Amanda Lays Rodrigues da Silva

Seleção de atributos para apoio ao diagnóstico do câncer de mama usando imagens termográficas, algoritmos genéticos e otimização por enxame de partículas

Dissertação de Mestrado

Recife

2019

Amanda Lays Rodrigues da Silva

Seleção de atributos para apoio ao diagnóstico do câncer de mama usando imagens termográficas, algoritmos genéticos e otimização por enxame de partículas

Trabalho submetido ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestra em Engenharia Biomédica.

Orientador: Prof. Wellington Pinheiro dos Santos

Recife

2019

Catalogação na fonte
Bibliotecária Margareth Malta, CRB-4 / 1198

S586s Silva, Amanda Lays Rodrigues da.
Seleção de atributos para apoio ao diagnóstico do câncer de mama usando imagens termográficas, algoritmos genéticos e otimização por enxame de partículas / Amanda Lays Rodrigues da Silva. - 2019.
62 folhas, figs., gráfs., tabs.

Orientador: Prof. Dr. Wellington Pinheiro dos Santos.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Biomédica, 2019.
Inclui Referências e Apêndice.

1. Engenharia Biomédica. 2. Câncer de mama. 3. Termografia. 4. Seleção de Atributos. 5. Algoritmos genéticos. 6. Otimização por enxame de partículas. 7. Máquina de vetor de suporte. I. Santos, Wellington Pinheiro dos. (Orientador). III. Título.

UFPE

610.28 CDD (22. ed.) BCTG/2020-32

Amanda Lays Rodrigues da Silva

Seleção de atributos para apoio ao diagnóstico do câncer de mama usando imagens termográficas, algoritmos genéticos e otimização por enxame de partículas

Trabalho submetido ao Programa de Pós-Graduação em Engenharia Biomédica do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestra em Engenharia Biomédica.

Aprovada em 21 de fevereiro de 2019:

Prof. Wellington Pinheiro dos Santos
Orientador

Prof. Ricardo Yara
Examinador Interno

Profa. Rita de Cássia Fernandes de Lima
Examinadora Externa

Recife
2019

Agradecimentos

Agradeço primeiramente a Deus, que durante toda minha vida esteve ao meu lado me ajudando e fortalecendo nos momentos difíceis.

Agradeço a minha mãe e ao meu padrasto, Marta e Júnior, que com muita paciência e amor me educaram e me proporcionaram momentos como este, sempre evidenciando que a educação e o conhecimento são a base para o sucesso profissional.

Agradeço ao meu orientador e amigo, professor Wellington, que depositou confiança em mim, me incentivou e me direcionou de como eu deveria agir para desenvolver um bom trabalho.

Agradeço aos meus amigos, por todo o apoio, incentivo que me deram forças para a conclusão deste projeto. Em especial a Maíra, Washington e Juliana por toda ajuda e paciência nessa trajetória.

Agradeço aos companheiros da pesquisa de termografia aplicada ao câncer de mama, pelos conhecimentos repassados e por todo o companheirismo durante o projeto. Agradeço a CAPES pelo apoio financeiro para execução deste trabalho.

