

## AVALIAÇÃO DE CLASSIFICADORES NO DIAGNÓSTICO DE CÂNCER DE MAMA

Cleiton Paiva Aquino<sup>1\*</sup>, Esp; Afonso Ueslei da Fonseca, Esp;  
Leandro Luís Galdino de Oliveira, PhD.

Instituto de Informática, Universidade Federal de Goiás.  
Goiânia, Goiás – Brasil.

**Resumo:** O câncer de mama é um tumor que se desenvolve como consequência de alterações genéticas em algum conjunto de células da mama, podendo atingir axilas e até mesmo outros órgãos. O diagnóstico é realizado por meio de informações coletadas através de exames e observações clínicas. Em sistemas informatizados tais informações são usadas como entrada para processos de auxílio ao diagnóstico. Objetivando aumentar a acurácia e precisão desses sistemas implementamos e comparamos dois classificadores de reconhecimento de padrões, nomeados: *Naïve Bayes (NB)* e *Quadratic Discriminant Analysis (QDA)*. Para condução dos experimentos utilizamos uma base de dados de 569 instâncias com 30 atributos e um rótulo identificador entre benignas e malignas, extraídas de imagens digitalizadas. Os resultados experimentais demonstram que, embora ambos classificadores sejam promissores, o classificador *QDA* apresenta, na média, melhores taxas de acurácia e sensibilidade.

**Palavras-chave:** *Câncer de Mama, Algoritmos de Classificação, Diagnóstico por Imagem.*

**Abstract:** *Breast cancer is a tumor that develops as a consequence of genetic alterations in a number of breast cancer cells, can reach the armpits and even other organs. The diagnosis is made based on information collected through tests and clinical observations. In computerized systems such information is used as input to aid the diagnosis process. Aiming to increase the accuracy and precision of these systems implemented and compared two pattern recognition classifiers appointed: Naïve Bayes (NB) and Quadratic Discriminant Analysis (QDA). To conduct the experiments we used a database of 569 instances with 30 attributes and a label identifier between benign and malignant, extracted from scanned images. The experimental results show that although both classifiers are promising the classifier Quadratic Discriminant Analysis has, on average, better accuracy rates and sensitivity.*

**Keywords:** *Breast Cancer, Classification Algorithm, Diagnostic Imaging.*

### Introdução

O câncer de mama é um tumor maligno que se desenvolve como consequência de alterações genéticas em algum conjunto de células da mama, que passam a se dividir descontroladamente, podendo atingir axilas e até mesmo outros órgãos. Ocorre o crescimento anormal das células mamárias, tanto do ducto quanto dos glóbulos mamários. Esse é o tipo de câncer que mais acomete as mulheres em todo o mundo, sendo 1,38 milhão de novos casos e 458 mil mortes pela doença por ano, de acordo com a Organização Mundial de Saúde (OMS). A proporção em homens e mulheres é de 1:100, ou seja, para cada 100 mulheres com câncer de mama, um homem terá a doença. No Brasil, o Instituto Nacional

<sup>1\*</sup> Autor correspondente: Cleiton Paiva Aquino. e-mail: cleitonpaivaquino@inf.ufg.br

de Câncer José Alencar Gomes da Silva estima 57.960 novos casos só em 2016. Segundo dados da Sociedade Brasileira de Mastologia, cerca de uma a cada 12 mulheres terão um tumor nas mamas até os 90 anos de idade<sup>[1],[23],[24]</sup>.

Diante dessa constatação, o uso de técnicas e ferramentas que tornem o tratamento mais eficiente é salutar, sendo imprescindível para tal um diagnóstico ou prognóstico preciso. Nesse sentido, o aprendizado de máquina (AM) e o reconhecimento de padrões (RP) podem contribuir substancialmente uma vez que fornecem soluções empíricas para problemas de classificação. No caso em questão, auxiliando nos diagnósticos ou prognósticos de exames de câncer de mama classificando-os entre benignos ou malignos<sup>[5],[6]</sup>.

Diversos autores têm tratado o problema do diagnóstico de câncer de mama por meio de reconhecimento de padrões. Mangassarian<sup>[5]</sup>, por exemplo, usou técnicas de aprendizado de máquina baseada em programação linear para aumentar a acurácia e objetivamente o prognóstico e diagnóstico do câncer de mama.

Pena-Reyes<sup>[6]</sup> combinou metodologias de sistemas *fuzzy* e algoritmos evolucionários com o objetivo de produzir um sistema automático de diagnóstico. Recentemente, Singh<sup>[7]</sup> abordou a detecção de câncer de mama usando geometria invariante para sensores de imagens de raio-X, construindo um novo arcabouço para sistema de Diagnóstico assistido por computador (CAD). Entretanto, seu sistema não é completamente automático uma vez que precisa de uma interação humana para a seleção das regiões de interesse (ROI) das imagens.

