

Método Computacional para Detecção da Cardiotoxicidade Utilizando Padrões Proteômicos, Análise de Componentes Independentes e Máquina de Vetores de Suporte

Jardiel Nunes Almeida, Lúcio Flávio A. Campos, José do Nascimento Linhares, Flávia Larisse da Silva Fernandes

Resumo—Resumo: Em virtude da importância da detecção precoce da Cardiotoxicidade, vários esquemas de diagnóstico auxiliados por computador estão sendo propostos com o objetivo de ajudar na identificação desta enfermidade. Isso porque o sucesso do tratamento no combate desta disfunção cardíaca depende de um diagnóstico rápido, pois quanto mais cedo iniciarse o tratamento, maiores serão as chances de cura. Propomos um método de Diagnóstico Auxiliado por Computador (CAD) para diagnosticar pacientes com Cardiotoxicidade, utilizando Análise de Componentes Independentes para extrair características de um sinal proteômico, depois fazendo uso da técnica de Máxima Relevância e Mínima Redundância para reduzir a dimensionalidade e com isso o custo computacional, e por fim a aplicação da Máquina de Vetores de Suporte para classificar as amostras entre presença ou ausência de Cardiotoxicidade. O método foi testado com a base de dados de padrões proteômicos SELDI-TOF, cujo melhor desempenho obtido foi com um vetor de 20 características, resultando em uma acurácia de 88,718%, com 85% de especificidade e 97,26% de sensibilidade

Palavras-Chave—Análise de Componentes Independentes, Cardiotoxicidade, Máxima Relevância e Mínima Redundância, Máquina de Vetores Suporte, Padrões Proteômico.

Abstract—Because of the importance of early detection of cardiotoxicity, several schemes computer aided diagnosis has been being proposed in order to help the identification of this disease. That's because the successful treatment to combat this cardiac dysfunction depends on rapid diagnosis because the earlier start the treatment, the greater the chances of healing. We propose a method of Diagnosis aided by computer (CAD) to diagnose patients with Cardiotoxicity, using Independent Component Analysis to extract characteristics of a proteomic signal. Then we use the technique of Maximum Relevance and Minimum Redundancy to reduce the dimensionality and thus the computational cost. And lastly the application of Support Vector Machine to classify the samples between the presence or absence of cardiotoxicity whose best performance was obtained with a vector of 20 features resulting in an accuracy of 88.718, with 85 of specificity and 97.26 of sensitivity.

Keywords—Independent Component Analysis, cardiotoxicity, Maximum and Minimum Redundancy Relevance, Support Vector Machine.

I. INTRODUÇÃO

Ao longo dos últimos anos, diferentes tratamentos para diversos tipos de câncer foram largamente desenvolvidos,

Jardiel Nunes Almeida, Departamento de Engenharia da Computação, Universidade Estadual do Maranhão, São Luis-MA, Brasil, E-mail: jardielguaiba@gmail.com .

levando a cura subsequente desta doença em alguns pacientes ou ao evidente aumento da sobrevida e qualidade de vida dos mesmos [1]. Porém, vários estudos comprovam que diferentes agentes antineoplásicos (antimetabólitos, antraciclina e agentes biológicos, hormonais, alquilantes e antimicrotúbulos) utilizados no tratamento, têm potencial cardiotoxíco[2]. Assim, eles podem causar o surgimento de uma nova enfermidade, conhecida como Cardiotoxicidade. Neste sentido, vários critérios de detecção e protocolos têm sido propostos para o tratamento e prevenção da mesma[2].

Sobre esta questão, vale ressaltar que a Cardiotoxicidade é definida pela situação na qual agentes externos (químicos ou físicos) interferem negativamente no coração, determinando alterações estruturais, elétricas e funcionais no miocárdio [3]. Este órgão torna-se mais fraco e não é tão eficiente em bombeamento, o que compromete a circulação do sangue [3]. Esta enfermidade pode ser causada por tratamentos de quimioterapia, complicações decorrentes da anorexia nervosa, efeitos adversos da ingestão de metais pesados, ou um medicamento administrado incorretamente como a bupivacaína [3].

