

Aplicação de Algoritmos de Aprendizagem Automática para a Previsão de Cancro de Mama



FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

U.PORTO

Pedro Miguel Ferreira – MIERSI

Orientação:
Prof. Doutora Inês Dutra
Doutor Nuno Fonseca

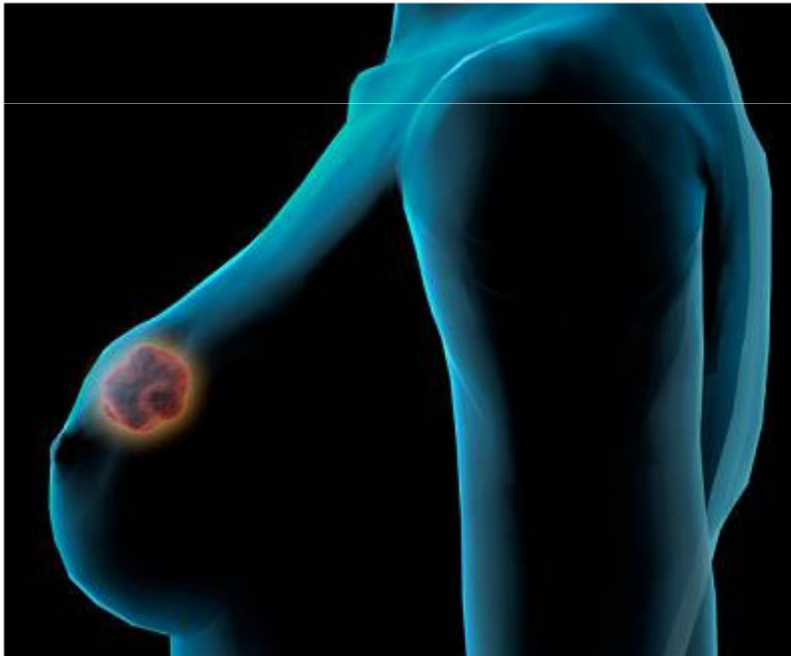
Organização

- Introdução
- Motivação
- Dados
- Objectivos
- Materiais e Métodos
- Resultados e Análise
- Conclusões e Trabalho Futuro

Organização

- Introdução
- Motivação
- Dados
- Objectivos
- Materiais e Métodos
- Resultados e Análise
- Conclusões e Trabalho Futuro

Cancro de Mama



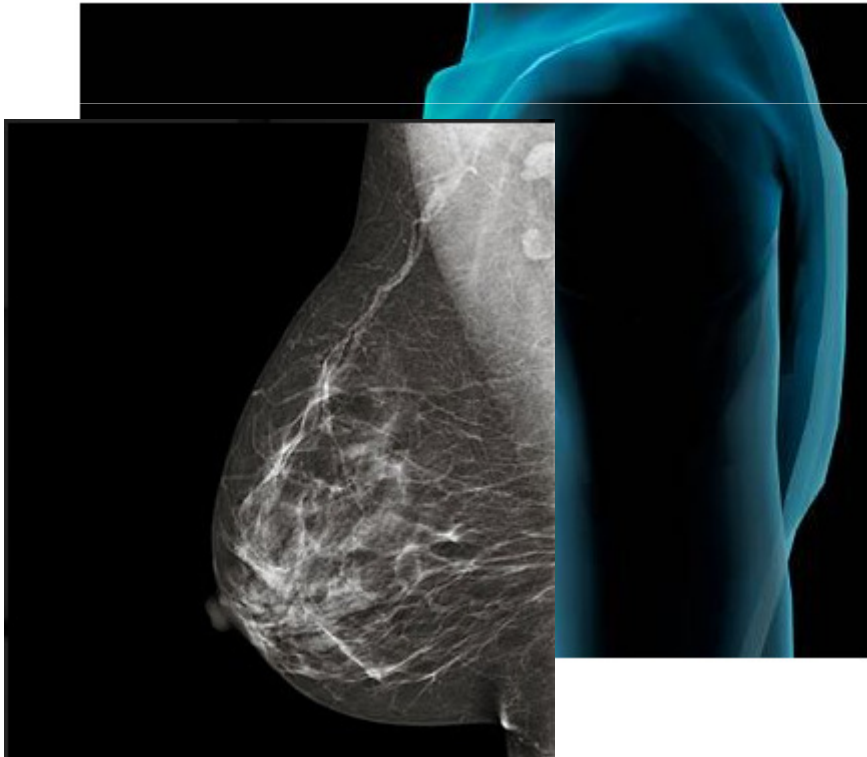
- EUA:
 - 1/8 mulheres desenvolve cancro de mama
 - Em 2006:
 - 191.410 com cancro
 - 40.820 ($\approx 21\%$) morreram

Fonte: *U. S. Cancer Statistics Working Group* - Outubro de 2010

- Portugal:
 - Por ano:
 - 4500 novos casos
 - 1500 óbitos (33%)

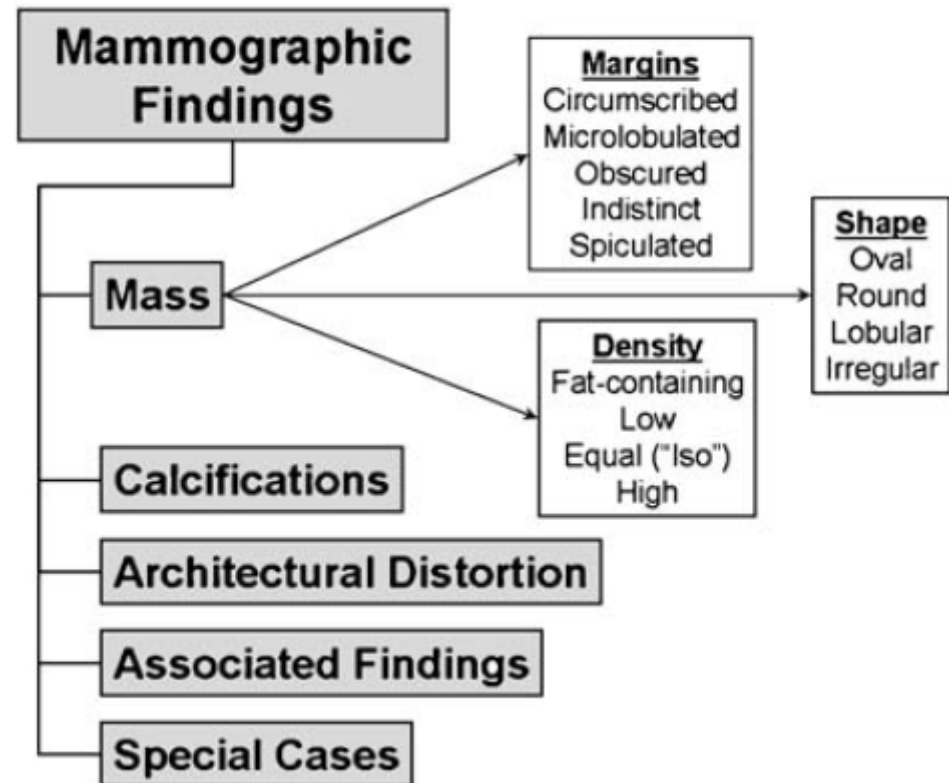
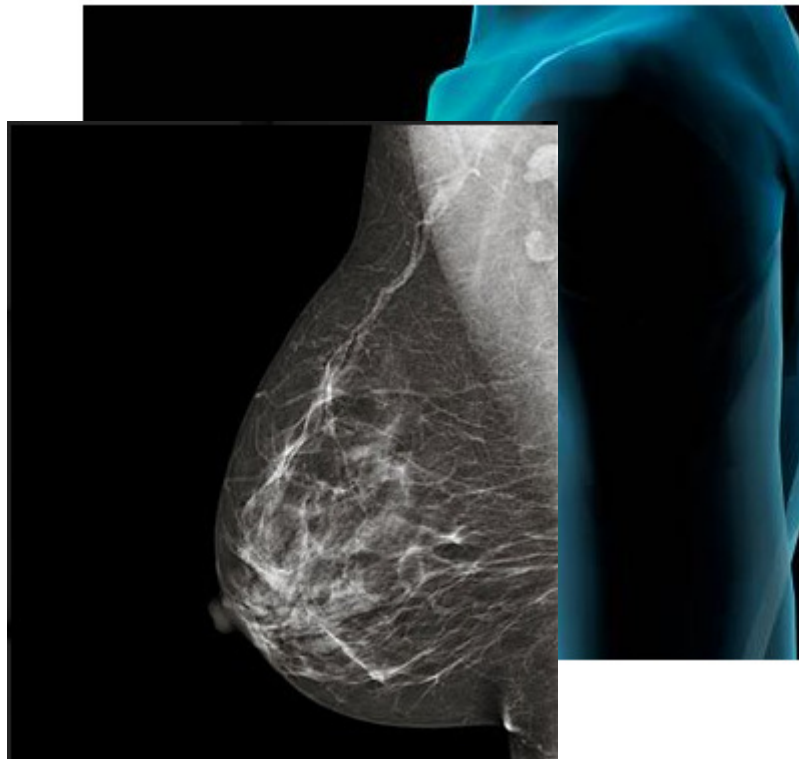
Fonte: *Liga Portuguesa Contra o Cancro* - Outubro de 2010

Rastreio



- Redução da taxa de mortalidade na ordem dos 30%
- Mamografia:
Método + económico e eficiente para a detecção de cancro de mama num estado precoce

Mamografia



Organização

- Introdução
- **Motivação**
- Dados
- Objectivos
- Materiais e Métodos
- Resultados e Análise
- Conclusões e Trabalho Futuro

Motivação

- Avaliar a **relevância de densidade de massa** dos nódulos como factor preponderante no diagnóstico de cancro de mama

Motivação

- Tentativa de construção de classificadores capazes de preverem **densidade de massa e malignidade**



- Reduzir número de biopsias



Organização

- Introdução
- Motivação
- **Dados**
- Objectivos
- Materiais e Métodos
- Resultados e Análise
- Conclusões e Trabalho Futuro

Dados

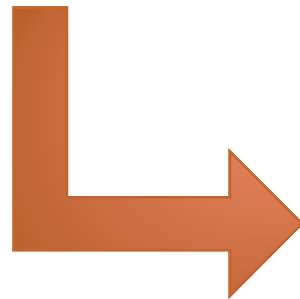
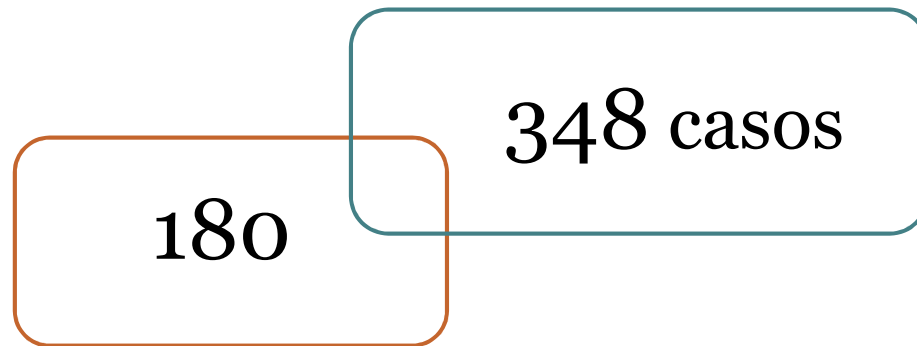


