



UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
CURSO DE ENGENHARIA DA COMPUTAÇÃO

PAOLA LAUANDE OLIVEIRA MENDES

**DIAGNÓSTICO AUXILIADO POR COMPUTADOR APLICADO A MARCADORES
DE CARDIOTOXICIDADE**

São Luís – MA

2019

PAOLA LAUANDE OLIVEIRA MENDES

**DIAGNÓSTICO AUXILIADO POR COMPUTADOR APLICADO A MARCADORES
DE CARDIOTOXICIDADE**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia da Computação da Universidade Estadual do Maranhão, como registro para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Prof. Dr. Lúcio Flávio de Albuquerque Campos.

São Luís – MA

2019

Mendes, Paola Lauande Oliveira.

Diagnóstico auxiliado por computador aplicado a marcadores de cardiotoxicidade / Paola Lauande Oliveira Mendes. – São Luís, 2019.

71 f

Monografia (Graduação) – Curso de Engenharia da Computação, Universidade Estadual do Maranhão, 2019.

Orientador: Prof. Dr. Lúcio Flávio de Albuquerque Campos.

1.Cardiotoxicidade. 2.Análise de componentes independentes. 3.Análise discriminante linear. 4.Máquina de vetores de suporte. I.Título.

CDU: 004.891.3

PAOLA LAUANDE OLIVEIRA MENDES

**DIAGNÓSTICO AUXILIADO POR COMPUTADOR APLICADO A MARCADORES
DE CARDIOTOXICIDADE**

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia da Computação da Universidade Estadual do Maranhão, como registro para obtenção do grau de Bacharel em Engenharia da Computação.

Aprovada em: de julho de 2019.

BANCA EXAMINADORA

Prof. Dr. Lúcio Flávio de Albuquerque Campos (Orientador)
Universidade Estadual do Maranhão

Prof. Dr. Luis Carlos Costa Fonseca (Examinador)
Universidade Estadual do Maranhão

Prof. Me. Antônio Fernando Lavareda Jacob Junior (Examinador)
Universidade Estadual do Maranhão

AGRADECIMENTOS

A Deus pela vida.

Aos meus pais Paula Lauande e João Filho e ao meu irmão João Neto pelo amor, incentivo, compreensão e paciência durante todo o período do curso de graduação.

Aos meus avós Maria do Rosário e Benedito Coelho e à minha tia Patrícia Lauande pelo afeto e apoio de sempre, concedendo conforto nas horas mais complicadas desta jornada.

Ao meu orientador, prof. Dr. Lúcio Flávio de Albuquerque Campos pelo acompanhamento, confiança e auxílio destinados a mim durante todo o período de desenvolvimento deste trabalho.

Aos meus amigos da Engenharia da Computação, em especial a Marcos Guilhon, João Francisco Vieira, Rodrigo Albuquerque, Eduardo Andrade e Luís Eduardo pela amizade e por termos superados juntos os momentos difíceis do curso.

Ao curso de Engenharia da Computação pela oportunidade de obter formação na área desejada e aos professores por terem colaborado com o conhecimento adquirido durante o curso.

A todos que contribuíram direta ou indiretamente para a elaboração deste trabalho.

“O interesse pela doença e pela morte é apenas outra expressão do interesse pela vida.”

(Thomas Mann)

RESUMO

A cardiotoxicidade se dá pela ocorrência de um dano muscular ou disfunção de eletrofisiologia do coração. Várias são as causas dessa enfermidade, entre as quais está a quimioterapia. Sabe-se que esta doença compromete o coração tornando-o ineficiente em suas funções que são primordiais para manter o bombeamento do sangue no corpo, e, se for descoberto em estágio avançado, pode vir a ser letal. O ecocardiograma é o método de diagnóstico mais utilizado, mas ainda não é eficaz para o diagnóstico precoce da cardiotoxicidade. Este trabalho apresenta uma proposta de CAD (*Computer-Aided Diagnosis*) para auxiliar no diagnóstico precoce da doença em questão. A metodologia utiliza sinais proteômicos cujas características serão extraídas por uma técnica chamada Análise de Componentes Independentes (ICA). As características mais significantes são selecionadas pelo algoritmo de Máxima Relevância Mínima Redundância (MRmR), e, então, as amostras são classificadas através dos métodos de Análise Discriminante Linear (LDA) e Máquinas de Vetores de Suporte (SVM), detectando a presença ou a ausência de cardiotoxicidade. Os métodos foram testados na base de dados proteômicos SELDI-TOF, com 62 amostras em alta resolução (373.257 pontos), sendo 34 com cardiotoxicidade e 28 normais. Os testes de classificação com LDA geraram 100% de acurácia, 100% de sensibilidade e 100% de especificidade para as 62 amostras, a partir de um vetor com 48 características. Os testes com SVM mostraram que o melhor desempenho foi obtido com um vetor de 20 características, resultando em uma taxa de acurácia de 88,718%, sensibilidade de 97,260% e especificidade de 85%. Comparando-se os resultados, foi observado que a LDA obteve melhor desempenho. Os resultados com LDA são promissores, porém deve ser testado com outras bases de dados para analisar sua real eficácia.

Palavras-chave: Cardiotoxicidade, Análise de Componentes Independentes, Análise Discriminante Linear, Máquinas de Vetores de Suporte.

ABSTRACT

Cardiotoxicity is caused by muscle injury or dysfunction of the electrophysiology of the heart. There are several causes of this disease, among which is chemotherapy. It is known that this disease compromises the heart, making it inefficient in its functions that are paramount to maintain the pumping of blood into the body, and if it is discovered at an advanced stage, can be lethal. The echocardiogram is the most used, but is not yet effective for the early diagnosis of cardiotoxicity. This work presents a CAD proposal (Computer-Aided Diagnosis) to aid in the early diagnosis of the disease in question. The methodology uses proteomic signals whose characteristics will be extracted by a technique called Independent Component Analysis (ICA). At most significant characteristics are selected by Maxima's algorithm Minimum Redundancy (MRmR), and then the samples are classified using the Linear Discriminant Analysis (LDA) methods and Support Vector Machines (SVM), detecting presence or absence cardiotoxicity. The methods were tested in the proteomic database SELDI-TOF, with 62 samples in high resolution (373,257 points), 34 with cardiotoxicity and 28 normal. The classification tests with LDA generated 100% accuracy, 100% sensitivity and 100% specificity for the 62 samples, from a vector with 48 characteristics. The tests with SVM showed that the best performance was obtained with a vector of 20 characteristics, resulting in an accuracy rate of 88.718%, sensitivity of 97.260%, and specificity of 85%. Comparing the results, it was observed that LDA performed better. The results with LDA are promising, but should be tested with other databases to analyze effectiveness.

Keywords: Cardiotoxicity, Analysis of Independent Components, Linear Discriminant Analysis, Support Vector Machines.

LISTA DE FIGURAS

Figura 1– Fluxograma da Teoria dos Múltiplos Golpes	22
Figura 2 - Esquema simplificado de um espectrômetro de massas	32
Figura 3 - Figura esquemática de um espectrômetro de massas MALDI-TOF	34
Figura 4 – Modelo do Problema de Separação Cega de Fontes.....	37
Figura 5 – Distribuição de duas variáveis gaussianas independentes	39
Figura 6 – Fluxograma do algoritmo FastICA.....	42
Figura 7 – A hierarquia do aprendizado	45
Figura 8 – Possíveis hiperplanos de separação e hiperplano ótimo.	46
Figura 9 – Conjunto de dados linearmente separável	47
Figura 10 – Conjunto de dados não linearmente separável	47
Figura 11 – Mapeamento de dados para um espaço de dimensão maior.....	48
Figura 12 - Diagrama em blocos da metodologia proposta.....	51
Figura 13 – Amostras disponíveis para conversão em padrões proteômicos	52
Figura 14 – Espectro de massas.....	53

LISTA DE TABELAS

Tabela 1 - Manifestações clínicas de cardiotoxicidade	21
Tabela 2 - Diferentes definições de cardiotoxicidade	23
Tabela 3 - Fatores de risco para cardiotoxicidade associada às antraciclinas	24
Tabela 4 - Classificação para a cardiomiopatia relacionada à quimioterapia	26
Tabela 5 - Incidência de insuficiência cardíaca por classe terapêutica	27
Tabela 6 – Funções Kernel	49
Tabela 7 - Desempenho do SVM para os vetores de características.....	58
Tabela 8 – Desempenho do LDA para os vetores de características.....	59

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
BSS	<i>Blind Source Separation</i> (Separação Cega de Fontes)
CAD	<i>Computer-Aided Diagnosis</i> (Diagnóstico Auxiliado por computador)
CTX	Cardiotoxicidade
EM	Espectrometria de Massas
ESI	<i>Electron Spray Ionization</i> (Ionização por Electrospray)
FEVE	Fração de Ejeção do Ventrículo Esquerdo
FN	Falso Negativo
FP	Falso Positivo
ICA	<i>Independent Component Analysis</i> (Análise de Componentes Independentes)
IECA	Inibidores da Enzima Conversora de Angiotensina
LDA	<i>Linear Discriminant Analysis</i> (Análise Discriminante Linear)
MALDI	<i>Matrix Assisted Laser Desorption Ionization</i> (Ionização e dessorção a laser assistida por matriz)
mRMR	Máxima Relevância e Mínima Redundância
pdf	Função densidade de probabilidade
QT	Quimioterapia
RBF	<i>Radial Basis Function</i> (Função de Base Radial)
RNM	Ressonância Nuclear Magnética
SELDI-TOF MS	<i>Surface Enhanced Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry</i>
SVM	<i>Support Vector Machine</i> (Máquina de Vetores de Suporte)

VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
VRN	Ventriculografia Radionuclídica

SUMÁRIO

AGRADECIMENTOS	5
RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
LISTA DE ABREVIATURAS E SIGLAS	11
1 INTRODUÇÃO	15
1.1 Justificativa.....	16
1.2 Trabalhos Relacionados.....	17
1.3 Objetivos	18
1.3.1 Objetivo Geral	18
1.3.2 Objetivos Específicos	19
1.4 Estrutura do Trabalho.....	19
2. REVISÃO DE LITERATURA	20
2.1 Cardiotoxicidade.....	20
2.1.1 Conceito e manifestação	20
2.1.2 Fatores de risco	23
2.1.3 Classificação	24
2.1.4 Agentes quimioterápicos e cardiotoxicidade	26
2.1.4.1 Antraciclina	26
2.1.5 Diagnóstico	27
2.1.5.1 Sinais e sintomas	28
2.1.5.2 Métodos de Imagem	28
2.1.5.3 Eletrocardiograma	29
2.1.5.4 Biomarcadores	29
2.1.5.5 Biópsia endomiocárdica	29
2.1.6 Tratamento	30
2.1.7 Prevenção	30
2.2 Análise Proteômica	31
2.2.1 Conceitos	31
2.2.2 Espectrometria de massa	32
2.2.3 Ionização por Electrospray (ESI)	32

2.2.3.1 MALDI	33
2.3 Biomedicina e estudo proteômico	35
2.4 Diagnóstico Auxiliado por Computador	35
2.5 Análise de Componentes Independentes	36
2.5.1 Separação Cega de Fontes	36
2.5.2 Definição e Método ICA	37
2.5.3 Restrições	39
2.5.4 Pré-processamento dos dados	40
2.5.4.1 Centralização	40
2.5.4.2 Branqueamento	40
2.5.5 Algoritmo FastICA	41
2.6 Seleção de Características Discriminativas	43
2.6.1 Máxima Relevância e Mínima Redundância	43
2.7 Aprendizagem de Máquina.....	44
2.8 Máquina de Vetores de Suporte.....	45
2.9 Análise Discriminante Linear	49
3. METODOLOGIA	51
3.1 Dataset	51
3.2 Extração de características	53
3.3 Seleção das características mais significativas.....	54
3.4 Classificação	54
3.5 Métricas de desempenho	55
4 RESULTADOS E DISCUSSÕES.....	57
4.1 Aquisição do Dataset.....	57
4.2 Extração de características	57
4.3 Seleção das características mais significantes	58
4.4 Classificação	58
4.4.1 Classificação SVM	58
4.4.2 Classificação LDA	59
5 CONCLUSÃO	61
REFERÊNCIAS.....	62
ANEXO A.....	70

1 INTRODUÇÃO

A medicina tem contribuído consideravelmente para o tratamento de várias doenças causadas por agentes biológicos, dentre elas, o câncer. Responsável por mais de seis bilhões de óbitos a cada ano no mundo, o câncer é um importante problema de saúde pública (GUERRA; GALLO; MENDONÇA, 2004). São poucas e complexas as formas de tratamento dessa enfermidade, como a radioterapia, hormonioterapia, imunoterapia, e a mais conhecida, quimioterapia (TARTARI; BUSNELLO; NUNES, 2009).

O tratamento quimioterápico é o método em que se utilizam substâncias químicas que controlam ou curam a doença, destruindo células malignas, bloqueando suas funções ou levando à morte celular. Entretanto, além de atingir as células cancerígenas, os quimioterápicos também causam danos às células saudáveis do organismo, que constituem os tecidos (TARTARI; BUSNELLO; NUNES, 2009). Por este motivo, os efeitos adversos do tratamento antitumoral são bastante frequentes, podendo causar danos temporários ou permanentes no paciente.

Um dos efeitos adversos do uso de quimioterápicos é a cardiotoxicidade, que, por não ter uma definição única, é entendida como qualquer alteração da homeostase do sistema cardiovascular induzida pelo tratamento oncológico (PINTO et al., 2017).

Atualmente, no mundo inteiro, as doenças cardiovasculares correspondem a 17,3 milhões de óbitos a cada ano, sendo a principal causa de morte (AMERICAN HEART ASSOCIATION, 2015). Em paralelo, ocupam 20% dos casos de morte em homens e mulheres com mais de 30 anos, no Brasil (MANSUR; FAVARATO, 2012).

Os principais sinais que a cardiotoxicidade apresenta são insuficiência cardíaca, hipertensão arterial, arritmias ventriculares, eventos tromboembólicos, entre outras disfunções cardíacas (SOUSA, 2014). Por vezes, pode ser assintomática, tendo seu diagnóstico tardio; e, dependendo da gravidade da doença, ela pode afetar a qualidade de vida do paciente, sendo possível até leva-lo à óbito (A CARDIOLOGIA NO SÉCULO 21: UMA VISÃO MULTIDISCIPLINAR, 2013).

Até os dias de hoje, não há protocolos bem definidos para a detecção precoce da cardiotoxicidade. Os métodos tradicionais que ajudam na sua prevenção são os exames cardíacos, tais como o ecocardiograma, dosagem de marcadores

bioquímicos e a biópsia endomiocárdica (COSTA et al., 2010). Esses e outros métodos de diagnósticos por imagem fazem a monitoração da função sistólica do ventrículo esquerdo do paciente, avaliando se a fração de ejeção é reduzida, situação em que os ventrículos do coração não se contraem adequadamente durante os batimentos. Isso faz com que o sangue não seja bombeado de forma eficiente para fora do coração. No entanto, só é possível fazer essa descoberta quando a lesão já está instalada no coração, o que não é responsivo o suficiente para uma doença grave (PIVETA; PAIVA; ANDRADE, 2012).

Levando em consideração que o sucesso do tratamento da cardiotoxicidade depende de um diagnóstico rápido, estão sendo propostos alguns esquemas de CAD (Diagnóstico Auxiliado por Computador) com a finalidade de ajudar na identificação precoce dessa enfermidade (ALMEIDA et al., 2019). Ainda não existem métodos CAD associados à essa patologia, porém as propostas lançadas dão uma margem considerável para o desenvolvimento completo de ferramentas que ajudem no seu diagnóstico precoce.

Neste trabalho será feito um estudo comparativo de técnicas de classificação, tais como Análise Discriminante Linear e Máquinas de Vetores de Suporte aplicados em CAD para auxiliar no diagnóstico de pacientes com Cardiotoxicidade, utilizando Análise de Componentes Independentes para extrair características de sinais proteômicos e posteriormente classificá-los através das técnicas mencionadas acima.

1.1 Justificativa

A Cardiotoxicidade é definida pela situação na qual agentes externos (químicos ou físicos) interferem negativamente no coração, determinando alterações estruturais, elétricas e funcionais no miocárdio. Devido à necessidade da detecção precoce dessa patologia, vários esquemas de CAD estão sendo propostos com o objetivo de ajudar na sua identificação, já que o sucesso do tratamento no combate à esta disfunção cardíaca depende de um diagnóstico rápido, pois quando mais cedo se iniciar o tratamento, maiores serão as chances de cura. Portanto, neste trabalho será apresentada uma proposta de CAD e, em seguida, esta será comparada com propostas já feitas em outros trabalhos. Esta ponte interdisciplinar entre assuntos das áreas de Computação e Medicina tem a finalidade de abrir espaço para novas ferramentas tecnológicas que ajudem no tratamento e cura da patologia em questão.