Assim, torna-se necessária a prevenção desta patologia, que é realizada através de uma avaliação inicial dos pacientes oncológicos submetidos a quimioterapia cardiotoxíca e esta avaliação tem como objetivos: excluir pacientes com evidências clínicas, laboratorial e radiológica de insuficiência cardíaca congestiva (IC) antes do início do tratamento quimioterápico, identificar pacientes com redução da fração de ejeção, associada a sintomas ou não, durante a quimioterapia [4]. E para fazer esta avaliação utiliza-se exames cardíacos, tais como: Eletrocardiografia, Eco-Dopplercardiografia, Cintilografia com Radionuclídeo, Teste Ergométrico, Biópsia Endomiocárdica, troponina T cardíaca, entre outros, menos utilizados.

O sucesso do tratamento no combate desta disfunção de eletro fisiologia do coração depende de um diagnóstico rápido, pois quanto mais cedo iniciar-se a assistência médica, maiores serão as chances de cura. Em virtude da importância da detecção precoce da ação cardiotoxíca provocada por essas drogas, vários esquemas de diagnóstico auxiliados por computador estão sendo propostos com o objetivo de ajudar na identificação precoce desta enfermidade. Vale ressaltar que não existem métodos CAD associados a essa Patologia.

Neste trabalho propõe-se um método de Diagnóstico Aux-

iliado por Computador (CAD) para ajudar no reconhecimento precoce da Cardiotoxicidade, utilizando dados ou sinais proteômicos. E para realizar a extração de característica destes sinais proteômicos será utilizada a técnica de Análise de Componentes Independentes (Independent Component Analysis-ICA), somada ao Algoritmo de Máxima Relevância Mínima Redundância (mRMR), para selecionar as características mais significantes e reduzir a dimensionalidade da matriz gerada. Após a seleção das características mais relevantes, estas serão classificadas utilizando-se a Máquina de Vetores Suporte (Support Vector Machine - SMV) entre duas classes, sendo que identificará o paciente com ou sem Cardiotoxicidade.

II. METODOLOGIA PROPOSTA

A. Método Proposto

O método proposto é descrito pelo diagrama em blocos mostrado na Figura 1. Este consiste em: extrair as características significantes do sinal proteômico utilizando a Análise de Componentes Independentes (ICA), reduzir a dimensionalidade da matriz de característica gerada com o algoritmo de Máxima Relevância e Mínima Redundância (mRMR) e por fim, fazer a classificação com a de Máquina de Vetores de Suporte (SVM).

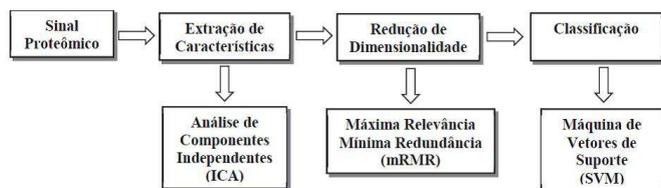


Fig. 1. Diagrama do método.

B. Espectrometria de massa e Sinal proteômico

Nos últimos anos diversos tipos de marcadores moleculares que auxiliam no diagnóstico precoce e no tratamento de várias doenças humanas, incluindo a Cardiotoxicidade como exposto em [5], vêm sendo analisados.

Para fazer a análise destes marcadores são utilizados conceitos como o da Espectrometria de Massa, uma técnica analítica utilizada para identificar compostos desconhecidos, modificar materiais conhecidos e elucidar as propriedades químicas e estruturais das moléculas. Nesta técnica, um composto é ionizado através de um método de ionização, os íons são separados na razão massa carga por meio de um método de separação, e o número de íons correspondentes a cada unidade de razão massa carga são registrados na forma de um espectro de massa. Para esse fim é necessário um espectrômetro de massa, um analisador que permite a determinação qualitativa e quantitativa dos compostos de uma amostra [6].

Neste sentido a Proteômica é entendida como sendo a análise em larga escala de um conjunto de proteínas, ou seja, a análise da expressão gênica de determinada célula, tecido ou organismo, sob determinadas condições ambientais, ou estágio de desenvolvimento que são responsáveis direta ou

indiretamente pelo controle de todos ou quase todos os processos biológicos. Isto permite a identificação e caracterização de marcadores biológicos, ou seja, moléculas endógenas ou exógenas específicas de um determinado estado patológico[5]. E nesta perspectiva gerar listas de proteínas que aumentam ou diminuem em expressão como causa ou consequência de patologia [5]. A natureza desta informação pode nos levar a causa ou a uma consequência, de processos de doenças e de toxicidade. Além do mais, o recente progresso de metodologias nessa área tem aberto novas oportunidades para obtenção de informações relevantes sobre processos normais e anormais que ocorrem no organismo humano [5].

