

**Artigo Original**

Recebido em 21/05/2009, aceito em 24/03/2010

## **A metodologia ROC na avaliação de um modelo *fuzzy* de predição do estágio patológico do tumor de próstata**

*The ROC methodology in the evaluation of a fuzzy model to prediction of the prostate tumor's pathological stage*

**Graciele Paraguaia Silveira\***  
**Laécio Carvalho de Barros**  
**Laércio Luis Vendite**

Departamento de Matemática Aplicada,  
IMECC / UNICAMP  
Universidade Federal de Goiás - Campus Jataí  
Rod. BR 364, Km 192 - CP 03  
75800-000 Jataí, Goiás - Brasil  
E-mail: [gracimat@ime.unicamp.br](mailto:gracimat@ime.unicamp.br)

**Ubirajara Ferreira**  
Disciplina de Urologia, FCM / UNICAMP

**Athanase Billis**  
Departamento de Patologia, FCM / UNICAMP

\*Autor para correspondência

**Resumo**

Nos últimos anos o aumento da incidência de casos de câncer de próstata configura-se como um importante problema de saúde pública e um desafio para a ciência médica. O objetivo deste trabalho é a avaliação do desempenho de um modelo matemático, desenvolvido por Silveira (2007) para prever o estadiamento patológico do câncer de próstata, por meio da metodologia ROC (*Receiver Operating Characteristic*). O modelo consiste num sistema baseado em regras *fuzzy* (SBRF), que combina os dados pré-cirúrgicos – estado clínico, nível de PSA e grau de Gleason – acionando um conjunto de regras linguísticas, elaboradas com base nas informações presentes nos nomogramas já existentes. A saída do sistema fornece as possibilidades do indivíduo, com determinado quadro clínico, se enquadrar em cada um dos estádios de extensão do tumor: *localizado*, *localmente avançado* e *metastático*. Para a análise do poder discriminatório do modelo *fuzzy* como um teste de diagnóstico, foi construída, a partir das medidas de sensibilidade e especificidade, a curva ROC e calculada a área total sob a curva, como medida de desempenho. Além disso, foram obtidos (de duas maneiras distintas) os pontos de corte mais “adequados”, isto é, um limiar de decisão entre a doença estar totalmente localizada no interior da glândula prostática ou não. Dados reais de pacientes do Hospital de Clínicas da UNICAMP foram usados nos cálculos e a cirurgia – prostatectomia radical – foi adotada como padrão-ouro. Os resultados alcançados mostraram que o modelo *fuzzy* em questão pode vir a ser utilizado para discriminar câncer de próstata localizado.

**Palavras-chave:** Câncer de próstata, Diagnóstico, Conjuntos *fuzzy*, Curva ROC.

**Abstract**

*In recent years, the increase in the incidence of prostate cancer has become a major public health problem and a challenge for medical science. The goal of this work is assessing the performance of a mathematical model, developed by Silveira (2007) to predict the pathological stage of the prostate cancer, through ROC methodology (Receiver Operating Characteristic). The model is a fuzzy rule-based system, that combines pre-surgical data – clinical stage, PSA level and Gleason score – availing of a set of linguistic rules made with base on information of the existents nomograms. The output of the system provides the possibilities of the individual, with certain clinical features, be in each stage of the tumor extension: localized, advanced locally and metastatic. To analyze the discriminatory power of the fuzzy model as a diagnosis test, was constructed from the measures of sensitivity and specificity, the ROC curve and calculated the total area under the curve, as measure of performance. Moreover, were obtained (in two different ways) the cutoff points most “appropriate”, that is a threshold for deciding between the disease is fully localized within the prostate gland or not. Real data of patients from the Clinics Hospital of UNICAMP were used in the calculations and the surgery – radical prostatectomy – was used as gold standard. The results showed that the fuzzy model in question can be used to discriminate localized prostate cancer.*

**Keywords:** Prostate cancer, Diagnosis, Fuzzy sets, ROC curve.

## Extended Abstract

### Introduction

The goal of this research is to evaluate the discrimination power of a fuzzy rule-based system, developed by Silveira (2007) – from a previous study done by Castanho (2005) – to predict the pathological stage of prostate cancer, using the ROC curve (Receiver Operating Characteristic).

