

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO E
SISTEMAS
MESTRADO PROFISSIONAL EM ENGENHARIA DA COMPUTAÇÃO E
SISTEMAS

JARDIEL NUNES ALMEIDA

*Proposta de um Método Computacional para
Detecção da Cardiotoxicidade*

São Luís
2016

JARDIEL NUNES ALMEIDA

*Proposta de um Método Computacional para
Detecção da Cardiotoxicidade*

Dissertação apresentada ao programa do Mestrado Profissional de Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão como parte dos requisitos para a obtenção do título de Mestre em Engenharia da Computação e Sistema.

São Luís
2016

JARDIEL NUNES ALMEIDA

*Proposta de um Método Computacional para
Detecção da Cardiotoxicidade*

Dissertação apresentada ao programa do Mestrado Profissional de Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão como parte dos requisitos para a obtenção do título de Mestre em Engenharia da Computação e Sistema.

Aprovada em de de 2016

BANCA EXAMINADORA

Prof. Lúcio Flávio A. Campos (Orientador)

Membro: Prof. Mauro Sérgio Silva Pinto

Membro: Prof. Pedro Brandão Neto

São Luís
2016

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 a esta disfunção cardíaca depende de um diagnóstico rápido, pois quanto mais cedo se iniciar 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 sinais proteômicos, depois utilizando a técnica de máxima relevância e mínima redundância para reduzir a dimensionalidade e com isto 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.

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.

Lista de Figuras

1	Esquema dos principais componentes de um espectrômetro de massas. Fonte [34]	20
2	Espectrômetro de massas. Fonte [57]	21
3	Conjunto de dados linearmente separável e um possível hiperplano separador. Fonte [85]	43
4	Ilustração da distância d entre os hiperplanos. Fonte [85]	45
5	Transformação do conjunto de dados no espaço de entrada (a) para o espaço de característica (b). Fonte [85]	48
6	Exemplo de conjunto não linearmente separável. Fonte [85]	48
7	Diagrama do método.	53
8	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: [32]	55
9	Espectro de Massa	56

Lista de Tabelas

1	Fatores de risco para cardiotoxicidade associada às antraciclinas	13
2	Recomendações para o tratamento da IC pós-QT	17
3	Resultados	57

Sumário

1	INTRODUÇÃO	10
2	Fundamentação Teórica	12
2.1	Cardiotoxicidade	12
2.1.1	Conceito e Manifestação	12
2.1.2	Fatores de Risco	13
2.1.3	Classificação	14
2.1.4	Agentes quimioterápicos e cardiotoxicidade	14
2.1.5	Diagnóstico	15
2.1.6	Tratamento da cardiotoxicidade	16
2.2	Análise Proteômica	18
2.2.1	Introdução	18
2.2.2	Espectrômetro de Massa	20
2.2.3	BIOINFORMÁTICA APLICADA AO ESTUDO DAS PROTEÍNAS	23
2.2.4	APLICAÇÕES BIOMÉDICAS	23
2.3	Análise de Componentes Independentes (ICA)	25
2.3.1	Introdução	25
2.3.2	Modelo e Definição	25
2.3.3	Independência Estatística e Descorrelação	26
2.3.4	Não Gaussianidade das Variáveis	28
2.3.5	Centralização	29
2.3.6	Branqueamento	29
2.3.7	Pré-Processamento Adicional	29
2.3.8	Princípios para Estimativa de Componentes Independentes	30
2.3.9	Maximização da Não Gaussianidade	30
2.3.10	Medida de Não Gaussianidade por Kurtosis	31
2.3.11	Negentropia	32
2.3.12	Aproximação da Negentropia	33
2.3.13	Estimativa de Máxima Verossimilhança	34
2.3.14	Informação Mútua	35
2.3.15	Principais Algoritmos para ICA	36
2.3.16	Algoritmo FastICA	36
2.4	Seleção de Características	37
2.5	Conceitos Básicos de Aprendizagem de Máquina	39
2.6	Máquina de Vetores de Suporte	40
2.6.1	Introdução	40
2.6.2	Bioinformática	41
2.6.3	Conjuntos Linearmente Separáveis	42

2.6.4	SVMs Lineares	43
2.6.5	Determinação do Hiperplano Ótimo	44
2.6.6	SVMs Não Lineares	47
2.6.7	Trabalhando com SVMs lineares no espaço de características	49
2.6.8	Determinação do hiperplano ótimo no espaço de características . .	49
2.6.9	Funções Kernel	50
2.6.10	Validação cruzada (VC)	51
3	Método e Materiais	53
3.0.1	Método Proposto	53
3.1	Base de Dados	54
3.2	Extração de Características	54
3.3	Seleção das características mais significantes	54
3.4	Avaliação do Método de Classificação	55
4	Resultados e Discussões	56
4.1	Aquisição de Dados	56
4.2	Extração de Características	56
4.3	Seleção das Características mais Significantes	57
4.4	Classificação	57
5	conclusão	59
	REFERÊNCIAS	60

1 INTRODUÇÃO

Ao longo dos últimos anos, diferentes tratamentos para diversos tipos de câncer foram largamente desenvolvidos, levando à 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, antraciclinas 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 a situação na qual agentes externos (químicos ou físicos) interferem negativamente no coração, determinando assim alterações estruturais, elétricas e funcionais no miocárdio[5]. Este órgão torna-se mais fraco e não é tão eficiente em bombeamento, o que compromete a circulação do sangue [3]. Tal 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, 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 o tratamento, 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 Auxiliado 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 - SVM)

entre duas classes, sendo que identificará o paciente com ou sem Cardiotoxicidade.

2 Fundamentação Teórica

2.1 Cardiotoxicidade

2.1.1 Conceito e Manifestação

Define-se Cardiotoxicidade pela situação na qual agentes externos (químicos ou físicos) interferem negativamente no coração, determinando assim alterações estruturais, elétricas e funcionais no miocárdio[5]. Ela é considerada como o efeito colateral mais grave decorrente do tratamento com quimioterápicos, principalmente relacionada à doxorrubicina. Dados consistentes indicam que a morte do cardiomiócito ocorre principalmente por apoptose e necrose. Grande estresse oxidativo, danos ao DNA nuclear, modificações na bomba de cálcio e contratilidade celular, supressão da transcrição de fatores que regulam a sobrevivência da célula e a síntese de proteínas, bem como a perda da estabilidade do sarcômero e estímulo à autofagia também são fatores identificados como importantes nos mecanismos de morte do cardiomiócito [5].

De maneira mais específica, a maioria dos estudos relaciona o conceito de cardiotoxicidade diretamente às medidas de função sistólica ventricular esquerda, principalmente avaliadas ao ecocardiograma, considerando como toxicidade cardíaca [11]:

- Fração de encurtamento do ventrículo esquerdo $< 28\%$ ou uma diminuição, em valor absoluto, > 10 unidades em relação ao valor basal pré-quimioterapia [5];
- Fração de ejeção ventricular esquerda:
 - Grau I: redução assintomática da FEVE em 10 a 20% do exame basal
 - Grau II: redução da FEVE mais que 20% do exame basal ou redução abaixo dos limites normais
 - Grau III: Insuficiência cardíaca clinicamente sintomática
- Fração de ejeção menor que 50% contraindica o uso do antracíclico.

As crianças e os adolescentes são particularmente mais susceptíveis aos efeitos Cardiotóxicos dos quimioterápicos, especialmente aos antracíclicos, uma vez que a perda de miócitos prejudica o crescimento cardíaco, resultando em inadequada massa ventricular esquerda residual e conseqüente cardiomiopatia no decorrer dos anos[5]. Também contribuem para a disfunção cardíaca a atrofia e desarranjo das miofibrilas observadas após a quimioterapia [10].

Reportando-se a esta questão, não existe um conceito padronizado a respeito da Cardiotoxicidade que englobe achados clínicos e de exames complementares. De forma geral, conceitua-se cardiotoxicidade secundária ao tratamento do câncer como achado de sinais e sintomas clínicos de insuficiência cardíaca não atribuída a outras conhecidas causas como: sepsis, insuficiência renal ou disfunção cardíaca prévia ao tratamento, documentada por exames complementares [11].

Como manifestações de Cardiotoxicidade, podemos encontrar discretas variações da pressão arterial, alterações eletrocardiográficas sutis até arritmias, pericardite, miocardite levando à cardiomiopatia com falência funcional do ventrículo esquerdo e insuficiência cardíaca congestiva, eventos tromboembólicos e até mesmo isquemia miocárdica e infarto agudo do miocárdio.

Podem ocorrer no início ou durante o tratamento, dias a semanas após o tratamento ou não ser aparente por meses ou até anos.

A partir do exposto, pode-se afirmar que é possível detectar a Cardiotoxicidade por meio de análise criteriosa dos fatores de risco, monitorização dos sinais e sintomas durante e após a quimioterapia e contínua avaliação por exames complementares proporcionando identificação precoce e intervenção clínica em tempo hábil. Todos os pacientes devem ser orientados, antes do início do tratamento, quanto aos riscos de Cardiotoxicidade, da necessidade de acompanhamento a longo prazo e da importância da manutenção de um estilo de vida saudável para prevenção ou melhor prognóstico de doenças cardiovasculares e metabólicas [12].

2.1.2 Fatores de Risco

Existem fatores de risco, bem documentados, que predispoem à cardiotoxicidade e são considerados indicadores independentes para disfunção ventricular. Estes incluem dose cumulativa (antraciclina, mitomicina), dose total administrada em um dia ou ciclo (ciclofosfamida, isofosfamida, carmustina, fluorouracil, cytarabina), esquema de administração em bolus ou lentamente (antraciclina), radiação mediastinal, idade, gênero feminino, administração concomitante de outros agentes cardiotoxícos, doença cardíaca preexistente, distúrbios eletrolíticos (principalmente hipocalcemia e hipomagnesemia), um resumo sobre estes fatores de risco é exposto na tabela 1:

Tabela 1: Fatores de risco para cardiotoxicidade associada às antraciclina

Fatores de Risco	Maior Risco
Idade	Menor Idade
Sexo	Feminino
Modo de Administração	Injeção Rápida
Dose Cumulativa	Dose Acima de:
Daunorrubicina	550 a 800mg/m ²
Doxorrubicina	400 a 550mg/m ²
Epirubicina	900 a 1.000mg/m ²
Idarrubicina	150 a 225mg/m ²
Irradiação Mediastinal	Precoce ou concomitante
Doenças cardiovasculares prévias	C.C.,miocardite, HAS, TEV,DAC
Distúrbios eletrolíticos	Hipocalcemia, hipomagnesemia

2.1.3 Classificação

A Cardiotoxicidade pode apresentar-se de forma aguda, subaguda ou crônica, na dependência da época de aparecimento das manifestações clínicas e achados aos exames complementares.

Cardiotoxicidade Aguda: forma rara de cardiotoxicidade que ocorre imediatamente após uma única dose de quimioterápico, ou no curso da terapia, onde as manifestações clínicas se apresentam dentro da primeira semana de tratamento.

Cardiotoxicidade Subaguda: semanas ou meses após o início do tratamento e geralmente de forma insidiosa.

Tanto a forma aguda quanto a subaguda podem se apresentar como alterações transitórias ao eletrocardiograma (ECG), síndrome miopericárdica ou falência aguda do ventrículo esquerdo. Ao ECG, pode-se encontrar alterações inespecíficas da repolarização, diminuição da voltagem dos complexos QRS, prolongamento do intervalo QT. A taquicardia sinusal é a alteração do ritmo mais frequente.

Alterações ao ECG são encontradas em cerca de 20 a 30% dos pacientes. Arritmias incluem taquicardias supraventriculares, ventriculares e juncionais (0,5 – 3%). Arritmias mais graves como flutter ou fibrilação atrial são raras.

Cardiotoxicidade Crônica pode ser dividida em dois tipos: crônica precoce (dentro do primeiro ano pós-tratamento) crônica tardia (depois do primeiro ano e até vários anos após o término da terapia).

Geralmente, a Cardiotoxicidade crônica apresenta-se como disfunção ventricular sistólica ou diastólica. Enquanto a cardiomiopatia dilatada (disfunção sistólica do ventrículo esquerdo) pode ocorrer imediatamente após altas doses de quimioterápicos ou no decorrer de meses ou anos, a cardiomiopatia restritiva é de ocorrência sempre crônica. A cardiomiopatia restritiva (disfunção diastólica do ventrículo esquerdo), que ocorre mais frequentemente pós-radioterapia, desenvolve-se em muitos sobreviventes ao longo do tempo, podendo levar à insuficiência cardíaca com função sistólica normal.

2.1.4 Agentes quimioterápicos e cardiotoxicidade

Os agentes quimioterápicos: Antraciclinas, Agentes alquilantes e Alcaloides da vinca são há muito conhecidos pelos seus efeitos adversos. Nos princípios gerais da quimioterapia, sabe-se que a maioria das drogas anticâncer tem mecanismos de ação não seletivos para células neoplásicas e células normais, e que a maioria delas produz seu efeito citotóxico interferindo com a síntese ou função dos ácidos nucleicos RNA e DNA.

Alguns desses efeitos, como a toxicidade medular (mielossupressão), náuseas e vômitos, alopecia, mucosite oral e intestinal, anormalidades de função hepática e renal, reações alérgicas sistêmica ou cutânea, são consequências inevitáveis do mecanismo de ação dos agentes citostáticos e/ ou citotóxicos em células que se dividem rapidamente. Tais efeitos

podem ser revertidos espontaneamente, ser tratados com medicação apropriada ou mudança no esquema de administração da droga e, via de regra, deixam poucas sequelas[5].

Outros efeitos são mais insidiosos e tendem a surgir em tecidos normais, como o coração, composto por células com uma capacidade regenerativa limitada. Na última década, a comunidade científica assistiu a uma acelerada aprovação de novas drogas que claramente melhoraram a sobrevivência dos pacientes com câncer. Infelizmente, entretanto, o objetivo de combinar melhor atividade antineoplásica com redução da toxicidade não foi totalmente obtido por esses novos medicamentos[5].

2.1.5 Diagnóstico

A avaliação inicial dos pacientes oncológicos submetidos a quimioterapia cardiotoxicidade tem como objetivos: excluir pacientes com evidências clínicas, laboratorial e radiológica de insuficiência cardíaca (IC) congestiva antes do início do tratamento quimioterápico, identificar pacientes com redução da fração de ejeção, associada a sintomas ou não, durante a quimioterapia [18]. É fundamental diagnosticar IC para evitar piora na qualidade de vida e aumento do risco de mortalidade dos pacientes[5]. Para fazer o diagnóstico da Cardiotoxicidade, destacam-se:

- **Sinais e sintomas:** Os sinais e sintomas de insuficiência cardíaca são importantes para o diagnóstico. No entanto, muitas vezes podem ser semelhantes às complicações do próprio câncer [19]. Para o diagnóstico da miocardiopatia associada à quimioterapia, é importante definir a classe e o quimioterápico utilizado, sua dose cumulativa, o uso prévio de outros quimioterápicos cardiotoxicos, e a presença de outros fatores de risco cardiovasculares [19]. São fatores de risco para Cardiotoxicidade de quimioterápicos: hipertensão, idade maior que 60 anos, disfunção do ventrículo esquerdo prévia, irradiação torácica prévia [20].
- **eletrocardiograma:** O eletrocardiograma é realizado rotineiramente na avaliação do paciente com fatores de risco para Cardiotoxicidade. Na insuficiência cardíaca, pode demonstrar baixa voltagem, bloqueio do ramo direito ou esquerdo, sobrecargas ventriculares e arritmias. É utilizado, também, para excluir distúrbios de condução, prolongamento do QT, anormalidades de repolarização ventricular [21]. A detecção de arritmias ventriculares e supraventriculares, como a fibrilação atrial, deve alertar o clínico para a presença de lesão estrutural cardíaca.
- **Biomarcadores:** A utilização de biomarcadores cardioespecíficos vem sendo apontada como ferramenta útil na identificação precoce de lesão cardíaca por quimioterápicos e seguimento dos pacientes oncológicos [22]. Estudos demonstram que a troponina pode representar um marcador sensível e específico de injúria miocárdica nos pacientes em uso de doses elevadas de quimioterápicos cardiotoxicos [23]. Ela

mostrou-se capaz de prever o desenvolvimento de disfunção ventricular, pois, nessa situação, seu nível sérico pode manter-se elevado após um mês do uso de antraciclina [24].