Resumo

A incidência de câncer de mama aumenta a cada ano. A detecção precoce da doença é fundamental já que quanto mais cedo a doença é descoberta melhores são os tratamentos e as chances de cura. Atualmente, a mamografia é o padrão ouro para o diagnóstico do câncer de mama, porém este exame apresenta algumas limitações. A termografia infravermelha é uma técnica que vem sendo bastante estudada devido aos seus benefícios. Os sistemas de classificação de tumores são detalhados e complexos e de difícil utilização pelos patologistas. Portanto, a combinação de profissionais especializados e métodos de análise digital de imagens de termografias de mama pode contribuir para a melhoria do diagnóstico. A partir disso, áreas computacionais têm se dedicado à pesquisa e à proposta de métodos para tratar esses dados. A seleção de atributos desempenha uma tarefa fundamental nesse processo, pois representa um problema de fundamental importância em aprendizado de máquina. Uma das principais áreas da Inteligência Computacional é a Computação Evolucionária (CE), que se fundamenta em estratégias para resolução de problemas baseando-se em métodos evolutivos oriundos da Teoria da Evolução de Darwin, tais como os mecanismos de seleção natural, cruzamento e mutações, além do comportamento adaptativo. Neste trabalho foi proposta a seleção de atributos em imagens termográficas com lesões mamárias utilizando os Algoritmos Genéticos (AG) e a Otimização por Enxame de Partículas (PSO). O principal objetivo dessa pesquisa foi analisar principalmente as etapas de seleção de atributos e de classificação, as quais são essenciais para a obtenção de um sistema capaz de interpretar as informações de entrada e generalizar a tomada de decisão. Para avaliar o desempenho dos subconjuntos selecionados foram usados diversos classificadores, no qual o Máquina de Vetor de Suporte foi mais efetivo. Foi possível uma redução de 169 atributos com acurácia de 91,115% para 57 atributos com acurácia de 87,082% utilizando AG. Com o algoritmo PSO foi encontrado um subconjunto de 60 atributos e uma acurácia de 86,157%. Os resultados mostraram que a nossa abordagem foi positiva, sendo evidenciada por uma significativa redução na quantidade de atributos sem diminuição considerável na acurácia em relação a classificação com todos os atributos.

Palavras-chave: Câncer de mama. Termografia. Seleção de Atributos. Algoritmos genéticos. Otimização por enxame de partículas. Máquina de vetor de suporte.

Abstract

The incidence of breast cancer increases each year. Early detection of the disease is critical since the sooner the disease is discovered the better the treatments and the chances of a cure. Currently, mammography is the gold standard for the diagnosis of breast cancer, but this test has some limitations. Infrared thermography is a technique that has been widely studied due to its benefits. Tumor classification systems are detailed, complex, and difficult for pathologists to use. Therefore, the combination of specialized professionals and methods of digital analysis of breast thermography images can contribute to the improvement of the diagnosis. From this, computational areas have been dedicated to research and the proposal of methods to treat this data. Attribute selection plays a fundamental role in this process, as it represents a fundamentally important problem in machine learning. One of the main areas of Computational Intelligence is Evolutionary Computing (EC), which is based on problem solving strategies based on evolutionary methods derived from Darwin's Theory of Evolution, such as the mechanisms of natural selection, crossing and mutations, beyond adaptive behavior. In this work, the selection of attributes in thermographic images with breast lesions using Genetic Algorithms (AG) and Particle Swarm Optimization (PSO) was proposed. The main objective of this research was to analyze mainly the stages of attribute selection and classification, which are essential for obtaining a system capable of interpreting the input information and generalizing decision making. To evaluate the performance of the selected subsets, several classifiers were used, in which the Support Vector Machine was more effective. It was possible to reduce 169 attributes with 91,115% accuracy to 57 attributes with 87,082% accuracy using GA. With the PSO algorithm, a subset of 60 attributes was found and an accuracy of 86.157%. The results showed that our approach was positive, being evidenced by a significant reduction in the number of attributes without a considerable decrease in accuracy in relation to the classification with all attributes.

Keywords: Breast cancer. Thermography. Attribute Selection. Genetic algorithms. Distance swarm optimization. Support vector machine.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

AG	Algoritmos Genéticos
AM	Aprendizado de Máquina
CE	Computação Evolutiva
IC	Inteligência Computacional
PSO	Otimização por Enxame de Partículas
FD	Diversidade Funcional
TIR	Termografia Infravermelha
SVM	Máquina de Vetor de Suporte
MLP	Multilayer Perceptron

Sumário

1 Introdução

No século XIX, a medicina, assim como outras áreas do conhecimento, era baseada principalmente em um saber empírico, decorrente da experiência acumulada ao longo das várias gerações. Tinha uma eficiência reduzida em relação ao seu objetivo principal, o tratamento das doenças, limitando-se, na maioria das situações, a cuidados paliativos (KOOP, 2006). No começo do século XX, a área médica começou a tratar e prevenir doenças a um ritmo crescente, conquistando, assim, um papel fundamental na vida dos indivíduos e na sociedade. Uma das consequências mais nítidas dessa evolução é o aumento da expectativa de vida. Pode-se dizer que esta longevidade é consequência da utilização de fármacos e dispositivos médicos, que por sua vez são decorrentes de um diálogo contínuo entre a medicina, de um lado, e a ciência e a engenharia, de outro (SANTOS, 2018).