1.2 Trabalhos Relacionados

Durante o trabalho desenvolvido, foram revisados alguns trabalhos científicos relacionados ao tema. Estas bibliografias são fontes de estudos com outras abordagens para a implementação de técnicas de classificação aplicadas a sinais proteômicos. Neste tópico serão destacadas e analisadas as principais contribuições de alguns dos trabalhos que ajudaram este a ser complementado.

Almeida (2016) apresentou um método de CAD para detectar a cardiotoxicidade que usa Análise de Componentes Independentes (ICA) para a extração de características, e Máquina de Vetores Suporte (SVM) para a classificação de pacientes portadores ou não da enfermidade em questão.

O estudo foi feito com 62 amostras de baixa resolução, cada amostra tendo 373257 níveis de características diferentes. Utilizando a técnica SELDI-TOF, obteve resultados com acurácia de 88,71%, com especificidade de 85% e sensibilidade de 97,26%, utilizando vetor de 20 características. A base de dados utilizada pode ser encontrada no site (PROGRAM, 2004).

Costa (2008), em sua tese de dissertação voltada ao câncer de mama, apresentou uma metodologia que também utiliza ICA e SVM, porém, acrescenta na pesquisa o classificador Análise Discriminante Linear (LDA), para distinguir as imagens entre nódulos ou não-nódulos e os tecidos em benignos ou malignos. Utilizou duas bases de dados disponíveis publicamente: a Digital Database for Screening Mammography (DDSM) (HEATH et al., 1998) e a Mammographic Image Analysis Society (mini-MIAS) (SUCKLING et al., 1994).

Os resultados obtidos com LDA foram de 90,11% de acurácia na discriminação entre nódulos e não-nódulos, e 95,38% entre tecidos benignos e malignos na base de dados DDSM; enquanto na base mini-MIAS, a discriminação entre os respectivos grupos de imagens foi de 85% e 92% (COSTA, 2008).

Com SVM, a taxa foi até 99,55% na discriminação entre nódulos e não-nódulos e também entre tecidos benignos e malignos, na base de dados DDSMM. Por outro lado, na base de dados mini-MIAS, foi obtido 98% e 100% nas duas discriminações, respectivamente (COSTA, 2008).

Reis e Campos (2017) propuseram um método CAD para a detecção de lesões de câncer de pele melanoma e não melanoma. Utilizam Análise de Componentes Independentes (ICA) e Análise Discriminante Linear (LDA) para extração e classificação das imagens. Os dados adquiridos para o estudo são de domínio público e podem ser acessados em Dermis (2012) e Dermquest (2012).

O método proposto obteve resultados satisfatórios, visto que, com o aumento da quantidade de características, de 20 para 185, aumentam-se também as porcentagens de acurácia, sensibilidade e especificidade. Para o vetor com 185 características, a exatidão é de 100%, a sensibilidade de 100% e a especificidade de 100% (REIS; CAMPOS, 2017).

Araujo (2018) direcionou seu estudo aos cânceres de ovário e próstata. Utilizou os classificadores LDA e SVM. Com a técnica SELDI-TOF para extração das amostras, foram utilizadas duas bases de dados: Ovarian Dataset 8-7-02 e JNCI Dataset 7-3-02.

Para a base de dados do câncer de ovário, a taxa de acurácia foi de 98,80%, com especificidade de 95,65% e sensibilidade de 100%, em um estudo com 253 amostras com baixa resolução e vetor de 10 características, para classificação câncer ou normalidade, utilizando-se do SVM. Já com LDA, obteve-se acurácia de 100%, especificidade de 100% e sensibilidade de 100%, sendo 253 amostras com 25 características.

Para a base de dados do câncer de próstata, levando em conta que foram 259 amostras e o vetor teria 20 características, as taxas de acurácia, sensibilidade e especificidade foram de 90,73%, 94,87% e 90,18% para o classificador SVM. Com LDA, alcançou-se um vetor de 85 características, obtendo-se acurácia, especificidade e sensibilidade de 100%.

1.3 Objetivos

1.3.1 Objetivo Geral

Realizar um estudo comparativo dos classificadores de padrão mais utilizados na atualidade, com o intuito de melhorar a precisão do diagnóstico de cardiotoxicidade.

1.3.2 Objetivos Específicos

- Realizar um estudo analítico sobre classificadores;
- Estudar sobre a característica dos sinais biológicos;
- Desenvolver algoritmos de extração de características aplicados ao problema proposto;
- Encontrar melhores resultados de Acurácia, Sensibilidade e Especificidade relacionados ao problema.

1.4 Estrutura do Trabalho

Este trabalho está dividido em 5 capítulos, conforme apresentado no sumário.

O capítulo 1 trata-se da Introdução, em que o tema é introduzido dentro de um contexto, deixando claro qual o problema em questão. Além disso, é apresentada a justificativa/motivação principal para a realização do trabalho. Posteriormente, alguns trabalhos relacionados são comentados em paralelo com seus resultados que contribuíram para a o desenvolvimento deste trabalho. Por conseguinte, são apresentados os objetivos gerais e específicos deste trabalho de conclusão de curso.

No capítulo 2, Revisão de Literatura, é feita uma fundamentação teórica dos principais assuntos que envolvem o trabalho. Conceitos de cardiotoxicidade e afins, Análise Proteômica, Análise de Componentes Independentes, o algoritmo de Máxima Relevância e Mínima Redundância, Conceitos básicos de Aprendizagem de Máquina, Máquina de Vetores de Suporte e Análise Discriminante Linear.

A metodologia proposta e os materiais/técnicas utilizadas nas etapas deste trabalho são descritos no capítulo 3.

O capítulo 4 lista os resultados obtidos, a análise das técnicas utilizadas e a validação do método pelas métricas de desempenho: Acurácia, Especificidade e Sensibilidade.

Por fim, no capítulo 5 faz-se a conclusão de todo o processo feito, com direito a discussões e análises. Além disso, mostra qual dos dois métodos utilizados é mais eficiente em sua proposta. Finalmente, dá-se uma margem de sugestões para trabalhos futuros.

2. REVISÃO DE LITERATURA

Neste capítulo, serão apresentados assuntos que servirão como base para o desenvolvimento do método proposto. Para compreender todo o processo de elaboração da proposta de Diagnóstico Auxiliado por Computador aplicado em cardiotoxicidade, é essencial que se tenha um entendimento sólido em cima de conceitos como a cardiotoxicidade e sua classificação, fatores de risco, agentes quimioterápicos, diagnóstico e tipos de tratamentos.

Em seguida, será feito um breve estudo sobre análise proteômica, assunto que engloba informações relevantes sobre determinado estado patológico de uma estrutura, para que, assim, ela seja classificada dentro de uma categoria.

Finalmente, serão revisados conceitos sobre as técnicas computacionais utilizadas ao longo do processo deste trabalho. Análise de Componentes Independentes (ICA), algoritmo de Máxima Relevância e Mínima Redundância (MRmR), Máquina de Vetores de Suporte (SVM) e Análise Discriminante Linear (LDA).

2.1 Cardiotoxicidade

2.1.1 Conceito e manifestação

A oncologia obteve avanços importantes nas últimas décadas com o desenvolvimento de tratamentos mais eficazes e menos tóxicos. Porém, ainda é notável o número de complicações cardiovasculares pós tratamento, relacionadas à grande exposição de toxicidade de quimioterápicos (MESQUITA, 2017). Dessa forma, surgiu uma nova especialidade chamada cardio-oncologia, com a finalidade de priorizar estudos de danos da terapia do câncer ao sistema cardiovascular, permitindo, assim, o desenvolvimento de formas mais eficazes de prevenção e tratamento (MESQUITA, 2017).

A cardiotoxicidade se dá pela alta taxa tóxica que interfere negativamente no coração, resultando em alterações estruturais, elétricas e funcionais no miocárdio. Ela é considerada o efeito colateral mais grave do tratamento da quimioterapia. A enfermidade não tem um conceito padronizado que se compõe de descobertas clínicas e exames complementares, mas, no geral, a cardiotoxicidade ligada ao tratamento do câncer é conceituada como um conjunto de sinais e sintomas clínicos de insuficiência cardíaca que não está relacionada a outras causas, como por exemplo

sepsis e insuficiência renal (I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA PEDIÁTRICA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2013).

A manifestação mais frequente e temida da cardiotoxicidade é a disfunção ventricular esquerda, que pode levar à insuficiência cardíaca. Além desta, há casos de arritmia, eventos tromboembólicos, hipertensão arterial sistêmica, entre outras, como mostra a Tabela 1 (I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2011).

Tabela 1 - Manifestações clínicas de cardiotoxicidade

Cardiotoxicidade
Insuficiência Cardíaca
Arritmias ventriculares e supraventriculares
Isquemia miocárdica aguda com ou sem supra de ST
Disfunção ventricular esquerda assintomática
Hipertensão arterial sistêmica
Doença pericárdica
Eventos tromboembólicos

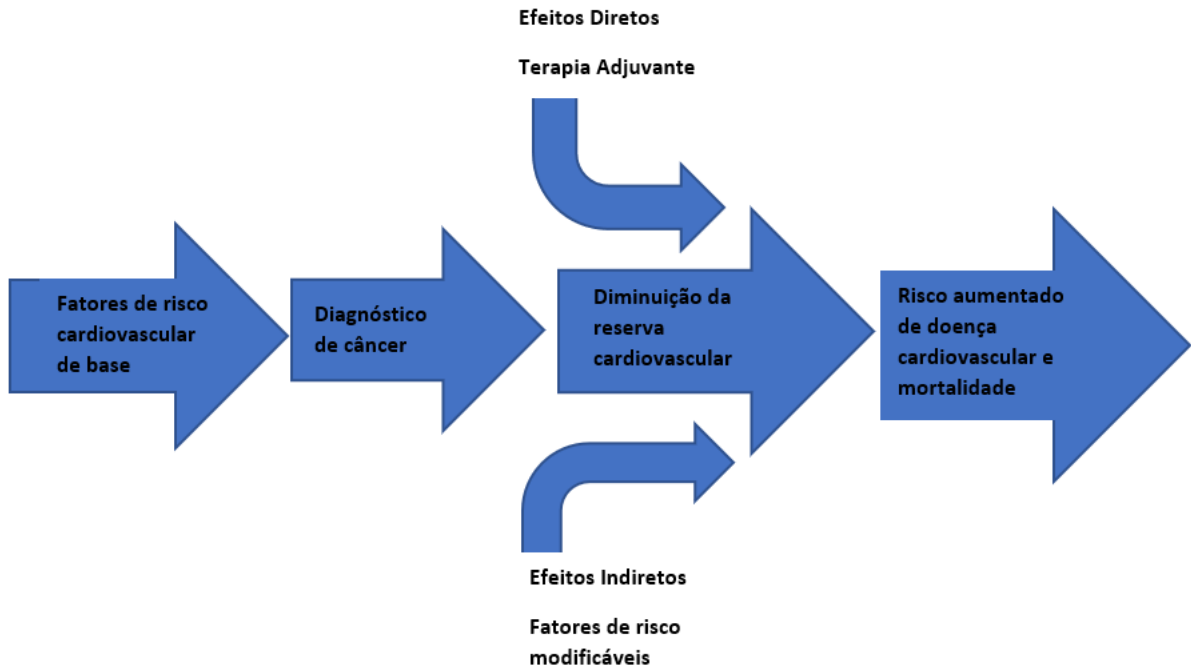
Fonte: FILHO et al (2011)

Todas essas manifestações estão inter-relacionadas e podem ocorrer isolada ou simultaneamente, acentuando o quadro de saúde do paciente. Além do mais, elas podem ocorrer no início, durante ou após o tratamento; até mesmo podem não ser aparente por meses ou até anos (I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA PEDIÁTRICA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2013).

Há uma teoria convincente que supõe que vários fatores juntos deixam os pacientes mais vulneráveis às agressões ao sistema cardiovascular, em tratamento quimioterápico (SCARLATELLI; LOPES JUNIOR; PONTES, 2015). A teoria dos múltiplos golpes explica que além das agressões tóxicas que o organismo recebe da quimioterapia, a mudança desfavorável no estilo de vida do enfermo pós-diagnóstico de câncer também passa a ser um fator de risco. É comum o paciente parar de se exercitar fisicamente, ter depressão (fator de risco cardiovascular) e ter tendência ao aumento de peso (JONES et al., 2007). Todos esses fatores tornam os pacientes mais

vulneráveis e aumentam seu risco de doença cardiovascular precoce, como mostra a Figura 1.

Figura 1– Fluxograma da Teoria dos Múltiplos Golpes



Fonte: Adaptado de JONES et al (2007)

Como não existe clareza na literatura da cardiotoxicidade, é importante homogeneizar as definições para um melhor entendimento e familiarização com o assunto, como na Tabela 2.

Tabela 2 - Diferentes definições de cardiotoxicidade

<i>Fonte</i>	<i>Definição de Cardiotoxicidade</i>	
<i>Instituto Nacional de Saúde, em I Diretriz Brasileira de Cardio-Onco (2011).</i>	Grau I	Redução assintomática da FEVE entre 10-20%.
	Grau II	Redução da FEVE abaixo de 20% ou abaixo do normal.
	Grau III	Insuficiência cardíaca sintomática.
American Society of Echocardiography Consensus (2014).	Redução $\geq 10\%$ da FEVE para um basal de FEVE final abaixo de 53% (sugestão: repetir o exame de imagem)	
Cardiac Review and Evaluation Committee (2002).	<ol style="list-style-type: none"> 1. Cardiomiopatia caracterizada por redução na FEVE globalmente ou mais severa em região septal; 2. Sinais e sintomas compatíveis de Insuficiência Cardíaca; 3. Redução $\geq 5\%$ da FEVE em pacientes sintomáticos ou $\geq 10\%$ da FEVE em pacientes assintomáticos para um basal menor do que 55%. 	

Fonte: MEDEIROS, WIEHE (2018). FEVE: Fração de Ejeção do Ventrículo Esquerdo

A FEVE é a porcentagem de sangue do ventrículo esquerdo que é ejetada a cada batimento cardíaco. Portanto, é uma variável importante para caracterizar a síndrome clínica IC (Insuficiência Cardíaca) (PINHEIRO et al., 2012).

2.1.2 Fatores de risco

Os fatores relacionados ao aumento do risco de se desenvolver uma doença são os chamados fatores de risco. Eles podem ser identificados no ambiente físico, na genética ou até mesmo em hábitos e costumes. Há alguns fatores de risco que podem deixar um paciente em tratamento quimioterápico ainda mais vulnerável à cardiotoxicidade. Por exemplo: idade, sexo, história prévia de doença cardíaca, modo de administração, dose cumulativa, tipo de antraciclina, entre outros. A Tabela 3 expõe mais detalhadamente os fatores de risco comentados.

Tabela 3 - 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:	Excedendo a dose cumulativa de:
<i>Daunorrubicina</i>	<i>550 a 800mg/m²</i>
<i>Doxorrubicina</i>	<i>400 a 550mg/m²</i>
<i>Epirubicina</i>	<i>900 a 1000mg/m²</i>
<i>Idarrubicina</i>	<i>150 a 225mg/m²</i>
Irradiação Mediastinal	Precoce ou concomitante
Doenças Cardiovasculares Prévias	Hipertensão arterial, doença coronária
Distúrbios eletrolíticos	Hipocalcemia, hipomagnesemia

Fonte: I Diretriz Brasileira de Cardio-Oncologia da SBC 2011

Além disso, os pacientes que recebem tratamento quimioterápico costumam desenvolver condições de risco tais como hipertensão, obesidade, síndrome metabólica, que acabam deixando ainda mais escassa a reserva cardiovascular, aumentando as chances de desenvolver cardiotoxicidade (RICHARDS et al., 2011).

2.1.3 Classificação

A cardiotoxicidade manifestada por antracíclicos é a mais comum na prática clínica, pelo fato desses agentes quimioterápicos serem bastante utilizados vários tipos de neoplasias e por terem um alto potencial de lesão cardíaca (PINTO et al., 2017). Dessa forma, em relação ao tempo de apresentação clínica, a cardiotoxicidade por antracíclicos divide-se em aguda, subaguda ou crônica.

A cardiotoxicidade aguda é considerada uma forma rara (menos do que 1% dos casos) e reversível, no geral. As manifestações são observadas geralmente logo após a infusão da droga, com arritmias supraventriculares (PINTO et al., 2017). Por esses motivos, ela é definida como toxicidade ocorrida durante o ciclo de quimioterapia; enquanto as formas subaguda e crônica são definidas como toxicidade ocorrida fora do ciclo de quimioterapia (SÁ et al., 2009).

A forma subaguda tem uma incidência de 1,6% a 2,1% dos casos, um pouco mais do que a aguda. Geralmente é irreversível, podendo se manifestar semanas ou

meses após o início do tratamento. Caracteriza-se principalmente pela disfunção ventricular (PINTO et al., 2017).