- Facultados por:
 - Dra. Elizabeth Burnside
 - Dr. Ryan Woods



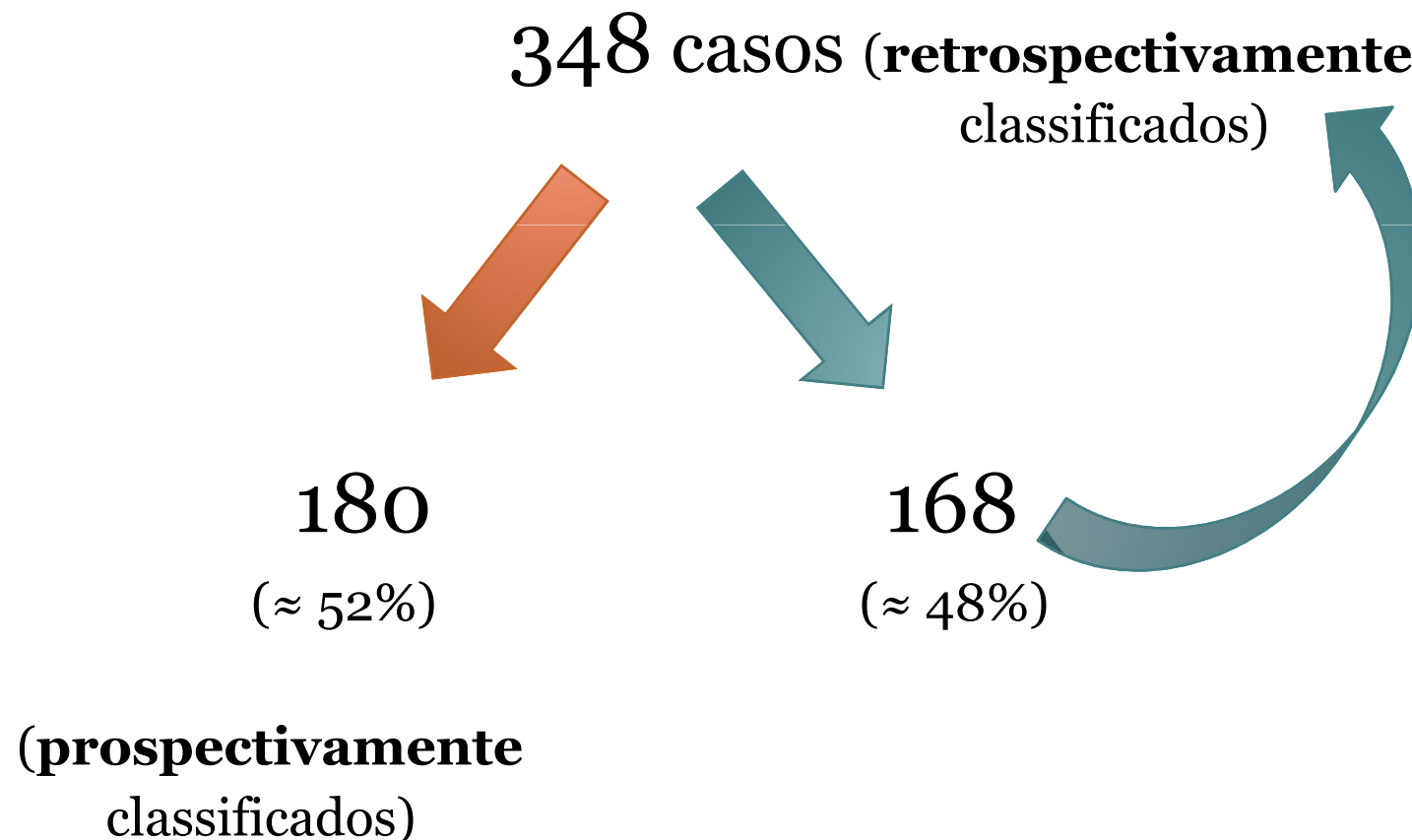
- 348 casos
- Cada caso corresponde a um nódulo da mama, sendo classificado **retrospectivamente** de acordo com sistema BI-RADS®
- Fruto dos resultados de mamografias
- Recolhidos entre Outubro 2005 e Dezembro 2007

Classificação dos nódulos



classificados **prospectivamente**
em termos de **densidade** por
um único radiologista

Classificação dos nódulos



Classificação dos nódulos

Prospectiva

- Classificação do atributo **densidade de massa** para os 180 casos, por apenas um radiologista:
 - *low density*;
 - *iso-dense*;
 - *high density*;
- Espécie de relatório médico breve e superficial;
- **Classificação** efectuada **sob stress**.

Retrospectiva

- Obtida numa espécie de reunião periódica entre radiologistas e médicos experientes em que **reavaliam exames**;
- **Revista a classificação de densidade de massa** efectuada pelo radiologista (estudo prospectivo);
- **Classificação** obtida **sem stress**;
- Valores de **referência** para atributo **densidade de massa**;

Atributos

- Estudo Prospectivo
 - Composto por 33 atributos
 - ↓ densidade
 - ***density_num***
- Estudo Retrospectivo
 - Composto por 35 atributos
 - ↓ densidade
 - ***retro_density***

density_num Vs *retro_density*



QUADRANT_LOCATION_def	SIDE_def	DEPTH_def	SIZE	OVERALL_BREAST_COMPOSITION	Density_num	retro_density	outcome_num
Upper Outer	Right	Posterior	11	heterogeneously dense	iso	iso	malignant
Upper Inner	Right	Middle	8	scattered fibroglandular densities	iso	iso	benign
Upper Inner	Right	Middle	8	heterogeneously dense	high	iso	benign
Upper Inner	Right	Middle	8	heterogeneously dense	iso	iso	benign
Upper Outer	Left	Middle	8	heterogeneously dense	high	iso	benign
Lower Inner	Right	Middle	10	scattered fibroglandular densities	high	iso	benign
	Left	Middle	10	scattered fibroglandular densities	high	iso	malignant
Upper Inner	Left	Middle	20	heterogeneously dense	high	high	benign
	Left		15	heterogeneously dense	iso	high	benign
Upper Outer	Right	Middle	20	extremely dense	iso	iso	benign
Upper Outer	Left	Anterior	20	heterogeneously dense	iso	iso	benign
Upper Inner	Right	Posterior	6	heterogeneously dense	iso	iso	malignant
Upper Outer	Right	Posterior	12	almost entirely fat	iso	iso	malignant
Lower Inner	Left	Posterior	18	scattered fibroglandular densities	high	high	malignant
Upper Outer	Left	Anterior	8	almost entirely fat	high	iso	benign
Lower Inner	Right	Posterior	15	heterogeneously dense	high	iso	benign
Upper Inner	Left	Middle	10	scattered fibroglandular densities	iso	iso	benign
	Left		30	heterogeneously dense	high	high	malignant

Atributos Utilizados

Atributos Utilizados
reread_group
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



Classificação dos nódulos em termos de **malignidade** fruto de resultados de **biopsias**

Organização

- Introdução
- Motivação
- Dados
- **Objectivos**
- Materiais e Métodos
- Resultados e Análise
- Conclusões e Trabalho Futuro

Objectivos

- **Foco:**
Forma como **densidade de massa** dos nódulos poderá influenciar **malignidade**;
- Encontrar relações entre atributos
 - Aplicação de técnicas de Aprendizagem Automática;
- “Aprender” modelos capazes de auxiliarem os médicos na avaliação de mamografias.

Organização

- Introdução
- Motivação
- Dados
- Objectivos
- **Materiais e Métodos**
- Resultados e Análise
- Conclusões e Trabalho Futuro

Materiais e Métodos

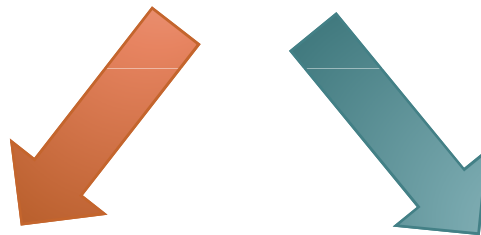
- WEKA



- ***10-fold cross-validation***
- ***Paired Corrected T-Tester de 0.01***

Materiais e Métodos

348 casos



180



Aprendizagem
10-fold cross-validation

168



Teste

Resultados e Análise

- Q_1 – Será densidade de massa um factor relevante no diagnóstico de cancro de mama?
- Q_2 – Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?
- Q_3 – Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

Resultados e Análise

- Q_1 – Será densidade de massa um factor relevante no diagnóstico de cancro de mama?
- Q_2 – Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?
- Q_3 – Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

Materiais e Métodos

348 casos



180



Aprendizagem
10-fold cross-validation

168



Teste

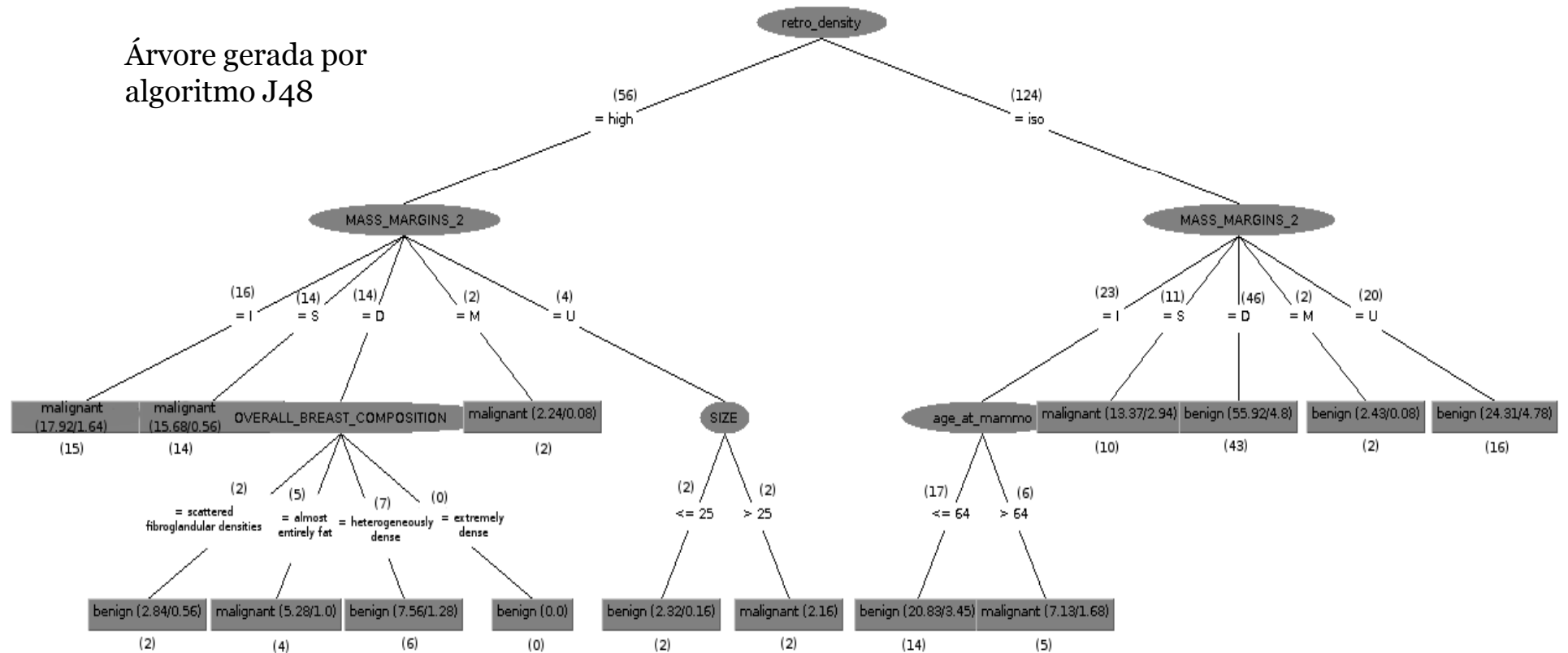
Q_1 - Será densidade de massa um factor relevante no diagnóstico de cancro de mama?