A partir desta evolução científica e tecnológica é possível antecipar uma nova era na prestação de cuidados à saúde. Essa abordagem permite detectar o aparecimento da doença em estágios iniciais, antecipar a sua progressão e aumentar a eficiência do tratamento. Por exemplo, às patologias oncológicas como melanoma ou leucemia já são oferecidos “diagnósticos moleculares”, que permitem aos clínicos efetuarem tratamento à medida que melhoram as chances de sobrevivência. Para que este paradigma tenha sucesso, é indispensável a continuação e o aprofundamento da comunicação entre a medicina e diversas áreas científicas e de engenharia (SANTOS, 2018; GAMBINO et al., 2019; VASCONCELOS; SANTOS; LIMA, 2018; CORDEIRO; BEZERRA; SANTOS, 2017; CORDEIRO et al., 2012; CORDEIRO; SANTOS; SILVA-FILHO, 2016; CORDEIRO; SANTOS; SILVA-FILHO, 2017; CORDEIRO; SANTOS; SILVA-FILHO, 2016; CORDEIRO; SANTOS; SILVA-FILHO, 2013a; CORDEIRO; SANTOS; SILVA-FILHO, 2013b; LIMA et al., 2015; LIMA; SILVA-FILHO; SANTOS, 2014).

A engenharia vem dando uma significativa contribuição nessa evolução, servindo como base de estudo para a elaboração de novas técnicas medicinais. Atualmente, existem várias áreas de inserção da engenharia na medicina, pode-se mencionar: o projeto de próteses; a biotecnologia, com o desenvolvimento da engenharia genética; a aplicação da nanotecnologia na confecção de tecidos sintéticos não rejeitados pelo organismo; o projeto de novas técnicas e equipamentos ligados ao desenvolvimento de formas de diagnósticos; o processamento de imagens médicas; entre outros (VILA-NOVA, 2017; SANTOS et al., 2008; SANTOS et al., 2009; SANTOS; ASSIS; SOUZA, 2009; COMMOWICK et al., 2018; ANDRADE et al., 2020; SILVA-JÚNIOR et al., 2020; SILVA-JÚNIOR et al., 2019).

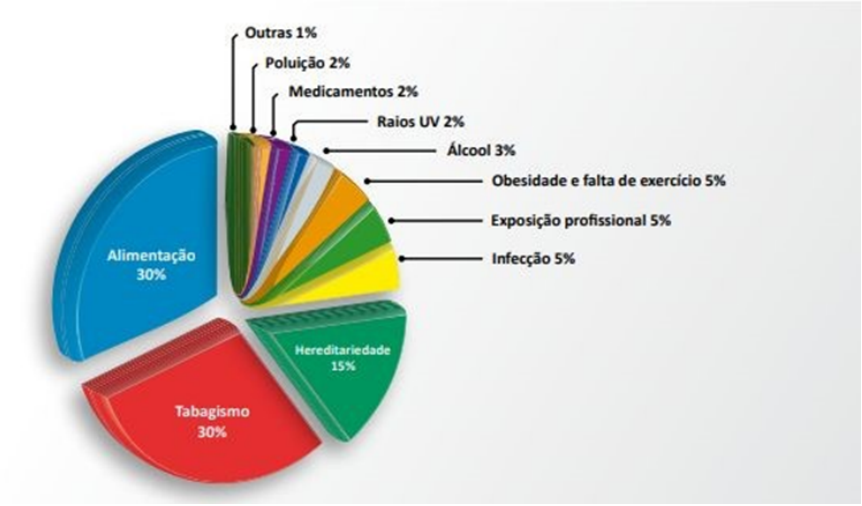
A aplicação de conhecimentos científicos e tecnológicos na medicina é fortemente fundamentada em métodos matemáticos, estatísticos e computacionais, objetivando a

