UFRRJ
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

DISSERTAÇÃO

MODELOS DE INTELIGÊNCIA COMPUTACIONAL APLICADO NA PREDIÇÃO DE CÂNCER DE MAMA

MARCELLE ROSA RIBEIRO LEAL

2018

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO

INSTITUTO DE CIÊNCIAS EXATAS

PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

MODELOS DE INTELIGÊNCIA COMPUTACIONAL APLICADO NA PREDIÇÃO DE CÂNCER DE MAMA

MARCELLE ROSA RIBEIRO LEAL

Sob a Orientação do Professor

Dr. Robson Mariano da Silva.

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Modelagem Matemática e Computacional, no Curso de Pós-graduação em Modelagem Matemática e Computacional, Área de concentração em Inteligência Computacional e Otimização.

Seropédica, RJ

Maio de 2018

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE CIÊNCIAS EXTAS
CURSO DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

MARCELLE ROSA RIBEIRO LEAL

Dissertação submetida como requisito parcial para obtenção do grau de Mestre em Ciências, no Curso de Pós-graduação em Modelagem Matemática e Computacional, área de concentração em Inteligência Computacional e Otimização.

Dissertação aprovada em ____/____/____

_____
Prof. Dr. Robson Mariano da Silva.
(Orientador)

_____
Prof. Dr. Rafael Bernado Texeira.

_____
Prof. Dr. Marcos Azevedo Benac.

_____
Prof. Dr. Reinaldo Belline.

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

d788m

de Oliveira Rosa Ribeiro, Marcelle, 1988-
Modelo de Inteligência Computacional Aplicado na Predição no Câncer de Mama / Marcelle de Oliveira Rosa Ribeiro. - 2018.
61 f.: il.

Orientador: Robson Mariano da Silva.
Dissertação(Mestrado). -- Universidade Federal Rural do Rio de Janeiro, Mestrado em Matemática Computacional, 2018.

1. Câncer de mama. 2. RNAs. 3. SVM. I. Mariano da Silva, Robson, 1963-, orient. II Universidade Federal Rural do Rio de Janeiro. Mestrado em Matemática Computacional III. Título.

Aos meus pais, Josué e Leide, que me deram a grandeza da vida e sempre me ensinaram o valor do estudo e do trabalho.

AGRADECIMENTOS

Este percurso de formação não teria sido possível se eu não tivesse encontrado ajuda e apoio para concretização deste sonho. Por isso agradeço:

Ao **Criador**, pela vida, pela chama de esperança que me fez expandir o espírito e ampliar os conhecimentos;

Ao meu querido orientador, **Professor Robson**, que me incentivou e orientou todo o processo de construção deste trabalho;

Ao meu amado esposo **Bruno** que sonhou comigo essa grande conquista acadêmica;

A minha inspiradora filha, **Melissa**, que com sua chegada, me fez ser uma pessoa mais produtiva;

A **todos os professores** do curso de Pós-Graduação em Modelagem Matemática e Computacional da Universidade Federal Rural do Rio de Janeiro, que me conduziram a aprender a aprender.

RESUMO

LEAL, Marcelle Rosa Ribeiro. **Modelos de Inteligência Computacional Aplicado na Predição de Câncer de Mama**.2018. 66p. Dissertação (Mestrado em Ciência em Modelagem Matemática e Computacional). Instituto de Ciências Exatas, UniversidadeFederal Rural do Rio de Janeiro, Seropédica, RJ, 2018.

O câncer de mama é a segunda neoplasia mais frequente no mundo. Segundo dados do Instituto Nacional de Câncer (INCA), no ano de 2014 foram diagnosticados 59.700 novos casos no Brasil, número este que corresponde a um aumento de 22% em relação ao ano de 2013. Sendo responsável por aproximadamente 39% dos óbitos das mulheres portadores de câncer. Para um diagnóstico preciso, exige-se muita experiência e, principalmente, que a classificação do estadiamento clínico do tumor (estágio do câncer) esteja correta. Desta forma, torna-se necessário o desenvolvimento de sistemas integrados que combinados com a experiência dos profissionais da área, possibilite realizar o diagnóstico preciso na detecção do câncer de mama. O objetivo do presente trabalho é aplicar as técnicas RNAs e SVM de sorte a auxiliar na interpretação diagnóstica das microcalcificações detectadas em mamografia de rastreamento. O conjunto utilizado nesse estudo consiste de 569 dados, proveniente de pacientes com suspeita de câncer de mama obtidos junto ao Instituto de Radiologia da Universidade Erlangen-Nuremberg, no período de 2003 a 2006. O banco de dados possui informações clínicas sobre raio, textura, perímetro, área, suavidade, compacidade, concavidade, côncavo, simetria e dimensão fractal. Os dados foram divididos em dois grupos: o conjunto de treinamento composto por 75% das amostras de exames mamográficos e o conjunto de teste independente, com 25% das amostras restantes. As técnicas desenvolvidas foram implementadas utilizando-se o software R. De acordo com a análise dos resultados foi possível evidenciar o desempenho promissor da SVM, que obteve na sua melhor simulção uma acurácia acima de 98%, no que tange aos valores de Falsos Negativos o melhor valor obtido foi 1,96%. Contudo, o modelo utilizando as Redes Neurais MLP apresentou na sua melhor simulação uma acurácia acima de 96% e no que tange aos valores de Falsos Negativos o melhor valor obtido também foi de 2%, sendo assim sua utilização relevante. Houve diferença estatística significante a nível de 95% (p-valor <0,05) no desempenho do modelo das Redes SVM e Rede Neural MLP na métrica acurácia. Indicando um melhor desempenho da Rede SVM. Não houve diferença estatística significativa entre os resultados referentes a determinação dos valores Falso Negativo entre as Redes.

Palavras-chave: câncer de mama, RNAs, SVM.

ABSTRACT

LEAL, Marcelle Rosa Ribeiro. **Models of Applied Computational Intelligence in Breast Cancer Prediction.**2018. 66p. Dissertation (Master in Science in Mathematical and Computational Modeling). Institute of Exact Sciences, Federal Rural University of Rio de Janeiro, Seropédica, RJ, 2018.

Breast cancer is the second most frequent neoplasm in the world. According to data from the National Cancer Institute (INCA), 52,680 new cases were diagnosed in Brazil in 2014, an increase of 22% compared to the year 2013. It accounts for approximately 39% of women's deaths cancer patients. For an accurate diagnosis, a lot of experience is required, and especially that the classification of the clinical staging of the tumor (stage of the cancer) is correct. Thus, it is necessary to develop integrated systems that, combined with the experience of the professionals of the area, make it possible to carry out the precise diagnosis in the detection of breast cancer. The objective of the present study is to apply the RNA and SVM techniques to assist in the diagnostic interpretation of microcalcifications detected in screening mammography. The set used in this study consists of 569 data, from patients with suspected breast cancer obtained from the Institute of Radiology of Erlangen-Nuremberg University from 2003 to 2006. The database has clinical information about lightning, texture, perimeter, area, smoothness, compactness, concavity, concave, symmetry and fractal dimension. The data were divided into two groups: the training set consisting of 75% of the mammographic examination samples and the independent test set, with 25% of the remaining samples. The techniques developed were implemented using software R. According to the analysis of the results it was possible to evidence the promising performance of the SVM, which obtained in its best simulation an accuracy above 98%, in relation to the values of False Negatives The best value obtained was 1.96%. However, the model using the MLP Neural Networks presented in its best simulation an accuracy of over 96% and in relation to the values of False Negatives, the best value obtained was also 2%, and therefore its relevant use. There was a statistically significant difference at the level of 95% (p-value <0.05) in the performance of the SVM Networks and MLP Neural Network model in the metric accuracy. Indicating better performance of the SVM Network. There was no statistically significant difference between the results regarding the determination of the false negative values among the Networks.

Key words: breast cancer, RNAs, SVM.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Em anos recentes, cresce no Brasil a mobilização pela detecção precoce do câncer de mama, visto que é o tipo de câncer que aparece com maior frequência entre as brasileiras (INCA 2013). A cada 8 brasileiras uma é diagnosticada com essa neoplasia (GODA, 2017), nas quais a maior incidência ocorre em mulheres de 35 a 50 anos. No entanto, é preciso observar que a incidência de câncer de mama não está restrita a essa faixa etária.

Estudos atuais estimam que cerca de 30% dos casos de câncer de mama poderiam ser evitados por meio da alimentação saudável, controle do peso corporal e atividade física regular. O autoexame das mamas fora o método propagado nos anos 80 para a detecção precoce do câncer de mama (Artigo de Opinião publicado no Correio Braziliense em 31/10/2016, p.13.), entretanto, atualmente a recomendação do autoexame fora substituída pela noção de *Breast Awareness*, que pode ser traduzida como "estar atenta às próprias mamas".

Segundo a Organização Mundial de Saúde, uma das estratégias para a detecção precoce do câncer de mama, além do *Breast Awareness*, é o exame mamografia. A mamografia é uma forma particular de radiografia capaz de registrar imagens da mama para assim diagnosticar a presença ou ausência de estruturas que possam indicar a doença (GODA, 2017).

No Brasil, o direito à mamografia vem sendo tratado como um componente do direito à saúde e cresce o movimento pela ampliação da cobertura e da faixa etária da população-alvo do rastreamento mamográfico. Mas estudos têm revelado que apenas 20% a 30% de nódulos mamográficos identificados pelo especialista, considerados suspeitos e submetidos a exame de biópsia, são malignos, isto significa que, a interpretação médica da mamografia leva a cerca 70% de biópsias desnecessárias com resultados benignos (HERMANN et al. 1997; 1988; JACOBSON; EDEIKEN, 1990; ARAUJO, 2000), o que leva a um custo desnecessário ao sistema de saúde além de causar danos emocionais ao paciente.

Vários desenvolvimentos tecnológicos estão surgindo de forma a tentar resolver esta problemática, desta forma tem surgido os sistemas CAD (Diagnóstico Auxiliado por Computador em inglês, *Computer-Aided-Diagnosis*), esses utilizam-se de técnicas provenientes da área de conhecimento da inteligência artificial, que inclui métodos para seleção de atributos e reconhecimento de padrões. (Castleman KN. Digital image processing. New Jersey: Prentice-Hall, 1996.)

Em geral, os sistemas CAD atuam na imagem da região de interesse ROI (em inglês, Regions of Interest) pré-identificada pela mamografia utilizando técnicas de processamento de imagens para determinar e extrair determinadas características do nódulo (as ROIs), e assim, são aplicadas técnicas de Aprendizado de Máquina (AM), que as classificam em benignas ou malignas (GODA, 2017).

Desse modo, as técnicas AM, com seu alto poder computacional, se tornam exemplo na área da inteligência artificial que vem sendo utilizada para a classificação das características extraídas. Sendo assim, arquiteturas de RNA como Rede Perceptron de Múltiplas Camadas (MLP em inglês Multi-Layer Perceptron) (HALKIOTIS; MANTAS; BOTSIS, 2002; YU; 2 GUAN,2000; ANDRE; RANGAYYAN, 2003), Máquinas de Vetores de Suporte (SVMs em inglês Support Vector Machines) são aplicadas na classificação de nódulos.

1.1 TRABALHOS RELACIONADOS

Na literatura pesquisada é possível encontrar uma série de trabalhos com propostas de modelos ou técnicas computacionais para apoiar o diagnóstico de câncer de mama. Em SARITAS (2012), encontramos a aplicação de rede neural visando determinar a malignidade de achados mamográficos, tendo como informação o parâmetro BI-RADS. Este estudo foi realizado em 800 pacientes diagnosticado através de biópisia. A avaliação do modelo proposto foi realizada utilizando matriz de Confusão e análise da curva ROC. A taxa de previsão da doença obtida no estudo foi de 90,5%, indicando que o modelo proposto em rede neural é confiável podendo ser de grande ajuda para os profissionais da área da saúde. Em HUANG *et all* (2018), encontramos uma revisão da aplicação das Máquinas de Vetor de Suportes (SVMs) na classificação genômica do câncer e sua perspectiva futura. No estudo realizado por NASCIMENTO *et all* (2016) encontramos uma avaliação entre as técnicas de SVMs e Rede Neural Artificial (RNA) com diferentes combinações de núcleos e critérios de parada visando classificar nódulos de câncer de mama. Os melhores resultados obtidos em termos da precisão e área sob a curva ROC foram 96,98% e 0,980, respectivamente. Quando comparadas as técnicas propostas o modelo estruturado em SVMs obteve melhor desempenho. Já em (OLIVEIRA, FACEROLI, FERNANDES, SANTOS, 2014) encontramos um estudo a respeito da classificação de nódulos a partir da análise de mamografias digitais, sendo assim desenvolvida uma rede neural artificial para reconhecimento de padrões em descritores de imagens mamográficas. Observou-se resultados superiores a 90% de acerto na classificação dos nódulos com as configurações de rede testadas e 96% de acerto na classificação dos nódulos ao se adicionar os descritores de forma e localização ao conjunto de treinamento.

Foram utilizadas três técnicas de AM em (ESTEVES, LORENA, NASCIMENTO, 2011) na tarefa de diagnosticar câncer a partir de imagens mamográficas da base de dados Digital Database for Screening Mammography (DDSM). Com o intuito de avaliar os classificadores para o uso futuro em um sistema CAD para o diagnóstico de câncer de mama as técnicas empregadas foram as Redes Neurais Artificiais (RNAs), as Máquinas de Vetores de Suporte (Support Vector Machines - SVMs) e as Random Forests (RFs), sendo as SVMs, apresentando o melhor desempenho sobre dados.

No (SAMPAIO, 2009) tem-se uma metodologia CAD para ajudar o especialista na tarefa de detecção de massas em imagens mamográficas. Esta metodologia utiliza Redes Neurais Celulares para segmentar as regiões de interesse, em seguida extrai características de geometria dessas regiões e descreve sua textura através da função K de Ripley, índice de Moran e coeficiente de Geary, após isso, são treinadas Máquina de Vetores de Suporte (MVS) para classificar as regiões candidatas em um das classes, massas ou não-massa, com sensibilidade de 80,00%, especificidade de 85,68%, acurácia de 84,62%, fora obtido uma taxa de 0,84 falsos positivos por imagem e 0,20 falsos negativos por imagem pertencentes à base de imagens DDSM e uma área sob da curva ROC de 0,870.

É apresentada em (ROCHA, JÚNIOR, SILVA, PAIVA, 2011) uma revisão bibliográfica de trabalhos voltados para detecção e diagnóstico de massas. Desta forma, destaca a utilização de sistemas CAD e CADx como contribuição para o aumento das chances de uma detecção e diagnósticos corretos, auxiliando os especialistas na tomada de decisões em um tratamento do câncer de mama.

A geometria fractal (dimensão fractal) associada à excentricidade de elipse e o índice de compacidade, são utilizadas em (PADILHA, ANTONIAZZI, HAAS E BRONDANI, 2017) para realizar uma análise quantitativa na diferenciação de tumores de mama. Sendo assim constatada uma fundamental importância desta abordagem para explorar os tumores de mama nas mamografias investigadas, classificando-os em benignos ou malignos.

Os trabalhos relacionados acima indicam que são promissoras pesquisas que abordam novas técnicas de detecção do Câncer de Mama. Um importante fato é o uso de banco de imagens públicas, como o DDSM e o MIAS, onde se destacam pela quantidade de trabalhos onde foram utilizados. Entretanto, o banco do Instituto de Radiologia da Universidade Erlangen-Nuremberg, possui uma grande quantidade de dados, de boa qualidade, além de informações extras referentes ao exame. Devido a isto, este banco foi utilizado neste trabalho.

Nota-se a relevância das técnicas de Aprendizado de Máquina (AM) para segmentação e extração de características, especialmente na análise forma e textura, assim como a utilização dos métodos: Redes Neurais (RN), Máquinas de Vetores de Suporte (SVMs) e Regressão Linear (RL) como classificadores.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Este trabalho tem como principal contribuição aplicar metodologia baseadas em Aprendizado de Máquina, como técnicas das Redes Neurais Artificiais e Máquina de Vetores de Suporte de sorte a auxiliar o especialista na classificação de microcalcificações mamárias.