Algoritmos de reconhecimento de padrões são capazes de construir fronteiras de decisão acerca da distribuição de dados de treinamento e possibilitam, com base nessas fronteiras, classificar novos dados de entrada<sup>[2]</sup>.

Nesse trabalho implementamos e apresentamos uma comparação entre os classificadores para reconhecimento de padrões *Naïve Bayes* e *Quadratic Discriminant Analysis* com o objetivo de aumentar a acurácia e precisão dos diagnósticos de câncer de mama.

### **Naïve Bayes**

O classificador *Naïve Bayes*<sup>[3]</sup> (NB) é um dos mais usados em *machine learning*<sup>[4]</sup>. É denominado de ingênuo (*naive*) porque assume que os eventos são independentes. Thomas Bayes desenvolveu o chamado teorema de Bayes em meados do século XVIII, mas ele só foi publicado por Richard Price em dezembro de 1763, dois anos depois da morte de Bayes.

O Naïve Bayes é um classificador probabilístico que define a classificação pela probabilidade posteriori de uma nova amostra pertencer a uma das classes conhecidas<sup>[3]</sup>.

A probabilidade  $P(c_i|\vec{v})$  da classe  $c_i$ , dado um vetor de característica  $\vec{v}$ , é determinada utilizando o teorema de Bayes, definido como:

$$P(c_i|\vec{v}) = \frac{P(c_i|\vec{v})P(c_i)}{P(\vec{v})} \quad (1-1)$$

Considerando que as probabilidades de ambas as classes são iguais, temos que  $P(c_i) = 0,5$  tanto para  $c_0$  quanto para  $c_1$ . Porém, ainda temos que calcular as probabilidades  $P(c_i|\vec{v})$  que não são conhecidas. O método mais utilizado para calcular computacionalmente a probabilidade é supor que as amostras possuem uma função de densidade gaussiana, definida da seguinte forma:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (1-2)$$

Assim, quanto mais próxima a distribuição das amostras tiverem a distribuição normal, melhor o resultado da classificação com o Naïve Bayes.

Esse método, geralmente, possui uma boa eficiência computacional e robustez. porém, depende muito da distribuição das informações extraídas pelos descritores, pois o classificador elabora sua função de decisão a partir da suposição de que os dados possuem distribuição normal, então, quanto mais próxima a distribuição for da normal, melhor o desempenho do classificador.

### ***Quadratic Discriminant Analysis***

O *Quadratic Discriminant Analysis*<sup>[3]</sup> (QDA) é um classificador paramétrico não linear em que não existe pressuposição de que os grupos possuem matrizes de covariância iguais. Assim como na *Linear Discriminant Analysis* (LDA), uma observação é classificada no grupo com a menor distância quadrada. Quando supomos que a normalidade é verdade, o melhor teste possível para determinar se uma hipótese é de uma determinada classe é o teste da taxa de probabilidade.

Suponha que há apenas dois grupos, ( $y \in \{0,1\}$ ), e as médias de cada classe são definidas com  $\mu_{y=0}$ ,  $\mu_{y=1}$  e as covariâncias são definidas como  $\Sigma_{y=0}$ ,  $\Sigma_{y=1}$ . Então a taxa de probabilidade será dada por

$$\text{Taxa de probabilidade} = \frac{\sqrt{|2\pi \Sigma_{y=1}|}^{-1} \exp\left(-\frac{1}{2}(x-\mu_{y=1})^T \Sigma_{y=1}^{-1}(x-\mu_{y=1})\right)}{\sqrt{|2\pi \Sigma_{y=0}|}^{-1} \exp\left(-\frac{1}{2}(x-\mu_{y=0})^T \Sigma_{y=0}^{-1}(x-\mu_{y=0})\right)} < t \quad (2-1)$$

para algum limiar  $t$ . Após alguns rearranjos, isto pode mostrar que o resultado da superfície de separação entre as classes é quadrática. O exemplo estimado do vetor de médias e das matrizes de variância-covariância irá substituir a quantidade de populações nesta fórmula.

Assim, o classificador QDA, ao contrário do LDA, produz fronteiras de decisão mais flexíveis, tendo, entretanto, mais parâmetros para estimar.

O trabalho encontra-se organizado nas seguintes seções: materiais e métodos, análise dos classificadores NB e QDA, comparação de resultados e conclusão.

