# Studying the relevance of Breast Imaging Features

HEALTHINF 2011 – January 28th – Rome, Italy




Pedro Miguel Ferreira

Inês Dutra
Nuno A. Fonseca
Ryan Woods
Elizabeth Burnside

CRACS
INESCPORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- Conclusions and Future Work

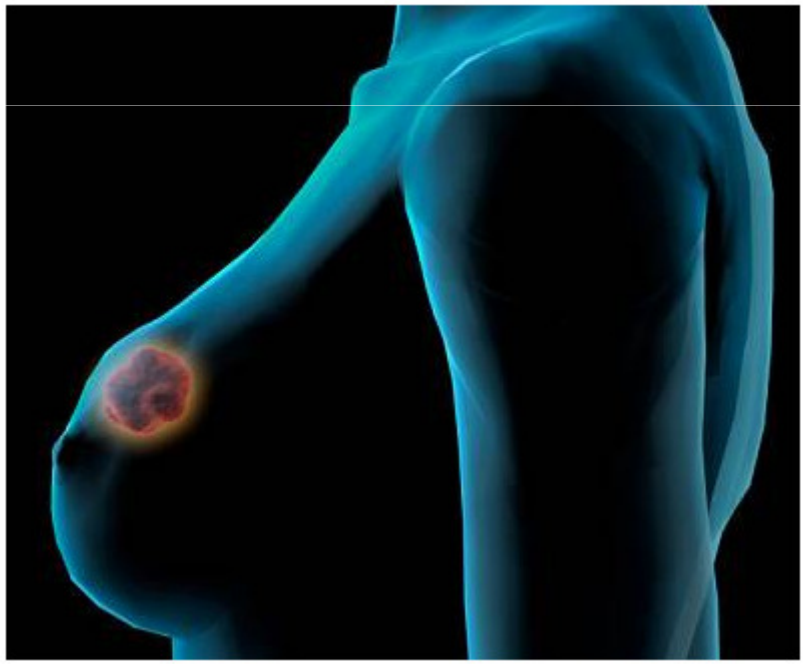
CRACS
INESCPORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- Conclusions and Future Work

CRACS
INESCPORTO LA

# Breast Cancer



- USA:
  - 1/8 women develops breast cancer
  - In 2006:
    - 191.410 with cancer
    - 40.820 (≈ 21%) died

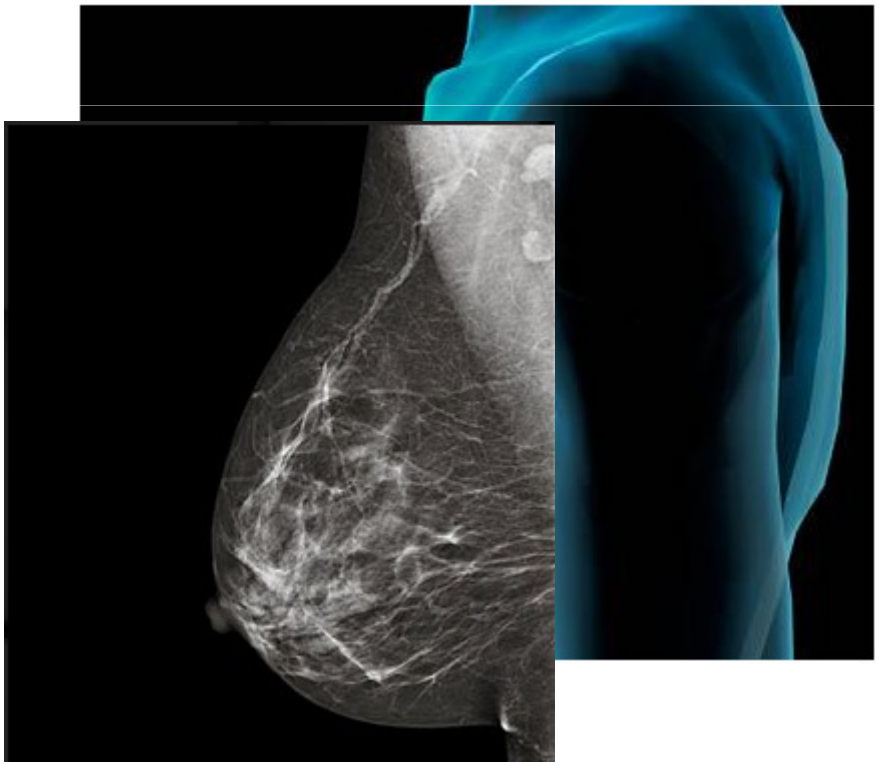
Source: *U. S. Cancer Statistics Working Group* - October 2010