redução de imprecisões e incertezas, validando o procedimento utilizado. Uma área da medicina que vem avançando com a utilização dessas técnicas de diagnóstico é a detecção do câncer de mama (VILA-NOVA, 2017; BARBOSA et al., 2020; SILVA et al., 2020; LIMA; FILHO; SANTOS, 2020; SANTANA et al., 2020; PEREIRA et al., 2020b; PEREIRA et al., 2020c; PEREIRA et al., 2020a; OLIVEIRA et al., 2019; SOUZA et al., 2019; AZEVEDO et al., 2015b; AZEVEDO et al., 2015a; LIMA et al., 2015; LIMA; SILVA-FILHO; SANTOS, 2014; CORDEIRO; BEZERRA; SANTOS, 2017; CORDEIRO et al., 2012; CORDEIRO; SANTOS; SILVA-FILHO, 2016; CORDEIRO; SANTOS; SILVA-FILHO, 2017; CORDEIRO; SANTOS; SILVA-FILHO, 2016; CORDEIRO; SANTOS; SILVA-FILHO, 2013a; CORDEIRO; SANTOS; SILVA-FILHO, 2013b).

O câncer é uma expressão genérica para um amplo conjunto de doenças que podem acometer qualquer parte do corpo. Outros termos utilizados são tumores malignos e neoplasias. Uma particularidade que caracteriza o câncer é a acelerada formação de células anormais que crescem além de seus limites habituais, podendo invadir partes adjacentes do corpo e se espalhar para outros órgãos, processo referido como metástase. A metástase é a principal causa de morte por câncer. De acordo com a Organização Pan–Americana de Saúde (OPAS), essa patologia representa uma das principais causas de morte no mundo, tendo sido responsável por cerca de 9,6 milhões de mortes em 2018. A nível global, uma em cada seis mortes são relacionadas à doença (BRASIL, 2018e).

A origem do câncer ainda é desconhecida (VILA-NOVA, 2017) mas, de acordo com o Instituto Nacional de Câncer (INCA), são inúmeros os fatores que podem favorecer o câncer, tais como: tabaco, álcool, dieta inadequada, vida sedentária, alguns vírus como o da hepatite B e fatores genéticos (BRASIL, 2018d). Na Figura 1 são ilustradas as principais causas do câncer.

Figura 1 – Estatísticas sobre as principais causas de câncer



Fonte: BRASIL (2018e)

Em termos mundiais, excluindo-se os cânceres de pele não melanoma, o câncer de mama é o mais frequente e comum tumor maligno entre as mulheres. Para o Brasil é estimado cerca de 59.700 novos casos de câncer de mama para cada ano do biênio 2018-2019, com um risco estimado de 56,33 casos a cada 100 mil mulheres (BRASIL, 2018e; BRASIL, 2018b). Outro fator preocupante é que o número de casos de câncer de mama tem aumentado anualmente em mulheres jovens (BORCHARTT et al., 2013). Esse aumento pode ser resultante, em parte, de uma maior exposição das mulheres a fatores de risco, decorrente do processo de urbanização e de mudanças no estilo de vida (PORTER, 2008), agravados pelo envelhecimento populacional que vem ocorrendo no Brasil de forma intensa (VICTORA et al., 2011).

Devido a esta grande quantidade de influências, a detecção precoce da doença é essencial, já que quanto mais cedo ela é descoberta, mais eficaz tende a ser o tratamento e maior é a chance de cura da paciente. Como consequência disso, ocorrerá uma redução na taxa de mortalidade em virtude desse tipo de câncer (BORCHARTT et al., 2013). De acordo com Lessa e Marengoni (2016), a probabilidade de cura do câncer de mama diminui consideravelmente se a doença não for detectada nos estágios iniciais.