## **Materiais e métodos**

Utilizamos nesse trabalho para a implementação dos algoritmos o software Octave<sup>[8]</sup> versão 4.0, acrescido e ativado com os seguintes pacotes: control-3.0.0, general-2.0.0, geometry-2.0.0, signal-1.3.2 e statistics-1.2.4.

Uma base de dados foi obtida de Lichman<sup>[17]</sup> e conta com 30 atributos calculados sobre dez características de valor real para cada núcleo da célula, descritas a seguir:

- a. raio (média de distâncias do centro de pontos no perímetro),
- b. textura (desvio padrão dos valores em escala de cinza),
- c. perímetro,
- d. área,
- e. suavidade (variação local em comprimentos de raio),
- f. compacidade ( $\text{perímetro}^2 / \text{área} - 1$ ),
- g. concavidade (gravidade de porções côncavas do contorno) alíneas,
- h. côncavas (número de porções côncavas do contorno),
- i. simetria,
- j. dimensão fractal (aproximação litoral - 1).

Para cada uma das 10 características três valores foram computados, quais sejam: a média, o desvio padrão, e a “pior” ou maior (média dos três maiores valores), resultando assim os 30 atributos. Além desses, cada instância ainda conta com um identificador e um rótulo que diz se a amostra é benigna ou maligna. Todos os valores de características são recodificados com quatro algarismos significativos. A base conta com um total de 569 instâncias, sendo que destas 357 representam amostras benignas e as 212 restantes amostras malignas. Informações adicionais sobre a base de dados podem ser encontradas em Lichman<sup>[17]</sup> e Wolberg<sup>[25]</sup>.

A Figura 1 traz um exemplo de uma das imagens usada na extração dos atributos, enquanto a Figura 2 mostra a dispersão do atributo “média do raio do núcleo da célula” para as 569 amostras separadas entre benignas e malignas.

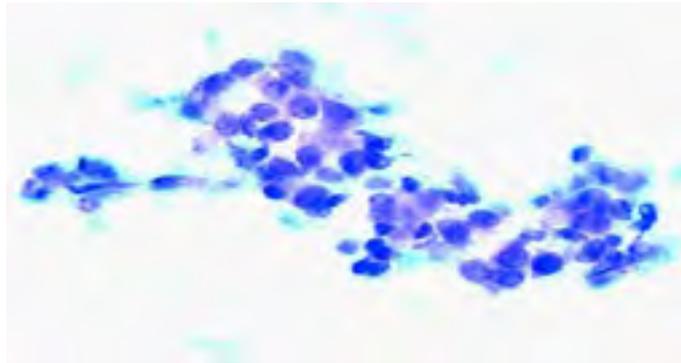


Figura 1. Imagem digitalizada de uma punção aspirativa por agulha fina (FNA) da massa da mama. Os 30 atributos são calculados a partir da imagem e descrevem as características dos núcleos celulares presentes.

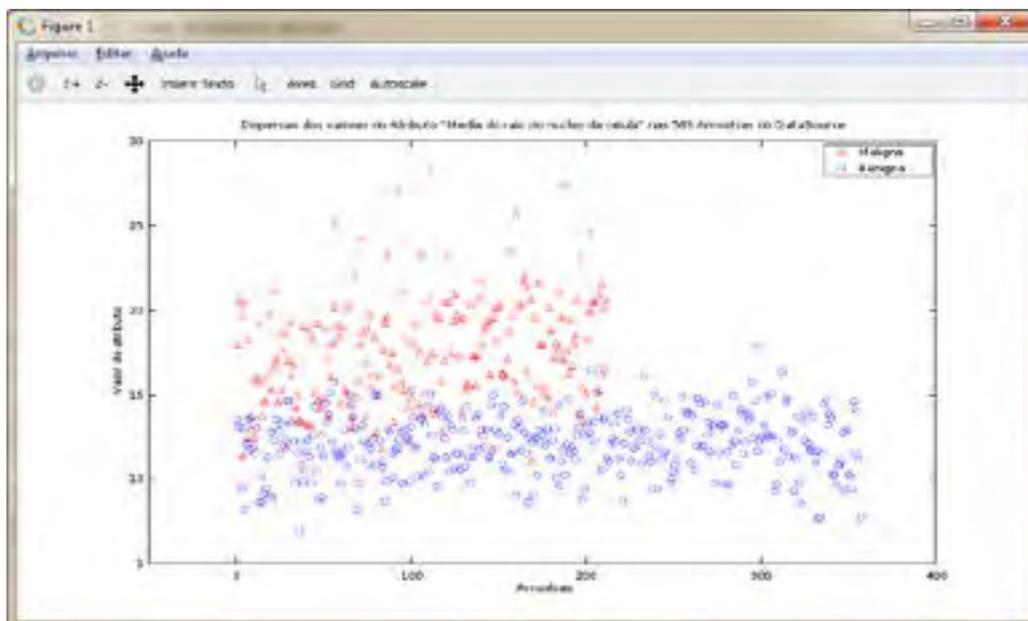


Figura 2. Dispersão dos valores do Atributo “média do raio do núcleo da célula” nas 569 amostras