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

Source: *Liga Portuguesa Contra o Cancro* - October 2010

CRACS
INESCPORTO LA

# Breast Screening Programs



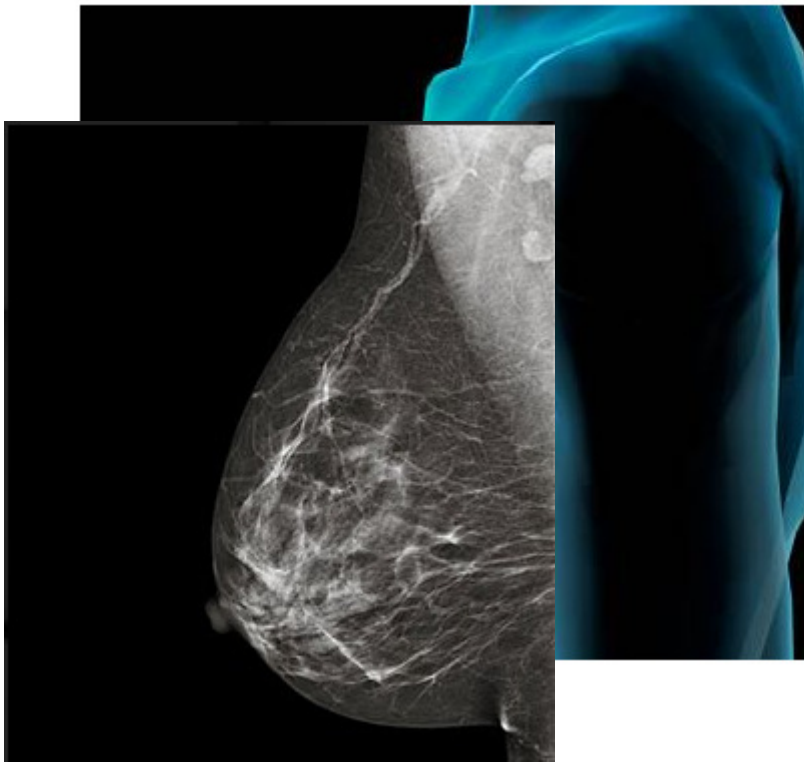
- Reduction of death rate in 30%

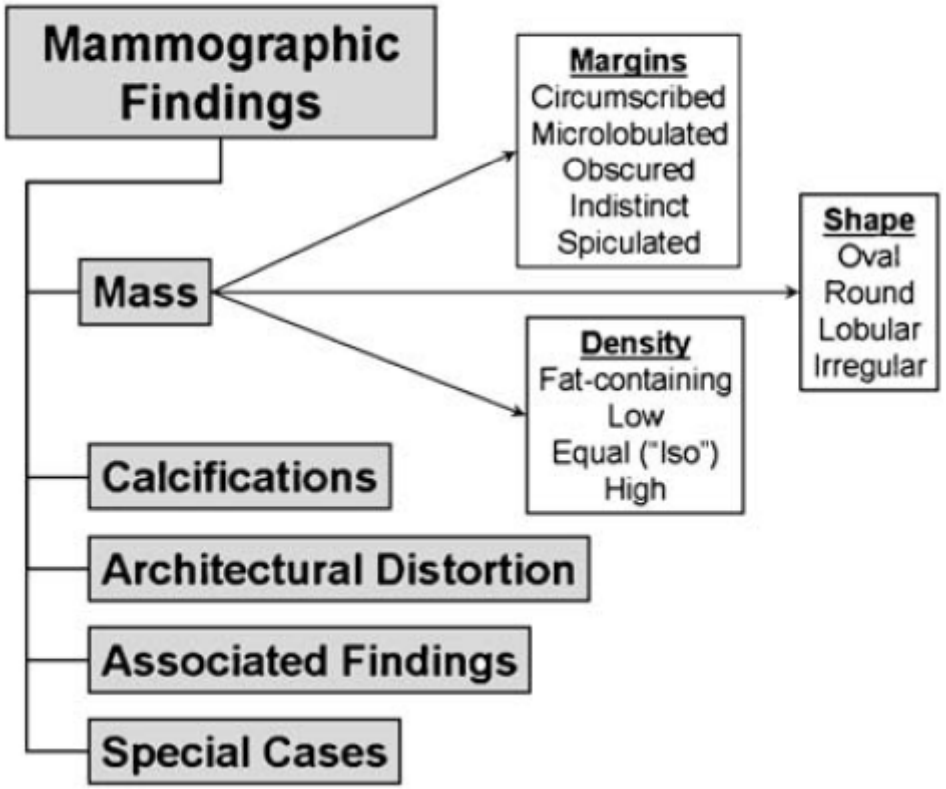
- **Mammography**:

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

CRACS
INESCPORTO LA

# Mammography





CRACS
INESCPORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- Conclusions and Future Work

CRACS
INESCPORTO LA

# Objectives

- Study the influence of **mass density** when predicting malignancy
- Find relations among features
  - Machine Learning techniques
- "Learn" models capable of helping physicians in the analysis of thousands of mammographies

CRACS
INESCPORTO LA

# Objectives

- Try to build classifiers capable of predicting **mass density** and **malignancy**

- Reduce the number of unnecessary biopsies

CRACS
INESCPORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- Conclusions and Future Work

