

Aplicação de um Método Computacional para o Diagnóstico Precoce do Câncer de Próstata Usando Reconhecimento de Padrões Proteômicos

Elzenir Montes, Lúcio Campos, Wesley Araujo, Ewaldo Santana, Claudyane Araujo
Universidade Estadual do Maranhão – Cidade Universitaria Paulo VI, MA, Brasil
elzenir.mtm@hotmail.com, lucioflavio@gamil.com, wesleydominicies@gmail.com,
ewaldoeder@gmail.com, claudy.sa@gmail.com

Resumo

O presente trabalho apresenta um método baseado em reconhecimento de padrões proteômicos para o diagnóstico precoce do câncer de próstata, utilizando técnicas computacionais, aplicadas na base de dados de padrões proteômicos *SELDI-TOF*. O método baseia-se em classificar o indivíduo quanto ao estágio de portabilidade do câncer de próstata. Para tanto, usa-se a técnica de Análise de Componentes Independente (ICA) para extrair as características, depois o algoritmo de Máxima Relevância e Mínima Redundância para reduzir o custo computacional, e finalmente a Máquina de Vetores de Suporte para a classificação. Obtêm-se o melhor resultado do método com um vetor de 27 características, alcançando acurácia, especificidade e sensibilidade, respectivamente de 89,21%, 83,68% e 95,08%.

Palavras-chave: Análise de Componentes Independentes. Câncer de próstata. Máquina de vetor de Suporte. Máxima relevância e mínima redundância. Reconhecimento de padrões.

Title: Application of a Computational Method for the Early Diagnosis of Prostate Cancer Using Proteomic Pattern Recognition

Abstract

This paper presents a method based on the recognition of proteomic patterns for the early diagnosis of prostate cancer, using computational techniques, applied in the database of SELDI-TOF proteomic patterns. The method is based on classifying the individual as to the portability stage of prostate cancer. To do so, the Independent Component Analysis (ICA) technique is used to extract the characteristics, after which are utilized the algorithm of Maximum Relevance and Minimum Redundancy to reduce the computational cost, and finally the Support Vector Machine to obtain the classification. The best result of the method was obtained with a vector of 27 characteristics, achieving accuracy, specificity and sensitivity, respectively of 89.21%, 83.68% and 95.08%.

Keywords: Independent component analysis. Prostate cancer. Support vector machine. Maximum relevance and minimum redundancy. Pattern Recognition

1. Introdução

Atualmente, o câncer de próstata é considerado um problema de saúde pública em muitos países do mundo devido a sua grande incidência e o seu crescimento exponencial. O Brasil vem passando, nas últimas décadas, por alterações de contexto social, econômico e, conseqüentemente, de saúde. O aumento da expectativa de vida, a melhoria e a evolução dos métodos diagnósticos podem explicar o crescimento das taxas de incidência ao longo dos anos no país. Além disso, a melhoria da qualidade dos sistemas de informação do país e a ocorrência de sobrediagnóstico, em função da disseminação do rastreamento do câncer de próstata com PSA e toque retal, também influenciam na magnitude da doença (INCA, 2015).

Na sua fase inicial esta doença possui uma evolução bastante sigilosa, tanto que a maioria dos pacientes não apresentam sintomas, e quando estes aparecem são confundidos com o crescimento benigno da próstata.

A detecção precoce de um câncer compreende duas diferentes estratégias: uma, destinada ao diagnóstico do câncer em indivíduos que apresentam sintomas ou sinais iniciais da doença (diagnóstico precoce) e outra, voltada para pessoas sem qualquer sinal ou sintoma e aparentemente saudáveis (rastreamento) (INCA, 2016).

No Brasil, em 2013, ocorreram quase 14 mil mortes por câncer de próstata. O Ministério da Saúde do Brasil, assim como ocorre em outros países (Austrália, Canadá e Reino Unido), não recomenda a organização de programas de rastreamento de câncer de próstata. Tal prática não está indicada, pois ainda existe considerável incerteza sobre a existência de benefícios associados. Portanto, ações de controle da doença devem focar em outras estratégias, como a prevenção primária e o diagnóstico precoce (INCA, 2015).

O tratamento do câncer de próstata, como nos demais tipos de câncer, é bastante invasivo, trazendo consigo efeitos colaterais que diminuem a qualidade de vida do paciente e familiares; e seu tratamento depende do estadiamento clínico.