Para avaliar a capacidade de generalização do modelo lançamos mão da técnica de validação cruzada sobre os classificadores. Essa técnica é amplamente empregada em problemas onde o objetivo da

modelagem é a predição. Busca-se, então, estimar o quão preciso é este modelo na prática, ou seja, o seu desempenho para um novo conjunto de dados.

O conceito central das técnicas de validação cruzada é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos e, posteriormente, utiliza-se alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos (dados de validação ou de teste) são empregados na validação do modelo <sup>[9],[10],[11]</sup>. Entre as formas mais usadas de particionamento estão os métodos *Holdout* e *Kfold*.

*Holdout*<sup>[12]</sup> é um método de validação cruzada que consiste em dividir o conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação). O conjunto de dados pode ser separado em quantidades iguais ou não. Uma proporção muito comum é considerar 2/3 dos dados para treinamento e o 1/3 restante para teste.

*K-fold*<sup>[12],[13]</sup> é um método de validação cruzada que consiste em dividir o conjunto total de dados em  $k$ , subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os  $k-1$  restantes são utilizados para estimação dos parâmetros e calcula-se a acurácia do modelo. Este processo é realizado  $k$  vezes alternando de forma circular o subconjunto de teste.

Sobre cada um dos algoritmos de classificação, um processo de treinamento e teste foi realizado em duas baterias de testes com as validações configuradas da seguinte forma:

#### Primeira bateria:

Validação *Holdout* – As 569 instâncias foram divididas em um conjunto de teste e um de treinamento. Esse processo foi realizado de 5 formas diferentes, sendo que a cada divisão o tamanho dos conjuntos de teste e treinamento foram modificados. A Tabela 1, a seguir, exemplifica a quantidade de cada conjunto em cada uma das divisões.

Tabela 1. Divisão das instâncias para validação cruzada pelo método de *Holdout*

Divisão Treino   Teste	Conjunto de Treinamento		Conjunto de Teste		Total
	Malignas	Benignas	Malignas	Benignas	
50%   50%	106	178	106	179	569
60%   40%	127	213	85	144	569
70%   30%	148	249	64	108	569
80%   20%	169	285	43	74	569
90%   10%	190	321	22	36	569

#### Segunda bateria:

Validação *Kfold* – dividimos nossas instâncias em  $k$  grupos, sendo  $k = 5$ , formando assim quatro grupos de 114 instâncias e um grupo de 113 instâncias. Em cada grupo se manteve a proporção entre amostras benignas e malignas. A cada etapa da validação *Kfold* um grupo era selecionado como conjunto de testes e os  $k-1$  grupos restantes como conjunto de treinamento, mantendo-se assim uma percentagem constante de 80% das instâncias de cada classe para treinamento e 20% para teste.

Para o treinamento e teste no classificador NB a base de dados foi, primeiramente, discretizada utilizando a seguinte estratégia:

- Crie uma matriz vazia para receber atributos discretizados;
- P1- Para cada atributo da base de dados (1 a 30)
- Encontre seu mínimo, seu máximo e seu desvio padrão.
- P2- Para cada amostra da base (1 a 569)
- (P3) Obtenha um atributo discretizado para cada um dos atributos (P1) com valores que vão de mínimo a máximo com intervalos de 2 vezes o desvio padrão.  
Exemplo: Atributo1: mínimo = 1; máximo=23, desvio padrão=2;  
Atributo1 discretizado: [ 1, 5,9, 13, 17, 21 ]
- Adicione ao final da nova Matriz um numero de zeros igual a quantidade de elementos do atributo discretizado.
- Compare o valor do atributo obtido na amostra corrente (P2) com cada um dos valores do atributo discretizado (P3)
- Se o resultado da comparação for verdadeiro para menor ou igual setar na mesma posição da nova Matriz o valor 1 e finalize a comparação.
- Se todas as comparações forem falsas setar 1 na última posição.  
Exemplo Valor do Atributo na amostra corrente = 17  
Matriz discretizada: [ 0, 0, 0, 0, 1, 0 ] ..

A discretização dos atributos se faz necessária, uma vez que os mesmos estão representados por valores contínuos e o classificador NB por padrão não prevê valores contínuos como entrada.