1.2.2 Objetivos Específicos

A. Verificar a diferença estatística entre os resultados obtidos pelos modelos propostos;

B. Realizar um estudo das variáveis envolvidas no processo;

C. Verificar o impacto das variáveis no desempenho da rede.

1.3 ESTRUTURA DO TRABALHO

No capítulo 2 encontra-se o Referencial Teórico, capítulo que apresenta os conceitos fundamentais sobre o Câncer, Câncer de Mama, Redes Nuerais Artificiais e Máquinas de Vetor de Suporte, com o objetivo de um embasamento teórico para melhor entendimento da proposta deste estudo.

No capítulo 3 são apresentados Materiais e Métodos, isto é, o conjunto de dados utilizados, as formas de como os dados estão distribuídos, a Modelagem Computacional do método proposto.

Os resultados obtidos e a comparação entre os modelos computacionais propostos são apresentados no capítulo 4. Ao final (capítulo 5), apresentamos recomendações finais e algumas propostas de trabalhos futuros.

2 REFERENCIAL TEÓRICO

2.1 CÂNCER

Com mais de 14 milhões de novos casos por ano, o câncer vem se tornando uma das doenças mais incidentes do mundo (OMS, 2014). As causas e tipos de malignidades humanas variam em regiões geográficas e tipos populacionais distintos, mas na maioria dos países, dificilmente encontramos uma família sem uma vítima de câncer.

O documento *World Cancer Report* (OMS, 2014), que, entre outros, descreve a frequência, tendências de incidência e mortalidade e causas conhecidas do câncer humano em diversos países – afirma que a razão de câncer pode aumentar em 50%, atingindo 20 milhões de novos casos, até o ano de 2025.

No ano de 2012, tumores malignos foram responsáveis por 15% das quase 56 milhões de mortes no mundo (IARC, 2014). As malignidades mais fatais diferem das três formas mais prevalentes (Tabela 1), com o câncer de pulmão responsável por 19,4% de todas as mortes por câncer, estômago 8,8% e fígado, 9,1%

No Brasil (INCA, 2018), as estimativas para o ano de 2018 apontam para mais de 600 mil novos casos de câncer. A incidência entre homens e mulheres, segundo a localização primária, pode ser vista na tabela 2.

O câncer é capaz de se espalhar pelo corpo por dois mecanismos: invasão e metástase. Invasão refere-se à migração direta e penetração por células cancerosas nos tecidos circunvizinhos, enquanto a metastase refere-se à habilidade das células cancerosas de penetrar nos vasos linfáticos e sanguíneos, circulante por essas vias, e então invadirem tecidos em outas partes do corpo. Metástases são a maior causa de morte por doenças malignas (WITTEKIND e NEID, 2005). Dependendo da presença ou ausência da capacidade de se espalhar por invasão ou metastase, os tumores são classificados como benignos ou malignos. Tumores benignos são tumores que não se espalham por invasão ou metastase; portanto, crescem apenas localmente, raramente constituindo um risco de vida. Tumores malignos são tumores capazes de se espalharem por invasão ou metastase (INCA, 2015). Por definição o termo câncer é apenas aplicado a tumores malignos, ou malignidades.

Tabela 1 – Prevalência mundial aproximada do número de casos novos de câncer segundo localização primária.

Localização Primária Neoplasia Maligna	**Estimativa Novos Casos/Ano**
Pulmão	1,8 milhão
Mama	1,7 milhão
Cólon e Reto	1,3 milhão
Estômago	950 mil
Fígado	780 mil
Colo de Últero	470 mil
Próstata e Testículo	1,2 milhão
Pancreático	216 mil
Linfoma Maligno não-Hodgkin	385 mil
Renal	340 mil
Subtotal	9,1 milhões
Outras Localizações	5,0 milhões
Todas as Neoplasias	> 14 milhões

Fonte: (IARC, 2014)

Tabela 2 – Estimativa nacional para o ano de 2018 de número de casos novos por câncer em homens e mulheres, segundo localização primária.

Localização Primária	**Estimativa dos Casos Novos**		
Neoplasia Maligna	**Masculino**	**Feminino**	**Total**
Mama Feminina	**	59700	59700
Laringe e Pulmão	25130	13810	38940
Estômago	13540	7750	21290
Colo do Útero	**	16370	16370
Próstata	68220	**	68220
Cólon e Reto	17380	18980	36360
Esôfago	8240	2550	10790
Leucemias	5940	4860	10800
Cavidade Oral	11200	3500	14700
Pele Melanoma	2920	3340	6260
Outras Localizações	74450	76540	150990
Subtotal	227020	207400	434420
Pele não Melanoma	85170	80410	165580
Todas as Neoplasias	312190	287810	600000

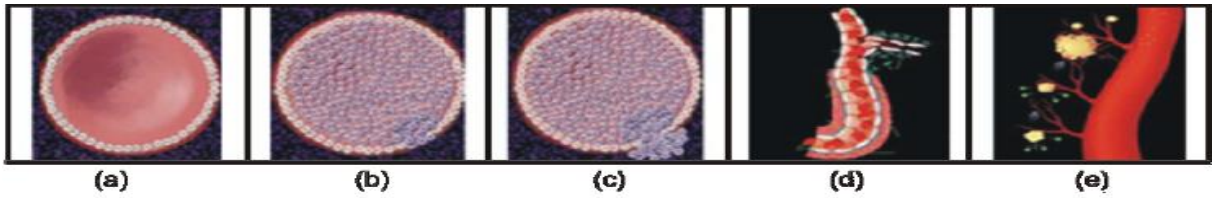
Fonte: (INCA, 2018).

2.2 CÂNCER DE MAMA

A estimativa mundial, realizada em 2012, pelo projeto Globocan/Iarc, é que o segundo tipo de câncer mais incidente no mundo foi o de mama com 1,7milhão de casos. Este câncer consiste em um crescimento descontrolado de células da mama que adquiriram características anormais (células dos lóbulos, produtores do leite, ou dos ductos, por onde é drenado o leite), anormalidades estas causadas por uma ou mais mutações no material genético de uma célula destas estruturas. Existem mutações que fazem com que uma célula apenas se divide exageradamente, mas não tenha a capacidade de invadir outros tecidos. Isto leva aos chamados tumores benignos ou não cancerosos.

A OMS em 2012 publicou a Classificação para Tumores de Mama – 4ª edição, na qual reconhece mais de 20 subtipos diferentes da doença. A maioria dos tumores de mama origina-se no epitélio ductal e são conhecidos como carcinoma ductal invasivo. Entretanto, como o câncer de mama se caracteriza por ser um grupo heterogêneo de doença, existem ainda outros subtipos de carcinomas que podem ser diagnosticados, como o lobular, o tubular, o mucinoso, o medular, o micropapilar e o papilar. (MISTÉRIO DA SAUDE, INCA, 2015)

Figura 1 - Etapas do desenvolvimento do câncer. (a) Ducto normal; (b) Carcinoma *in situ*; (c) Carcinoma ductal invasor; (d) Invasão de tecidos conectivos; (e) Embolização tumoral no vaso sanguíneo.

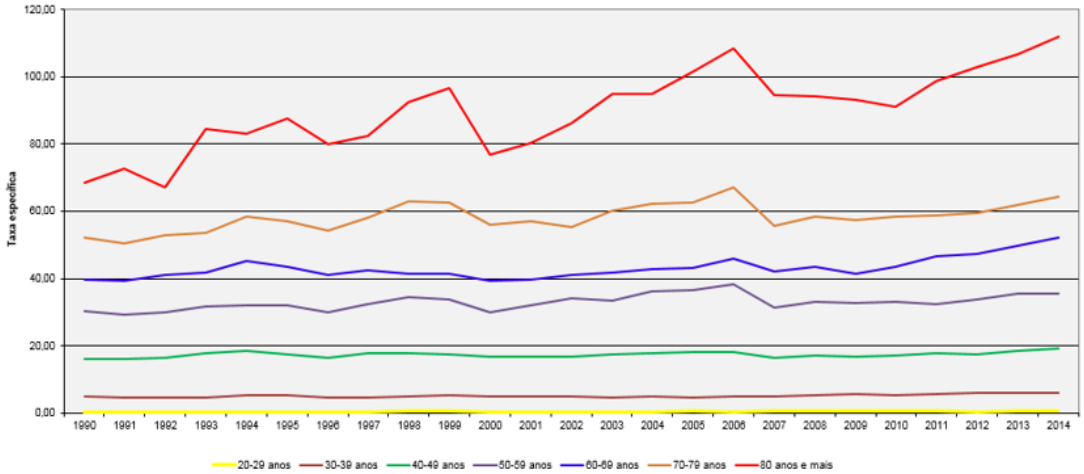


Fonte: (MAMAINFO, 2009).

O câncer de mama é um tipo de câncer considerado multifatorial, pois há muitos agentes potenciais para o desenvolvimento desse câncer, a saber: fatores biológico-endócrinos, vida reprodutiva, comportamento e estilo de vida, envelhecimento, fatores relacionados à vida reprodutiva da mulher, história familiar de câncer de mama, alta densidade do tecido mamário (razão entre o tecido glandular e o tecido adiposo da mama), consumo de álcool, excesso de peso, sedentarismo e exposição à radiação ionizante.

A incidência do câncer de mama tende a crescer progressivamente a partir dos 40 anos (ADAMI, HUNTER e TRICHOPOULOS, 2008) e a mortalidade também aumenta progressivamente com a idade. No Brasil, conforme os dados do gráfico a seguir, na população feminina abaixo de 40 anos, ocorrem menos de 10 óbitos por câncer de mama a cada 100 mil mulheres, enquanto na faixa etária a partir de 60 anos o risco é 20 vezes maior. (BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Informática do SUS (Datasus). Sistema de Informações sobre Mortalidade - SIM. http://www2.datasus.gov.br/DATASUS/index.php?area=0205 Acesso em: 10/11/2017)

Gráfico 1 - Taxas de mortalidade por câncer de mama feminina, específicas por faixas etárias, por 100.000 mulheres. Brasil, 1990 – 2014**.**



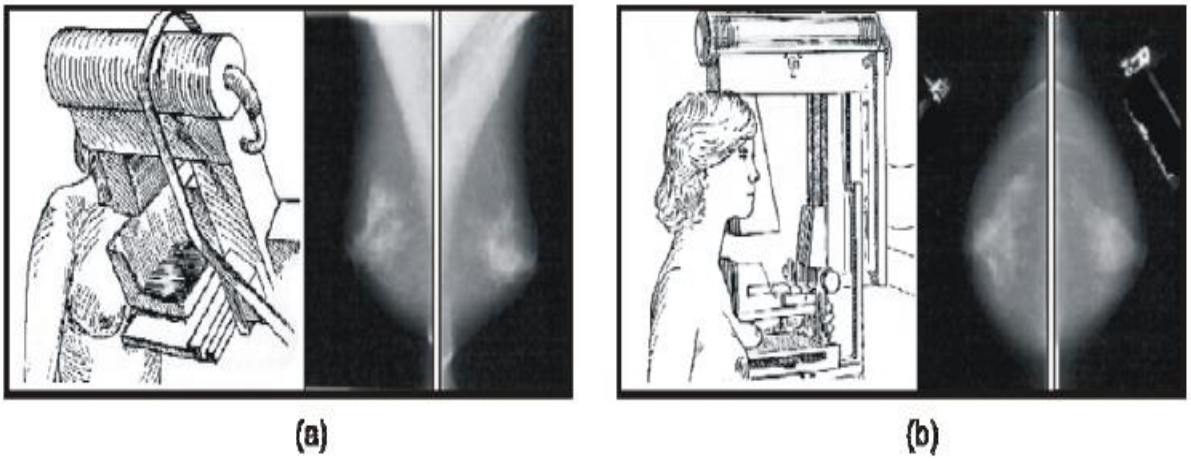
Fonte: Sistema de Informação de Mortalidade/DATASUS.

A prevenção primária dessa neoplasia tem como medidas uma alimentação saudável, prática de atividade física regular e manutenção do peso ideal, podendo evitar cerca de 30% dos casos de câncer de mama (MISTÉRIO DA SAUDE, INCA, 2015), entretanto, como a idade continua sendo um dos mais importantes fatores de risco, a mamografia bienal para mulheres entre 50 a 69 anos é a estratégia recomendada pelo Ministério da Saúde no Brasil.

2.2.1 Mamografia

A mamografia é uma radiografia das mamas que tem como objetivo produzir imagens de alta resolução das estruturas internas da mama, a fim de permitir a detecção de anomalias. Esse exame é feito por equipamento de raios x chamado mamógrafo e normalmente é bilateral, isto é, é feita uma radiografia de cada mama. Além disso, em um exame de mamografia, duas projeções de cada mama são indispensáveis: uma visão médio-lateral oblíqua (MLO) e uma crânio-caudal (CC) (SAMPAIO,2009), conforme a Figura 2.

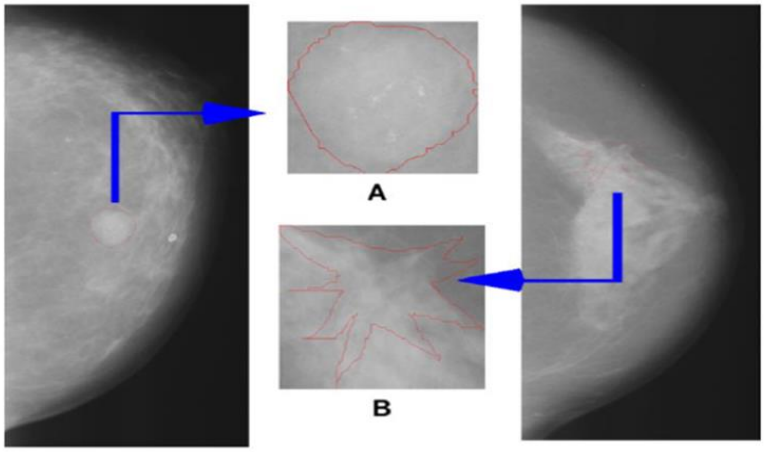
Figura 2 - Exames mamográficos. (a) Imagem MLO; (b) Imagem CC.



Fonte: adaptado de (ANGELO, 2007).

Conforme (WANG e KARAYIANNIS, 1998) há uma grande dificuldade em se trabalhar com imagens mamográficas, visto que as mesmas apresentam baixo contraste pelo fato de que as diferenças entre tecidos doentes e normais são muito pequenas. Em (SAMPAIO, 2009) revela que a mamografia é um exame de alta sensibilidade que está diretamente relacionada à idade da mulher, mostrando assim que a mamografia é suscetível a uma alta taxa de falsos positivos como também falsos negativos, causando uma alta proporção de mulheres sem câncer que sofrem uma nova avaliação clínica, enfrentam uma biópsia ou perdem o melhor intervalo de tempo para o tratamento do câncer (BIRD; WALLACE e YANKASKAS. 1992), (KERLIKOWSKE, 2000).

Figura 3 - (a) Mamografia com nódulo benigno selecionado. (b) Mamografia com nódulo maligno selecionado.



Fonte: (HEATH et al. 1998).

Diante das colocações precedentes, nota-se a necessidade de se analisar adequadamente as microcalcificações detectadas pela mamografia, com uma integração entre os especialistas da área, diversas técnicas de processamento de imagem têm sido aplicadas para ajudar a tornar a mamografia um recurso mais preciso e eficiente. Uma das contribuições nessa área são os descritores geométricos da imagem, conforme discutida na próxima seção.

2.2.2 Descritores Geométricos

Na literatura podem ser encontrados vários descritores ou medidas geométricas (WIRTH, 2001), visto que uma maneira de descrever e analisar objetos é através de sua forma. Desta forma, iremos conceituar os descritores geométricos utilizados nesse trabalho, a saber: raio, textura, perímetro, área, suavidade, compacidade, concavidade, côncavo, simetria e dimensão fractal.