Para a doença localizada, podem ser oferecidos a cirurgia, radioterapia e até mesmo observação vigilante (em algumas situações especiais). Para doença localmente avançada, a radioterapia ou a cirurgia em combinação com tratamento hormonal têm sido utilizadas. Para doença metastática, quando o tumor original já se espalhou para outras partes do corpo, o tratamento de eleição é a terapia hormonal (INCA, 2016).

Dessa forma, quanto mais cedo essa doença for detectada mais fácil e eficaz será seu tratamento, evitando o óbito e aumentando a qualidade de vida do paciente. Os estudos existentes na literatura para avaliação da efetividade do rastreamento do câncer da próstata apresentam problemas metodológicos em seu desenho. O baixo valor preditivo positivo dos testes de rastreamento levam a uma elevada proporção de resultados falso-positivos, e, conseqüentemente, um alto índice de realização de biópsias desnecessárias (INCA, 2015).

Além dos métodos de diagnósticos tradicionais, tais como exame clínico e dosagem do PSA, que possuem baixa sensibilidade, deve-se investir em outras formas de detectar esta enfermidade para auxiliar o profissional da saúde no diagnóstico precoce e mais preciso.

Na última década, pesquisadores de várias áreas vêm realizando de forma bem sucedida trabalhos envolvendo o diagnóstico precoce de doenças através de técnicas computacionais (CADs), como câncer de ovário (Araujo, Campos, & Furtado, 2014), diabetes (Ribeiro, 2015) e Granulomatose de Wegener (Linhares (Linhares, Campos, & Santana, 2016).

O método proposto foi realizado fazendo uso de sinais proteômicos, no qual usou-se a técnica de Análise de Componentes Independentes, para a extração de características, juntamente com o algoritmo de Máxima Relevância e Mínima Redundância (mRMR), a fim de selecionar as características mais significativas. Por fim, usou-se a Máquina de Vetores de Suporte, para a classificação. Todas as técnicas descritas aqui serão apresentadas nas próximas seções juntamente com os resultados e a validação do método proposto.

2. Método Proposto

O método proposto engloba basicamente três etapas principais: extrair as características do sinal proteômico, fazendo uso da técnica Análise de Componentes Independentes (ICA), posteriormente reduzir a dimensionalidade do conjunto de características, através da técnica de Máxima Relevância e Mínima Redundância (mRMR) e realização da classificação final, por meio da Máquina de Vetores de Suporte (SVM). As técnicas computacionais empregadas em cada etapa deste método serão descritas nas subseções subsequentes. A figura 1 mostra o diagrama de bloco do método proposto.

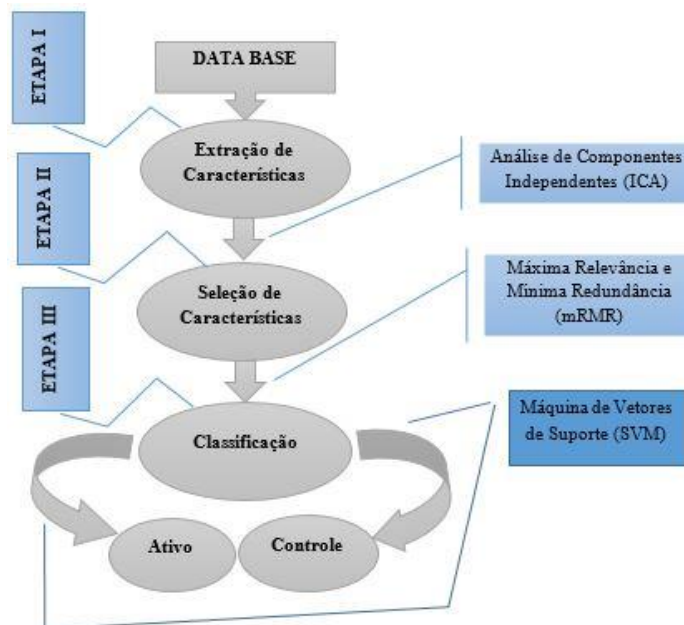


Figura 1 – Diagrama de blocos do método proposto

A implementação e os testes do método proposto foram efetuados usando a linguagem MatLab®, R2013, fazendo uso ainda dos pacotes *fastICA* e *mRMR*, disponíveis em: (AAPO, 2015), (Matworks, 2015); e o pacote SVM obtido em (Araujo, 2014).

