

Avaliação de um teste em medicina usando uma curva *ROC fuzzy*

Maria José P. Castanho¹,
DEMAT, UNICENTRO, 85.010-990, Guarapuava, PR.

Akebo Yamakami²,
DT-FEEC, UNICAMP, 13.083-970, Campinas, SP.

Laécio C. de Barros³, Laércio L. Vendite⁴,
DMA-IMECC, UNICAMP, 13.083-970, Campinas, SP.

Resumo. Neste artigo desenvolvemos a metodologia ROC (*Receiver Operating Characteristic*) Fuzzy, que combina a teoria dos conjuntos fuzzy e a metodologia ROC convencional. Aplicamos esta ferramenta para avaliar o poder de discriminação de um sistema baseado em regras fuzzy, construído para prever o estado patológico do câncer de próstata. Na curva ROC fuzzy utilizamos graus de pertinência para descrever o padrão-ouro.

Palavras-chave: *Curva ROC, Conjuntos Fuzzy, Câncer de Próstata*

1 Introdução

Quando um teste para diagnóstico é desenvolvido em medicina, é necessário avaliar sua habilidade em classificar corretamente os indivíduos em dois subgrupos clinicamente relevantes.

Uma ferramenta amplamente difundida para avaliar o desempenho de testes de diagnóstico é a análise da curva ROC (*Receiver Operating Characteristic*). Tal ferramenta originou-se na Teoria de Detecção de Sinais, no início dos anos 50, e na década de 60 foi aplicada pela primeira vez em medicina (Metz, 2003). Desde então foram sugeridas várias alterações. Metz (1978) apresenta uma técnica para gerar dados de resposta, que

¹doutoranda da FEEC, zeza@dt.fee.unicamp.br

²akebo@dt.fee.unicamp.br

³laeciocb@ime.unicamp.br

⁴vendite@ime.unicamp.br

são usados para traçar a curva ROC, em testes de diagnóstico que devem ser interpretados subjetivamente. Chakraborty (1993) aplicou o método FROC, uma extensão do método ROC, para avaliar detecção de nódulos em radiografias da mama onde havia a necessidade de especificar a localização do nódulo detectado. De Leo e Campbell (1990) introduziram a metodologia ROC fuzzy para avaliar a hipertrofia ventricular esquerda através de um eletrocardiograma.

O objetivo deste trabalho é avaliar o poder de discriminação de um sistema baseado em regras fuzzy construído para prever o estado patológico do câncer de próstata. Para isto será utilizado uma curva ROC (*Receiver Operating Characteristic*) Fuzzy, que combina a teoria dos conjuntos fuzzy e a metodologia ROC convencional.

2 Curva ROC

Quando consideramos o resultado de um teste de diagnóstico em duas populações, uma população com a doença e outra sem, raramente observamos uma perfeita separação entre estes grupos. Em geral, há uma sobreposição entre as duas curvas que representam cada um destes grupos, conforme Figura 1.

Portanto, qualquer que seja o “valor de corte” (ponto que separa as duas populações) escolhido, alguns indivíduos com a doença serão classificados corretamente como positivos (VP - total de indivíduos Verdadeiros Positivos) e alguns serão classificados como negativos (FN - total de indivíduos Falsos Negativos). Por outro lado, alguns casos sem a doença serão classificados como positivos (FP - total de indivíduos Falsos Positivos) mas alguns serão corretamente classificados como negativos (VN - total de indivíduos Verdadeiros Negativos).

As medidas de *sensibilidade* e *especificidade* fornecem a probabilidade de que o teste classifique corretamente um indivíduo doente e um indivíduo saudável, respectivamente. Tais medidas são dadas por

$$\text{Sensibilidade} = \frac{VP}{VP + FN}, \quad (2.1)$$

$$\text{Especificidade} = \frac{VN}{VN + FP}. \quad (2.2)$$

Supondo que os resultados de testes médicos sejam traduzidos por um valor em um determinado intervalo contínuo, o ponto de corte que separa os diagnósticos verdadeiros e falsos deve pertencer a este intervalo. Para avaliar a eficiência do teste, o procedimento é estudar o efeito de diversos pontos de corte em termos de sensibilidade e especificidade. Para isto constrói-se a curva ROC que é a representação dos pares (1-Especificidade, Sensibilidade) obtidos ao considerar todos os possíveis valores de corte do teste. Ver Figura 2.

