Benchmarking Datasets for Breast Cancer CAD

Miguel Angel Guevara López

BioIDA: Biomedical Image and Data Analysis Group

INEGI - Institute of Mechanical Engineering and Industrial Management, Faculty of Engineering, University of Porto

Team Members working at INEGI:
Naimy Gonzalez de Posada
Pedro Cunha
Daniel C. Moura
Miguel A. Guevara López

Research Lines/Applications

Main Research Lines

- Pattern Recognition – Machine Learning
- Data Mining
- Digital Image Processing
- Artificial Intelligence

Application Area

- Biomedical Sciences

- BCDR – Breast Cancer Digital Repository
- MIWAD – Mammography Image Workstation for Analysis and Diagnosis.
- A methodology for exploring massively machine learning classifiers.

- 3D Reconstruction System for Spine
- Proper tool for evaluating the success of dental implant placement.
- Bone age determination (on development)

Results/Products

Benchmarking Datasets for Breast Cancer CAD

GOALS

- Building an Iberian reference (golden standard) breast cancer digital repository
- Designing reliable systems that provide a 2nd opinion to radiologists, based on data extracted from:
  - + *Clinical records*
  - + *Medical images*
- Training medical students, formed physicians and other medical-related professionals
- Exploring and extending achieved findings to others medical imaging areas (lung, liver, etc.)

IMED Project

(for Development of Algorithms for Medical Image Analysis)

Hospital S. João

Faculty of Medicine,
University of Porto

CETA – CIEMAT, Spain

Partners

BREAST CANCER DIGITAL REPOSITORY

Homepage | More About | Publications | Contacts | Logout

Select the characteristics desired for the search. Blank form show all the patients in the repository.

Gender
☑ Female ☐ Male

Segmentations
With Segmentations? ☐

Age(Value)
◉ Value

Age(Interval) ○

Breast Density
☐ n/a ☐ < 25%
☐ 25% to 50% ☐ 51% to 75%
☐ > 75%

Breast Location
☐ Left ☐ Right
☐ QSI ☐ QSE
☐ QII ☐ QIE
☐ Axillary ☐ Central
☐ Retroareolar

Mammography
☐ Normal ☑ Anomaly
☑ Nodule
☐ Microcalcification
☐ Calcification
☐ Axillary Adenopathy
☐ Architectural Distortion
☐ Stroma Distortion

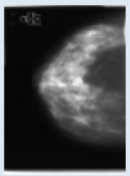
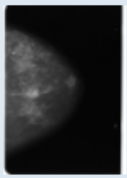
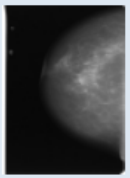
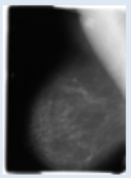
Biopsy Result
☐ No
☑ Benign
☐ Suspect
☐ Insufficient / Unrepresentative
☐ Malignant

Definitive Diagnosis
☐ No ☐ Benign
☐ C.I.S. ☐ Invasive C.
☐ Micro C. ☐ Others
☐ Undetermined

Classification
☐ Unnasigned ☐ Birads 0
☐ Birads 1 ☐ Birads 2
☐ Birads 3 ☐ Birads 4A
☐ Birads 4B ☐ Birads 4C
☐ Birads 5 ☐ Birads 6

Search | Reset search

Patient 2	Patient 4	Patient 36	Patient 52	Patient 59
Age 53	Age 62	Age 63	Age 35	Age 56
Female	Female	Female	Female	Female
1 studie(s)	1 studie(s)	1 studie(s)	1 studie(s)	1 studie(s)


1 2 3 4 5
119 patient(s) found - page **1/24**

BCDR

http://bcdr.inegi.up.pt

BREAST CANCER DIGITAL REPOSITORY

Homepage | More About | Publications | Contacts | Logout

Patient: 4
Age: 62
Gender: Female
Lesion | Images

Study	Date	N. Lesions	Breast Density
4	1994-03-09	1	< 25%

Back to Patient Search

Lesions

Lesion	Breast	Diagnosis
4	Left	No

Breast Location
(•) Left ( ) Right
[x] QSI [x] QSE
[ ] QII [ ] QIE
[ ] Axillary [ ] Central
[ ] Retroareolar

Mammography
( ) Normal (•) Anomaly
[x] Nodule
[ ] Microcalcification
[ ] Calcification
[ ] Axillary Adenopathy
[ ] Architectural Distortion
[ ] Stroma Distortion

Ecography
( ) Normal (•) Anomaly ( ) No
[x] Cystic Nodule [ ] Solid Nodule
[ ] Microcalcification [ ] Ectasia
[ ] Other

Previous Surgery
(•) No ( ) Yes

Biopsy
[x] Aspiration [ ] Vacuum
[ ] Core [x] Echography
[ ] Stereotactic [ ] Palpation
[ ] MRI

Anatomical Pathology
( ) No
(•) Benign
( ) Suspect
( ) Insufficient / Unrepresentative
( ) Malignant

Lesion Characterization
(•) Benign
( ) P-Benign
( ) Malignant
( ) P-Malignant
( ) Indeterminate

Definitive Diagnosis
(•) No
( ) Benign
( ) C.I.S.
( ) Invasive C.
( ) Micro C.
( ) Others
( ) Undetermined