2.1. Extração de Características Através da Análise de Componentes Independentes

A análise de componentes independentes é uma técnica estatística e computacional, que revela componentes ou fatores subjacentes a um conjunto de variáveis aleatórias, medições, ou sinais observados multivariados. Difere das técnicas tradicionais utilizadas nesta área, porque tais componentes são estatisticamente independentes (ou têm dependência estatística minimizada) e são não gaussianos (AAPO, 2015).

A ICA é um poderoso método utilizado principalmente no tratamento de problemas de separação cega de sinais de fontes independentes, também conhecido como *Blind Source Separation* (BSS). O problema é definido na estimação da saída de uma fonte conhecida ao receber vários sinais desconhecidos e misturados. A ICA tem sido aplicada em várias áreas, tais como: comunicação móvel, engenharia biomédica, instrumentação, dentre outras.

Segundo Comon (1994), a ICA de um vetor aleatório $X = [x_1, x_2, \dots, x_m]^T$ consiste na determinação de uma transformação linear $Y = WX$, que minimiza a dependência estatística entre os elementos do vetor aleatório $Y = [y_1, y_2, \dots, y_n]^T$. Assumindo-se que os elementos do vetor aleatório X sejam modelados como combinações lineares dos elementos do vetor aleatório $S = [s_1, s_2, \dots, s_n]^T$. Pode-se escrever na forma matricial

$$X = AS \quad (1)$$

Os elementos de S são estatisticamente independentes entre si, e A corresponde a uma matriz de coeficiente de mistura de dimensão $m \times n$.

Para se estimar, a matriz de mistura A é necessário satisfazer a restrição, que a matriz A seja quadrada e para fins de simplicidade foi assumido, que o número de componentes independentes equivale ao número de misturas observadas, neste caso tem-se $n = m$, e A passa a corresponder a uma matriz de quadrada, dimensão $m \times m$. O objetivo da técnica é recuperar as fontes S , através de X , sem nenhuma informação sobre as propriedades de A .

Sendo, $x_i = a_{i1}.s_1 + \dots + a_{im}.s_m$ para todo $i = 1, \dots, m$.

Define-se X, A e S como:

$$X = [x_1, x_2, \dots, x_m]^T \quad (2)$$

$$A = \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} \end{bmatrix} \quad (3)$$

$$S = [s_1, s_2, \dots, s_m] \quad (4)$$

O modelo representado acima descreve dados observados pelo processo de mistura das componentes independentes s_i , que não podem ser observados diretamente. Como a matriz A é quadrada, pode possuir inversa. O objetivo da ICA é, portanto, a partir apenas dos sinais capturados X , estimar os sinais individuais S e a matriz A .

Afim de estimar as componentes independentes e a matriz de características A , na equação (1), multiplica-se ambos os membros por $W = A^{-1}$, obtendo $Y = WX$, sendo Y a estimativa de S . Sendo X mais gaussiano que S , uma componente independente é estimada ao se achar uma matriz W , que projeta os elementos de X em uma distribuição de probabilidade não gaussiana (Linhares, Campos, & Santana, 2016). Com o objetivo de estabelecer a matriz A e as componentes independentes S , utilizou-se o algoritmo fastICA.

No algoritmo fastICA, a não gaussianidade é calculada fazendo uso de aproximações da negentropia (J), que são mais robustas do que as medidas de curtose e possuem menor custo computacional (Marchini, Heatom, & Ripley, 2004). Essa aproximação é dada pela equação (5).

$$J(y) = |E_y\{G(y) - E_v\{G(v)\}}|^p \quad (5)$$

Onde v é uma variável aleatória gaussiana normalizada, y é assumido normalizado e com variância unitária, e o expoente $p = 1, 2$, geralmente.

2.2. Seleção das características mais significativas

A etapa de seleção de características é de suma importância em um sistema de reconhecimento de padrões, pelo fato do grande número de variáveis encontradas nos problemas atuais. Objetivando uma melhor classificação de modo a reduzir o custo computacional, torna-se necessário algum tipo de redução de dimensionalidade do conjunto de características.

As características irrelevantes podem ser removidas sem comprometer o resultado da classificação, pois neste contexto, são consideradas redundantes, ou seja, implicam na presença de outra característica com a mesma funcionalidade, e não trazem nenhuma informação nova ao vetor de características (Araujo, 2014).