CRACS
INESCPORTO LA

# Data

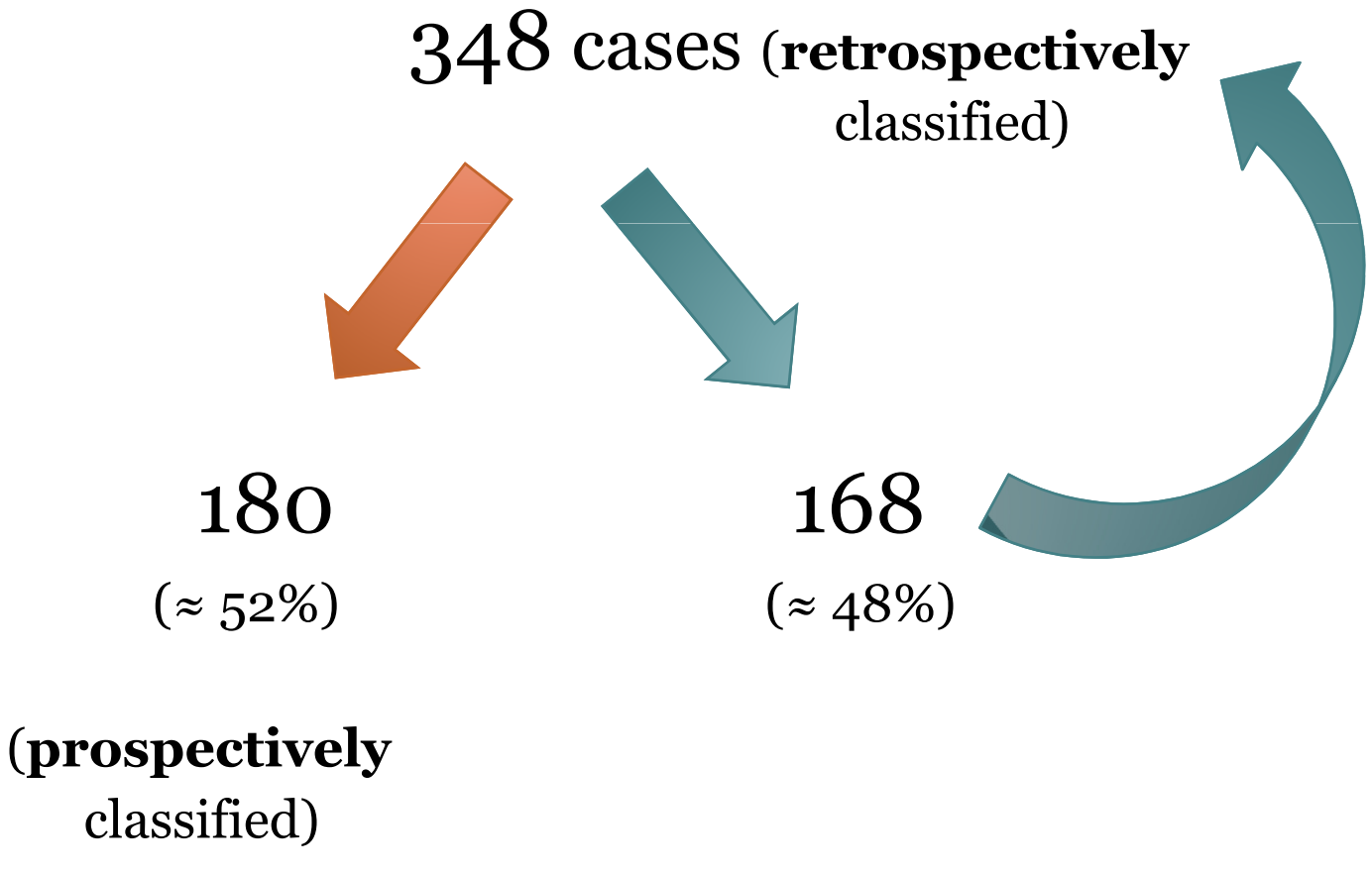
- Provided by:
  - Dr. Ryan Woods
  - Dra. Elizabeth Burnside

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

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

CRACS
INESCPORTO LA

# Masses classification



CRACS INESCPORTO LA

# 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** in a periodic meeting in which they **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
INESCPORTO LA

# Outline

- Breast Cancer
- Objectives
- Data
- Methodology
- Results and Analysis
- Conclusions and Future Work