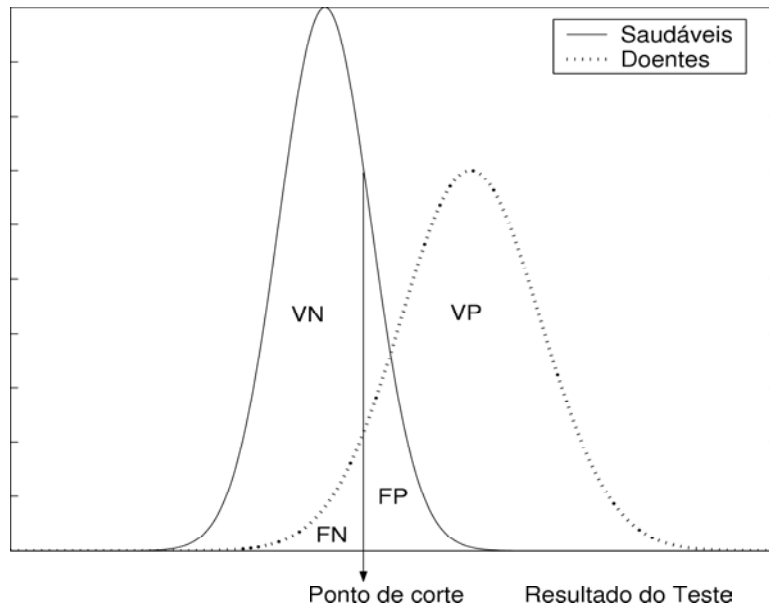


Figura 1: Distribuição dos resultados de um teste em indivíduos doentes e saudáveis. Supõe-se que as distribuições teóricas são normais com médias e desvios padrões diferentes.

A área sob a curva é, então, empregada como uma medida do desempenho do teste como discriminador de pacientes doentes e saudáveis. Um teste ideal é aquele cuja área sob a curva ROC é igual a 1. Quando a curva ROC é a bissetriz, ou seja, área igual a 0.5, o teste não permite distinguir entre os grupos.

3 Padrão-ouro

A análise da curva ROC é baseada na idéia de que é possível saber, com certeza, se um paciente está doente ou não. Esta certeza é obtida por algum procedimento, conhecido como padrão-ouro que pode ser: uma biópsia, cirurgia, análise *post-mortem*, um teste mais preciso ou mesmo um diagnóstico feito por um ou mais clínicos experientes na área.

Algumas vezes, o padrão-ouro não permite uma classificação dicotômica: positivo ou negativo. Se ele é dado por um teste que não pode ser interpretado objetivamente ou se é feito por um grupo de especialistas existe uma subjetividade que deve ser considerada sob risco de produzir um modelo distorcido forçando a classificação dicotômica.

A teoria dos conjuntos fuzzy fornece ferramentas que permitem preservar a subjetividade ao invés de impor dicotomia.

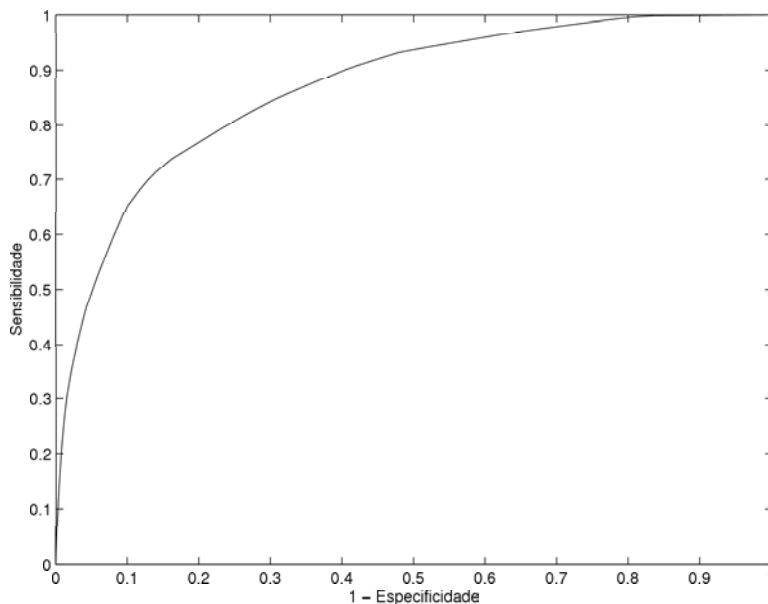


Figura 2: Exemplo de Curva ROC

4 Conjuntos Fuzzy

A idéia central da teoria dos conjuntos fuzzy é a pertinência gradual de elementos a um conjunto. No caso clássico há somente duas possibilidades (dois graus de pertinência), ou o elemento pertence ao conjunto (grau de pertinência igual a 1) ou não (grau de pertinência igual a 0).