- **Métodos de imagem:** São necessárias a avaliação e a quantificação da função ventricular por métodos de imagem antes do início da quimioterapia cardiotoxicidade. Essa mensuração deve ser feita por meio do ecodopplercardiograma ou da ventriculografia radioisotópica [26]. O método escolhido deve ser mantido por todo o seguimento. Quando encontrados valores da FE < 50%, não se recomenda iniciar drogas com alto potencial de Cardiotoxicidade, sendo adequado discutir com o oncologista a possibilidade de indicar esquema quimioterápico com menor risco cardiovascular. O ecocardiograma tem sido a opção mais utilizada pelo baixo custo, fácil acesso e caráter não invasivo [26]. Além disso, permite avaliar não só a função sistólica, como a função diastólica, as válvulas cardíacas e o pericárdio.
- **Biópsia** A biópsia é um método altamente sensível e específico na detecção da cardiomiopatia induzida por antraciclina, mostrando alteração irreversível da arquitetura celular[29]. No entanto, seu uso tem sido desencorajado, não só por seu caráter invasivo e pelos riscos associados, mas também pela alta acurácia obtida pelos métodos de imagem na avaliação da função cardiovascular.

2.1.6 Tratamento da cardiotoxicidade

A disfunção ventricular pós-quimioterapia, foi durante muito tempo, considerada uma situação clínica irreversível. Esse conceito se baseava em estudos antigos, nos quais o diagnóstico somente era feito em fases bastante avançadas da afecção. Atualmente, com os fármacos utilizados para insuficiência cardíaca, e com avanços diagnósticos, tornou-se possível identificar a doença em fases mais precoces e até mesmo proporcionar melhora da função ventricular, mesmo quando ela já está instalada[8].

As drogas que efetivamente mudaram o prognóstico de pacientes com insuficiência cardíaca são aquelas que podem atuar no processo de remodelamento, proporcionando melhora de função e redução dos diâmetros ventriculares. Nesse contexto, os inibidores da enzima conversora de angiotensina (IECA), bloqueadores de receptores AT2, betabloqueadores e bloqueadores da aldosterona são drogas de primeira linha para o tratamento da IC, independentemente da etiologia. Uma vez que a discussão em questão é o tratamento da disfunção pós-QT, é racional que esses conceitos sejam extrapolados para essa população[8].

Os IECA constituem uma classe de medicamentos com comprovados benefícios na evolução clínica de pacientes com IC, em relação a morbidade, qualidade de vida e mortalidade. Essa afirmação baseia-se em numerosos ensaios randomizados, placebo-controlados, que demonstraram os benefícios dos IECA nos diferentes estágios evolutivos

vos da IC, desde os mais avançados aos moderados, e mesmo na disfunção ventricular sistólica assintomática; portanto, agindo também no processo de prevenção da disfunção ventricular[8].

O uso do IECA é indicado nas diferentes etiologias de IC, a partir do diagnóstico da disfunção ventricular, mesmo que os sintomas não estejam instalados ainda. Esse conceito é exatamente o que recomendamos em pacientes submetidos a tratamento quimioterápico. Durante a monitorização periódica desses pacientes, ao detectar sinais de disfunção sistólica e ou diastólica, deve-se introduzir essa classe de medicamentos, na maior dose tolerada. Para pacientes que não podem receber IECA em razão de intolerância, indicasse o uso de bloqueadores do receptor AT2 da angiotensina (BRA), para obtenção do efeito vasodilatador e de bloqueio neuro-hormonal (Tabela 10)[8].

Tabela 2: Recomendações para o tratamento da IC pós-QT

Classe	Indicação	N. de evidência
I	IECA em pacientes com IC, na maior dose tolerada	C
IIa	IECA em pacientes com disfunção diastólica	C
I	BRA devem ser recomendados a pacientes com IC	C
I	Betabloqueadores em pac. com IC e disf. sistólica assint.	C
I	Espironolactona em pac. sintomáticos, com disf. sistólica	C
IIa	Indicação de transplante cardíaco para pacientes com IC	C
III	IECA em pac. com insuficiência renal, hipotensão sint.	C

A introdução dos betabloqueadores no tratamento farmacológico da IC constitui um dos maiores avanços dos últimos anos no manejo clínico dessa doença. Seu uso tem sido fundamentado na atenuação da hiperatividade simpática aumentada na IC, que contribui para a progressão da disfunção miocárdica. Seus efeitos benéficos incluem inibição da cardiotoxicidade das catecolaminas, redução do consumo energético pelo miocárdico e melhora do relaxamento diastólico. Além disso, inibe a vasoconstrição periférica, reduz a frequência cardíaca, atua no remodelamento ventricular, e exerce efeitos anti-hipertensivos, antianginosos, antiarrítmicos e antiproliferativos[8].

Existem três betabloqueadores disponíveis para o tratamento da IC com efetividade comprovada: metoprolol, bisoprolol e carvedilol. Esses seriam os medicamentos a serem utilizados no tratamento clínico de pacientes com disfunção ventricular sistólica e/ou diastólica após QT[30].

Os betabloqueadores apresentam benefícios clínicos comprovados de melhora da classe funcional, redução da progressão dos sintomas de IC e redução de internação hospitalar, em pacientes com insuficiência cardíaca com disfunção sistólica, classe funcional I a IV da NYHA [8]. Quando associados com IECA ou com BRA, os betabloqueadores apresentam atuação terapêutica no aumento da sobrevida em longo prazo por redução na mortalidade por insuficiência cardíaca e morte súbita por arritmia ventricular[8].

Os antagonistas do receptor da aldosterona (espironolactona e eplerenone) também são drogas que se demonstram eficazes na redução da morbidade e mortalidade dos pacientes com disfunção ventricular em classes funcionais II a IV, com efeitos favoráveis também no processo de remodelação ventricular e, portanto, estão indicadas em todos os pacientes com disfunção ventricular sintomática[31]. A Tabela 10 traz as evidências do tratamento da IC após QT.

Também é importante discutir que hoje, cada vez mais, se fala da indicação de transplante cardíaco para pacientes com IC avançada, não responsiva ao tratamento clínico otimizado. É sabido que a terapêutica imunossupressora pode predispor ao aumento da incidência de câncer, ou até mesmo induzir recorrência da doença prévia. Assim, é fundamental a discussão conjunta com a oncologia para definição de critério de cura e rastreamento completo para posterior indicação e inclusão do paciente na fila de transplante.

2.2 Análise Proteômica

2.2.1 Introdução

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 [32], vêm sendo desenvolvidos.

Para desenvolver estes marcadores são utilizados conceitos como o da Espectrometria de Massa, é 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, que é um gráfico que mostra a abundância (intensidade) relativa de cada íon que aparece como picos com m/z definidos. Para esse fim é necessário um espectrômetro de massa que é um analisador que permite a determinação qualitativa e quantitativa dos compostos de uma amostra [35].

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. Nesta perspectiva gerar listas de proteínas que aumentam ou diminuem em expressão como causa ou consequência de patologia [32]. 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 [32].

2.2.2 Espectrômetro de Massa

O espectrômetro de massa é um analisador que permite a determinação qualitativa e quantitativa dos compostos de uma amostra [35], como ilustrado nas figuras 1 e 2. Este equipamento faz uma medida da razão massa sobre a carga (m/z) e não da massa em si. Se por exemplo, uma biomolécula é ionizada pela adição de um ou mais prótons (íons H^+) o instrumento mede o m/z após a adição de $1Da$ (Dalton) para cada próton adicionado. Assim uma molécula com massa de $1000Da$ que seja ionizada pela adição de um próton aparecerá no espectro com m/z de $1001 [(1000 + 1)/1]$. Por outro lado se a mesma molécula sofrer adição de dois prótons durante a ionização, essa molécula será detectada com m/z $501 [(1000 + 2)/2]$. Moléculas também podem ser ionizadas perdendo um próton, adquirindo carga negativa. Nesta situação o espectrômetro passa a detectar íons de carga negativa. Assim uma molécula com massa de $1000Da$ que seja ionizada pela perda de um próton aparecerá no espectro com m/z de $999 [(1000 - 1)/1]$. Por outro lado se a mesma molécula sofrer perda de dois prótons durante a ionização, essa molécula será detectada com m/z $499 [(1000 - 2)/2]$.

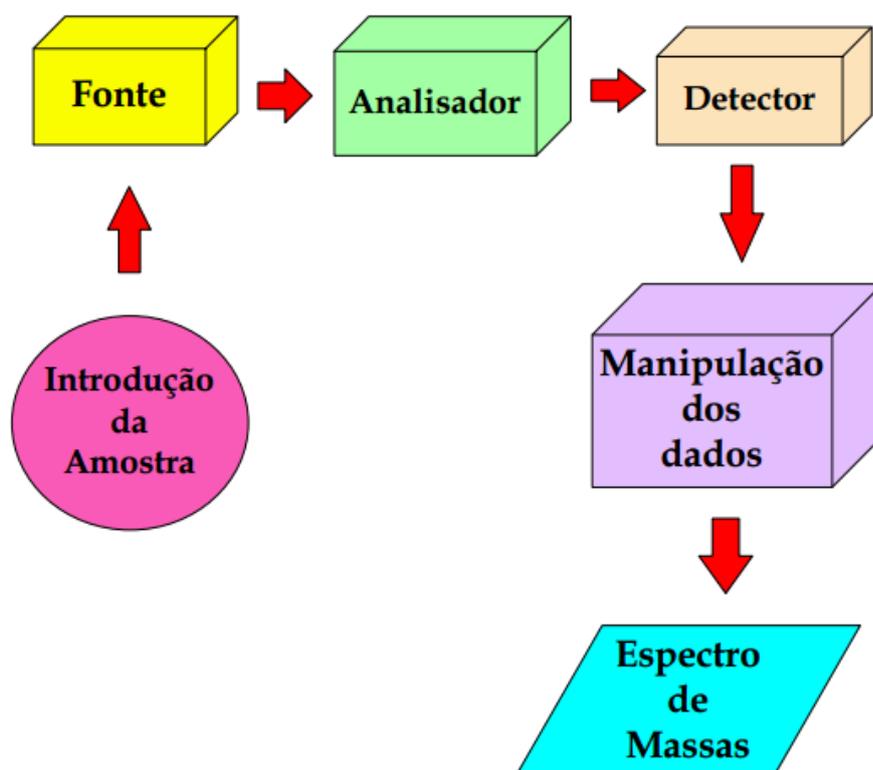


Figura 1: Esquema dos principais componentes de um espectrômetro de massas. Fonte [34]



Figura 2: Espectrômetro de massas. Fonte [57]

O desenvolvimento de duas técnicas de ionização, o electrospray (ESI, electrospray ionization) e ionização/dessorção a laser assistida por matriz (MALDI, matrix-assisted laser desorption/ionization), tornou possível a determinação precisa de massas de compostos de alto peso molecular, bem como de moléculas de baixo peso molecular e revolucionou a aplicabilidade da espectrometria de massas para quase todas as moléculas biológicas. As aplicações da técnica incluem a área da Proteômica bem como a de descoberta de drogas (*Drug Discovery*).

As características fundamentais de um espectrômetro de massas são:

- produção de íons em fase gasosa;
- aceleração dos íons a velocidades específicas em um campo elétrico;
- separação dos íons por um analisador de massas;
- detecção de cada íon em um m/z específico.

O instrumento é calibrado com padrões cujas massas são conhecidas com alta precisão. No espectrômetro de massas utiliza-se uma escala de carbono com $^{12}\text{C} = 12,000000$. Este nível de precisão pode ser atingido em equipamentos de alta resolução.

O analisador de massas pode separar os íons pelo uso de campos magnéticos ou elétricos. Alternativamente o tempo que os íons de diferentes massas levam para migrar distâncias definidas pode ser medido precisamente em um analisador de massas por tempo de voo (TOF, time-of-flight).

Qualquer material que possa ser ionizado e cujos íons possam ser transferidos para a fase gasosa pode ser analisado por espectrometria de massas (MS), lembrando que a análise requer pressões extremamente baixas (ex. alto vácuo de aproximadamente 10 – 6 Torr). A maioria das análises de moléculas biológicas como proteínas, oligossacarídeos, e ácidos nucleicos são realizadas em analisadores do tipo quadrupolo, quadrupolo-ion trap e TOF.

Componentes de um espectrômetro

Todos os espectrômetros de massa são basicamente similares. Eles consistem dos seguintes itens:

1. Sistema de alto vácuo. Para isso utiliza-se uma série de bombas (ex. bombas turbo moleculares, etc.);
2. Sistema de entrada de amostras (amostras líquidas podem entrar por capilares ou amostras sólidas podem ser colocadas em placas específicas);
3. Fonte de ionização onde as amostras são convertidos em íons e transferidos para a fase gasosa;
4. Analisador de massas. Pode ser de vários tipos TOF, quadrupolo, quadrupolo-ion trap, etc.;
5. Detector.

2.2.3 BIOINFORMÁTICA APLICADA AO ESTUDO DAS PROTEÍNAS

Os experimentos desenvolvidos no laboratório necessitam ser complementadas pelas análises virtuais feitas com o auxílio de um computador [41]. Além de softwares para analisar a separação eletroforetica, ferramentas bioinformáticas foram desenvolvidas. Algumas destas estão disponíveis através da Internet ao usuário: ExpASy (www.expasy.ch/www/tools.html)[41].

Estas bases de dados incluem mapas de proteína do plasma humano, da urina, do líquido cérebro-espinhal, e dos tecidos da mama, do coração, e da bexiga e dos carcinomas, assim como de vários micro-organismos. Finalmente, dado a natureza dinâmica do proteoma, os detalhes experimentais e os resultados relacionados devem ser indicados como importantes chaves bioquímicas ou implicações de relevância nas doenças [42].

Estes softwares permitem não somente a identificação das proteínas, mas também de uma caracterização adicional que varia do cálculo das propriedades físico-químicas básicas até a predição de potenciais modificações pós-traducionais e de estruturas tridimensionais 3D. A proteína encontrada e as bases de dados dos géis bidimensionais são o núcleo da bioinformática na pesquisa de proteomas. SWISS-PROT é um exemplo típico de uma base de dados [42].

2.2.4 APLICAÇÕES BIOMÉDICAS

A perspectiva oferecida pela proteômica tem sido utilizada na pesquisa de diferentes áreas da medicina, incluindo a biomedicina. Estas pesquisas poderiam se classificar de distintas formas: em função do tipo de amostra empregada, da doença ou do tipo de doenças que abordam, da técnica ou das técnicas utilizadas, do uso ou da aplicação, etc.

As pesquisas clínicas com proteômica se baseiam em torno ao tipo de amostra utilizada. Nesta base de pesquisas destaca-se as pesquisas com linhagens celulares, com tecidos e proteômica de fluidos. As linhagens celulares são variantes derivadas das células cancerosas ou de células normais fixadas pela inibição dos mecanismos celulares que causam a interrupção do crescimento. Diferente da maioria das células isoladas diretamente dos organismos, estas células são capazes de se dividirem indefinidamente quando se mantêm em cultura no laboratório [43].

Na pesquisa com linhagens celulares, a proteômica tem caracterizado os perfis de expressão de proteínas em diferentes tipos destas linhagens como base para futuros experimentos comparativos [44] [45]. Estudos têm demonstrado que o número de proteínas identificadas em cada proteoma varia de forma importante, o que provavelmente seja o reflexo da existência de diferenças no nível de expressão das distintas proteínas em cada linhagem celular [46].

As linhagens celulares tumorais humanas podem constituir um modelo importante para o estudo do câncer. Deste modo, as pesquisas que comparam os proteomas das

diversas linhagens tumorais humanas (ou células obtidas diretamente de um tumor) com suas correspondentes linhagens celulares normais começam a ser habituais para identificar marcadores da doença ou biomarcadores que permitam a detecção precoce, classificação e escolha do prognóstico dos tumores assim como para propor novas formas terapêuticas de tratamento. Sendo assim, a proteômica pode realizar pesquisas para avaliar o potencial maligno de um tumor (potencial metastático segundo seja a expressão de uma proteína específica), a quimiosensibilidade ou quimioluminescência deste [47]. O uso da proteômica para o desenho, indicação e predição da resposta de um fármaco ou a fármacos seria também extrapolável a outras doenças, é o que se conhece como farmacoproteômica [48].

Além das pesquisas em linhagens celulares humanas, a proteômica tem-se aplicado atualmente na análise de tecidos e fluidos para o estudo de situações fisiológicas (durante o desenvolvimento, em diferentes estágios metabólicos, frente a diversas respostas ambientais) ou patológicas (câncer, auto-imunidade, infecções) [49]. Os objetivos destas pesquisas utilizando tecidos e fluidos são similares às pesquisas com linhagens celulares (identificação de biomarcadores, farmacoproteômica).

A falta de acesso de amostras teciduais é um problema nas pesquisas proteômicas clínicas, mas a informação obtida pode ser determinante e, às vezes, é a melhor escolha [50].

Na proteômica de fluidos, apresenta-se maior facilidade de obtenção e processamento das amostras. A aplicação da proteômica tem-se utilizado para o estudo das diversas doenças (neoplásicas, reumáticas, endocrinológicas, toxicológicas) [32] [49]. Outro tipo de amostra de mais difícil acesso é a análise proteômica do líquido cefalorraquidiano. Este se tem utilizado recentemente no estudo das bases etiopatogênicas e na identificação de biomarcadores de doenças neurológicas, como transtornos neuropsiquiátricos, tumores cerebrais e síndromes dolorosas lombares [51].