Os dados usados neste trabalho foram baseados em padrões proteômicos usando a técnica *SELDI-TOF*, que se mostrou um padrão de informação preciso para auxiliar no diagnóstico de pacientes com Cardiotoxicidade [5]. As amostras utilizadas no trabalho foram adquiridas em [7].

A Figura 2 ilustra a amostra que foi extraída através de um espectrômetro de massa e foi posteriormente convertida em um sinal multinível através dos níveis de intensidade proteômicos encontrados no espectro de massa.

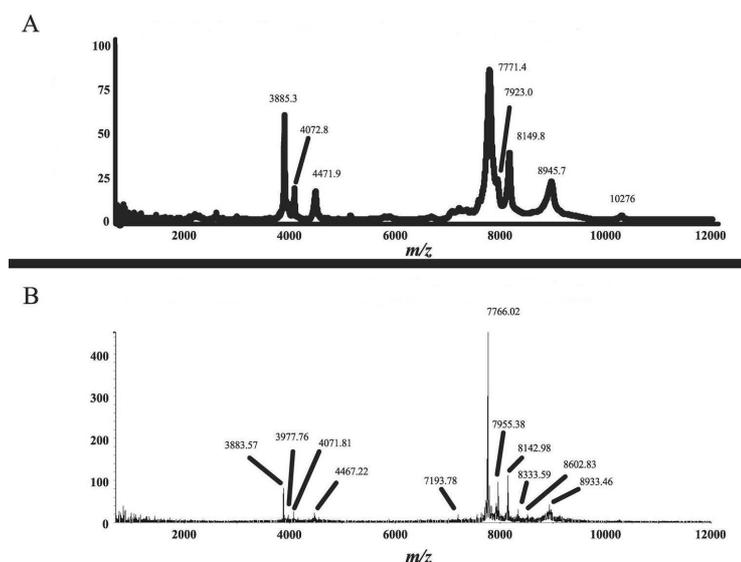


Fig. 2. Espectro de Massa. Comparação entre baixa resolução e alta resolução de um espectro de massa SELDI-TOF. Sendo, a figura 2.A corresponde a um espectro de massa de alta resolução e a figura 2.B a um de baixa resolução. O eixo vertical corresponde ao nível de intensidade do espectro de massa, enquanto o eixo horizontal corresponde à razão massa / carga. Fonte: [5]

C. Análise de Componentes Independentes

A análise de componentes independentes (ICA) é uma técnica estatística e computacional capaz de revelar componentes desconhecidos a um conjunto de variáveis aleatórias, medições, ou sinais observados multivariados [10]. Por isso ela é utilizada em: processamento de sinais Biomédicos, Telecomunicação e processamento de imagem. este trabalho é baseado em processamento de sinais Biomédicos, que justifica a utilização da ICA. No modelo ICA, considera-se que dado sinal observado x_i pode ser representado como uma combinação linear de n variáveis aleatórias s_i que são estatisticamente independentes e não-gaussianas [8].

Desta forma pode-se escrever cada sinal x_i como:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n, \forall i, j = 1, 2, \dots, n. \quad (1)$$

Onde os a_{ij} são os coeficientes de mistura (característica) e os s_i são os componentes independentes.

Este modelo pode ser expresso na forma matricial como:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (2)$$

Sendo \mathbf{A} a matriz de características e \mathbf{S} as componentes independentes.

Verifica-se a partir da equação 2 que o problema da Análise de Componentes Independentes consiste em encontrar \mathbf{A} e \mathbf{S} conhecendo-se apenas \mathbf{X} . A solução deste problema pode ser obtida explorando-se a propriedade de independência ou não gaussianidade das componentes independentes [10]. O algoritmo *fastICA*, proposta por *Hyvärine*[10], é utilizado para estas matrizes. Este algoritmo tem rápida convergência, e, se comparado com algoritmos baseados em gradiente, é mais simples, pois não necessita de ajuste no passo de adaptação [10]. O *fastICA* usa como medida de não gaussianidade uma versão aproximada da negentropia dada pela equação 3

$$J(y) \propto [E(G(y)) - E(G(y_{gaus}))]^2. \quad (3)$$

Sendo os k_i constantes positivas, E é o operador esperança, y_{gaus} variáveis gaussianas com variância unitária e média zero e os G_i são funções não quadráticas. Segundo [10], as funções G_1 e G_2 , representadas nas equações 4 e 5, garantem boas aproximações da negentropia e melhoram a convergência do algoritmo *fastICA*.