A cardiotoxicidade crônica é a forma mais frequente da doença, com incidência de 1,6% a 5%, e na maioria dos casos, é irreversível. Aparece mais tardiamente do que a subaguda, o que faz com que seja ainda mais severa, manifestando-se no primeiro ano de tratamento; os sintomas mais acentuados aparecem no 7º ano após o tratamento (PINTO et al., 2017). Essa forma ainda se divide em duas categorias: precoce e tardia. A primeira manifesta-se dentro do primeiro ano pós-tratamento; enquanto a segunda é diagnosticada depois do primeiro ano e até vários anos depois do término da terapia.

Geralmente, a cardiotoxicidade crônica pronuncia-se como disfunção ventricular sistólica ou diastólica, que é quando o enchimento ventricular está prejudicado e a pressão do ventrículo esquerdo pode estar elevada. A primeira pode ocorrer imediatamente ou no decorrer de meses ou anos após doses alta de quimioterápicos; a segunda ocorre comumente pós-radioterapia. As duas podem levar à insuficiência cardíaca (I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA PEDIÁTRICA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2013).

Por outro lado, a cardiotoxicidade também pode ser classificada conforme os danos dos agentes quimioterápicos, como mostra a Tabela 4.

Tabela 4 - Classificação para a cardiomiopatia relacionada à quimioterapia

<i>Cardiotoxicidade</i>	<i>Protótipo</i>	<i>Relação com dose cumulativa</i>	<i>Achados na biópsia endomiocárdica</i>	<i>Reversibilidade</i>
Tipo I	Doxorrubicina	Não	Vacúolos, destruição dos sarcômeros, necrose	Não
	Ciclofosfamida			
Tipo II	Trastuzumabe	Não	Aparência benigna ultraestrutural	Sim (maioria dos casos)
	Sunitinibe			
	Sorafenibe			

Fonte: I Diretriz Brasileira de Cardio-Oncologia da SBC 2011

2.1.4 Agentes quimioterápicos e cardiotoxicidade

Será feita uma breve explanação da Antraciclina, um dos principais agentes quimioterápicos usados no tratamento da oncologia e que foi objeto de estudo da base de dados utilizada neste trabalho.

2.1.4.1 Antraciclinas

As antraciclinas são os fármacos mais relacionados à cardiotoxicidade. Essa classe é bastante utilizada no tratamento do câncer, e tem como principais exemplos a doxorrubicina, epirrubicina e a idarrubicina (HAJJAR; HOFF; KALIL FILHO, 2012). Possui grande eficácia em tumores comuns, como linfomas e câncer de mama, inclusive em crianças com leucemia e sarcomas. A partir das antraciclinas que a cardiotoxicidade foi classificada em aguda, subaguda e crônica, como já comentado no tópico anterior (SÁ et al., 2009).

Apesar de sua eficiência no tratamento de algumas neoplasias, as antraciclinas possuem alto dano cardíaco. Além de dose cumulativa de antraciclinas, há fatores que também favorecem a ocorrência da cardiotoxicidade, como por exemplo regime de infusão, fatores de risco para doença cardiovascular, uso concomitante de outras

drogas, extremos de idade, etc. Na Tabela 5 é possível visualizar a maior incidência de disfunção ventricular ou insuficiência cardíaca para as antraciclínas.

Tabela 5 - Incidência de insuficiência cardíaca por classe terapêutica

AGENTE QUIMIOTERÁPICO	INSUFICIÊNCIA CARDÍACA	FREQUÊNCIA DE USO
ANTRACICLINAS (DOXORRUBICINA, EPIRRUBICINA, IDARRUBICINA)	5%-35% dos casos (dose acima de 500mg/m ²)	++++
AGENTES ALQUILANTES (CICLOFOSFAMIDA, IFOSFAMIDA)	5%-25% dos casos	++++
AGENTES ANTIMICROTÚBULOS (DOCETAXEL, PACLITAXEL)	2%-10% dos casos	+++
ANTICORPOS MONOCLONAIS E INIBIDORES DA TIROSINA-QUINASE		
TRASTUZUMABE	2%-28% dos casos	++
BEVACIZUMABE	2%-10% dos casos	++
SUNITINIBE	3%-10% dos casos	++

Fonte: I Diretriz Brasileira de Cardio-Oncologia da SBC 2011

2.1.5 Diagnóstico

Atualmente, ainda não existem protocolos bem estipulados que determinem o procedimento ideal para fazer a detecção precoce de alterações cardiovasculares que estão associadas à quimioterapia (YEH et al., 2004). O mais essencial de tudo é encontrar achados de sinais e sintomas clínicos de disfunções cardíacas não atribuídas a outras causas conhecidas; além de descobrir fatores de riscos que predispõem à cardiotoxicidade e que são considerados indicadores independentes para essas disfunções.

Para isto, é preciso haver uma avaliação inicial dos pacientes oncológicos que estão sob tratamento quimioterápico, com a finalidade de excluir pacientes com evidências clínicas, laboratorial e radiológica de insuficiência cardíaca antes do início do tratamento, identificar pacientes com redução da fração de ejeção, associadas a sintomas ou não, durante a quimioterapia (ALTENA, 2009).

2.1.5.1 Sinais e sintomas

É muito importante detectar sinais e sintomas de alguma das manifestações que a cardiotoxicidade traz para o paciente. Porém, muitos casos se assemelham às complicações do próprio câncer, o que faz com que a avaliação de riscos seja importante para determinação do diagnóstico. É necessário definir a classe do quimioterápico usado, sua dose cumulativa, uso prévio de outros fármacos cardiotoxícos e, não menos importante, a presença de outros fatores de risco cardiovasculares (CARDINALE, 2010).

2.1.5.2 Métodos de Imagem

Os métodos de imagem mais utilizados para detecção da CTX (cardiotoxicidade) causada pela QT (quimioterapia) são os seguintes: ecocardiograma, ventriculografia e ressonância magnética.

O ecocardiograma é a opção mais utilizada, pelo fato de ter um custo baixo, ter alta disponibilidade e não ser invasivo. Por outro lado, tem uma elevada variabilidade temporal, o que significa que não se detecta com segurança algumas alterações na FEVE. Portanto, o diagnóstico funciona, mas só com alterações tardias (RIBEIRO et al., 2018).

A VRN (Ventriculografia radionuclídica), junto com o ecocardiograma, também é um dos métodos mais utilizados no estudo da FEVE, possuindo elevada acurácia. É muito recomendada para avaliação inicial de pacientes em tratamento com antraciclinas, porém, a exposição à radioatividade é alta. Além disso, limita-se na identificação de outras anormalidades cardíacas estruturais que são altamente reconhecidas pelo ecocardiograma e ressonância nuclear magnética (PIVETA; PAIVA; ANDRADE, 2012).

Considerado o método padrão de referência na avaliação de função ventricular, a RNM (Ressonância nuclear magnética) permite a avaliação da perfusão e função miocárdica, pericárdio e detecção de massas miocárdicas. O seu custo já é mais elevado do que os métodos mencionados acima, e tem uma disponibilidade limitada (CRUZ; RODRIGUES; CAMPELO, 2015).

2.1.5.3 Eletrocardiograma

O exame de eletrocardiograma avalia o ritmo cardíaco e detecta sinais de isquemia. Não é invasivo, tem baixo custo e é um complemento a outros exames. Detecta situações clínicas como arritmias, miocardiopatias, isquemia miocárdica, entre outras. Os primeiros sinais de danos cardíacos são reconhecidos logo mesmo antes de qualquer detecção feita pelo ecocardiograma. Por ser um exame complementar fundamental no processo de diagnóstico do paciente oncológico, é importante que o eletrocardiograma seja realizado várias vezes durante o tratamento, em intervalos de tempo variados de acordo com a situação clínica (I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA PEDIÁTRICA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2013).

2.1.5.4 Biomarcadores

Os biomarcadores cardíacos vem sendo uma ferramenta bastante útil para identificar precocemente a lesão cardíaca causada por quimioterápicos em pacientes oncológicos. Eles são moléculas proteicas altamente utilizadas para detectar e medir processos biológicos normais ou patogênicos, como alterações cardiovasculares, à exemplo da insuficiência cardíaca. A partir da observação dos níveis de troponina, proteína C reativa e peptídeo natriurético cerebral, é possível medir a morbidade e mortalidade na insuficiência cardíaca (SANTOS; HAJJAR, 2012).

2.1.5.5 Biópsia endomiocárdica

Esse método de biópsia endomiocárdica é considerado um dos mais valiosos para o diagnóstico, pela sua capacidade de ser sensível e específico (BACAL; NUSSBAUM; PIRES, 2012). Ele analisa um fragmento do miocárdio in vivo, portanto é um método invasivo e não é utilizado como exame de rotina. É bastante

recomendado seu uso nas situações em que métodos não invasivos, em especial a ressonância nuclear magnética, não trazerem resultados eficientes para o paciente.

2.1.6 Tratamento

Não existem tratamentos específicos para a cardiotoxicidade induzida por qualquer que seja o fármaco utilizado na quimioterapia. Em vez disso, existem terapias padrão para determinada manifestação de dano cardíaco que o paciente desenvolve, como por exemplo a insuficiência cardíaca (VOLKOVA; RUSSELL, 2012).

Os tratamentos que serão escolhidos para cada paciente envolvem muitas variáveis, como os fatores de risco cardíaco, o tratamento farmacológico da insuficiência cardíaca e mudanças na quimioterapia para minimizar a toxicidade miocárdica (COSTA et al., 2010). Entretanto, a base dos tratamentos dispõe da administração de diuréticos, inibidores da enzima conversora de angiotensina (IECA) e beta-bloqueadores.

2.1.7 Prevenção

As estratégias de prevenção para os efeitos cardiotoxícos são de extrema importância para melhorar o prognóstico dos pacientes sobreviventes do câncer. Antes da quimioterapia, é importante avaliar parâmetros hematológicos e bioquímicos, pressão arterial sistêmica, eletrocardiograma e ecocardiograma, entre outros, para um maior monitoramento do organismo. Durante e após o tratamento, é necessária a vigilância cardiológica constante do paciente, pois há riscos de mudança da função cardíaca (COSTA et al., 2010).

Apesar de alguns fármacos serem potencialmente tóxicos para o coração, a utilização de outras medicações pode ser benéfica para a prevenção da cardiotoxicidade. À exemplo disto, tem-se o uso de medicações cardioprotetoras e o uso de antraciclinas modificadas estruturalmente, como a Epirubicina e a Idarrubicina (RIBEIRO, 2012).

Por último, mas tão importante quanto às outras medidas, o paciente precisa adotar um estilo de vida saudável. Alimentação com baixo teor de gorduras saturadas, diminuir o consumo de sal, evitar ingestão de álcool e o uso do tabaco. Muitos estudos

mostram que alimentação balanceada e prática regular de atividades físicas, aumentam a reserva cardiovascular (CRUZ; RODRIGUES; CAMPELO, 2015).

2.2 Análise Proteômica

Este tópico abordará conceitos sobre os estudos proteômicos, que, nos últimos anos, estão adquirindo grande relevância em relação à saúde, no que se refere à aplicação na doença humana.

2.2.1 Conceitos

A proteômica é o estudo em larga escala das proteínas, relacionado com um conjunto de tecnologias que tem a finalidade de separar e identificar proteínas em amostras biológicas complexas. Então, a análise proteômica por si, tem o intuito de detectar diferenças quantitativas e qualitativas na expressão protéica (GALDOS-RIVEROS et al., 2010).

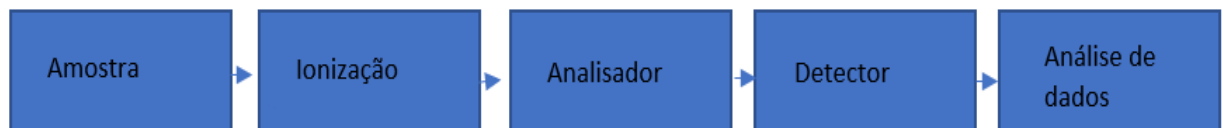
O campo de pesquisa da proteômica abre espaço à descoberta de novas drogas, terapias, diagnósticos, microbiologia, bioquímica. À medida que é possível identificar e caracterizar marcadores biológicos, ou seja, moléculas específicas de determinado estado patológico, o estudo proteômico colabora significativamente no diagnóstico precoce de doenças e na evolução do tratamento de doenças (GALDOS-RIVEROS et al., 2010).

Nos últimos anos, tipos variados de marcadores moleculares que colaboram com diagnóstico precoce e tratamento de várias patologias, como a Cardiotoxicidade, vem sendo desenvolvidos. A Espectrometria de Massa, por exemplo, é uma técnica analítica utilizada para identificar compostos desconhecidos, modificar materiais conhecidos e elucidar as propriedades químicas e estruturais da molécula (PETRICOIN et al, 2004).

2.2.2 Espectrometria de massa

A técnica de espectrometria de massa consiste na ionização de um composto e avaliação da razão massa/carga (m/z) dos íons. Utiliza-se um equipamento com fonte de ionização, um ou mais analisadores de massas e um detector. A fonte de ionização gera íons peptídicos ou proteicos, geralmente transferindo prótons para as moléculas sem alterar sua estrutura química. Dessa forma, um campo elétrico acelera o íon; o analisador de massas separa-o por m/z , ou então é selecionado de acordo com uma razão m/z já determinada e fragmentada em um processo “em tandem”. Por fim, os íons passam pelo detector, que é conectado a um computador com software de análise de dados (BARBOSA et al., 2011). O EM é esquematizado como mostra a Figura 2.

Figura 2 - Esquema simplificado de um espectrômetro de massas



Fonte: CARVALHO et al (2006)

A espectrometria de massa (EM) é uma tecnologia bastante antiga, usada no campo das engenharias e da física, mas era um pouco deixada de lado quanto às ciências da saúde pelo fato de degradar biomoléculas ao ionizá-las (CARVALHO et al., 2006). Porém, recentemente, foram desenvolvidas duas técnicas chamadas Matrix Assisted Laser Desorption Ionization (MALDI) e Electron Spray Ionization (ESI), que tornou possível a determinação precisa de massas de compostos de alto peso molecular, e de baixo peso também. Além disso, a aplicabilidade da espectrometria de massas direcionou-se para quase todas as moléculas biológicas (CARVALHO et al., 2006).

2.2.3 Ionização por Electrospray (ESI)

O electrospray, apesar de ser conhecido como uma fonte de ionização, na verdade é um processo de transferência de íons já existentes em solução para a fase

gasosa. Esse tipo de ionização forma um “spray” eletrolítico, de onde são geradas pequenas gotas carregadas, e, destas, são liberados os íons (URZEDO, 2008).

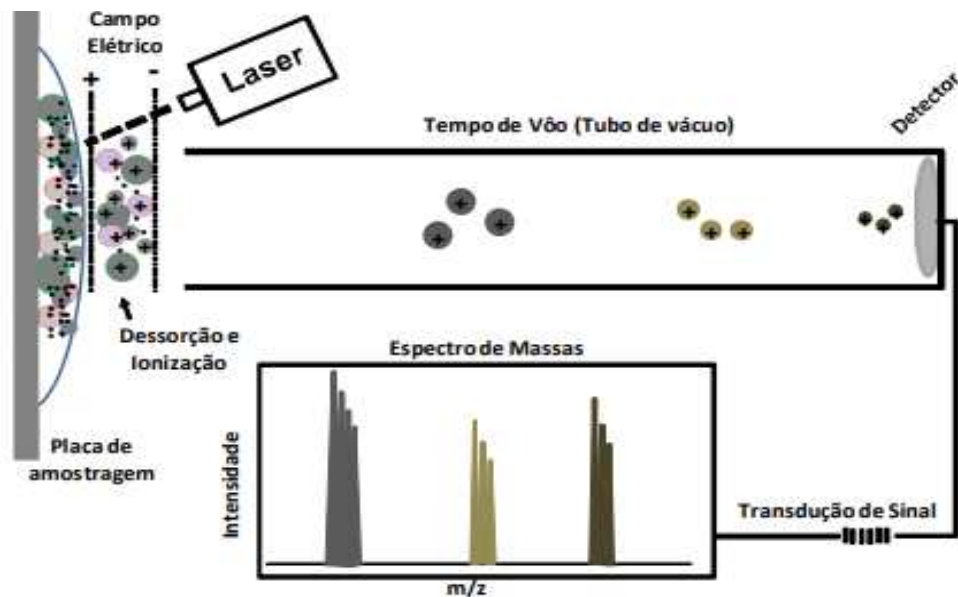
A ESI ocorre sob pressão atmosférica e alta tensão. Primeiramente, gotas altamente carregadas em condições que permitam que elas evaporem, são dispersas na solução contendo eletrólitos. Essa solução é bombeada através de um capilar, onde é aplicado um potencial positivo ou negativo, forçando um processo de oxirredução. Dessa forma, a gota formada na ponta do capilar estará rodeada de íons positivos ou negativos, dependendo do potencial aplicado. A gota ganha forma de cone, à medida que a densidade de carga aumenta, provocando uma deformação nela. A partir do momento em que a densidade de carga supera a tensão superficial, a gota desprende-se do capilar, subdividindo-se. Finalmente, os íons ficam completamente dessolvatados (DINIZ, 2011).