Num conjunto fuzzy um elemento pode pertencer parcialmente a ele, com um grau de pertinência no intervalo $[0, 1]$.

Seja \mathbf{X} o universo de discurso. Um subconjunto fuzzy A de \mathbf{X} é um conjunto de pares ordenados definidos por

$$A = \{(x, \mu_A(x)) | x \in X \text{ e } \mu_A \rightarrow [0, 1]\},$$

onde $\mu_A(x)$ denota o grau de pertinência de x em A .

Esta definição permite “relaxar” a restrição imposta na escolha dicotômica preservando a característica incerta do conhecimento que está sendo modelado e evitando perda de informações.

5 Metodologia Fuzzy para Curva ROC

Como já mencionamos, para avaliar um teste de diagnóstico, é necessário definir um *padrão-ouro* que dará o verdadeiro diagnóstico o qual será utilizado como referência para comparar com o resultado do teste. Com este procedimento podemos classificar o indivíduo submetido ao teste numa das seguintes categorias: verdadeiro positivo, falso negativo, verdadeiro negativo ou falso positivo.

Quando o padrão-ouro não permite uma classificação dicotômica, podemos expressá-lo através de uma função de pertinência ao conjunto fuzzy *Doentes*. Esta função pode ser construída baseando-se em dados estatísticos e/ou conhecimento de especialistas, variando o tipo e a complexidade com a aplicação. Porém, deve manter um caráter monotônico indicando que, quanto mais doente um indivíduo, maior seu grau de pertinência a este conjunto.

Seja \mathbf{X} a população submetida ao teste. O subconjunto fuzzy D de \mathbf{X} é o conjunto de pares ordenados definidos por:

$$D = \{(x, \mu_D(x)) | x \in X \text{ e } \mu_D \rightarrow [0, 1]\},$$

onde $\mu_D(x)$ denota o grau de pertinência do indivíduo com diagnóstico x ao conjunto dos *Doentes*.

O complementar, \bar{D} , do conjunto fuzzy D com respeito ao universo de discurso \mathbf{X} , é definido por

$$\bar{D} = \{(x, \mu_{\bar{D}}(x)) | \mu_{\bar{D}}(x) = 1 - \mu_D(x)\}, \forall x \in X.$$

Será considerado aqui que o complementar do conjunto dos *Doentes* é o conjunto fuzzy dos *Saudáveis*. Assim, $\mu_{\bar{D}}(x)$ denota o grau de pertinência do indivíduo com diagnóstico x ao conjunto dos *Saudáveis*.

O resultado do teste também pode ser *Positivo* ou *Negativo* com um determinado grau de certeza, ou seja, pode ser dado através de uma função de pertinência ao conjunto fuzzy *Positivo*, definido por

$$P = \{(x, \mu_P(x)) | x \in X \text{ e } \mu_P \rightarrow [0, 1]\},$$

onde $\mu_P(x)$ denota o grau de pertinência do indivíduo com diagnóstico x ao conjunto *Positivo*.

Seu complementar, \bar{P} , é definido por

$$\bar{P} = \{(x, \mu_{\bar{P}}(x)) | \mu_{\bar{P}}(x) = 1 - \mu_P(x)\}, \forall x \in X,$$

onde $\mu_{\bar{P}}(x)$ denota o grau de pertinência do indivíduo x ao conjunto cujo resultado do teste é *Negativo*.

As funções de pertinência para o padrão-ouro e resultado do teste devem ser reduzidas às funções características (contradomínio em $\{0, 1\}$) quando estes são binários pois os casos fuzzy são extensões dos casos clássicos (ou, alternativamente, o caso clássico é um caso especial do caso fuzzy).