Com isso, nossa base que inicialmente era representada por uma matriz de 569x30 atributos saltou para uma nova matriz na forma discretizada de 569x129 atributos. Sobre essa matriz discretizada foi então realizado o processo de treinamento e teste já descrito anteriormente.

Para o classificador QDA nenhum processo de discretização foi necessário sendo realizado o processo de treinamento e testes, igualmente, realizado no classificador NB diretamente na base de dados.

## Resultados

Para a obtenção dos resultados cada um dos algoritmos de classificação foi submetido às baterias de testes de validação descritas anteriormente e, então, foram avaliados com a medida de acurácia, que é calculada da seguinte forma:

$$ACC = \frac{VP+VN}{Total\_de\_amostras} \quad (3)$$

Onde VP é o número de verdadeiros positivo, VN é o número de verdadeiros negativo e Total\_de\_amostras é o número total de amostras utilizadas no teste, uma medida tradicional na avaliação de algoritmos de classificação.

A Tabela 2 ,a seguir, mostra as taxas de acurácia do classificador NB nas diferentes configurações de cada uma das baterias de teste:

Tabela 2: Acurácia do classificador NB nas baterias de teste.

1ª. Bateria: <i>Holdout</i>	2ª. Bateria – <i>Kfold</i> (k=5)
-----------------------------	----------------------------------

50%   50%	60%   40%	70%   30%	80%   20%	90%   10%	80%   20%	80%   20%	80%   20%	80%   20%	80%   20%
<b>91.228</b>	<b>94.737</b>	<b>95.322</b>	<b>93.860</b>	<b>94.737</b>	<b>93.043</b>	<b>92.920</b>	<b>94.783</b>	<b>95.575</b>	<b>92.035</b>
Acurácia Média:					Acurácia Média:				
<b>93.977</b>					<b>93.671</b>				

A Tabela 3, a seguir, mostra as taxas de acurácia do classificador QDA nas diferentes configurações de cada uma das baterias de teste:

Tabela 3: Acurácia do classificador QDA nas baterias de teste.

1ª. Bateria: Holdout					2ª. Bateria – Kfold (k=5)				
50%   50%	60%   40%	70%   30%	80%   20%	90%   10%	80%   20%	80%   20%	80%   20%	80%   20%	80%   20%
<b>94.386</b>	<b>93.421</b>	<b>97.661</b>	<b>98.246</b>	<b>94.737</b>	<b>93.913</b>	<b>94.690</b>	<b>96.552</b>	<b>95.575</b>	<b>98.230</b>
Acurácia Média:					Acurácia Média:				
<b>95.690</b>					<b>95.786</b>				

## Análise e comparação dos resultados

A sensibilidade é a capacidade de um teste revelar altas proporções de resultados positivos nas pessoas que apresentam a condição investigada. Portanto, quanto menor a proporção de falsos negativos, erro de classificação complementar da sensibilidade, maior a sensibilidade do teste.

A especificidade é a capacidade de um teste em mostrar resultados positivos apenas nas pessoas com a condição investigada e não naquelas que não a possuem. Portanto, quanto menor a proporção de falsos positivos, erro de classificação complementar da especificidade, maior é a especificidade do teste.

A sensibilidade e a especificidade são critérios de validade resultantes da comparação do teste com a verdade. Portanto, para se expressar, objetivamente, os dois critérios, com relação a um teste, é preciso comparar seus resultados com aqueles obtidos por um exame de referência, cujos resultados possam ser considerados como corretos ou reais.

Na prática, esta é uma questão delicada, uma vez que a verdade nem sempre pode ser obtida com absoluta segurança. Por isso, é comum utilizar-se, como exame de referência indicador da verdade, uma bateria exaustiva de testes complementares ou repetidos ou, ainda, o melhor teste conhecido. A aplicação deste exame de referência é, em geral, de alto custo e, portanto, as avaliações de validade de um teste são feitas em pequenas amostras de estudos, especialmente delineados para este fim<sup>(14),(15)</sup>. A Tabela 4 mostra um esquema representativo do exame de referência, também conhecido como matriz de confusão ou tabela da verdade.

Tabela 4. Exame de referência

	+	-	Total
+	VP	FN	VP + FN
-	FN	VN	FN + VN
Total	VP + FN	FN + VN	Total_de_amstras

Onde:

**VP** = verdadeiros positivos

**VN** = verdadeiros negativos

**FN** = falsos negativos

**FP** = falsos positivos

**VP + FN** = total real de positivos

**VP + FN** = total aparente de positivos

#### Validade

Sensibilidade :  $1 - \beta = VP / (VP + FN)$

Especificidade:  $1 - \alpha = VN / (FN + VN)$

#### Erros de Classificação

$\beta = b / (VP + FN)$

$\alpha = FN / (FN + VN)$

A AUC, área sob a curva ROC (do inglês Area Under ROC Curve), é uma medida padrão para comparação de sistemas. Quanto maior a área sob a curva ROC maior a exatidão do sistema.