O efeito máximo das abordagens proteômicas na pesquisa biomédica não foi conseguido ainda, em parte por causa da falta de recursos e informação entre os pesquisadores sobre os avanços tecnológicos. Entretanto, um grande progresso está sendo feito, e importantes revisões, informações de diversas áreas biomédicas são dadas para ilustrar o potencial desta abordagem. As abordagens proteômicas oferecem um grande potencial para decifrar problemas biológicos complexos tais como a natureza de complexos moleculares particulares ou dos caminhos na patogênese da doença [52]), e de complexos nucleares de leveduras [53]. A descoberta de biomarcadores biológicos de doenças como câncer [54], mal de Chagas [55] e para a comparação dos padrões de proteína da catarata humana [56].

2.3 Análise de Componentes Independentes (ICA)

2.3.1 Introdução

A Análise de Componentes Independentes(ICA) tem como objetivo principal, encontrar uma representação linear de dados não gaussianos, minimizando a dependência estatística entre eles, de forma que os componentes resultantes sejam estatisticamente independentes, ou tão independentes, quanto possível.

Neste contexto, a ICA diferencia-se das outras técnicas existentes, pois ela trabalha com componentes que são ao mesmo tempo não gaussianos e estatisticamente independentes entre si. Esta representação pode ser usada para capturar a estrutura dos componentes, ou dados, em várias aplicações como extração de características ou separação de sinais.

Neste capítulo será realizada uma revisão do estado da arte dos princípios da Análise de Componentes Independentes, bem como uma revisão dos principais algoritmos que utilizam a ICA com o objetivo de realizar a separação cega de sinais. É apresentada ainda uma revisão de conceitos estatísticos indispensáveis para a compreensão do problema da ICA.

2.3.2 Modelo e Definição

A Análise de Componentes Independentes é um poderoso método utilizado principalmente para o tratamento de problemas de separação cega de sinais de fontes independentes. Nesta perspectiva, a ICA estima quantidades desconhecidas a partir de medidas conhecidas, sejam elas parâmetros, sequências de dados ou sinais. Neste trabalho, as medidas foram obtidas a partir de um sinal proteômico.

Reportando-se a esta questão, vale ressaltar que para fazer a definição do modelo ICA, utilizaremos variáveis que não são observadas diretamente, elas são inferidas através de outras variáveis observadas [117]. Assim, sejam n variáveis aleatórias observadas x_1, x_2, \dots, x_n , que entendemos como sendo misturas e este modelo tem como objetivos principais encontrar uma representação linear de dados não gaussianos, minimizando a dependência estatística entre eles de forma que os componentes resultantes sejam estatisticamente independentes, ou tão independentes quanto possível, além de identificar e separar fontes estatisticamente independentes a partir de uma determinada mistura de fontes originais. E isto tem como consequência a extração das características desta mistura. Neste trabalho cada dado ou sinal proteômico x_i é modelado como uma combinação linear de n variáveis aleatórias s_1, s_2, \dots, s_n , tal que:

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

O modelo linear pode ser expresso na forma matricial, por:

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

Desta maneira o problema consiste em encontrar o sinal \mathbf{S} , então para isto bastaria multiplicarmos a equação 2 pela inversa da matriz \mathbf{A} e assim teríamos 3, com isso o problema estaria resolvido.

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \quad (3)$$

Porém, não é possível obter a matriz \mathbf{A}^{-1} já que nada é conhecido a respeito da matriz \mathbf{A} . O que se consegue no máximo, é fazer uma estimativa da matriz \mathbf{A}^{-1} a partir das considerações de independência e não-gaussianidade das fontes. Assim, a ICA procura estimar um conjunto de fontes y_i também independentes, através de um sistema de separação \mathbf{W} quando somente as medidas do sinal x_i são conhecidas.

Desta forma, podemos redefinir o problema de encontrar os Componentes Independentes como:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} \quad (4)$$

sendo \mathbf{Y} o vetor de Componentes Independentes que aproxima \mathbf{S} e \mathbf{W} uma estimativa da matriz \mathbf{A}^{-1} .

Desta forma, pode-se dizer que a não-gaussianidade das fontes originais é o ponto de partida para a elaboração de métodos para a solução do problema da ICA, pois só assim será possível a determinação dos coeficientes de \mathbf{W} e dos componentes de \mathbf{Y} .

Para melhor compreensão de como são feitas as medidas de não-gaussianidade e como estas medidas são utilizadas para a separação de fontes estatisticamente independentes. Serão apresentados ao longo deste capítulo, os princípios matemáticos da ICA e os principais algoritmos desenvolvidos com base nestes princípios.

Obs: O operador \mathbf{W} na equação 4 pode ser visto, como um operador de rotação que busca representar as componentes de \mathbf{X} na direção de máxima independência estatística.

Nas seções seguintes serão apresentados as principais considerações estatísticas para solucionar o problema da Análise de Componentes Independentes.

2.3.3 Independência Estatística e Descorrelação

Para que se possa utilizar ICA é necessário que os sinais originais sejam estatisticamente independentes. Assim sendo, a matriz de separação \mathbf{W} é determinada de tal forma que os componentes de $y(k)$ também sejam estatisticamente independentes. Isto significa que o valor de um componente não fornece nenhuma informação sobre o valor dos outros componentes.

Sejam $p(y_1, y_2, \dots, y_n)$ a função densidade de probabilidade conjunta f_{dp} variáveis

aleatórias y_1, y_2, \dots, y_n e $p(y_n)$ a função densidade de probabilidade marginal de y_n , ou seja, a função densidade de probabilidade de y_n quando somente esta é considerada. É dito que y_1, y_2, \dots, y_n são estatisticamente independentes se e somente se, a função densidade de probabilidade conjunta de $y(k)$ puder ser fatorada no produto de suas f.d.p.'s marginais, como mostra a equação 5.[71]

$$p(y_1, y_2, \dots, y_n) = p(y_1).p(y_2) \dots p(y_n) \quad (5)$$

A equação 5 pode ser utilizada para definir a propriedade mais importante das variáveis aleatórias independentes, ou seja, [58]:

$$E[g(y_1), g(y_2), \dots, g(y_n)] = E[g(y_1)].E[g(y_2)] \dots E[g(y_n)] \quad \text{para } y_1 \neq y_2 \neq \dots \neq y_n \quad (6)$$

Sendo $g(\cdot)$ funções não lineares quaisquer.

Através da equação 6 pode-se verificar que se uma determinada v.a. independente sofrer uma transformação não linear a condição de independência não é afetada.

Uma forma fraca de independência é a não correlação. Um conjunto de n v.a. (y_1, y_2, \dots, y_n) é dito ser não correlacionado se sua covariância é igual a zero; como definido pela equação 7.

$$E[g(y_1), g(y_2), \dots, g(y_n)] - E[g(y_1)].E[g(y_2)] \dots E[g(y_n)] = 0 \quad (7)$$

Como base em 5 e 7 pode-se comprovar que se as v.a. são estatisticamente independentes, elas são não correlacionadas, mas a não correlação não implica em independência. O único caso em que não correlação implica em independência estatística ocorre quando as v.a. são gaussianas, uma vez que estas variáveis são completamente descritas por suas estatísticas de segunda ordem [58]. Como a não correlação não garante a independência, ela também não pode garantir a separação em componentes independentes. Assim, para garantir a separação é necessário recorrer a estatísticas de ordem superior. A utilização de estatísticas de ordem superior, por sua vez, faz com que a separação, utilizando ICA, somente seja garantida para no máximo uma fonte gaussiana.

Não obstante, como a independência estatística implica não correlação, a maioria dos métodos de ICA restringe o problema da estimativa de fontes originais de tal forma que os componentes independentes estimados sejam sempre não correlacionados. O procedimento de não correlação reduz o número de parâmetros livres a serem determinados na separação, simplificando o problema.

2.3.4 Não Gaussianidade das Variáveis

Uma das principais restrições para a realização da ICA é que os componentes independentes sejam não gaussianos ou que no máximo um dos componentes seja gaussiano. Nesta seção são apresentadas as principais razões que justificam esta restrição.

Considere um sistema de mistura (2×2) ortogonal e duas fontes gaussianas ($s_1 e s_2$). Assim, os sinais misturados resultantes ($x_1 e x_2$) são gaussianos, decorrelacionados e tem variância unitária.

Como a f.d.p. conjunta de ($x_1 e x_2$) é simétrica, ela não contém nenhuma informação sobre a matriz de mistura \mathbf{A} . Desta forma, se mais do que uma das fontes originais for gaussiana, pode-se dizer que a matriz \mathbf{A} não pode ser estimada [58]. De forma mais rigorosa, pode-se comprovar que a distribuição das variáveis gaussianas não é afetada por qualquer transformação ortogonal e, que as variáveis são independentes. Assim, para variáveis gaussianas, pode-se apenas estimar o modelo para ICA a menos de uma transformação ortogonal. Isto confirma e reforça a afirmação de que a separação em componentes independentes só é possível se no máximo uma fonte original for gaussiana. Além disso, é importante reforçar a ideia de que a separação somente é possível através da utilização de estatísticas de ordem superior e que, para variáveis gaussianas, as estatísticas de ordem maior que dois são zero, uma vez que as variáveis gaussianas são totalmente caracterizadas por suas estatísticas de segunda ordem.

Por outro lado, para variáveis não gaussianas, a independência estatística é a principal garantia de que as fontes estimadas serão distintas e com base neste critério são definidas a maioria das estratégias para a separação cega utilizando ICA.

A seguir são apresentadas mais informações sobre as f.d.p. 's gaussianas e não gaussianas para melhor compreensão das definições.

A função densidade de probabilidade para a distribuição gaussiana ou normal é definida em 8.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

Onde μ e δ são, respectivamente, a média e o desvio padrão.

Uma densidade de probabilidade sub-gaussiana ou Platykurt é tipicamente mais plana que a distribuição gaussiana. Um exemplo é a f.d.p. uniforme.

Já a f.d.p. super-gaussiana ou *Leptokurtic* tem tipicamente extremidades mais prolongadas e picos mais proeminentes que os da distribuição gaussiana. Um exemplo é distribuição Laplaciana. Sinais de voz e de música apresentam distribuição super-gaussiana.

Assim, vários métodos foram desenvolvidos, baseados no princípio de que para variáveis não gaussianas a independência é suficiente para garantir que as fontes estimadas serão distintas, validando desta forma a separação das fontes.

Nas próximas seções, serão apresentados os principais métodos utilizados para fazer

medidas de independência que permitem a separação dos sinais recebidos pelos sensores em componentes independentes. A maioria dos métodos para ICA, necessita que os dados passem por pré-processamento. Os principais métodos de pré-processamento empregados pelas técnicas de análise de componentes independentes são a Centralização e o Branqueamento. Outros métodos como, por exemplo, a filtragem tipo passa faixa [72], também podem ser empregados, mas não serão discutidos neste trabalho.

2.3.5 Centralização

O processamento inicial é feito através da centralização do vetor $x(k)$. Isto é feito subtraindo a média $E[x(k)]$ de cada elemento de $x(k)$, fazendo com que $x(k)$ se torne uma variável de média zero. Isto implica que $s(k)$ também tenha média zero [72].

Este tipo de pré-processamento não é obrigatório, mas simplifica os algoritmos. Após a estimativa da matriz \mathbf{A} , com os dados centralizados, pode-se completar a estimativa adicionando vetor média de $s(k)$ de volta ao vetor $s(k)$ centralizado.

2.3.6 Branqueamento

As informações estatísticas de segunda ordem não permitem a separação em componentes independentes. Mas, por outro lado, a decorrelação espacial ou independência de segunda ordem permite expressar o espaço dos sinais medidos em um sistema ortogonal, ou branco, que ainda não implica na independência estatística destas medidas, mas que irá simplificar o problema [68].

No branqueamento, também denominado esferamento, o vetor de observações $x(k)$ é linearmente transformado em um novo vetor $z(k)$ que é branco e é definido como: $z(k) = Qx(k)$, onde Q é uma matriz de transformação linear.

O branqueamento faz com que os componentes de $z(k)$ sejam decorrelacionados e que sua matriz de covariância seja unitária, ou seja, igual à identidade.

Portanto, se Q é uma matriz de branqueamento para $x(k)$, então $z(k) = Qx(k)$, é espacialmente branco, seus componentes são decorrelacionados e sua variância é unitária.

O branqueamento é aplicado, portanto, após a centralização e antes do algoritmo principal e sua aplicação é sempre possível.

2.3.7 Pré-Processamento Adicional

Para simplificar ainda mais o problema da ICA alguns algoritmos utilizam métodos de pré-processamento adicionais.

Em alguns casos, por exemplo, pode ser útil reduzir a dimensão dos dados, ao mesmo tempo em que é feito o branqueamento, descartando os autovalores de $E[x(k)x(k)^T]$ que sejam muito pequenos (*da ordem de* 10^{-10}), resultando, em geral, na redução do ruído. A

redução de dimensão pode prevenir ainda a sobreposição que algumas vezes ocorre na ICA [73].

Além disso, a filtragem pode ser útil, já que não altera a matriz de mistura, uma vez que os componentes independentes são filtrados pelo mesmo filtro que irá filtrar a mistura.

O filtro ótimo a ser utilizado é aquele que aumenta a independência dos componentes enquanto reduz o ruído, ou seja, é aquele que une as vantagens dos filtros passa-alta e passa-baixa enquanto minimiza suas desvantagens. Isto leva a filtragem passa-faixa, na qual as frequências mais altas e mais baixas são filtradas, deixando uma faixa de frequência adequada [72].

2.3.8 Princípios para Estimativa de Componentes Independentes

A estimativa dos dados independentes é realizada, na maioria dos casos, pela minimização ou maximização de uma dada função custo, de contraste ou objeto [63]. Neste trabalho, será utilizado o termo função custo para definir qualquer função cuja otimização permite a estimativa de componentes independentes. Em [74], como neste trabalho, é feita uma distinção entre a função custo e o algoritmo para implementar a ICA, sendo que eles estão relacionados.

Desta forma, o modelo para ICA depende de ambos, algoritmo de otimização e função custo, ou, em outras palavras, as propriedades estatísticas - como consistência e robustez - do modelo ICA dependem da escolha da função custo e as propriedades computacionais - como velocidades de convergência, requisitos de memória e estabilidade - dependem da escolha do algoritmo de otimização, sendo que algoritmo e função custo são independentes, ou seja, podem-se utilizar diferentes algoritmos de otimização para a mesma função custo e vice e versa.

As principais funções custo para estimativa dos componentes independentes são baseadas na medida e maximização da não-gaussianidade das variáveis medidas, principalmente por meio de HOS, minimização da informação mútua e estimativa da máxima verossimilhança.

Embora conceitualmente diferentes, pode-se demonstrar que matematicamente estas funções são equivalentes [75].

A partir destas funções custos foram desenvolvidos inúmeros algoritmos como os apresentados em ([68], [67], [64], [63], [62], [74]).

Serão apresentadas a seguir as principais funções custo utilizadas para Análise de Componentes Independentes baseadas na abordagem estatística.

2.3.9 Maximização da Não Gaussianidade

Uma das formas de estimar os componentes independentes é através da medida da não gaussianidade dos componentes, considerando que estes são centralizados (média zero) e

de variância unitária.

A medida da não gaussianidade pode ser feita por meio do kurtosis ou achatamento da curva da distribuição de probabilidade dos componentes sob estudo ou utilizando o princípio da negentropia como será apresentado nas seções seguintes [72].

2.3.10 Medida de Não Gaussianidade por Kurtosis

As funções custo mais simples para ICA são expressas através de cumulantes quarta ordem, ou kurtosis.

O kurtosis ($kurt(.)$) de uma variável aleatória $y(k)$ é definido como:

$$kurt(y(k)) = E[y(k)^4] - 3E[y(k)^2]^2 \quad (9)$$

Considerando que $y(k)$ tem variância unitária, 9 pode ser escrita por 10:

$$kurt(y(k)) = E[y(k)^4] - 3 \quad (10)$$

A equação 10 mostra que o kurtosis é simplesmente uma normalização do momento de quarta ordem, ou seja, $E[y(k)^4]$.

Existe uma propriedade muito importante com relação aos kurtosis e que justifica sua utilização em ICA, ou seja, para uma variável gaussiana, o momento de quarta ordem é dado por $3E[s^2(k)]^2$, como a variância $E[s^2(k)] = 1$; conseqüentemente, o kurtosis é nulo, o que reforça a ideia de que variáveis gaussianas são proibidas para ICA.