No caso clássico, onde $\mu_P(x)$ toma o valor 0 se o teste é *Negativo* ou 1 se o teste é *Positivo* e $\mu_D(x)$ toma o valor 0 se o verdadeiro diagnóstico é *Saudável* ou 1 se o verdadeiro diagnóstico é *Doente*, as quatro categorias podem ser especificadas através de uma tabela verdade baseada na lógica de proposições, como mostrado na Tabela 1.

Tabela 1: Tabela Verdade para o Caso Clássico

| Teste $\mu_P(x)$ | Padrão-Ouro $\mu_D(x)$ | VP | FP | VN | FN |
|------------------|------------------------|----|----|----|----|
| 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 |

Quando consideramos o resultado do teste fuzzy e/ou o padrão-ouro fuzzy encontramos a pertinência do indivíduo em cada uma das categorias mencionadas anteriormente: verdadeiro positivo, falso negativo, verdadeiro negativo ou falso positivo. Parasuraman et al. (2000) definem esta pertinência através dos operadores fuzzy máximo e mínimo.

Para que o indivíduo seja classificado corretamente pelo teste ele deve pertencer ao conjunto D e, também, ao conjunto P ou ao conjunto \bar{D} e ao conjunto \bar{P} . Matematicamente temos

$$VP = D \cap P$$

$$VN = \bar{D} \cap \bar{P}$$

onde VP é o conjunto *Verdadeiro Positivo* e VN o conjunto *Verdadeiro Negativo*. Como estamos supondo que D e P são subconjuntos fuzzy de \mathbf{X} , então VP e VN são, também, subconjuntos fuzzy de \mathbf{X} cujas funções de pertinência são, respectivamente, μ_{VP} e μ_{VN} . Temos

$$\mu_{VP}(x) = \mu_{(D \cap P)}(x) = \min[\mu_D(x), \mu_P(x)], x \in X$$

$$\mu_{VN}(x) = \mu_{(\bar{D} \cap \bar{P})}(x) = \min[\mu_{\bar{D}}(x), \mu_{\bar{P}}(x)] = \min[1 - \mu_D(x), 1 - \mu_P(x)].$$

Se $D = P$ o teste classifica corretamente todos os indivíduos submetidos a ele e

$$\mu_{VP}(x) = \mu_P(x) = \mu_D(x),$$

$$\mu_{VN}(x) = 1 - \mu_D(x) = 1 - \mu_P(x) \text{ e}$$

$$\mu_{FP}(x) = \mu_{FN}(x) = 0, \forall x \in X.$$

Quando $D \neq P$, as categorias de erro FP e FN devem refletir o grau de superestimação ou subestimação do resultado do teste com relação ao padrão ouro. Se o padrão-ouro é mais próximo à certeza de positivo do que o teste então a pertinência ao conjunto FN deve ser diferente de zero mas não deve ter qualquer pertinência ao conjunto FP , ou seja,

$$\mu_D(x) > \mu_P(x) \Rightarrow \mu_{FN}(x) \neq 0 \text{ e } \mu_{FP}(x) = 0.$$

Por outro lado, se o padrão-ouro estiver mais próximo à certeza de negativo do que o teste, temos

$$\mu_D(x) < \mu_P(x) \Rightarrow \mu_{FP}(x) \neq 0 \text{ e } \mu_{FN}(x) = 0.$$

Então, Parasuraman et al. (2000) definem:

$$\mu_{FN}(x) = \max[\mu_D(x) - \mu_P(x), 0]$$

$$\mu_{FP}(x) = \max[\mu_P(x) - \mu_D(x), 0].$$

Como a união das quatro categorias representam o universo de resultados possíveis temos $\mu_{VP}(x) + \mu_{FP}(x) + \mu_{VN}(x) + \mu_{FN}(x) = 1$.

Considerando que as populações *Doentes* e *Saudáveis* são distribuídas normalmente, podemos construir a curva ROC.

Quando o padrão-ouro e/ou o resultado do teste é dado por uma função de pertinência a um conjunto fuzzy, denominaremos esta curva como curva ROC fuzzy.