$$G_1(y) = \frac{1}{\beta} \log(\cosh(\beta y)), \text{ com } 1 \leq \beta \leq 2 \quad (4)$$

$$G_2(y) = -\exp\left(-\frac{y^2}{2}\right). \quad (5)$$

D. Seleção das características mais significantes

Após a determinação da matriz de características \mathbf{A} , faz-se necessário reduzi-la selecionando as características mais discriminantes para melhorar a performance do classificador. Além disso, o uso de muitas características podem aumentar o erro de classificação e o custo computacional.

Neste trabalho a redução de características foi realizada através do algoritmo de Máxima Relevância e Mínima Redundância, que seleciona as características mais relevantes, através da equação 6, e retira as redundantes através da equação 7.

$$\max M(\mathbf{a}, c), R = \frac{1}{|\mathbf{a}|} \sum_{a_i \in \mathbf{A}} I(a_i, c) \quad (6)$$

$$\min R(\mathbf{a}), R = \frac{1}{|\mathbf{a}|^2} \sum_{a_i a_j \in v} I(a_i, a_j) \quad (7)$$

Sendo \mathbf{a} um vetor de características, c o vetor de classe, a_i e a_j duas características individuais e I a informação mútua. A informação mútua mede quanta informação uma variável aleatória (VA) possui sobre outra.[11]

O algoritmo mRMR combina M e R (equação 8) para obter simultaneamente as características mais relevantes e menos redundantes.

$$\max \Phi(M, R), \Phi = M - R, \quad (8)$$

Com o vetor de características reduzido pela técnica de Máxima Relevância e Mínima Redundância, pode então ser feita a classificação das amostras. O que será mostrado na próxima Seção.

E. Classificação

Na classificação das amostras utilizou-se a Máquina de Vetores de Suporte. A classificação foi realizada a partir da análise do vetor de características já reduzido através da técnica mRMR, onde as amostras foram rotuladas em negativo (grupo controle) ou positivo (com Cardiotoxicidade).

No entanto, para mensurar a aprendizagem da Máquina e assim aumentar a confiabilidade dos resultados, utilizou-se o *Cross-Validation* (validação cruzada) [14], uma técnica de partilhamento de amostragem randômica, utilizada para estimar com maior precisão a acurácia (probabilidade de classificação correta de uma instancia selecionada estatisticamente) de um classificador. Existem basicamente três métodos distintos de validação cruzada, são eles: *Holdout*, *K-fold* e *Leave-one-out* [14]. Neste trabalho o método utilizado foi *K-fold*. Onde o conjunto de dados (exemplos) é aleatoriamente dividido em k partições mutuamente exclusivas (*folds*) e de tamanho aproximadamente igual a $\frac{n}{k}$ dados. As $(k - 1)$ folds são utilizadas para para treinamento e o fold restante para testes. Este processo é repetido k vezes, e a cada vez é considerado um *fold* diferente para teste.

1) *Máquinas de Vetores de Suporte*: A Máquina de Vetores de Suporte (SVM) é um método de aprendizagem supervisionada, capaz de classificar a partir de n indivíduos observados pertencentes a diversos subgrupos, a que classe um indivíduo que deve ser classificado pertence [15].

As Máquinas de Vetores Suporte (SVMs), são algoritmos de aprendizagem bastante utilizados na área de aprendizagem de máquina. Elas constituem uma técnica embasada na Teoria de Aprendizado Estatístico [15] que vem recebendo grande atenção nos últimos anos [16]. As SVMs têm algumas características que tornam seu uso muito atrativo, tais como: Boa capacidade de generalização, robustez em grandes dimensões, Convexidade da função objetivo e Teoria bem definida. Além do mais a SVM pode ser utilizada para fazer a classificação e trabalha bem em espaço de alta dimensionalidade, atuando em problema de duas classes e assim podendo fazer a identificação dos pacientes com Cardiotoxicidade e os sem Cardiotoxicidade.