180 casos

- (E_1) – Previsão de *outcome_num* com *retro_density*
 - (E_2) – Previsão de *outcome_num* com *density_num*
 - (E_3) – Previsão de *outcome_num* sem densidade de massa
- } Previsão de malignidade

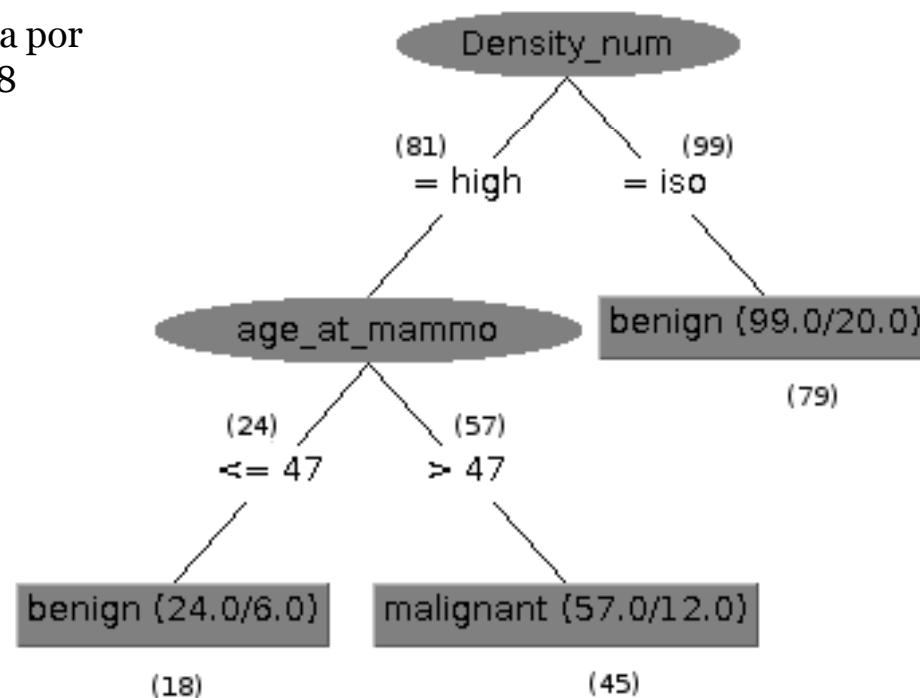
Q₁ - Previsão de *outcome_num* com *retro_density* (E₁)

Árvore gerada por
algoritmo J48



Q_1 - Previsão de *outcome_num* com *density_num* (E_2)

Árvore gerada por
algoritmo J48



SVM's

Q₁

180	Previsão de <i>outcome_num</i>		
	com densidade de massa		E_3
	E_1	E_2	
Métrica	Retrospectiva (<i>retro_density</i>)	Prospectiva (<i>density_num</i>)	sem densidade de massa
Instâncias Correctamente Classificadas	84.78% (7.96)	82.72% (8.32)	81.39% (8.81)
Estatística <i>Kappa</i>	0.68 (0.17)	0.63 (0.17)	0.60 (0.18)
Precisão	0.84 (0.12)	0.82 (0.13)	0.81 (0.14)
<i>Recall</i>	0.78 (0.15)	0.75 (0.15)	0.72 (0.15)
<i>F-Measure</i>	0.80 (0.11)	0.77 (0.11)	0.75 (0.12)

óstico de

- **Densidade de massa** tem alguma **influência** sobre **malignidade**, quando a densidade é a observada nos dados retrospectivos (E_1)
- (E_1) -> *retro_density* -> ICC = 84.78% (+/- 7.96)
- (E_3) -> sem densidade -> ICC = 81.39% (+/- 8.81)
- Resultados estatisticamente diferentes ($p = 0.01$)

Resultados e Análise

- Q_1 – Será densidade de massa um factor relevante no diagnóstico de cancro de mama?
- Q_2 – Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?
- Q_3 – Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

Materiais e Métodos

348 casos



180



Aprendizagem
10-fold cross-validation

168



Teste

Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

180 casos

- (E_4) – Previsão de *retro_density*
 - (E_5) – Previsão de *density_num*
- } Previsão de densidade

Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?



naïve Bayes

Id_Exam	reread_group	age_at_mammo	MASS_MARGINS_worst	SIZE	OVERALL_BREAST_COMPOSITION	retro_density	Density_num
94	salkowski	41	Circumscribed	13	scattered fibroglandular densities	iso	iso
95	salkowski	69		12	scattered fibroalldular densities	iso	iso

180	Previsão de densidade de massa	
Métrica	E_4 <i>retro_density</i>	E_5 <i>density_num</i>
Instâncias Correctamente Classificadas	72.83% (9.89)	67.22% (12.14)
Estatística Kappa	0.37 (0.23)	0.33 (0.25)
Precisão	0.58 (0.20)	0.66 (0.16)
Recall	0.58 (0.22)	0.60 (0.17)
F-Measure	0.56 (0.18)	0.62 (0.15)

145	salkowski	49	Circumscribed	13	scattered fibroglandular densities	iso	iso
146	burnside	35	Indistinct	28	heterogeneously dense	iso	iso
148	sisney	56	Indistinct	6	heterogeneously dense	iso	iso

126 rows fetched in 0:00.1131

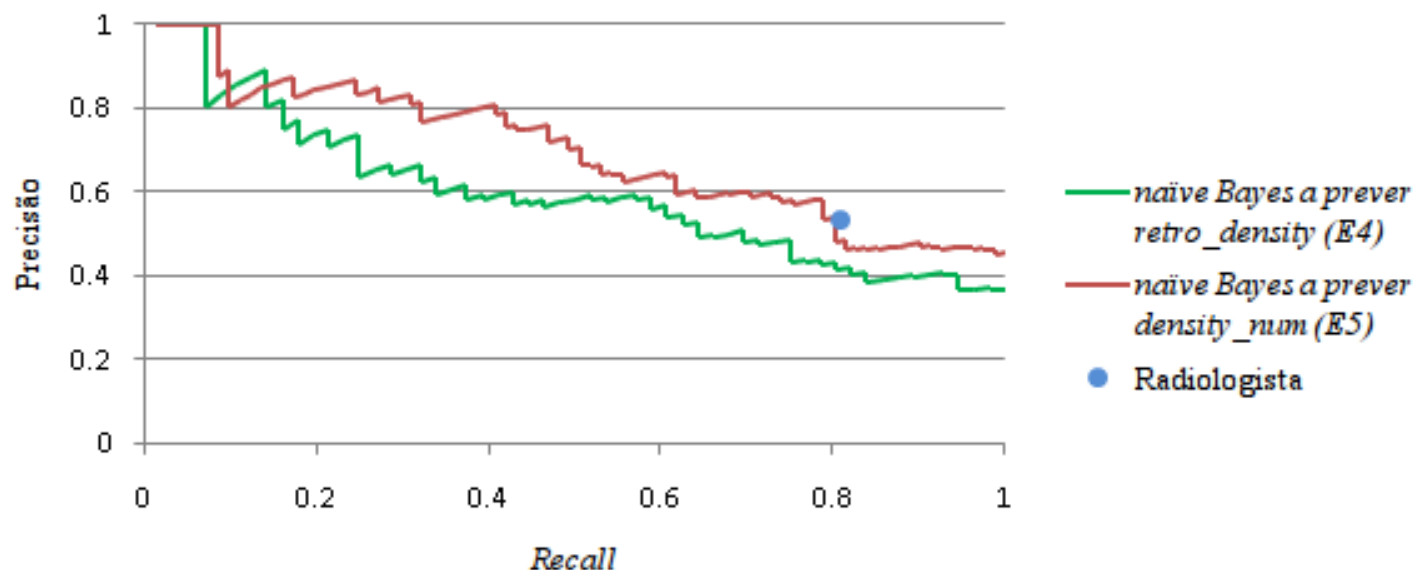
Start Editing Apply Changes

- Radiologista classificou em termos de densidade 70% das instâncias de acordo com a classificação retrospectiva

Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

Curvas *Precision-Recall*

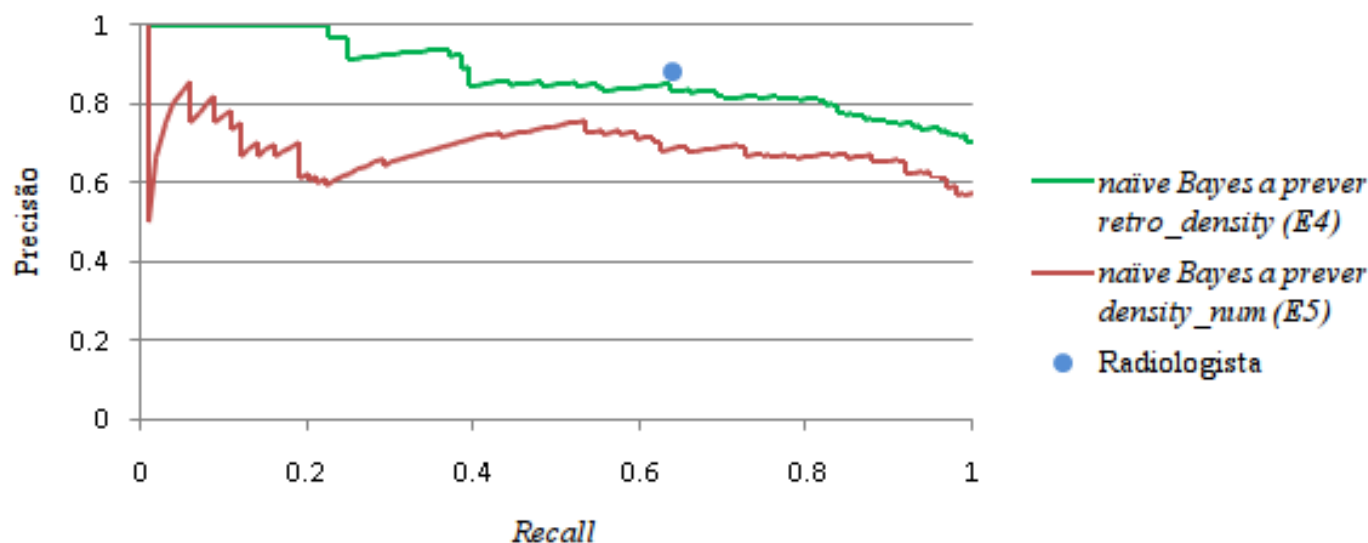
Previsão de *retro_density* e *density_num*
(180 casos - classe *high*)



Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

Curvas *Precision-Recall*

Previsão de *retro_density* e *density_num*
(180 casos - classe iso)

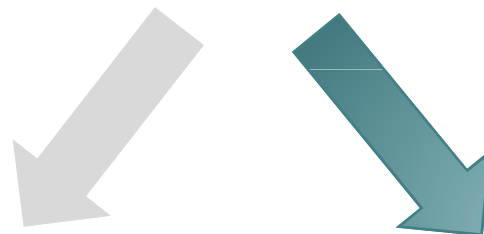


Resultados e Análise

- Q_1 – Será densidade de massa um factor relevante no diagnóstico de cancro de mama?
- Q_2 – Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?
- Q_3 – Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

Materiais e Métodos

348 casos



180



Aprendizagem
10-fold cross-validation

168



Teste

Q_3 . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

168 casos

- (E_6) – Previsão de *retro_density*
 - (E_7) – Previsão de *density_num*
- } Previsão de densidade

• E_8, E_9, E_{10}, E_{11}

} Previsão de malignidade

Q_3 . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

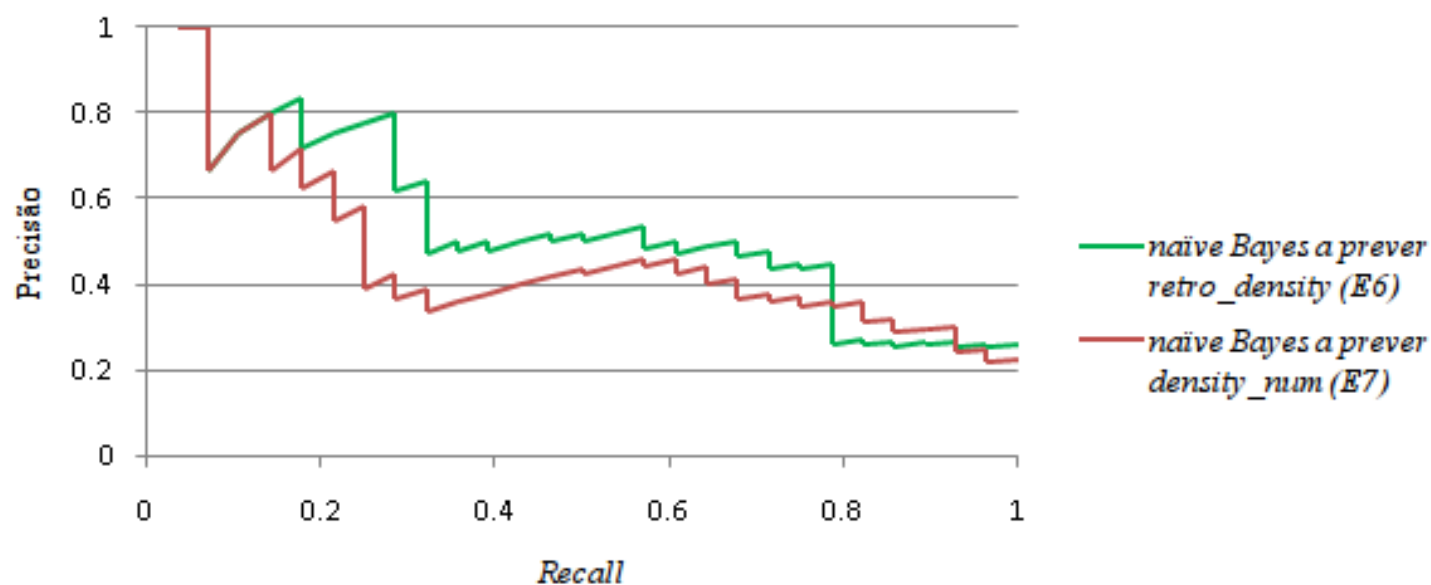
168	Previsão de densidade de massa	
Métrica	E_6 <i>retro_density</i>	E_7 <i>density_num</i>
Instâncias Correctamente Classificadas	82.14%	75.60%
Estatística <i>Kappa</i>	0.45	0.35
Precisão	0.48	0.38
<i>Recall</i>	0.68	0.71
<i>F-Measure</i>	0.56	0.49

- Classificadores baseados no algoritmo *naïve Bayes*

Q₃ . Qual o comportamento conjunto de dados

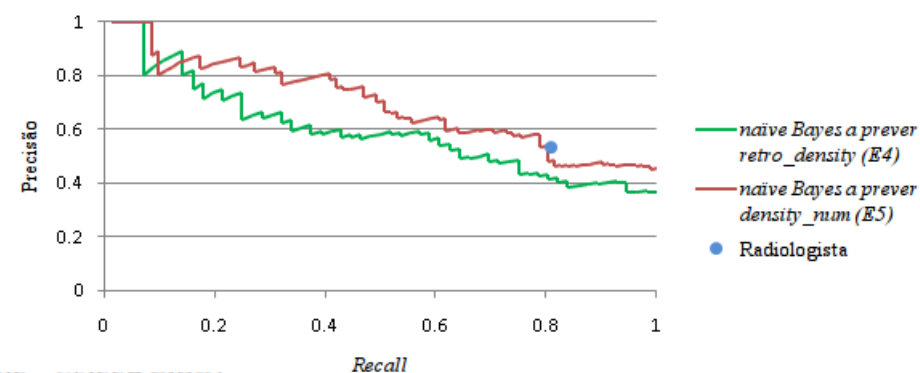
Curvas P_i

Previsão de *retro*
(168 cas



Curvas *Precision-Recall*

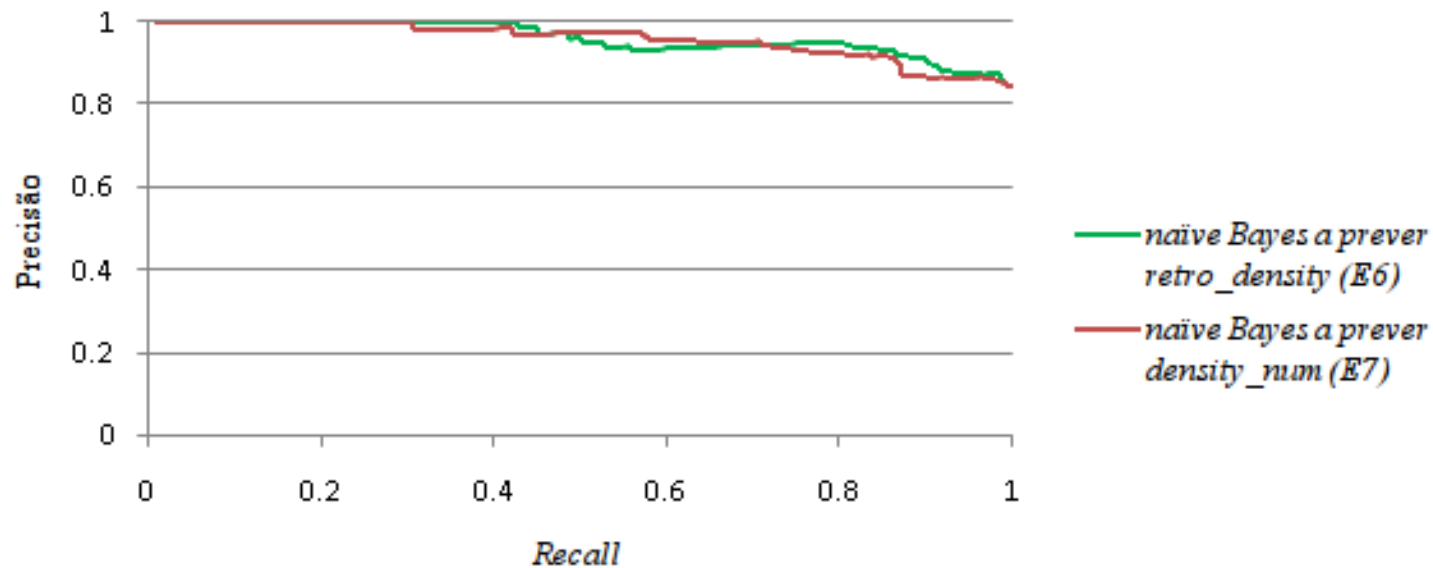
Previsão de *retro_density* e *density_num*
(180 casos - classe *high*)



Q₃ . Qual o comportamento
conjunto de dados

Curvas Precision-Recall

Previsão de *retro_density*
(168 casos)



Q_3 . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

168 casos

- (E_8) – Previsão de *outcome_num* com *retro_density*
- (E_9) – Previsão de *outcome_num* com *retro_density* prevista em E_6
- (E_{10}) – Previsão de *outcome_num* com *density_num* prevista em E_7
- (E_{11}) – Previsão de *outcome_num* sem densidade de massa