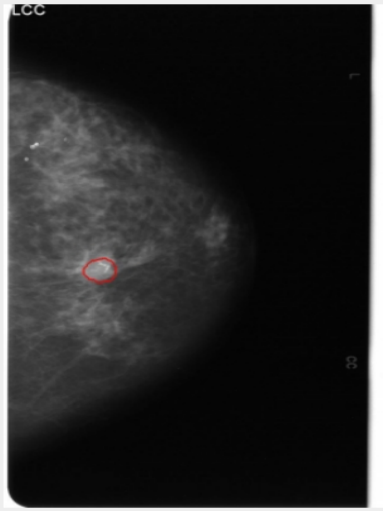
Classifications

Id	Family	Class	Classifier
170	BIRADS	BIRADS 2	Medic 1

Segmentations

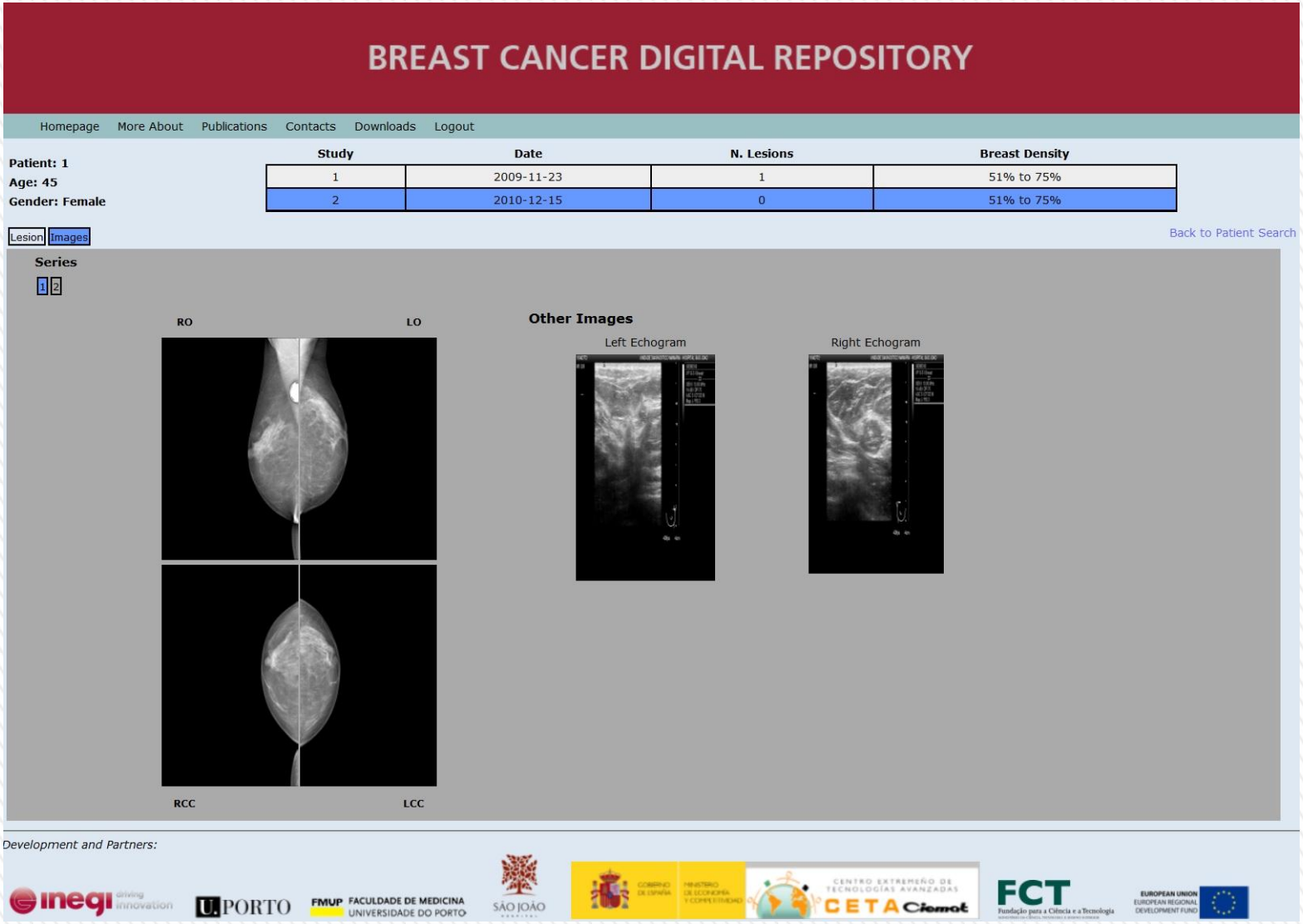
Id	Image	Date
336	Left Cranium Caudal	2011-08-02
337	Left Oblique	2011-08-02

Area:	2451.000	Skewness:	-0.558
Perimeter:	190.734	Kurtosis:	0.888
Correlation:	0.001	Elongation:	0.927
Minimum:	59.000	Roughness:	1.181
Maximum:	211.000	Shape:	1.088
Median:	139.000	Circularity:	0.847
Mean:	134.807	Contrast:	28.586
Standard deviation:	25.412	Entropy:	7.178
Statistical mode:	143.000		
Inverse difference moment:	0.265	Angular second moment:	0.001
X center mass:	181.564	X centroid:	182.402
Y center mass:	616.123	Y centroid:	617.451