Essa técnica é considerada distinta dentre as outras técnicas de ionização. Ela tem capacidade de produzir íons multiplamente carregados, de tal forma que seja possível analisar compostos com alta massa molecular. Além disso, é considerada uma técnica suave, pelo motivo de não provocar tanta fragmentação dos analitos estudados (DINIZ, 2011).

2.2.3.1 MALDI

A ionização/dessorção a laser assistida por matriz (MALDI) consiste em um sistema em que a amostra biológica é colocada em uma placa que dispõe de uma matriz com composto orgânico. Já na placa, a amostra é irradiada com um laser a fim de ser vaporizada; assim, há ionização de várias moléculas. Essas moléculas são aspiradas em um tubo de vácuo e levadas a um detector de massas, sendo o mais utilizado o detector tipo TOF. Este é mais utilizado por ser o melhor detector que possui uma faixa de detecção de massas ilimitada (WILSON; WALKER, 2010). A Figura 3 mostra a esquematização de um MALDI-TOF.

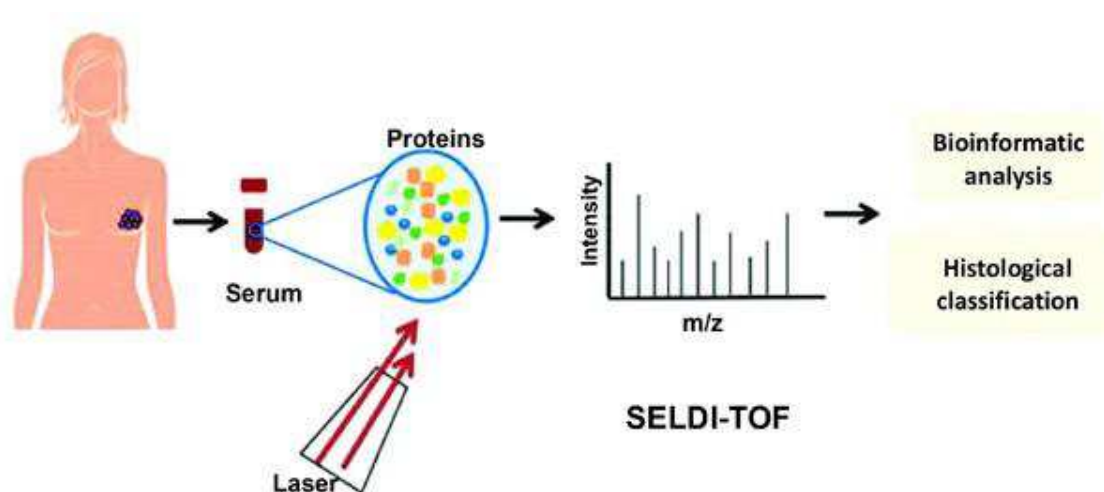
Figura 3 - Figura esquemática de um espectrômetro de massas MALDI-TOF



Fonte: ASSIS; JULIANO; JULIANO, 2011.

Existe uma tecnologia chamada SELDI, que também é uma ionização/dessorção a laser, porém, de superfície. É uma tecnologia de retenção cromatográfica baseada no arranjo de chips de proteína, que permite a análise direta de espectrometria de massa de analitos retidos na matriz. Fornece um método de triagem sensível, rápido e de alto rendimento, com a capacidade de detectar e analisar amostras complexas de proteínas (PETRICOIN; A LIOTTA, 2004).

Figura 4 - Técnica SELDI-TOF-MS aplicada em câncer de mama



Fonte: YIGITBASLI et al., 2018

A Figura 4 ilustra o passo-a-passo, em um exemplo de amostra de câncer de mama sendo convertida em sinais proteômicos para futuras análises bioinformáticas e classificações histológicas.

2.3 Biomedicina e estudo proteômico

A bioinformática, de um modo geral, é a utilização de software e análise de bancos de dados que, juntos, visam dar um norte à novas pesquisas que precisam de uma análise de dados junto com simulações de experimentos. A partir do conhecimento sobre os processos biológicos e sobre as prioridades dos seres vivos, e a necessidade de resolver problemas criando produtos de utilidade, surge a biotecnologia, que está ligada ao desenvolvimento de novas tecnologias e produtos (HERNANDEZ; SAN MIGUEL; MARTÍN-GIL, 2010).

A biotecnologia e a bioinformática estão interligadas a partir do momento em que são gerados novos dados e conhecimentos que podem ser aplicados para desenvolver novos produtos e novas soluções.

Os experimentos feitos em laboratórios necessitam do complemento de análises virtuais feitas com o auxílio do computador (HOCHSTRASSER, 1998). Para isso, foram desenvolvidos softwares para analisar a separação eletroforetica, e ferramentas bioinformáticas, que podem ser encontradas na Internet. Essas bases de dados consistem em mapas de proteína do plasma humano, da urina, dos tecidos da mama, do coração, de acordo com o que for mais interessante para a pesquisa (BAIROCH; APWEILER, 1998).

As pesquisas da biomedicina com base na proteômica, se fundamentam em cima do tipo de amostra utilizada, em que se destacam as pesquisas com linhagens celulares, com tecidos e proteômica de fluidos. A maior facilidade de obtenção e processamento das amostras vem da pesquisa com fluidos. A proteômica de fluidos é bastante utilizada para o estudo de diversas doenças, inclusive as neoplásicas e toxicológicas (GALDOS-RIVEROS et al., 2010).

2.4 Diagnóstico Auxiliado por Computador

O CAD (Diagnóstico Auxiliado por Computador) é um esquema computacional que auxilia o diagnóstico de determinadas doenças e complementa interpretações do

especialista que está sujeito a erros de detecção (DOI, 2007). Diante desses fatos, é muito importante que haja uma integração entre ferramentas computacionais e interpretação de imagens pelo radiologista (FERREIRA JUNIOR; OLIVEIRA, 2007), claramente sem subestimar a capacidade humana, que é necessária na tomada de decisões.

Ao longo dos anos, foram desenvolvidos dois tipos diferentes de esquemas de CAD, um para auxílio à detecção de lesões e outro para diagnóstico com base na classificação de lesões malignas e benignas (DOI, 2007). O primeiro tipo funciona a partir da localização de padrões anormais por intermédio da varredura da imagem pelo computador. Já o outro, é executado através da quantificação de características da imagem e sua classificação correspondente a padrões normais ou anormais (AZEVEDO-MARQUES, 2001).

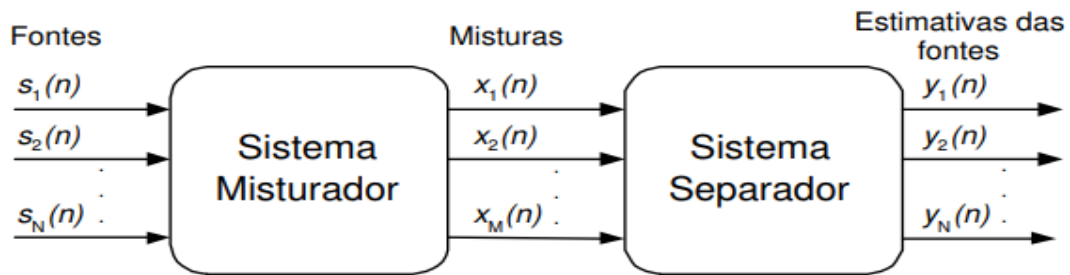
Os sistemas CAD são provenientes de temas da Visão Computacional e da Inteligência Artificial. A Visão Computacional é a ciência em que computadores são treinados para extrair informações/atributos de uma imagem em formato digital, dividida em uma etapa de processamento de imagem para realce e segmentação das lesões, e uma etapa de quantificação de atributos da imagem. De uma forma geral, primeiramente a segmentação vai descrever a imagem através de atributos, e, então, estes serão quantificados a partir de propriedades métricas, topológicas e de textura dos objetos. Finalizado esse processo, é feita uma classificação das estruturas presentes na imagem, papel da Inteligência Artificial, utilizando técnicas de algoritmos genéticos e classificadores (AZEVEDO-MARQUES, 2001).

2.5 Análise de Componentes Independentes

2.5.1 Separação Cega de Fontes

A Análise de Componentes Independentes (ICA) está intimamente relacionada com o problema de Separação Cega de Fontes (BSS, Blind Source Separation), devido à sua grande importância no contexto da BSS (DUARTE, 2006). A Figura 5 ilustra o problema geral comentado.

Figura 5 – Modelo do Problema de Separação Cega de Fontes



Fonte: DUARTE, 2006.

O problema é descrito da seguinte forma: um conjunto de N sinais fontes é posto na entrada de um sistema misturador, ou seja, um sistema em que vai misturar os sinais fontes, gerando M saídas. O objetivo da separação de fontes é recuperar os sinais originais, através de um sistema separador com as amostras das misturas. O problema denomina-se “cego” pela finalidade de recuperar estes sinais, mesmo com a falta de informação sobre o sistema misturador e as fontes (DUARTE, 2006).

2.5.2 Definição e Método ICA

A ICA é uma técnica que corresponde a um caminho viável para identificar e separar fontes estatisticamente independentes, a partir de uma mistura de fontes originais (FRANCO, 2008). O seu objetivo é extrair informações úteis ou fornecer sinais de dados, que podem estar na forma de imagens, ações, sons. Dessa forma, a ICA é uma técnica com aplicações em sinais médicos, ensaios biológicos e sinais de áudio (THARWAT, 2018).

Para ser realizada, a ICA necessita que as fontes sejam sinais aleatórios e independentes entre si (OLIVEIRA, 2015). Portanto, para executar o método ICA, uma condição deve ser satisfeita para que elementos de um vetor s sejam estaticamente independentes entre si:

$$p_s(s) = p_{s_1}(s_1) \cdot p_{s_2}(s_2) \dots p_{s_N}(s_N), \quad (2.1)$$

onde $p_s(s)$ é a função densidade de probabilidade (pdf) conjunta dos elementos s_1, s_2, \dots, s_N e $p_{s_i}(s_i)$.

Sejam n variáveis aleatórias observadas x_1, \dots, x_N , sendo as misturas captadas pelos sensores no problema BSS. Cada x_i será representado como combinação linear de n variáveis aleatórias s_1, s_2, \dots, s_N , tal que:

$$x_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{iN}s_N, \text{ para todo } i, j = 1, \dots, n \quad (2.2)$$

onde os a_{ij} são os coeficientes (pesos) da mistura linear, que podem ser fatores externos que influenciam na estrutura das fontes (SILVA, 2010).

Considerando um modelo matricial, onde A é a matriz formada pelos coeficientes de mistura, tem-se:

$$x = A.s, \quad (2.3)$$

sendo a matriz A uma matriz constante de dimensão $M \times N$, e os elementos de $s = [s_1, s_2, \dots, s_N]$ estaticamente independentes entre si (DUARTE, 2006). O modelo definido na Equação 2.2 é chamado de modelo de análise de componentes independentes, em que é preciso estimar tanto s quanto a matriz de mistura A , que também é desconhecida, pois apenas o vetor x é conhecido. Dessa forma, a finalidade do método ICA é estimar a matriz A apenas com informações contidas na matriz x (HYVARINEN; KARHUNEN; OJA, 2001).

Diante do problema de estimar a matriz com pouca informação, é necessário que sejam feitas suposições gerais, discutidas no próximo tópico, a partir das considerações de independência e não-gaussianidade das fontes. A ICA estima um conjunto de fontes y_i também independentes, através de um sistema de separação W quando somente as medidas do sinal x_i forem conhecidas. Assim, pode-se restabelecer o problema de encontrar os componentes independentes como:

$$y = W.x, \quad (2.4)$$

em que y é o vetor de Componentes Independentes que aproxima s e o W uma estimativa da matriz inversa A^{-1} .

Depois que a matriz de mistura A é estimada, é possível a obtenção dos componentes independentes a partir da sua matriz inversa A^{-1} (FRANCO, 2008):

$$s = A^{-1}.x \quad (2.5)$$

Além da aplicação em BSS, o método ICA também se aplica a outros problemas que não tem relação com problema de separação de sinais, como: extração de características, compressão de dados, econometria, estágio de pré-processamento em sistemas classificadores, etc (DUARTE, 2006).

2.5.3 Restrições

Para que haja eficácia no desenvolvimento do método ICA, é necessário que sejam colocadas algumas restrições nas variáveis. São feitas as seguintes pressuposições (FRANCO, 2008):

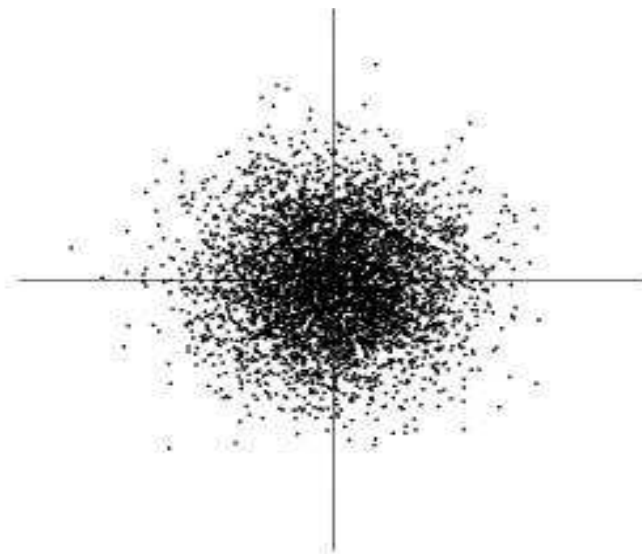
1. As fontes devem ser estatisticamente independentes.

Variáveis aleatórias são estatisticamente independentes se a informação de uma não fornece qualquer informação em relação a outra, como foi demonstrado no subtópico anterior.

2. As fontes independentes precisam ter distribuições de probabilidade não-gaussianas.

A distribuição de duas variáveis gaussianas gera uma figura de densidade completamente simétrica, como mostra a Figura 6.

Figura 6 – Distribuição de duas variáveis gaussianas independentes



Fonte: Hyvärinen e Oja (2000, p.26)

Como o gráfico é considerado isotrópico, ou seja, suas propriedades físicas são iguais em todas as direções, não é viável que se estime a matriz de mistura das fontes originais.

3. A matriz de mistura deve ser quadrada.

Se a matriz de mistura está na sua forma quadrada, impõe-se que o número de fontes independentes seja igual ao número de sinais observados; assim, pode possuir matriz inversa.

2.5.4 Pré-processamento dos dados

A maioria dos métodos ICA tem a necessidade de que seus dados passem por um pré-processamento antes de utilizar o algoritmo ICA. Existem variados tipos de técnicas, porém, as principais que serão discutidas são a Centralização e o Branqueamento (HYVARINEN; KARHUNEN; OJA, 2001). Esta etapa é bastante útil pelo motivo de que elimina e simplifica parte do trabalho do algoritmo que será escolhido.

2.5.4.1 Centralização

Neste primeiro passo, as misturas observadas passam pela fase de centralização, ou seja, subtrai-se das misturas a sua média, para que seja observado que elas possuem média zero. Considerando que tem média zero, as fontes independentes também terão. Não é um tipo de pré-processamento obrigatório, mas simplifica os algoritmos (HYVARINEN; KARHUNEN; OJA, 2001).

2.5.4.2 Branqueamento

A fase de branqueamento é feita logo após a centralização, sendo um pouco mais complexa de ser calculada. Nesta, o vetor de misturas observadas é linearmente transformado em um novo vetor, em que seus componentes são descorrelacionados e sua matriz de covariância é igual a matriz identidade, ou seja, sua variância é igual a 1 (OLIVEIRA, 2015).

2.5.5 Algoritmo FastICA

Após a fase de pré-processamento, o algoritmo de resolução da ICA ajuda na otimização das estruturas. Algumas versões de algoritmos foram implementadas para otimizar o método ICA, mas será enfatizado uns dos algoritmos mais utilizados para essa finalidade: o Algoritmo FastICA, desenvolvido por Hyvärinen e Oja.

O Algoritmo FastICA, ou Algoritmo Rápido de Ponto Fixo, é uma implementação simples e relativamente rápida para separar fontes independentes de uma mistura linear (CASTANEDA-VILLA; CALDERON-RIOS; JIMENEZ-GONZALEZ, 2015). Ele baseia-se em uma estratégia iterativa de ponto fixo com o objetivo de determinar sequencialmente os componentes não-gaussianos máximos. Além disso, é o algoritmo ICA que tem menos interações na precisão da separação (CARMO et al., 2017).