As medidas de sensibilidade e especificidade em cada ponto de corte, x , necessárias para a construção da curva ROC fuzzy, são calculadas da mesma forma que no caso clássico, ou seja,

$$\text{Sensibilidade}(x) = \frac{\sum \mu_{VP}(x_i)}{\sum \mu_{VP}(x_i) + \sum \mu_{FN}(x_i)}, \quad (5.3)$$

$$\text{Especificidade}(x) = \frac{\sum \mu_{VN}(x_i)}{\sum \mu_{VN}(x_i) + \sum \mu_{FP}(x_i)}, \quad (5.4)$$

$\forall i, i = 1, 2, \dots, n$ onde x_i é o i -ésimo indivíduo submetido ao teste e n é o número total de indivíduos da população.

Devemos observar que no caso clássico $\mu_{VP}(x_i)$, $\mu_{VN}(x_i)$, $\mu_{FP}(x_i)$ e $\mu_{FN}(x_i)$ assumem os valores 0 ou 1. Assim, (5.3) e (5.4) reduzem-se a (2.1) e (2.2), respectivamente.

A área sob a curva ROC fuzzy pode ser usada para avaliar o desempenho do teste como discriminador de pacientes doentes e saudáveis. Porém, quando os graus de pertinência entre 0 e 1 estão presentes no conjunto de dados, a área máxima está restrita à área sob a curva ROC “perfeita”, A_p , (onde o teste classifica corretamente todos os indivíduos) e

$$A_p \leq 1$$

sendo igual a 1 no caso clássico. Leo e Campbell (1995) propõe converter a área, A , sob a curva ROC fuzzy numa área normalizada, A_n , dada por

$$A_n = 0.5 + \frac{A - 0.5}{2(A_p - 0.5)}.$$

Notemos que, para dados fuzzy, quando $A = A_p$ então $A_n = 1$ e para dados binários, $A_p = 1$ e, portanto, $A_n = A$.

A curva ROC também pode ser utilizada para comparar o desempenho de dois ou mais testes utilizando o mesmo padrão-ouro. O teste será melhor quanto maior for a área sob a curva (Galparsoro e Fernández, 1998).

6 Teste a ser avaliado

Nosso objetivo é avaliar o poder de discriminação de um sistema baseado em regras fuzzy (SBRF) (Castanho et al., 2003) construído para prever o estágio do câncer de próstata considerando três variáveis:

- níveis de Antígeno Prostático Específico (PSA) no sangue;
- grau de diferenciação do tumor dado pelo score de Gleason (biópsia);
- estado clínico: avaliação clínica através do toque retal complementado, ou não, por outros exames.

A saída do SBRF (leitura que indica o resultado do teste) é um número no intervalo $[0, 1]$ que representa o paciente com determinado estágio do câncer: confinado à glândula prostática ou não. Escolhendo um ponto de corte, $x \in [0, 1]$, os pacientes representados por valores maiores do que x têm resultado do teste *Positivo*, ou seja, há envolvimento do câncer além da próstata, e para valores menores do que x têm resultado do teste *Negativo*,

quer dizer, os indivíduos têm câncer confinado na próstata. Consideramos, desta forma, o resultado do teste de maneira clássica: $\mu_P(x)$ igual a 0 ou 1.

Determinar o estágio do câncer é importante para auxiliar o médico a tomar a decisão sobre o tipo de terapia a ser administrada. A prostatectomia radical, por exemplo, apresenta alto índice de cura desde que o câncer esteja confinado à próstata.

Para fazer a avaliação do SBRF foi utilizada a base dados SEER 1973-2000 (SEER, 2003). Tendo em vista que alguns pacientes com as mesmas características: níveis de PSA, escore de Gleason e estado clínico se encontram em diferentes estágios de desenvolvimento de câncer, optamos por construir uma função de pertinência para o padrão-ouro levando em conta a frequência com que ocorre cada estágio. Assim, construímos uma função de pertinência ao conjunto dos pacientes cujo câncer não está confinado na próstata, conforme Figura 3. Tomamos uma amostra com 400 pacientes e submetemos ao SBRF. A cada