Q_3 . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

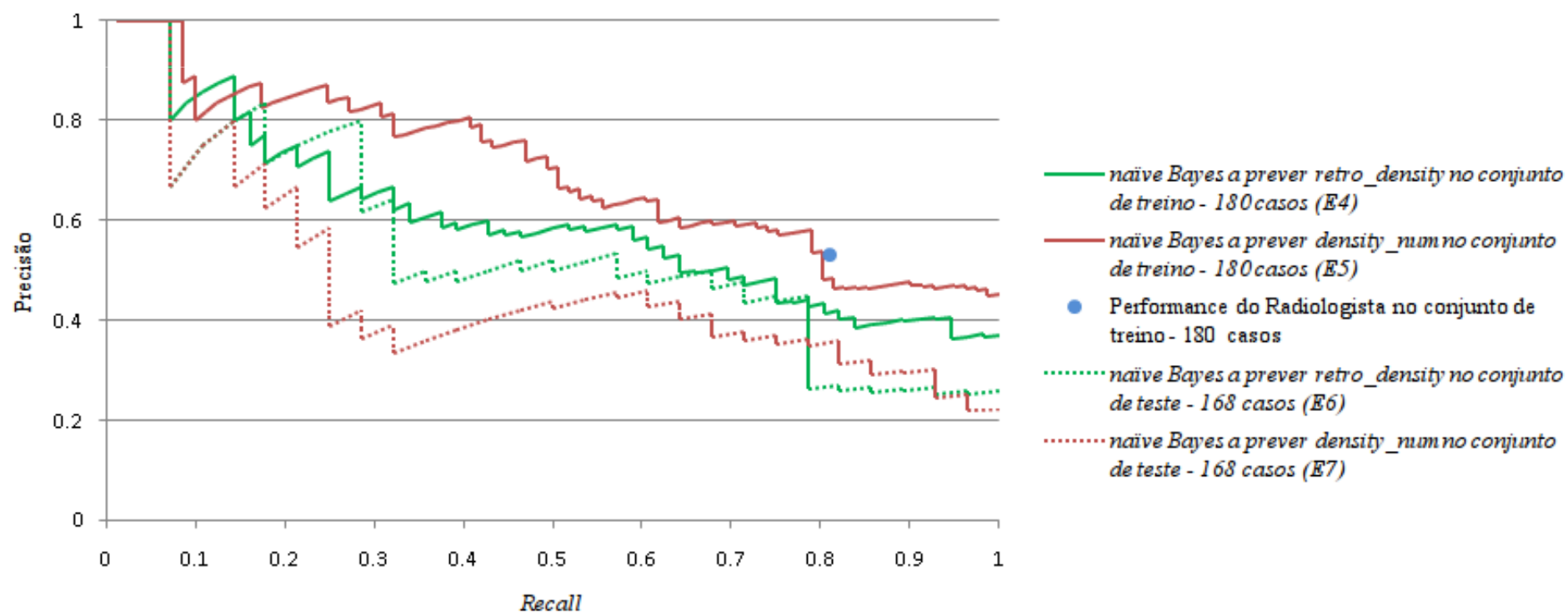
168	Previsão de <i>outcome_num</i>			
	com densidade de massa			E_{11}
Métrica	E_8 Retrospectiva (<i>retro_density</i>) (actual)	E_9 Retrospectiva (<i>retro_density</i>) (preenchida pelo classificador <i>naive Bayes</i>)	E_{10} Prospectiva (<i>density_num</i>) (preenchida pelo classificador <i>naive Bayes</i>)	sem densidade de massa
Instâncias Correctamente Classificadas	81.55%	79.76%	79.17%	77.38%
Estatística <i>Kappa</i>	0.52	0.48	0.46	0.42
Precisão	0.70	0.65	0.65	0.61
<i>Recall</i>	0.60	0.60	0.55	0.53
<i>F-Measure</i>	0.64	0.62	0.60	0.57

- Classificadores baseados em SVM's

Resumo (Previsão de densidade)

Curvas *Precision-Recall*

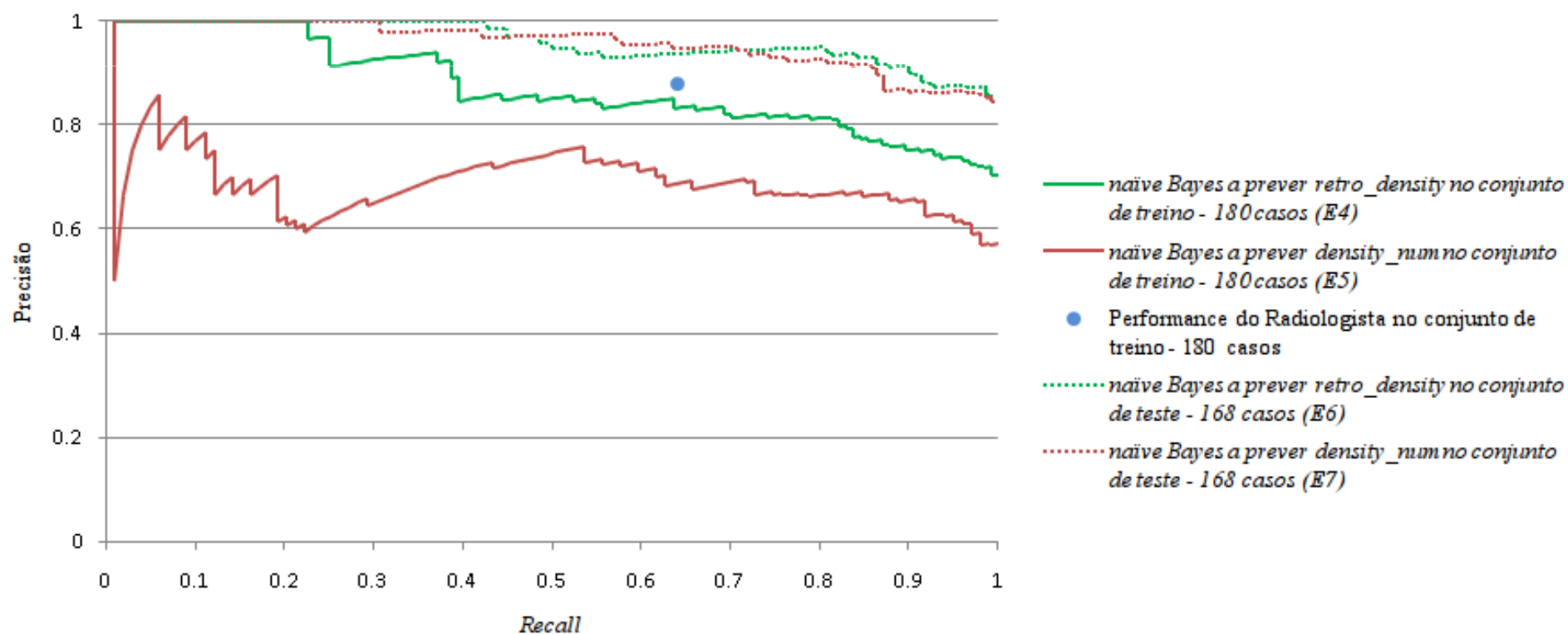
Previsão de *retro_density* e *density_num*
(classe *high*)



Resumo (Previsão de densidade)

Curvas *Precision-Recall*

Previsão de *retro_density* e *density_num*
(classe *iso*)



Resumo (Previsão de malignidade)

180/168	Previsão de <i>outcome_num</i>						
	com densidade de massa					E_3 (180)	E_{11} (168)
Métrica	E_1 (180) Retrospectiva (<i>retro_density</i>)	E_8 (168) Retrospectiva (<i>retro_density</i>) (actual)	E_9 (168) Retrospectiva (<i>retro_density</i>) (preenchida pelo classificador <i>naive</i> <i>Bayes</i>)	E_2 (180) Prospectiva (<i>density_num</i>)	E_{10} (168) Prospectiva (<i>density_num</i>) (preenchida pelo classificador <i>naive</i> <i>Bayes</i>)	sem densidade de massa	sem densidade de massa
Instâncias Correctamente Classificadas	84.78% (7.96)	81.55%	79.76%	82.72% (8.32)	79.17%	81.39% (8.81)	77.38%
Estatística <i>Kappa</i>	0.68 (0.17)	0.52	0.48	0.63 (0.17)	0.46	0.60 (0.18)	0.42
Precisão	0.84 (0.12)	0.70	0.65	0.82 (0.13)	0.65	0.81 (0.14)	0.61
<i>Recall</i>	0.78 (0.15)	0.60	0.60	0.75 (0.15)	0.55	0.72 (0.15)	0.53
<i>F-Measure</i>	0.80 (0.11)	0.64	0.62	0.77 (0.11)	0.60	0.75 (0.12)	0.57

- Classificadores baseados em SVM's

Conclusões e Trabalho Futuro

- a) Classificação automática de uma mamografia poderá alcançar resultados semelhantes ou mesmo superiores aos obtidos por radiologistas sob stress;
- b) Densidade de massa assume-se como um bom indicador de malignidade;
- c) Os classificadores de aprendizagem automática são capazes de prever densidade de massa com um nível qualitativo semelhante ao de um radiologista.

Conclusões e Trabalho Futuro

- a) Estender este estudo a universos de dados maiores e geograficamente distintos;
- b) Aplicar técnicas de aprendizagem automática baseadas em aprendizagem estatística relacional.

Divulgação

- Ferreira P., Dutra I., Fonseca N. A., Woods R., and Burnside E.
Studying the relevance of Breast Imaging Features – HealthInf
2011 – International Conference on Health Informatics. January
26th - 29th, Rome, Italy.

Obrigado



FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

U.PORTO

Pedro Miguel Ferreira – MIERSI

Orientação:
Prof. Doutora Inês Dutra
Doutor Nuno Fonseca

Anexos

Materiais e Métodos

10-fold cross-validation

Iteração

1



2



3



4



5

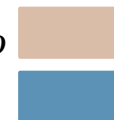


(...)

(...)

Treino

Teste



Q_1 - Será densidade de massa um factor relevante no diagnóstico de cancro de mama?

180	Previsão de <i>outcome_num</i>		
	com densidade de massa		E_3
Métrica	E_1 Retrospectiva (<i>retro_density</i>)	E_2 Prospectiva (<i>density_num</i>)	sem densidade de massa
Instâncias Correctamente Classificadas	84.78% (7.96)	82.72% (8.32)	81.39% (8.81)
Estatística <i>Kappa</i>	0.68 (0.17)	0.63 (0.17)	0.60 (0.18)
Precisão	0.84 (0.12)	0.82 (0.13)	0.81 (0.14)
<i>Recall</i>	0.78 (0.15)	0.75 (0.15)	0.72 (0.15)
<i>F-Measure</i>	0.80 (0.11)	0.77 (0.11)	0.75 (0.12)

- Classificadores baseados em **SVM's**

Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

180	Previsão de densidade de massa	
Métrica	E_4 <i>retro_density</i>	E_5 <i>density_num</i>
Instâncias Correctamente Classificadas	72.83% (9.89)	67.22% (12.14)
Estatística <i>Kappa</i>	0.37 (0.23)	0.33 (0.25)
Precisão	0.58 (0.20)	0.66 (0.16)
<i>Recall</i>	0.58 (0.22)	0.60 (0.17)
<i>F-Measure</i>	0.56 (0.18)	0.62 (0.15)