Para encontrar os componentes independentes, o algoritmo FastICA utiliza o procedimento de maximização da negentropia, que é uma medida baseada na entropia diferencial (DAMASCENO, 2011). Há diversas formas de se estimar a Negentropia de um sinal, mas a abordagem baseada em expectância E (HYVARINEN; KARHUNEN; OJA, 2001) é descrita pela equação 2.6:

$$J(y) \approx \sum_{i=1}^N k_i \left[E(G_i(y)) - E(G_i(y_{gaus})) \right]^2 \quad (2.6)$$

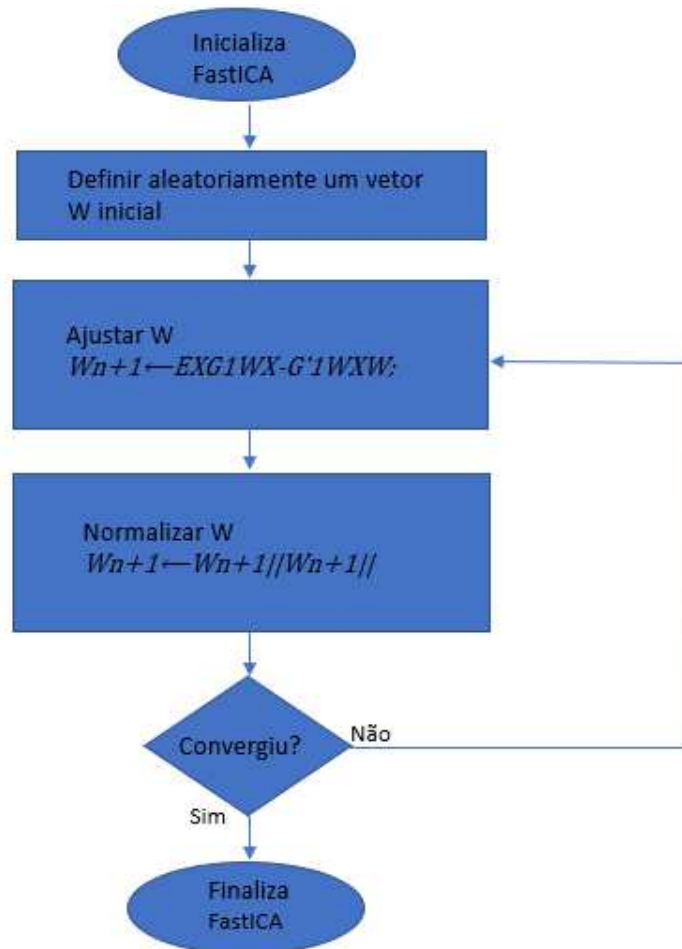
em que k_i são constantes positivas, E é o operador expectância, y_{gaus} variáveis gaussianas com variância unitária e média zero, os G_i são funções não quadráticas (LINHARES et al., 2016). As funções G_1 e G_2 , descritas nas equações 2.7 e 2.8, asseguram boas aproximações da negentropia e melhoram a convergência do algoritmo FastICA (HYVARINEN; KARHUNEN; OJA, 2001).

$$G_1(y) = \frac{1}{\beta} \log(\cosh(\beta y)), \text{ com } 1 \leq \beta \leq 2 \quad (2.7)$$

$$G_2(y) = -\exp\left(-\frac{y^2}{2}\right) \quad (2.8)$$

Para executar o FastICA, são seguidos alguns passos, como mostra o fluxograma da Figura 7.

Figura 7 – Fluxograma do algoritmo FastICA



Fonte: Autor.

Neste caso, a convergência acontece quando os valores atual e anterior de W tendem para a mesma direção, ou seja, o produto interno entre esses valores é aproximadamente igual a 1. Além de ter uma convergência mais rápida do que de outros algoritmos, não há necessidade de definir parâmetros, tornando-o fácil de usar. Os componentes podem ser estimados cada um de uma vez, facilitando a análise dos dados, e, assim, diminuindo o custo computacional, o que existe um espaço de memória reduzido (HYVARINEN; KARHUNEN; OJA, 2001).

2.6 Seleção de Características Discriminativas

O processo de seleção de características mais relevantes em um sistema de reconhecimento de padrões é de grande importância. Ele pode reduzir o problema da dimensionalidade, ou seja, diminuir o número necessário de amostras de treinamento para obter-se bons resultados com um classificador. A remoção de características irrelevantes não compromete o resultado da classificação e melhora sua performance, diminuindo custos computacionais e reduzindo tempo de classificação (LINHARES et al., 2016).

A seleção de características consiste na escolha de determinadas características provenientes de um conjunto produzido a partir do sinal original, sem perder sua capacidade discriminante (CATARINO, 2009). As características podem ser selecionadas de várias formas, entre elas a Máxima Relevância, que seleciona as características que mais se correlacionam com a variável de classe (CATARINO, 2009), e a Mínima Redundância, que exclui informações redundantes sem comprometer o conjunto de características (DING; PENG, 2005).

A medida de Máxima Relevância é dada pela informação mútua I entre a variável de classe c e cada característica x_i , em um conjunto A de características, como mostra a equação 2.8 (CATARINO, 2009):

$$\max D(A, c), D = \frac{1}{|A|} \sum_{x_i \in A} I(x_i; c). \quad (2.8)$$

Enquanto a medida de Mínima Redundância, em termos da informação mútua I , entre duas características x_i e x_j , é dada pela equação 2.9 abaixo:

$$\min R(A), R = \frac{1}{A^2} \sum_{x_i x_j \in A} I(x_i; x_j). \quad (2.9)$$

2.6.1 Máxima Relevância e Mínima Redundância

O algoritmo de Máxima Relevância e Mínima Redundância (mRMR) é utilizado como meio para reduzir o conjunto de características, selecionando um conjunto com características mais relevantes e menos redundantes (CATARINO, 2009). O mRMR combina as equações 2.8 e 2.9 para encontrar a equação 2.10, que apresentará um

conjunto de características mais relevantes e menos redundantes, após um processo de otimização (DING; PENG, 2005).

$$\max \phi(D, R), \phi(D, R) = D - R \quad (2.10)$$

Depois que a dimensionalidade do vetor de características é reduzida pela técnica de mRMR, é feita a classificação das amostras com o tipo de classificador adequado (PENG; LONG; DING, 2005).

2.7 Aprendizagem de Máquina

O Aprendizado de Máquina (AM), ou Learning Machine, é uma ciência proveniente da Inteligência Artificial que tem como objetivo desenvolver técnicas ou sistemas capazes de adquirir conhecimento e aprender com erros através de experiências acumuladas (FERIS et al., 2001). Em outras palavras, é um conjunto de técnicas desenvolvidas capazes de extrair conhecimento a partir de amostra de dados (MITCHELL, 1997).

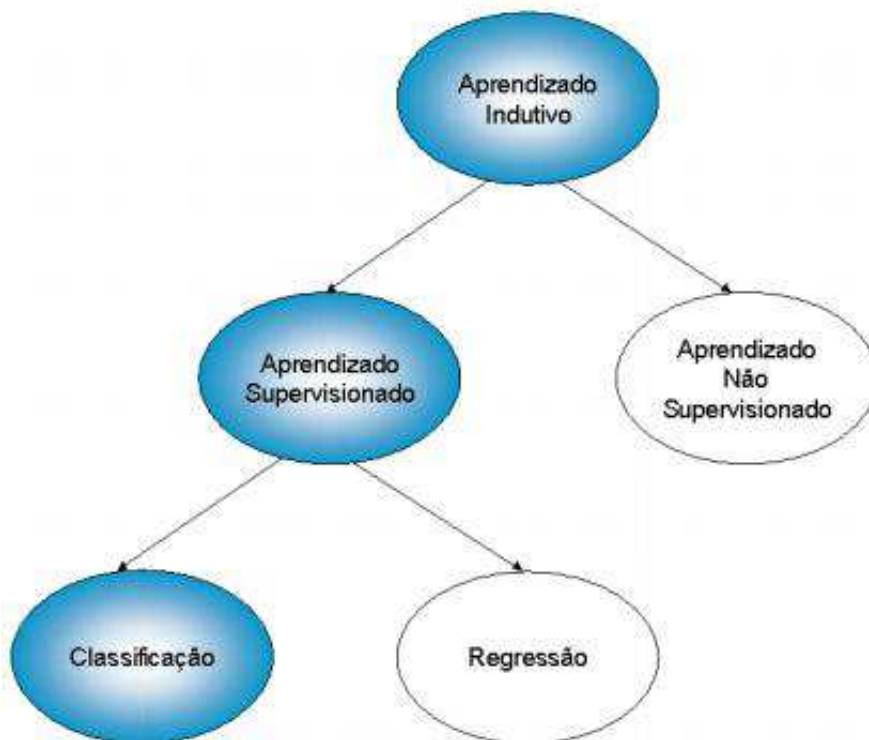
O AM divide-se em duas categorias: aprendizado supervisionado e não supervisionado (HAYKIN, 1999). No aprendizado de máquina supervisionado os algoritmos de AM utilizam um conjunto de dados, sejam de entrada ou saída, para treinamento, e, então, será extraído uma representação do conhecimento desses dados; dessa forma, essa representação poderá ser capaz de produzir saídas corretas para novas entradas. Já o AM não supervisionado, não há presença de exemplos rotulados da função a ser aprendida pela rede (HAYKIN, 1999). Será dada ênfase apenas no AM supervisionado.

Como foi esclarecido, o AM supervisionado adota paradigmas cujos dados são rotulados. Para cada entrada de dados, um agente externo indica se a saída calculada é boa ou ruim; assim, a resposta fornecida pela rede neural é comparada à resposta esperada (FERNEDA, 2006). Dessa forma, se houver erro, este é informado à rede para que sejam feitos reparos e, assim, ter respostas mais eficientes.

Problemas de aprendizado supervisionado são classificados em problemas de regressão e classificação (MCCUE, 2007), como é observado na Figura 8 (MONARD; BARANAUSKAS, 2003). Eles são caracterizados a partir do atributo que descreverá uma característica de interesse do problema. Se o atributo tiver valor discreto, o

problema é de Classificação; se for contínuo, será um problema de Regressão (PRATI, 2006). Enquanto a classificação consiste em descobrir uma função que classifique um conjunto de registro de uma classe já pré-estabelecida (GALVAO; MARIN, 2009), a regressão tem o objetivo de prever o valor do atributo a partir das características dos exemplos (PRATI, 2006).

Figura 8 – A hierarquia do aprendizado



Fonte: BARANAUSKAS, MONARD, 2003, p.41.

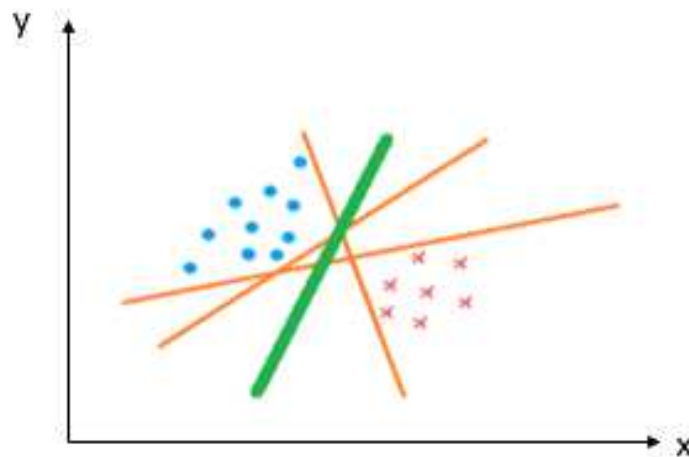
2.8 Máquina de Vetores de Suporte

A Máquina de Vetores de Suporte, *Support Vector Machines* (SVM), é uma técnica de aprendizado de máquina do tipo supervisionada, criada por Vapnick em 1965, que tem a finalidade de resolver problemas de regressão e classificação (LINHARES et al., 2016). Ela vem sendo cada vez mais utilizada na solução de reconhecimento de caracteres, detecção de faces e imagens e categorização de textos (FERIS, 2001).

A ideia do SVM está em estimar uma função para classificar os dados em duas classes, com base no princípio de separação ótima de classes. Isso é feito através de um hiperplano que vai ser considerado uma superfície de decisão para maximizar a margem e o ponto mais próximo de uma das classes predefinidas (GAROFALO,

2015), como é mostrado na Figura 9. Essa maximização faz com que a SVM seja robusta em relação a uma grande dimensão de dados, tendo assim, uma boa capacidade de generalização e suporte ruídos nos dados (RODRIGUES, 2015).

Figura 9 – Possíveis hiperplanos de separação e hiperplano ótimo.

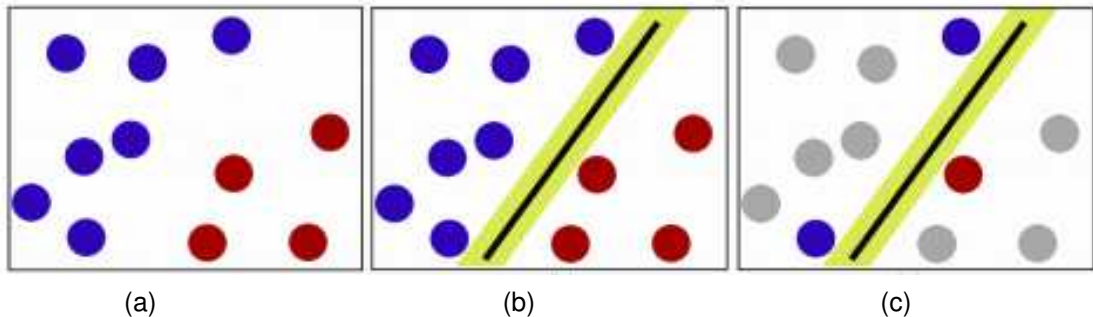


Fonte: Adaptado de TAKAKURA et al., 2017, p.80.

Na Figura 8 existe um conjunto de classificadores lineares que separam duas classes (bolinhas azuis e "x" vermelhos), que são os possíveis hiperplanos (linhas laranjas) inclusive um que maximiza a margem de separação, que é o hiperplano ótimo (linha verde), que representa o classificador do SVM (TAKAKURA et al., 2017). Esse hiperplano tende a separar o conjunto de pontos de cada classe de modo que a maioria dos pontos de uma mesma classe estejam do mesmo lado (MARINS, 2008).

Existem conjuntos linearmente separáveis e os não linearmente separáveis. Na Figura 10 (a) é possível notar pontos de dois tipos de classes, de cores vermelha e azul. Elas são linearmente separáveis pois existe uma linha capaz de dividi-las, de forma que as margens entre as classes seja a maior possível (NORONHA; FERNANDES, 2016), de acordo com a Figura 10 (b). Constata-se que é necessário apenas os dados mais próximos da linha de classificação, como observa-se na Figura 10 (c); e esses dados relevantes são os Vetores de Suporte (NORONHA; FERNANDES, 2016).

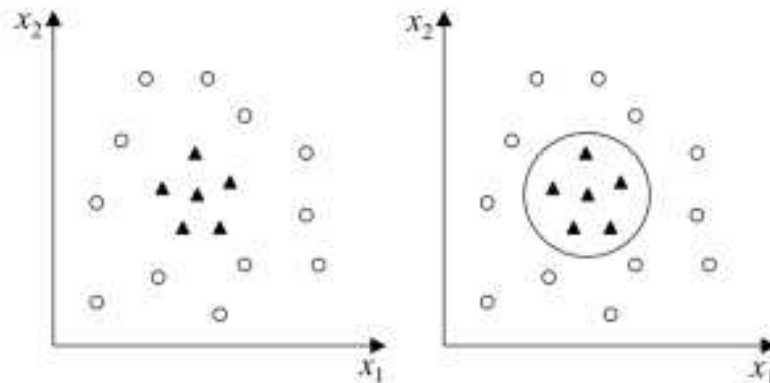
Figura 10 – Conjunto de dados linearmente separável



Fonte: NORONHA, FERNANDES, 2016.

Há casos em que não é possível separar classes através de um hiperplano, que são os chamados conjuntos não linearmente separáveis. Ao observar a Figura 11 percebe-se que não existe nenhuma de separar as duas classes de forma linear; situação em que o uso de uma fronteira curva seria mais adequado para separar as classes (LORENA; CARVALHO, 2007).

Figura 11 – Conjunto de dados não linearmente separável



Fonte: Adaptado de LORENA, de CARVALHO, 2007, p.60.