A redução de características consiste na escolha de um subconjunto das características mais informativas produzidas a partir dos sinais originais sem que se perca sua capacidade discriminante (Catarino, 2009), isto é, é realizada a escolha de um subconjunto ótimo de características, que representa a informação principal, contida nos dados, segundo algum critério. Esta seleção é importante em casos que a medição do conjunto de características é custoso, pois pode permitir que um subconjunto representativo, e, menor que o original, seja selecionado, melhorando a qualidade dos dados e os modelos construídos, Deste modo, tornando-se mais compreensível, e o processamento se torna mais rápido.

Para diminuir o esforço computacional, pode-se utilizar um critério, baseado em máxima Relevância e Mínima Redundância (mRMR), que maximiza a informação mútua e

minimiza a medida de redundância (Peng, Long, & Ding, 2005). Neste trabalho utilizou-se o critério mRMR para reduzir o conjunto de características.

As características do conjunto A podem ser selecionadas através da Máxima Relevância, que é dada pela informação mútua I entre a variável de classes c e cada característica v_i , de acordo com a equação (6),

$$\max D(A, c), D = \frac{1}{|A|} \sum_{v_i \in A} I(v_i; c) \quad (6)$$

muitas destas características selecionadas são redundantes, isto é, não trazem nenhuma informação nova para o problema em questão, logo as mesmas podem ser excluídas do conjunto de características sem comprometer os resultados. Para resolver o problema das características redundantes aplica-se juntamente com a Máxima Relevância, a condição de Mínima Redundância, que seleciona mutuamente apenas as características mutuamente exclusivas (Ding & Peng, 2005). A mínima redundância é dada pela equação (7):

$$\min R(A), R = \frac{1}{|A|^2} \sum_{v_i, v_j \in A} I(v_i, v_j) \quad (7)$$

De modo geral, os critérios descritos nas equações (6) e (7) são denominados conjuntamente de Máxima Relevância e Mínima Redundância (mRMR) (Peng, Long, & Ding, 2005), os quais fornecem, após um processo de otimização, as características mais relevantes e menos redundantes. O operador $\emptyset(D, R)$ que combina D e R , de modo a otimizá-los simultaneamente é representado pela equação (8).

$$\max \emptyset(D, R), \emptyset = D - R \quad (8)$$

2.3. Classificação

A Máquina de vetor suporte, do inglês Support Vector Machine (SVM), é uma técnica de aprendizado estatístico, baseada no princípio da Minimização do Risco Estrutural (SRM), e pode ser usada para resolver problemas de classificação (Cristianini & Shawe, 2000). Este método de aprendizado supervisionado é capaz de classificar, a partir de n indivíduos observados e pertencentes a diversos subgrupos, a que classe um indivíduo que deve ser classificado pertence (Bishop, 2006).

As SVM são consideradas sistemas de aprendizagem que utilizam um espaço de hipóteses de funções lineares em um espaço de muitas dimensões. Em casos em que o conjunto de amostras é composto por duas classes separáveis, um classificador SVM é capaz de encontrar um hiperplano com base em um conjunto de pontos, denominados vetores de suporte, o qual maximiza a margem de separação entre as classes. Mesmo quando as duas classes não são separáveis, a SVM é capaz de encontrar um hiperplano através do uso de conceitos pertencentes à teoria da otimização (VAPNIK, 1998).

Em algumas situações em que os elementos do conjunto de amostras não são linearmente separáveis, a SVM faz um mapeamento desses dados para um espaço de dimensão maior. E nesse novo espaço existe uma maior probabilidade de que eles sejam classificados por um hiperplano (Haykin, 2007). Para realizar a mudança do espaço de representação dos dados do conjunto a ser classificado utiliza-se as funções de kernel. Neste trabalho fez-se uso da função de kernel gaussiano (base radial), dada pela equação (9).

$$k(x_i, x_j) = e^{-\lambda|x_i-x_j|^2} \quad (9)$$

2.4. Validade de um teste de diagnóstico

Em processamento de sinais biomédicos e reconhecimento de padrões, a metodologia de desempenho usual é medida calculando algumas medidas estatísticas sobre o resultado dos testes (Bushberg & Boone, 2011). A validade de um teste dá-se em termos quantitativos ou qualitativos, e possui por finalidade diagnosticar um evento ou predizê-lo.