A curva ROC (Receiver operating characteristic) teve origem na avaliação de qualidade em sinais com ruídos, mas vem sendo muito utilizado na medicina para avaliar a qualidade de um teste clínico. Já outros estudos utilizaram o uso da curva ROC na avaliação de classificadores e demonstraram sua eficiência e qualidade superior, indicando a substituição da acurácia pela AUC <sup>[14],[15]</sup>.

A curva ROC é uma representação gráfica para os valores de sensibilidade (eixo Y) e 1-especificidade (eixo X). Usualmente, o uso da curva ROC na medicina é constituído a partir de um conjunto de pontos que indicam diversos resultados dependentes de alguma variável que influencie no resultado. Com tudo no caso de classificadores binários (que possuem somente duas classes) testados neste trabalho, usamos apenas um ponto, resultante da média entre as iterações de cada bateria.

As análises e comparações mostradas a seguir foram aplicadas sobre os resultados das duas baterias do processo de treinamento e teste em ambos os classificadores e mostra à sensibilidade, a especificidade, a taxa de falsos positivos (erro  $\alpha$ ), a taxa de falso negativo (erro  $\beta$ ) e a AUC obtidas.

**Classificador NB – 1ª e 2ª Baterias (1ª B e 2ª B):**

Tabela 5 - Análise dos resultados do classificador NB

% Dados Treino	Sensibilidade	Especificidade	$\beta$	$\alpha$	AUC
1ª B. 50%	0.896	0.955	0.104	0.045	0.97
1ª B. 60%	0.894	<b>0.958</b>	0.106	<b>0.042</b>	0.97
1ª B. 70%	0.860	0.981	0.140	0.019	0.98
1ª B. 80%	0.907	<b>0.958</b>	0.093	<b>0.042</b>	0.98
1ª B. 90%	0.905	1.000	0.095	0.000	0.99
2ª B. 80%	0.860	0.972	0.139	0.028	0.96
2ª B. 80%	0.833	0.986	0.167	0.014	0.99
2ª B. 80%	0.907	0.972	0.093	0.028	0.98
2ª B. 80%	0.905	0.986	0.095	0.014	0.98
2ª B. 80%	0.929	0.915	0.071	0.084	0.99

A Figura 3 mostra as curvas ROC médias para o classificador NB em cada uma das baterias de teste.

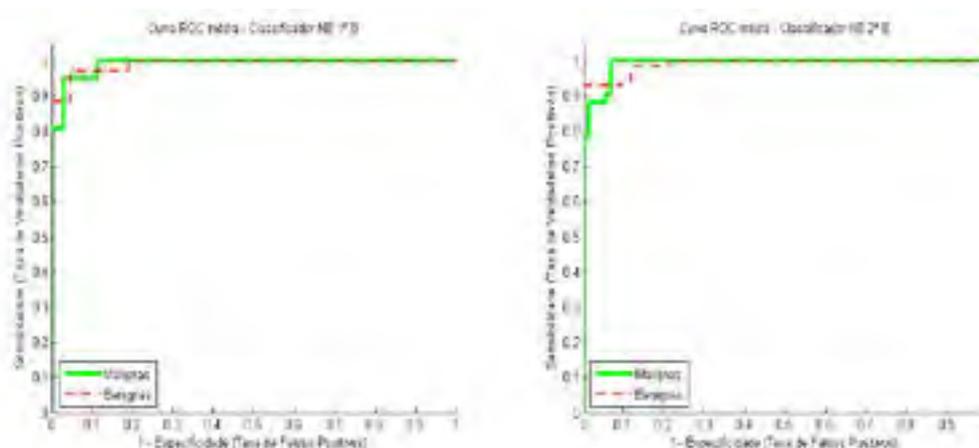


Figura 3. Curvas ROC médias do classificador NB para 1ª B e 2ª B.