Dado um conjunto de pontos de treino (x_i, y_i) , em um espaço n -dimensional, com $x_i \in \mathbb{R}^n, i = 1, 2, \dots, n$, onde cada ponto x_i são as entradas e pertencem a uma das duas classes identificadas pelo rótulo $y_i \in \{-1, 1\}$ (OLIVEIRA, 2013). O objetivo é estimar uma função $f: \mathbb{R}^n \rightarrow \{-1, +1\}$, que separe as duas classes através do hiperplano, de forma que a distância para os vetores de suporte é maximizada. A

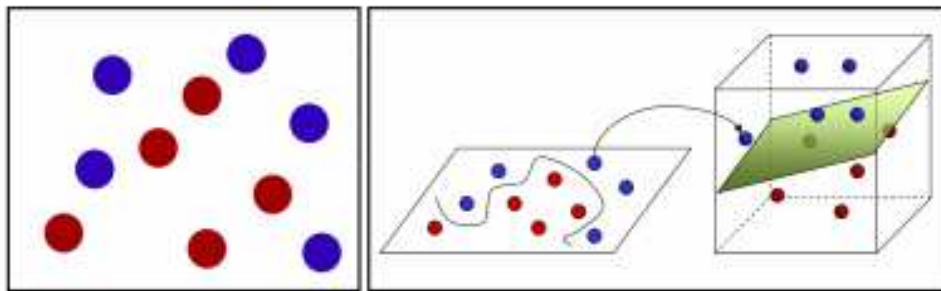
função que separa as classes através de hiperplanos pode ser definida como (TAKAHASHI, 2012):

$$w^T \cdot x + b = 0 \quad (2.11)$$

em que $w^T \cdot x$ é o produto escalar entre os vetores w e x , e o w é o vetor de pesos ajustáveis, e b é um limiar conhecido como bias.

Em situações de conjuntos não linearmente separáveis, faz-se o mapeamento dos dados para um espaço de dimensão maior, como exibido na Figura 12. Nesse espaço, existe uma alta probabilidade que sejam classificados por um hiperplano (HAYKIN, 2008). Para isso, utilizam-se funções Kernel,

Figura 12 – Mapeamento de dados para um espaço de dimensão maior



Fonte: NORONHA, FERNANDES, 2016.

que tem a capacidade de representar espaços abstratos. A Tabela 6 mostra as funções Kernel que são mais utilizadas e apresentam bons resultados de classificação (LINHARES et al., 2016). A função de base radial (RBF) é um tipo de função de núcleo bastante utilizada em classificações com SVM, mostrando resultados satisfatórios.

Tabela 6 – Funções Kernel

TIPO DE FUNÇÃO	FORMA MATEMÁTICA
FUNÇÃO DE BASE RADIAL	$k(x_i, x_j) = e^{-\gamma x_i-x_j ^2}$
FUNÇÃO POLINOMIAL	$k(x_i, x_j) = (1 + x_i \cdot x_j)^n$
FUNÇÃO SIGMOIDAL	$k(x_i, x_j) = \tanh(ax_i \cdot x_j + b)$

Fonte: LINHARES et al, 2016, p.284.

2.9 Análise Discriminante Linear

Análise Discriminante é uma técnica utilizada para discriminar e classificar objetos (KHATTREE; NAIK, 2000), utilizando funções discriminantes lineares ou quadráticas. Portanto, a Análise Discriminante Linear, *Linear Discriminant Analysis* (LDA), é um critério de maximização da separação entre duas ou mais classes de amostras (KITANI; THOMAZ, 2007), que utiliza funções discriminantes lineares. A LDA tem sido usada com sucesso como técnica de extração de características estatísticas em vários problemas de classificação (THOMAZ; KITANI; GILLIES, 2006).

Para desenvolver a técnica LDA, é necessário seguir alguns passos. Considerando que as classes sejam π_1 e π_2 , primeiramente calcula-se a separabilidade entre diferentes classes, ou seja, a distância entre a média de diferentes classes:

$$D(X) = L' \cdot X = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot X \quad (2.12)$$

em que, $X = [X_1, X_2, \dots, X_p]$, $\pi = [\pi_1, \pi_2]$, L é o vetor discriminante, X é o vetor aleatório de características das populações, μ o vetor de médias e Σ a matriz comum de covariâncias das classes (JOHNSON; WICHERN, 1999). A função discriminante linear para uma dada observação x_0 é:

$$D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 \quad (2.13)$$

e o ponto médio entre as duas médias μ_1 e μ_2 das classes é:

$$m = \frac{1}{2} [\mu_1 - \mu_2]' \cdot \Sigma^{-1} [\mu_1 + \mu_2] \quad (2.14)$$

$$m = \frac{1}{2} [D(\mu_1) + D(\mu_2)] \quad (2.15)$$

Dessa forma, classifica-se baseado nas condições abaixo:

Aloca-se x_0 em π_1 se $D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 \geq m$

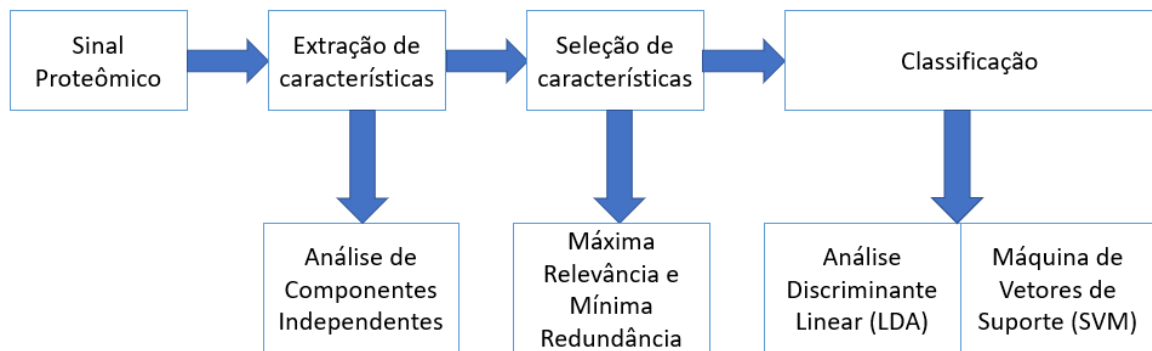
Aloca-se x_0 em π_2 se $D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 < m$.

3. METODOLOGIA

A metodologia deste trabalho será dedicada exclusivamente a explicar o processo de desenvolvimento do método proposto, que tem a intenção de auxiliar o diagnóstico de cardiotoxicidade, a partir dos padrões proteômicos da base utilizada. A elaboração deste método é composta por três etapas: a extração das características dos sinais proteômicos através da Análise de Componentes Independentes (ICA), a redução da dimensionalidade das características utilizando a técnica de Máxima Relevância e Mínima Redundância (mRMR), e a classificação final pelas técnicas de SVM E LDA.

Todos os métodos discutidos neste trabalho serão relatados nas subseções abaixo. A Figura 13 ilustra todo o procedimento.

Figura 13 - Diagrama em blocos da metodologia proposta



Fonte: Autor.

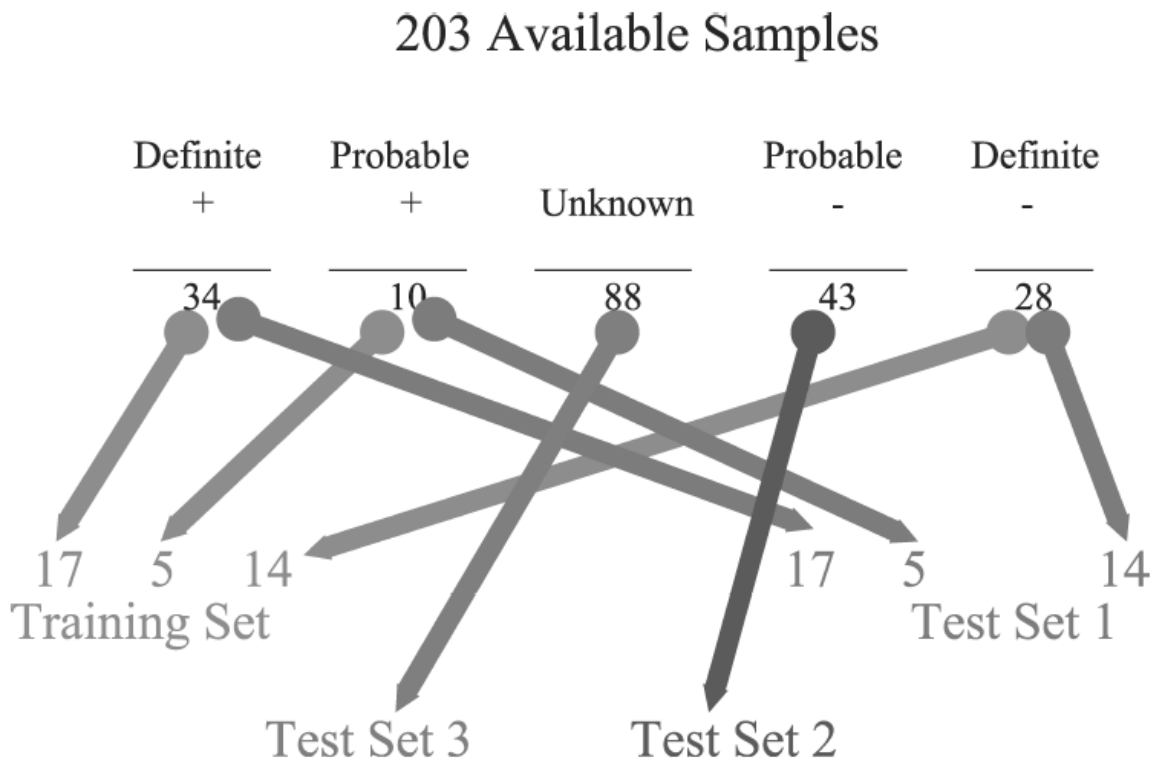
A implementação e os testes foram efetuados usando a linguagem MatLab, R2017b e fazendo uso dos pacotes *fastICA* (AAPO, 2015) e *mRMR* (Mathworks,2015), e, finalmente o pacote SVM obtido em Almeida (2016).

3.1 Dataset

Em primeiro lugar, foi feita a aquisição de dados, etapa em que é realizada a obtenção das amostras que serão utilizadas (PETRICOIN et al., 2004), que são provenientes de soro de modelos de ratos induzidos por quimioterápicos. A base de dados utilizada é de domínio público e pode ser adquirida no site (PETRICOIN; LIOTTA, 2004). Essa e outras bases de dados são disponibilizadas para estudos e avaliações do efeito das mudanças nas condições experimentais sobre as

características desses espectros. A Figura 14 esclarece o cenário de estudo das amostras recolhidas da base de dados. Foi um total de 203 espécimes de ratos, sendo 34 definidas como positivas e 28 definidas como negativas, as quais foram os objetos de estudo deste trabalho. O resto foi considerado como provável ou desconhecido.

Figura 14 – Amostras disponíveis para conversão em padrões proteômicos

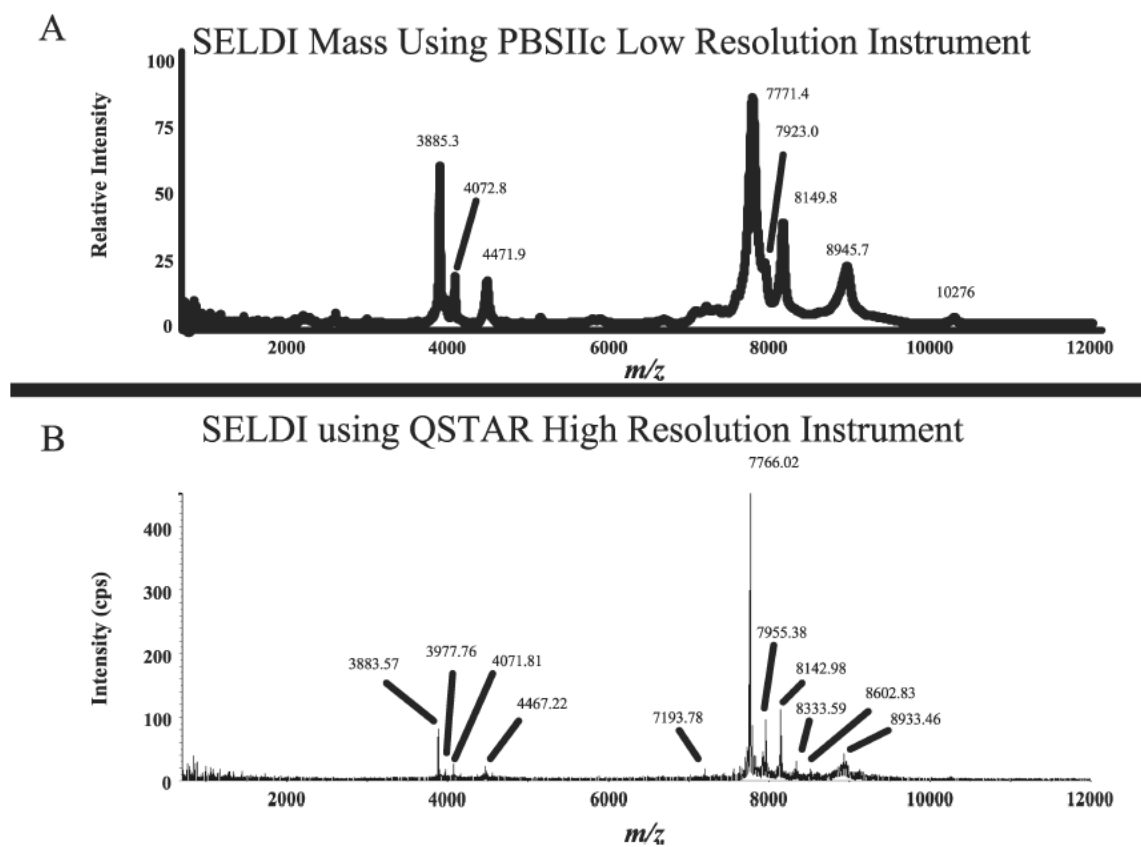


Fonte: PETRICOIN et al., 2004.

Os dados utilizados foram baseados em padrões proteômicos usando a técnica de espectrometria de massas SELDI-TOF, que é conhecida por ser um padrão de informação preciso para auxiliar no diagnóstico de pacientes com cardiotoxicidade. No total são 62 amostras, sendo 34 delas com diagnóstico de cardiotoxicidade e 28 com diagnóstico normal (grupo de controle); cada amostra com 373257 características diferentes. Dessa forma, obteve-se uma matriz X_{nm} de ordem 62×373257 contendo as características de cada amostra.

A Figura 15 aponta a forma como as amostras extraídas pelo espectrômetro de massas foram convertidas em um sinal multinível, em que é possível ver níveis de intensidade proteômicos encontrados no espectro de massa. Mostra a comparação entre um espectro de massa SELDI-TOF com baixa resolução (A) e alta resolução (B). O eixo y corresponde ao nível de intensidade do espectro, e o eixo x equivale à razão massa/carga.

Figura 15 – Espectro de massas



Fonte: PETRICOIN et al., 2004.

3.2 Extração de características

Após as amostras serem adquiridas nas bases de dados, o próximo passo é o de extração de características. Essa etapa tem o objetivo de extrair as características mais relevantes que representam de maneira significativa as diferenças entre os dados dos grupos analisados, ou seja, classificar corretamente se a amostra (sinal proteômico) é de um paciente portador ou não da cardiotoxicidade.

Para extração das características, será utilizada a técnica ICA, vista na Seção 2.5. Com a junção da matriz dos casos com cardiotoxicidade, de dimensões 34×373257 , com a matriz dos casos normais de dimensões 28×373257 , obtém-se a matriz X de ordem 62×373257 para ser utilizada como entrada na técnica ICA. Cada linha da matriz X corresponde a uma amostra e cada coluna a uma característica de cada amostra.

A matriz X servirá de entrada para o algoritmo FastICA para que possam ser obtidas as funções de base da matriz A , de dimensão 62×60 , que contém as características de cada amostra. Cada linha da matriz A corresponde a uma amostra, e cada coluna a uma característica, ou seja, um parâmetro de entrada para o classificador.

3.3 Seleção das características mais significativas

O objetivo desta fase é identificar as características mais significativas dentro do vetor de características observado, visto que representarão melhor os dados gerados na etapa de extração de características. Esta tarefa é considerada essencial pois diminui o custo computacional e evita que o resultado seja insatisfatório e com baixa acurácia. A etapa será realizada com a ajuda do algoritmo de Máxima Relevância e Mínima Redundância (mRMR), revisada na Seção 2.6.1.

3.4 Classificação

Durante a etapa de classificação serão utilizados dois tipos de classificadores: Máquina de vetores de suporte (SVM) e Análise Discriminante Linear (LDA), descritos nas Seções 2.8 e 2.9, respectivamente.