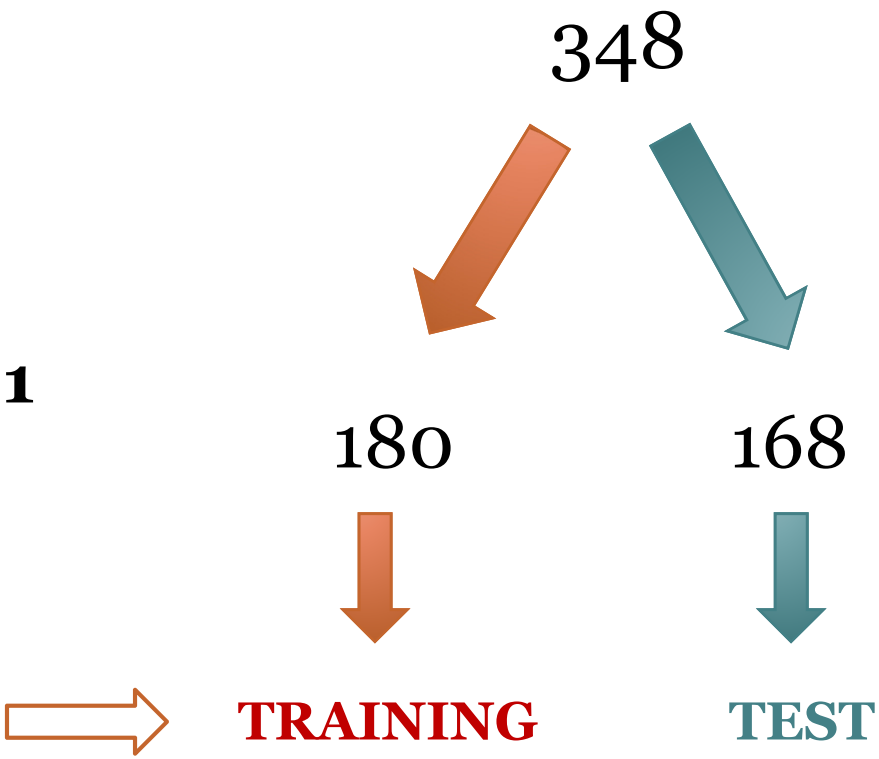
CRACS INESCPORTO LA

# Methodology

- WEKA


- Paired Corrected T-Tester
  - **Significance level: 0.01**

- **10-fold stratified cross-validation**



CRACS
INESCPORTO LA

# Results and Analysis

- **$Q_1$** – Is mass density predictive of malignancy?

  Prediction of malignancy with and without mass density — 180

- **$Q_2$** – Can we obtain classifiers that predict mass density as well as a radiologist?

  Prediction of mass density — 180

- **$Q_3$** – Can the generated classifiers behave well on unseen data?

  Prediction of mass density
  Prediction of malignancy with and without mass density — 168

CRACS
INESCPORTO LA

# Results and Analysis

- **$Q_1$** – Is mass density predictive of malignancy?

  Prediction of malignancy with and without mass density 180

- $Q_2$ – Can we obtain classifiers that predict mass density as well as a radiologist?

  Prediction of mass density 180

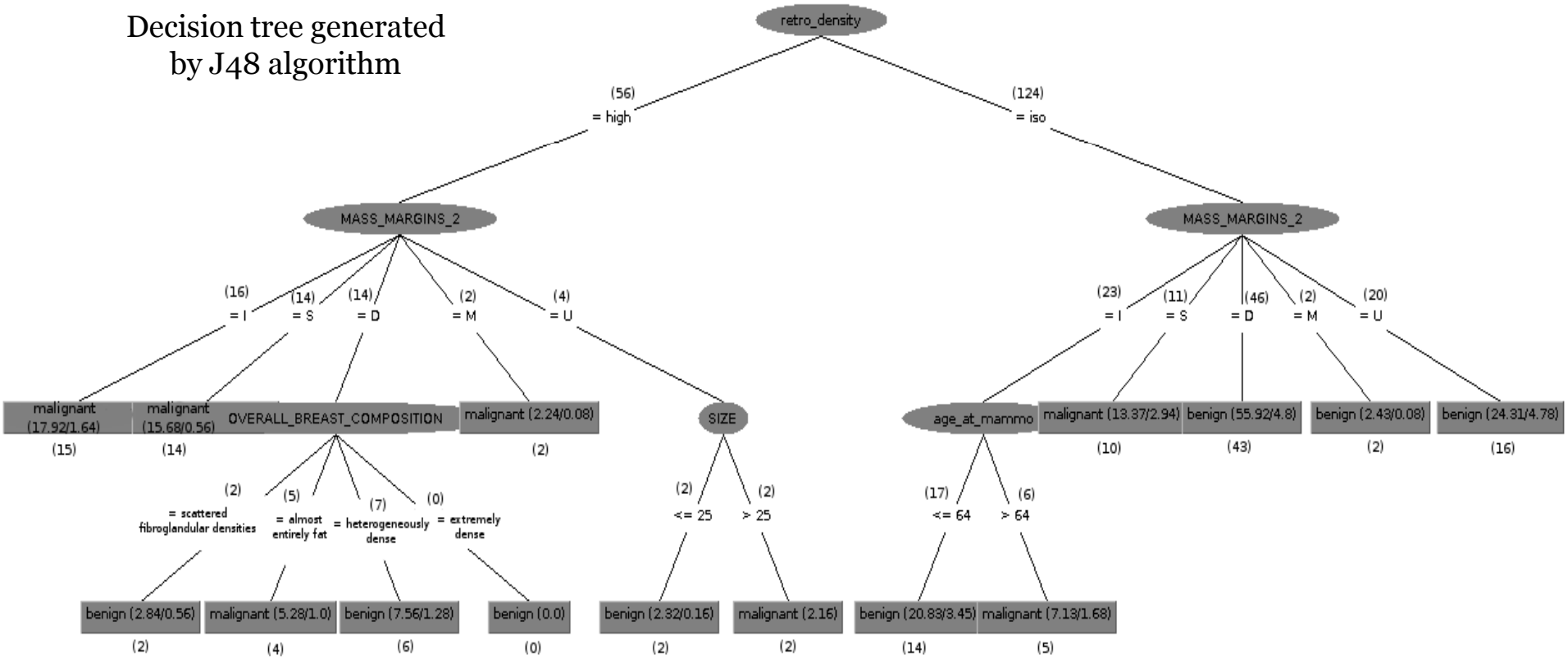
- $Q_3$ – Can the generated classifiers behave well on unseen data?