Atualmente, a mamografia é o exame de referência para o diagnóstico do câncer de mama, porém este método apresenta algumas limitações, tais como a dificuldade na detecção da doença no caso de mamas densas, ou seja, mamas formadas principalmente por tecido glandular, que ocorre na maioria das pacientes jovens (BORCHARTT et al., 2013); são ainda registrados altos índices de falsos positivos na utilização dessa técnica, além da exposição da paciente à radiação ionizante, fator que pode inclusive aumentar as chances da mulher desenvolver a doença. Desse modo, percebe-se a necessidade de outras técnicas que deem suporte ao diagnóstico do câncer de mama (LELES et al., 2015).

O autoexame e o exame clínico são habitualmente os primeiros procedimentos a serem realizados na busca de modificações suspeitas. Os dois procedimentos são manuais. No autoexame, a paciente apalpa a mama procurando as alterações. Já no exame clínico, um especialista apalpa a mama também em busca de alterações. Porém, esses exames só detectam alterações em estágio já bastante avançado (BRASIL, 2018e; BRASIL, 2018b). Vários outros exames baseados em imagem também podem ser utilizados para detecção do câncer de mama. Entre eles a ressonância magnética, cuja principal desvantagem é o alto custo para sua realização, o que pode ser impeditivo em muitos casos (BORCHARTT et al., 2013).

A termografia infravermelha (TIR) é uma técnica de triagem não invasiva, barata e rápida que tem como objetivo registrar a radiação emitida pela superfície da pele da paciente ao longo de um período de tempo (DOURADO, 2014). Alguns estudos indicam que essa técnica pode detectar o câncer de mama mais precocemente que outros métodos, oferecendo o potencial de detectá-lo em até anos antes da mamografia (BORCHARTT et al.,

2013). O corpo humano emite energia térmica que pode ser transformada em temperatura. Essa energia pode ser detectada utilizando-se uma câmera termográfica. As câmeras de infravermelho, ou câmeras termográficas, percebem a radiação térmica emitida pelo corpo e a converte em uma imagem que representa a distribuição de temperaturas superficiais desse corpo (GONÇALVES, 2017). As variações de temperatura do tecido canceroso em relação ao tecido vizinho saudável se dão pelo processo de angiogênese. Por esse processo, o tumor estimula a criação de novos vasos sanguíneos para sua alimentação. Com mais vasos alimentando o tumor, a temperatura da região se mostra superior em relação à sua redondeza (DOURADO-NETO, 2014).

Na década de 50, iniciaram os primeiros trabalhos e relatos sobre o uso da TIR para detecção do câncer de mama. Porém, como consequência dos resultados sem relevâncias obtidos na época, a técnica entrou em desuso. Entretanto, com os avanços tecnológicos nos equipamentos de câmeras infravermelhas, nos anos 2000, os pesquisadores voltaram a considerar o uso desse método de imagem, que tem se mostrado uma técnica promissora (KANDLIKAR et al., 2017). Segundo (KEYSERLINGK et al., 1998), a sensibilidade para a detecção de câncer do tipo carcinoma ductal é significativamente melhorada quando combinado o exame de mamografia com a imagem térmica, alcançando a margem de 95%. Portanto, pelos fatos expostos acima, a TIR configura-se como uma atraente técnica de auxílio à detecção precoce do câncer de mama.

As ferramentas de classificação de tumores são específicas e complexas e, frequentemente, os patologistas têm alguma dificuldade na utilização das mesmas, o que limita seu uso (FERREIRA; OLIVEIRA; MARTINEZ, 2011). Portanto, a combinação de profissionais especializados e técnicas de análise digital de imagens termográficas de mama pode auxiliar o diagnóstico, o prognóstico e o tratamento do câncer de mama (BANDYOPADHYAY, 2010).