Para validar os resultados das técnicas usadas, será utilizada a técnica estatística de validação cruzada *k-fold-cross validation* (KOVAHI, 1995), cujo objetivo é dividir aleatoriamente o conjunto de dados em k subconjuntos aproximadamente iguais para testar o modelo durante a fase de treinamento. O conjunto de dados é dividido em 10 subconjuntos (10-fold-cross), em que 9 deles são mantidos para treinar o desempenho do modelo, e o subconjunto restante é usado para o teste. Este processo é realizado k vezes, ou seja, 10 vezes, alternando o conjunto de teste de forma circular. No final de cada iteração, calcula-se a acurácia sobre os erros, e, assim, a acurácia final é calculada com a média de cada fase.

3.5 Métricas de desempenho

O processo de avaliação dos resultados depende de algumas métricas de desempenhos que são utilizadas para fazer medidas estatísticas sobre os resultados dos testes (BUSHBERG, 2012). Os resultados de testes são divididos em: Verdadeiro Positivo (*VP*), Verdadeiro Negativo (*VN*), Falso positivo (*FP*) e Falso Negativo (*FN*). *VP* e *VN* é o número de amostras classificadas de forma correta (FLETCHER, 2006), enquanto *FP* e *FN* representam a quantidade de amostras classificadas de modo errôneo.

Os tipos de resultados mencionados acima são utilizados para medir o desempenho do modelo de classificação aplicado através das métricas. Cada métrica é relacionada a uma característica específica da capacidade e desempenho do sistema. As três métricas utilizadas neste trabalho foram: Acurácia, Especificidade e Sensibilidade.

A Acurácia (*A*) é a proporção de acertos do classificador durante a fase de teste (PEREIRA, 2008), que calcula a capacidade do modelo de acertar o diagnóstico. Ela é definida pela seguinte proporção:

$$A = \frac{VP+VN}{VP+VN+FP+FN} \quad (3.1)$$

A Especificidade (*E*) é a taxa de verdadeiros negativos que são classificados corretamente pelo teste, ou seja, demonstra a capacidade do modelo em reconhecer as “amostras saudáveis”. Sua definição é calculada por:

$$E = \frac{VN}{VN+FP} \quad (3.2)$$

A Sensibilidade (*S*) é a taxa de verdadeiros positivos que são classificados corretamente pelo teste. É a capacidade do modelo em reconhecer as “amostras doentes”, definida por:

$$S = \frac{VP}{VP+FN} \quad (3.3)$$

Por conseguinte, o resultado final da classificação dentro da matriz conterá todos os valores, desde o primeiro teste com uma característica até o último com 60

características. O resultado estará disposto da seguinte forma: $FINAL = [k \ VP \ VN \ FP \ FN \ Acurácia \ Sensibilidade \ Especificidade]$, em que “k” é o número de características.

4 RESULTADOS E DISCUSSÕES

Este capítulo é destinado a analisar e discutir os resultados obtidos nas abordagens utilizadas expostas no Capítulo 3. Para o método proposto, foi criado um projeto com um arquivo principal (ANEXO A) que foi implementado usando a linguagem MatLab, R2017b, e que contém as chamadas para outras funções do trabalho. Durante o estudo, vários testes foram realizados para julgar o método proposto; nas Seções seguintes serão mostradas todas as avaliações em cima do método.

4.1 Aquisição do *Dataset*

A base de dados utilizada '*Toxicoproteomic analysis of anthracycline-induced cardiotoxicity*' contém amostras de padrões proteômicos extraídas através da técnica SELDI-TOF, conforme a Seção 3.1. São 62 casos, dos quais 34 casos têm diagnóstico positivo (grupo cardiotoxicidade) e 28 casos tem diagnóstico negativo (grupo controle). Cada caso tinha inicialmente 373257 características diferentes, porém só havia informação até 90000 pontos, depois disso só havia ruídos. Portanto, para uma uniformização dos sinais reduziu-se para 100000 características.

4.2 Extração de características

Nesta etapa, o objetivo é obter parâmetros que discriminem de maneira significativa os dois subgrupos, de forma a identificar a presença ou ausência da cardiotoxicidade; isso garante que a classificação ocorra corretamente.

Foi utilizada a técnica de Análise de Componentes Independentes (ICA), revisada na Seção 2.5. Primeiramente, é necessário obter a matriz X , que é a junção dos casos de cardiotoxicidade e dos casos normais. Dessa forma, a matriz X resultante possuirá dimensões 62×100000 , sendo 62 a quantidade de amostras da base de dados e 100000 os níveis de intensidade (características) por cada amostra.

Depois que a matriz X é obtida, ela vai servir como entrada para o algoritmo *FastICA*, visto na Seção 2.5.5, para que possam ser obtidas as funções de base da matriz de características A de ordem 62×60 , que contém as características de cada uma das amostras. Cada linha da matriz A corresponde a uma amostra, e cada coluna

corresponde a uma característica. Assim, o classificador vai ter um parâmetro para diferenciar os pacientes enfermos e os saudáveis.

4.3 Seleção das características mais significantes

Para reduzir a dimensionalidade da matriz de características gerada na etapa anterior, foi utilizado o algoritmo de Máxima Relevância e Mínima Redundância. Foram feitos testes incrementando, de 1 em 1, o número de características selecionadas até o limite 60. É uma etapa importante antes da classificação, pois reduz o número de parâmetros que serão utilizados na entrada do classificador, melhorando os resultados. Cada vetor gerado através do Mrmr, foi testado com o classificador Máquina de Vetores de Suporte (SVM) e Análise Discriminante Linear (LDA), com a finalidade de encontrar o vetor de melhor desempenho.

4.4 Classificação

Como última etapa, serão utilizados dois classificadores: SVM e LDA, descritos nas Seções 2.8 e 2.9, respectivamente.

4.4.1 Classificação SVM

A Tabela 7 mostra os melhores resultados obtidos no processo de classificação de cardiotoxicidade pela Máquina de Vetores de Suporte (SVM), em destaque. Utilizou-se a SVM que tem núcleo baseado em RBF (*Radial-Basis Function*).

Tabela 7 - Desempenho do SVM para os vetores de características

Características	VP	FP	VN	FN	Especificidade (E)%	Sensibilidade (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	97,260	88,718
49	33	11	17	1	80	88,897	82,057
53	33	11	17	1	82,143	86,943	80,597

Fonte: ALMEIDA, 2016.

Os resultados da Tabela 7 foram obtidos através do método 10-fold-cross-validation, técnica de validação cruzada que é utilizada logo após a classificação da SVM. Os melhores resultados foram para os vetores com 14, 16, 20, 49 e 53 características, que obtiveram melhor desempenho durante o período de testes do classificador. Na tabela é notável que com apenas 20 características das 60 existentes, o método obteve 85,000% de especificidade, 97,260% de sensibilidade e 88,718% de acurácia.

Além dos resultados das métricas do vetor de 20 características, atentou-se também que das 34 amostras positivas para cardiotoxicidade, 33 foram classificadas corretamente (VP), ou seja, apenas 1 foi classificada como normal (FN). Por outro lado, dos 28 casos com diagnóstico normal, apenas em 10 casos (FP) houve erro classificatório, sendo diagnosticados com cardiotoxicidade.

4.4.2 Classificação LDA

A classificação da cardiotoxicidade utilizando a técnica LDA, se baseia na separação linear dos dados (ANEXO A). A Tabela 8 mostra os resultados obtidos no processo de classificação, para algumas das características analisadas.

Tabela 8 – Desempenho do LDA para os vetores de características

<i>Características</i>	<i>VN</i>	<i>VP</i>	<i>FP</i>	<i>FN</i>	<i>Acurácia (A)%</i>	<i>Sensibilidade (S)%</i>	<i>Especificidade (E)%</i>
25	22	27	6	7	79,032	75,862	81,818
28	25	29	3	5	87,096	83,333	90,625
33	27	31	1	3	93,548	90	96,875
45	28	33	0	1	98,387	96,551	100
48	28	34	0	0	100	100	100
49	28	34	0	0	100	100	100

Fonte: Autor.

Ao analisar a Tabela 8, foi possível conferir que de acordo com o aumento da quantidade de características, aumentam-se também as porcentagens de acurácia, sensibilidade e especificidade. Para o vetor com 48 características, o método LDA obteve exatamente 100% nas métricas de especificidade, sensibilidade e acurácia. Além disso, a partir do vetor de 48 características, foi possível concluir que todas as 62 amostras foram classificadas corretamente, sendo 34 com cardiotoxicidade (VP) e 28 normais (VN), obtendo-se 100% de acurácia, sensibilidade e especificidade.

Para efeito de comparação, observou-se que o método LDA gerou melhores resultados com esse conjunto de amostras, visto que a separação linear dos dados se adaptou melhor ao caso do que o SVM, que se baseia na técnica que maximiza a distância de classes diferentes. Entretanto, é fundamental lembrar que o método deve ser testado com outras bases de dados, a fim de analisar a sua real eficácia.

5 CONCLUSÃO

Neste trabalho, foi apresentado um método de detecção de cardiotoxicidade auxiliado por computador, utilizando sinais proteômicos que foram extraídos pela técnica SELDI-TOF. As características desses sinais foram extraídas através da Análise de Componentes Independentes (ICA), com a ajuda do algoritmo FastICA. As características mais relevantes foram selecionadas pelo algoritmo de Máxima Relevância e Mínima Redundância, e posteriormente classificadas pelas técnicas Máquina de Vetores de Suporte e Análise Discriminante Linear, discriminando as amostras em normais ou com cardiotoxicidade.

A técnica de mRMR possibilitou a redução da dimensionalidade da matriz de características sem afetar negativamente os resultados mesmo excluindo algumas características. Reduzir a dimensionalidade da matriz gerou uma diminuição do custo computacional da máquina ao classificar-se as amostras.

Sobre a relação com o diagnóstico clínico, o especialista no diagnóstico da cardiotoxicidade deve seguir os procedimentos habituais avaliando de forma subjetiva cada paciente e recorrer aos métodos padrões de diagnóstico, como os métodos de imagens. A proposta do CAD é funcionar como uma segunda leitura ou opinião, com o propósito de aumentar a sensibilidade do diagnóstico.

Considerando que a detecção precoce é extremamente importante para obtenção da cura da cardiotoxicidade, os métodos CAD que auxiliam no diagnóstico, propõem-se a facilitar a detecção da doença no seu estágio inicial. Dessa maneira, assegura-se o tratamento adequado aos pacientes de forma antecipada, tornando as chances de cura ainda maiores. Isto explica a importância de um método com alta taxa de acerto, como o LDA mostrou-se ser altamente eficaz em comparação com o SVM.

As técnicas apresentadas geraram bons resultados. Desta forma, algumas propostas poderão ser desenvolvidas em trabalhos futuros, em diferentes contextos, com outras bases de dados mais complexas e comparando com outros métodos, a fim de aumentar a confiabilidade do método apresentado.

REFERÊNCIAS

AAPO, Hyvarinen. **Independent Component Analysis (ICA) and Blind Source Separation (BSS)**, 2016. Disponível em: <<http://research.ics.aalto.fi/ica/fastica/>>. Acesso em: 03 maio 2019.

ALMEIDA, Jardiel Nunes et al. **Método Computacional para Detecção da Cardiotoxicidade Utilizando Padrões Proteômicos, Análise de Componentes Independentes e Máquina de Vetores de Suporte**. Disponível em: <<http://sbprt.org.br/sbprt2016/anais/ST06/1570281054.pdf>>. Acesso em: 02 jun. 2019.

ALMEIDA, Jardiel Nunes. **Método de Detecção da Cardiotoxicidade**. 2016. 59 f. Dissertação (Mestrado) - Curso de Física, Universidade Estadual do Maranhão, São Luís, 2016.

ALTENA, Renske et al. **Cardiovascular toxicity caused by cancer treatment: strategies for early detection**. Lancet Oncol. 2009;
AMERICAN HEART ASSOCIATION (AHA); AMERICAN STROKE ASSOCIATION. **Heart Disease and Stroke Statistics – At-a-Glance**. Circulation, 131 (4), p. 29-322, 2015.

ARAUJO, Claudyane da Silva. **DIAGNÓSTICO AUXILIADO POR COMPUTADOR APLICADO EM CÂNCER DE OVÁRIO E PRÓSTATA**. 2018. 71 f. TCC (Graduação) - Curso de Engenharia da Computação, Cct, U, São Luís, 2018.

AZEVEDO-MARQUES, Paulo Mazzoncini de. Diagnóstico auxiliado por computador na radiologia. **Radiol Bras**, São Paulo, v. 34, n. 5, p. 285-293, Out 2001.

BACAL, Fernando; NUSSBAUM, Ana Cláudia Aparecida Santos; PIRES, Philippe Vieira. Tratamento da cardiotoxicidade. **Revista da Sociedade de Cardiologia do Estado de São Paulo**, São Paulo, v. 22, n. 3, p.20-24, 2012.

BAIROCH, A.; APWEILER, R. **The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998**. Nucleic Acids Research., v. 26, 1998.

BARBOSA, Eduardo Buzolin et al. Proteômica: metodologias e aplicações no estudo de doenças humanas. **Revista da Associação Médica Brasileira**, São José do Rio Preto, v. 58, n. 3, p.366-375, 2011.

BUSHBERG, J T et al. **The essential physics of medical imaging**. 3. ed. Philadelphia: Lippincott William e Wilkins, 2012.

CARDINALE, Daniela et al. **Anthracycline-induced cardiomyopathy: clinical relevance and response to pharmacologic therapy**. J Am Coll Cardiol. 2010;

CARMO, S. K. S. et al. Control Proposal for a High Purity column based on the separation of variables by the independent component analysis method. **Brazilian Journal Of Chemical Engineering**, [s.l.], v. 34, n. 1, p.317-330, jan. 2017.

CARVALHO, Paulo Costa et al. Marcadores séricos e espectrometria de massa no diagnóstico do câncer. **O Jornal Brasileiro de Patologia e Medicina Laboratorial**. Rio de Janeiro, p. 431-436. dez. 2006.

CASTANEDA-VILLA, N.; CALDERON-RIOS, E. R.; JIMENEZ-GONZALEZ, A. On the Identification of an ICA Algorithm for Auditory Evoked Potentials Extraction: A Study on Synthetic Data. **Rev. mex. ing. bioméd**, México, v. 36, n. 2, p. 107-119, agosto 2015.

CATARINO, Francisco Manuel Inácio Ferreira. **Segmentação da íris em imagens com ruído**. 2009. 83 f. Tese (Doutorado) - Curso de Engenharia Informática, Departamento de Informática, Universidade da Beira Interior, Covilhã, 2009.

COSTA, Álvaro Teixeira da et al. Cardiotoxicidade induzida por quimioterápicos. **Revista Médica Minas Gerais**, Belo Horizonte, p.11-14, 2010.

COSTA, Daniel Duarte. **Classificação de Lesões em Mamografias por Análise de Componentes Independentes, Análise Discriminante Linear e Máquina de Vetores de Suporte**. 2008. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Maranhão, São Luís.

CRUZ, Margarida; RODRIGUES, Joana Duarte; CAMPELO, Manuel. Cardiotoxicidade na terapêutica com antraciclinas: estratégias de prevenção. **Revista Portuguesa de Cardiologia**, Porto, p.359-371, 2015.

DAMASCENO, Nielsen Castelo. Análise Comparativa entre FastICA por Maximização da Negentropia e AG-ICA usando Negentropia de Rényi. In: SIMPÓSIO DE TECNOLOGIA DA INFORMAÇÃO DA REGIÃO NOROESTE DO RIO GRANDE DO SUL (STIN 2011), 2., 2011, Santo Ângelo. **Simpósio**. Santo Ângelo: Uri, 2011. p. 1 - 6.

DERMIS, S. D. I. (2012). Online atlases. Disponível em: <<http://dermis.net/dermisroot/en/home/index.htm>>. Acesso em 08 maio 2019.

DERMQUEST. (2012). Online database. Disponível em: <<https://www.dermquest.com/>>. Acesso em 08 maio. 2019.

DING, C.; PENG, H. Minimum redundancy feature selection from microarray gene expression data. **Journal of Bioinformatics and Computational Biology**, Imperial College Press, v. 3, n. 2, p. 185–205, 2005.

DINIZ, Maria Elisa Romanelli. **Uso da técnica de espectrometria de massas com ionização por eletrospray (ESI-MS) para o estudo do mecanismo de reações orgânicas e avaliação do perfil de fragmentação de bis-hidroxiiminas aromáticas**. 2011. 93 f. Dissertação (Mestrado) - Curso de Química, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

DOI, Kunio. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. **Computerized Medical Imaging And Graphics: The**

Official Journal Of The Computerized Medical Imaging Society, Chicago, v. 31, p.198-211, jul. 2007.

DUARTE, Leonardo Tomazeli. **Um Estudo sobre Separação Cega de Fontes e Contribuições ao Caso de Misturas Não-lineares**. 2006. 140 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 2006.

FERIS, R. S., et al. (2001). Efficient Real-Time Face Tracking in Wavelet Subspace. **Proceedings of International Conference on Computer Vision, Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems**, Vancouver, BC., pages 113–118.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ci. Inf.**, Brasília, v. 35, n. 1, p. 25-30, Abr.2006.