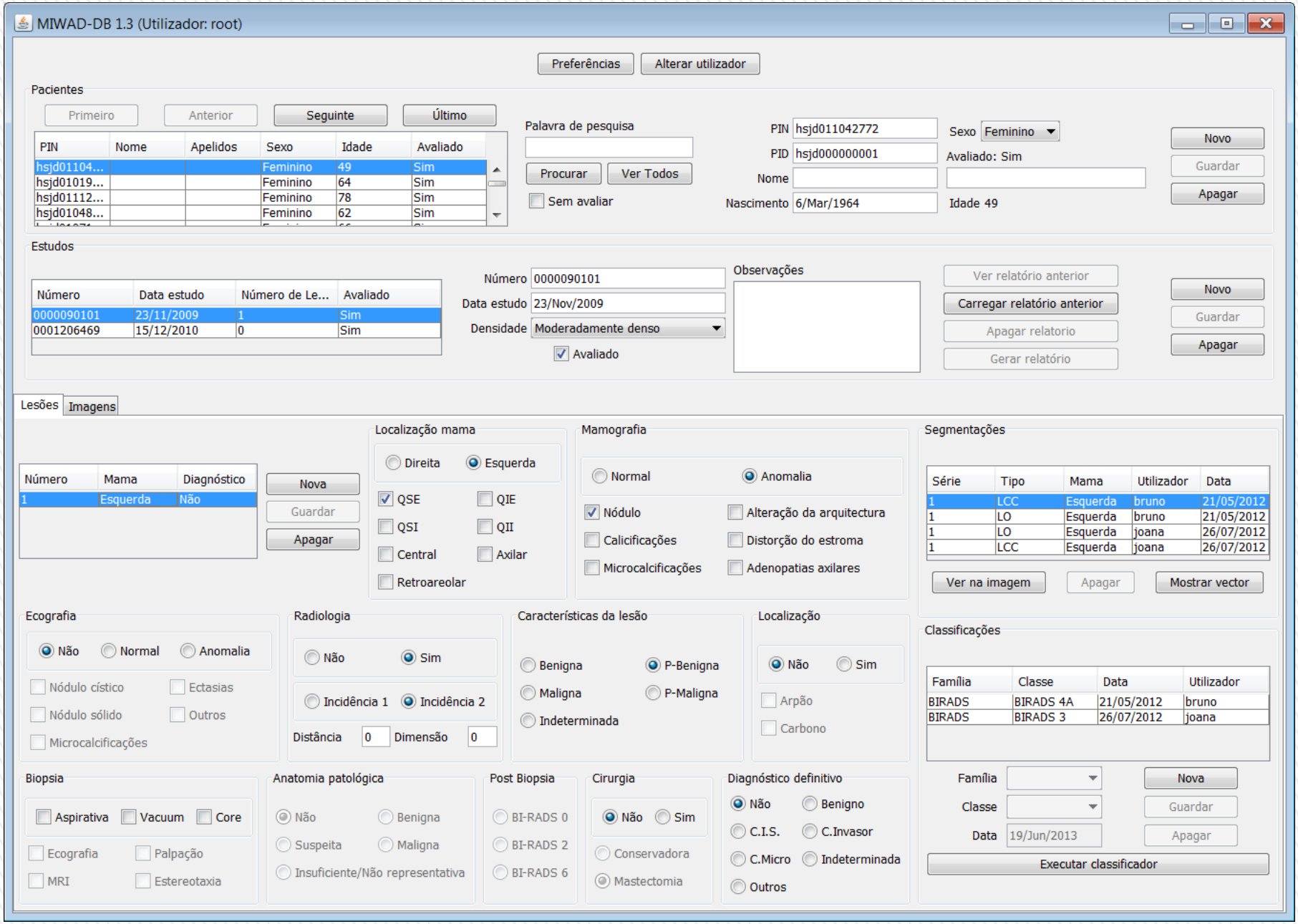
BCDR

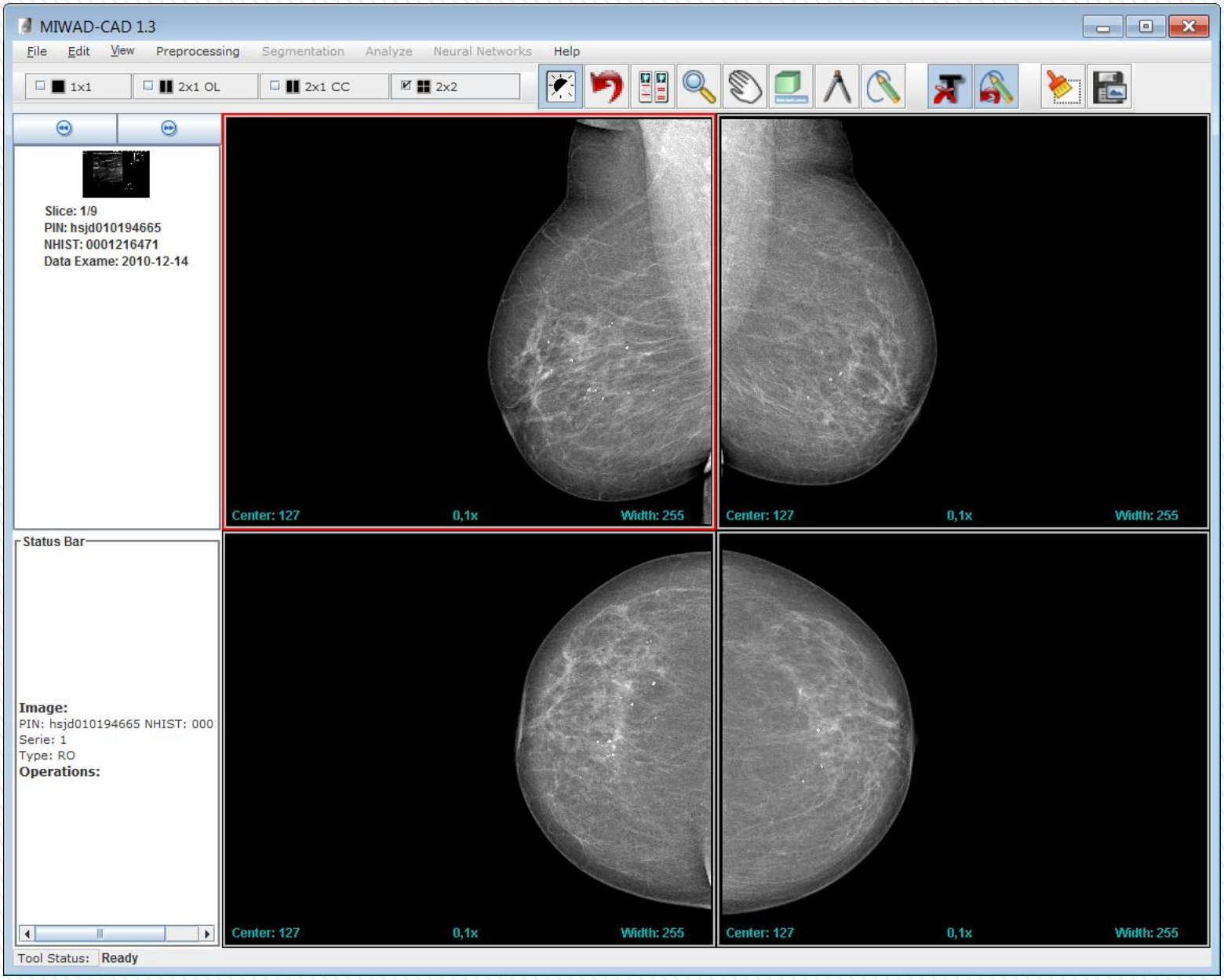
http://bcdr.inegi.up.pt



BCDR

http://bcdr.inegi.up.pt


MIWAD-DB 1.3 (Utilizador: root)
Preferências
Alterar utilizador
Pacientes
Primeiro
Anterior
Seguinte
Último
PIN Nome Apelidos Sexo Idade Avaliado
hsjd01104... Feminino 49 Sim
hsjd01019... Feminino 64 Sim
hsjd01112... Feminino 78 Sim
hsjd01048... Feminino 62 Sim
Palavra de pesquisa
Procurar
Ver Todos
Sem avaliar
PIN hsjd011042772
PID hsjd000000001
Nome
Nascimento 6/Mar/1964
Sexo Feminino
Avaliado: Sim
Idade 49
Novo
Guardar
Apagar
Estudos
Número Data estudo Número de Le... Avaliado
0000090101 23/11/2009 1 Sim
0001206469 15/12/2010 0 Sim
Número 0000090101
Data estudo 23/Nov/2009
Densidade Moderadamente denso
Avaliado
Observações
Ver relatório anterior
Carregar relatório anterior
Apagar relatorio
Gerar relatório
Novo
Guardar
Apagar
Lesões
Imagens
Número Mama Diagnóstico
1 Esquerda Não
Nova
Guardar
Apagar
Localização mama
Direita
Esquerda
QSE
QIE
QSI
QII
Central
Axilar
Retroareolar
Mamografia
Normal
Anomalia
Nódulo
Alteração da arquitectura
Calicificações
Distorção do estroma
Microcalcificações
Adenopatias axilares
Segmentações
Série Tipo Mama Utilizador Data
1 LCC Esquerda bruno 21/05/2012
1 LO Esquerda bruno 21/05/2012
1 LO Esquerda joana 26/07/2012
1 LCC Esquerda joana 26/07/2012
Ver na imagem
Apagar
Mostrar vector
Ecografia
Não
Normal
Anomalia
Nódulo cístico
Ectasias
Nódulo sólido
Outros
Microcalcificações
Radiologia
Não
Sim
Incidência 1
Incidência 2
Distância 0
Dimensão 0
Características da lesão
Benigna
P-Benigna
Maligna
P-Maligna
Indeterminada
Localização
Não
Sim
Arpão
Carbono
Classificações
Família Classe Data Utilizador