- **Raio**

Segmento de reta que une o centro de uma circunfereência de uma superfície da imagem a qualquer ponto da circunferência.

- **Textura**

Segundo (COSTA, YANDRE, 2014), textura é um importante atributo visual presente em imagens, mas que não tem definição formal, visto que a diversidade de texturas torna impossível estabelecer uma definição universal para a mesma. Entretanto, o padrão visual de uma região e características do objeto da imagem relata detalhes sobre o conteúdo da imagem o que é muito útil em visão computacional.

- **Perímetro**

O perímetro é a medida do contorno da superfície da imagem.

- **Área**

Extensão limitada da supefície da imagem.

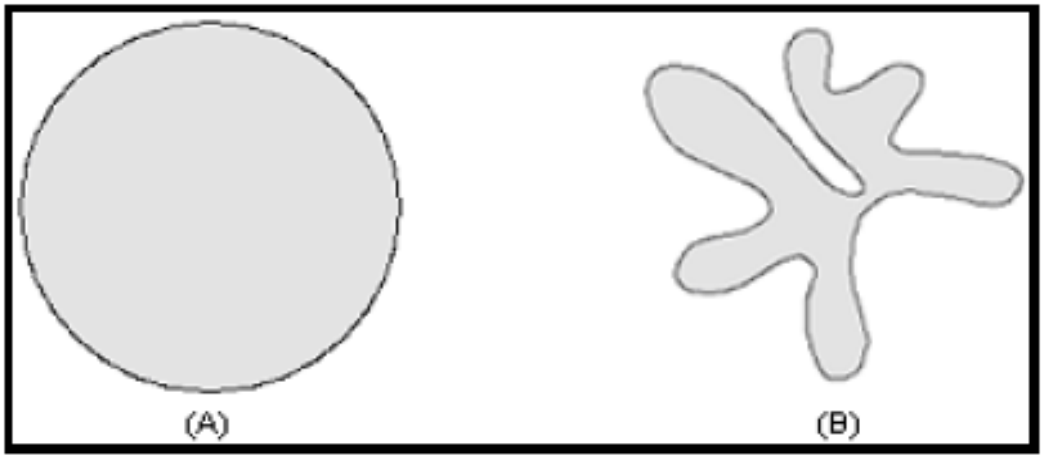
- **Suavidade**

Conceito referente a suavidade da parte analizada da imagem.

- **Compacidade**

É a medida geométrica que mede a densidade do objeto, em comparação com uma figura perfeitamente densa, ou seja, um círculo (SCHOUTEN, 2003).

Figura 4 - Comparação de objetos através da compacidade: (a) compacto; (b) não-compacto.



Fonte: (SCHOUTEN, 2003).

A compacidade é definida por (SAMPAIO, 2009):

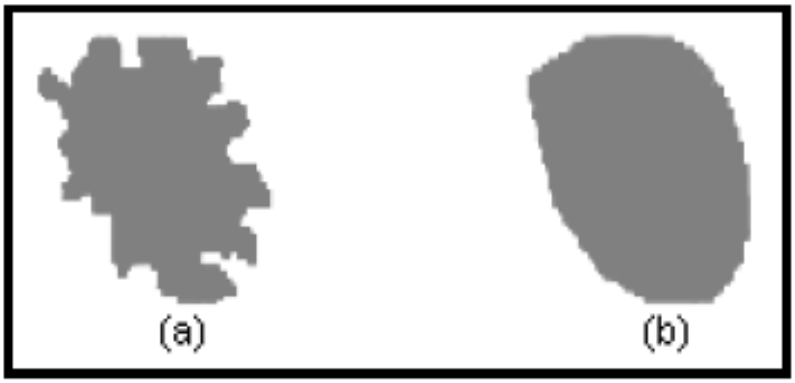
$$C_o = \frac{p^2}{4\pi A} \quad (1)$$

Sendo A a área do objeto em estudo e p o seu perímetro.

- **Concavidade**

Em (FONSECA, 2011), concavidade é a medida geométrica que define o grau de proximidade do objeto com um círculo. Trata-se da razão entre a área do objeto e o perímetro convexo, onde o perímetro convexo é calculado sobre a região que engloba o objeto de forma a não deixar picos ou defeitos no contorno.

Figura 5 - Exemplo de perímetro convexo. (a) Objeto com picos e defeitos. (b) Objeto com superfície convexa.



Fonte: (FONSECA, 2001).

- **Côncavo**

Característica da superfície da imagem que é mais profunda no centro do que na extremidade, ou borda.

- **Simetria**

Característica da superfície da imagem que permite que a imagem seja dividida em partes de igual formato e tamanho.

- **Dimensão fractal**

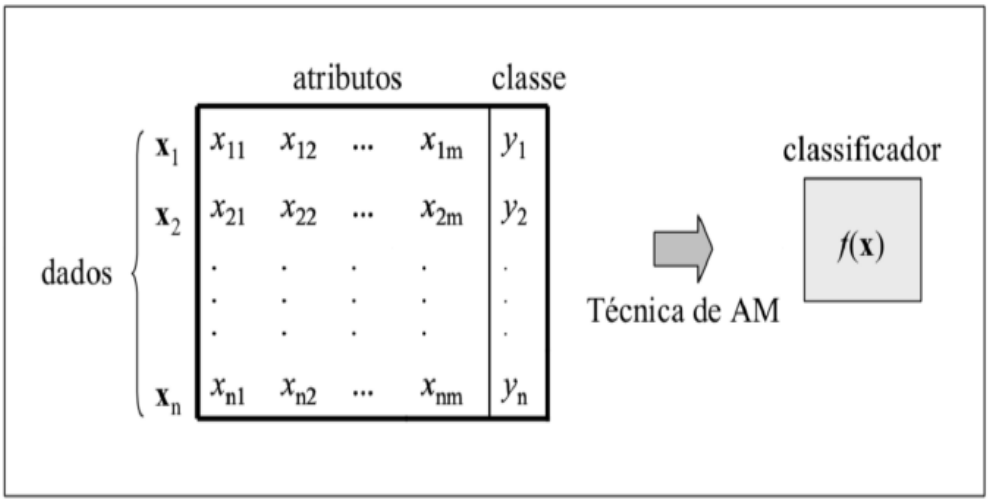
Conforme MANDELBROT (1991), é o número que quantifica o grau de irregularidade ou de fragmentação de um conjunto geométrico, de uma figura ou de um objeto natural e assume, no caso dos objetos da geometria clássica de Euclides, as suas dimensões usuais inteiras.

2.3 APRENDIZADO DE MÁQUINA

Como as técnicas de AM utilizam o princípio de inferência denominado indução, esse pode ser dividido em dois tipos principais: supervisionado e não-supervisionado. O tipo de aprendizado abordado neste trabalho é o supervisionado, que consiste abstratamente, a ideia da figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada, saída desejada (HAYKI, 2001).

Desta forma, dado um conjunto de exemplos rotulados na forma (xi, yi), em que xi representa um exemplo e yi denota o seu rótulo, deve-se produzir um classificador, também denominado modelo, preditor ou hipótese, capaz de predizer precisamente o rótulo de novos dados. Esse processo de indução de um classificador a partir de uma amostra de dados é denominado treinamento (GODA, 2017). O classificador obtido também pode ser visto como uma função f, a qual recebe um dado x e fornece uma predição y.

Figura 6 - Indução de classificador em aprendizado supervisionado.



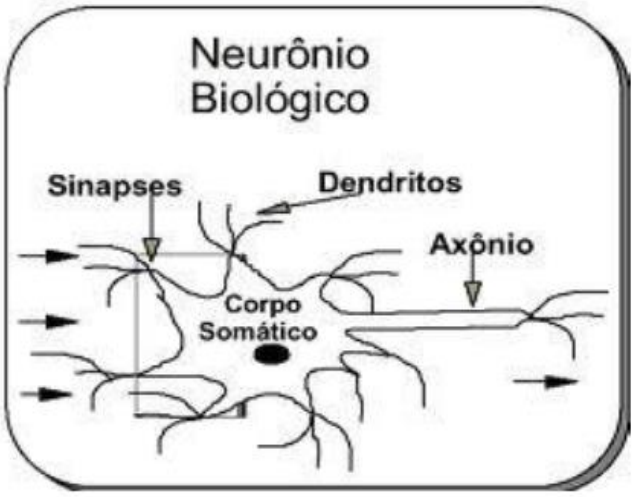
Fonte: (GODA, 2017).

Sendo os rótulos ou classes o fenômeno de interesse sobre o qual se deseja fazer previsões. Se os rótulos possuem valores contínuos, tem-se uma regressão (T. MITCHELL, 1997). Um problema de classificação no qual k = 2 é denominado binário, quando k > 2, se configura um problema multiclasses.

2.3.1 Redes Neurais Artificiais

Em (BRAGA, 2006), tem-se que um neurônio biológico é composto por dendritos que têm por função receber os estímulos elétricos (informações) transmitidos por outros neurônios e encaminhá-los ao corpo do neurônio, também chamado de corpo somático. O corpo somático, por sua vez, é responsável por coletar, combinar e armazenar essas informações até atingir um limiar de excitação (*threshold*), quando, finalmente, acontece uma descarga elétrica através do *axônio*, que é constituído de uma fibra tubular que pode alcançar até alguns metros, e é responsável por transmitir os estímulos elétricos para outros neurônios, conforme mostra a figura 7.

Figura 7 - Representação de um neurônio biológico destacando os seus principais componentes.



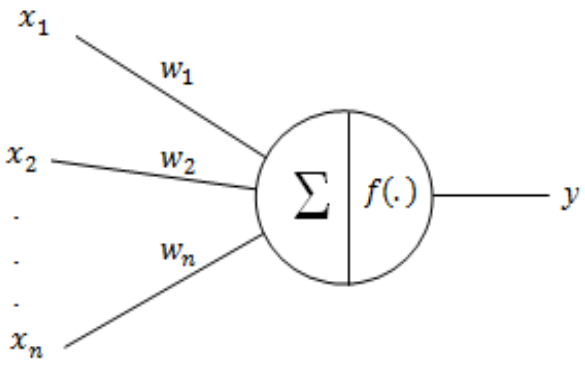
Fonte: Google (2017).

Através de técnicas inspiradas na Natureza, as Redes Neurais Artificiais (RNAs), buscam o desenvolvimento de sistemas inteligentes artificiais que imitem alguns aspectos do comportamento humano, tais como: percepção, raciocínio, aprendizado, evolução e adaptação. (KOVÁCS, 1996; HAYKIN, 2001).

2.3.1.1 Estrutura Básica de um Neurônio Artificial

A primeira estrutura de um neurônio artificial foi proposta pelos pesquisadores McCulloch e Pitts, que possuía n entradas que recebem valores $x_1, x_2, \dots, x_n$ e apenas uma camada de saída y (representado o axônio). Para representar o comportamento das sinapses os terminais de entrada dos neurônios possuem pesos acoplados $w_1, w_2, \dots, w_n$, cujos valores podem ser positivos ou negativos, dependendo se as sinapses correspondentes forem inibitórias ou excitatórias (BRAGA et al., 2012). A Figura 8 demonstra o modelo proposto McCulloch e Pitts.

Figura 8 - Estrutura de um neurônio Artificial.



Fonte: BRAGA *et al.*, 2012

Entretanto, o seguinte modelo de RNA tinha uma grande limitação. Seus pesos eram fixos e não ajustáveis, o que possibilitava apenas a resolução de funções linearmente separáveis (TORRES JUNIOR et al., 2005). O processamento de uma informação neste modelo é dado pela atuação de uma sinapse w_i em uma entrada x_i expressa pela multiplicação $w_i x_i$. Os pesos determinam em que grau o neurônio deve considerar sinais de disparo ou ativação. O disparo do neurônio ocorre através da aplicação de uma função de ativação que ativa ou não a saída, dependendo do valor da soma ponderada das entradas pelos pesos (BRAGA et al., 2012; AGGARWAL; SONG, 1998) (Equação 2).

$$\sum_{i=1}^{n} x_i \, w_i \tag{2}$$

A função de ativação *f* (.) é responsavel por gerar a saída *y* do neurônio *i*, a partir do valor de ativação. O seguinte modelo utilizava uma função de ativação degrau deslocada do limiar (threshold) de ativação θ em relação à origem (Equação 3), o que impossibilitava a resolução de problemas de maior complexidade como os casos não lineares (ROCHA et al., 2008).

$$f(.) = \begin{cases} 1 \sum_{i=1}^{n} x_i \, w_i \;\geq\; \theta \\ 0 \sum_{i=1}^{n} x_i \, w_i \;<\; \theta \end{cases} \tag{3}$$

Essas características fizeram desse modelo um alvo de críticas por não conseguirem resolver problemas mais complexos. Contudo, o trabalho de outros pesquisadores como

Donald Hebb, Frank Rosenblatt, Minsky, Papert, John Hopfield, Rumelhart e outros, re-definindo o paradigma das RNAs através da criação de metodologias de aprendizagem e treinamento, dando origem ao grande número de estruturas de RNAs utilizadas hoje, que podem ser empregadas tanto na solução de problemas de baixa complexidade como os casos lineares, quanto na solução de problemas de alta complexidade como os casos não lineares (LIPPMANN, 1987).

2.3.1.2 Formas de Aprendizagem

Uma RNA possui a característica de aprender por meio de exemplos estraindo conhecimento de um determinado conjunto de dados. O conhecimento é adquirido a partir do processo pelo qual os parâmetros livres de uma rede neural são ajustados por meio de uma forma continuada de estímulo pelo ambiente externo. Este processo é definido como aprendizado que é obtido por meio do treinamento das RNAs.

Para tanto, existem muitos algoritmos de treinamentos que podem ser definidos em dois conjuntos, aprendizado supervisionado, que necessita de um supervisor, e o aprendizado não supervizionado, onde não há um supervisor (HAYKIN, 2001). Dentro do contexto de aprendizado supervisionado, um dos mais típicos processos é aprendizagem por correção de erros, que tem por finalidade ajustar os pesos das sinapses por meio do cálculo da diferença do valor esperado ou desejado pela saída de um neuronio *i* $y_{di}(t)$ no instante *t*, menos o valor predito pela RNA $y_{pi}(t)$ no instante *t* produzindo desta forma um erro $e(t)$ (Equação 4) (BRAGA et al., 2012).

$$e(t) = y_{di}(t) - y_{pi}(t) \tag{4}$$

Este resultado $e(t)$ irá auxiliar no calculo do ajuste aplicado aos pesos, que tem como objetivo minizar o erro da próxima iteração (t+1) (Equação 5).

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \tag{5}$$

Na equação, $w_{ij}(t)$ representa o valor do peso no tempo *t*, e $\Delta w_{ij}(t)$ o valor do ajuste a ser aplicado ao peso, que segundo Yi; York (2004), é alcançado através da minimização da função de custo $E(t)$ dada pelo erro médio quadrático (Equação 6):

$$E(t) = \frac{1}{2}\sum_{i=1}^{n} e_i^2(t) \tag{6}$$

Na equação anterior, *n* é o numero de amostras do treinamento, $e_i^2(t)$ é o erro ao quadrado originado pela saída esperada $y_{di}(t)$ menos a predita pela RNA $y_{pi}(t)$. Esta minimização resulta na regra do delta ou regra de Widrow-Hoff, em que o valor do ajuste e $\Delta w_{ij}(t)$ é a originado da multiplicação de uma constante positiva que determina a taxa de aprendizado η, pelo produto do erro $e(t)$ e a entrada da sinapse x_i (Equação 7) (BRAGA et al., 2012; MULLUHI et al., 1993).