Existem quatro possíveis interpretações para o resultado: positivo na presença da doença, resultado verdadeiro positivo (VP), ou negativo na ausência da doença, resultado verdadeiro negativo (VN), ou negativo na presença da doença, resultado falso positivo (FP), e por fim negativo na presença da doença, resultado falso negativo (FN). Este último pode ser considerado um resultado péssimo, pois pode fazer com que o paciente não inicie o tratamento na fase inicial da patologia, o que prejudicaria seriamente o tratamento, levando a doença a um estágio mais avançado e algumas vezes ao óbito.

Sensibilidade, Especificidade e Acurácia são as medidas mais utilizadas para descrever um sistema de diagnóstico. Sendo a Sensibilidade (S) a capacidade de um teste detectar os indivíduos verdadeiramente positivo, isto é, diagnosticar corretamente os doentes. Especificidade (E), capacidade que o teste tem em detectar os verdadeiros negativos, ou seja, diagnosticar corretamente os indivíduos sadios. Acurácia (A), proporção de acertos, ou seja, o total de verdadeiros positivos e verdadeiros negativos em relação à amostra estudada.

Segundo (Guilhon, 2006), as equações para calcular a sensibilidade, especificidade e acurácia são respectivamente:

$$S = \frac{Vp}{Vp+Fn} \quad (10)$$

$$E = \frac{Vn}{Vn+Fp} \quad (11)$$

$$A = \frac{Vp+Vn}{Vp+Vn+Fp+Fn} \quad (12)$$

3. Resultados e Discussões

3.1. Base de dados e a espectrometria de massa

Os dados utilizados, referenciados em padrões proteômicos, para a aplicação do método proposto, foram obtidos por um espectrômetro de massas, que é baseado na tecnologia de dessorção e ionização no tempo (*Surface Enhanced Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry -SELD-TOF*). A base de dados é pública, gratuita e pode ser obtida em (Seldi-MS, 2002). A técnica SELDI-TOF, a qual, por meio de dados obtidos no espectro massas, permite detectar e identificar moléculas a partir de sua razão massa/carga (m/z).

A base de dados aqui utilizada contém várias amostras de arquivos no formato “csv” e cada amostra contém duas informações: relação entre a massa de um determinado íon e o número de cargas elementares que ele carrega (m/z) e intensidade do sinal do íon, estas informações estão separadas por vírgula e foram extraídas através de um espectrômetro de massas, onde posteriormente foi convertida em sinais multinível, através dos níveis de intensidade encontrados nas amostras.

Na presente base de dados foram utilizados 259 sinais proteômicos que estão divididos em: 190 casos sem a patologia, denominado como grupo de controle; e 69 casos com a patologia, denominado como grupo ativo.

Na figura 3, tem-se quatro sinais multinível, que foram escolhidos de forma aleatória da base de dados, e, através do software Matlab[®], desenhou-se os gráficos desses sinais, onde dois desses sinais são do grupo controle (sem câncer de próstata) e os outros dois do grupo ativo (com câncer de próstata). Tem-se ainda que o eixo horizontal corresponde à razão massa carga e o eixo vertical à intensidade do sinal.

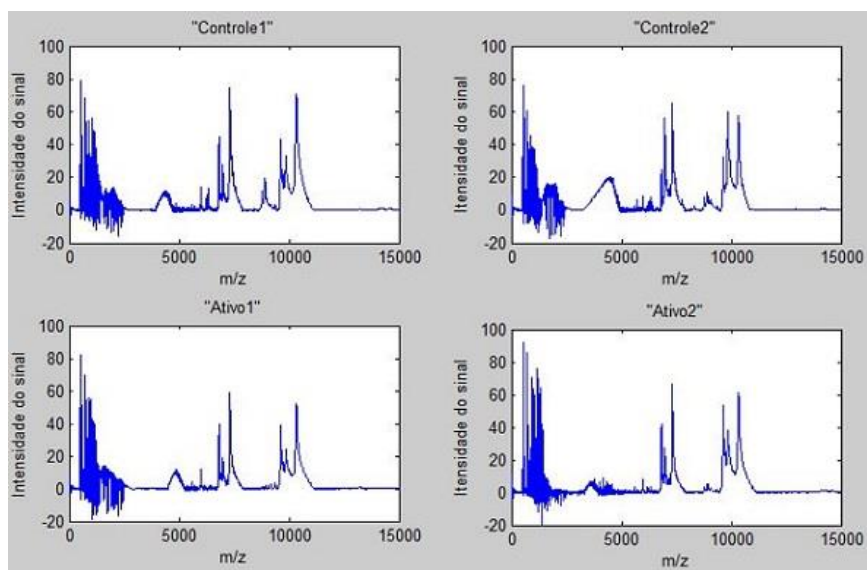


Figura 3 - Sinais extraídos da base de dados (Seldi-MS, 2002), dos casos ativo e controle.