  Prediction of mass density 168
  Prediction of malignancy with and without mass density

CRACS INESCPORTO LA

# $Q_1$ - Predicting malignancy with *retro_density* ($E_1$)

Decision tree generated by J48 algorithm



SVM's

| 180 | Prediction of malignancy | | |
|---|---|---|---|
| | with mass density | | $E_3$ |
| Metric | $E_1$ Retrospective (*retro_density*) | $E_2$ Prospective (*density_num*) | without mass density |
| Corrected Classified Instances | 84.78% (+/- 7.96) | 82.72% (+/- 8.32) | 81.39% (+/- 8.81) |
| Kappa Statistic | 0.68 (+/- 0.17) | 0.63 (+/- 0.17) | 0.60 (+/- 0.18) |
| F-Measure | 0.80 (+/- 0.11) | 0.77 (+/- 0.11) | 0.75 (+/- 0.12) |

- **Mass density** has some **influence** when predicting **malignancy**, especially if we use density from the retrospective study ($E_1$)

- ($E_1$) -> *retro_density* -> CCI = 84.78% (+/- 7.96)
  K = 0.68 (+/- 0.17)

- ($E_3$) -> without density -> CCI = 81.39% (+/- 8.81)
  K = 0.60 (+/- 0.18)

CRACS INESCPORTO LA

# Results and Analysis

- $Q_1$ – Is mass density predictive of malignancy?

  Prediction of malignancy with and without mass density 180

- **$Q_2$** – Can we obtain classifiers that predict mass density as well as a radiologist?

  Prediction of mass density **180**

- $Q_3$ – Can the generated classifiers behave well on unseen data?

  Prediction of mass density 168
  Prediction of malignancy with and without mass density

CRACS INESCPORTO LA

# $Q_2$ - Can we obtain classifiers that predict mass density as well as a radiologist?

- 70% of masses classified by the radiologist in the prospective study (180 findings) agreed to the classified masses in the retrospective study
- **Radiologist's accuracy: 70%**

*naïve Bayes*

| 180 | Prediction of mass density |
|---|---|
| Metric | $E_4$ *retro_density* |
| Corrected Classified Instances | 72.83% (+/- 9.89) |
| Kappa Statistic | 0.37 (+/- 0.23) |
| F-Measure | 0.56 (+/- 0.18) |

CRACS
INESCPORTO LA

# Results and Analysis

- $Q_1$ – Is mass density predictive of malignancy?

  Prediction of malignancy with and without mass density 180

- $Q_2$ – Can we obtain classifiers that predict mass density as well as a radiologist?

  Prediction of mass density 180

- **$Q_3$** – Can the generated classifiers behave well on unseen data?

  Prediction of mass density
  Prediction of malignancy with and without mass density 168

CRACS INESCPORTO LA

# $Q_3$ - Can the generated classifiers n data?

*naïve Bayes*

| 180 | Prediction of mass density |
|---|---|
| Metric | $E_4$ *retro_density* |
| Corrected Classified Instances | 72.83% (+/- 9.89) |
| Kappa Statistic | 0.37 (+/- 0.23) |
| F-Measure | 0.56 (+/- 0.18) |