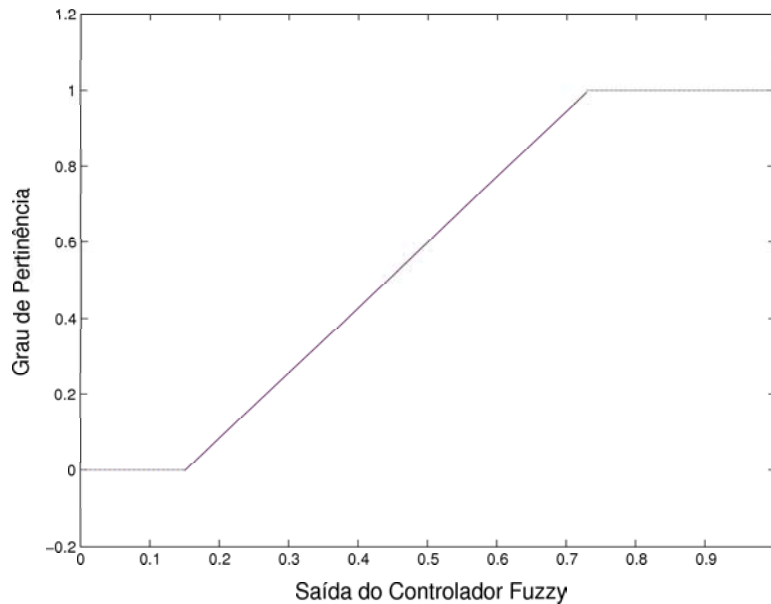


Figura 3: Função de Pertinência ao Conjunto dos Indivíduos com Câncer Não Confinado ao Órgão

paciente foi associado um resultado no intervalo $[0, 1]$. Calculamos, então, as medidas de sensibilidade e especificidade para cada ponto de corte possível neste intervalo. Alguns pontos são mostrados na Tabela 2.

Estas medidas incorporam a pertinência definida no padrão-ouro. Podemos então falar em *sensibilidade fuzzy* e *especificidade fuzzy* o que nos permite construir a curva ROC

Tabela 2: Valores de Sensibilidade e Especificidade para diversos pontos de corte para o Resultado do Teste

| Ponto de Corte | Especificidade | Sensibilidade |
|----------------|----------------|---------------|
| 0.2555 | 0.0153 | 0.9945 |
| 0.2836 | 0.0861 | 0.9615 |
| 0.3121 | 0.5420 | 0.6909 |
| 0.3759 | 0.7161 | 0.5514 |
| 0.3944 | 0.7203 | 0.5465 |
| 0.4097 | 0.8014 | 0.4403 |
| 0.4307 | 0.8122 | 0.4246 |
| 0.4379 | 0.8232 | 0.4079 |
| 0.4738 | 0.8861 | 0.2944 |
| 0.5000 | 0.9816 | 0.0827 |
| 0.5725 | 0.9978 | 0.0141 |
| 0.6078 | 0.9999 | 0.0014 |
| 0.6732 | 1.0000 | 0.0000 |

fuzzy mostrada na Figura 4.

Observando a Tabela 2 vemos que o valor de corte determina a relação entre a sensibilidade e especificidade do teste.

Uma maneira de escolher um ponto de corte é calcular a eficiência, ou seja, a média aritmética entre sensibilidade e especificidade. O melhor ponto teria maior eficiência. No nosso caso o ponto escolhido seria 0.3759 que tem eficiência igual a 0.6338. Este ponto seria usado para separar os pacientes que têm câncer confinado na próstata daqueles em que o câncer já envolveu a cápsula prostática e/ou estruturas adjacentes. Neste caso, o teste terá uma sensibilidade de 0.5514 e uma especificidade de 0.7161.

Porém, antes de decidir pelo ponto, o médico deverá ponderar sobre qual é o erro mais tolerável: falsos negativos (câncer não confinado não detectado) ou falsos positivos (pacientes com câncer confinado classificados erroneamente como não confinado). Por exemplo, se o médico achar conveniente diminuir os falsos positivos, ou seja, aumentar a especificidade do teste, (o que só será possível com a diminuição da sensibilidade), deve aumentar o valor de corte.

A área sob a curva ROC fuzzy é igual a 0.6620. Normalizando, temos $A_n = 0.7436$. Isto significa que, um paciente escolhido ao acaso tem 74% de chance de ser classificado corretamente utilizando o SBRF.