3.2. Extração de características

Para se obter a matriz X do modelo ICA descrito na seção (2.1), uniu-se a matriz dos casos controle, dimensão 190×15154 , com a matriz dos casos ativo, dimensão 69×15154 , gerando, assim, a matriz X , dimensão 259×15154 .

Utilizou-se a matriz X como entrada para o algoritmo *FastICA*, objetivando as funções de base da matriz quadrada A de ordem 259×259 , que contém todas as características dos sinais proteômicos, onde cada linha representa uma amostra e cada coluna um nível de intensidade do sinal proteômico, isto é, as características dos vetores de cada sinal, que permitirá a futura classificação do paciente, com a presença ou ausência do câncer de próstata.

3.3. Seleção das características mais significativas e redução da dimensionalidade

Como já descrito, o objetivo desta etapa é reduzir o custo computacional e o tempo de processamento, selecionando as características que melhor representem o conjunto de dados após a etapa de extração. Para reduzir a dimensionalidade da matriz de características A , utilizou-se o algoritmo de Máxima Relevância e de Mínima Redundância.

Os testes realizados com o objetivo de reduzir o vetor de características ocorreram incrementando, de cinco em cinco, o número de características até o limite de 100, sendo que cada vetor gerado foi testado com o classificador da Máquina de Vetores de Suporte (SVM), com o objetivo de encontrar o vetor de melhor desempenho.

Percebeu-se que a partir da centésima característica não ocorria mais variações positivas para os resultados do classificador, e com isso, gerou-se uma nova matriz A_r , de dimensão 100×259 , com as características reorganizadas da mais relevante e menos redundante para a menos relevante e mais redundante.

3.4. Classificação e validação do método proposto

A etapa de classificação analisa cada vetor de característica, isto é, as linhas da matriz A_r , e após essa análise pode realizar-se a sua devida classificação: em controle ou ativo dos indivíduos.

Com o objetivo de aumentar a confiabilidade nos resultados do classificador, usou-se uma técnica estatística de validação cruzada *10-fold-cross validation*, na qual divide-se as 100 amostras igualmente em 10 grupos de 10, utilizando 9 grupos para treino e 1 para teste, esse procedimento foi repetido várias vezes, permutando os grupos até que todos fossem utilizados.

A média dos melhores resultados encontrados pela técnica de validação cruzada em conjunto com a SVM pode ser vista na tabela 1. Observa-se que os vetores com os melhores desempenhos foram os de 26, 27 e 30 características, durante a fase de classificação pela SVM.

Características	VP	FP	VN	FN	Acurácia(%)	Sensibilidade(%)	Especificidade(%)
25	48	8	182	21	88,81	80,24	94,56
26	50	6	184	19	88,80	84,28	93,76
27	52	7	183	17	89,21	83,68	95,08
28	49	9	181	20	88,81	83,63	95,00
29	48	9	181	21	88,45	80,43	95,04
30	49	7	183	20	89,62	81,99	95,30

Tabela 1 – Desempenho do classificador para os melhores resultados do método proposto.

Assim, considerando o vetor de 27 características, representado na linha 3 da tabela 1, conclui-se que das 190 amostras sem o câncer de próstata, 183 foram classificadas corretamente (Verdadeiro Negativo), errando apenas dezessete casos, dos quais os indivíduos eram do grupo controle, e foram classificados no grupo ativo; já nos 69 casos com o câncer de próstata, 52 foram classificadas corretamente quanto a presença da doença (Verdadeiros positivos), obtendo métricas para acurácia, sensibilidade e especificidade, respectivamente de 89,21%, 83,68% e 95,08%.

Também testou-se na base de dados [Seld-MS, 2002] somente o conjunto de técnicas ICA e SVM. Os resultados deste teste é apresentado na tabela 2.