BIRADS BIRADS 4A 21/05/2012 bruno
BIRADS BIRADS 3 26/07/2012 joana
Família
Classe
Data 19/Jun/2013
Nova
Guardar
Apagar
Executar classificador
Biopsia
Aspirativa
Vacuum
Core
Ecografia
Palpação
MRI
Estereotaxia
Anatomia patológica
Não
Benigna
Suspeita
Maligna
Insuficiente/Não representativa
Post Biopsia
BI-RADS 0
BI-RADS 2
BI-RADS 6
Cirurgia
Não
Sim
Conservadora
Mastectomia
Diagnóstico definitivo
Não
Benigno
C.I.S.
C.Invasor
C.Micro
Indeterminada
Outros


MIWAD-CAD 1.3
File
Edit
View
Preprocessing
Segmentation
Analyze
Neural Networks
Help
1x1
2x1 OL
2x1 CC
2x2
Slice: 1/9
PIN: hsjd010194665
NHIST: 0001216471
Data Exame: 2010-12-14
Center: 127
0,1x
Width: 255
Center: 127
0,1x
Width: 255
Status Bar
Image:
PIN: hsjd010194665 NHIST: 000
Serie: 1
Type: RO
Operations:
Center: 127
0,1x
Width: 255
Center: 127
0,1x
Width: 255
Tool Status:
Ready

BCDR-FMR (Film Mammography)

- Anonymous cases from HSJ – FMUP medical historical archives.
- 1010 patients cases (with ages between 20 and 90 years old).
- **1125 studies** (3703 mammography incidences, 1044 lesions clinically described (820 already identified by radiologists in MLO and CC images, of which 276 are biopsy proven).
- 1517 segmentations were made on MLO and CC images and classified (BI-RADS) by specialized radiologists.
- Low resolution images.

BCDR-DMR (Full Field Digital Mammography)

- Now in construction, with anonymous cases from HSJ – FMUP.
- At present is composed by 730 patients cases, including **828 studies**, 2837 mammography incidences (MLO and CC) and 2073 ultrasound pictures.
- High resolution images (3328x4084 pixels and a bit depth of 14 bits per pixel).
- Expected 2000 or 3000 patients' cases