The prostate cancer is the more frequent in men in Brazil (Ferreira and Nardi, 1999; SBU, 2008). To the complex decision-making by the specialist, regarding the extension of tumor, examinations are available: rectal examination, serum dosages of PSA (Prostatic Specific Antigen) and biopsy (Gleason score).

In an attempt to achieve greater accuracy in predictions, statistical tools (namely nomograms) have been developed by researchers. In the case of the prostate cancer, the medical use the nomograms of Stephenson and Kattan (2006), which provide the probabilities that the cancer is localized, advanced locally and metastatic, from the data pre-surgical.

However, these nomograms do not include most of the uncertainties involved in the problem, and according to specialists, the results do not reflect the Brazilian reality because were gotten from foreign populations.

The fuzzy sets theory is a good alternative for dealing with uncertainties and was adopted in the development of the mathematical prediction model of the prostate cancer stage.

For the model can be effectively regarded as a diagnostic test, the cutoff point (threshold for deciding whether the disease is localized or not) and the analysis of the ROC curve have to be determined.

### Materials and Methods

The ROC curve is a graphical representation of the ordered pairs (1-Specificity, Sensitivity). To calculate the measures of sensitivity and specificity of the test, we need that its possible results are classified in Positive (localized) and Negative (not localized).

Sensitivity is the probability that the test gives positive and the individual really has localized tumor. Specificity is the probability that the test gives negative and the individual really has not localized tumor.

The gold standard that provides the true diagnosis for comparison with the test in question is the surgery – radical prostatectomy.

The choice of the cutoff point was made from the calculation of the maximum efficiency of the test, obtained from two different ways: arithmetic mean between sensitivity and specificity (I) and the product between sensitivity and specificity (II).

Real data of 170 patients of the Clinics Hospital of UNICAMP were used in the constructing of the ROC curve.

### Results

For the construction of the ROC curve, the patients were submitted to the fuzzy rule-based system and the possible cutoff points

determined, for the calculation of measures of sensitivity and specificity.

The gold standard is the radical prostatectomy and the results of this procedure, divided the population into two groups: Localized and Not Localized.

Considering the efficiency (I), the cutoff point obtained is 0.2322 and their measures of sensitivity and specificity in this point are 0.9091 and 0.5732. Already for the efficiency (II) got the cutoff point 0.2723 and their sensitivity and specificity are 0.8409 and 0.6219, respectively.

The total area under the ROC curve is 0.7956 with confidence interval (CI) 95% (0.76-0.83).

### Discussion

For the first cutoff point obtained (0.2322), we can interpret that in 100 patients with localized cancer, would be detected approximately 91 and 9 would be classified as advanced tumor. On the other hand, of 100 patients with not localized cancer, 57 had negative results and 43 were incorrectly classified as localized. The interpretation concerning the second cutoff point is analogous.

An important observation is that whatever the choice of the cutoff point, errors occur and therefore should be assessed the more tolerable: false positive or false negative results.

The area under the ROC curve is 0.7956 with CI 95% (0.76-0.83). This shows that given two patients, one with localized cancer and one with advanced cancer, the probability that the test will have greater magnitude for the individual with localized cancer is 0.7956. A test totally incapable of discriminate the individuals would have an area under the curve of 0.5 (Martinez et al., 2003).

One advantage of the ROC curve is the direct representation of the ability of a variable in discriminate subjects, with empirical cutoff points. However, a limitation is the fact that the classification of individuals is restricted to two groups: clinical condition present or absent. In real situations, there are a variety of information and many are neglected when this classification is dichotomous. An option that can be adopted is the use of a membership function for the Positive set, thereby obtaining a fuzzy result for the test. This is a proposal of future work.

### Conclusion

The decision of a specialist, regarding the extension of the prostate tumor, is complex and surrounded by uncertainty. A fuzzy mathematical model was developed with the intention to include these uncertainties and the own empirical knowledge of the professional. In this work, the model in question was evaluated through ROC curve. We determined two cutoff points, between the disease is localized or not.

The ROC curve was made and the calculation of the area under the curve showed that the model can be used as discriminator of localized tumor.