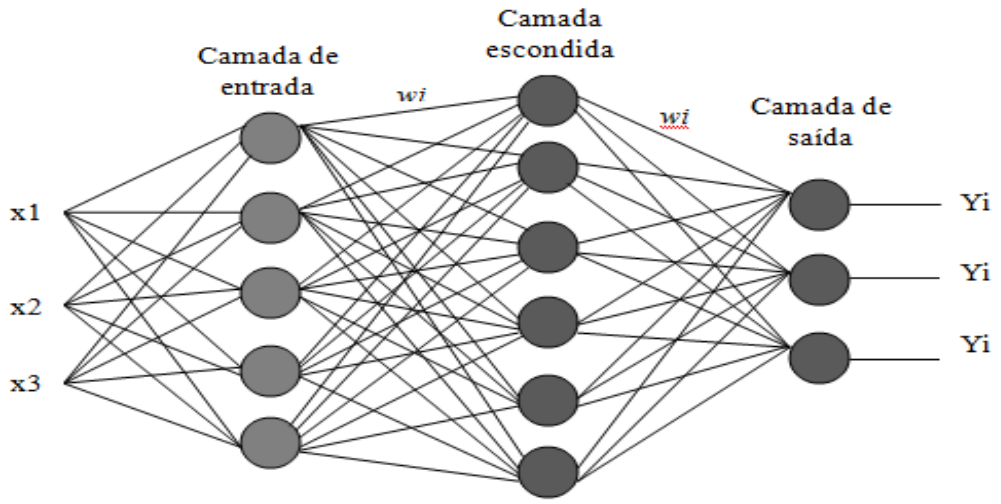
$$\Delta w_{ij}(t) = \eta e_i(t) x_i(t) \tag{7}$$

2.3.1.3 Redes Perceptrons de Múltiplas Camadas (MLP)

A RNA do tipo MLP ou em inglês *multilayer perceptron*, que pertence a uma classe de RNAs conhecida como *feedfoward*, é uma aproximadora universal de funções que inicialmente foi implementada para a resolução do problema não linear do XOR. Devido ao seu sucesso, ela vem sendo aplicada em diferentes problemas combinatórios e na solução de diferentes tarefas como processamento de informações, reconhecimento de padrões, previsões de tempo, problemas de classificação, processamento de imagens, previsões de atividades sísmicas e outros (SHAH; GHAZALI, 2011).

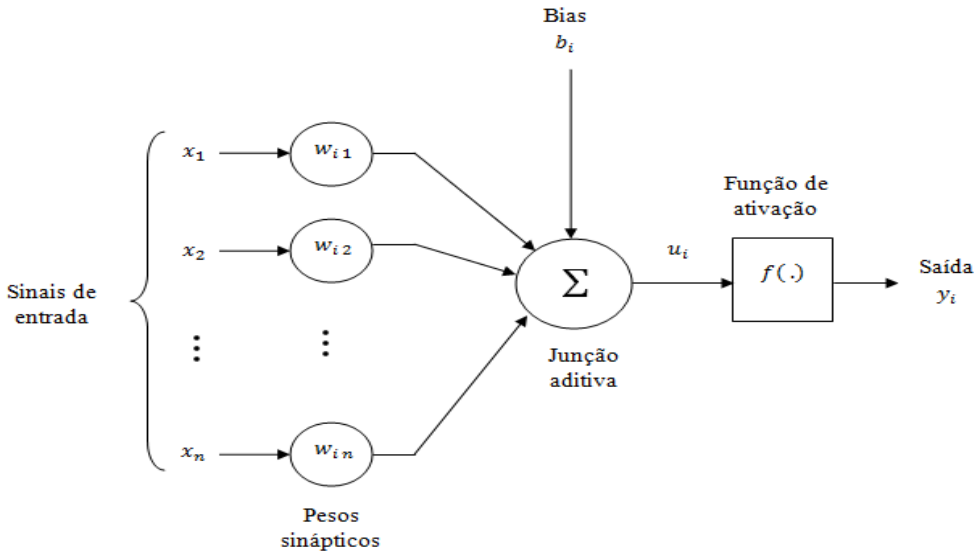
A estrutura de uma MLP é formada por um conjunto de neurônios dispostos em camadas, contendo uma camada de entrada, uma ou mais camadas intermediárias ou escondidas, e uma camada de saída. Cada um dos neurônios da camada de entrada está conectado a todos os neurônios da camada intermediária. Da mesma forma, cada neurônio da camada intermediária está conectado com todos os neurônios da camada de saída (MACHADO, 2005; ASADUZZAMAN et al., 2010). A Figura 9 demonstra a estrutura de uma rede MLP.

Figura 9 - Estrutura de um RNA do tipo MLP.



Diferente do modelo proposto por McCulloch e Pitts onde o neurônio simplesmente comparava o somatório ponderado das entradas pelos pesos a um limiar que gerava a saída binária, ou seja, zero ou um. A RNA do tipo MLP possui neurônios capazes de gerar qualquer saída (HAYKIN, 2001; BRAGA et al, 2012; GUARNIERI, 2006) . Desta forma cada neurônio da RNA do tipo MLP deve ser interpretado como modelo de um neurônio não linear expresso na Figura 10.

Figura 10 - Modelo de neurônio não linear.



Fonte adaptado de Haykin (2001).

Na figura (10), u_i é a saída da combinação linear do somatório das entradas ponderadas pelos pesos w_i, mais o bias (b_i). A saída do neurônio y_i é originada pela aplicação de uma função $f(.)$ no resultado originado da combinação linear do somatório das entradas ponderadas pelos pesos w_i, mais o bias (b_i) que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo (HAYKIN, 2001). Desta forma, o neurônio *i* é expresso pelas (Equações 8, 9):

$$u_i = \sum_{i=1}^{n} w_{ij} x_i + b_i \tag{8}$$

$$y_i = f(u_i) \tag{9}$$

Existem muitos tipos de funções de ativação. Contudo, as mais empregadas são as funções degrau, linear, sigmoides (Equações 10, 11, 12, 13) (HAYKIN, 2001; BRAGA et al., 2012; AGGARWAL; SONG, 1998).

a) Função de ativação degrau:

$$f(u_i) = \begin{cases} 1 & se \ u_i \geq 0 \\ 0 & se \ u_i < 0 \end{cases} \tag{10}$$

b) Função Linear:

$$f(u_i) = u_i \tag{11}$$

c) Funções Sigmóides:

1) Função Logística:

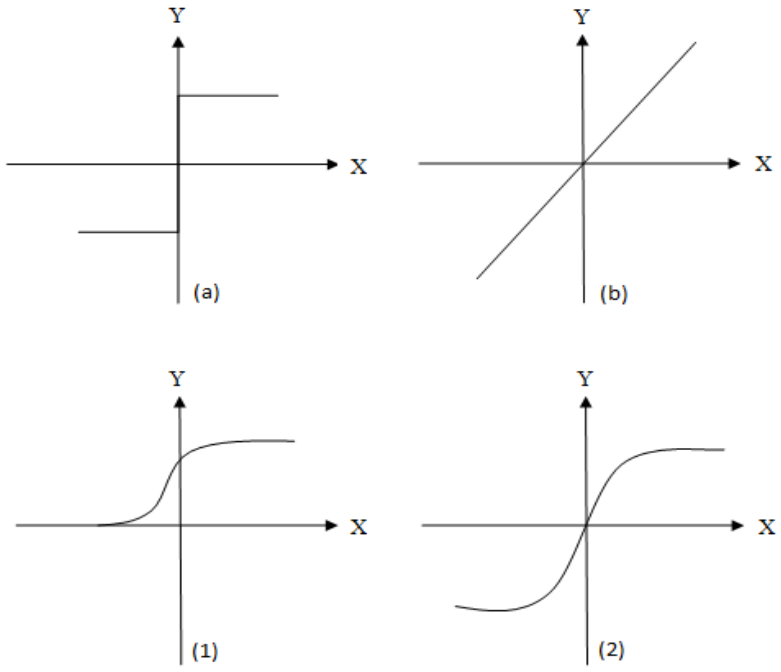
$$f(u_i) = \frac{1}{(1 + exp^{(-ui)})} \tag{12}$$

2) Função Tangente Hiperbólica:

$$f(u_i) = tgh\left(\frac{u_i}{2}\right) = \frac{1 - exp(-ui)}{1 + exp(-ui)} \tag{13}$$

O comportamento das funções é expresso na Figura 11, onde as entradas dos neurônios são representadas pelo eixo x e as saídas pelo eixo y, como demonstrado acima a função degrau pode originar saídas no intervalo entre 0 e 1 (AGGARWAL; SONG, 1998; LIPPMANN, 1987; SKODZIK et al., 2013). A função linear tem como saída mesmo valor de u_i, o que resulta num gradiente igual a 1 e as funções sigmóides, como a logística e a tangengente hiperbólica produzem saídas entre dois intervalos 0 a 1 e -1 e 1.

Figura 11 - Gráficos dos comportamentos das funções de transferência. (a) Função Degrau, (b) Função Linear, (1) Função Logística, (2) Função Tangente Hiperbólica



No entanto, o resultado da saída final da RNA é obtido após o processo de treinamento que faz o ajuste dos pesos através de um algoritmo. Segundo BRAGA, et al. (2012), o treinamento da RNA do tipo MLP por meio do algoritmo *back-propagation* padrão pode ser lento para várias aplicações, e seu desempenho ainda pode piorar para problemas mais complexos. Por isso, desde a criação deste algoritmo, várias modificações foram propostas tanto para acelerar seu tempo de treinamento quanto para melhorar seu desempenho. Por meio destas variações foram desenvolvidos os métodos de treinamento: *Quickprop, Levenberg-Marquardt, back-propagation* com Newton, Rprop e outros.

2.3.2 Máquinas de Vetor de Suporte (SVMs)

Máquinas de vetores de suporte são sistemas de reconhecimento de padrões, usualmente empregados em problemas de classificação binária, que definem funções discriminantes de margem ótima de separação de classes.

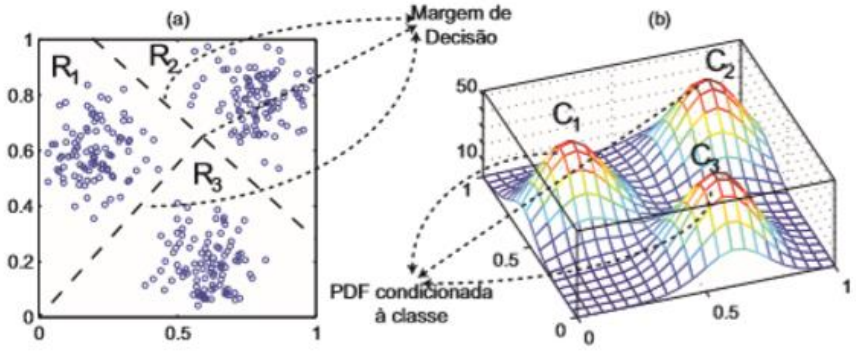
Sistemas de reconhecimento de padrões desempenhm classificações multi-classes, com múltiplos atributos, independente do tipo de regra de decisão aplicada. Um classificador padrão $\boldsymbol{x}$ à uma classe w_k, $k \in \{1,2,3, \ldots, c\}$, particionando o espaço de atributos em segmentos lineares, áreas, volumes, e hiper-volumes, denominadas regiões de decisão (figura 4.1). A região de decisão (R_k) de uma classe pode ser descontínua, e as margens entre regiões de decisão adjacentes são denominadas margens de decisão, ou de separação.

Decisões de classificação baseadas em vetores de atributos $\boldsymbol{x}$ podem ser definidas pelo uso de funções discriminantes explicitamente definidas como,

$$d_k(\boldsymbol{x}), \qquad k = 1, 2, \ldots, c, \tag{14}$$

onde cada função discriminante está associada com uma classe particular conhecida w_k, $k = 1,2, \ldots, c$.

Figura 12: Regiões e Margens de decisão. (a) Regiões de decisão R_k e Margens de decisão para três classes não-sobrepostas; (b) Função de densidade de probabilidade associadas às classes.



Em problemas dicotômicos, o de classificação binária, emprega-se uma única função discriminante $d(\boldsymbol{x})$, que atribui classe de acordo com o sinal da resposta:

$$d(\boldsymbol{x}) = d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}), \quad (15)$$

ou seja, $d_1(\boldsymbol{x})$ é maior que $d_2(\boldsymbol{x})$, $d(\boldsymbol{x}) > 0$, e o padrão $\boldsymbol{x}$ é atribuído à classe1, caso contrário este é atribuído à classe 2.

Em problemas lineares dicotômicos, as tuplas de treinamento, dados por $(\boldsymbol{x_1},y_1)$, $(\boldsymbol{x_2},y_2)$, . . ., $(\boldsymbol{x_t},y_t)$, $\boldsymbol{x} \in \Re^n$, $y \in \{-1, 1\}$, são linearmente separáveis, e há diferentes hiperplanos capazes de separar as classes. Usando os dados de treinamento durante a aprendizagem, o sistema de reconhecimento de padrões, expresso por uma máquina de aprendizado, encontra pesos $\boldsymbol{w} = [w_1, w_2, \ldots, w_n]^T$ e o bias b, ou 1 x w_0, de uma função discriminante dada por:

$$d(\boldsymbol{x}, \boldsymbol{w}, b) = \boldsymbol{w}^T \boldsymbol{x} + b = \sum_{i=1}^{n} w_i x_i + b, \text{ onde: } \boldsymbol{x}, \boldsymbol{w} \in \Re^n. \quad (16)$$

Após treinamento bem sucedido, usando-se os parâmetros obtidos (eq. 16), a máquina de aprendizado, dado um padrão $\boldsymbol{x}$ desconhecido, produz uma saída f_o de acordo com a função indicadora dada por:

$$i_F = f_o = sgn\,(d(\boldsymbol{x}, \boldsymbol{w}, b)). \quad (17)$$

ou:

$$\text{Se } d(\boldsymbol{x}, \boldsymbol{w}, b) > 0, \quad \boldsymbol{x} \in 1 \text{ (i.e., } o = y_1 = +1), \text{ e}$$

$$\text{Se } d(\boldsymbol{x}, \boldsymbol{w}, b) < 0, \quad \boldsymbol{x} \in 2 \text{ (i.e., } o = y_1 = -11).$$

A margem de decisão é então obtida pela intersecção entre a função discriminante (eq. 17), $d(\boldsymbol{x}, \boldsymbol{w}, b)$, e o espaço de atributos, sendo dada por,

$$d(\boldsymbol{x}, \boldsymbol{w}, b) = 0. \quad (18)$$

O aprendizado pode ser considerado como a descoberta da melhor função f_o de parâmetros ajustáveis $\boldsymbol{w}$ e b, utilizando os dados de treinamento disponíveis. Para se medir a qualidade de uma função qualquer f_o, deve-se definir medidas apropriadas (função de erro, custo ou perda), sendo as mais comuns: (1) Em regressão: O erro quadrático (norma L_2), definido como $L(y, f_o(\boldsymbol{x},y)) = (y - f_o)^2$; e o erro adsoluto (norma L_1), definido com $L(y, f_o(\boldsymbol{x},y)) = |\, y - f_o|$; e (2) Em classificação (binária).

$$L(y, f_o(\boldsymbol{x},y)) = 0, \text{ se } f_o(\boldsymbol{x},y) = y;$$

$$L(y, f_o(\boldsymbol{x},y)) = 1, \text{ se } f_o(\boldsymbol{x},y) \neq y.$$

O erro médio, ou risco esperado, de uma função teórica de regressão $f(.)$, com função de erro dada pela norma L_2 é dada por:

$$R[f] = \boldsymbol{E}[(y - f(\boldsymbol{x}))^2] = \int (y - f(\boldsymbol{x}))^2 P(\boldsymbol{x},y) D\,\boldsymbol{x} Dy, \tag{19}$$

onde, ($\boldsymbol{x}$, y) são tuplas de treinamento e $f(\boldsymbol{x})$ é definida como a média de P ($\boldsymbol{x}|y$).