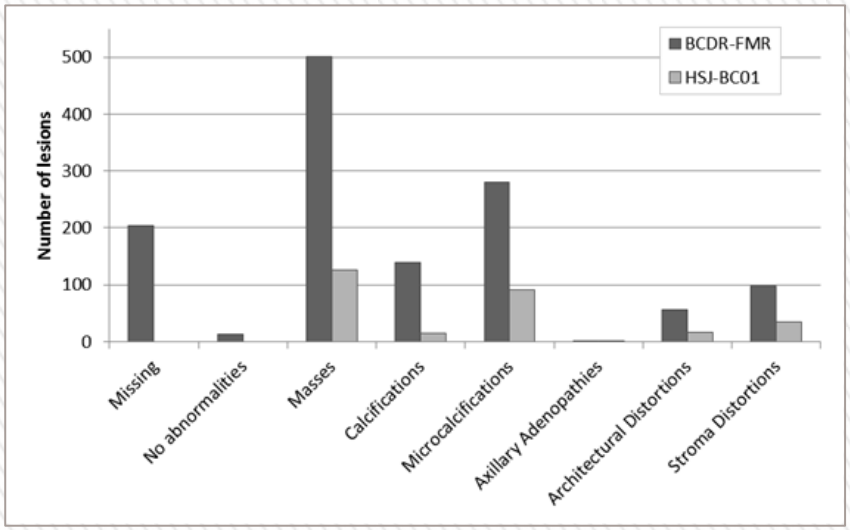
Breast Cancer Digital Repository

Public Datasets

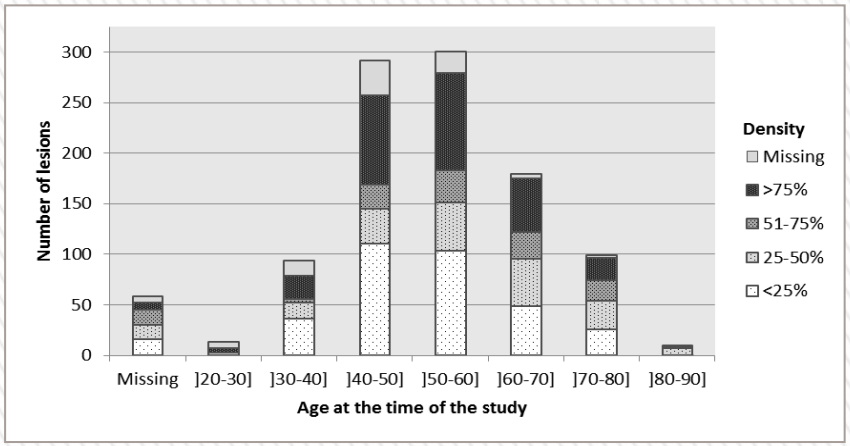
- Currently, 2 public datasets, biopsy-proven
  - BCDR-F01: Film mammography
  - BCDR-D01: Digital mammography (soon)
- Datasets presented in two flavours
  - CSV file with Metadata and Pre-computed features
    - Ready to use with machine learning software (e.g. Weka)
    - For the machine learning researcher that does not want to spend time with the hassle of reading images and computing features
  - CSV file with Metadata + Mammographies
    - For the computer vision and pattern recognition researchers
    - Outlines of the lesions available

›

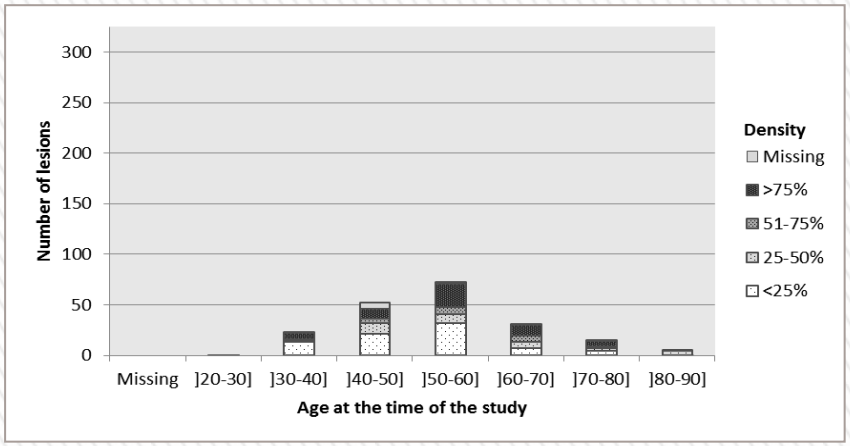
- Formed from **200 lesions**: 100 benign and 100 malignant (biopsy proven)
- Composed by **358 features vectors** (184 instances related to the 100 benign lesions and 174 instances related to the 100 malignant lesions)
- Each vector includes clinical, intensity, texture and shape features.



Distribution of abnormalities in BCDR-FMR and in the BCDR-F01



Age distribution and breast density distribution per age interval in the BCDR-FMR



Age distribution and breast density distribution per age interval in the BCDR-F01