## Introdução

A incidência de casos de câncer tem aumentado de maneira considerável, principalmente a partir do século passado, configurando-se, na atualidade, como um dos mais importantes problemas de saúde pública mundial.

O câncer de próstata é o mais frequente em homens adultos e a segunda principal causa de morte por doença maligna no Brasil (Ferreira e Nardi, 1999; SBU, 2010). O exame clínico (toque retal) e a dosagem sérica do Antígeno Prostático Específico (PSA) são indicados visando o diagnóstico precoce. Os resultados podem sugerir a realização da biópsia prostática. A tomada de decisão do especialista, com relação ao estadiamento patológico da doença, ocorre com base nos resultados dos exames citados.

Na tentativa de obter maior precisão nas predições, os pesquisadores têm desenvolvido ferramentas baseadas em modelos estatísticos, denominadas nomogramas. No caso do câncer de próstata, um nomograma que vem sendo utilizado é o de Stephenson e Kattan (2006). Esse nomograma combina os dados pré-cirúrgicos e fornece as probabilidades de que o câncer de próstata esteja localizado totalmente no interior da glândula, localmente avançado ou metastático (espalhando-se pelo organismo). No entanto, determinar a extensão do tumor é um processo complexo e as opções disponíveis não incluem a maioria das incertezas e subjetividades envolvidas.

O exame clínico é útil na identificação de regiões suspeitas, mas depende essencialmente da experiência do médico. As dosagens séricas de PSA auxiliam na detecção da doença, entretanto ainda não permitem a descoberta precoce de todos os casos. Além disso, há pacientes com mesmo PSA, em estádios patológicos diferentes e, portanto, o estadiamento do câncer não pode ser determinado somente a partir dessa medida.

Da mesma forma, as biópsias podem não revelar a presença de tumor na glândula prostática, retardando o diagnóstico e sendo muitas vezes realizadas desnecessariamente.

Já os nomogramas estatísticos conhecidos, de Stephenson e Kattan (2006), não apresentam resultados satisfatórios para a população brasileira, pois foram desenvolvidos a partir de dados de populações estrangeiras.

É de importância vital descobrir em qual estágio de extensão o tumor se encontra, pois a cirurgia de retirada da glândula (prostatectomia radical)

apresenta um alto índice de cura, se o câncer estiver localizado.

Diante disso, foi desenvolvido um sistema baseado em regras *fuzzy* – tomando como referência o estudo de Castanho (2005) – que combina as informações encontradas através dos exames, para prever o estadiamento patológico do câncer de próstata (Silveira, 2007; Silveira *et al.*, 2008).

A teoria de conjuntos *fuzzy* vem sendo aplicada em medicina desde a década de 70 (Adlassnig, 1980; Kalmanson e Stegall, 1975) e foi adotada na construção do modelo devido à capacidade de lidar com as imprecisões, indo assim de encontro às reais necessidades clínicas.

Foram feitas simulações com dados reais de pacientes do Hospital de Clínicas da UNICAMP e os resultados, em princípio, foram validados por especialistas. Entretanto é necessária uma análise formal e objetiva do desempenho do modelo, para que efetivamente seja considerado como teste de diagnóstico.

O objetivo deste trabalho é a avaliação do modelo *fuzzy* em sua capacidade de discriminar pacientes com câncer localizado, daqueles com envolvimento além da glândula prostática, por meio da curva ROC (*Receiver Operating Characteristic*).

## O modelo *fuzzy*

Um sistema baseado em regras *fuzzy* (SBRF) compreende quatro módulos principais: um módulo de fuzzificação, no qual se encontram os valores *fuzzy* (funções de pertinência) das variáveis de entrada e saída do sistema; uma base de regras; um módulo de inferência, que utiliza um método de processamento das variáveis de entrada e saída e a base de regras, método este escolhido previamente; e um módulo de defuzzificação, que transforma a saída em um valor numérico (Barros e Bassanezi, 2006; Pedrycz e Gomide, 1998).