No aprendizado baseado em dados a função de densidade de probabilidade $P(\boldsymbol{x},y)$ (eq. 19) é desconhecida, e o aprendizado deve basear-se nos dados amostrais (tuplas) de treinamento. O algoritmo de aprendizado deve de uma máquina de aprendizado deve descobrir a relação entrada-saída, função $f_o(\boldsymbol{x})$, usando apenas os dados de treinamento, o que, até certo ponto, aproxima-se da minimização do risco esperado $R[f]$ no espaço de busca T, ou seja,

$$f_o(\boldsymbol{x}) = \underset{fo\,\epsilon T}{arg\ min}\ R[f]. \tag{20}$$

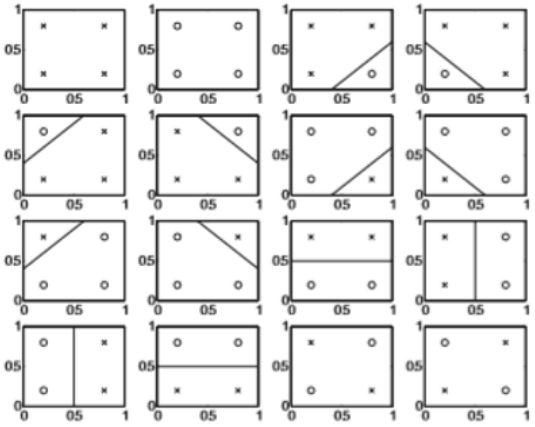
Uma maneira de proceder é usar os dados de treinamento para aproximar a integral no risco esperado (eq. 19) por uma soma finita. Isso leva à definição do risco empírico:

$$R_{emp}[f] = \frac{\sum_{i=1}^{l} (\, y - f(x_i\ , w))^2}{l} \tag{21}$$

onde l é o tamanho da amostra de treinamento e o conjunto de parâmetros w é o sujeito do aprendizado.

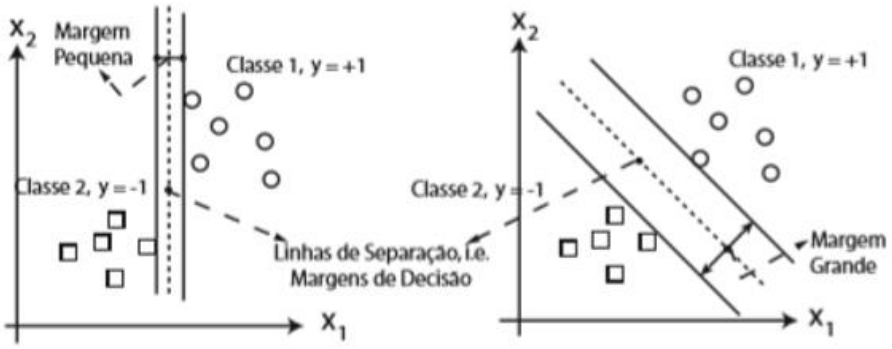
A capacidade de um conjunto de funções indicadoras i_F ($\boldsymbol{x}$, $\boldsymbol{w}$) é expressa pela dimensão Vc (figura 13), que é o número máximo de pontos, h, que podem ser separadas em todas as formas possíveis por essas funções. Por exemplo, a dimensão VC do hiperplano orientado em um espaço n-dimensional é igual a $n + 1$ (i.e., $h = n+ 1$). Enquanto a capacidade de $f_{n,1}$ aumenta (ex.: aumento do número de vetores de suporte), a capacidade de aproximação da máquina de aprendizado aumenta pelo uso de parâmetros ajustáveis adicionais, ou seja, menor graus de liberdade.

Figura 13: Dimensão VC para classificadores lineares. 4 pontos em $\Re^{n=2}$separadas em todas as formas 2^4= 16 possíveis pela função indicadora i_F ($\boldsymbol{x}$, $\boldsymbol{w}$) = *sgn(u)*, representada pela linha $u = 0$. Para i_F ($\boldsymbol{x}$, $\boldsymbol{w}$), $h = n + 1$= 3. Os dois últimos casos inferiores correspondem ao problema clássico *XOR*, que não é separável por funções lineares.



Com a disponibilidade apenas das tuplas de treinamento, além de encontramos o classificador que melhor aproxima o risco esperado pela minimização do risco empírico, desejamos encontrar entre todos os hiperplanos que minimizam o erro empírico, aquele de maior margem de separação, *M*, entre classes (figura 14).

Figura 14: Planos de separação admissíveis. Direita, uma boa solução com margem grande, e esquerda, uma menos aceitável com margem pequena



Geometricamente, a margem de separação *M* que deve ser maximizada durante o treinamento, é a projeção, no plano normal de pesos, da distância entre dois vetores de suporte quaisquer, pertencentes às duas diferentes classes. Essa margem é igual a:

$$M = (\boldsymbol{x_1} - \boldsymbol{x_2})_{\mathrm{w}} = (\boldsymbol{x_1} - \boldsymbol{x_3})_{\mathrm{w1}} \qquad (22)$$

podendo ser encontrada como segue (fig. 14)

$$M = D_1 - D_2, \qquad (23)$$

$$D_1 = \| \boldsymbol{x_1} \| \cos(\alpha) \quad \text{e} \quad D_2 = \| \boldsymbol{x_2} \| \cos(\beta), \qquad (24)$$

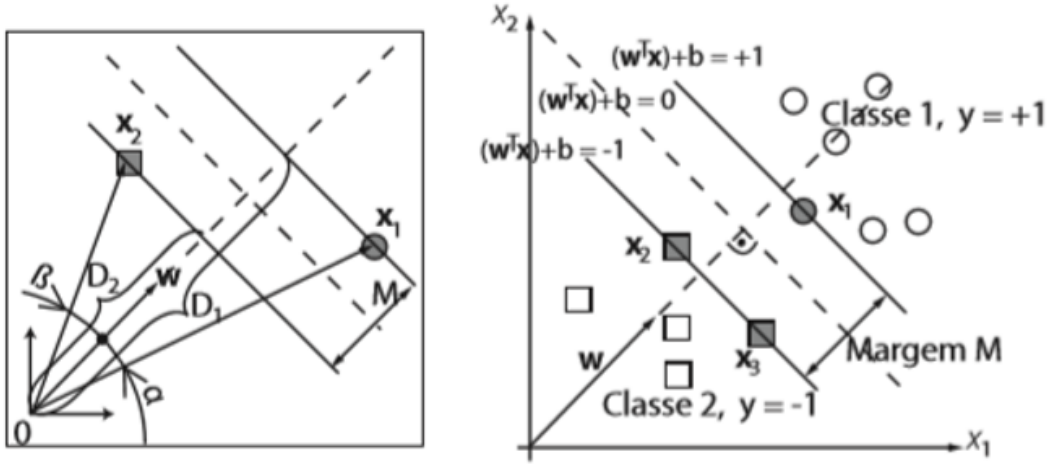
$$\cos(\alpha) = \frac{x_1^T w}{\|x_1\|\|w\|} \qquad \text{e} \qquad \cos(\beta) = \frac{x_2^T w}{\|x_2\|\|w\|} \qquad (25)$$

$$\therefore M = \frac{x_1^T w - x_2^T w}{\|w\|} \qquad (26)$$

Usando o fato de $\boldsymbol{x_1}$ e $\boldsymbol{x_2}$ serem vetores de suporte, i.e., $\boldsymbol{w^T x_1} + b = 1$ e $\boldsymbol{w^T x_2} + b = -1$, obtemos:

$$M = \frac{2}{\|w\|} \quad \text{, onde: } \|w\| = \sqrt{\langle w, w \rangle} = \sqrt{w_1^2 + w_2^2 + \ldots + w_n^2}\,. \qquad (27)$$

Figura 15: O hiperplano canônico de separação ótima e os vetores de suporte no espaço primordial. OS pontos satisfazendo $\boldsymbol{w^T x_1} + b = 1$ e $\boldsymbol{w^T x_2} + b = -1$ são vetores de suporte e o hiperplano canônico de separação ótima, de maior margem, satisfaz $y_i \mid \boldsymbol{w^T x_1} + b \mid \geq 1$, $i = 1, \ldots, l$.



A minimização da norma do vetor de peso $\|w\|$ (eq. 27) é igual a minimização de $\boldsymbol{w^T w} = \langle w, w \rangle = \sum_{i=1}^{n} w_i^2$, que leva à maximização da margem M. O hiperplano canônico de separação ótima definido pela margem $M = 2 / \|\boldsymbol{w}\|$, específica vetores de suporte, que satisfazem $y_j \mid \boldsymbol{w^T x_j} + b \mid = 1$, $j = 1, \ldots, N_{sv}$ (Figura YY1). Ao mesmo tempo o hiperplano canônico de separação ótima satisfaz as inequações:

$$y_i \mid \boldsymbol{w^T x_i} + b \mid \geq 1, \; i = 1, \ldots, N_{sv}, \tag{28}$$

onde N_{sv} indica o número de vetores de suporte.

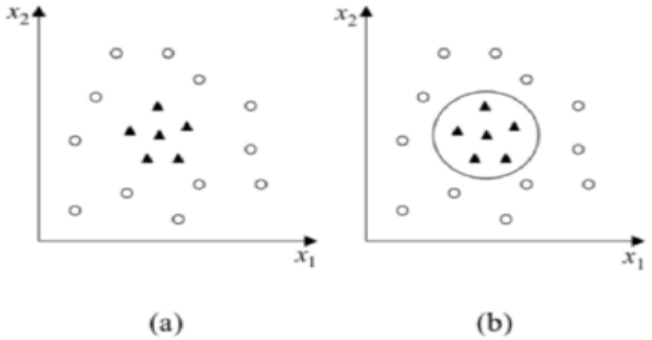
Para o caso de amostras não linearmente separáveis as SVMs irá classifica-las usando uma transformação não linear que transforme o espaço entrada (dados) para um novo espaço (espaço de características). Esse espaço deve apresentar dimensão suficientemente grande, e através dele, a amostra pode ser linearmente separável. Sendo assim, o hiperplano canônico de separação ótima é definido como uma função linear de vetores retirados do espaço de característica ao invés do espaço de entrada original (HAYKIN, 2001).

2.3.2.1 Maquinas de Vetores de Suporte (SVMs) Não Lineares

O problema de classificação inicial abordado pelas SVMs (problema de classificação binária) trata da classificação de duas classes, sem perda de generalidade, através de um hiperplano canônico de separação ótima a partir de conjunto de treinamento linearmente separável se for possível separar os padrões de classes diferentes contidos no mesmo por pelo menos um hiperplano (Haykin 2001, Semolini 2002).

Entretanto, há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano. Um exemplo é apresentado na figura 16, em que o uso de uma fronteira curva seria mais adequado na separação das classes.

Figura 16- Fronteira de curva para separação de classes: (a) Conjunto de dados não linear; (b) Fronteira não linear no espaço de entradas.



Segundo (M. A. HEARST, B. SCHÖLKOPF, S. DUMAIS, E. OSUNA, AND J. PLATT, 1998) as SVMs lidam com problemas não lineares mapeando o conjunto de treinamento de seu espaço original, referenciado como de entradas, para um novo espaço de maior dimensão, denominado espaço de características (feature space).

Sendo assim, seja Φ: X → £ um mapeamento, em que X é o espaço de entradas e £ denota o espaço de características; observa-se que a escolha apropriada de Φ faz com que o conjunto de treinamento mapeado em £ possa ser separado por uma SVM linear. Usa-se esse procedimento motivado pelo teorema de Cover (HAYKIN, 2001).

Dado um conjunto de dados não linear no espaço de entradas X, o teorema de Cover, afirma que X pode ser transformado em um espaço de características £ no qual com alta probabilidade os dados são linearmente separáveis, sendo que para isso duas condições devem ser satisfeitas, a saber: a primeira condição é que a transformação seja não linear, enquanto a segunda condição é que a dimensão do espaço de características seja suficientemente alta.

De forma a ilustrar esses conceitos, considere o conjunto de dados apresentado na Figura 16(a) (K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, 2001); transformando os dados de R^2 para R^3 com o mapeamento representado na Equação 21, o conjunto de dados não linear em R^2 torna-se linearmente separável em R^3 (Figura 16b). Sendo assim possível encontrar um hiperplano capaz de separar esses dados, descrito na Equação 22. Pode-se verificar que a função apresentada, embora linear em R^3 (Figura 16b), corresponde a uma fronteira não linear em R^2.

$$\phi(\mathrm{x}) = \phi\,(x_1, x_2) = (x_1^2, \sqrt{2}\,x_1x_2, x_2^2) \tag{29}$$

$$\mathrm{f(x)} = w\,.\phi(\mathrm{x}) + \mathrm{b} = w_1x_1^2 + w_2\sqrt{2}\,x_1x_2 + w_3x_2^2 + \mathrm{b} = 0 \tag{30}$$

Sendo assim, inicialmente se mapeia os dados para um espaço de maior dimensão aplicando a função ϕ e em seguida aplica-se a SVM Lienar sobre esse espaço, esta por sua vez, encontra o hiperplano com maior margem de separação, garantindo assim uma boa generelização.

2.3.2.2 Hiperplano Ótimo para Classes não Linearmente Separáveis

De um modo geral observa-se que o problema de classificação onde a classe não é linearmente separável a construção de um hiperplano de separação, sendo dado os padrões de treinamento, provavelmente se gerará erros de classificação. Tento em vista isto o objetivo da SVM é encontrar um hiperplano que minimiza a probabilidade de erro de classificação junto ao conjunto de treinamento.

Há casos onde não é necessário fazer um mapeamento de características no conjunto de treinamento. Esses casos tratados pela SVM linear com margens de separação suaves ou flexíveis, pois poderão existir pontos (x_i,d_i) que violaram a (eq.30).

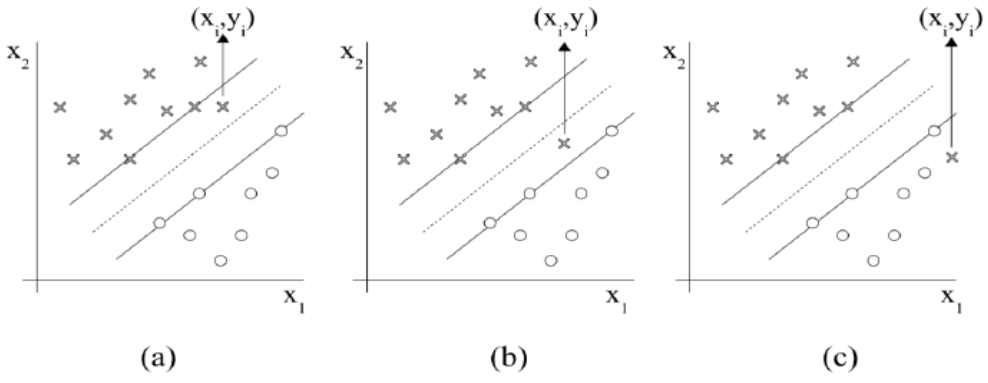
$$d_i(w^T x_i + \mathrm{b}) \geq 1 \tag{31}$$

Onde: x_i é o padrão de entrada o i-ésimo exemplo e d_i é a resposta desejada, d_i = {+1, -1} que representa as classes linearmente separáveis; $w^T x_i$ é produto escalar entre os vetores *w* e *x*, em que *x* é o vetor de entrada que representa os padrões de entrada do conjunto e *w* é vetor de pesos ajustáveis e *b* é um limiar conhecido como bias.

Esta violação pode ocorrer em três diferentes situações:

- O ponto (x_i, d_i) se encontra dentro da região de separação e no lado correto da superfície de decisão, como mostra na figura 17 (a). Neste caso, houve uma escolha incorreta do hiperplano.
- O ponto (x_i, d_i) se encontra dentro da região de separação e no lado incorreto da superfície de decisão, como mostra na figura 17 (b). Neste caso, houve uma escolha incorreta do hiperplano de margem maior.
- O ponto (x_i, d_i)) se encontra fora da região de separação e no lado incorreto da superfície de decisão, como mostra a figura 17 (c).