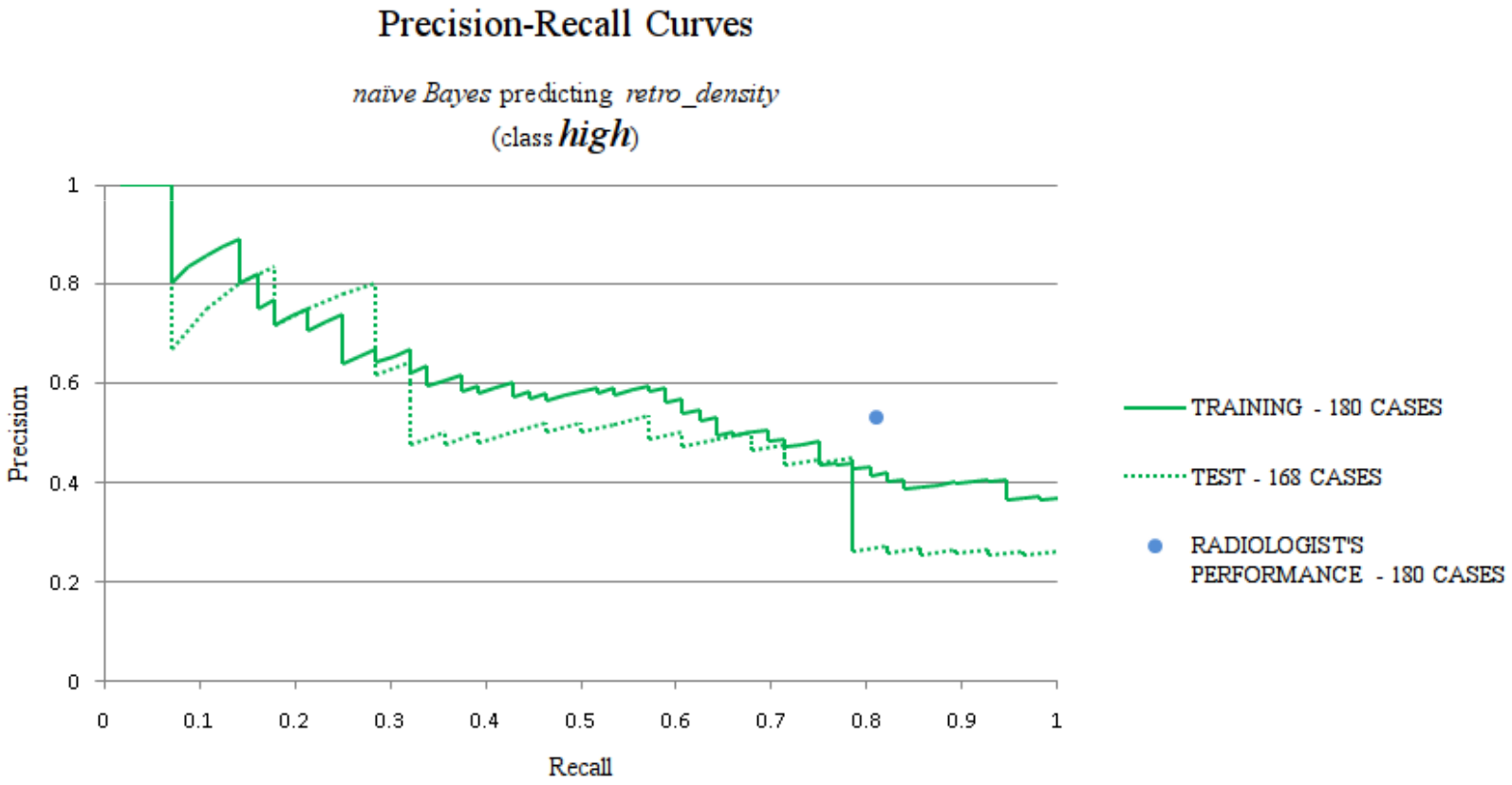
*naïve Bayes*

| 168 | Prediction of mass density |
|---|---|
| Metric | $E_6$ *retro_density* |
| Corrected Classified Instances | 82.14% |
| Kappa Statistic | 0.45 |
| F-Measure | 0.56 |