No modelo *fuzzy* em questão, as variáveis de entrada consideradas foram: **Estádio Clínico**, classificado linguisticamente como T1, T2a, T2b, T2c e T3a, de acordo como o sistema internacional TNM de classificação de tumores malignos; **Nível de PSA**, considerado Normal (até 4 ng/mL), Levemente Elevado (4-10 ng/mL), Moderadamente Elevado (10-20 ng/mL) e Altamente Elevado (acima de 20 ng/mL), de acordo com a classificação presente na literatura; **Score de Gleason**, classificado em Bem Diferenciado (graus 2, 3 e 4), Médio Diferenciado (grau 5), Médio-Baixo Diferenciado (grau 6), Pouco Diferenciado (grau 7) e Não Diferenciado (graus 8, 9 e 10).

O sistema de Gleason é utilizado pelos patologistas para análise de amostras de tecido retiradas na biópsia prostática.

Para a elaboração da base de regras *fuzzy* foram feitas todas as combinações entre estado clínico, nível de PSA e escore de Gleason, levando em conta todos os termos linguísticos atribuídos às variáveis e todas as probabilidades presentes nos nomogramas de Stephenson e Kattan (2006) para os diferentes estádios da doença. As funções de pertinência para as variáveis de entrada, escolhidas juntamente com especialistas, foram trapezoidais.

No modelo de Castanho (2005), que serviu de referência, foram utilizados os nomogramas de Partin *et al.* (1997).

O processo de inferência adotado foi o método de Mamdani e a defuzzificação foi feita pelo método do Centro de Gravidade (Massad *et al.*, 2008).

A variável de saída *Estádio da Doença* foi classificada por meio dos seguintes termos linguísticos: Localizado, Localmente Avançado e Metastático. Suas funções de pertinência foram ajustadas a partir da adaptação de uma função triangular, adotada como ponto de partida para o estudo. O método usado foi o dos quadrados mínimos nos pontos que representam as probabilidades (Kattan *et al.*) de cada estágio, para várias amostras de pacientes submetidos ao SBRF.

Em seguida, foi utilizado um método que transforma grau de pertinência em probabilidade (Klir e Yuan, 1995). As funções ajustadas foram normalizadas e concentradas, e estão representadas na Figura 1.

O sistema foi implementado utilizando o Toolbox *Fuzzy* do MATLAB®, v. 7.0.

Os resultados obtidos pelo SBRF representam o grau com o qual o quadro clínico do paciente é compatível

com os estádios extensão do tumor. A intenção é que este modelo *fuzzy* seja apresentado aos especialistas da área de uro-oncologia, como uma alternativa na predição do estágio patológico do câncer de próstata, frente aos nomogramas tradicionais existentes.

Nesse sentido, um *software* foi elaborado, a partir do modelo *fuzzy* descrito, e consiste em uma interface gráfica que faz a interação com as sub-rotinas que efetuam os cálculos (Silveira *et al.*, 2008).

O próximo passo é a avaliação formal e objetiva dos resultados, visando conhecer o desempenho deste teste diagnóstico. A metodologia indicada para isto, e que será descrita a seguir, é a análise ROC (*Receiver Operating Characteristic*).

### Materiais e Métodos

A metodologia ROC surgiu no contexto de detecção de sinais eletrônicos e problemas com radares, durante a segunda guerra mundial (Zweig e Campbell, 1993). A partir dos anos 70 a metodologia se disseminou amplamente em vários ramos da pesquisa biomédica (Martinez *et al.*, 2003).

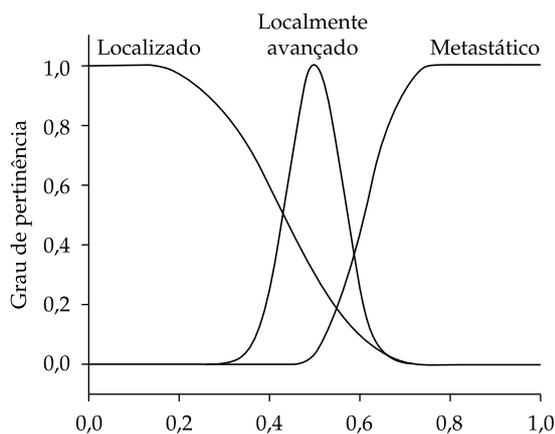
A curva ROC é uma representação gráfica dos pares (1 - *Especificidade*, *Sensibilidade*), no plano cartesiano. Para calcular sensibilidade e especificidade de um teste de diagnóstico é necessário que seus possíveis resultados sejam classificados em *Positivo* ou *Negativo*.