O kurtosis de uma variável aleatória pode ser positivo, negativo ou nulo, sendo que:

- $k > 0$ a variável é sub-gaussiana é um exemplo de função densidade de probabilidade é a plana, uniforme;
- $k < 0$ a variável é super-gaussiana é um exemplo de função densidade de probabilidade é a Laplaciana;
- $k = 0$ a variável é gaussiana é um exemplo de função densidade de probabilidade gaussiana ou normal.

A não gaussianidade pode ser medida através do valor absoluto do kurtosis, sendo que:

- $|Kurt| = 0$ as variáveis são gaussianas;
- $|kurt > 0|$ as variáveis são não-gaussianas.

Apenas em alguns casos especiais podem-se encontrar variáveis não gaussianas com $|kurt| = 0$ [71].

Assim, as componentes independentes são encontrados maximizando ou minimizando o kurtosis, de acordo com algoritmo de otimização empregado.

Na prática, para se encontrar a matriz de separação \mathbf{W} que fornecerá os componentes independentes $y(k)$, tais que, $y(k) = s(k) = \mathbf{W}^T x(k)$, com base na medida dos kurtosis das fontes, deve-se, como descrito em [58]:

- Branquear o sinal medido $x(k)$, de forma a se obter $z(k)$;
- Inicializar \mathbf{W} com um valor qualquer;
- Calcular a direção em que kurtosis de $y(k)$ está crescendo (ou decrescendo) mais fortemente;
- Baseado nas medidas disponíveis $x(k)$, utilizar um método de gradiente, ou equivalente, para encontrar um novo vetor \mathbf{W} .
- O processo acima descrito continua até que todas as fontes sejam estimadas.

Embora existam vários algoritmos que utilizam kurtosis para solucionar o problema da ICA, na prática, eles apresentam algumas desvantagens, pois sua implementação computacional é complexa, não é uma medida robusta de não gaussianidade e é sensível a outliers (amostras inconsistente que não se adaptam a curva).

Além disso, como se devem estimar os kurtosis a partir de dados observados, ou amostrados, os valores podem variar bastante em função do número de amostras e alterar a estimativa das fontes.

2.3.11 Negentropia

Uma outra forma de se medir a não gaussianidade de uma função é através da negentropia ou entropia diferencial.

A entropia de uma variável aleatória pode ser interpretada como o grau de desordem ou desorientação deste sistema; assim, quanto mais imprevisível e desestruturada for a variável, maior é sua entropia. Para a mesma variância, as variáveis gaussianas são as que apresentam a maior entropia [71]; o significa que a distribuição gaussiana é mais aleatória ou menos estruturada de todas as distribuições e, portanto, a entropia pode ser utilizada como medida de não gaussianidade.

Com base na teoria da informação e sob determinadas restrições, a entropia H definida pela equação 25 define o comprimento do código de uma variável aleatória. Em outras palavras, a entropia define o grau de informação de uma dada variável aleatória [72].

$$H(y) = - \sum_i p(y = a_i) \log(p(y = a_i)) \quad (11)$$

Onde $P(y = a_i)$ é a probabilidade de $y(k) = y$ ser igual a a_i . A definição 25 pode ser generalizada para variáveis aleatórias contínuas e vetores e, nestes casos, ela passa a ser denominada entropia diferencial. Matematicamente, a entropia diferencial pode então ser definida por:

$$H(y) = - \int f(y) \log(f(y)) dy \quad (12)$$

Onde: $f(y)$ é a f.d.p. de $y(k)$.

Por simplicidade, para se obter uma medida de não gaussianidade, utiliza-se frequentemente uma modificação da entropia diferencial denominada negentropia (J), que é equacionada em 27 [58].

$$J(y) = H(y_{GAUSS}) - H(y) \quad (13)$$

Onde: y_{GAUSS} = variável aleatória gaussiana com a mesma matriz de covariância que y .

Da equação 27 obtém-se que:

- quando a $|Kurt| = 0$ as variáveis são gaussianas;
- quando a $|kurt > 0|$ as variáveis são não-gaussianas.

A vantagem de se utilizar a negentropia como medida de não gaussianidade é sua base teórica e matemática bem definida. Outra característica importante da negentropia é que ela é invariante para transformações lineares inversíveis [63].

O inconveniente de se utilizar a negentropia como função custo, é que ela requer um grande esforço computacional, pois necessita de uma estimativa da f.d.p. da função analisada. Uma solução é utilizar uma aproximação da negentropia.

2.3.12 Aproximação da Negentropia

Um método clássico para o cálculo da negentropia aproximada é através de momentos de ordem superior, como mostra a equação 28 [58].

$$J(y) = \frac{1}{12} E[y^3]^2 - \frac{1}{48} \text{curtosis}(y)^2 \quad (14)$$

Em 28 considera-se que y tem média zero e variância unitária e a validade desta aproximação é limitada, sofrendo com a não robustez do *kurtosis*, chegando algumas vezes a ser imprecisa e em muitos casos sensível a *outliers*.

Para evitar problemas como estes, novas técnicas foram propostas. No caso mais simples, utiliza-se uma função não-quadrática, $G(\cdot)$, resultando na aproximação mostrada em 16.

$$J_G(y) = \alpha[E[G(y)]^2 - E[G(b)]]^2 \quad (15)$$

A equação 16 é válida para praticamente qualquer função $G(\cdot)$ não-quadrática e considerando que b é uma variável aleatória gaussiana com média zero e variância unitária.

A escolha adequada de $G(\cdot)$, faz com que 16 seja uma aproximação conceitualmente simples, computacionalmente rápida e com propriedades estatísticas interessantes, como a robustez. Valores de $G(\cdot)$, que se mostram adequados, foram:

$$G_1(u) = \frac{1}{a} \log(\cosh(a_1 u)), \quad G_2(u) = -\exp\left(\frac{-u^2}{2}\right) \quad (16)$$

Um dos principais algoritmos que utilizam a aproximação da negentropia para medida de não-gaussianidade é o FastICA [76].

2.3.13 Estimativa de Máxima Verossimilhança

Uma das abordagens mais conhecidas para Análise de Componentes Independentes é através da estimativa da Máxima Verossimilhança (MLE - Maximum Likelihood Estimation). Os estimadores baseados em MLE são muito utilizados por sua eficiência e variação mínima. Mas, por outro lado, a ML requer que as funções de densidade de probabilidade dos sinais de entrada sejam conhecidas, o que nem sempre é possível e tem sido um desafio para os pesquisadores. Contudo, o princípio da ML pode ser aplicado para solucionar o problema de separação cega utilizando ICA. Para um modelo sem ruído aditivo, pode-se determinar a verossimilhança e então estimar os componentes independentes pela maximização da verossimilhança como descrito em [72].

O estimador ML é calculado buscando parâmetros que maximizem a verossimilhança das observações e frequentemente é mais prático e mais simples utilizar o logaritmo da verossimilhança. A utilização do logaritmo não causa nenhuma alteração uma vez que o máximo do logaritmo é obtido no mesmo ponto da máxima verossimilhança. Assim, sendo $\mathbf{W} = w_1, w_2, \dots, w_n$, a inversa da matriz de mistura \mathbf{A} , o logaritmo da verossimilhança é dado por:

$$L = \sum_{t=1}^T \sum_{i=1}^m \log[f_i(\mathbf{W}_i^T x(k)) + N \log(| \det \mathbf{W} |)] \quad (17)$$

Onde $f_i(\cdot)$ é a função densidade de probabilidade de $s(k)$, considerada conhecida, e $x(k)$, $k = 1, \dots, N$, as observações ou medidas de $s(k)$. O termo $\log(| \det \mathbf{W} |)$ é derivado da regra clássica para transformação das variáveis aleatórias e suas densidades [71].

Em geral, para qualquer vetor aleatório $x(k)$ com densidade de probabilidade p_x e para uma qualquer matriz \mathbf{W} a densidade de $y(k) = \mathbf{W}x(k)$ é dada por 18.

$$p_y = p_x \mathbf{W}^{-1} y \log(\det \mathbf{W}^{-1}) \quad (18)$$

A abordagem de ICA baseada em MLE propõe calcular a matriz \mathbf{W} que maximize 18. A vantagem dos métodos que utilizam ML, é que, sob condições regulares, eles produzem resultados eficientes, bem definidos pela teoria da estimação [74].

Por outro lado, estes métodos apresentam também algumas desvantagens, como:

- Necessita-se do conhecimento *a priori* da densidade de probabilidade dos componentes independentes, o que leva ao emprego de métodos consideravelmente complexos;
- A solução pode ser sensível a *outlier*, dependendo da forma da f.d.p. escolhida

2.3.14 Informação Mútua

Uma outra abordagem para ICA, também baseada na teoria da informação, é feita através da minimização da informação mútua (I), que é uma forma natural de se medir a independência entre variáveis aleatórias [72].

A informação mútua (I) entre m variáveis aleatórias $y_i(k), i = 1...m$, pode ser definida a partir do conceito de entropia diferencial como mostrado em 19.

$$I(y_1, \dots, y_n) = \sum_{i=1}^m H(y_i) - H(y) \quad (19)$$

Onde $H(\cdot)$ é a entropia diferencial [72]. A informação mútua pode ser entendida utilizando a interpretação de entropia como comprimento de código, que diz que a entropia fornece, a grosso modo, o comprimento médio mínimo do código de uma variável aleatória. Desta forma, $H(y_i)$ fornece o comprimento do código para y_i , quando estas variáveis são codificadas separadamente e $H(y)$ fornece o comprimento do código quando y_i é codificado como um vetor aleatório, ou seja, todos, os componentes de y_i são codificados sob o mesmo código.

Com base na equação 19, pode-se definir a informação mútua entre os sinais medidos $x(k)$ e os sinais a serem estimados $y(k)$ pela equação 20.

$$I(y_1, \dots, y_n) = H(x) - H\left(\frac{x}{y}\right) \quad (20)$$

A informação mútua é uma forma natural de medida da dependência entre variáveis aleatórias sendo sempre não negativa e zero se, e somente se, as variáveis são estatisticamente independentes. Efetivamente isto ocorre uma vez que a observação de uma variável independente não fornece nenhuma informação sobre outra variável independente.

Pode-se verificar, a partir de 20 que para $I(x, y) = 0$, $H(x) = H\left(\frac{x}{y}\right)$. Desta forma, pode-se utilizar a informação mútua como um critério confiável para estimar um modelo para ICA minimizando a informação mútua entre os componentes [74].

Uma propriedade importante da informação mútua [71] é que para uma transformação linear inversível como $y(k) = \mathbf{W}x(k)$, pode-se escrever 21.

$$I(y_1, \dots, y_n) = \sum_{i=1}^m H(y_i) - H(y) - \log(| \det \mathbf{W} |) \quad (21)$$

Se $y(k)$ for considerada descorrelacionada e com variância unitária, tem-se que:

$$E[yy^T] = \mathbf{W}E[xx^T]\mathbf{W}^T = I \quad (22)$$

Onde \mathbf{I} é a matriz identidade. A equação 22 implica em:

$$\det I = \det(\mathbf{W}E[xx^T]\mathbf{W}^T) = (\det \mathbf{W})(\det E[xx^T])(\det \mathbf{W}^T) \quad (23)$$

A partir de 23, conclui-se que \mathbf{W} é constante. Além disso, para $y(k)$ com variância unitária, a entropia e a negentropia diferem somente por uma constante. Assim, pode-se escrever 21 como:

$$I(y_1, \dots, y_n) = C - \sum_i J(y_i) \quad (24)$$

Onde C é uma constante que não depende de \mathbf{W} e $J(\cdot)$ é a negentropia. A equação 24 define a principal relação entre negentropia e I [71]. Pela equação 24 pode-se concluir que encontrar a matriz \mathbf{W} que minimize a I é, a grosso modo, equivalente a encontrar a direção na qual a negentropia é maximizada. Pode-se observar ainda, a partir de 24, que a estimativa dos componentes independentes através da minimização da I é equivalente à maximização do somatório das não gaussianidades das estimativas, quando estas são consideradas descorrelacionadas; trata-se mais uma justificativa da idéia de encontrar a maximização da não gaussianidade para solucionar o problema da ICA.

2.3.15 Principais Algoritmos para ICA

Após a escolha do princípio de estimativa para ICA é necessário escolher um método de otimização para sua implementação. Neste contexto, existem vários algoritmos que implementam o ICA tais como: Algoritmo de Héroult-Jutten, Algoritmo de Descorrelação Não Linear, Algoritmo INFOMAX, Algoritmos Baseados em Gradiente Natural, PCA Não Linear, Algoritmos Baseados em Estatísticas de Ordem Superior, Algoritmo FastICA. Em nosso trabalho o algoritmo utilizado foi o Algoritmo FastICA.

2.3.16 Algoritmo FastICA

Os algoritmos adaptativos baseados em gradiente estocástico podem apresentar problemas quando utilizados em casos práticos, pois sua convergência torna-se lenta, além de ser dependente de uma escolha adequada do passo de adaptação. Uma escolha não adequada do passo de adaptação pode levar a problemas na convergência do algoritmo.

Para tentar minimizar este problema foi desenvolvida uma família de algoritmos baseados em iterações de ponto fixo denominada FastICA [72].

Os algoritmos desta família são diferenciados pela abordagem e pela função custo utilizada; todos eles, porém visam encontrar componentes independentes através da maximização da negentropia.

Com relação à abordagem, foram desenvolvidas duas versões do FastICA: a primeira permite a recuperação de todas as fontes (multi-unit) e a segunda, encontra um dos componentes (one-unit). Ambas abordagens podem utilizar qualquer função custo não paramétrica para estimar a negentropia das fontes [77]. O algoritmo inicialmente proposto utilizava kurtosis, sendo, posteriormente, generalizado para outras funções custo [72].

O algoritmo FastICA procura encontrar uma direção, ou seja, um vetor, cuja projeção $WTx(k)$ maximiza a função $JG(W)$; ou seja, uma aproximação da negentropia.

2.4 Seleção de Características

Faz-se necessário identificar as características mais importantes dentre um vetor de características observado visto que é uma das tarefas mais críticas encontradas em sistemas de reconhecimento de padrões. Dentro desta esfera de conhecimento tal atividade é considerada de essencial importância para diminuir o erro de classificação e o custo computacional [117].

Reportando-se a esta questão, vale mencionar as características irrelevantes que 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 supracitado.

E para elucidar melhor este aspecto: Seja v um vetor de dados, com n amostras e m características, representado por $v = \{v_i, i = 1, \dots, m\}$, e c um vetor de classe (rótulo).

Pode-se observar que o problema de seleção de características é encontrar do conjunto de observação de dimensão v , um subconjunto de m características que represente c , de maneira satisfatória e efetiva.

Com base nessas orientações dada a condição de encontrar um mapeamento "ótimo", o algoritmo deve buscar a melhor forma de encontrar este subconjunto. Há de convir que as duas formas mais comuns são: classificando-as por algum critério e selecionando as k melhores características, a segunda é escolher um subconjunto mínimo dentro do conjunto de características sem afetar a precisão da classificação.

Sendo assim, na seleção de um subconjunto eficiente, os algoritmos podem automaticamente determinar o número de características, ou o operador pode estipular o tamanho do subconjunto de características.

A condição de caracterização significativa implica em um erro de classificação mínimo possível, que requer a máxima dependência estatística entre o subconjunto m selecionado,

e o vetor de classe c . Tal esquema é chamado de Máxima Dependência.

Em termos de informação mútua I , que vem da teoria da informação, a proposta de realizar a seleção de características para encontrar um vetor v , com m características $\{v_i\}$, que conjuntamente tenham a maior dependência possível com o vetor de classe c , é dada por:

$$\max D(v, c), D = I(\{v_i, i = 1, \dots, m\}) \quad (25)$$

Entretanto, a equação 25 pode demandar um esforço computacional grande, quando os dados são multivariados, já que para estimar o vetor v , é necessário calcular inúmeras vezes a inversa da matriz de covariância. Por este motivo, o algoritmo para encontrar a máxima dependência entre variáveis é considerado de grande esforço computacional, ocasionando um custo computacional elevado.

Para diminuir o esforço computacional, utiliza-se uma técnica baseada em Máxima Relevância e Mínima Redundância (mRMR), que maximiza a informação mútua e minimiza a medida de redundância.

Máxima Relevância e Mínima Redundância

Uma das formas de selecionar características é por meio da Máxima Relevância descrita por:

$$\max M(\mathbf{a}, c), R = \frac{1}{|\mathbf{a}|} \sum_{a_i \in A} I(a_i, c) \quad (26)$$

Sendo v um vetor de características, c o vetor de classe e v_i uma característica individual. É provável que as características selecionadas de acordo com o critério descrito anteriormente tenham muita redundância, ou seja, a dependência entre estas características pode ser grande. Para resolver tal problema, aplica-se em conjunto, a condição de Mínima Redundância, que seleciona mutuamente apenas as características mutuamente exclusivas [Ding and Peng 2003], tem-se, portanto

$$\min R(\mathbf{a}), R = \frac{1}{|\mathbf{a}|^2} \sum_{a_i a_j \in v} I(a_i, a_j) \quad (27)$$

Os critérios descritos nas equações 26 e 26 são chamados conjuntamente de Máxima Relevância e Mínima Redundância (mRMR) [Peng, Long and Ding 2005]. Pode-se definir o operador $\phi(D, R)$ para combinar D e R, para em seguida otimizá-los simultaneamente, obtendo assim:

$$\max \Phi(D, R), \Phi = D - R \quad (28)$$

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.