Características	VP	FP	VN	FN	Acuracia(%)	Sensibilidade(%)	Especificidade(%)
50	46	3	187	23	89,98	73,32	98,82
63	47	6	184	22	89,20	76,78	96,72
71	45	4	186	24	89,22	75,14	97,46
75	48	4	186	21	90,35	74,09	97,73
82	45	4	186	24	89,23	74,65	99,04

Tabela 2 - Melhores resultados da aplicação do conjunto de técnicas ICA com a SVM

Considerando o vetor de 75 características, representado na linha 4 da tabela 2, as métricas de acurácia, sensibilidade e especificidade são respectivamente 90,35%, 74,09% e 97,73%. Este conjunto de técnicas possui um pior desempenho e maior custo computacional, pois o seu melhor resultado foi com um vetor de 75 características, e isto se justifica pelo fato das características não estarem organizada da mais relevante e menos redundante para a menos relevante e mais redundante.

4. Considerações finais

As técnicas computacionais aqui empregadas mostraram-se bastante eficazes, como a Análise de Componentes Independentes (ICA), na fase de extração das características dos sinais extraídos do espectro de massas pela técnica SELDI-TOF; assim como a técnica de seleção de atributos Máxima Relevância e Mínima Redundância (mRMR) que selecionou as melhores características e mostrou que a redução não afetou de forma negativa os resultados, e ainda diminuiu o custo computacional, permitindo um bom desempenho do

classificador Máquina de Vetores de Suporte (SVM), que foi capaz de classificar com certa eficiência os pacientes nos grupos ativo ou controle.

O exame clínico mais utilizado é o toque retal, apesar de suas limitações, uma vez que somente as porções posterior e lateral da próstata podem ser palpadas, deixando de 40% a 50% dos tumores fora do seu alcance. As estimativas de sensibilidade variam entre 55% e 68%. O valor preditivo positivo é estimado entre 25% e 28%. Quando utilizado em associação à dosagem do PSA com valores entre 1,5 ng/ml e 2,0 ng/ml, sua sensibilidade pode chegar a 95% (INCA, 2015).

Os resultados aqui apresentados demonstram que o método proposto alcançou um bom desempenho. A capacidade de generalização da SVM em conjunto com o ICA e mRMR foi de 96,31% de acerto, com relação ao grupo controle, e 75,36% de acertos com relação a classificação no grupo ativo. A taxa geral de acerto deste classificador foi de 90,73%.

Porém, deve-se levar em consideração o tabu ainda existente entre os homens e o teste do toque retal, cercado, ainda no século XXI, de muitos preconceitos, levando a comunidade masculina ao adiamento deste teste, e até mesmo a sua não realização, quando ocorre a suspeita da presença dessa doença, o que pode gerar complicações futuras para o tratamento, prejudicando na qualidade de vida do paciente e dos seus familiares.

Assim, os resultados apresentados neste trabalho nos encorajam a continuar as pesquisas com uma perspectiva de que os próximos resultados sejam ainda melhores do que os atuais e novas aplicações em outras bases de dados possam ser realizadas para se obter um melhor modelo do problema e posteriormente o desenvolvimento de um software que possa auxiliar os profissionais da saúde em um diagnóstico precoce e eficaz, diminuindo com isso, os impactos negativos causados pelo câncer de próstata e até mesmo outras doenças.

Referências

AAPO (2016). “Independent Component Analysis (ICA) and Blind Source Separation (BSS). Disponível em: <http://research.ics.aalto.fi/ica/fastica/>. [2 de setembro de 2015].

Araujo W. B. D. “Método de detecção de câncer de ovário utilizando análise de componentes independentes, algoritmo de máxima relevância e mínima redundância e máquina de vetores de suporte”. Dissertação (Mestrado em Engenharia de Computação e Sistemas) – Universidade Estadual do Maranhão, São Luís, 2014.

Araujo W.B.D., Campos L.F.A, Aline S.F. (2014), “Método de Detecção de Câncer de Ovário Utilizando Padrões Proteômicos, Análise de Componentes Independentes e Máquina de Vetores de Suporte”. XIV WORKSHOP DE INFORMÁTICA MÉDICA, 14. Anais do congresso da sociedade brasileira de computação. Brasília: CSBC. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/wim/2014/011.pdf>. [20 de abril de 2016].

Bishop, C. M (2006). “Pattern Recognition and Machine Learning”. New York: Springer.

Bushberg J. T and Boone J. M. (2011). “The essential physics of medical imaging. Lippincott Williams & Wilkins.

Catarino F.M.I.F. “Segmentação da íris em imagens com ruído”. (Dissertação de mestrado) – Universidade da Beira Interior, Covilhã, 2009.