Normalmente os resultados dos testes são dados por variáveis contínuas num determinado intervalo. Assim, atribuem-se os termos *Positivo* ou *Negativo*, a partir de um valor escolhido  $k$  (ponto de corte), para indicar se o indivíduo apresenta ou não o fator pesquisado. Além disso, deve-se conhecer o verdadeiro diagnóstico, obtido através de algum procedimento denominado padrão-ouro.

Um teste classifica corretamente um indivíduo se seu resultado, dado pelo padrão-ouro, for *Positivo* e o indivíduo pertencer ao conjunto *Doente* – denotado por Verdadeiro Positivo (VP) – ou se o resultado for *Negativo* e o indivíduo pertencer ao conjunto *Saudável* – denotado por Verdadeiro Negativo (VN) – (Castanho, 2005).

Entretanto, há casos em que o resultado do teste é positivo, quando na realidade o indivíduo está saudável ou o resultado é negativo e o indivíduo está doente. Esses erros são denotados por Falso Positivo (FP) e Falso Negativo (FN), respectivamente.

*Especificidade* é a probabilidade de que o teste dê negativo para um indivíduo saudável. Logo, considerando um ponto de corte  $k$ , segue que:



**Figura 1.** Funções de pertinência – variável Estádio da Doença. **Figure 1.** Membership functions – variable Disease Stage.

$$\text{Especificidade} = \frac{\sum_{i=1}^n VN(x_i)}{\sum_{i=1}^n VN(x_i) + \sum_{i=1}^n FP(x_i)} \quad (1)$$

onde  $x_i$  é o  $i$ -ésimo indivíduo submetido ao teste, classificado como negativo para  $x_i < k$ , e  $n$  é o número total de indivíduos da população.

*Sensibilidade* é a probabilidade de que o teste forneça resultado positivo para um indivíduo que tenha a doença. Assim, para um ponto de corte  $k$ , tem-se:

$$\text{Sensibilidade} = \frac{\sum_{i=1}^n VP(x_i)}{\sum_{i=1}^n VP(x_i) + \sum_{i=1}^n FN(x_i)} \quad (2)$$

onde  $x_i$  é o resultado do teste para o  $i$ -ésimo indivíduo submetido a ele, considerado como positivo para  $x_i \geq k$ , e  $n$  é o número total de indivíduos da população.

A escolha do ponto de corte envolve a combinação ótima entre especificidade e sensibilidade. Se o propósito principal do teste é a identificação da doença, então se deve selecionar um ponto de corte com alta sensibilidade. Já para a confirmação da doença, os casos positivos encontrados serão testados novamente e, portanto, no segundo teste deve-se exigir alta especificidade.

Uma das formas de se escolher o ponto de corte é por meio do cálculo da média aritmética entre sensibilidade e especificidade (Castanho, 2005), isto é:

$$\text{Eficiência} = \frac{\text{Sensibilidade}(k) + \text{Especificidade}(k)}{2} \quad (3)$$

Outro método de seleção do ponto de corte é proposto por Duarte (2004) e consiste em obter o ponto para o qual se tenha o maior produto entre sensibilidade e especificidade, ou seja:

$$\text{Eficiência} = \text{Sensibilidade} \times \text{Especificidade} \quad (4)$$

A área sob a curva ROC constitui-se numa medida-resumo usual do desempenho de um teste de diagnóstico, pois é estimada levando-se em consideração todas as medidas de sensibilidade e especificidade (Martinez *et al.*, 2003).

Se o resultado do teste for positivo para valores maiores do que o ponto de corte, e este resultado se encontrar em um intervalo contínuo, pode-se interpretar a área sob a curva ROC como a probabilidade de que o resultado do teste seja maior para um indivíduo doente do que para um indivíduo saudável (Zweig e Campbell, 1993). Para os casos em

que for considerado positivo, isso significa a presença da doença.