Figura 17- Hiperplano ótimo para classes não linearmente separáveis: (a) o ponto (x_i, d_i)) se encontra na região de separação, mas do lado correto. (b) O ponto (x_i, d_i)) se encontra na região de separação, mas do lado incorreto. (c) O ponto (x_i, d_i) se encontra fora da região de separação, mas do lado incorreto.



Com isso introduz-se uma variável negativa $\{\xi_i\}$ $1 \leq i \leq n$ na definição do hiperplano de separação.

$$d_i(w^T x_i + b) \geq 1 + \xi_i \tag{32}$$

As variáveis ξ_i são chamadas de variáveis soltas, e medem os desvios dos pontos (x_i, d_i) para a condição ideal de separação de classes. Sabe-se que:

- Quando ξ_i satisfazer $0 \leq \xi_i \leq 1$ o ponto encontra-se dentro da região de separação, mas do lado correto da superfície de decisão.
- Quando $\xi_i > 1$ o ponto encontra-se do lado incorreto do hiperplano de separação.

Os vetores de suporte são pontos que o resultado da (eq. 32) é igual a $1-\xi_i$ mesmo que $\xi_i > 0$. Ao retirar um padrão do conjunto treinamento em que $\xi_i > 0$ a superfície de decisão tem possibilidade de mudança, porém, ao retirar um padrão em que $\xi_i = 0$ e se o resultado da (eq. 32) for maior que 1, a sua superfície de decisão permanecerá inalterada.

Sendo assim o objetivo é encontrar um hiperplano de separação onde o erro de classificação incorreta seja mínimo perante o conjunto de treinamento, podendo ser feito minimizando a equação:

$$\phi(\xi) = \sum_{i=1}^{N} (\xi_i - 1) \tag{33}$$

em relação ao vetor peso $\boldsymbol{w}$, sujeito a restrição da equação do hiperplano de separação da (eq. 30) e a restrição sobre $w^T w$. A função $I(\xi-1)$ é uma função indicadora, definida por:

$$I(\xi-1) = \begin{cases} 0 \; se \; \xi \leq 0 \\ 0 \; se \; \xi > 0 \end{cases} \tag{34}$$

A minimização de $\phi(\xi)$ é um problema de otimização não convexo de classe NP-completo não determinístico em tempo polinomial (HAYKIN, 2001). Logo para fazer este problema de otimização matematicamente tratável, aproxima-se a função $\phi(\xi)$ por:

$$\phi(\xi) = \sum_{i=1}^{n} \xi_i \tag{35}$$

Para a simplificação de cálculos computacionais a função a ser minimizada em relação ao vetor peso $\boldsymbol{w}$ segue:

$$\phi(\boldsymbol{w}, \xi) = \frac{1}{2} \boldsymbol{w}^T . \boldsymbol{w} + C \sum_{i=1}^{n} \xi_i \tag{36}$$

onde o parâmetro C controla a quantidade da complexidade do algoritmo e o número de amostras do conjunto de treinamento classificados incorretamente, sendo denominado *parâmetro de penalização.* A minimização do primeiro termo da (eq. 36) está relacionada á

minimização da dimensão VC da SVM. O segundo termo pode ser visto como um limitante superior para o número de erros no padrão treinamento apresentados a máquina. Dái, a (eq. 36) satisfaz os princípios de minimização do risco estrutural.

O problema de otimização em sua representação primal para encontrar o hiperplano ótimo de separação para classes não linearmente separáveis pode ser escrito como:

$$Minimizar: \frac{1}{2} w^T . w + C \sum_{i=1}^{N} \xi_i \quad (37)$$

Sujeito as restrições: $\begin{cases} (1) & d_i(w^T x_i + b) - \xi_i, para\ i = 1, \dots, N \\ (2) & \xi_i \geq 0, \forall\ i = 1, \dots, N \end{cases}$

Há casos também onde, há a necessidade de mapear o espaço de entrada não linear para um espaço de características. Para realizar esse mapeamento, as funções Kernel ou produto do núcleo interno são utilizados.

2.2.2.3 Funções de Kernel

Um kernel *k* é uma função que recebe dois pontos $\boldsymbol{x_i}$ e $\boldsymbol{x_j}$ do espaço de entrada e computa o produto escalar $\varphi^{\boldsymbol{T}}(\boldsymbol{x_i}).\ \varphi(\boldsymbol{x_j})$ no espaço de características.

O termo $\varphi^{\boldsymbol{T}}(\boldsymbol{x_i}).\ \varphi(\boldsymbol{x_j})$ representa o produto interno dos vetores $\boldsymbol{x_i}$ e $\boldsymbol{x_j}$, sendo o kernel representado por:

$$\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \varphi^{\boldsymbol{T}}(\boldsymbol{x_i}).\ \varphi(\boldsymbol{x_j}) \quad (38)$$

A utilização de kernels está na simplicidade de cálculos e na capacidade de representar espaços muitos abstratos. As funções φ devem pertencer a um domínio em que seja possível o cálculo de produtos internos. No geral, utiliza-se o teorema de Mercer para satisfazê-las. Segundo o teorema, os *kernels* devem ser matrizes posititivamente definidas, isto é, $\boldsymbol{k_{i,j}} = \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j})$, para todo $i,j = 1, ..., N$, deve ter auto-vetores maiores que 0.

As funções φ devem pertencer a um domínio em que seja possível o cálculo de produtos internos. No geral, utiliza-se o teorema de Mercer para satisfazê-las. Segundo o teorema, os *kernels* devem ser matrizes positivamente definidas, isto é, $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, para todo $i,j = 1, ..., N$, deve ter auto-vetores maiores que 0.

Alguns *kernels* mais utilizados são: os polinomiais, os gaussianos ou RBF (*Radial Basis Function*) e o sigmoidais.

Tabela 3: Principais *kernels* utilizados nas SVMs

Kernel	Função k ($\boldsymbol{x}_i.\boldsymbol{x}_j$)	Comentários
Polinomial	$(\boldsymbol{x}_i^T \cdot \boldsymbol{x}_j + 1)^p$	p é especificado *a priori* pelo usuário
RBF	$e^{(-\frac{1}{2\sigma^2}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)}$	a largura de σ^2 é especificada *a priori* pelo usuário
Sigmoidal	$\tanh(\beta_0 \boldsymbol{x}_i^T \cdot \boldsymbol{x}_j + \beta_1)$	Teorema de Mercer satisfeito somente para β_0 e β_1

A obtenção de um classificador por meio do uso de SVMs envolve a escolha de uma função *kernel* apropriada, além de parâmetros desta função e do algoritmo de determinação do hiperplano ótimo. A escolha do *kernel* e de seus parâmetros afetam o desempenho do classificador através da superfície de decisão.

3 MATERIAL E MÉTODOS

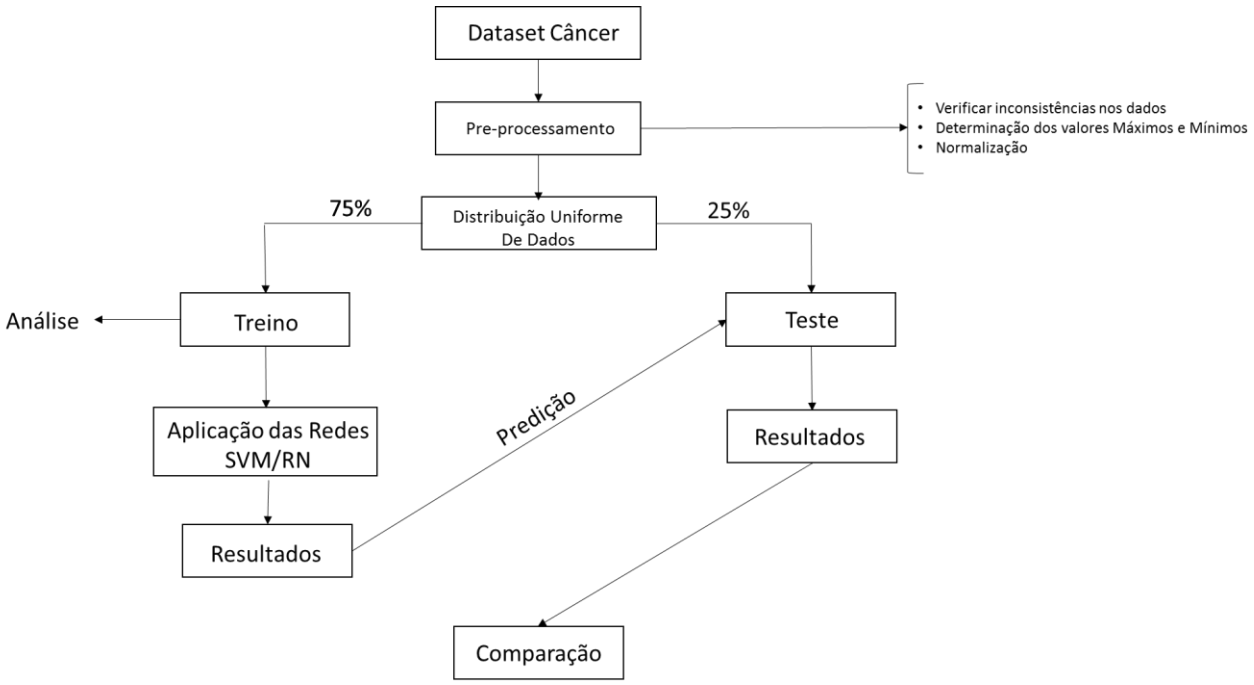
3.1 Base de Dados

O conjunto utilizado nesse estudo consiste de 569 dados, proveniente de pacientes com suspeita de câncer de mama obtidos junto ao Instituto de Radiologia da Universidade Erlangen-Nuremberg, no período de 2003 a 2006. O banco de dados possui informações clínicas sobre raio, textura, perímetro, área, suavidade, compacidade, concavidade, côncavo, simetria e dimensão fractal.

3.2 Modelagem Computacional

O método proposto neste estudo consiste em um modelo computacional baseado em: Rede Neural Artificial (RNA) e Máquinas de Vetor de Suportes não Linear (SVMs-não linear) na classificação de malignidade em dados mamográficos. O fluxograma do modelo proposto é exemplificado na figura 18 e suas etapas são descritas a seguir:

Figura 18 – Fluxograma do modelo proposto.



Os modelos propostos foram desenvolvidos utilizando o software R (http://www.R-project.org/). O pacote *Kernellab* foi utilizado na implementação do Modelo SVMs-não linear e no Modelo de Rede Neural MLP foi utilizado o *Neuralnet*. O pré-processamento dos dados consistiu em analisar a existência de amostras com dados faltantes. O conjunto de treinamento e teste foram divididos de forma uniformemente distribuídos com proporção de 75% e 25% respectivamente para os grupos em tela, além da normalização dos dados.

Em Rocha *et all* (2008), tem-se que o procedimento de normalização também pode ser designado por escalonamento, que é responsável por realizar uma transformação nos dados de forma que o algoritmo de treinamento ou ajuste possa obter melhores resultados. Sendo assim esse procedimento codifica todos atributos em intervalos semelhantes fazendo com que todos dados tenham a mesma importância.

Sabe-se que esta semelhança de valores é obita pela mudança da escala original dos valores para um intervalo, podendo ser entre 0 a 1, -1 a 1 ou outros. Isso aumenta a compatibilidade dos dados com as funções de transferência (MACHADO, 2005; NETO; PELLI, 2004).

Para tanto, neste estudo foi adotada a metodologia de normalização das variáveis que altera a escala real dos valores de entrada para o intervalo entre 0 a 1 (eq. 39).

$$x_j^{norm} = \frac{x_j - x^{min}}{x^{max} - x^{min}} \tag{39}$$

- onde x_j^{norm} é a variável normalizada.
- x_j é a variável na posição j.
- x^{min} valor mínimo observado entre as variáveis.
- x^{max} valor máximo observado entre as variáveis.

Um resumo das características do conjunto de dados para o treinamento e teste pode ser vista nas tabelas 4 e 5.

Tabela 4 – Análise exploratória do conjunto de treinamendo

	Ráio	Textura	Perímetro	Área	Suavidade	Compacidade	Concavidade	Côncavo	Simetria	Dimensão Fractal
Mínimo	7,69	9,71	48,34	170,4	0,06	0,02	0	0	0,1	0,05
1° Quartil	11,71	16,39	75,27	422,3	0,09	0,06	0,03	0,02	0,16	0,06
Mediana	13,46	19,07	87,19	559,2	0,1	0,09	0,06	0,03	0,18	0,06
Médio	14,18	19,43	92,28	659,3	0,1	0,1	0,09	0,05	0,18	0,06
3° Quartil	16,08	21,79	106	798,3	0,1	0,13	0,13	0,07	0,2	0,07
Máximo	28,11	39,28	188,5	2501	0,15	0,35	0,43	0,2	0,3	0,1

Tabela 5 – Análise exploratória do conjunto de teste

	Ráio	Textura	Perímetro	Área	Suavidade	Compacidade	Concavidade	Côncavo	Simetria	Dimensão Fractal
Mínimo	7	10,38	48,79	143,5	0,05	0,02	0	0	0,1	0,05
1° Quartil	11,6	15,45	74,61	415,6	0,09	0,07	0,03	0,02	0,16	0,06
Mediana	13,11	18,34	84,59	530,4	0,1	0,09	0,05	0,03	0,18	0,06
Médio	14	18,87	91,03	641,8	0,1	0,11	0,09	0,05	0,18	0,06
3° Quartil	15,53	21,81	102,92	798,2	0,1	0,13	0,13	0,07	0,2	0,07
Máximo	27,22	32,47	182,1	2250	0,16	0,31	0,41	0,2	0,3	0,01

Para avaliar a performace dos modelos propostos nesse estudo foi utilizada a seguinte métrica. A acurácia de classificação, ou precisão total, de um classificador *f*(.) é dada por $\frac{V_P+V_N}{n}$ onde V_p são as amostras de rótulo positivo (+1) preditas como positivas, $V_N$são amostras de rótulo negativo (-1) preditas como negativas e *n* é o número total de amostras. Já a taxa de erro das classes ± 1 é dada por $\frac{F_N}{V_P+F_N}$ e $\frac{F_P}{F_P+V_N}$, respectivamente, para as classes de rótulo positivo e negativo, onde F_N são amostras de rótulo positivo preditas como negativas e F_P são amostras de rótulo negativo preditas como positivas. A relação dessas métricas para um problema de classificação binário são representadas na tabela

Tabela 6 – Matriz de confusão para classificação binária.

Classe	Predita C_+	Predita C_-	Taxa de Erro da Classe
C_+	Verdadeiros positivos	Falsos negativos	$\frac{F_N}{V_P+F_N}$
	V_P	F_N	
C_-	Falsos positivos	Verdadeiros negativos	$\frac{F_P}{F_P+V_N}$
	F_P	V_N	

Ao final das 50 simulações, são selecionadas as que apresentarem os melhores resultados em termos do erro de treino. Para tal, os valores do erro de treino são ordenados em ordem crescente e as simulações que apresentarem resultado superior ao limite de distribuição do primeiro quartil serão selecionadas. Após essa seleção, será calculado a média nesse conjunto de simulações, visando obter um resultado mais robusto do modelo, na classificação das microcalcificações mamárias.

Sobre o modelo SVMs não linear (que é um algoritmo de classificação), tem-se que escolher uma função kernel, além de parâmetros dessa função e do valor da constante de regularização *C*. Sendo assim, utilizando a validação cruzada, a função de kernel utilizada e

os valores dos parâmetros *C* e σ foram estimados através da avaliação dos resultados de simulações. A relação dos parâmetros está sumarizada na tabela 5.