Diante disso, sistemas inteligentes têm sido desenvolvidos como método de auxílio em diferentes áreas da saúde, e podem colaborar com patologistas para uma classificação mais efetiva dos tumores analisados, diminuindo as limitações de uso impostas pelos sistemas de classificação atuais, acelerando, portanto, o trabalho desses profissionais (FERREIRA; OLIVEIRA; MARTINEZ, 2011).

Na busca por um dispositivo ou método que seja apto à detecção precoce da doença em questão, alguns estudiosos do âmbito computacional têm empregado técnicas computacionais de mineração de dados. Dentre os mecanismos empregados, a criação e a seleção de atributos se destacam (KOHAVI; JOHN et al., 1997). A criação de atributos baseia-se em gerar novos atributos a partir de outros existentes, porém, é essencial que as informações importantes sejam capturadas em um conjunto de dados mais efetivo. Portanto, a seleção de atributos é utilizada para diminuir a dimensão dos dados, favorecendo a aplicação de algoritmos de mineração. A redução de dimensionalidade acarreta em uma representação mais concreta, destacando a atenção do usuário sobre os fatores mais

relevantes (WITTEN; FRANK; HALL, 2011). O problema desse método pode ser delineado em como descobrir um subconjunto de atributos de um conjunto de dados original que gere uma classificação mais efetiva (OLIVEIRA-JÚNIOR et al., 2017).

Uma das principais áreas da Inteligência Computacional (IC) é a Computação Evolucionária (CE), que fundamenta-se em estratégias para resolução de problemas baseando-se em métodos evolutivos oriundos da Teoria da Evolução de Darwin (FERREIRA, 1990), tais como os mecanismos de seleção natural, cruzamento e mutações, além do comportamento adaptativo (EBERHART; SHI, 2007). O objetivo fundamental da CE é desenvolver ferramentas para a construção de sistemas inteligentes para modelar comportamento inteligente (EBERHART; SHI, 2007).

Neste trabalho foi proposta a seleção de atributos de imagens termográficas com lesões mamárias utilizando os Algoritmos Genéticos (AG) e a Otimização por Enxame de Partículas (PSO). O principal objetivo dessa pesquisa foi analisar principalmente as etapas de seleção de atributos e de classificação, as quais são essenciais para a obtenção de um sistema capaz de interpretar as informações de entrada e generalizar a tomada de decisão, para, assim, construir um sistema mais eficaz e robusto.

1.1 Objetivos

O objetivo deste trabalho é desenvolver um sistema inteligente utilizando imagens termográficas para auxiliar no diagnóstico de lesões mamárias, servindo assim como dispositivo de triagem e apoio à detecção precoce do câncer de mama. Para alcançar o objetivo deste trabalho, os seguintes objetivos específicos foram buscados:

- O pré-processamento das imagens termográficas e a utilização de técnicas de extração de características;
- Com a extração de características, foi realizada a separação das classes em seus respectivos diagnósticos: normal, tumor benigno, maligno e cisto;
- Seleção e aplicação de algoritmos para seleção de atributos das imagens;
- Realização de experimentos buscando os melhores parâmetros para os algoritmos;
- Testes com diversos classificadores, avaliando a acurácia dos subconjuntos de atributos selecionados;
- Comparação entre a acurácia da classificação com todos os atributos e com a classificação dos subconjuntos de atributos encontrados com os algoritmos;
- Análise da efetividade da classificação dos subconjuntos de atributos selecionados;

- Contribuir para construção de uma máquina de aprendizado conexionista para classificação de imagens termográficas de mama, com habilidade de detectar lesões;
- Propor uma solução para dispositivo móvel que auxilie na triagem e diagnóstico do câncer de mama e que funcione via internet.