- Classificadores baseados no algoritmo *naïve Bayes*

Conclusões e T

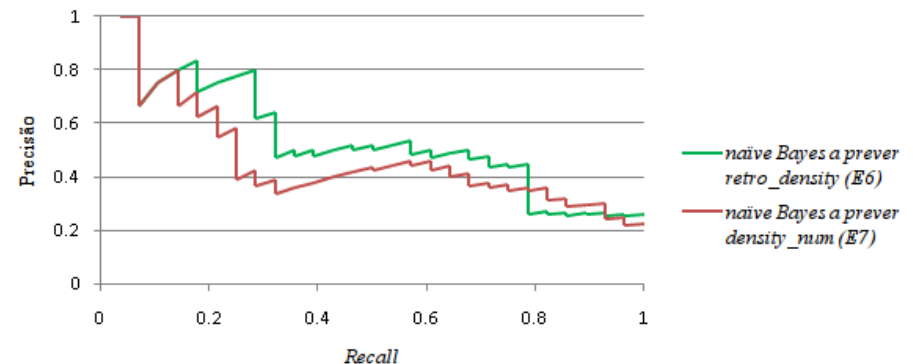
- Considerações finais so

- Para previsão instâncias *high*:

- Informação relativa ao estudo retrospectivo (*retro_density*) não auxilia classificador ao longo da aprendizagem. Poderá ajudar na fase de classificação de novas instâncias (fase de teste);

Curvas Precision-Recall

Previsão de *retro_density* e *density_num*
(168 casos - classe *high*)

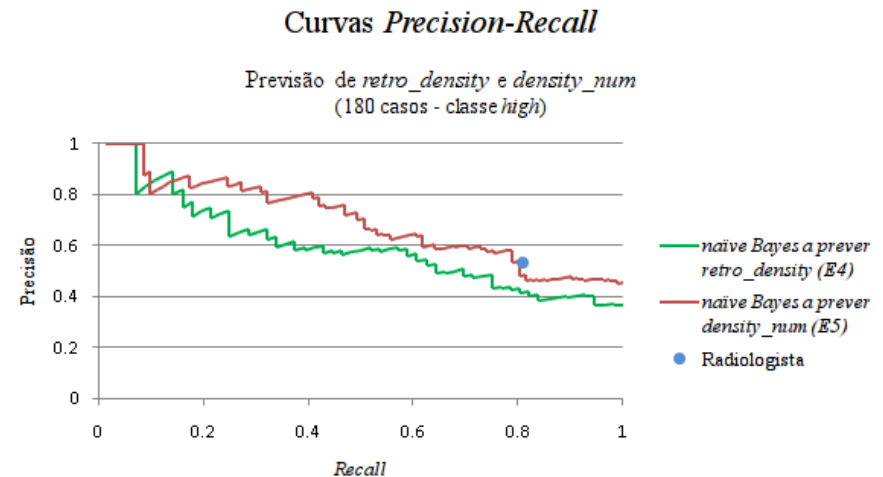


Conclusões e Ti

- Considerações finais sol

- Para previsão instâncias *high*:

- Classificadores muito próximos da performance do radiologista;
- Classificadores podem alcançar valores de *recall* + elevados, caso seja possível comprometer algumas das instâncias negativas (perder precisão).



Conclusões e T

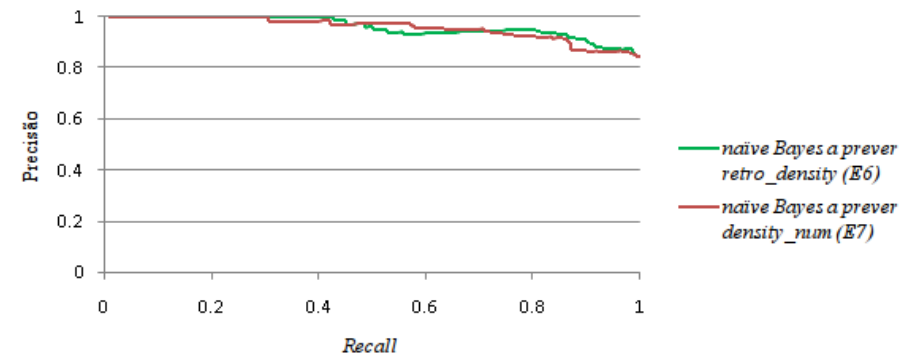
- Considerações finais sc

- Para previsão instâncias iso:

- Classificadores apresentam desempenho com erro baixo na presença de novos dados;
- Performance dos classificadores superior ao desempenho do radiologista.

Curvas Precision-Recall

Previsão de *retro_density* e *density_num*
(168 casos - classe iso)

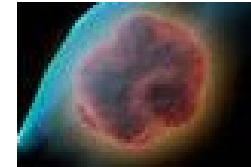


Previsão de densidade (resumo)

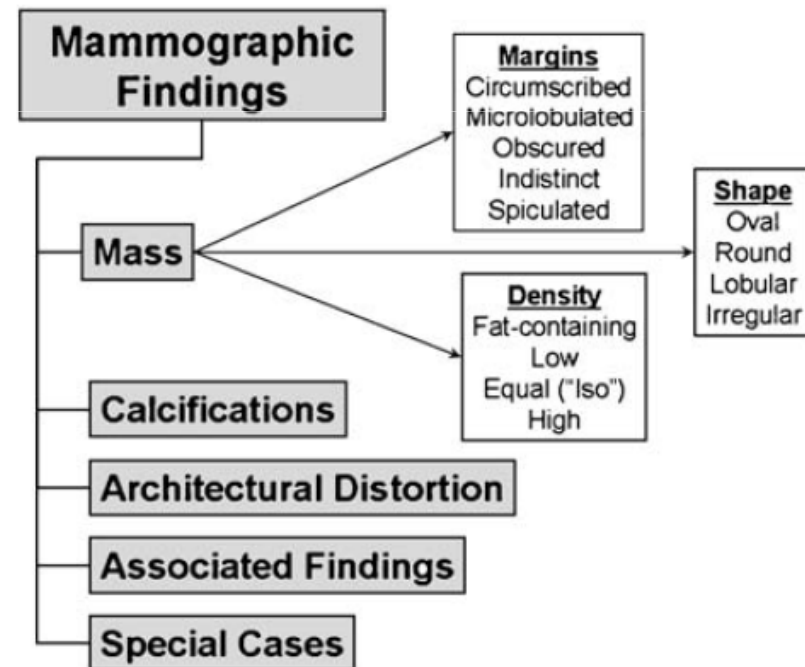
180/168	Previsão de densidade de massa				
Métrica	Radiologista (180)	E_5 <i>density_num</i> (180)	E_7 <i>density_num</i> (168)	E_4 <i>retro_density</i> (180)	E_8 <i>retro_density</i> (168)
Instâncias Correctamente Classificadas	70.00%	67.22% (12.14)	75.60%	72.83% (9.89)	82.14%
Estatística <i>Kappa</i>	0.52	0.33 (0.25)	0.35	0.37 (0.23)	0.45
Precisão	0.53	0.66 (0.16)	0.38	0.58 (0.20)	0.48
<i>Recall</i>	0.81	0.60 (0.17)	0.71	0.58 (0.22)	0.68
<i>F-Measure</i>	0.64	0.62 (0.15)	0.49	0.56 (0.18)	0.56

- Classificadores baseados no algoritmo *naïve Bayes*

Classificação dos nódulos BI-RADS®



- Sistema BI-RADS® (*Breast Imaging Reporting and Data System*)
- Léxico que visa padronizar relatórios médicos
- Constituído por 43 descritores



Q_1 - Será densidade de massa um factor relevante no diagnóstico de cancro de mama?

- 2 tipos de densidades de massa:

- *retro_density*
- *density_num*



utilizamos ambos para
construir classificadores

Métricas

$$\text{Precisão} = \frac{TP}{(TP + FP)}$$

$$\text{Recall} = \text{Sensibilidade} = \text{TVP} = \frac{TP}{(TP + FN)}$$

Estado da Arte

- Trabalhos desenvolvidos na aplicação de métodos de aprendizagem automática a dados relativos a cancros de mama:
 - Desenvolvimento de um método multi-superfície de separação de padrões para diagnósticos médicos aplicados à citologia da mama por Wolberg e Mangasarian no início da década de 90 [WHW90];
 - Aplicação de redes neuronais artificiais a dados de mamografias [YWu93, HAA02, Aye10];

Estado da Arte

- Prognóstico da doença, recorrendo a métodos de aprendizagem indutiva [WNS95];
- Extracção de informação a partir de textos de relatórios médicos referentes a mamografias [HNa09];
- Estudo sobre a influência da idade no carcinoma ductal *in situ* (cancro de mama não invasivo + frequente) [HNa10];

Estado da Arte

- Estudos que comprovam importância da densidade de massa dos nódulos no diagnóstico de cancro de mama:
 - Investigação levada a cabo por Davis *et al* [JDa05];
 - Aplicação de programação lógica indutiva a conjunto de dados referentes a cancro de mama [RW009];
 - Aplicação de regressão logística e estatística *Kappa* a conjunto de dados de mamografias, chegando à conclusão que densidade de massa elevada é um indicador de malignidade importante com $k = 0.53$ [Woo10];

Materiais e Métodos

10-fold cross-validation

- Método de validação em que os dados são divididos em N subconjuntos (blocos de dimensão semelhante - *folds*) para aprendizagem de N iterações;
- No treino são utilizados $N-1$ blocos, e apenas 1 para teste, sendo este diferente a cada iteração;
- Processo repetido para as N amostras;
- Performance do classificador definida de acordo com média dos N testes;
- Vantagem ***cross-validation***: todos os dados são utilizados.

SVM's

- Métodos supervisionados utilizados para previsão e classificação;
- Na aprendizagem de 2 classes, uma SVM (*Support Vector Machine*) tenta encontrar a melhor função de classificação que permita a distinção entre membros de 2 classes num conjunto de treino;
- Para um conjunto de dados linearmente separados, uma função de classificação linear corresponde a um hiperplano que atravessa as 2 classes, dividindo-as;

SVM's

- Razão pela qual uma SVM procura encontrar uma margem máxima num hiperplano:
 - oferecer uma melhor capacidade de generalização;
 - melhor performance de classificação nos dados de treino, assim como bons indicadores para uma correcta classificação de dados futuros;
- Classificadores SVM extremamente precisos, no entanto lentos aquando do processamento de extensos conjuntos de dados.

naïve Bayes

- Algoritmo fácil de construir e de interpretar;
- Simples e robusto;
- Atinge altos níveis de performance;
- Métodos bayesianos associam uma probabilidade a cada previsão, o que representa o nível de confiança do classificador na classificação final.

