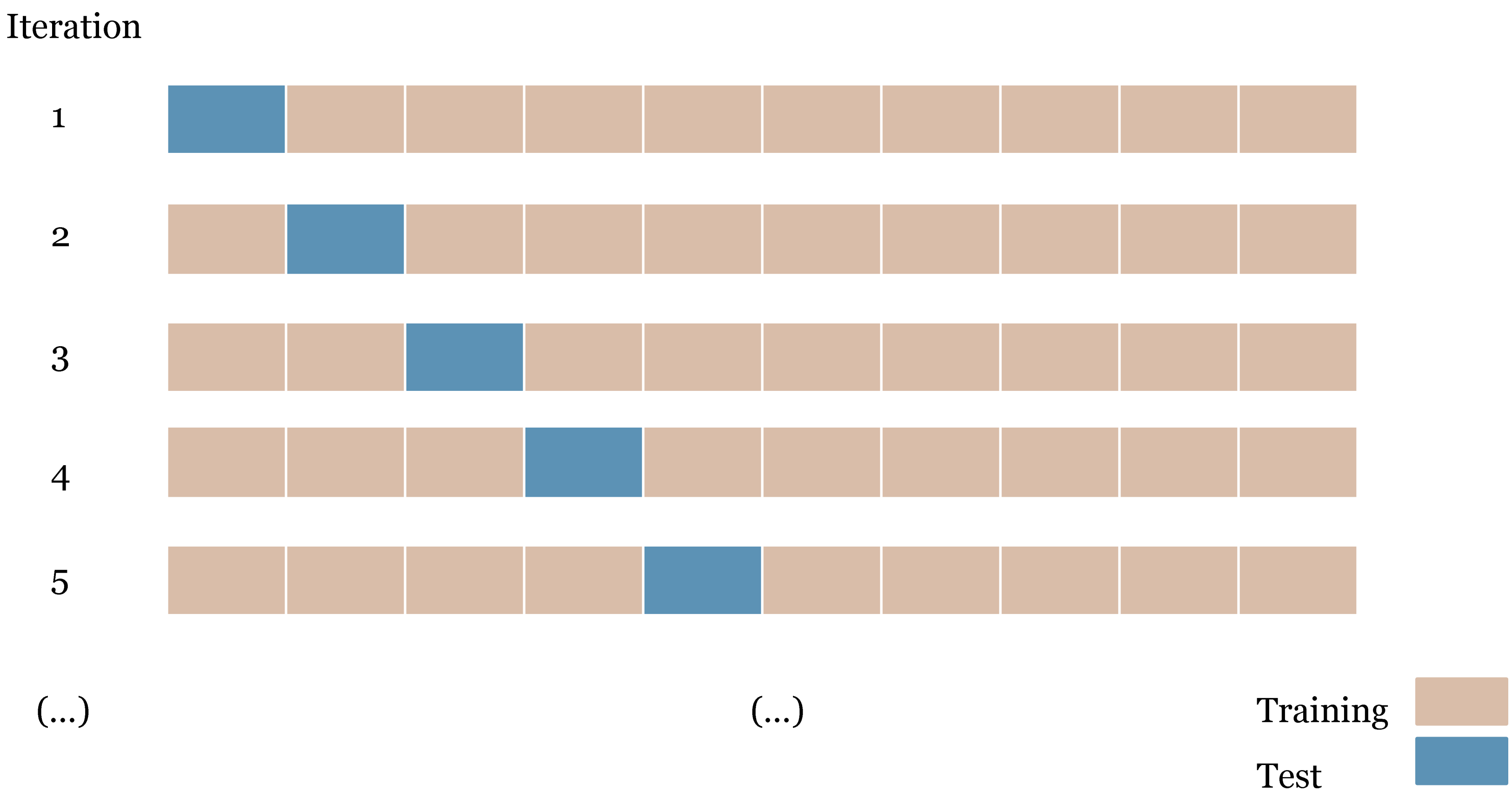
pedroferreira@dcc.fc.up.pt
nunofonseca@acm.org
ines@dcc.fc.up.pt

http://cracs.fc.up.pt

# Appendices

# Methodology

## 10-fold stratified cross-validation



# 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%)** | |

CRACS INESCPORTO LA

# 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%)** | |

# WEKA algorithms used

CLASSIFIERS' PERFORMANCE FOR EACH TASK. VALUES NOT IN BOLD ARE STATISTICALLY SIGNIFICANTLY WORSE THAN THE CLASSIFIER WITH HIGHEST ACCURACY (USING PAIRED T-TEST WITH $\alpha = 0.05$).

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

# Parameter variation in WEKA algorithms

Best parameter selection

($E_1$) Predicting *outcome_num* with *retro_density*; → SMO '-C 0.05 –N1 PK –E 1.0'

($E_2$) Predicting *outcome_num* with *density_num*;

($E_3$) Predicting *outcome_num* without mass density;

($E_4$) Predicting *retro_density*;

($E_2$)–($E_4$) → SMO '-C 0.05 –N2 PK –E 1.0'

($E_5$) Predicting *density_num*; → *naïve Bayes default*

**Parameters Selection:**

**SMO:**
*-C (complexity parameter)*
*-N (filterType)*
- *0 - Normalize training data*
- *1 - Standardize training data*
- *2 - No normalization/standardization*

*PK (PolyKernel)*
*-E (exponent value)*