FERREIRA JUNIOR, José Raniery; OLIVEIRA, Marcelo Costa. Unidade de processamento gráfico no diagnóstico auxiliado por computador. **Journal Of Health Informatics**, [s. l.], v. 3, n. 9, p.103-108, out. 2007.

FLETCHER, Robert H; FLETCHER, Suzanne W; BARTLETT, Jones &. **Epidemiologia clínica: Elementos Essenciais**. 4. ed. Porto Alegre: Artmed, 2006.

FRANCO, Alexandre Lerch. **Aplicação da Análise de Componentes Independentes em Estudo de Eventos em Finanças**. 2008. 154 f. Tese (Doutorado) - Curso de Administração, Ufrs, Porto Alegre, 2008.

GALDOS-RIVEROS, Alvaro Carlos et al. **PROTEÔMICA: NOVAS FRONTEIRAS NA PESQUISA CLÍNICA**. 11. ed. São Paulo: Enciclopédia Biosfera, 2010. 24 p.

GALVAO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de mineração de dados: uma revisão da literatura. **Acta paul. enferm.**, São Paulo, v. 22, n. 5, p. 686-690, Out. 2009.

GAROFALO, D. F. T.; MESSIAS, C. G.; LIESENBERG, V.; BOLFE, E. L.; FERREIRA, M. C. Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático. **Pesquisa Agropecuária Brasileira** (Online), v.50, p. 593-604, 2015.

GUERRA, Maximiliano Ribeiro; GALLO, Cláudia Vitória de Moura; MENDONÇA, Gulnar Azevedo e Silva. **Risco de câncer no Brasil: tendências e estudos epidemiológicos mais recentes**. 2004.

HAJJAR, Ludhmila Abrahão; HOFF, Paulo Marcelo Gehm; KALIL FILHO, Roberto. O impacto da doença cardiovascular no paciente com câncer: aspectos epidemiológicos e prognóstico. **Revista da Sociedade de Cardiologia do Estado de São Paulo**, São Paulo, p.2-7, 2012.

HAYKIN, S. **Neural Networks - A Comprehensive Foundation**. Prentice-Hall, New Jersey, 2nd edition, 1999

HAYKIN, Simon. **REDES NEURAIS: PRINCÍPIOS E PRÁTICA**. 2. ed. Hamilton, Ontario: Bookman, 2008. 908 p. Tradução de Paulo Martins Engel.

HERNANDEZ, A. SM.; SAN MIGUEL, R.; MARTÍN-GIL, F. J. **Importancia de las aplicaciones clínicas de la Proteômica**. *Revista del Laboratorio Clínico*, v. 3, n. 1, 2010.

HOCHSTRASSER, D. F. **Proteome in perspective**. *Clinical Chemistry and Laboratory Medicine*, v. 36, 1998.

HYVARINEN, A., KARHUNEN, J., OJA, E. **Independent Component Analysis**, Nova York, John Wiley & Sons, p. 481, 2001.

HYVARINEN and OJA. Independent component analysis: algorithms and applications. **Neural networks**, Finland, v.13, n.4-5, p.411-430, 2000.

I Diretriz Brasileira de Cardio-Oncologia da Sociedade Brasileira de Cardiologia. *Arq Bras Cardiol* 2011; 96(2 supl.1): 1-52. Disponível em: <http://publicacoes.cardiol.br/consenso/2011/diretriz_cardio_oncologia.pdf>. Acesso em: 02 jun. 2019.

I DIRETRIZ BRASILEIRA DE CARDIO-ONCOLOGIA PEDIÁTRICA DA SOCIEDADE BRASILEIRA DE CARDIOLOGIA. Rio de Janeiro: Sbc - Núcleo Interno de Publicações, v. 100, n. 1, maio 2013. Disponível em: <http://publicacoes.cardiol.br/consenso/2013/i_diretriz_cardio-oncologia_pediatica.pdf>. Acesso em: 02 set. 2019.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, p. 815. Kalil Filho R, Hajjar LA, Bacal F, Hoff PM, Diz M del P, Galas FRBG, et al. I Diretriz Brasileira de Cardio-Oncologia da Sociedade Brasileira de Cardiologia. *Arq Bras Cardiol* 2011; 96(2 supl.1): 1-52.

JONES, Lee W. et al. Early Breast Cancer Therapy and Cardiovascular Injury. **Journal Of The American College Of Cardiology**. Durham, Edmonton, p. 1435-1441. out. 2007. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0735109707022401?via%3Dihub#!>>. Acesso em: 01 maio 2019.

KHATTREE, Ravindra; NAIK, Dayanand N. **Multivariate Data Reduction and Discrimination with SAS Software**. North Carolina: Sas Institute Incorporated, 2000. 558 p.

KITANI, Edson C.; THOMAZ, Carlos E. ANÁLISE DE DISCRIMINANTES LINEARES PARA MODELAGEM E RECONSTRUÇÃO DE IMAGENS DE FACE. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 2007, Rio de Janeiro. **Anais...** Rio de

Janeiro: Congresso Brasileiro da Sociedade Brasileira de Computação, 2007. p. 962 - 971.

LINHARES, José do Nascimento et al. Método Computacional para o Diagnóstico Precoce da Granulomatose de Wegener. **Revista de Informática Teórica e Aplicada**, [S.l.], v. 23, n. 1, p.277-292, 29 maio 2016. Universidade Federal do Rio Grande do Sul.

LORENA, Ana Carolina; CARVALHO, André C. P. L. F. de. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, São Carlos, n. 2, p.43-67, 2007.

MANSUR, A. P.; FAVARATO, D. **Mortalidade por doenças cardiovasculares no Brasil e na região metropolitanade**. São Paulo: Atualização 2011. Arq Bras Cardiol, 2012.

MATHWORKS. **Minimum-redundancy maximu-relevance feature selection**. Disponível em: <<http://www.mathworks.com/matlabcentral/fileexchange/>>. Acesso em: 10 maio 2019.

MCCUE, C. **Data Mining and Predictive Analysis - Intelligence Gathering and Crime Analysis**. Elsevier, 2007.

MARINS; A. I. M.; VELLASCO, M. M. B. R. **Implementação do FREx_SVM: Máquinas de vetor suporte para classificação em múltiplas classes**. 2008. Disponível em: <http://www.pucrio.br/pibic/relatorio_resumo2008/resumos/ctc/ele/ele_aaraoim.pdf>. Acesso em: 28 junho 2016.

MEDEIROS, Alexandre Kreling; WIEHE, Mário. **CARDIOTOXICIDADE INDUZIDA POR QUIMIOTERAPIA**. 2018. Disponível em: <<http://ebooks.pucrs.br/edipucrs/acessolivre/periodicos/acta-medica/assets/edicoes/2018-2/arquivos/pdf/9.pdf>>. Acesso em: 02 jun. 2019.

MESQUITA, Claudio Tinoco. Editorial: O Surgimento da Cardio-Oncologia: The Rise of Cardio-Oncology. **International Journal Of Cardiovascular Sciences**, [s.l.], p.419-421, 2017.

Mitchell, T. (1997). **Machine Learning**. McGraw Hill.

MONARD, Maria Carolina, BARANAUSKAS, José Augusto. **Conceitos Sobre Aprendizado de Máquina**. Sistemas Inteligentes Fundamentos e Aplicações. 1 ed. Barueri-SP: Manole Ltda, 2003. p. 89--114. ISBN 85-204-168.

NORONHA, Daniel Holanda; FERNANDES, Marcelo Augusto Costa. Implementação em FPGA de Máquina de Vetores de Suporte (SVM) para Classificação e Regressão. In: BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEM, 5., 2016, Recife. **Conference**. Recife: BRACIS, 2016. p. 1 - 12.

OLIVEIRA, Fernando Soares Sérvulo de. **CLASSIFICAÇÃO DE TECIDOS DA MAMA EM MASSA E NÃO-MASSA USANDO ÍNDICE DE DIVERSIDADE TAXONÔMICO E MÁQUINA DE VETORES DE SUPORTE**. 2013. 70 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Centro de Ciências Exatas e Tecnologia, Universidade Federal do Maranhão, São Luís, 2013.

OLIVEIRA, Henrique Evangelista de. **Um Método para Separação Cega de Fontes Esparsas Utilizando Evolução Diferencial**. 2015. 63 f. Dissertação (Mestrado) - Curso de Ciências Aplicadas, Universidade Estadual de Campinas, Limeira, 2015.

PENG, Hanchuan; LONG, Fuhui and Ding, C. **Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy**. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(8):1226–1238, Aug. 2005.

PEREIRA, Mauricio Gomes. **Epidemiologia Teoria e Prática**. Rio de Janeiro: Guanabara, 2008.

PETRICOIN, Emanuel F; A LIOTTA, Lance. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. **Current Opinion In Biotechnology**, [s. l.], v. 15, n. 1, p.24-30, fev. 2004. Disponível em: <<https://home.ccr.cancer.gov/ncifdaproteomics/pdf/CurrentOpinionBiotech.pdf>>. Acesso em: 20 jun. 2019.

PETRICOIN, Emanuel F. et al. Toxicoproteomics: Serum Proteomic Pattern Diagnostics for Early Detection of Drug Induced Cardiac Toxicities and Cardioprotection. **Toxicologic Pathology**, Clinical Proteomics Program Website, v. 32, n. 1, p.122-130, jan. 2004.

PINHEIRO, Tadeu Batista et al. Insuficiência Cardíaca com Fração de Ejeção Normal em Paciente Adulto Jovem. **Bras Cardio**, Niterói, v. 26, n. 1, p.54-57, nov. 2012.

PINTO, Carolina Maria et al. Quimioterapia e cardiotoxicidade. **Rev. Soc. Cardiol. Estado de São Paulo**, v.4, n.27, p.266-273, dez. 2017.

PIVETA, Rafael B.; PAIVA, Marcelo G.; ANDRADE, José L. Diagnóstico por imagem da cardiotoxicidade. **Revista da Sociedade de Cardiologia do Estado de São Paulo**, São Paulo, v. 22, n. 3, p.14-19, set. 2012.

PRATI, Ronaldo Cristiano. **Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos**. 2006. 167 f. Tese (Doutorado) - Curso de Ciências da Computação, Icmc, Usp, São Carlos, 2006.

PROGRAM, C. P. **Toxicoproteomic analysis of anthracycline-induced cardiotoxicity**. 2004. Disponível em: <<https://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp>>. Acesso em: 03 Mar. 2019.

REIS, Luciano Marques Brito; CAMPOS, Lúcio Flávio de A. Detecção de Lesões de Câncer de Pele Utilizando Análise de Componentes Independentes e Análise Discriminante Linear. **Revista de Ciências da Computação**, São Luís, v. 12, p.1-15, 2017.

RIBEIRO, Mario Luiz et al. Detecção Precoce e Monitoramento da Disfunção Ventricular Esquerda Secundária à Quimioterapia do Câncer pelos Métodos de Imagem. **Arquivos Brasileiros de Cardiologia**, [s.l.], v. 3, n. 112, p.309-316, dez. 2018.

RIBEIRO, Silvia Moulin. Prevenção da cardiotoxicidade. **Revista da Sociedade de Cardiologia do Estado de São Paulo**, São Paulo, v. 22, n. 3, p.25-28, 2012.

RICHARDS, C. J.; JE, Y.; SCHUTZ, F. A. B.; HENG, D. Y. C.; DALLABRIDA, S. M.; MOSLEHI, J. J.; CHOUEIRI, T. K. - Incidence and Risk of Congestive Heart Failure in Patients With Renal and Nonrenal Cell Carcinoma Treated With Sunitinib. **J. Clin. Oncol.**, 29 (2011) 3450– 3456.

RODRIGUES, T. A. O. et al. Predição de função de proteínas através da extração de características físico-químicas. **Revista de Informática Teórica e Aplicada**, v. 22, n. 1, p. 29–51, 2015.

SÁ, Michel Pompeu Barros de Oliveira et al. Cardiotoxicidade e quimioterapia. **Revista da Sociedade Brasileira de Clínica Médica**, São Paulo, v. 7, n. 5, p.326-330, out. 2009.

SANTOS, Marília Harumi Higuchi dos; HAJJAR, Ludhmila Abrahão. Pesquisa em cardio-oncologia - o futuro. **Revista da Sociedade de Cardiologia do Estado de São Paulo**, São Paulo, v. 22, n. 3, p.36-41, set. 2012.

SCARLATELLI, Ariane Vieira; LOPES JUNIOR, Gilberto de Lima; PONTES, Lucíola. **Cardiotoxicidade no tratamento de adultos com câncer - como abordar?** 2015.

SILVA, Alan Paulo Oliveira da. **Uma Implementação da Análise de Componentes Independentes em Plataforma de Hardware Reconfigurável**. 2010. 85 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Ufrn, Natal, 2010.

SOUSA, João Paulo Pereira. **Cardiotoxicidade Induzida por Fármacos Anti-Cancerosos**. 2014. 24 f. Dissertação (Mestrado) - Curso de Farmácia, Faculdade de Farmácia, Universidade de Coimbra, Coimbra, 2014.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., et al, The mammographic images analysis society digital mammogram database. **Exerpta Medical**, v. 1069, p. 375-378, 1994.

TAKAHASHI, Adriana. **Máquina De Vetores: Suporte Intervalar**. 2012. 72 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2012.

TAKAKURA, Aline Miki et al. USO DO APRENDIZADO DE MÁQUINA NO DIAGNÓSTICO MÉDICO DE PATOLOGIAS. **Colloquium Exactarum**, Presidente Prudente, v. 10, n. 1, p.78-90, out. 2017.

TARTARI, Rafaela Festugatto; BUSNELLO, Fernanda Michielin; NUNES, Claudia Helena Abreu. **Perfil Nutricional de Pacientes em Tratamento Quimioterápico em um Ambulatório Especializado em Quimioterapia**. 2009. Disponível em: <http://www1.inca.gov.br/rbc/n_56/v01/pdf/07_artigo_perfil_nutricional_paciente_oncologico.pdf>. Acesso em: 02 jun. 2019.

THARWAT, Alaa. Independent component analysis: An introduction. **Applied Computing And Informatics**, Frankfurt, p.1-15, ago. 2018.

THOMAZ, Carlos Eduardo; KITANI, Edson Caoru; GILLIES, Duncan Fyfe. A maximum uncertainty LDA-based approach for limited sample size problems - with application to face recognition. **Journal Of The Brazilian Computer Society**, Campinas, v. 12, n. 2, p.7-18, set. 2006.

URZEDO, Ana Paula Fonseca Maia de. **Degradação de Substâncias de Relevância Ambiental por Processos Oxidativos e Redutivos com Monitoramento por Espectrometria de Massas com Ionização Electrospray**. 2008. 158 f. Tese (Doutorado) - Curso de Ciências Químicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

VOLKOVA, Maria; RUSSELL, Raymond. Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. **Current Cardiology Reviews**, [s.l.], v. 7, n. 4, p.214-220, 14 fev. 2012. Bentham Science Publishers Ltd.

WILSON, K; WALKER, J. **Principles and Techniques of Biochemistry and Molecular Biology**. 2010. 7ed. Pag. 352

YEH, T.H, Edward et al. **Cardiovascular Complications of Cancer Therapy - Diagnosis, Pathogenesis, and Management**. *Circulation*. 2004; 109: 3122-31.

YIGITBASLI, Turkan et al. Perfil de SELDI-TOF-MS do Fenótipo Metastático em Subtipos Histopatológicos do Câncer de Mama, **Current Proteomics**, v.15, n.3, p. 214 – 220, 2018.

ANEXO A

```
% Universidade Estadual do Maranhão

% Graduação em Engenharia da Computação

% Aluna: Paola Lauande Oliveira Mendes

% Orientador: Prof. Dr. Lúcio Flávio de A. Campos

% Classificação LDA

clear all

clc

FINAL=[];

Acuracia = [];

Sensibilidade = [];

Especificidade = [];

% carrega pasta organizada com as matrizes 'A' e 'At'

load organizada

    for k=1:60

        VP = 0;

        VN = 0;

        FP = 0;

        FN = 0;

Af=A(:,1:k);

quadisc = fitcdiscr(Af,At);
```

```
cvmodel = crossval(quadisc,'kfold',10);

obj = fitcdiscr(Af,At);

R = confusionmat(obj.Y,resubPredict(obj));

VP=R(1,1);VN=R(2,2);FP=R(1,2);FN=R(2,1);

    Acuracia = 100*((VP+VN)/(VP+VN+FP+FN));

    Sensibilidade = 100*((VP)/(VP+FN));

    Especificidade = 100*((VN)/(VN+FP));

    x=[k VP VN FP FN Acuracia Sensibilidade Especificidade];

% matriz final com todos resultados do primeiro ao último teste

FINAL=[FINAL;x];

    end
```