F. Métricas e Desempenho

Em processamento de sinais biomédicos e reconhecimento de padrões, a metodologia de desempenho usual é avaliada calculando-se algumas medidas estatísticas sobre o resultado dos testes [18]. Neste trabalho os resultados da classificação a partir da realização de testes são divididos em: Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN).

Sendo VP e VN números de amostras que são corretamente identificadas como positivas e negativas, respectivamente pelo classificador. FP e FN representam o número de amostras correspondentes aos pacientes que são diagnosticados erroneamente como positivo (portador da doença) ou negativo (não portador da doença), respectivamente. Estes números são utilizados para gerar medidas capazes de quantificar o desempenho da metodologia, para avaliar a eficiência do nosso método e se os objetivos foram alcançados. As medidas de desempenho aqui utilizadas são: Acurácia, Especificidade, Sensibilidade.

III. RESULTADOS E DISCUSSÕES

A. Aquisição de Dados

A base de dados usada neste trabalho é composta por 62 sinais proteômicos divididos em: 28 amostras com diagnóstico positivo para a Cardiotoxicidade e 34 amostras com diagnóstico normal (controle). Esses sinais proteômicos foram obtidos a partir da técnica *SELDI-TOF*, que se mostrou ser um padrão de informação preciso para auxiliar no diagnóstico de pacientes com Cardiotoxicidade [5]. As amostras utilizadas no trabalho foram adquiridas em [7]. Cada amostra da base possui 373257 níveis de intensidade diferentes, no entanto para melhorar os resultados, reduzimos cada amostra para 100000 níveis de intensidade diferentes como ilustrado na figura 3

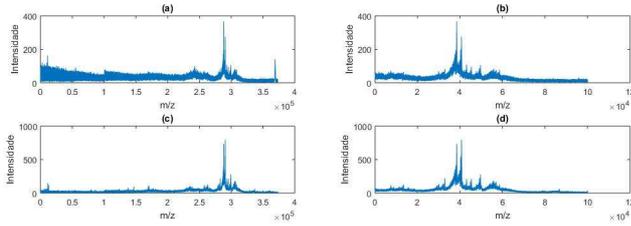


Fig. 3. Espectro de Massa. A figura (3.a) representa um sinal proteômico original de um paciente portador da Cardiotoxicidade contendo 373257 níveis de intensidade e a figura (3.b) mostra o sinal representado na figura (3.a) já reduzido para 100000 níveis de intensidades. Enquanto, a figura (3.c) ilustra um sinal proteômico de um paciente não portador da Cardiotoxicidade no tamanho original e a figura (3.d) mostra este sinal reduzido para 100000 níveis de intensidades.

B. Extração de Características

O objetivo desta etapa foi obter parâmetros a partir dos sinais proteômicos para discriminar os dois grupos (portador e não portador de Cardiotoxicidade) de informações capazes de identificar a Cardiotoxicidade. As características selecionadas devem garantir que as amostras de um paciente sejam corretamente classificadas como portadora ou não da Cardiotoxicidade.

Para extrair as características que distinguem os casos com Cardiotoxicidade dos casos do grupo controle foi utilizada a Análise de Componentes Independentes (ICA), e para esse fim criou-se a matriz \mathbf{X} de ordem 62 por 100000 que é a junção das matrizes com amostras de casos com Cardiotoxicidade e grupo controle. A matriz \mathbf{X} compõe o modelo ICA.

O algoritmo *FastICA* usou a matriz \mathbf{X} para obter uma outra matriz que contém as características de cada uma das amostras,

e esta matriz é denominada matriz \mathbf{A} de ordem 62 por 62. Sendo que cada linha desta matriz \mathbf{A} corresponde a uma amostra, e cada coluna corresponde a uma característica. Desta maneira o classificador terá um parâmetro para distinguir pacientes com Cardiotoxicidade daqueles que não possuem a doença.