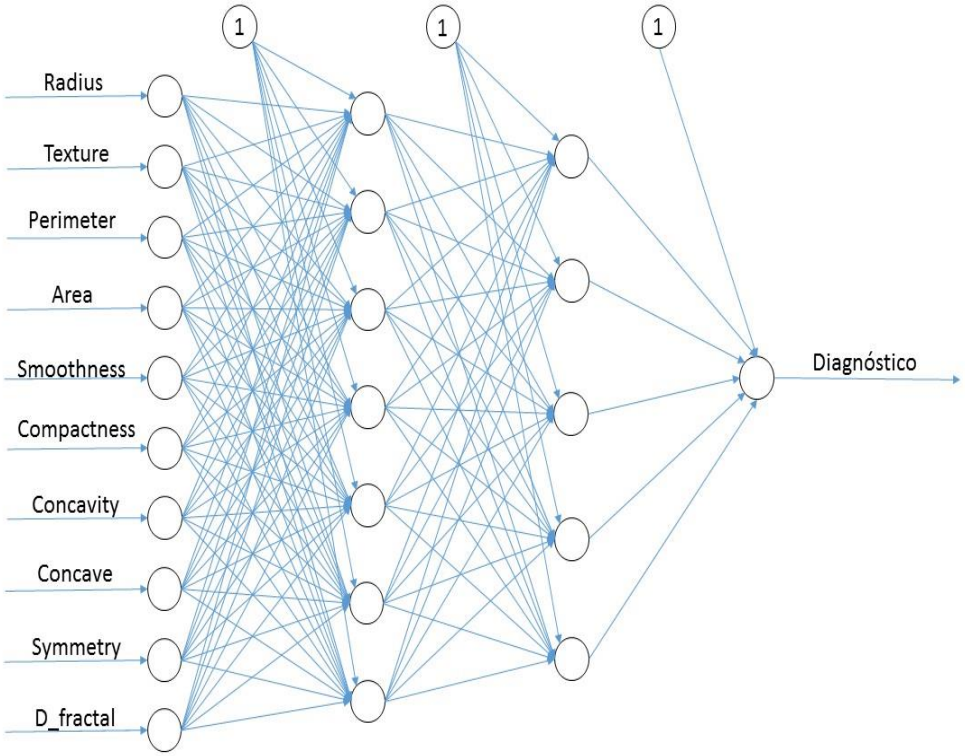
Tabela 7 – Tabela de parâmetros do modelo SVM não linear

Parâmetros	Valor
Simulações	50
Tipo de classificador	C-svc
Função de kernel	rbf
Variância da função de kernel (σ)	0,5
Parâmetro de regularização C	1
Critério de parada (tol)	0,001
Erro de validação cruzada	1

É importante ressaltar que na etapa de treinamento gera-se o modelo com os vetores de suporte que é utilizado pelo modelo proposto (SVM não linear) na etapa de testes, sendo que a etapa de treinamento desconhece totalmente as amostras contidas no conjunto de teste, sendo dessa forma conjuntos independentes. Após o modelo gerado, é aplicado no conjunto de teste com o objetivo de verificar sua eficácia na classificação das microcalcificações mamárias.

Sobre a utilização do modelo computacional de redes neurais MLP, proposto nesse trabalho foi avaliado com a incorporação de todas as variáveis. A configuração da rede utilizada esta representada na figura 19.

Figura 19– Configuração da rede MLP utilizada.



A rede proposta foi implementada usando script desenvolvido em R com a utilização da biblioteca “neuralnet. A mesma constituída de duas camadas ocultas, sendo a primeira com 7 (sete) neurônios e a segunda com 4 (quatro), com pesos inciais definidos de forma aleatória entre (0 e 1). Os dados foram normalizados utilizando a escala min-max no intervalo [0,1] e a avaliação do desempenho do modelo proposto foi mensurado através do erro médio quadrático. A função de ativação utilizada na camada oculta foi a função logística e o algoritmo de aprendizado utilizado foi o Rprop+r(backpropagation resiliente). A relação dos parâmetros está sumarizada na tabela 8.

Tabela 8 – Tabela de parâmetros do modelo Redes Neurais MLP.

Parâmetros	**Valor**
Nº máximo de épocas	100.000 iterações
Função	Logistic
Likelihood	Falsa
Nº de repetições da rede no treinamento	1
Algoritmo de treinamento	Backpropagation
Métrica usada no erro de treinamento	SSE
Nº de camadas ocultas	2
Nº de neurônios da primeira camada oculta	7
Nº de neurônios da segunda camada oculta	4
Inicialização dos pesos	Aleatoriamente com valores [0,1]
Threshold (Critério de parada)	0,01

4 RESULTADOS E DISCUSSÕES

4.1 Resultados do modelo SVM não linear

A utilização o modelo computacional SVMs não linear proposto nesse trabalho foi avaliado com a incorporação de todas as variáveis (raio, textura, perímetro, área, suavidade, compacidade, concavidade, côncavo, simetria e dimensão fractal) do conjunto da dados de pacientes portadores de microcalcificação mamárias.

O desempenho obtido no conjunto das 50 simulações realizadas pelo modelo SVMs não linear na categorização de malignidade de microcalcificação mamária, no que tange aos parâmetros do erro de treinamento, erro de validação cruzada (*leave-one-out*), número de vetores de suporte, acurácia e valor falso negativo, são apresentados na tabela 9.

Tabela 9 – Tabela de desempenho do Modelo SVM não linear

Erro_treino (%)	Erro_validação (%)	N° vetor suporte	Acurácia (%)	Valor FN (%)
2,58	4,75	120	94,37	7,55
2,11	5,55	128	92,25	14,89
3,75	5,05	113	95,07	8,00
3,28	4,95	128	92,96	15,52
3,51	5,70	113	94,37	11,54
3,51	5,25	122	96,48	0,00
2,81	4,15	126	93,66	3,57
3,98	5,95	129	97,18	1,79
4,45	6,60	122	95,77	7,41
3,04	6,05	119	95,77	8,33
3,04	5,25	124	96,48	6,12
3,75	5,85	115	95,07	7,55
3,75	5,95	119	95,77	9,80
3,51	5,55	125	94,37	6,56
3,51	6,15	118	95,07	7,55
3,75	5,85	133	97,89	3,70
3,04	4,45	125	94,37	10,34
3,51	5,60	123	97,18	8,70
3,98	5,60	127	98,59	1,96
3,51	5,35	115	92,96	15,91
3,04	4,90	126	95,07	10,71
3,51	5,55	120	94,37	5,77
3,28	5,55	121	94,37	11,54
3,51	5,80	118	96,48	6,00
3,98	4,70	131	92,25	15,00
3,28	5,45	119	94,37	12,28

3,04	5,65	117	94,37	7,69
3,98	5,75	116	92,96	12,96
3,98	4,70	125	92,96	19,61
3,98	5,40	127	95,77	7,84
3,51	5,90	115	95,77	8,33
3,04	5,20	124	94,37	14,55
3,04	4,70	121	95,07	10,00
3,04	5,15	121	91,55	21,15
3,28	4,45	123	97,18	3,92
2,81	5,25	117	93,66	10,20
3,75	4,45	124	95,77	7,14
4,22	6,15	121	96,48	5,45
3,51	5,20	128	96,48	11,63
2,81	4,80	115	95,07	10,00
2,58	4,20	125	91,55	10,17
3,28	5,55	121	95,07	6,12
2,58	4,80	117	92,96	9,80
2,11	3,25	118	90,85	16,36
4,22	7,25	124	97,18	5,77
4,45	6,55	126	96,48	10,42
3,04	5,50	127	96,48	7,27
4,22	6,30	127	98,59	3,85
3,51	4,60	125	94,37	9,62
3,75	5,40	129	96,48	2,08

A simulação que apresentou melhor desempenho no conjunto de validação em termos da acurácia obteve valor de 98,59% na caracterização referente a malignidade de microcalcificação mamária, com número de vetores de suporte igual a 127. No que tange aos valores de Falsos Negativos, classificação referente a não malignidade da microcalcificação mamária sendo que há o melhor valor obtido foi 1,96. Na tabela 10 apresentamos os resultados médio obtidos nas 50 simulações do modelo.

Tabela 10 – Tabela do desempenho médio das 50 simulações do Modelo SVM não linear

	Erro treino	**Erro validação**	**n° de vetor**	**Acuracia (%)**	**Valor FN (%)**
Media	0,0341	0,0535	122	95,00	9,00
Mediana	0,0351	0,0543	123	95,07	8,00
Desvio Padrao	0,0055	0,0071	4,89	0,02	4,00
Pior Simulação	0,0258	0,0530	113	90,85	5,63
Melhor Simulação	0,0398	0,0560	127	98,59	1,96

4.2 Resultados do modelo Redes Neurais MLP

A utilização o modelo computacional Redes Neurais MLP proposto nesse trabalho também foi avaliado com a incorporação de todas as variáveis (raio, textura, perímetro, área, suavidade, compacidade, concavidade, côncavo, simetria e dimensão fractal) do conjunto da dados de pacientes portadores de microcalcificação mamárias.

O desempenho obtido no conjunto das 50 simulações realizadas pelo modelo de rede neural MLP na categorização de malignidade de microcalcificação mamária, no que tange aos parâmetros do erro de treinamento, número de épocas, *thereshold*, acurácia e valor falso negativo, são apresentados na tabela 11.

Tabela 11 – Tabela de desempenho do modelo de rede neural MLP.

Erro treino (%)	N° de épocas	Critério de parada	Acurácia (%)	Valor FN (%)
2,57	8666	0,00993	91,55	3,70
49,80	4513	0,00700	92,25	7,84
112,26	7373	0,00953	93,66	10,87
101,96	7032	0,00916	94,37	9,30
54,11	18278	0,00948	90,85	22,22
50,63	29934	0,00934	90,85	13,21
2,85	16995	0,00998	93,66	5,26
100,81	10527	0,00930	92,96	16,00
50,47	6626	0,00898	95,07	7,32
2,78	20910	0,00879	91,55	14,89
96,26	11702	0,00996	93,66	3,92
7,29	15818	0,00977	93,66	9,26
5,10	24453	0,00980	95,07	2,17
95,80	4930	0,00702	92,25	12,50
104,28	5505	0,00917	96,48	4,26
64,20	64570	0,00991	90,14	9,80
2,57	22326	0,00985	94,37	7,55
99,95	21945	0,00975	89,44	16,33
1,72	21175	0,00961	92,96	5,26
62,58	91173	0,00988	94,37	3,51
49,25	12844	0,00942	92,25	9,68
54,81	6958	0,00849	89,44	10,17
6,05	8759	0,00956	90,14	10,71
102,64	9188	0,00944	90,14	10,00
3,84	16585	0,00962	95,07	3,77
143,95	4198	0,00800	95,07	3,77
2,16	9877	0,00979	94,37	2,00
51,84	9987	0,00977	94,37	10,71
0,95	15114	0,00959	92,25	10,17
0,87	10524	0,00948	90,14	9,26
50,79	9364	0,00965	92,25	12,77

102,72	28125	0,00938	97,18	4,26
3,37	7771	0,00991	91,55	5,66
49,78	15894	0,00966	89,44	15,22
53,09	16492	0,00759	92,25	6,00
143,40	8662	0,00962	92,25	6,12
152,45	8370	0,00917	90,14	12,73
2,39	7638	0,00868	89,44	12,73
49,00	24965	0,00927	90,14	15,91
184,92	26034	0,00960	95,07	7,27
50,76	10055	0,00978	92,96	13,21
1,56	15862	0,00918	92,25	12,00
157,14	6318	0,00967	95,77	10,20
99,15	5539	0,00862	92,25	8,93
2,25	9532	0,00947	93,66	6,25
6,37	6822	0,00931	95,07	6,12
78,78	10865	0,00979	92,25	10,53
50,08	13702	0,00976	90,14	14,81
2,19	24029	0,00902	91,55	7,69
65,91	41237	0,00993	90,85	15,52

A melhor simulação apresentou desempenho no conjunto de teste em termos da acurácia obteve valor de 94,37% e a pior simulação apresentou 90,85% na caracterização referente a malignidade de microcalcificação mamária. No que tange aos valores de Falsos Negativos, classificação referente a não malignidade da microcalcificação mamária sendo que há o melhor valor obtido foi 2,00 na melhor simulação e a média obtida no conjunto das 50 simulações foi 9,38%, conforme apresentado na tabela 12.

Tabela 12 – Tabela de desempenho médio das 50 simulações do Modelo Redes Neurais MLP

	SSE Treinamento	**Thereshold**	**N°. épocas**	**Acurácia (%)**	**Valor FN (%)**
Media	0,5580	0,0093	16315	92,58	9,38
Mediana	0,5069	0,0095	10696	92,25	9,49
Desvio Padrão	0,5028	0,0006	***	0,02	0,04
Pior Simulação	0,5411	0,0094	18278	90,85	22,22
Melhor Simulação	0,0216	0,0098	9877	94,37	2,00

O conjunto de pesos obtidos na melhor das simulações realizadas está representado tabelas 13, 14 e 15.

Tabela 13 – Tabela dos pesos da primeira camada oculta

Neurônio 1	Neurônio 2	Neurônio 3	Neurônio 4	Neurônio 5	Neurônio 6	Neurônio 7
-0,219	-1,903	-1,745	-8,181	0,086	10,397	-3,907
1,749	3,354	0,279	-3,263	-0,486	-1,364	-0,247
-3,487	0,558	-17,075	18,461	2,839	-5,408	11,44
0,018	-0,419	-2,139	-0,377	0,735	0,744	-1,032
-0,536	-1,761	6,872	27,495	-7,451	-10,408	18,228
0,471	0,463	50,464	-18,293	-0,949	-7,7	-16,458
0,2	-0,387	-13,742	2,571	0,809	2,9	-4,226
1,785	-0,024	-29,981	18,77	1,59	-7,05	116,177
-2,436	5,661	-11,455	1,063	-4,819	-4,251	-73,728
-0,579	-0,946	5,773	4,819	1,575	-1,631	1,741
-0,018	-1,13	-3,742	-2,029	-0,391	1,842	5,176

Tabela 14 – Tabela dos pesos da segunda camada oculta

Neurônio 1	Neurônio 2	Neurônio 3	Neurônio 4
-0,455	-0,494	0,156	0,547
-2,421	-7,501	-23,925	24,382
0,697	2,953	-0,71	-11,206
-1,791	5,128	15,628	-8,77
-2,305	3,719	14,401	-11,423
-0,622	-9,394	-13,878	13,106
2,219	-5,241	-23,297	17,042
-7,912	64,11	11,916	-7,386

Tabela 15 – Tabela dos pesos da ultima camada

Neurônio 1
-1,498
1,228
0,298
2,198
1,189

O conjunto de pesos obtidos na pior das simulações realizadas está representado tabelas 16, 17 e 18.

Tabela 16 – Tabela dos pesos da primeira camada oculta

Neurônio 1	Neurônio 2	Neurônio 3	Neurônio 4	Neurônio 5	Neurônio 6	Neurônio 7
2,608	1,452	2,031	-3,901	-12,427	0,233	2,340
-0,189	-0,017	1,030	4,025	1,367	-0,708	-1,415
-13,287	-0,492	-5,871	2,581	-0,974	-0,747	-5,620
-5,576	0,510	4,571	2,241	-1,822	1,059	-0,078
8,369	-5,904	-11,864	-10,382	23,244	-2,105	-7,581
14,556	0,815	-3,272	-14,696	27,447	-2,828	7,037
-18,141	-2,496	0,757	6,846	-17,701	1,128	2,007
42,803	-3,536	-5,651	8,312	-15,978	2,804	-24,243
17,939	-1,368	-8,862	34,509	50,287	-0,503	14,939
1,394	0,854	1,054	2,343	-13,587	-6,630	-2,363
-19,063	0,305	7,910	-7,599	10,261	1,354	-5,826

Tabela 17 – Tabela dos pesos da segunda camada oculta

Neurônio 1	Neurônio 2	Neurônio 3	Neurônio 4
0,142	-3,626	-1,123	12,302
6,510	-0,347	-14,742	-4,750
-26,682	32,269	10,247	51,740
95,752	-79,862	77,558	59,599
-12,491	12,217	12,233	-14,672
-3,192	-20,768	-15,683	-48,305
26,079	-43,457	51,342	16,253
2,215	16,475	26,443	26,029

Tabela 18 – Tabela dos pesos da última camada

Neurônio 1
1
1,312
1,275
-1,808
-0,499

Tabela 19 – Comparação dos resultados médios obtidos pelos modelos SVM-não linear e RN-MLP na categorização de malignidade no conjunto das 50 simulações.