2.5 Conceitos Básicos de Aprendizagem de Máquina

As técnicas de aprendizagem de máquina (AM) empregam um princípio de inferência denominado indução, no qual obtém-se conclusões genéricas a partir de um conjunto particular de exemplos. O aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não-supervisionado.

No aprendizado supervisionado tem-se a figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada, saída desejada [111]. O algoritmo de AM extrai a representação do conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente.

No aprendizado não-supervisionado não há a presença de um professor, ou seja, não existem exemplos rotulados. O algoritmo de AM aprende a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade. Essas técnicas são utilizadas principalmente quando o objetivo for encontrar padrões ou tendências que auxiliem no entendimento dos dados [112].

O tipo de aprendizado abordado neste trabalho é o supervisionado. Neste caso, dado um conjunto de exemplos rotulados na forma (x_i, y_i) , em que x_i representa um exemplo e y_i denota o seu rótulo, deve-se produzir um classificador, também denominado modelo, preditor ou hipótese, capaz de prever precisamente o rótulo de novos dados. Esse processo de indução de um classificador a partir de uma amostra de dados é denominado treinamento. O classificador obtido também pode ser visto como uma função f , a qual recebe um dado x e fornece uma predição y .

Os rótulos ou classes representam o fenômeno de interesse sobre o qual se deseja fazer previsões. Neste trabalho, considera-se o caso em que os rótulos assumem valores discretos $1, \dots, k$. Tem-se então um problema de classificação. Caso os rótulos possuam valores contínuos, tem-se uma regressão [113]. Um problema de classificação no qual $k = 2$ é denominado binário. Para $k > 2$, configura-se um problema multiclassismo.

Cada exemplo, também referenciado por dado ou caso, é tipicamente representado por um vetor de características. Cada característica, também denominada atributo, expressa um determinado aspecto do exemplo [114]. Normalmente, há dois tipos básicos de atributos: nominal e contínuo. Um atributo é definido como nominal (ou categórico) quando não existe uma ordem entre os valores que ele pode assumir (por exemplo, entre cores). No caso de atributos contínuos, é possível definir uma ordem linear nos valores assumidos (por exemplo, entre pesos $\in \mathbb{R}$).

Um requisito importante para as técnicas de AM é que elas sejam capazes de lidar com dados imperfeitos, denominados ruídos. Muitos conjuntos de dados apresentam esse tipo de caso, sendo alguns erros comuns a presença de dados com rótulos e/ou atributos incorretos. A técnica de AM deve idealmente ser robusta a ruídos presentes nos dados,

procurando não fixar a obtenção dos classificadores sobre esse tipo de caso. Deve-se também minimizar a influência de outliers no processo de indução. Os *outliers* são exemplos muito distintos dos demais presentes no conjunto de dados. Esses dados podem ser ruidos ou casos muito particulares, raramente presentes no domínio.

A obtenção de um classificador por um algoritmo de AM a partir de uma amostra de dados também pode ser considerada um processo de busca [113].

Procura-se, entre todas as hipóteses que o algoritmo seja capaz de gerar a partir dos dados, aquela com melhor capacidade de descrever o domínio em que ocorre o aprendizado.

Para estimar a taxa de predições corretas ou incorretas (também denominadas taxa de acerto e taxa de erro, respectivamente) obtidas por um classificador sobre novos dados, o conjunto de exemplos é, em geral, dividido em dois subconjuntos disjuntos: de treinamento e de teste. O subconjunto de treinamento é utilizado no aprendizado do conceito e o subconjunto de teste é utilizado para medir o grau de efetividade do conceito aprendido na predição da classe de novos dados.

Um conceito comumente empregado em AM é o de generalização de um classificador, definida como a sua capacidade de prever corretamente a classe de novos dados. No caso em que o modelo se especializa nos dados utilizados em seu treinamento, apresentando uma baixa taxa de acerto quando confrontado com novos dados, tem-se a ocorrência de um superajustamento (*overfitting*). É também possível induzir hipóteses que apresentem uma baixa taxa de acerto mesmo no subconjunto de treinamento, configurando uma condição de subajustamento (*underfitting*). Essa situação pode ocorrer, por exemplo, quando os exemplos de treinamento disponíveis são pouco representativos ou quando o modelo obtido é muito simples [114].

2.6 Máquina de Vetores de Suporte

2.6.1 Introdução

A Máquina de Vetor Suporte (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 [78]. Os fundamentos desta técnica foram desenvolvidos por Vapnik e colaboradores [79].

Esta técnica já foi utilizada com sucesso em diversas aplicações de reconhecimento de padrões, como por exemplo: categorização de textos, categorização de SPAM, reconhecimento de caracteres manuscritos, reconhecimento de textura, análise de expressões de genes, etc.

Basicamente o funcionamento de uma SVM pode ser descrito da seguinte forma: dadas duas classes e um conjunto de pontos que pertencem a essas classes, uma SVM determina o hiperplano que separa os pontos de forma a colocar o maior número de pontos da mesma classe do mesmo lado, enquanto maximiza a distância de cada classe a esse hiperplano.

A distância de uma classe a um hiperplano é a menor distância entre ele e os pontos dessa classe e é chamada de margem de separação. O hiperplano gerado pela SVM é determinado por um subconjunto dos pontos das duas classes, chamado vetores suporte.

As SVMs vêm sendo aplicadas com sucesso a diversos problemas computacionais [86].

Existem diversas aplicações das SVMs, tais como: Visão Computacional, Reconhecimento de Dígitos, Bioinformática. Este trabalho é em Bioinformática.

2.6.2 Bioinformática

Por Bioinformática entende-se a aplicação de técnicas computacionais, envolvendo desde o armazenamento de informações à análise das mesmas, no gerenciamento de informações biológicas [116]. Os dados neste domínio em geral são volumosos e complexos (possuem grande número de atributos, ou seja, grande dimensão), o que torna o uso das SVMs adequadas. Seguem algumas das principais aplicações de interesse em Bioinformática e como as SVMs forma empregadas em cada uma delas:

- **Identificação de genes:** A identificação de genes envolve o reconhecimento e a localização de cada gene em sequências de DNA. Uma abordagem comumente utilizada na realização desta tarefa envolve determinar sinais que podem ser identificados nas sequências e que indicam situações especialmente adequadas á localização de genes [87]. Um exemplo de sinal é o sítio de início de tradução (SIT) em sequências de mRNA (Ácido Ribonucleico mensageiro). Zien [88], cita a utilização de SVMs no reconhecimento desses sítios. são obtidos na comparação entre SVMs e ADs.
- **Estrutura de proteínas:** As funções e propriedades de uma proteína são determinadas por sua estrutura tridimensional [92]. Esta estrutura, por sua vez, é definida pela sequência de genes que codificam a proteína em questão. Uma área de pesquisa que tem despertado grande interesse é a análise dos dados genéticos sequenciados com o *m* de determinar a estrutura das proteínas codificadas. Estudos semelhantes são conduzidos para determinar a estrutura tridimensional de moléculas de *Ribonucleic Acid* (RNA). Uma das tarefas envolvidas na previsão da estrutura de proteínas é o reconhecimento de tipos de dobras que estas possuem. Métodos para comparação de sequências como o *pairwise* são comumente utilizados nessa tarefa, buscando sequências similares e atribuindo o tipo de dobra destas à proteína sendo consultada. Porém, existem bases com menos de 35% de semelhança entre as proteínas em nível sequencial. Para estes casos, o uso de algoritmos de AM tem se mostrado mais adequado.
- **Análise de Expressão Gênica:** O objetivo neste caso é analisar o nível de expressão de diferentes genes (a quantidade de proteínas que este produz). Com isto pode-se, por exemplo, comparar mudanças na expressão de alguns genes frente a

alguma doença, possibilitando assim identificar alvos para futuros medicamentos e/ou terapias gênicas[94]. Muitos trabalhos utilizam SVMs na análise de dados provenientes de experimentos de DNA microarray, técnica de laboratório utilizada para medir o nível de expressão de genes. Esses dados caracterizam-se por possuir uma grande dimensionalidade (possuem medidas de milhares de genes). As SVMs, portanto, são apropriadas a essa tarefa. A medida do nível de expressão dos genes em geral é útil no diagnóstico de moléstias, por meio da identificação de genes relacionados a uma determinada doença. Furey [94] aplicaram SVMs sobre dados de *Deoxyribonucleic AcidDNA microarray* para classificação de tecidos cancerosos e, por meio de adaptações no algoritmo original, determinaram exemplos do conjunto de treinamento que continham erros em sua classificação, que é usualmente feita a priori por um Biólogo. Os resultados observados se mostraram comparáveis a outros métodos, tais como as RNAs [95]. Em um experimento semelhante, [96] aplicaram SVMs, *AdaBoost* e clusterização sobre dados de *microarray* para classificação de tumores, sendo os resultados obtidos também promissores. Brown [97], por sua vez, utiliza as SVMs sobre dados de DNA *microarray* para classificação de genes. Experimentos sobre esses dados sugerem que genes com função similar possuem padrão de expressão semelhante. Nos estudos conduzidos as SVMs se mostraram superiores a outras técnicas de AM, tais como as Árvores de Decisão e o *Kernel Fisher Discriminant*.

2.6.3 Conjuntos Linearmente Separáveis

Um conjunto de treinamento S é linearmente separável, se for possível separar os padrões das classes diferentes contidos no mesmo por pelo menos um hiperplano [98]. Classificadores que separam os dados por meio de um hiperplano são denominados lineares, podendo ser definidos por: 29.

$$w.x + b = 0 \quad (29)$$

Onde $w.x$ é o produto escalar entre os vetores w e x , em que w é o vetor normal ao hiperplano e b é um termo "compensador". O par (w, b) é determinado durante o treinamento do classificador. Esta equação divide o espaço de entradas em duas regiões: $w.x + b > 0$ e $w.x + b < 0$,

levando à 30 [99].

$$A = \begin{cases} y_i = +1 & \text{se } wx_i + b > +1 \\ y_i = -1 & \text{se } wx_i + b < -1 \\ i = 1, \dots, n \end{cases} \quad (30)$$

Uma função sinal $g(x) = \text{sgn}(f(x)) = \text{sgn}(w.x + b)$, pode ser aplicada sobre 30,

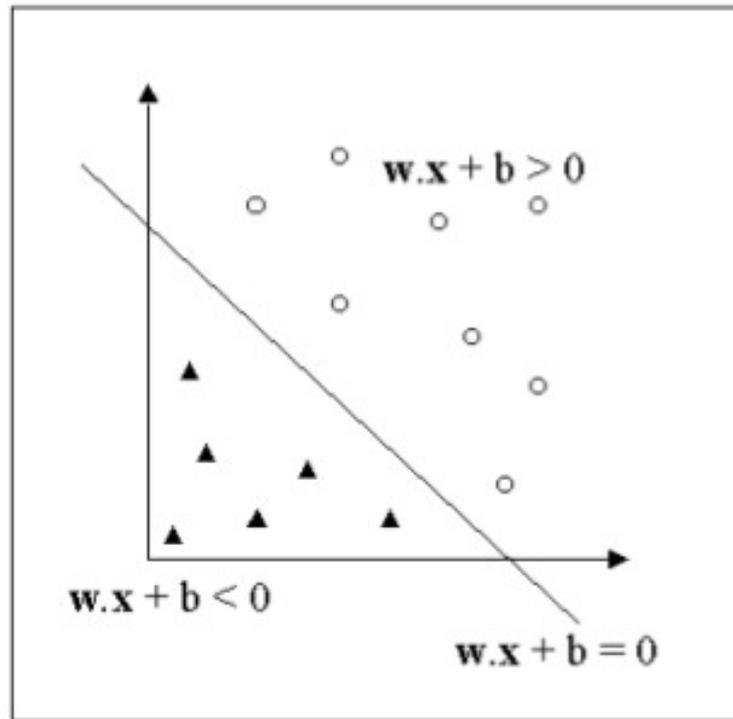


Figura 3: Conjunto de dados linearmente separável e um possível hiperplano separador. Fonte [85]

levando à classificação $+1$ se $f(x) > 0$ e -1 se $f(x) < 0$.

Logo, um conjunto de treinamento é linearmente separável se é possível determinar pelo menos um par (w, b) tal que a função sinal $g(x) = \text{sgn}(f(x))$ consiga classificar corretamente todos os exemplos contidos neste grupo. Este fato pode ser observado na Figura 3

2.6.4 SVMs Lineares

Na aplicação da Teoria de Aprendizado Estatístico (TAE), deve-se escolher o classificador com o menor risco empírico possível e que também satisfaça a restrição de pertencer a uma família F com dimensão VC pequena. A primeira condição, no caso de conjuntos de treinamento linearmente separáveis, é satisfeita para pelo menos um par (w, b) definido por 30. Em geral, há diversas combinações de valores de w e b capazes de separar corretamente os dados deste tipo de conjunto. Para satisfazer a segunda restrição, utiliza-se o resultado da TAE que relaciona o risco funcional de uma função à margem (ρ) de separação entre os dados de treinamento e o hiperplano separador. Uma definição formal do conceito de margem é apresentada a seguir [100].

Seja f uma hipótese utilizada para classificação de entradas na forma (x_i, y_i) , em que y_i representa a classe do padrão x_i . Então a Equação 31 define a margem com a qual o padrão x_i é classificado. A margem γ de um classificador é dada então por 32.

$$\rho f(x_i, y_i) = y_i f(x_i) \quad (31)$$

$$\rho = \min(y_i f(x_i)) \quad (32)$$

Portanto, entre os classificadores que minimizam o risco empírico, deve-se escolher aquele que possui a maior margem ρ' . O hiperplano que possui esta margem ρ' é denominado ótimo.

2.6.5 Determinação do Hiperplano Ótimo

Esta Seção apresenta os passos seguidos por uma SVM linear na determinação do hiperplano ótimo para conjuntos linearmente separáveis.

Como assumiu-se que o conjunto de treinamento é linearmente separável, pode-se reescalar w e b de forma que os pontos mais próximos do hiperplano separador satisfaçam $|wx + b = 1|$ [99]. Obtém-se com isto a representação canônica do hiperplano, adotada para facilitar as considerações subsequentes realizadas na determinação do hiperplano ótimo.

A partir dessa consideração, a desigualdade 33 caracteriza os classificadores lineares que separam o conjunto de treinamento com uma margem positiva. Segundo este sistema, não há pontos entre $|wx + b = 0|$ e $|wx + b = \pm 1|$, ou seja, supõe-se que a margem ρ é sempre maior que a distância entre os hiperplanos $wx + b = 0$ e $|wx + b = 1|$ [101]. Devido a essa suposição, as SVMs obtidas são usualmente denominadas SVMs com margens rígidas.

$$B = \begin{cases} wx_i + b \geq +1, & \text{se } y_i = +1 \\ wx_i + b \leq -1, & \text{se } y_i = -1 \\ i = 1, \dots, n \end{cases} \quad (33)$$

Sejam x_1 um ponto sobre $w.x + b = -1$ e x_2 um ponto sobre $w\Delta x + b = +1$ (Equação 34). Supõe-se também que x_1 intercepta a reta perpendicular a x_2 , como indicado na Figura 4 [86]. Tem-se assim que:

$$B = \begin{cases} wx_1 + b = -1 \\ wx_2 + b = 1 \end{cases} \quad (34)$$

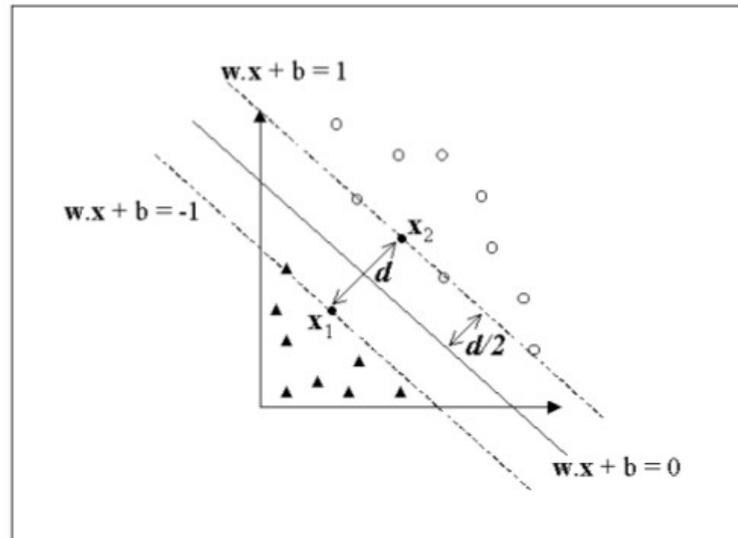


Figura 4: Ilustração da distância d entre os hiperplanos. Fonte [85]

A partir de 34, obtém-se 35.

$$w\Delta(x_2 - x_1) = 2 \quad (35)$$

Como w e $x_2 - x_1$ são ortogonais ao hiperplano separador, esses vetores são paralelos entre si. Pode-se desta forma deduzir 36.

$$|w.(x_2 - x_1)| = \|w\| \|x_2 - x_1\| \quad (36)$$

em que $\|\cdot\|$ representa a norma de um vetor. Substituindo 35 em 36, obtém-se 37.

$$\|x_2 - x_1\| = \frac{2}{\|w\|} \quad (37)$$

A norma $\|x_2 - x_1\|$ mede a distância entre os hiperplanos $w.x + b = -1$ e $w.x + b = +1$. Logo, por 37 pode-se afirmar que a distância entre os mesmos é dada por $\frac{2}{\|w\|}$. Da mesma forma, pode-se deduzir que a distância entre os hiperplanos $w.x + b = 0$ e $w.x + b = 1$ ou ($w.x + b = -1$) é dada por $\frac{2}{\|w\|}$. Como foi suposto que a margem é sempre maior que esta última distância, a minimização de $\|w\|$ leva a uma maximização da margem.