- Classifier based on ***naïve Bayes*** algorithm

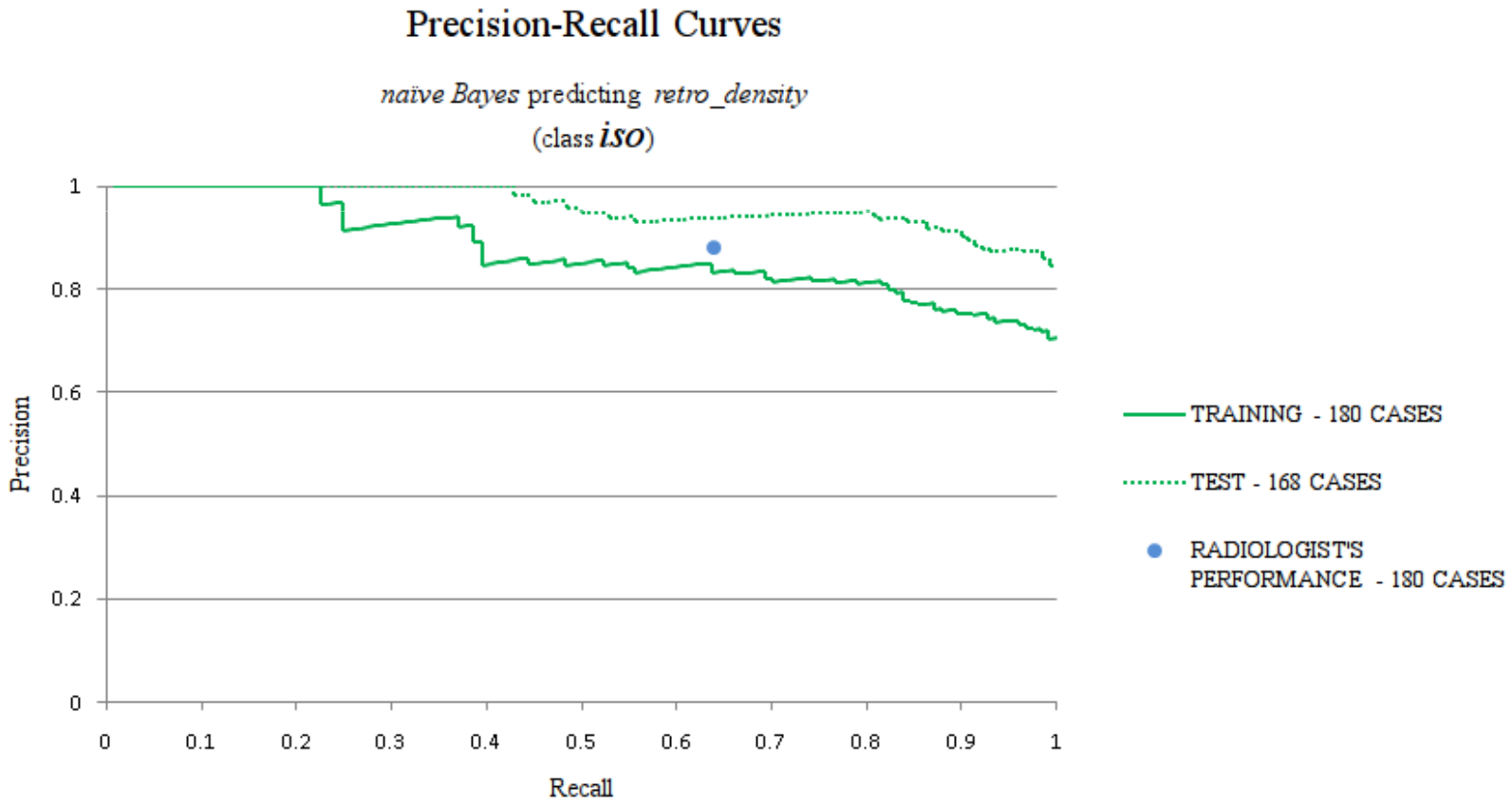
CRACS INESCPORTO LA

# Summary (Predicting density – class *high*)



CRACS
INESCPORTO LA

# Summary (Predicting density – class *iso*)



CRACS INESCPORTO LA

# $Q_3$ – Can the generated classifiers behave well on unseen data?

| 168 | Prediction of malignancy | | |
|---|---|---|---|
| | with mass density | | $E_{11}$ |
| Metric | $E_8$ Retrospective (real values) | $E_9$ Retrospective (predicted values by classifier *naive Bayes*) | without mass density |
| Corrected Classified Instances | 81.55% | 79.76% | 77.38% |
| Kappa Statistic | 0.52 | 0.48 | 0.42 |
| F-Measure | 0.64 | 0.62 | 0.57 |

- Classifiers based on **SVM's**

CRACS
INESCPORTO LA

# Conclusions and Future Work

a) Automatic classification of a mammography can achieve results as good as specialists;

b) Mass density seems to be a good evidence of malignancy;

c) Using machine learning classifiers to predict mass density can reach equal or better results than the ones obtained by radiologists.

CRACS
INESCPORTO LA

# Conclusions and Future Work

a) Extend this work to larger and geographically distinct datasets ;

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

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

d) Incorporate the generated models into a mammography classification system.