	SMV-Não Linear		**RN-MLP**	
	Acurácia (%)	**Valor FN (%)**	**Acurácia (%)**	**Valor FN (%)**
Média ± desvio padrão	95,00 ± 0,02	9,00 ± 0,04	92,58 ± 0,02	9,38 ± 0,04
Mediana	95,07	8,00	92,25	9,48
Pior Simulação	90,85	5,63	90,85	22,22
Melhor Simulação	98,58	1,96	94,37	2,00

5 CONCLUSÕES

Para o Brasil, estimam-se 59.700 casos novos de câncer de mama, para cada ano do biênio 2018-2019, com um risco estimado de 56,33 casos a cada 100 mil mulheres (INCA 2018). Desta forma, pensando na elevada taxa de incidência e mortes causadas pelo câncer de mama, atualmente, no Brasil e no mundo, justifica o desenvolvimento de pesquisas científicas voltadas para estratégias de auxílio na detecção precoce da doença, fator determinante para o sucesso do tratamento. Dentro deste contexto, o presente trabalho teve como objetivo principal aplicar e analisar o modelo utilizando as redes SVM e Redes Neurais MLP na predição de neoplasias mamárias.

De acordo com a análise dos resultados foi possível evidenciar o desempenho promissor do modelo estrututurado em SVM não linear. Visto que obteve na sua melhor simulção uma acurácia acima de 98% e no que tange aos valores de falsos negativos, esta simulação obteve valor de inferior a 2% (1,96%).

O modelo de rede neural MLP, na sua melhor simulação obteve uma acurácia superior a 94%, com valor de falso negativo de 2% indicando uma precisão do modelo de 98% em termos de sensibilidade.

Em termos da acurácia média obtida no conjunto das 50 simulações realizadas, verificamos que o modelo estruturado em SVM-não linear apresentou desempenho superior ao modelo utilizando redes neurais MLP. Calculando o p-valor a nível de significância de 95% (p-valor $< 0,05$), pode-se verificar que há diferença estatística significante entre os resultados referente à acurácia entre os modelos aplicados. Indicando assim que o modelo SVM tem melhor performace. Entretanto, calculando o p-valor a mesmo nível de significância pode-se observar que não existe diferença estatística significante entre os resultados referentes a determinação dos valores falso negativo entre os modelos, isto é, apesar do modelo SVM apresentar um valor médio menor essa diferença não é estatisticamente relevante.

O modelo SVM-não linear, nas 50 (cinquenta) simulações realizadas no conjunto de teste obteve simulação com valor de 100% de sensibilidade ou seja 0% na determinação de valores falsos negativos. Fato que não foi verificado com o modelo de rede neural MLP, que obteve valor máximo para sensibilidade de 98%.

É importante ressalatar que a acurácia obtida pelo modelo na classificação de microcalcificação mamária, encontra-se próxima aos valores obtidos na literatura com a utilização de técnicas baseada em inteligência computacional (SHANTHV, BHASKARAN, 2013).

Apesar dos bons resultados obtidos com a aplicação do modelo das Redes SVM e Redes Neurais MLP na classificação de microcalcificações mamárias com dados de mamografias, há necessidade de aumentar o número de amostras.

Entretanto, apesar dos modelos aplicados não terem alcançado em todos casos 100% de precisão em suas predições, o mesmo obteve resultados satisfatórios. Por isso, pode-se concluir que as Redes SVM e Redes Neurais MLP são ferramentas úteis na previsão de microcalcificações mamárias.

5.1 TRABALHOS FUTUROS

Como trabalhos futuros, pretende-se verificar: A aplicação da técnica de fuzzificação na codificação dos dados; Aplicação de outras ferramentas baseadas em inteligência computacional de sorte a comparar os resultados com o modelo proposto.

6 REFERÊNCIAS BIBLIOGRÁFICAS

AGGARWAL, R.; SONG, Y. Artificial Neural Networks in Power Systems. Power Engineering Journal. v.12, p. 279-287, 1998.

ALBARAKATI, N.; KECMAN, V. Fast Neural Network Algorithm for Solving Classification Tasks. IEEE-Southeastcon Proceedings. p.1 - 8, 2013.

ALMEIDA F. F. M. Relatório técnico: Support Vector Machine. Universidade Federal de Campina Grande, Centro de Ciência e Tecnologia. 2007.

AYER, T., ALAGOZ, O. CHATWAL, J., SHAVLIK, J. W. KAHN, E. J., BURNSIDE, E. S. Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. Vol. Cancer. 2010.

ASADUZZAMAN, MD.; AHMED, S. U.; KHAN, F. E.; SHAHJAHAN, MD.; MURASE, K. Making Use of Damped Noisy Gradient in Training Neural Network. IEEE - Neural Networks (IJCNN), The 2010 International Joint Conference on. p. 1- 5, 2010.

BARCELLOS, J. C. H. Algoritmos Genéticos Adaptativos: Um estudo comparativo. 143 p. Dissertação (Mestre em Engenharia). Escola Politécnica da Universidade de São Paulo, 2000. São Paulo - SP.

BECKAMANN, M. Algoritmos Genéticos como Estratégia de Pré-processamento em Conjuntos de Dados Desbalanceados. 112 p. Dissertação (Programa de Pós-graduação em Engenharia Civil - Mestre em Engenharia Civil) - Universidade Federal do Rio de Janeiro – UFRJ, 2010. Rio de Janeiro-RJ.

BISHOP, C. M. Neural Networks for Pattern Recognition. Oxford University Press. 1997.

BOUGHRARA, H.; CHTOUROU, M.; AMAR, C. B. MLP Neural Network Based Face Recognition System Using Constructive Training Algorithm. IEEE - Multimedia Computing and Systems (ICMCS), International Conference on. p.233-238, 2012.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. Redes Neurais Artificiais - Teoria e Aplicações. Ed. 2, p. 225, 2012.

BRAZ JÚNIOR, G.; SILVA, E. C.; PAIVA, A. C.; SILVA, A. C.; Gattass, M. Breast Tissues Mammograms Images Classification using Moran s Index, Geary s Coefficient and SVM . In: International Conference on Neural Information Processing, 2007, Kitakyushu. Lecture Notes in Computer Science, LNCS, 2007.

CAMPOS, L. F. A., SILVA, A. C., BARROS, A. K. Independent Component Analysis and Neural Networks Applied for Classification of Malignant, Benign and Normal Tissue in Digital Mammography. Methods of Information in Medicine, 2007.

CHANG, C. E LIN, C. Libsvm – A Library for Support Vector Machines, 2003.

COSTA, D. D., BARROS, A. K., E SILVA, A. C. Independent Component Analysis in Breast Tissues Mammograms Images Classification using LDA and SVM . Information Technology Applications in Biomedicine - ITAB2007 - Tokyo. Conference on 6th International Special Topic. 2007.
CRISTIANINI, N. E., SHAWE-TAYLOR, J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press. 2000.

C. J. C. Burges. A tutorial on support vector machines for pattern recognition. Know- ledge Discovery and Data Mining, 2(2):1–43, 1998.

CASTRO, M. C. F. Predição Não - Linear de Séries Temporais Usando Redes Neurais RBF por Decomposição em Componentes Principais. 192 p. Tese (Doutor em Engenharia Elétrica). Universidade Estadual de Campinas - UNICAMP, 2001. Campinas -SP.

CAUDILL, M. Neural Network Primer, Miller Freeman Publications, 1990.

CERQUEIRA, E. O.; ANDRADE, J. C.; POPPI, R. J. Redes Neurais e suas Aplicações em Calibração Multivariada. Quim Nova. vol. 24, n. 6. p.864-873, 2001.

COVER, T. M. Geometrical and Statistical Properties of Systems of Linear Inequalities With Applications in Pattern Recognition. IEEE - Electronic Computers, Transactions on, v. 14, n. 3. p. 326-334, 1965.

DDSM. The Digital Database for Screening Mammography , Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer, in Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., Medical Physics Publishing. 2001. Disponível em http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

FONSECA, J. S.; MARTINS, G. A.; TOLEDO, G. L. Estatística Aplicada. Ed. 2, p. 267, 2012.

FENTON, J. J., TAPLIN, S. H., CARNEY, P. A., ABRAHAM, L., SICKLES, E. A., D'ORSI, C., BERNS, E. A., CUTTER, G., HENDRICK, R. E., BARLOW, W. E. , ELMORE, J. G. Influence of Computer-Aided Detection on Performance of Screening Mammography . Breast Diseases: A Year Book Quarterly. 2007.

FREER, T. W. E ULISSEY, M. J. Screening Mammography with Computer-Aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center Radiology. 2001.

GIGER, M. L. Computer-aided diagnosis of breast lesions in medical images. Computing in Science & Engineering. 2000.

GODA, Rúpila Rami da Silva. Inteligência Computacional Aplicada em Microcalcificações Mamárias. Dissertação (Mestrado em Modelagem Matemática e Computacional). Instituto de

Ciências Exatas, Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro, Seropédica, 2016.

GOLDBERG, D. E. Genetic Algorithms in Search, Optimization and Machine Learning. New York: Addison-wesley. Ed.1, p. 432, 1989.

GOLDBERD, D. Genetic Algorithms in Search, Optimization, and Machine Learning. EUA: Addison-Wesley. 1989.

GONÇALVES, R. M.; COELHO, L. S.; KRUEGER, C. P.; HECK, B. Modelagem Preditiva de Linha de Costa Utilizando Redes Neurais Artificiais. Bol. Ciênc. Geod., sec. Artigos, Curitiba, v. 16, n. 3, p.420-444, 2010.

GONZALEZ, R. C., WOODS, R. E. Digital Image Processing. 3rd Edition. Prentice Hal Gunn S. R. (1998). Support Vector Machine for Classification and Regression. Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science. 2007.

GUARNIERI, R. A. Emprego de Redes Neurais Artificiais e Regressão Linear Múltipla no Refinamento das Previsões de Radiação Solar do Modelo ETA. 171 p. Dissertação (Programa de Pós-graduação em Meteorologia). Instituto Nacional de Pesquisas Espaciais-INPE, 2006. São José dos Campos-SP.

HAYKIN, S. Redes Neurais Princípios e Pratica, Ed. 2, p. 902, 2001.

HAGAN, M. T.; MENHAJ, M. B. Training Feedforward Networks with the Marquardt Algorithm. IEEE Transactions on Neural Networks, v. 5, p. 989-993, 1994.

HEATON, J. Programming Neural Networks with Encog 2 in Java. p. 481, 2010.

HUANG, S. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics and Proteomics, vol 15, p. 41-51, 2018.

NASSIF, D., PAGE, M., AYVACI, SHAVLIK, J., BURNISIDE, E. S. Uncovering age-specific invasive and dcis breast cancer rules using inductive logic programming. Proceedings of ACM International Health Informatics Symposium (IHI 2010), ACM Digital Library, 2010.

HANSE, D. Fundamentos da Patologia, Câncer de Mama, pág. 553 - 557. Rio de Janeiro Guanabara Koogan, 2007.

HEATH, M., BOWTER, K., KOPNAS, D., MOORE, R., KEGELMEYER, W.P. Current Status of the Digital Database for Screening Mammography. Digital Mammography. 2008. Disponível em: http://www.icadmed.com/.

INSTITUTO NACIONAL DO CÂNCER. Estimativas 2018. Incidência de câncer no Brasil. Brasil, 2018.

CASTRO, J. L., FLORES-HIDALGO, L. D., MANTAS, C. J., PUCHE, J. M. Extraction of fuzzy rules from support vector machines. Fuzzy Sets and Systems archive, 2007.

JAIN, A. K., DUIN, R. P. W., MAO, J. Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis e Machine Intelligence. 2000.

KARATZOGLOU, A., SMOLA, A., HORNIK, K., & KARATZOGLOU, M. A. Package 'kernlab'.R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.

KOPANS, DB. Imagem da mama. 2ª ed. Rio de Janeiro: Revinter Medsi; 2000.

LORENA A. C. e CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. 2007.

LIPPMANN, R. P. An Introduction to Computing with Neural Nets. IEEE ASSP MAGAZINE. p.4- 22, 1987.

MARTINS, L. O., 2007. Detecção de Massas em Imagens Mamográficas Através do Algoritmo Growing Neural Gas e da Função K De Ripley. Dissertação de mestrado. Universidade Federal do Maranhão, Departamento de Engenharia de Eletricidade, Programa de Pós-Graduação em Engenharia de Eletricidade. São Luís, 2007.

MINISTERIO DA SAÚDE. Política Nacional de Câncer de Mama. Brasília: Ministério da Saúde; 2008.

MOAYEDI, F., BOOSTANI, R., AZIMIFAR, Z., KATEBI, S. A Support Vector Based Fuzzy 54 Neural Network Approach for Mass Classification in 81 Mammography. Digital Signal Processing, 2007. 15th International Conference.

NASCIMENTO, C. D. L. Breast tumor classificationin ultrasound images using support vector machines and neural networks. In Research on Biomedical Engineering.2016

OSTA, H., QAHWAJI, R., IPSON, S. Wavelet-based Feature Extraction and Classification for Mammogram Images using RBF and SVM. In Proceedings of International Conference on Visualization, Imaging, and Image Processing (VIIP). Palma de Mallorca, Spain. 2008.

PADWAL, M. Elements of breasting imaging basics. 2007. Disponivel em: http://www.gehealthcare.com/usen/ultrasound/education/products/cme_breast.html.

HERBRICH, R. Learning Kernel Classifiers: Theory and Algorithms. MIT Press, 2001.

ROCHA, M.; CORTEZ, P.; NEVES, J. M. Análise Inteligente de Dados Algoritmo e Implementação em Java. p. 177, 2008.

WITTEKIND, C., NEID, M. "Cancer invasion and metatisis", Oncology, v.69, n. suppl 1, pp. 14-16, Sep 19 2005.

WOODS, R., OLIPHANT, L., SHINKI, K., Shinki., PAGE, D., SHAVLIK, J., BURNSIDE, E. Validation of results from knowledge discovery: Mass density as a predictor of breast cancer. J Digit Imaging, 2009.

WOODS, R., BURNSIDE, E. The mammographic density of a mass is a significant predictor of breast cancer. Radiology, USA, 2010.

WORLD HEALTH ORGANIZATION, INTERNATIONAL AGENCY FOR RESEARCH ON CANCER, *World Cancer Report*. Geneva, Switzerland, WHO Library, 2014.

SAMPAT, M. P., MARKEY, M. K., BOVIK, A. C. Computer-Aided Detection and Diagnosis in Mammography. Handbook of Image and Vídeo Processing. 2005.

SAMPAIO, W. B. Detecção de massas em imagens mamográficas usando redes neurais celulares, funções geoestatísticas e máquinas de vetores de suporte/ Wener Borges de Sampaio. – São Luís, 2009

SARITAS, I. Prediction of Breast Cancer Using Artificial Neural Networks. In: Journal of Medical Systems, vol 36. pp 2901-2907. 2012

SHANTHU, S., BHASKARAN, A.V. A Novel Approach for detecting and Classifying Breast Cancer. In: International Journal of Intelligent Information Technologies, 2013.

SOUSA, J. R., SILVA, A. C., PAIVA, A. C. Lung Structures Classification Using 3D Geometric Measurements and SVM . In: 12th Iberoamerican Congress on Pattern Recognition, 2007, Valparaiso. Lecture Notes Computer Science - LNCS. Berlin: Springer-Verla.

VAPNIK, V. N., CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications, 1971.

VAPNIK, V. N. Statistical Learning Theory. John Wiley and Sons, 1998.

ZURRIDA, S. et al. A dissecção axilar no carcinoma de mama. In: VERONESI, U. et al. Mastologia Oncológica. Rio de Janeiro: MEDSI, 2002.