C. Seleção das Características mais Significantes

Nesta etapa foram obtidos os parâmetros que melhor representam as informações geradas a partir da extração de características. Pois, caso todos os parâmetros obtidos pela extração de características sirvam de entrada para o classificador, poderíamos ter resultados insatisfatórios, com baixa acurácia e grande esforço computacional. E para selecionar as características que melhor represente o banco de dados, foram realizados testes para a redução do vetor de características de cada amostra incrementando, de cinco em cinco, o número de características selecionadas através da técnica de Máxima Relevância e Mínima Redundância (mRMR) até 62, sendo que para encontrar o vetor de melhor desempenho, cada vetor gerado foi testado com a Máquina de Vetores de Suporte (SVM).

D. Métricas de Avaliação

Para entendermos melhor o que significa especificidade, sensibilidade e acurácia, pois são utilizadas neste trabalho, vamos definir estas variáveis que nos auxiliarão para melhor compreensão dos resultados obtidos.

Sendo Acurácia (A) a taxa de acerto do classificador durante a fase de teste, definida por:

$$A = (VP + VN)/(VP + VN + FP + FN) \quad (9)$$

A Especificidade (E) é a proporção de verdadeiros negativos que são corretamente classificados pelo teste, definida por:

$$E = VN/(VN + FP) \quad (10)$$

A Sensibilidade (S) é a proporção de verdadeiros positivos que são corretamente classificados pelo teste, definida por:

$$S = VP/(VP + FN) \quad (11)$$

E. Classificação

Na última etapa, foi utilizada a SVM como classificador das amostras dos pacientes em controle e pacientes com Cardiotoxicidade.

Utilizou-se para isto um classificador que tem núcleo baseado em RBF (*Radial-Basis Function*), com a configuração padrão dos parâmetros, sem otimização dos mesmos. As amostras estão contidas em um conjunto apenas, com o objetivo de realizar os testes de validação cruzada *10-fold cross-validation*.

Foram realizados vários testes para verificação da eficácia dos resultados demonstrados na tabela 1, contendo: especificidade, sensibilidade e acurácia. Os melhores resultados obtidos através do método *10-fold cross-validation* foram para os

vetores com 14, 16, 20, 53 e 49 características, eles tiveram bom desempenho durante o período de testes do classificador. Baseado nos resultados das tabelas, verifica-se que com 20 características das 62 possíveis, o método obteve 88,718% de acurácia, 85,000% de especificidade e 97,260% de sensibilidade.

TABELA I
RESULTADOS

Carct	VP	FP	VN	FN	Esp(E)(%)	Sen(S)(%)	Acu(A)(%)
14	32	11	17	2	83,333	97,521	87,566
16	33	10	18	1	86,905	95,426	83,412
20	33	10	18	1	85,000	97,260	88,718
49	33	11	17	1	80,000	88,897	82,057
53	33	11	17	1	82,143	86,943	80,597

Considerando o vetor de 20 características, observou-se também, que das 34 amostras positivas para Cardiotoxicidade, 33 foram classificadas corretamente (VP), logo dos 34 caso de Cardiotoxicidade 1 foram classificados como normal (FN). Dos 28 casos com diagnóstico normal, somente em 10 casos (FP) houve erro de classificação, diagnosticando-os positivo para a Cardiotoxicidade.

IV. CONCLUSÃO

Este artigo propõe um método computacional para fazer o diagnóstico precoce da Cardiotoxicidade, através da classificação de padrões proteômicos, utilizando: Análise de Componentes Independentes (ICA), Algoritmo de Máxima Relevância e Mínima Redundância (mRMR), e Máquina de Vetores de Suporte.

Os resultados encontrados demonstraram que o conjunto de técnicas aplicadas é eficiente para diagnosticar a Cardiotoxicidade, pois conseguiu identificar os indivíduos normais e os portadores desta enfermidade. Estes resultados são vistos nas métricas de desempenho encontradas: 88,718% acurácia, 85% de especificidade e 97,26% de sensibilidade, em um estudo que utilizou 62 amostras com baixa resolução. Das 62 amostras utilizadas, 51 foram classificadas corretamente (VP + VN) e 11 classificadas incorretamente (FP+FN). Apenas uma amostra foi classificada como negativa sendo positiva.

Para confirmar a eficácia do método, novos testes devem ser realizados em bases de dados maiores.

REFERÊNCIAS

[1] MORAES, A. d. J. P. Viabilidade do treinamento Físico Aeróbio por pacientes com Câncer Hematológico antes do transplante Autólogo de Células-Tronco Hematopoéticas. Tese (Mestrado em Ciências do Movimento Humano)? Universidade de Estado de Santa Catarina, Florianópolis, 2014.