# Thank you!




Pedro Miguel Ferreira

pedroferreira@dcc.fc.up.pt

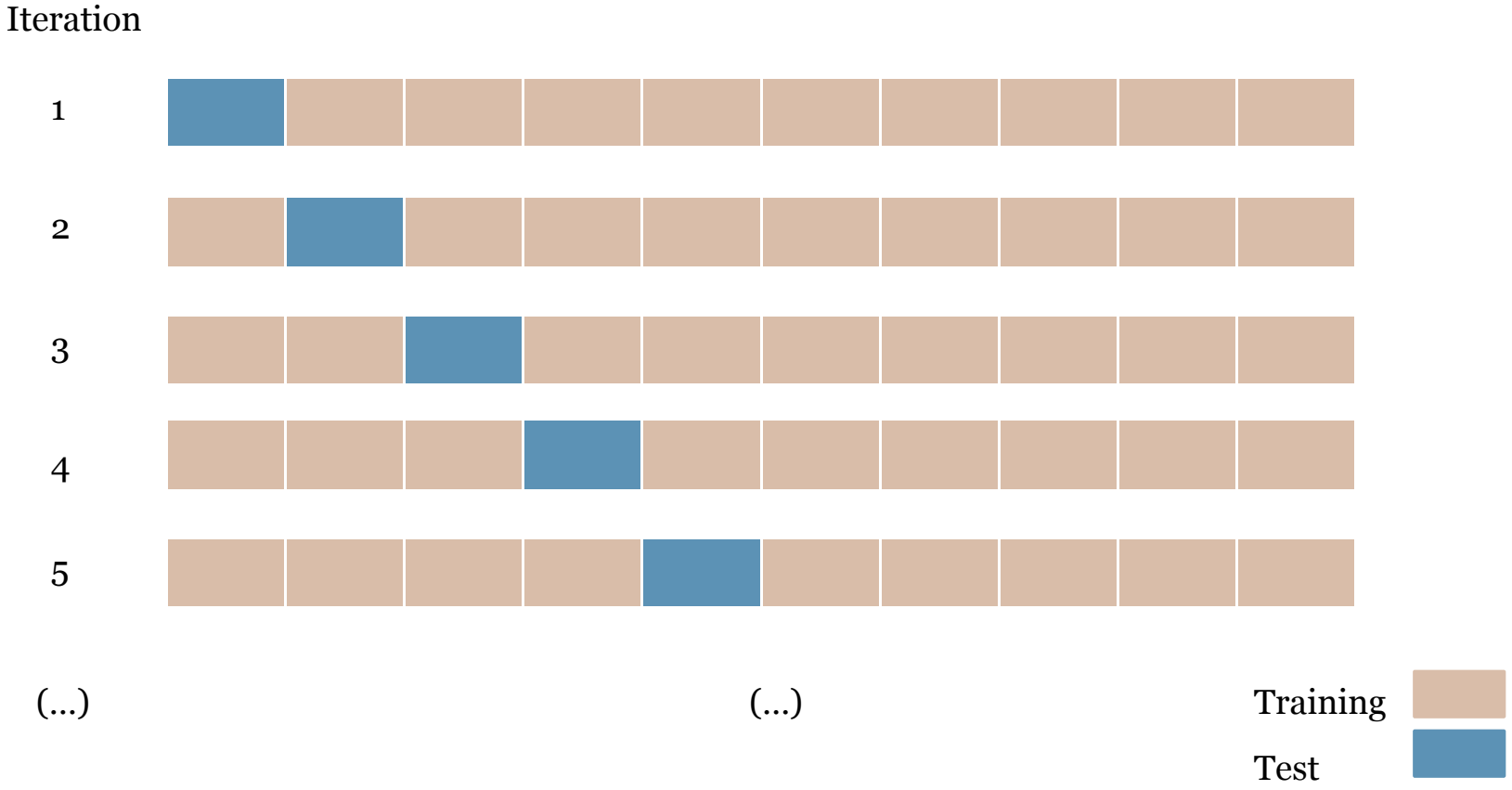
http://cracs.fc.up.pt

# Appendices

CRACS INESCPORTO LA

# Methodology

## 10-fold stratified cross-validation



CRACS INESCPORTO LA

# 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%)** | |

CRACS
INESCPORTO LA

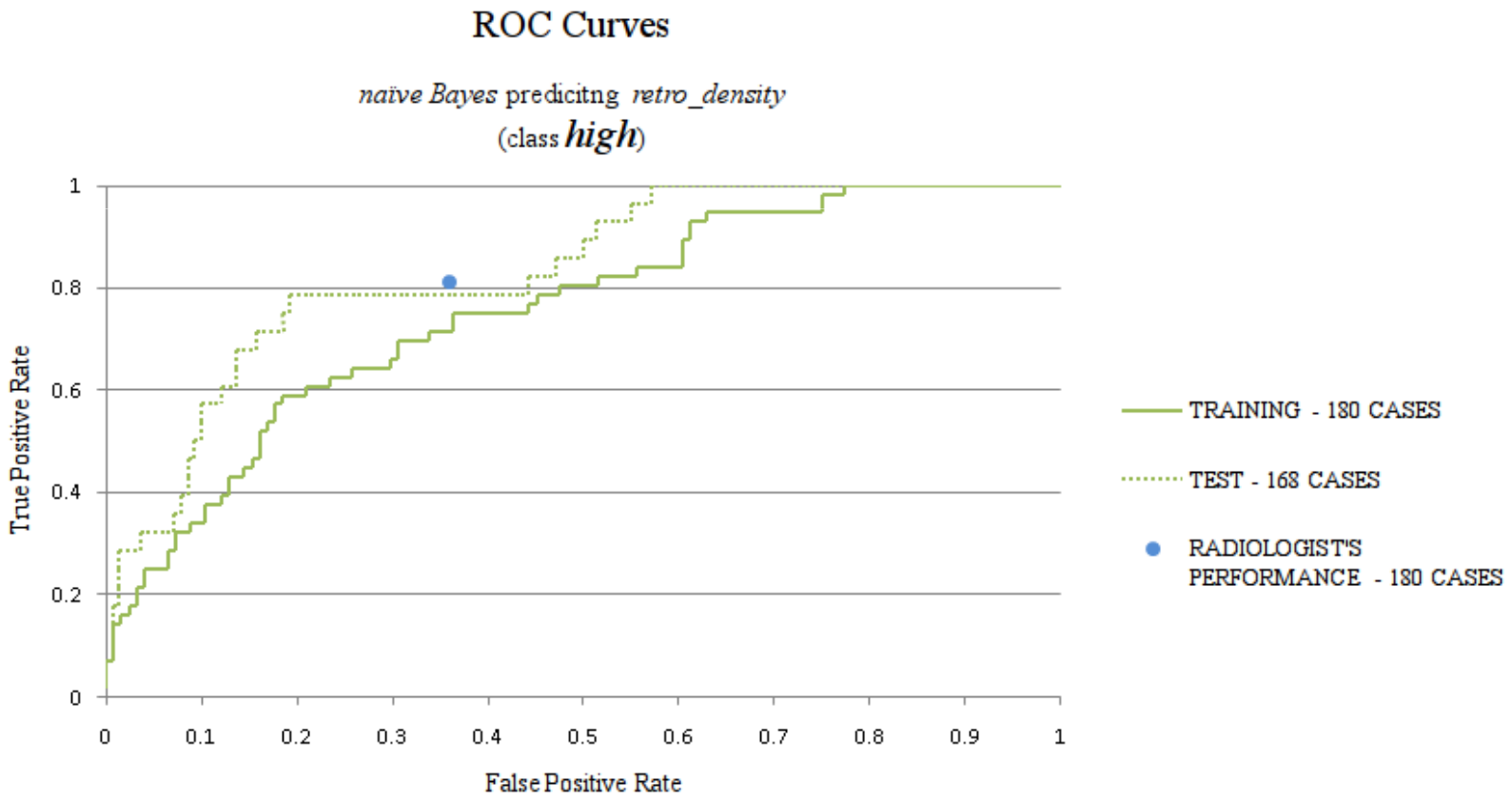
# Data distribution

- 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%)** | |

CRACS
INESCPORTO LA

# ROC Curves (Predicting density – class *high*)



CRACS
INESCPORTO LA

# ROC Curves (Predicting density – class *iso*)