Distribuição dos dados

- 348

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

Distribuição dos dados

- 180

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

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

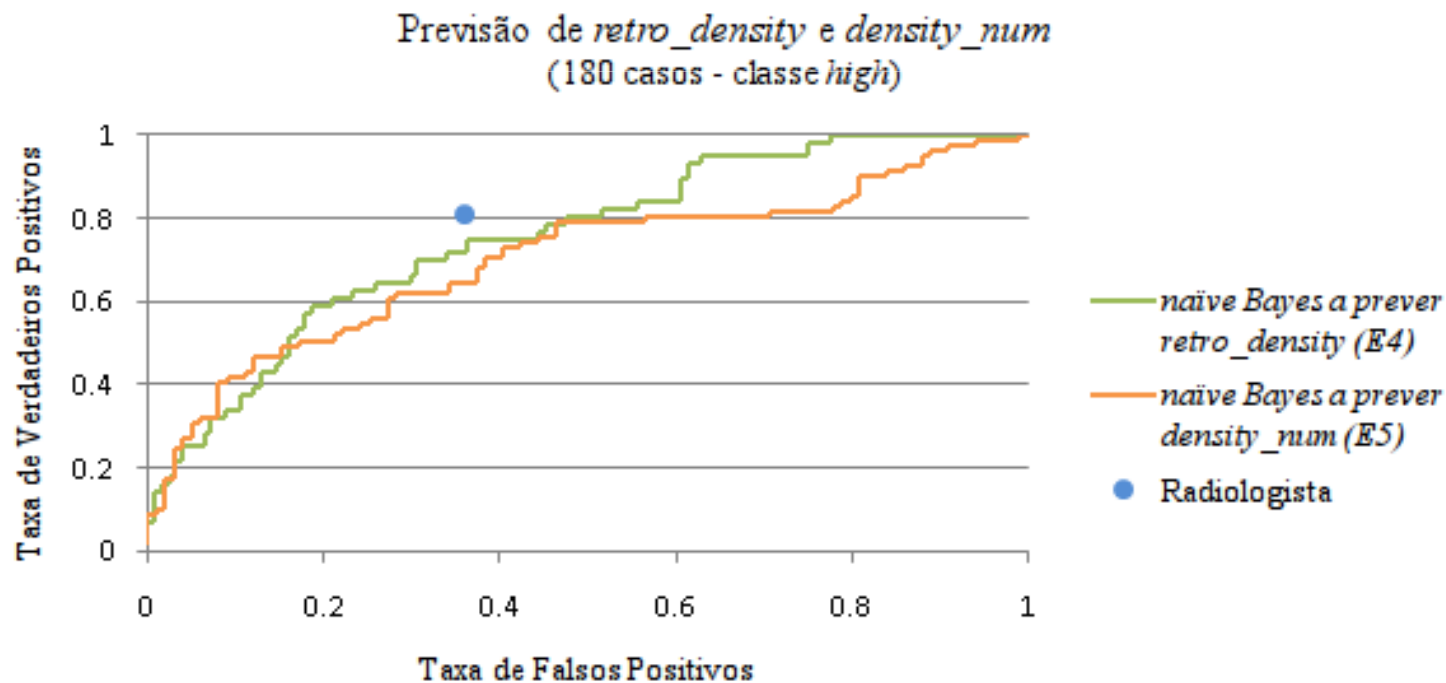
Distribuição dos dados

- 168

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

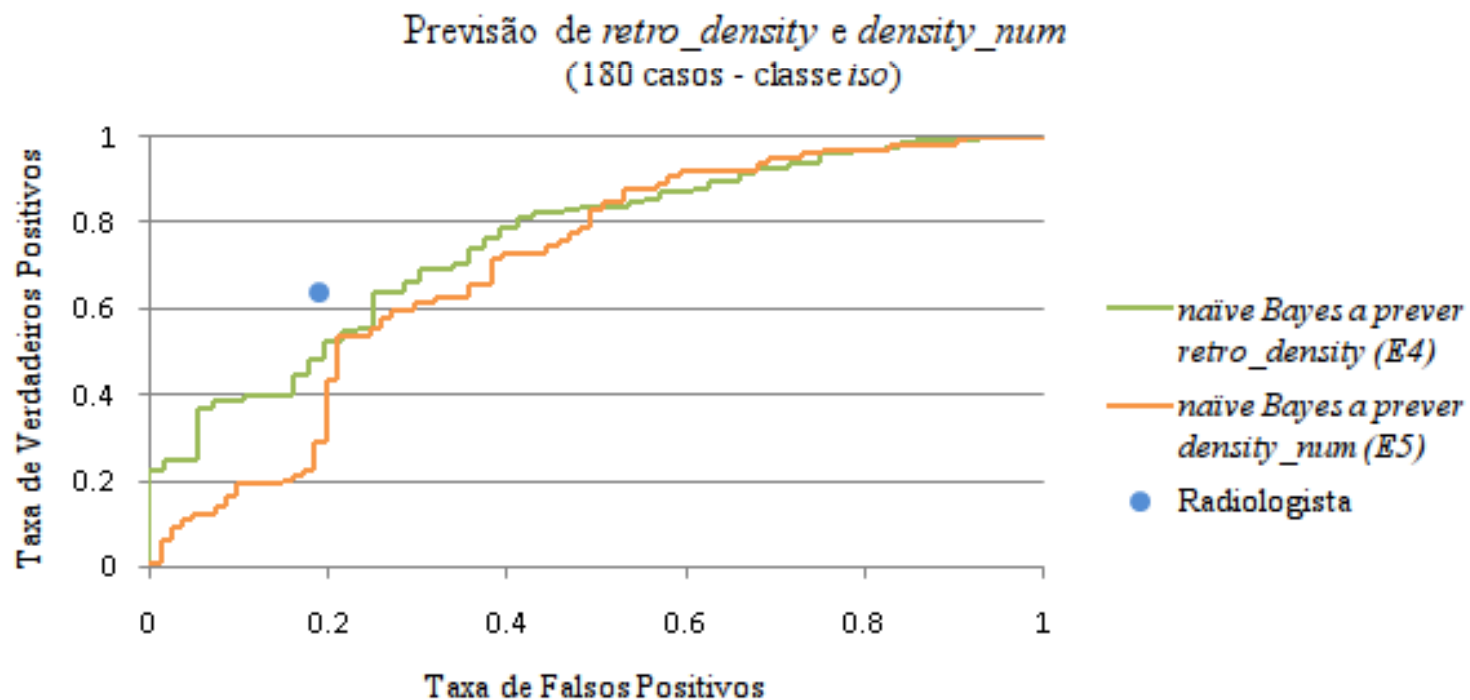
Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

Área ROC

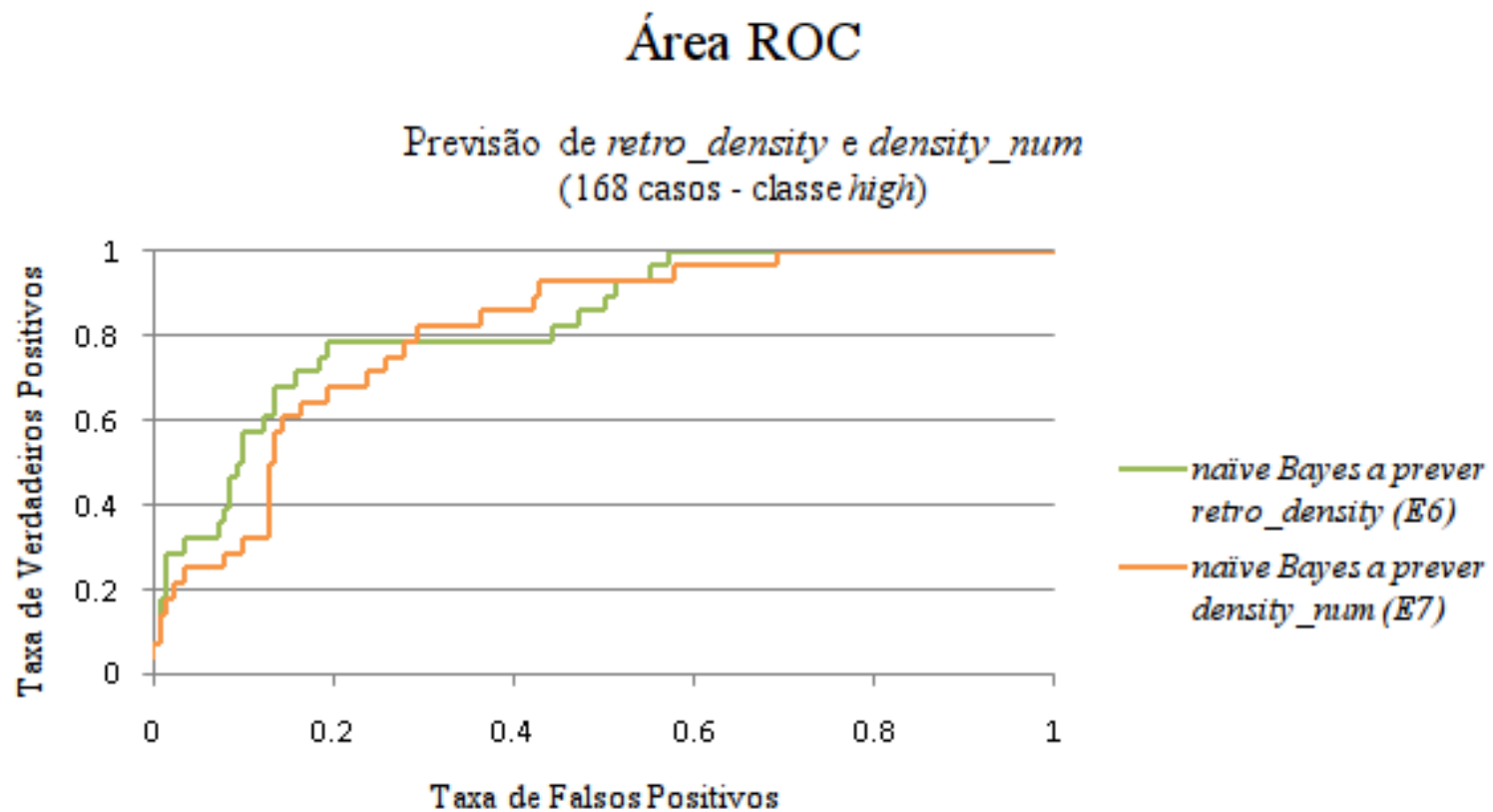


Q_2 - Será possível obter classificadores capazes de preverem densidade de massa com um nível qualitativo semelhante ao de um radiologista?