O objetivo deste trabalho é avaliar o poder de discriminação do modelo *fuzzy*, de predição do estágio patológico do câncer de próstata, em relação à doença estar localizada na glândula (positivo) ou não (negativo).

Para a construção da curva ROC foram utilizados dados reais de pacientes do Hospital de Clínicas da UNICAMP, obtidos em janeiro de 2009 junto a especialistas da área de uro-oncologia.

Na base de dados constam um total de 547 pacientes com câncer de próstata, com idades variando entre 47 e 87 anos (média: 69), que passaram pelo hospital entre 1991 e 2008.

Destes pacientes, 213 foram submetidos à cirurgia de retirada da glândula prostática, denominada prostatectomia radical. Contudo, apenas 170 pacientes contêm todas as informações necessárias para os cálculos das medidas de sensibilidade e especificidade.

A prostatectomia radical possibilita ao especialista conhecer o verdadeiro diagnóstico do paciente, ou seja, o real estágio patológico do tumor da próstata. Portanto, o diagnóstico a partir da cirurgia foi adotado como padrão-ouro.

## Resultados

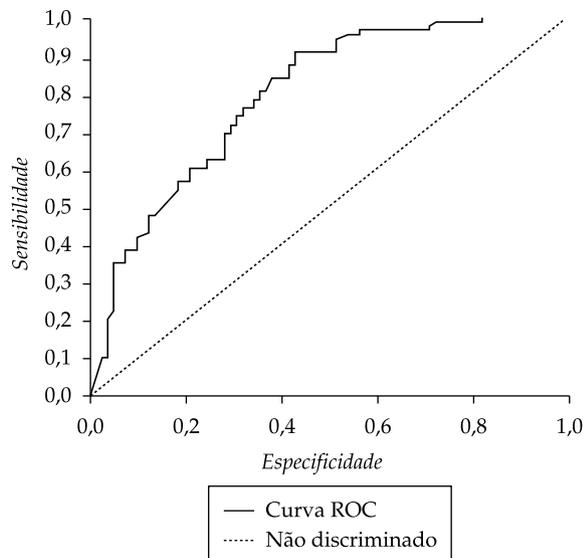
Para a obtenção da curva ROC, os 170 pacientes foram submetidos ao sistema baseado em regras *fuzzy* – SBRF – e os possíveis pontos de corte, no intervalo [0,1], foram encontrados, para o cálculo das medidas de sensibilidade e especificidade.

Os resultados da cirurgia, presentes no banco de dados considerado, dividiram a população em dois conjuntos: Localizado e Não Localizado.

Em seguida, buscou-se a maior medida de eficiência visando obter assim os pontos de corte “mais adequados”. A eficiência foi calculada, inicialmente, através da equação 3. O ponto de corte que apresentou maior eficiência foi 0,2322; a medida de sensibilidade do teste nesse ponto é 0,9091 e a medida de especificidade é 0,5732.

Calculando a eficiência também por meio da equação 4, o ponto de corte obtido foi 0,2723. Nesse ponto, a sensibilidade do teste é 0,8409 e a especificidade é 0,6219.

As interpretações desses resultados encontram-se na próxima seção.



**Figura 2.** Curva ROC para o estágio do câncer de próstata.  
**Figure 2.** ROC curve for the prostate cancer stage.

A curva ROC construída a partir dos dados dos 170 pacientes submetidos à prostatectomia radical encontra-se na Figura 2.

A área total sob a curva ROC é 0,7956, com intervalo de confiança (IC) 95% (0,76-0,83). Esses resultados são discutidos a seguir.

### Discussão

Pode-se interpretar que, para o primeiro ponto de corte obtido (0,2322), com sensibilidade de 0,9091, de cada grupo de 100 pacientes com câncer localizado, aproximadamente 91 seriam detectados e 9 seriam classificados como localmente avançado. Por outro lado, com especificidade de 0,5732, de 100 pacientes com câncer não localizado, 57 teriam resultados negativos, isto é, corretos. Entretanto, 43 seriam classificados incorretamente como localizados.