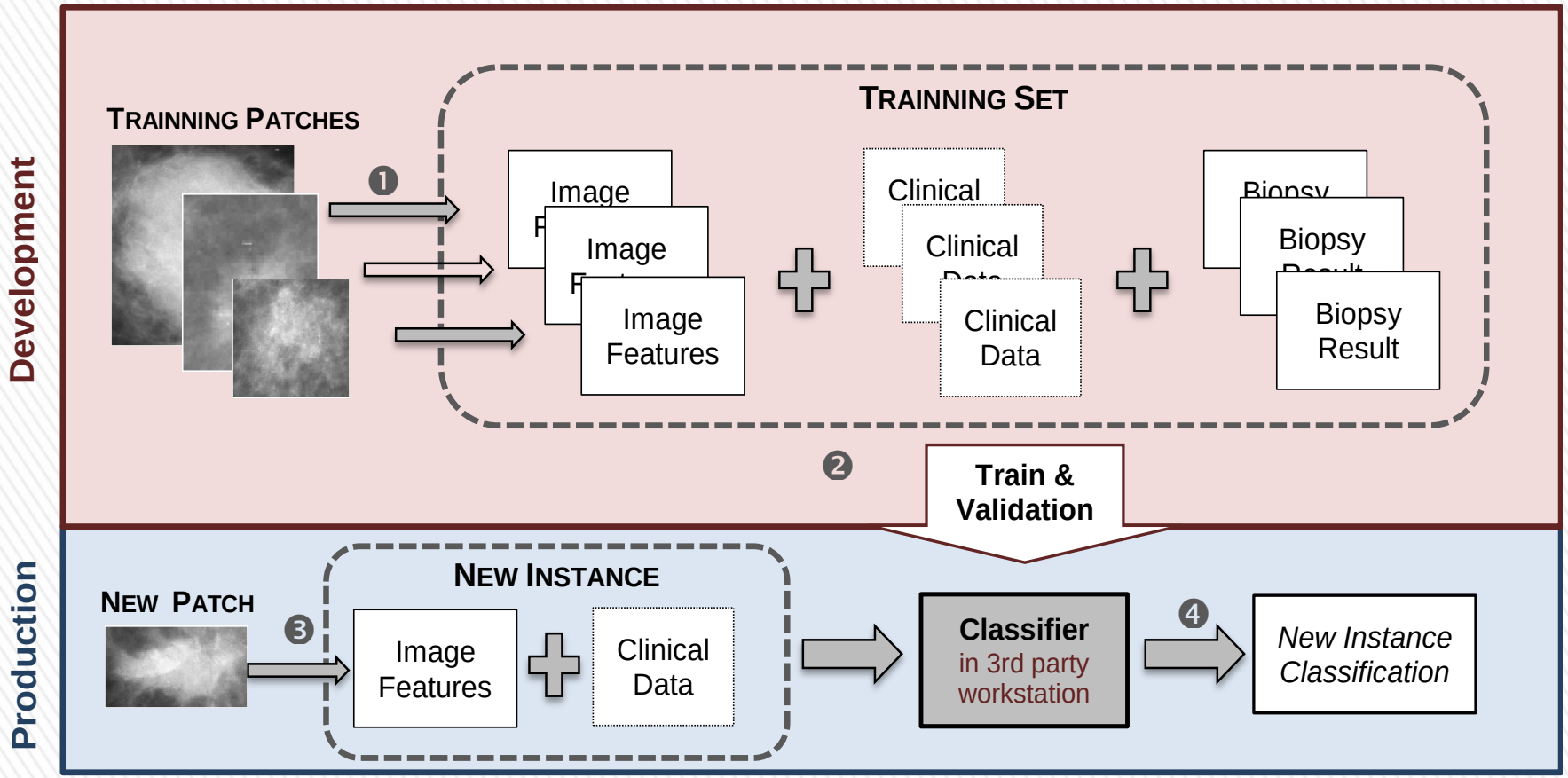
Dataset: BCDR-F01

Dataset: BCDR-D01

- Formed from **79 lesions**: 49 benign and 30 malignant (biopsy proven)
- Composed by **143 features vectors** (86 instances related to the 49 benign lesions and 57 instances related to the 30 malignant lesions)
- Each vector includes clinical, intensity, texture and shape features.

- To be released soon on www.bcdr.inegi.pt

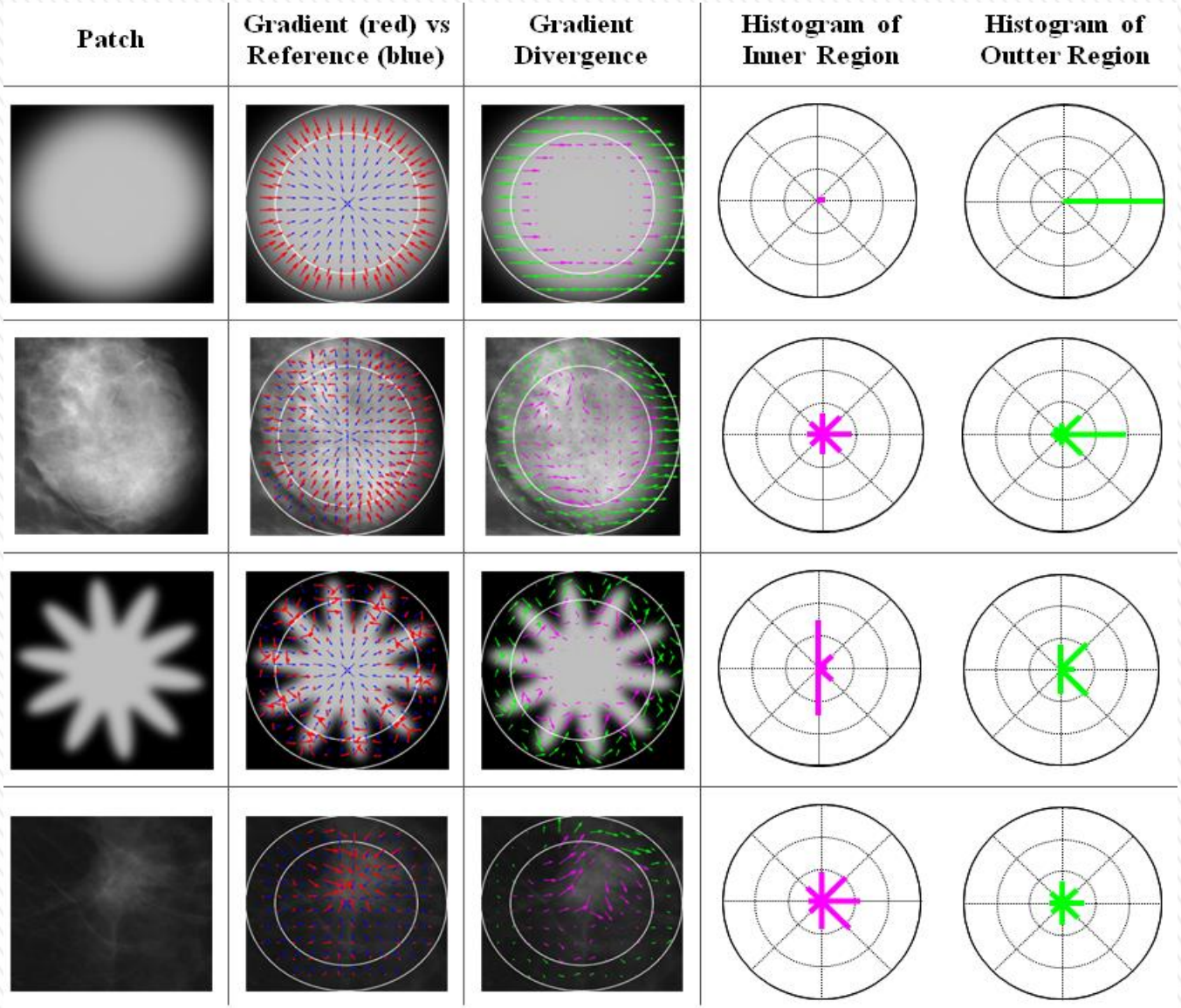
- New datasets are under construction.