**Classificador QDA – 1ª e 2ª Baterias (1ª B e 2ª B):**

Tabela 6 - Análise dos resultados do classificador QDA

% Dados Treino	Sensibilidade	Especificidade	$\beta$	$\alpha$	AUC
1ª B. 50%	0.943	0.955	0.057	0.045	0.98
1ª B. 60%	0.929	<b>0.958</b>	0.070	<b>0.042</b>	0.99
1ª B. 70%	0.937	0.981	0.062	0.019	0.99
1ª B. 80%	0.907	<b>0.958</b>	0.093	<b>0.042</b>	0.99
1ª B. 90%	0.857	1.000	0.143	0.000	0.99
2ª B. 80%	0.953	0.972	0.046	0.028	0.99
2ª B. 80%	0.905	0.986	0.095	0.014	1.00
2ª B. 80%	0.953	0.972	0.046	0.028	1.00
2ª B. 80%	0.929	0.986	0.071	0.014	0.98
2ª B. 80%	0.976	0.915	0.024	0.084	0.99

A Figura 4 mostra as curvas ROC médias para o classificador QDA em cada uma das baterias de teste.

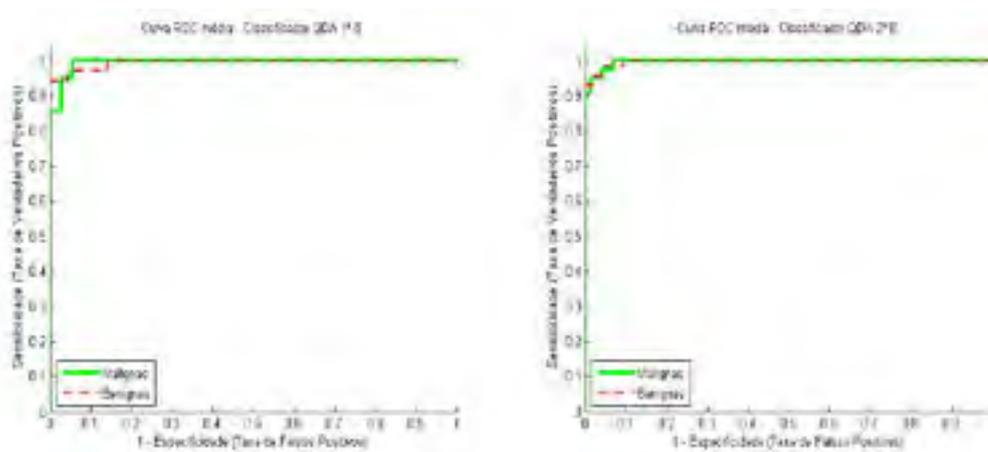


Figura 4. Curvas ROC médias do classificador NB para 1ª B e 2ª B.

Conforme pode ser observado nas Tabelas 5 e 6 fica evidente, em ambos os classificadores, que o aumento do tamanho do bloco de treinamento (1ª B) afeta tanto especificidade quanto o erro de classificação  $\alpha$ , que representa justamente o erro de classificar uma amostra negativa como positiva, porém, o mesmo não pode se dizer do erro  $\beta$ . Apesar de esta relação ser na maior parte dos testes diretamente relacionada ao aumento do tamanho da partição, isso não é uma regra, como pode ser observado

nos valores destacados em ambas as tabelas nas linhas 1ª B. 60% e 1ª B. 80% em que os valores de especificidade e erro de classificação  $\alpha$  são iguais. Isso pode ser explicado em razão da larga faixa de valores presente em cada um dos 30 atributos de nossas instâncias e da seleção das partições que ocorre de forma randômica.

Frente à análise dos resultados dos dois classificadores, observa-se que o classificador QDA tem melhores taxas de acurácia e sensibilidade (médias) e uma pequena vantagem nos valores de AUC, o que não se mantém quanto às taxas de especificidade e falsos positivos (erro  $\alpha$ ) onde suas taxas médias se equiparam às do classificador NB. As curvas ROC, por sua vez, evidenciam a qualidade dos testes e a eficiência dos classificadores.

## Conclusão

É possível inferir que o diagnóstico de câncer de mama auxiliado por sistemas e técnicas computacionais é totalmente possível, apesar de nada trivial, uma vez que os classificadores tiveram uma taxa média de erros de 7,1% no NB e de 5,1% no QDA (médias) sobre o conjunto de dados.

Pode-se ainda notar que o classificador NB, apesar da sua natureza “ingênua”, possui grande poder de classificação e, talvez, não tenha se saído tão bem quanto o QDA pelo método de discretização adotado, ou por não lidar tão bem com a presença de *outliers*<sup>2</sup>. Quanto ao classificador QDA este mostra grande desempenho na classificação mas assim como o NB peca no que se refere à taxa de falsos negativos, algo que também pode ser explicado pela presença de *outliers*.

Conclui-se por fim, que os dois classificadores são promissores na tarefa de classificar as amostras entre benignas e malignas, o que é comprovado pelos valores de AUC de ambos, ressaltando-se, ainda o fato de que nenhuma análise de *outliers* ter sido realizada sobre essas amostras.