Portanto, o hiperplano ótimo é definido para os valores de w e b que satisfazem as desigualdades 33 e para os quais a norma $\|w\|$ é mínima. Este é um problema de otimização com restrições e pode ser reescrito como:

$$\text{Minimizar : } \|w\|^2 \quad (38)$$

Sob as restrições: $y_i(w\Delta x_i + b) \geq 1$, para $i = 1, \dots, n$

Este é um problema clássico em otimização denominado programação quadrática [86]. O problema de AM é então reduzido a uma forma que pode ser analisada sob o ponto

de vista de otimização de função quadrática, para o qual há uma ampla e estabelecida teoria existente. Na resolução do mesmo, introduz-se uma função Lagrangiana, definida em termos de w e b , apresentada em 39 [101].

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w \Delta x_i + b) - 1) \quad (39)$$

em que os α_i são denominados multiplicadores de Lagrange. O problema passa a ser então a minimização de 39 em relação a w e b e a maximização dos α_i . Os pontos ótimos desta equação são obtidos por meio da resolução das igualdades apresentadas em 40 e 41.

$$\frac{\partial L}{\partial b} = 0 \quad (40)$$

$$\frac{\partial L}{\partial w} = 0 \quad (41)$$

As Equações 40 e 41 levam ao resultado representado por 42 e 43.

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (42)$$

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (43)$$

Substituindo 42 e 43 em 39, obtém-se o seguinte problema de otimização (denominado dual):

$$\text{Maximizar} : \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j$$

$$\text{Sujeito a} : \begin{cases} \alpha_i \geq 0, & i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

Calculado o valor ótimo α^* , w^* é encontrado por 43. Sendo determinada a direção do hiperplano ótimo (ou seja, w^*), b^* pode ser calculado por 44.

$$b^* = \frac{-1}{2} [\text{Max}_{i|y_i=-1}(w^* x_i) + \text{Min}_{i|y_i=+1}(w^* x_i)] \quad (44)$$

Resumindo os procedimentos matemáticos realizados, percebe-se que para fazer a determinação do hiperplano ótimo, tenta-se encontrar os valores de α^* para utilizá-los na determinação de (w^*, b^*) . É demonstrado que α_i^* assume valores positivos para exemplos do treinamento que estão a uma distância do hiperplano ótimo exatamente igual à margem. Para os outros padrões, o valor de α_i^* é nulo. Essa propriedade caracteriza uma esparsividade da solução [102].

Os exemplos para os quais $\alpha_i^* > 0$ são denominados vetores suporte (Support Vectors - SVs). A variável w^* pode então ser reescrita na forma indicada em 45.

$$w^* = \sum_{x_i \in SV}^n \alpha_i y_i x_i \quad (45)$$

Como consequência, o hiperplano ótimo é determinado unicamente pelos SVs, que por este motivo são considerados os exemplos mais informativos do conjunto de treinamento. Os padrões para os quais $\alpha_i = 0$ não participam na definição do hiperplano ótimo. Logo, se apenas o subconjunto dos dados de treinamento formado pelos SVs fossem empregados na indução, o classificador obtido seria o mesmo gerado com a utilização de todo o conjunto.

Fixados os valores dos parâmetros w^* e b^* por meio do treinamento de uma SVM, o classificador extraído é definido em 46.

$$g(x) = \text{sgn}(f(x)) = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i^* y_i x_i x + b^*\right) = \begin{cases} +1, & \text{se } \sum_{x_i \in SV} \alpha_i^* y_i x_i x + b^* > 0 \\ -1, & \text{se } \sum_{x_i \in SV} \alpha_i^* y_i x_i x + b^* < 0 \end{cases} \quad (46)$$

A partir de 46, verifica-se que a classificação de um novo padrão x requer apenas a computação do produto interno entre x e cada SV.

2.6.6 SVMs Não Lineares

Os classificadores lineares são bastante simples. Entretanto, sua utilização é limitada. Há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano. Um exemplo é apresentado na Figura 5 a), em que o uso de uma fronteira curva na separação das classes seria mais adequada. Este Capítulo descreve como as SVMs lineares podem ser generalizadas para lidar com tais situações.

Na realização desta tarefa, são definidas funções reais Φ_1, \dots, Φ_M no domínio do espaço de entrada. Por meio da utilização destas funções, mapeia-se o conjunto de treinamento S conforme 47 para um novo espaço, denominado espaço de características, da maneira apresentada em 48.

$$S = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\} \text{ em que } X_i = (x_{i1}, x_{i2}, \dots, x_{in}) \text{ } y_i \in \{-1, +1\} \quad (47)$$

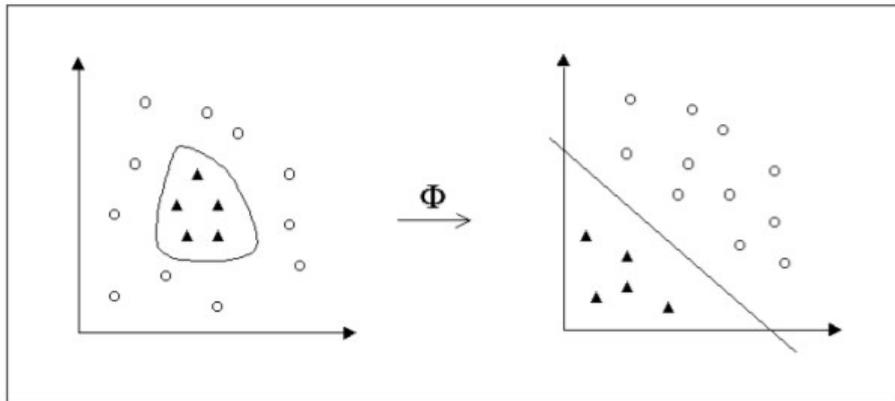


Figura 5: Transformação do conjunto de dados no espaço de entrada (a) para o espaço de característica (b). Fonte [85]

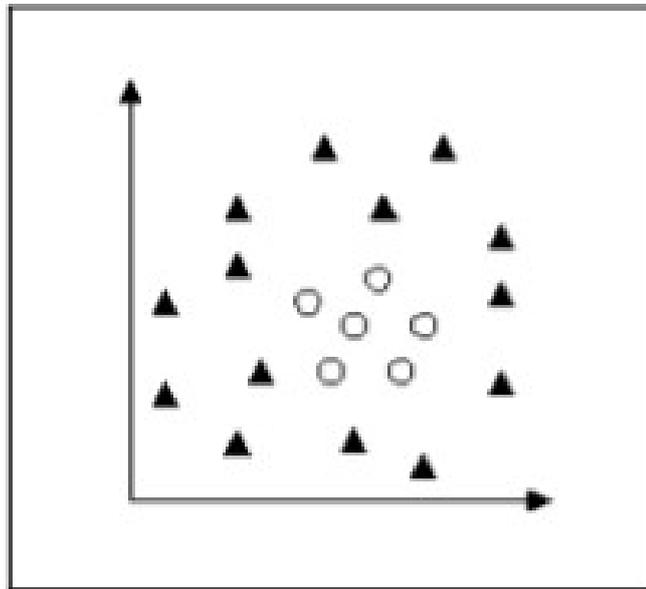


Figura 6: Exemplo de conjunto não linearmente separável. Fonte [85]

$$X_i (i = 1, \dots, n) \mapsto \Phi(X_i) = (\Phi_1(X_i), \Phi_2(X_i), \dots, \Phi_n(X_i)) \Rightarrow \Phi(S) = \{(\Phi(X_n), y_n)\} \quad (48)$$

Estas funções Φ_i podem ser não-lineares. Em particular, M pode ser muito maior que m , a dimensão do espaço de x . Uma característica singular do espaço de características que a escolha de uma função Φ apropriada torna o conjunto de treinamento linearmente separável. As SVMs lineares anteriormente apresentadas podem então ser utilizadas sobre o conjunto de treinamento mapeado nesse espaço[86]. Este fato é ilustrado na Figura 5.

Considere, por exemplo, os dados representados na Figura 6 [103]. Neste caso um limite circular é mais adequado para separação das classes. A equação 48 define essa fronteira.

$$x_1^2 + x_2^2 < r^2 \quad (49)$$

Utilizando a transformação definida pela equação 50 sobre a equação 49, tem-se como resultado uma classificação linear em \mathfrak{R}^3 . Esta é descrita por 51, em que $w = (1, 0, 1)$ e $b = r^2$.

$$\Phi(X) = (x_1^2, x_1x_2, x_2^2) \quad (50)$$

$$w \cdot \Phi(X) + b = 0 \quad (51)$$

Portanto, mesmo que a fronteira entre as classes seja polinomial no espaço de entradas, este problema pode ser mapeado para outro de classificação linear no espaço de características. Logo, os conceitos apresentados para o caso linear podem ser estendidos para o caso não linear, através da definição de uma função de mapeamento ϕ adequada.

2.6.7 Trabalhando com SVMs lineares no espaço de características

Como os dados no espaço de características podem ser eficientemente separados por meio de fronteiras lineares, é possível utilizar o SVM linear para separar os padrões neste novo espaço.

2.6.8 Determinação do hiperplano ótimo no espaço de características

Para determinar do hiperplano ótimo no espaço de características, é necessário fazer adaptações da SVM linear, apresentadas a seguir:

- Para qualquer conjunto de treinamento $\phi(S) = (\phi(x_1), y_1), \dots, (\phi(x_n), y_n)$
- Seja $\alpha = (\alpha_1^*, \dots, \alpha_n^*)$ a solução do seguinte problema de otimização com restrições:

- Maximizar: $\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j \alpha_i \alpha_j \Phi(X_i) \Phi(X_j)$

- Sob as restrições:

$$\begin{cases} \sum_{i=1}^n y_i \alpha_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n \end{cases}$$

- O par (w^*, b^*) apresentado a seguir define o hiperplano ótimo.

- $w^* \leftarrow \sum_{i=1}^n y_i \alpha_i^* \Phi(X_i)$

-

$$b^* \leftarrow \frac{-1}{2} [Max_{i|y_i=-1}(w^* x_i) + Min_{i|y_i=+1}(w^* x_i)]$$

O problema de classificação passa a ser a avaliação da função ??.

Além disso, as considerações de Karish-Kuhn-Tucker apresentadas para o SVM linear se mantêm para o caso não linear. A partir das equações listadas acima, percebe-se que a única informação necessária sobre o mapeamento ϕ é uma definição de como o produto interno $\phi(x_i) \cdot \phi(x_j)$ pode ser realizado, para quaisquer x_i e x_j pertencentes ao espaço de entradas. Isto é obtido com a introdução do conceito de Kernels.

2.6.9 Funções Kernel

Um Kernel K uma função que recebe dois pontos x_i e x_j do espaço de entradas e computa o produto escalar $\phi(x_i)\phi(x_j)$ no espaço de características, como descrito em 52 [104].

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j) \quad (52)$$

Como exemplo tem-se que o *Kernel* do mapeamento representado pela Equação 50 é dado por $K(x_1, x_2) = (x_1 \cdot x_2)^2$ [86]. Pode-se observar que esta função é mais simples que a função Φ . Isso ocorre de maneira geral, pois em grande parte dos casos, Φ assume formas bastante complexas. Por este motivo, é comum definir a função *Kernel* sem conhecer-se explicitamente o mapeamento Φ . A utilidade dos Kernels está, portanto, na simplicidade do cálculo e na capacidade de representar espaços muito abstratos.

As funções Φ devem pertencer a um domínio em que seja possível o cálculo de produtos internos. Há um conjunto de funções com esta propriedade, que pertencem ao espaço de *Hilbert* [105]. Este conhecimento é utilizado na definição dos *Kernels*. Geralmente faz-se uso das condições estabelecidas pelo Teorema de Mercer nesta definição. Segundo este Teorema, os *Kernels* devem ser matrizes positivamente definidas, isto é, a matriz \mathbf{K} , em que $\mathbf{K}_{ij} = \mathbf{K}(x_i, x_j)$ para todo $i, j = 1, \dots, n$, deve ter auto-valores maiores que 0 [106]. Essas funções são usualmente denominadas *Kernels de Mercer* [107].

Alguns dos Kernels mais utilizados são os polinomiais, os Gaussianos ou RBF (*RadialBasis Function*) e os Sigmoidais, apresentados na Tabela 2.6.9.

Tipo de Kernel	Função $K(x_i, x_j)$	Comentários
Polinomial	$(x_i^T \cdot x_j + 1)^p$	A pot. p deve ser especific. pelo usuário
Gaussiano	$\exp(-\frac{1}{\sigma^2} \ x_i - x_j\ ^2)$	A amplit. σ^2 é especific. pelo usuário
Sigmoidal	$\tanh(\beta_0 \cdot x_i \cdot x_j + \beta_1)$	Util. somente valores de β_0 e β_1

Alguns pontos podem ser destacados em relação aos *Kernels* a [103]:

- No caso de *Kernels* polinomiais, os mapeamentos Φ também são funções polinomiais com complexidade crescente a medida que o expoente p aumenta.
- O *Kernel* Gaussiano corresponde a um espaço de características de dimensão infinita. Pode-se afirmar que quase todas formas de mapeamento podem ser implemen-

tadas por esta função em particular. Além disso, por meio da utilização desse tipo de função, define-se Redes Neurais do tipo RBF (*Radial-Basis Function*), em que o número de funções base radiais e seus centros são determinados automaticamente pelo número de SVs e pelos valores de multiplicadores de Lagrange (α_i) associados aos mesmos, respectivamente.

- A utilização do *Kernel Sigmoidal*, por sua vez, permite explicitar uma Rede Neural Artificial do tipo Perceptron multicamadas. Neste caso, o número de neurônios da camada intermediária e os vetores de *bias* associados aos mesmos também são determinados pelo número de SVs e pelos valores dos (α_i) associados aos mesmos, respectivamente.

A obtenção de um classificador por meio do uso de SVMs envolve a escolha de uma função *Kernel*, além de parâmetros desta função e do algoritmo para determinação do hiperplano ótimo (como o valor da constante C, por exemplo). A escolha do Kernel e dos parâmetros considerados tem efeito no desempenho do classificador obtido [99], pois eles definem a fronteira de decisão induzida. Existem algumas técnicas para seleção do modelo, que provêm meios para determinação da função Kernel e parâmetros do algoritmo. Entre elas, tem-se a regra span [108], a regra de Jaakola e Haussler [109]) e a regra proposta em [110].

2.6.10 Validação cruzada (VC)

Os algoritmos de classificação buscam, de um modo geral, aprender com o passado para poderem generalizar sobre o futuro [110]. Dessa forma, o processo de aprendizagem tem grande importância já na sua inicial, ou seja, na seleção do conjunto de exemplos para o treinamento (teste) e também na escolha da parametrização usada pela função de kernel escolhida. Nesse contexto, a técnica validação cruzada (*cross-validation*) é bastante conhecida e utilizadas na estatística.

Na última fase deste trabalho, foi realizada a classificação das amostras utilizando-se a Máquina de Vetores de Suporte (SVM). Esta 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 nossos resultados, utilizamos o Cross-Validation (validação cruzada) [118] que é 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* [118]. 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.