Para o segundo ponto de corte encontrado (0,2723), a interpretação é feita de modo análogo ao anterior, isto é, de cada 100 pacientes com câncer localizado, 84 seriam apontados pelo teste e 16 seriam classificados como não localizados. Por outro lado, de 100 pacientes com câncer não localizado, 62 seriam identificados, enquanto que 38 seriam incorretamente considerados como câncer localizado.

Qualquer que seja a escolha do ponto de corte deve-se avaliar a relação custo-benefício dos tratamentos aos quais serão submetidos os pacientes e assim definir o que seria mais tolerável: mais falsos positivos (câncer avançado incorretamente classificado como localizado) ou mais falsos negativos (câncer localizado incorretamente classificado como avançado).

A área sob a curva ROC construída, ilustrada na Figura 2, é 0,7956, com IC 95% (0,76-0,83). Isso mostra que, dados dois pacientes, um com câncer localizado e outro com câncer avançado, a probabilidade de que o resultado do teste tenha maior magnitude para o indivíduo com câncer localizado é de 0,7956.

Quanto maior a capacidade de um teste em discriminar os indivíduos segundo estes dois grupos, mais a curva ROC se aproximaria do canto superior esquerdo do gráfico, e a área sob a curva seria próxima de 1. Um teste totalmente incapaz de discriminar os indivíduos teria uma área sob a curva de 0,5 (Martinez et al., 2003).

É importante ressaltar as vantagens e limitações do uso da curva ROC. Sem dúvida, uma vantagem é a possibilidade de se ter em um único gráfico uma representação direta da habilidade de uma variável em discriminar sujeitos em diferentes grupos, segundo pontos de corte empíricos. No entanto, qualquer estudo envolvendo sensibilidade e especificidade pressupõe a existência de um padrão-ouro “verdadeiro”. Muitas vezes a comparação ocorre entre um novo teste desenvolvido e outro teste que apresenta também problemas de grau desconhecido (Phelps e Hutson, 1995).

Além disso, outra limitação que pode ser apontada é o fato que a classificação de indivíduos em grupos está restrita a dois grupos: um em que a condição clínica está presente e outro em que tal condição está ausente.

Em situações reais, mais comuns, há multiplicidade de fatores envolvidos e seria importante que os resultados do teste de diagnóstico pudessem incorporá-las. Uma alternativa seria tentar obter uma função de pertinência para o conjunto *Positivo* e obter o resultado do teste *fuzzy*. Assim, ao invés de pacientes discriminados apenas como Localizado (grau 1) ou Não Localizado (grau 0), será possível conhecer o grau, entre 0 e 1, com o qual o indivíduo se enquadra no conjunto *Positivo* (Localizado).

Esse é um trabalho futuro a ser desenvolvido.

### Conclusões

A teoria de conjuntos *fuzzy* é uma alternativa para a inclusão de imprecisões, subjetividades e do próprio conhecimento prático dos especialistas em modelos matemáticos.

O objetivo inicial foi a análise do desempenho de um modelo *fuzzy*, aplicado ao problema de predição do estágio patológico do câncer de próstata, utilizando a metodologia ROC (*Receiver Operating*

*Characteristic*). A intenção foi avaliar a capacidade do teste em discriminar pacientes com câncer de próstata localizado, daqueles com envolvimento além da glândula.

Dois métodos – equações 3 e 4 – são encontrados na literatura, para a obtenção do ponto de corte “mais adequado”. Um ponto de corte é um limiar de decisão, ou seja, para resultados do teste maiores do que o ponto de corte, o sujeito é classificado como câncer localizado; para valores menores do que tal ponto, o paciente é classificado como câncer não localizado. No presente caso foram obtidos os pontos de corte 0,2322 e 0,2723.

Por fim, foi construída a curva ROC a partir das medidas de sensibilidade e especificidade, e calculada a área total sob a curva. O resultado obtido foi 0,7956, com IC 95% (0,76-0,83), significando que o modelo *fuzzy* de predição do estágio do câncer de próstata pode vir a ser utilizado para discriminar câncer localizado.

Esse trabalho contou com a participação efetiva de especialistas em uro-oncologia do HC da UNICAMP, que analisaram e validaram os resultados obtidos levando-se em conta suas experiências na prática médica.