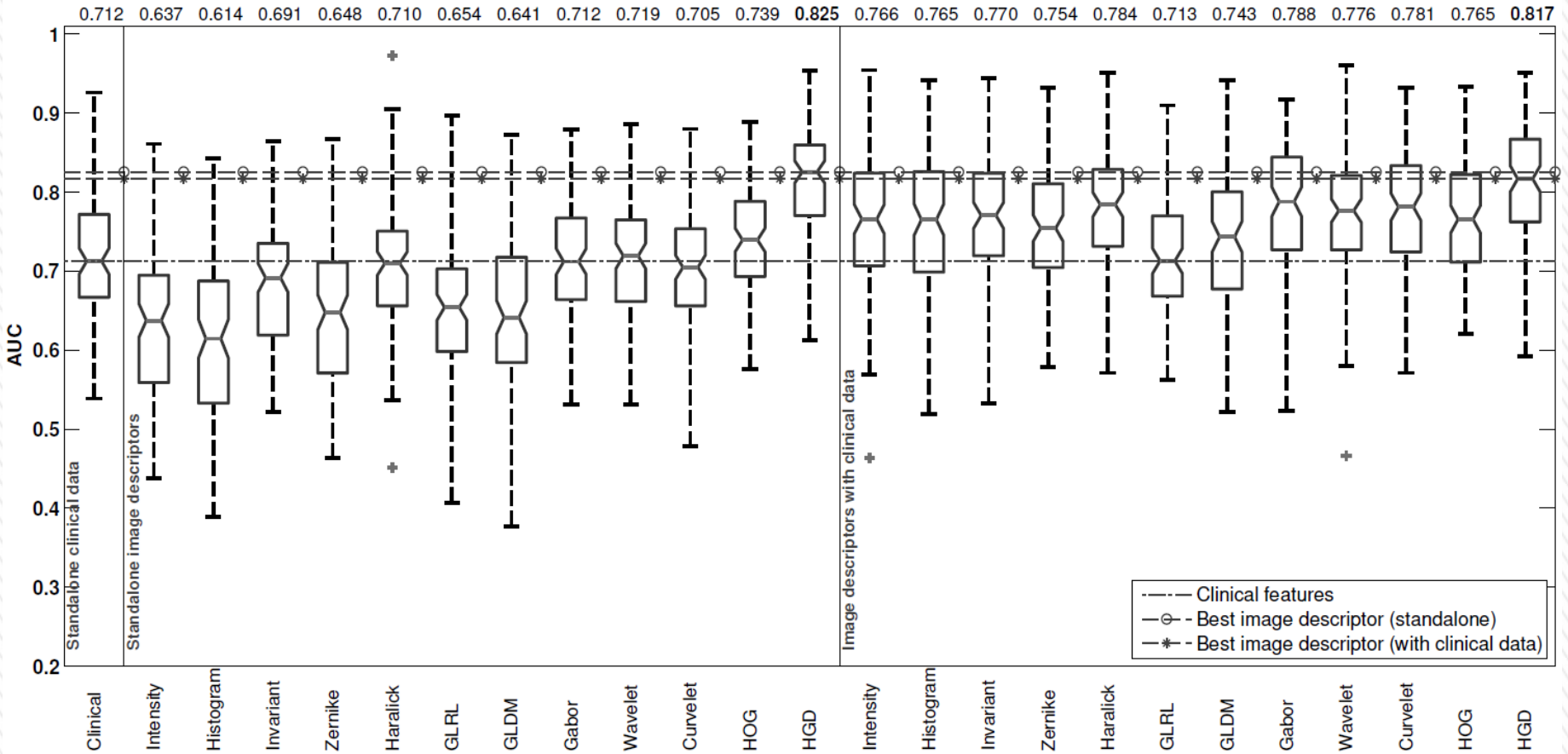
CADx
from development to production

- Different Datasets
  - e.g. size, population, abnormalities
- Several Descriptors
  - e.g. Intensity, Texture, Shape, Multi-scale
- Combining clinical data
  - e.g. Age, Breast Density, Observed abnormalities
- Several classifiers
  - e.g. SVM, Neural Networks, Random Forests