3 Método e Materiais

3.0.1 Método Proposto

O método proposto é descrito pelo diagrama em blocos mostrado na Figura 7. Este método tem por finalidade diagnosticar pacientes com cardiotoxicidade. Para este fim, consiste em: extrair as características de sinais proteômicos utilizando a Análise de Componentes Independentes (ICA), para realizar a redução de dimensionalidade da matriz de característica gerada usa-se a técnica de Máxima Relevância e Mínima Redundância (mRMR) e por fim, a classificação é feita através da Máquina de Vetores de Suporte (SVM).

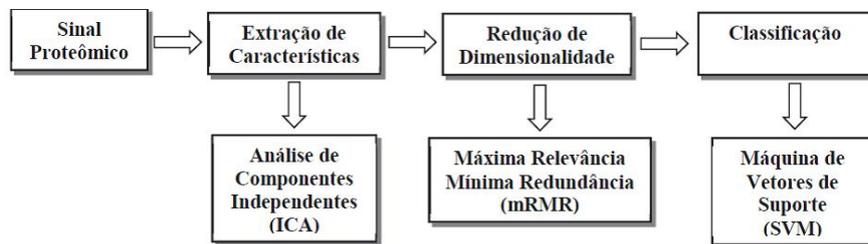


Figura 7: Diagrama do método.

3.1 Base de Dados

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 [32], vêm sendo desenvolvidos. Para desenvolver estes marcadores são utilizados conceitos como o da Espectrometria de Massa.

Neste contexto 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. E nesta perspectiva gerar listas de proteínas que aumentam ou diminuem em expressão como causa ou consequência de patologia [32]. A natureza desta informação pode nos levar a causa ou a uma consequência, de processos de doenças tal como a 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 [32].

Os dados usados neste trabalho foram baseados em padrões proteômicos usando a técnica de espectrometria de massa, que se mostrou ser um padrão de informação preciso para auxiliar no diagnóstico de pacientes com Cardiotoxicidade. As amostras utilizadas no trabalho foram adquiridas no site [119]. Esta base de dados contem 28 amostras com diagnóstico positivo para a Cardiotoxicidade e 34 amostras com diagnóstico normal (controle), cada amostra possui 373257 níveis de intensidade ou características diferentes.

A Figura 8 ilustra a amostra que foi extraída através de um espectrômetro de massa e que foi posteriormente convertida em um sinal multinível através dos níveis de intensidade proteômicos encontrados no espectro de massa.

3.2 Extração de Características

O objetivo desta etapa é obter parâmetros a partir de sinais proteômico que descrevam e discriminem os diferentes subgrupos de informações capazes de identificar determinadas doenças, tais como a Cardiotoxicidade. As características selecionadas devem garantir que as amostras de um paciente sejam corretamente classificadas como portadora ou não da Cardiotoxicidade.

3.3 Seleção das características mais significantes

Faz-se necessário identificar as características mais importantes dentre um vetor de características observado visto que é uma das tarefas mais críticas encontradas em sistemas

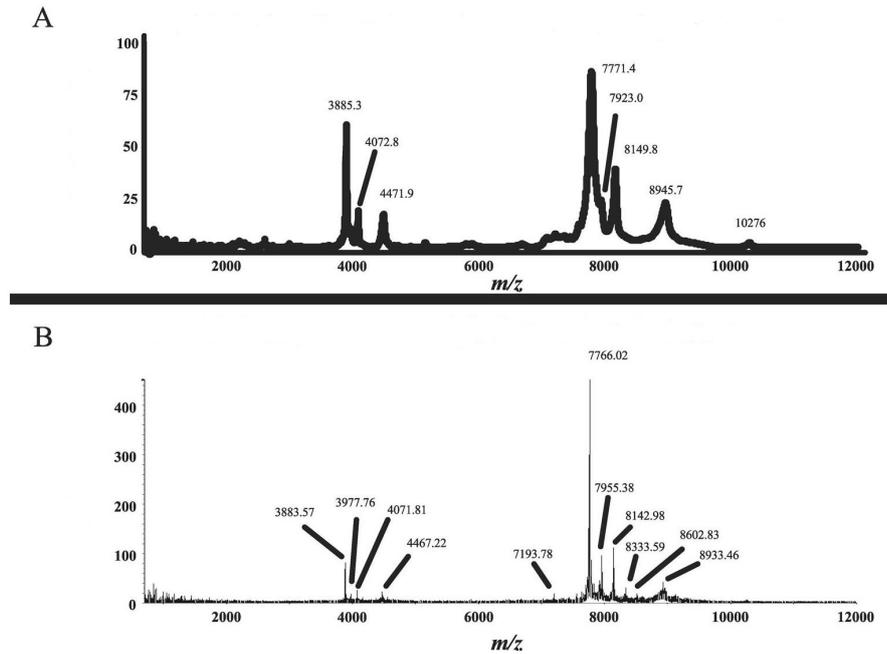


Figura 8: Espectro de Massa. Comparação entre baixa resolução e alta resolução de um espectro de massa SELDI-TOF. Sendo, a figura 2.A corresponde a um espectro de massa de alta resolução e a figura 2.B a um de baixa resolução. O eixo vertical corresponde ao nível de intensidade do espectro de massa, enquanto o eixo horizontal corresponde à razão massa / carga. Fonte: [32]

de reconhecimento de padrões[117]. Dentro desta esfera de conhecimento tal atividade é considerada de essencial importância para diminuir o erro de classificação e o custo e computacional [117].

Reportando-se a esta questão, vale mencionar que as características irrelevantes podem ser removidas sem comprometer o resultado da classificação, pois neste contexto, são consideradas redundantes, ou seja, existe outra característica com a mesma função, e não trazem nenhuma informação nova ao vetor supracitado. Esta etapa será realizada pelo algoritmo de Máxima Relevância e Mínima Redundância (mRMR).

3.4 Avaliação do Método de Classificação

Este trabalho é baseado em sinais biomédicos e reconhecimento de padrões, e a metodologia de desempenho usual é medida calculando algumas medidas estatísticas sobre o resultado dos testes [120]. Os resultados podem ser divididos em: Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN).

Sendo VP e VN o número de amostras que são corretamente identificadas, FP e FN representam o número de amostras correspondentes aos casos que são classificados erroneamente.

4 Resultados e Discussões

4.1 Aquisição de Dados

Esta base de dados consiste em 28 amostras com diagnóstico positivo para a Cardiotoxicidade e 34 amostras com diagnóstico normal (controle). Cada amostra tinha inicialmente 373257 níveis de intensidade ou características diferentes, no entanto para melhorar os resultados, reduziu-se a amostra para 100000 níveis de intensidade ou características diferentes como ilustrado na Figura 9.

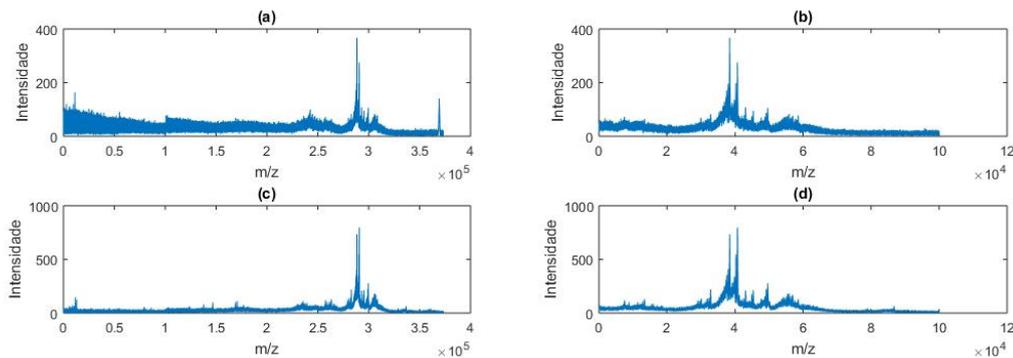


Figura 9: Espectro de Massa .

Na figura 9 verifica-se em (a) o sinal proteômico original de um paciente portador da Cardiotoxicidade contendo 373257 níveis de intensidade ou características diferentes e em (b) mostra o sinal representado em (a) já reduzido para 100000 níveis de intensidade ou características diferentes. Enquanto, (c) ilustra o sinal proteômico de um paciente não portador da Cardiotoxicidade no tamanho original e em (d) o sinal representado em (c) já reduzido para 100000 níveis de intensidade ou características diferentes.

4.2 Extração de Características

Para extrair as características que distinguem os casos com Cardiotoxicidade dos casos do grupo controle, foi utilizada a Análise de Componentes Independentes (ICA), para esse fim criou-se a matriz \mathbf{X} de ordem 62 por 373257 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 contem as característica de cada uma das amostras, e esta matriz é denominada matriz \mathbf{W} de ordem 62 por 62. Sendo que cada linha desta matriz \mathbf{W} corresponde a uma amostra, e cada coluna corresponde a uma característica. Desta maneira o classificador terá um parâmetro para distinguir os pacientes com Cardiotoxicidade dos pacientes sem a doença.

4.3 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 fossem utilizados na entrada do classificador, poderíamos ter resultados insatisfatórios, com baixa acurácia e grande esforço computacional. E para selecionar as características que melhor representem 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. Onde reduziu-se as amostras que antes tinham 373257 para 100000 níveis de intensidade, sendo que para encontrar o vetor de melhor desempenho, cada vetor gerado foi testado com a Máquina de Vetores de Suporte (SVM).

4.4 Classificação

Na última etapa, foi utilizada a SVM como classificador das amostras dos pacientes em controle e pacientes com Cardiotoxicidade. Utilizou-se 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 o teste de validação cruzada 10-fold cross-validation.

Tabela 3: Resultados

Caract.	VP	FP	VN	FN	Especif.(E)(%)	Sensib.(S)(%)	Acurácia(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

Foram realizados vários testes para verificação da eficácia dos resultados demonstrados na tabela 3, 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 como exposto pela tabela 3. Baseado nos resultados da tabela 3, verifica-se que com o vetor de 20 características das 62 possíveis, o método obteve 88,718% de acurácia, 85,000% de especificidade e 97,260% de sensibilidade.

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 apenas 1 foi classificado incorretamente(normal (FN)). Dos 28 casos com diagnóstico normal, somente em 10 casos houve erro de classificação(FP), diagnosticando-os positivo para a Cardiotoxicidade.

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, e é definida por:

$$A = (VP + VN)/(VP + VN + FP + FN) \quad (53)$$

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

$$E = VN/(VN + FP) \quad (54)$$

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

$$S = VP/(VP + FN) \quad (55)$$

5 conclusão

Este artigo propõe um método computacional para fazer o diagnóstico precoce em pacientes que estão sujeitos a desenvolver a Cardiotoxicidade, através da classificação de padrões proteômicos, utilizando: Análise de Componentes Independentes (ICA), Algoritmo de Máxima Relevância e Mínima Redundância (mRMR) e a Máquina de Vetores Suporte. O método proposto utilizou o Algoritmo FestICA para extrair um conjunto de características de um espectro de massa obtido a partir da espectrometria de massa. Percebeu-se que a seleção das características mais relevantes e menos redundantes, obtida pela mRMR, nos possibilitou reduzir a dimensionalidade da matriz de características e isso diminuiu o esforço computacional da máquina. Verificou-se que os resultados na classificação do grupo controle(normal) e do com a doença, obtidos pela SVM, foram satisfatórios.

Considerando que padrões proteômicos podem ser uma solução para a detecção precoce da Cardiotoxicidade, pois o conjunto das técnicas utilizadas neste trabalho permitiu que tivéssemos bons resultados. Desta forma, objetivo aqui é contribuir com um método que seja eficaz na detecção da Cardiotoxicidade.

Os resultados demonstraram que o conjunto de técnicas utilizadas neste trabalho conseguiu classificar os padrões proteômicos dos grupos controle (normal), e portadores de Cardiotoxicidade. Verificou-se que este método alcançou um bom desempenho, pois com um vetor de 20 características, obteve-se uma acurácia de 88,718%, com especificidade de 85% e sensibilidade de 97,26%, em um estudo que utilizou 62 amostras com baixa resolução.

No entanto, diante desse elenco se a base de dados fosse maior, este método poderia ter obtido um desempenho melhor na acurácia, especificidade e sensibilidade. Reportando-se ainda sobre esta questão, há de se convir que isto pode ser feito em trabalhos futuros que contenham uma base de dados maior.

REFERÊNCIAS

- [1] MORAES, A. d. J. P. Viabilidade do treinamento Físico Aeróbio por pacientes com Câncer Hematológico antes do transplante Autólogo de Células-Tronco Hematopoiéticas. 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 Cardio-Oncologia Pediátrica 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] I Diretriz Brasileira de Cardio-Oncologia Pediátrica da Sociedade Brasileira de Cardiologia
- [6] MORAES, A. d. J. P. Viabilidade do treinamento Físico Aeróbio por pacientes com Câncer Hematológico antes do transplante Autólogo de Células-Tronco Hematopoiéticas. Tese (Mestrado em Ciências do Movimento Humano) Universidade de Estado de Santa Catarina, Florianópolis, 2014.
- [7] 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.
- [8] I Diretriz Brasileira de Cardio-Oncologia Pediátrica da Sociedade Brasileira de Cardiologia.
- [9] 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.
- [10] Zhang YW, Shi J, Li YJ, Wei L. Cardiomyocyte death in doxorubicin-induced cardiotoxicity. Arch Immunol Ther Exp (Warsz). 2009;57(6):435-45.
- [11] Albin A, Pennesi G, Donatelli F, Cammarota R, De Flora S, Noonan DM. Cardiotoxicity of anticancer drugs: the need for cardio-oncology and cardiooncological prevention. J Natl Cancer Inst. 2010;

- [12] de Matos Neto RP, Petrilli AS, Silva CM, Campos Filho O, Oporto VM, Gomes Lde F, et al. Função sistólica do ventrículo esquerdo pela ecocardiografia em crianças e adolescentes com osteossarcoma tratados com doxorubicina com e sem dexrazoxane. *Arq Bras Cardiol.* 2006;
- [13] Robson D, Verma S. Anthracyclines in early-stage breast cancer: is it the end of an era *Oncologist.* 2009;
- [14] Floyd JD, Nguyen DT, Lobins RL, Bashir Q, Doll DC, Perry MC. Cardiotoxicity of cancer therapy. *J Clin Oncol.* 2005;
- [15] Yeh ET, Tong AT, Lenihan DJ, Yusuf SW, Swafford J, Champion C, et al. Cardiovascular complications of cancer therapy: diagnosis, pathogenesis, and management. *Circulation.* 2004;
- [16] Pai VB, Nahata MC. Cardiotoxicity of chemotherapeutic agents: incidence, treatment and prevention. *Drug Saf.* 2000;
- [17] Cappellano AM, Bouffet E, Silva F. Vinorelbine in progressive unresectable low-grade glioma in children. In: Annual Meeting of American Society of Clinical Oncology (ASCO). Chicago (Illinois), June ; 2011.
- [18] Altena R, Perik PJ, van Veldhuisen DJ, de Vries EG, Gietema JA. Cardiovascular toxicity caused by cancer treatment: strategies for early detection. *Lancet Oncol.* 2009;
- [19] Cardinale D, Colombo A, Lamantia G, Colombo N, Civelli M, De Giacomi G, et al. Anthracycline-induced cardiomyopathy: clinical relevance and response to pharmacologic therapy. *J Am Coll Cardiol.* 2010;
- [20] Yusuf SW, Razeghi P, Yeh ET. The diagnosis and management of cardiovascular disease in cancer patients. *Curr Probl Cardiol.* 2008;
- [21] Kilickap S, Barista I, Akgul E, Aytemir K, Aksoy S, Tekuzman G. Early and late arrhythmogenic effects of doxorubicin. *South Med J.* 2007;.
- [22] Cardinale D, Sandri MT. Role of biomarkers in chemotherapy-induced cardiotoxicity. *Prog Cardiovasc Dis.* 2010;
- [23] Cardinale D, Sandri MT, Colombo A, Colombo N, Boeri M, Lamantia G, et al. Prognostic value of troponin I in cardiac risk stratification of cancer patients undergoing high-dose chemotherapy. *Circulation.* 2004;
- [24] Lenihan D, Massey MR, Baysinger KB. Superior detection of cardiotoxicity during chemotherapy using biomarkers. *J Cardiac Fail.* 2007;

- [25] Salvatici M, Cardinale D, Spaggiari L, Veglia F, Tedesco CC, Solli P, et al. Atrial fibrillation after thoracic surgery for lung cancer: use of a single cut-off value of N-terminal pro-B type natriuretic peptide to identify patients at risk. *Biomarkers*. 2009;
- [26] Karanth NV, Roy A, Joseph M, de Pasquale C, Karapetis C, Koczwara B. Utility of prechemotherapy echocardiographical assessment of cardiac abnormalities. *Support Care Cancer*. 2010 Dec 15. [Epub ahead of print].
- [27] Appel JM, Sogaard P, Mortensen CE, Skagen K, Nielsen DL. Tissue-Doppler Assessment of Cardiac Left Ventricular Function during Short-Term Adjuvant Epirubicin Therapy for Breast Cancer. *J Am Soc Echocardiogr*. 2011;
- [28] Fallah-Rad N, Lytwyn M, Fang T, Kirkpatrick I, Jassal DS. Delayed contrast enhancement cardiac magnetic resonance imaging in trastuzumab induced cardiomyopathy. *J Cardiovasc Magn Reson*. 2008;
- [29] Sawyer DB, Peng X, Chen B, Pentassuglia L, Lim CC. Mechanisms of anthracycline cardiac injury: can we identify strategies for cardioprotection? *Prog Cardiovasc Dis*.
- [30] Bristow MR. beta-adrenergic receptor blockade in chronic heart failure. *Circulation*. 2000;
- [31] Bauersachs J, Fraccarollo D, Ertl G, Gretz N, Wehling M, Christ M. Striking increase of natriuresis by low-dose spironolactone in congestive heart failure only in combination with ACE inhibition: mechanistic evidence to support RALES. *Circulation*. 2000;.
- [32] 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, 2011.
- [33] DATABASE, howpublished=<https://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.aspx>, Note="acesado em jan. 2016.
- [34] Azevedo, Débora de Almeida : Espectrometria de massas.
- [35] 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.
- [36] PROGRAM, C. P. Toxicoproteomic analysis of anthracycline-induced cardiotoxicity. 2004.
- [37] O' FARRELL, P. H. High Resolution Two-Dimensional Electrophoresis of Protein. *The Journal of Biological Chemistry*, v. 250, 1975.