Área ROC



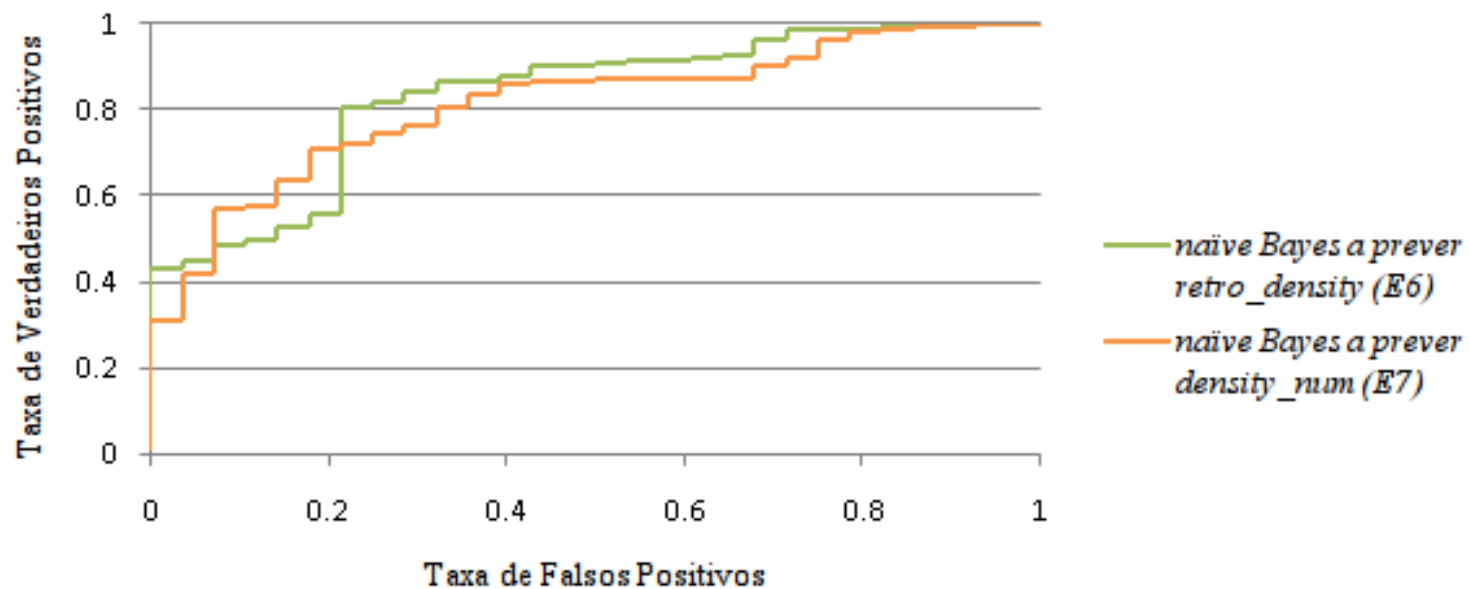
Q₃ . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?



Q₃ . Qual o comportamento dos classificadores gerados num conjunto de dados desconhecidos?

Área ROC

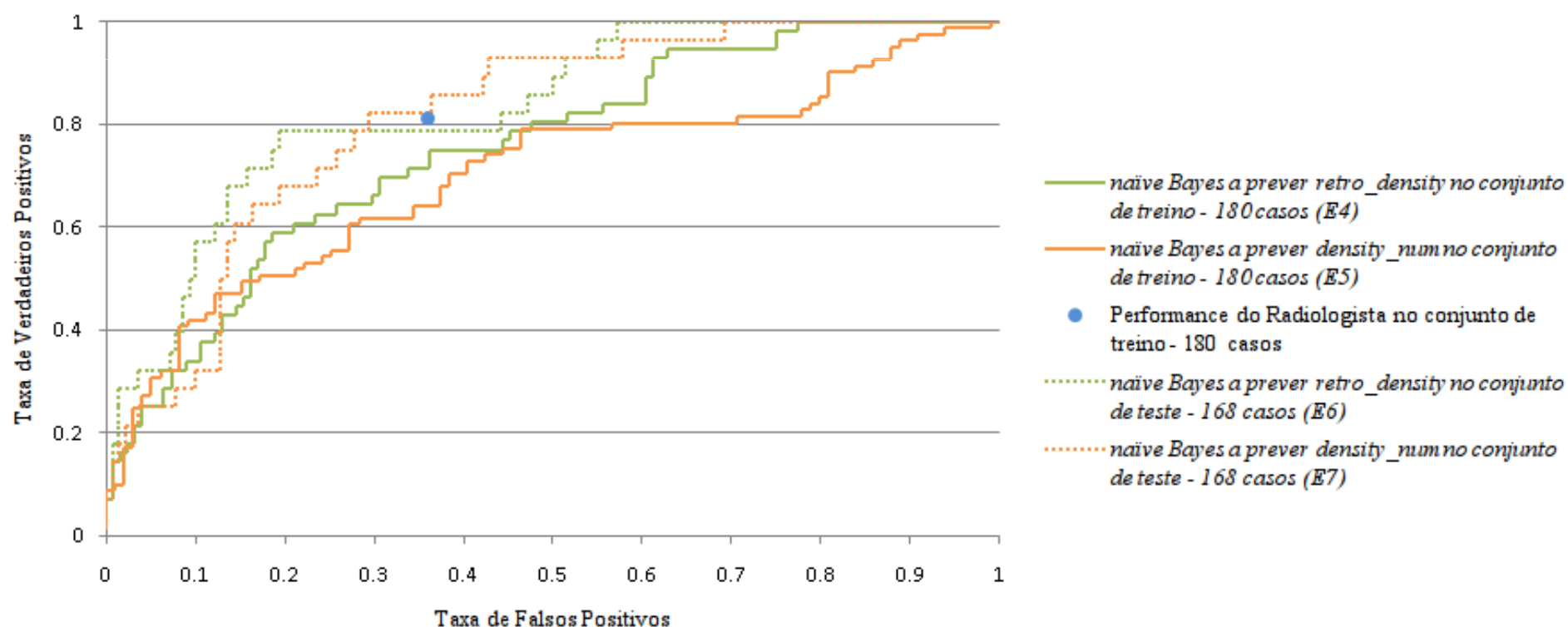
Previsão de *retro_density* e *density_num*
(168 casos - classe iso)



Conclusões e Trabalho Futuro

Área ROC

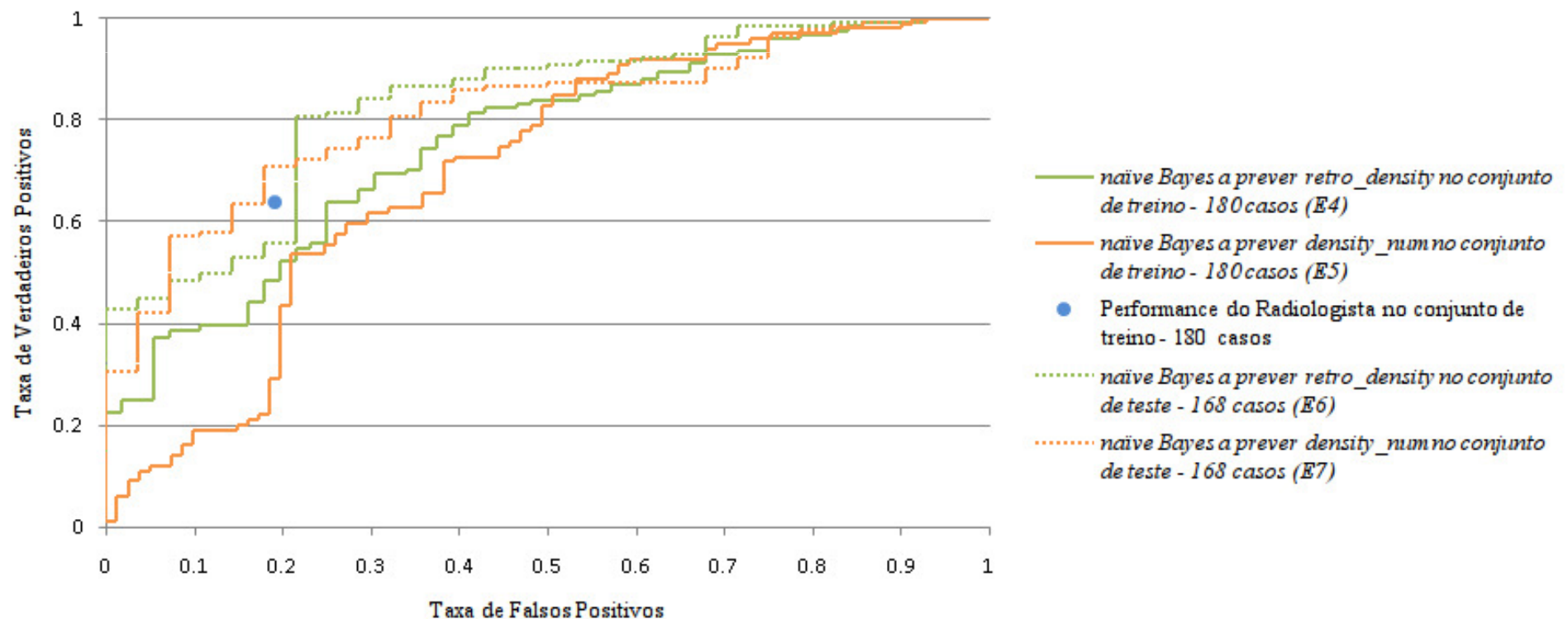
Previsão de *retro_density* e *density_num*
(classe *high*)



Conclusões e Trabalho Futuro

Área ROC

Previsão de *retro_density* e *density_num*
(classe iso)



Bibliografia

- [Aye10] - Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J. W., Kahn, C. E. J., and Burnside, E. S. Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. 2010. Vol. Cancer.
- [HAA02] - An evolutionary artificial neural networks approach for breast cancer diagnosis. Abbass, H. A. 2002, Artificial Intelligence in Medicine.
- [HNa09] - Information extraction for clinical data mining: A mammography case study. H. Nassif, R. Woods, E. Burnside, M. Ayvaci, J. Shavlik, and D. Page. ICDMW'09: Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, Washington, DC, USA : IEEE Computer Society, 2009.

Bibliografia

- [HNa10] - Uncovering age-specific invasive and dcis breast cancer rules using inductive logic programming. H. Nassif, D. Page, M. Ayvaci, J. Shavlik, and E. S. Burnside. Proceedings of 2010 ACM International Health Informatics Symposium (IHI 2010), s.l. : ACM Digital Library, 2010.
- [JDa05] - Knowledge discovery from structured mammography reports using inductive logic programming. J. Davis, E. S. Burnside, I. C. Dutra, D. Page, and V. S. Costa. 2005, American Medical Informatics Association 2005 Annual Symposium.
- [RWo09] - Validation of results from knowledge discovery: Mass density as a predictor of breast cancer. R. Woods, L. Oliphant, K. Shinki, D. Page, J. Shavlik, and E. Burnside. 2009, J Digit Imaging.

Bibliografia

- [WHW90] - Multisurface method of pattern separation for medical diagnosis applied to breast cytology. W. H. Wolberg, and O. L. Mangasarian. 1990, Proceedings of the National Academy of Sciences.
- [WNS95] - An inductive learning approach to prognostic prediction. W. N. Street, O. L. Mangasarian, and W. H. Wolberg. 1995, ICML.
- [Woo10] - The mammographic density of a mass is a significant predictor of breast cancer. Ryan Woods, and Elizabeth Burnside. USA : s.n., 2010, Radiology.
- [YWu93]Wu, Y., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., and Metz, C. E. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. 1993. Vol. Radiology.