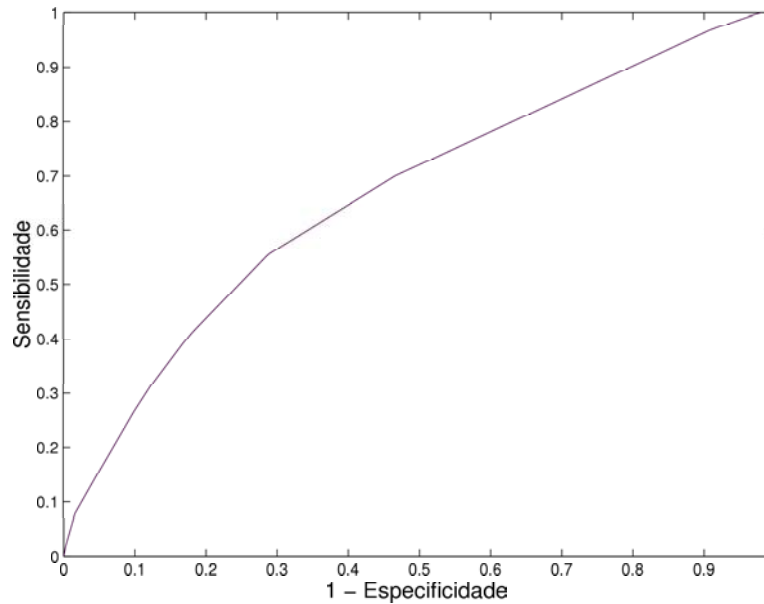


Figura 4: Curva ROC Fuzzy

7 Considerações Finais

A introdução da metodologia fuzzy na construção da curva ROC permite uma melhor representação da realidade nos casos em que não é possível uma classificação binária. Particularmente, no teste aqui avaliado, foi utilizado um padrão-ouro fuzzy que permitiu incluir a frequência com que determinado resultado do teste implica em determinado estágio do câncer. Isto foi necessário pois, nem sempre, a mesma condição do paciente, em termos de níveis de PSA, escore de Gleason e estado clínico, indicam o mesmo estágio do câncer o que deveria ocorrer no padrão-ouro clássico.

Qualquer ponto ao longo da curva ROC de um teste nos dá a sensibilidade e a especificidade do teste para o valor de corte associado ao ponto. Na curva ROC fuzzy nenhum ponto de corte tem sensibilidade e especificidade iguais a 1, o que indica um teste ideal. Acreditamos que a escolha do “melhor” ponto de corte deve ser feita com o auxílio de um especialista que tenha conhecimento detalhado dos riscos e benefícios da decisão médica derivada do resultado do teste.

A área sob a curva ROC fuzzy encontrada nos permite afirmar que o sistema baseado em regras fuzzy discrimina pacientes com câncer confinado ao órgão daqueles em que o câncer já envolveu a cápsula prostática e/ou vesículas seminais e/ou linfonodos.

Referências

- Castanho, M. J. P., Yamakami, A., Barros, L. C., e Vendite, L. L. (2003). Prognostic of the pathological stage of prostate cancer: An option using fuzzy sets theory. *Proceedings of the Third Brazilian Symposium on Mathematical and Computational Biology*, 1:247–253.
- Chakraborty, D. P. (1993). Observer performance methodology in medical imaging. in “Notes for Dr. Alavi’s Course”, Pennsylvania.
- De Leo, J. e Campbell, G. (1990). The fuzzy receiver operating characteristic function and medical decisions with uncertainty. *Proceedings of the First International Symposium on Uncertainty Modeling and Analysis, IEEE Computer Society Press*, páginas 694–699.
- Galparsoro, L. U. e Fernández, S. P. (1998). Curva roc. *Cad. Aten. Primaria*, 5(4):229–235.
- Leo, J. D. e Campbell, G. (1995). Biomedical applications of uncertainty modeling and analysis with fuzzy receiver operating characteristic methodology. *Proceedings of ISUMA-NAFIPS ’95*, páginas 192–197.
- Metz, C. E. (1978). Basic principles of roc analysis. *Seminars in Nuclear Medicine*, 8:283–297.
- Metz, C. E. (2003). Methodologic issues. in “Fourth National Forum on Biomedical Imaging in Oncology”.
- Parasuraman, R., Masalonis, A., e Hancock, P. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42:636–659.
- SEER (2003). *Surveillance, Epidemiology, and End Results (SEER) Program, Public-Use Data (1973-2000)*. National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, www.seer.cancer.gov.