- [38] WILKINS, M. R.; WILLIAMS, K. L.; APPEL, R. D.; HOCHSTRASSER, D. Proteome research: new frontiers in functional genomics. Germany: Springer-Verlag, 1997. p 243.
- [39] WILKINS, M. R.; PASQUALI, C.; APPEL, R. D.; OU, K.; GOLAZ, O.; SANCHEZ, J. C.; JAN, J. X.; GOOLEY, A. A.; HUGHES, E.; HUMPHERY-SMITH, I.; WILLIAMS, K. L.; HOCHSTRASSER, D. F. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature Biotechnology*, v.14, 1996.
- [40] GYGI, S. P.; RIST, B.; GERBER, S. A.; TURECEK, F.; GELB, M. H.; AEBSOLD, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*, v. 17, n. 10, 1999.
- [41] HOCHSTRASSER, D. F. Proteome in perspective. *Clinical Chemistry and Laboratory Medicine*, v. 36, 1998.
- [42] BAIROCH, A., APWEILER, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Research.*, v. 26, 1998.
- [43] HERNANDEZ, A. SM.; SAN MIGUEL, R.; MARTÍN-GIL, F. J. Importancia de las aplicaciones clínicas de la Proteómica. *Revista del Laboratorio Clínico*, v. 3, n. 1, 2010.
- [44] CHAURAND, P.; SANDERS, M. E.; JENSEN, R. A.; CAPRIOLI, R. M. Proteomics in diagnostic pathology: Profiling and imaging proteins directly in tissue sections, *Am. J. Pathol.*, v. 165, 2004).
- [45] ISSAQ, H. J.; CONRADS, T. P.; PRIETO, D.A.; TIRUMALAI, R.; VEENSTRA, T. D. SELDITOF MS for diagnostic proteomics, *Anal Chem.*, v. 75, 2003.
- [46] MOORE, G. E.; MINOWADA, J. Historical progress and the future of human cell culture research, *Hum Cell.*, v. 5, 1992.
- [47] LE NAOUR, F. Contribution of proteomics to tumor immunology, *Proteomics.*, v. 1, 2001.
- [48] JAIN, K. K. Role of pharmacoproteomics in the development of personalized medicine, *Pharmacogenomics.*, v. 5, 2004.
- [49] GONZÁLEZ-BUITRAGO, J. M., FERREIRA, L.; LORENZO, I. Urinary proteomics, *Clin Chim Acta.*, v. 375, 2007.
- [50] BHATTACHARYA, S. H.; GAL, A. A.; MURRAY, K. K. Laser capture microdissection MALDI for direct analysis of archival tissue, *J Proteome Res.*, v. 2, 2003.

- [51] ZHENG, P. P., LUIDER, T. M.; PIETERS, R.; AVEZAAT, C. J.; VAN DEN BENT, M. J.; SILLEVIS SMITT, P. A. Identification of tumor-related proteins by proteomic analysis of cerebrospinal fluid from patients with primary brain tumors, *J Neuropathol Exp Neurol.*, v. 62, 2003.
- [52] NEUBAUER, G.; KING, A.; RAPPSILBER, J.; CALVIO C.; WATSON, M.; AJUH P.; SLEEMAN, J.; LAMOND, A.; MANN, M. Mass spectrometry and STdatabase searching allows characterization of the multi-protein spliceosome complex. *Nature Genetics.*, v. 20, 1998.
- [53] ROUT, M. P.; AITCHISON, J.D.; SUPRAPTO, A.; HJERTAAS, K.; ZHAO, Y.; CHAIT, B. T. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *Journal of Cell Biology.*, 2000;
- [54] MA, Y.; PENG, J.; HUANG, L.; LIU, W.; ZHANG, P.; QIN, H. Searching for serum tumor markers for colorectal cancer using 2D-DIGE approach. *Electrophoresis.*, v. 30, 2009.
- [55] PABA, J.; SANTANA, J. M.; TEIXEIRA, A. R. L.; FONTES, W.; SOUSA, M. V.; RICART, C. A. O. Proteomic analysis of the human pathogen *Trypanosoma cruzi*. *Proteomics.*, v. 4, 2004.
- [56] BENNETT, K. L.; FUNK, M.; TSCHERNUTTER, M.; BREITWEISER, F. P.; PLANAYAVSKY, M.; MOHIEN, C. U.; MULLER, A.; TRAJANOSKY, Z.; COLINGE, J.; SUPERTI-FURGA, G.; SCHMIDT-ERFURTH, U. Proteomic analysis of human cataract aqueous humour: comparison of the one-dimensional gel LCMS with two-dimensional LCMS of unlabelled and iTRAQ®-labelled specimens. *J. Proteomics.*, 2010.
- [57] DATABASE, howpublished=<https://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>, Note="acesado em jan. 2016.
- [58] HYVARINEN, A., Oja, E., *Independent Component Analysis: Algorithms and Applications*. Neural Networks, 2000.
- [59] HAYKIN, S., *Neural Networks A Comprehensive Foundation*. Prentice Hall, 2 ed., 1998
- [60] CARDOSO, J-F., Source Separation using High Order Moments. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP 1989)*, v.4, Glaslow, UK, 1989.
- [61] COMON, P., Separation of Stochastic Process. *Proc. of Workshop on High Order Spectral Analysis*, p. 174-179, Vail, Colorado, Jun. 1989.

- [62] JUTTEN, C., Héroult, J., Blind Separation of Sources: An Adaptive Algorithm based on Neuromimetic Architecture. *Signal Processing*, v.24, 1991.
- [63] COMON, P., Independent Component Analysis, A New Concept? *Signal Processing, Special Issue on High Order Statistics*, n. 36, 1994.
- [64] BELL, A., Sejnowski, T. J. (1995). An Information Maximization Approach to Blind Separation and Deconvolution.
- [65] LEE, T-W., Sejnowski, T., Independent Component Analysis for Sub-Gaussian and Super-Gaussian Mixture. In *IV Joint Symposium on Neural Computation*, v. 7, p.132-140, Institute for Neural Computation, 1997.
- [66] CARDOSO, J. F., Laheld, B., Equivariant Adaptive Source Separation. *IEEE Trans. on Signal Processing*, 1996 (a).
- [67] AMARI, S-I, Cichocki, A., Yang, H. H., A New Learning Algorithm for Blind Signal Separation. In Michael C. Mozer, David S. Toretzky and Michael E. Hasselmo Editors, *Advances in Neural Information Processing Systems*, v. 8, MIT Press, Cambridge, MA, 1996.
- [68] CARDOSO, J-F., Blind Separation: Statistical Principles. *Proceedings of IEEE*, 86(10), 1998.
- [69] BLISS Project [on line <http://www.bliss-project.org>]
- [70] ICA and BSS Summer School, Berlin, Jun. 2003. [on line [http://www.first.fhg.de/harmeli/summer school](http://www.first.fhg.de/harmeli/summer%20school)].
- [71] PAPOULIS, A., Probability, Random Variables and Stochastic Process. Mc Graw Hill, 3ed., 1991.
- [72] HYVARINEN, A., Oja, E., Karhunen, J., Independent Component Analysis. John Wiley e Sons, 2001.
- [73] HYVÄRINEN, A., Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Trans. on Neural Networks*, 10(3), p. 626-634, 1999(c).
- [74] HYVARINEN, A., Survey on Independent Component Analysis. *Neural Computing Surveys*, 2, 94-128, 1999.
- [75] CARDOSO, J-F., Infomax and Maximum Likelihood for Source Separation. *IEEE Letters on Signal Processing*, v.4, n.4, Abr. 1997.
- [76] HYVÄRINEN, A., Oja, E., A Fast Fixed Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7), 1997.

- [77] HYVÄRINEN, A., One Unit Contrast Function for Independent Component Analysis: A Statistical Analysis. Proc. IEEE
- [78] Workshop on Neural Network for Signal Processing, p.388-397, Florida, 1997 (b). CRISTIANINI, N., SHAW-TAYLOR J., An introduction to support Vector Machines: and other kernel-based learning methods. New York: Cambridge University Press, 1999. 189p.
- [79] Vapnik, V. N; Lerner, A. Pattern recognition using generalized portrait method. Automation and Remote Control, vol. 24, 1963.
- [80] Vapnik V.N. The nature of statistical learning theory, Springer-Verlag, 1995.
- [81] Boser, B. E.; Guyon, I. M.; Vapnik, V. N. A training algorithm for optimal margin classifiers. In D. Haussler, editor, Annual ACM Workshop on COLT, ACM Press, 1992.
- [82] Schölkopf, B. Support Vector Learning. Munich, Germany: Oldenburg-Verlag, 1997.
- [83] Burges, C. J. C. A tutorial on support vector machines for pattern recognition. Data Mining Knowledge Discovery, vol. 2, no 2, 1998.
- [84] Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; Williamson, R. C. Estimating the support of a high-dimensional distribution. Microsoft Research Corporation Report MSR-TR, 1999.
- [85] Lorena, A. C.: Introdução às Máquinas de Vetores Suporte(Support Vector Machines). Instituto de Ciências Matemáticas e de Computação.
- [86] Hearst, M. A., Scholkopf, B., Dumais, S., Osuna, E., and Platt, J. (1998). Trends and controversies - support vector machines. IEEE Intelligent Systems, 13(4):1828.
- [87] Cravem, M. W. and Shavlik, J. W. (1994). Machine learning approaches to gene recognition. IEEE Expert, 9(2):210.
- [88] Zien, A., Ratsch, G., Mika, S., Schölkopf, B., Lengauer, T., and Müller, K. R. (2000). Engineering support vector machine kernels that recognize translation initiation sites in DNA. Bioinformatics, 16:906-914.
- [89] Scholkopf, B., Simard, O., Smola, A., and Vapnik, V. (1998). Prior knowledge in support vector kernels. In Jordan, M. I., Kearns, M. J., and Solla, S. A., editors, Advances in Neural Information and Processing Systems, pages 640-646. The MIT Press.

- [90] Lorena, A. C., Batista, G. E. A. P. A., de Carvalho, A. C. P. L. F., and Monard, M. C. (2002a). The influence of noisy patterns in the performance of learning methods in the splice junction recognition problem. In *IEEE Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN 2002)*, pages 31?36. IEEE Computer Society Press.
- [91] Lorena, A. C., Batista, G. E. A. P. A., de Carvalho, A. C. P. L. F., and Monard, M. C. (2002b). Splice junction recognition using machine learning techniques. In Bazzan, A. L. C. and de Carvalho, A. C. P. L. F., editors, *Proceedings of the First Brazilian Workshop on Bioinformatics*.
- [92] Casley, D. (1992). *Primer on molecular biology*. Technical report, U. S. Department of Energy, Office of Health and Environmental Research.
- [93] Ding, C. H. Q. and Dubchak, I. (2001). Multi-class protein fold recognition using Support Vector Machines and Neural Networks. *Bioinformatics*.
- [94] Furey, T., Cristianini, N., Duffy, N., Bednarski, D., Schummer, M., and Haussler, D. (2001). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*.
- [95] Tamayo, P., Slonim, D., Mesirov, J., Zhum, Q., Kitareewan, S., Dmitrovsky, E., Lander, E., and Golub, T. (1996). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. In *PNAS*.
- [96] Ben-Hur, A., ad H. T. Siegelmann, D. H., and Vapnik, V. N. (2000). A support vector clustering method. In *Proceedings of the International Conference on Pattern Recognition (ICPR00)*, volume 2.
- [97] Brown, M., Grundy, W., Lin, D., Christianini, N., Sugnet, C., Jr, M., and Haussler, D. (1999). Support vector machine classification of microarray gene expression data. Technical Report UCSC-CRL, Department of Computer Science, University California Santa Cruz, Santa Cruz, CA.
- [98] Russel and Norvig (1995). *Artificial Intelligence - a Modern Approach*. Prentice Hall.
- [99] Muller, K. R., Mika, S., Rˆatsch, G., Tsuda, K., and Schˆolkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*.
- [100] Smola, A. J., Barlett, P., Schˆolkopf, B., and Schuurmans, D. (1999b). Introduction to Large Margin Classifiers, chapter 1, pages 1?28. In Smola et al. (1999a).
- [101] Campbell, C. (2000). An introduction to kernel methods. In Howlett, R. J. and Jain, L. C., editors, *Radial Basis Function Networks: Design and Applications*, pages 155? 192. Springer Verlag.

- [102] Cristianini, N. and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.
- [103] Vert, J.-P. (2001). Introduction to support vector machines and applications to computational biology (draft). <http://web.kuicr.kyoto-u.ac.jp/~vert/research/semsvm/>.
- [104] Haykin, S. (1999). Neural Networks - A Comprehensive Foundation. Prentice-Hall, New Jersey, 2 edition.
- [105] Wahba, G. (2000). An introduction to model building with reproducing kernel hilbert spaces. Technical Report 1020, Statistics Department of University of Wisconsin-Madison.
- [106] Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. Philos. Trans. Roy. Soc. London, A 209:415-446.
- [107] Smola, A. J., Barlett, P., Schölkopf, B., and Schuurmans, D. (1999b). Introduction to Large Margin Classifiers, chapter 1, pages 1-28. In Smola et al. (1999a). Smola, A. J. and Schölkopf, B. (2002). Learning with Kernels. The MIT Press, Cambridge
- [108] Chapelle, O. and Vapnik, V. N. (2000). Model selection for support vector machines. In Solla, S. A., Leen, T. K., and Müller, K.-R., editors, Advances in Neural Information Processing Systems. The MIT Press.
- [109] Jaakola, T. and Haussler, D. (1999). Probabilistic kernel regression models. In Proceedings of the 1999 Conference on AI and Statistics. Morgan Kaufmann.
- [110] Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S. (2002). Choosing multiple parameters for support vector machines. Machine Learning.
- [111] S. Haykin. Neural Networks - A Comprehensive Foundation. Prentice-Hall, New Jersey, 2nd edition, 1999.
- [112] M. C. P. Souto, A. C. Lorena, A. C. B. Delbem, and A. C. P. L. F. Carvalho. Técnicas de Aprendizado de Máquina para problemas de Biologia Molecular. Minicursos de Inteligência Artificial, Jornada de Atualização Científica em Inteligência Artificial, XXIII Congresso da Sociedade Brasileira de Computação, 2003.
- [113] T. Mitchell. Machine Learning. McGraw Hill, 1997.
- [114] M. C. Monard and J. A. Baranauskas. Conceitos de aprendizado de máquina. In S. O. Rezende, editor, Sistemas Inteligentes - Fundamentos e Aplicações. Editora Manole, 2003.

- [115] J. A. Baranauskas and M. C. Monard. Reviewing some machine learning concepts and methods. Technical Report 102, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos. Disponível em: <ftp://ftp.icmc.usp.br/pub/BIBLIOTECA/retec/RT102.ps.zip>, Fevereiro 2000.
- [116] Baldi, P. and Brunak, S. (1998). *Bioinformatics - The Machine Learning Approach*. The MIT Press.
- [117] 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.
- [118] 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..
- [119] PROGRAM, C. P. *Toxicoproteomic analysis of anthracycline-induced cardiotoxicity*. 2004.
- [120] Bushberg, J. T., Seibert, A. J., Leidholdt, E. M. and Boone, J. M. (2001). "The Essential Physics of Medical Imaging", second ed, Lippincott William e Wilkins, Philadelphia, PA