## Referências

- [1] World Health Organization. Cancer control: knowledge into action. WHO guide for effective programmes. Vol. 2. World Health Organization, 2007.
- [2] Webb AR. Statistical pattern recognition. John Wiley & Sons. 2003.
- [3] Jain AK, Duin RPW, Mao J. Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine intelligence. 2000; 22(1), 4-37.
- [4] Andrieu C, De Freitas N, Doucet A, Jordan MI. An introduction to MCMC for machine learning. Machine learning. 2003; 50(1-2), 5-43.
- [5] Mangasarian, Olvi L, W. Nick Street, and William H. Wolberg. “Breast cancer diagnosis and prognosis via linear programming.” Operations Research. 1995; 43.4: 570-577.
- [6] Pena-Reyes CA, Sipper M. Designing breast cancer diagnostic systems via a hybrid fuzzy-genetic methodology. In Proceedings of the 1999 IEEE International Fuzzy Systems Conference. 1999 (Vol. 1, pp. 135-139).
- [7] Singh SP, Urooj S, Lay-Ekuakille A. Breast Cancer Detection Using PCPCET and ADEWNN: A Geometric Invariant Approach to Medical X-Ray Image Sensors. IEEE Sensors Journal. 2016; 16(12), 4847-4855.
- [8] GNU-OCTAVE, O. URL <http://www.gnu.org/software/octave/>. Acesso em 28 jun. 2016.
- [9] Geisser S. Predictive Inference. New York, NY: Chapman and Hall. 1993; ISBN 0-412-03471-9.

2 Em estatística, outlier, valor aberrante ou valor atípico, é uma observação que apresenta um grande afastamento das demais da série<sup>(122)</sup> (que está “fora” dela), ou que é inconsistente. A existência de outliers implica, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados às amostras.

- [10] Kohavi R. “A study of cross-validation and bootstrap for accuracy estimation and model selection”. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (San Mateo, CA: Morgan Kaufmann). 1995; 2 (12): 1137–1143. CiteSeerX: 10.1.1.48.529.
- [11] Devijver PA, Kittler J. Pattern Recognition: A Statistical Approach. London, GB: Prentice-Hall. 1982.
- [12] Refaeilzadeh P, Tang L, Liu H. Cross-validation. In Encyclopedia of database systems. Springer US. 2009; pp. 532-538.
- [13] McLachlan GJ, Do K-A; Ambrose C. Analyzing microarray gene expression data. Wiley. 2004
- [14] METZ C. Roc methodology in radiologic imaging. Invest Radiol. 1986; 21(9):720–733.
- [15] Souza C. “Análise de poder discriminativo através de curvas roc.” Disponível em: <http://crsouza.blogspot.com/2009/07/analise-de-poder-discriminativo-atraves.html>. Acesso em 05 set. 2016.
- [16] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International joint Conference on artificial intelligence. 1995; [S.l.: s.n.], v. 14, p. 1137–1145.
- [17] Lichman M. UCI Machine Learning Repository - Breast Cancer Wisconsin [<https://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, School of Information and Computer Science. 2013
- [18] Asuncion A, Newman D. UCI machine learning repository. 2007.
- [19] Geisser S. Predictive Inference. New York, NY: Chapman and Hall. 1993
- [20] Kohavi R. “A study of cross-validation and bootstrap for accuracy estimation and model selection”. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (San Mateo, CA: Morgan Kaufmann). 1995; 2(12): 1137–1143.
- [21] Vassallo P. “Thinking, Fast and Slow.” ETC.: A Review of General Semantics. 2012; 69.4:480-481.
- [22] Gladwell M. Outliers: The story of success. Hachette UK. 2008
- [23] Barros AC, et al. Mastologia: Conduas. Rio de Janeiro: Revinter. 1999; 62-5.
- [24] Instituto Nacional do Câncer (INCA). “Estimativa de incidência de câncer no Brasil para 2014”. 2014.
- [25] Wolberg WH, Street WN, Mangasarian OL. Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Letters. 1994; 77:163-171.

## Contato

Cleiton Paiva Aquino  
Instituto de Informática – UFG. Alameda  
Palmeiras, Quadra D,  
Câmpus Samambaia, CEP 74690-900 -  
Goiânia – GO  
E-mail: [cleitonpaivaaquino@inf.ufg.br](mailto:cleitonpaivaaquino@inf.ufg.br)  
[engcleitonpaiva@gmail.com](mailto:engcleitonpaiva@gmail.com)  
Fone: (62) 3269-4146 / Cel: (62) 98417-9261