Cristianini N. and Shawe-Taylor J. (2000). “ An Introduction to support vector machines and other kernel-based learning methods”. Cambridge university press.

Ding C., Peng H. “Minimum redundancy feature selection from microarray gene expression data”. Journal of Bioinformatics and Computational Biology, Imperial College Press, volume 3, nº2, pp. 185-205. ISSN 1757-6334. Disponível em: http://ranger.uta.edu/~chqing/papers/mRMR_JBCB.pdf. [29 de março de 2016].

Haykin S. (2007). “Redes neurais: princípios e prática”. Bookman, Porto Alegre.

INCA (2015). “Estimativa 2016: incidência de câncer no Brasil”/Instituto Nacional do Câncer. José Alencar Gomes da Silva, Rio de Janeiro. INCA (2016). Instituto Nacional do câncer, <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/prostata>. [15 de abril de 2016].

Marchini J.L., Heaton C. and Ripley B.D. (2004) “FastICA algorithms to perform ICA and Projection Pursuit”. Disponível em: <http://www.stats.ox.ac.uk/~marchini/software.html>. [22 de Março de 2016].

Matworks (2015). “Minimum-redundancy maximum-relevance feature selection.” Disponível em: <http://www.mathworks.com/matlabcentral/fileexchange/>. [27 de nov. De 2015].

Nascimento J.L. et al. (2016), “Método Computacional para o Diagnóstico Precoce da Granulomatose de Wegener”, Revista de Informática Teórica e Aplicada, volume 23, nº1, pp. 277-292.

Peng H., Long F., Ding C. (2005) “Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy”. IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 27, August.

Ribeiro, A.C. et al. (2015), “Diabetes classification using a redundancy reduction preprocessor”. Research on Biomedical Engineering, volume 31, nº2, pp. 97-106. ISSN 2446-4740. Disponível em: http://www.scielo.br/scielo.php?pid=S2446-47402015000200097&script=sci_arttext. [5 de julho de 2016].

Seldi-MS (2002). “Base de Dados”. Disponível em: <http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>. [25 de nov. de 2015].

Vapnik V. N. (1998). “Statistical Learning Theory”. Wiley, New York.



Elzenir Montes é licenciado em Matemática, e especialista em Estatística pela Universidade Estadual do Maranhão, encontrando-se atualmente na fase final do mestrado em Engenharia da Computação e Sistemas. Foi professor substituto na Universidade Estadual do Maranhão (2014 – 2016). Áreas de interesse: Processamento de sinais; Análise de Componentes Independentes.



Lúcio Campos é doutor em Biotecnologia pela RENORBIO. Possui Mestrado em Engenharia Elétrica pela UFMA (2006). Atualmente é Professor da UEMA, atuando no curso de Engenharia da Computação, no Mestrado Profissional em Engenharia da Computação e Sistemas e coordenando o curso técnico à distância de Redes de Computadores. Tem experiência na área de Engenharia Elétrica, com ênfase em Telecomunicações e Processamento de Sinais Biológicos, atuando principalmente nos seguintes temas: Processamento de Imagens Médicas, Análise de Componentes Independentes, Redes Neurais Artificiais.



Wesley Araujo é Mestre em Engenharia da Computação e Sistemas - UEMA (2014), Pós-graduado em Planejamento e Implementação de Redes LAN e WAN pelo Instituto de Ensino Superior de Londrina(2008). Graduação em Ciência da Computação pela Universidade Federal do Maranhão (2005). Certificação CCNA (2008-2011) - Cisco Systems - CSCO11300367. Tem experiência na área de Ciência da Computação, com ênfase em Redes de Computadores e Gerência de Projetos e Sistemas.



Ewaldo Santana é Graduado em matemática pela UFMA (1998), mestrado em Engenharia de Eletricidade pela UFMA (2006), doutorado em Engenharia Elétrica pela Unicamp (PB) (2009). Atualmente é Professor do mestrado em Engenharia da Computação e Sistema UEMA. Áreas de Interesse: Processamento de Sinais; Filtragem Adaptativa; Análise de Componentes Independentes; Neurociência visual.



Claudyane Araujo é licenciada em matemática pela Universidade Federal do Maranhão (2014). E atualmente cursa Engenharia da Computação na Universidade Estadual do Maranhão.

(esta página par está propositadamente em branco)