[2] Sá, M. P. B. de O. et al. Cardiotoxicidade e Quimioterapia. 2009. Disponível em: <http://files.bvs.br/upload/S/1679-1010/2009/v7n5/a010.pdf>. Acesso em: 3 set 2014.

[3] I Diretriz Brasileira de Cardi o-Oncologia Pediatrica da Sociedade Brasileira de Cardiologia. Disponível em:

[4] R, A. et al. Cardiovascular toxicity caused by cancer treatment: strategies for early detection. 2009. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/19341970>. Acesso em: 2 abr. 2009.

[5] PETRICOIN, E. F. P. et al. Toxicoproteomics: Serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. Toxicological Pathology, Clinical Proteomics Program Website, v. 32, n. 1, p. 179, 2011. ISSN 0192-6233 print / 1533-160. Disponível em: <http://home.ccr.cancer.gov/ncifdaproteomics/pdf/ToxPath.pdf>.

[6] BARBOSA, E. B. et al. Proteômica: Metodologias e aplicações no estudo de doenças humanas. Revista da Associação Médica Brasileira, Revista da Associação Médica Brasileira, v. 58, n. 3, p. 366?375, 2012. ISSN 0104-4230. Disponível em: <http://dx.doi.org/10.1590/S0104-42302012000300019>.

[7] PROGRAM, C. P. Toxicoproteomic analysis of anthracycline-induced cardiotoxicity. 2004. Disponível em: <http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>. Acesso em: 2 maio 2014.

[8] SILVA, A. P. O. d. Uma Implementação da Análise de Componentes Independentes em Plataforma de Hardware Reconfigurável. Tese ((Mestrado em Automação e Sistemas; Engenharia de Computação; Telecomunicações)?Universidade Federal do Rio Grande do norte, Natal, 2011.

[9] PROGRAM, C. P. Análise comparativa das abordagens de estimativa do modelo FestICA por maximização da negentropia verossimilhança. 2012. Disponível em: https://www.ime.usp.br/arquivos/4congresso/4120Henrique20Morimitsu_N.pdf. Acesso em: 3 set 2014.

[10] HYVARINEN J. KARHUNEN, E. O. A. (Ed.). Independent Component Analysis. USA: Pearson Prentice Hall, 2001.

[11] T. M. Cover and J. A Thomas, Elements of Information Theory. New York, USA: Wiley, 1991

[12] ARAUJO, W. B. D. Método de Detecção de Câncer de Ovário utilizando Padrões Proteômicos, Análise de Componentes Independentes e Máquina de Vetores de Suporte. Tese (Mestrado em Engenharia da Computação e Sistema)? Universidade Estadual do Maranhão, São Luis, 2014.

[13] DING, C.; PENG, H. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology, Imperial College Press, v. 3, n. 2, p. 185? 205, 2005. ISSN 1757-6334. Disponível em: http://penglab.janelia.org/papersall/docpdf/2004JBCB_feasel-04-06-15.pdf.

[14] KOHAVI, R. A. A study of cross-validation and bootstrap for accuracy estimation and model selection. Computational and Applied Mathematics, Digital Library, v. 2, n. 3, p. 1137 a 1143, 1995. ISSN 0101 – 205. Disponível em: <http://dl.acm.org/citation>.

[15] CHAPELLE, o.; VAPNIK, V. Model selection for support vector machines. In: SOLLA, S. A.; LEEN, T. K.; MÜLLER, K.-R. (Ed.). Advances in Neural Information Processing Systems 12. Cambridge, Mass: MIT Press, 2000. Disponível em: <http://www.ens-lyon.fr/~ochapell/ms/nips99.ps>.

[16] HEARST, M. A. et al. Trends and controversies - support vector machines. IEEE Intelligent Systems, v. 13, n. 4, p. 1828, 1998. Disponível em: <http://computer.org/intelligent/ex1998/pdf/x4018.pdf>.

[17] HAYKIN, S. O. (Ed.). Neural Networks A Comprehensive Foundation. USA: Pearson Prentice Hall, 1998.

[18] BUSCHBERG, J. T. et al. (Ed.). The Essential Physics of Medical Imaging. Philadelphia,PA, USA: Wolters Kluwer., 2012.