### Agradecimentos

Agradecemos ao Dr. Wagner Eduardo Matheus, médico urologista do HC da UNICAMP, pelo auxílio na interpretação do banco de dados. O segundo autor agradece o suporte financeiro da FAPESP – processo nº 06/05920-7 e CNPq – processo nº 307890/2006-6.

### Referências

ADLASSNIG, K. P. A Fuzzy logical model of computer-assisted medical diagnosis. **Methods of Information in Medicine**, v. 19, n. 3, p. 141-148, 1980.

BARROS, L. C.; BASSANEZI, R. C. **Tópicos de lógica Fuzzy e Biomatemática**. Campinas: UNICAMP/IMECC, 2006. 344 p.

CASTANHO, M. J. P. **Construção e Avaliação de um Modelo Matemático para prever a evolução do Câncer de Próstata e descrever seu crescimento utilizando a Teoria de Conjuntos Fuzzy**. 2005. 127 f. Tese (Doutorado em Engenharia Elétrica) - FEEC/UNICAMP, Campinas, 2005.

DUARTE, P. S. **Definição de conduta na investigação de doença coronária obstrutiva utilizando teoria de conjuntos Fuzzy**

**aplicada a dados clínico-epidemiológicos, ergométricos e cintilográficos**. 2004. 119 f. Tese (Doutorado em Medicina) - Faculdade de Medicina, Universidade de São Paulo – USP, São Paulo, 2004.

FERREIRA, U.; NARDI, A. C. Câncer de Próstata. In: NETTO, N. R. (Ed.). **Urologia Prática**. 4 ed. São Paulo: Atheneu, 1999.

KALMANSON, D.; STEGALL, H. F. Cardiovascular investigations and Fuzzy sets theory. **The American Journal of Cardiology**, v. 35, n. 1, p. 80-84, 1975.

KLIR, G.; YUAN, B. **Fuzzy sets and fuzzy logic theory and applications**. Upper Saddle River: Prentice Hall, 1995. 574 p.

MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA, B. B. A curva ROC para testes diagnósticos. **Caderno de Saúde Coletiva**, v. 11, n. 1, p. 07-31, 2003.

MASSAD, E.; ORTEGA, N. R. S.; BARROS, L. C.; STRUCHINER, C. J. **Fuzzy Logic in Action: Applications in Epidemiology and Beyond**. Heidelberg: Springer-Verlag Berlin, 2008. 348 p.

PARTIN, A. W.; KATTAN, M. W.; SUBONG, E.; WALSH, P. C.; WOJNO, K. J.; OESTERLING, J. E.; SCARDINO, P. T.; PEARSON, J. D. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. **Journal of the American Medical Association (JAMA)**, v. 277, n. 18, p. 1445-1551, 1997.

PEDRYCZ, W.; GOMIDE, F. A. C. **An introduction to Fuzzy sets: Analysis and design**. Massachusetts: The MIT Press, 1998. 465 p.

PHELPS, C. E.; HUTSON, A. Estimating diagnostic test accuracy using a “Fuzzy gold standard”. **Medical Decision Making**, v. 15, p. 44-57, 1995.

SILVEIRA, G. P. **Aplicação da Teoria de Conjuntos Fuzzy na predição do estadiamento patológico do Câncer de Próstata**. 2007. 105 f. Dissertação (Mestrado em Matemática Aplicada) - IMECC/UNICAMP, Campinas, 2007.

SILVEIRA, G. P.; VENDITE, L. L.; BARROS, L. C. Software desenvolvido a partir de um Modelo Matemático Fuzzy para prever o estágio patológico do Câncer de Próstata. **Biomatemática**, v. 18, p. 27-36, 2008.

SOCIEDADE BRASILEIRA DE UROLOGIA - SBU. **Câncer de Próstata**. Disponível em: <<http://www.sbu.org.br/2010/03/cancer-de-prostata/>>. Acesso em: 18 ago. 2010.

STEPHENSON, A. J.; KATTAN, M. W. Nomograms for prostate cancer. **BJU International**, v. 98, n. 1, p. 39-46, 2006.

ZWEIG, M. H.; CAMPBELL, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. **Clinical Chemistry**, v. 39, n. 4, p. 561-577, 1993.