Experimental strategies

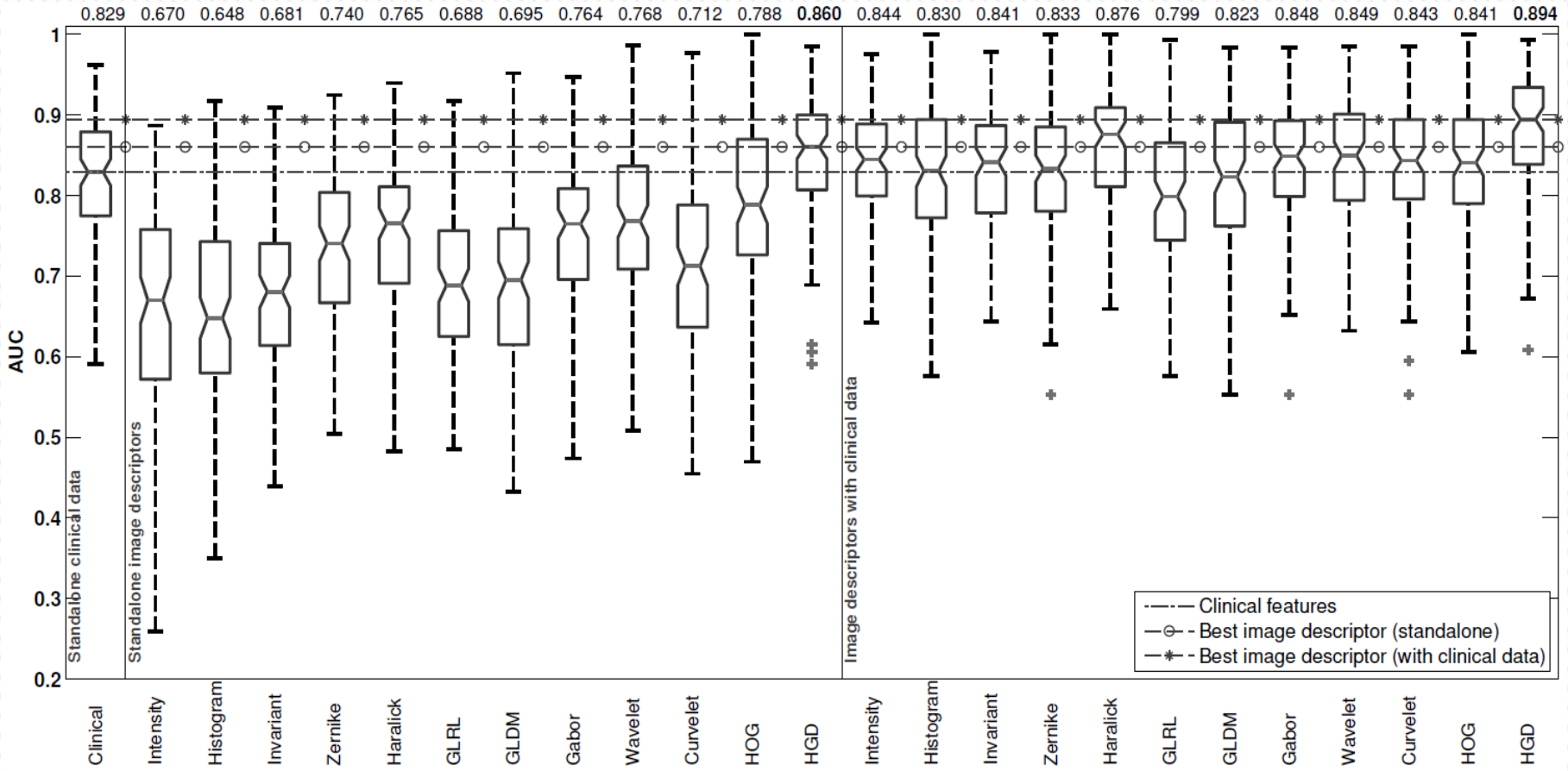


NEW!!! HGD – Histograms of Gradient Divergence



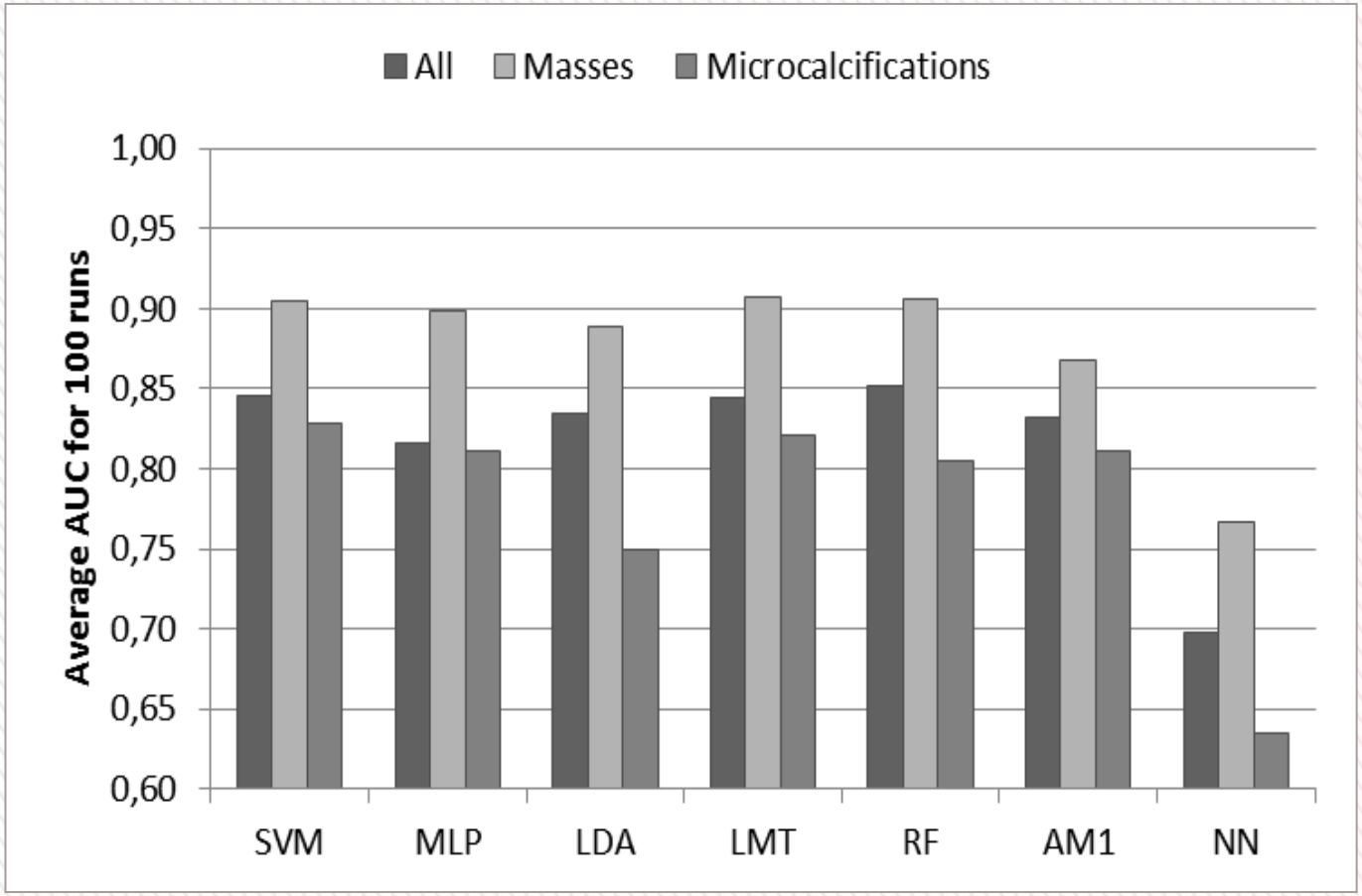
Moura, D. and M. Guevara López (2013). "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis." International Journal of Computer Assisted Radiology and Surgery: 1-14.

Results – All Lesions



Moura, D. and M. Guevara López (2013). "An evaluation of image descriptors combined with clinical data for breast cancer diagnosis." International Journal of Computer Assisted Radiology and Surgery: 1-14.

Results – Masses



SVM: Support Vector Machine ***MLP***: Multilayer Perceptron ***LDA***: Linear Discriminant Analysis ***LMT***: logistic Model Trees
RF: Random Forests ***AM1***: Adaboost M1 ***NN***: Nearest Neighbor

Benchmarking Results

›

Benchmarking Results

- Results for the Digital dataset are under construction
- Preliminary results (leave one out cross-validation) show AUC near 0.95 for single view evaluation
- Journal article under preparation with benchmarking results for the two datasets

Conclusion

- Successful collaboration between FMUP – INEGI – CIEMAT
  - Public repository with biopsy proven, clinically validated datasets
  - CADx workstation prototype
  - Good practices award from Fundación de la Ciencia y la Tecnología Española (FECYT)
- New descriptor (HGD) showing promising results for mass classification
- CADx evaluation and certification