1.2 Estrutura do Trabalho

Esse trabalho está organizado em 6 capítulos. O Capítulo 1 trata de uma introdução a respeito do câncer, em específico sobre o câncer de mama. Nele foi citada a importância do diagnóstico precoce da patologia e a utilização da termografia como método auxiliar para detecção. No Capítulo 2 encontram-se trabalhos relacionados. No Capítulo 3 há a revisão bibliográfica, capítulo que apresenta os conceitos fundamentais para o entendimento deste trabalho. O Capítulo 4 descreve a metodologia da pesquisa. No Capítulo 5, o detalhamento dos testes, os resultados obtidos e as discussões desses resultados são apresentados. Finalmente, no Capítulo 6, encontram-se a conclusão e as perspectivas para trabalhos futuros. No Apêndice A são apresentados os testes que resultaram nos melhores resultados apresentados no texto.

2 Trabalhos Relacionados

Há na literatura diversos trabalhos que utilizam imagens termográficas de mama para identificar a presença de alterações na mama, que podem ser benignas ou malignas.

No estudo de Gonçalves (2017), foi apresentada uma metodologia que classifica as pacientes como normais (sem lesão), com alterações benignas ou malignas em um banco de 70 imagens termográficas. Nestas imagens, foram realizadas técnicas de pré-processamento, segmentação e, logo em seguida, foram extraídas 17 características de interesse, através de medidas estatísticas e dimensão fractal. Posteriormente, foram utilizados classificadores como Máquina de Vetor de Suporte (SVM) e Redes Neurais Artificiais (RNA). O melhor resultado obtido foi 80,95% de acurácia utilizando o classificador SVM com função kernel cúbica. Esta pesquisa mostrou resultados favoráveis no sentido de comprovar a premissa de que as imagens termográficas podem contribuir na detecção do câncer de mama.

Diferentemente do trabalho de Gonçalves (2017), no estudo de Albonico (2017) é apresentada uma abordagem de seleção de atributos e classificação de lesões mamárias em imagens de ultrassom. O banco de dados utilizado é composto por 541 imagens, das quais 314 são de lesões benignas e 227 de lesões malignas, com diagnóstico comprovado por biópsia. O banco de dados contém a segmentação manual destas imagens, realizada por um médico especialista, e 22 atributos morfológicos previamente extraídos. Para facilitar a interpretação desses exames, surgiram os sistemas *Computer-Aided Diagnosis* (CAD), os quais visam fornecer uma segunda opinião para os médicos. Nesta pesquisa, foram desenvolvidas as etapas de seleção de atributos e de classificação presentes nestes sistemas. Foram utilizadas abordagem *Wrapper*, com estratégia de busca baseada em algoritmos genéticos, e duas estratégias em filtro, o teste t de Welch e o algoritmo *ReliefF*. Para avaliar o desempenho dos subconjuntos foi elaborado um classificador do tipo *Multilayer Perceptron* (MLP). A métrica utilizada para avaliar o desempenho de classificação de cada subconjunto de atributos foi a área sob a curva *Receiver Operating Characteristics* (Az). Os resultados encontrados pelas técnicas em filtro mostraram que a seleção de atributos é capaz de fornecer bons resultados na classificação, alcançando 0,731 para Az. Porém, o método *Wrapper* se mostrou mais efetivo, obtendo um valor de 0,835 para Az, utilizando 8 dos 22 atributos existentes na base de imagens. Portanto, esta pesquisa demonstrou que a seleção de atributos conseguiu diminuir a quantidade de atributos e aumentar o desempenho final da classificação.

Na pesquisa de Madhu et al. (2016), é desenvolvida uma ferramenta automática que busca aumentar a especificidade na análise das imagens termográficas. A base de dados

em estudo é composta por imagens de 265 pacientes, sendo 120 com mamas normais, 78 com anomalias malignas e 67 anomalias benignas, lactantes e tecidos sensíveis a hormônio. Este método teve como objetivo distinguir tumores malignos de outras lesões não malignas que causam aquecimento localizado na mama. A metodologia foi desenvolvida utilizando como características a presença de regiões anormais (regiões que manifestam aumento considerável de temperatura comparado com suas regiões vizinhas), temperatura relativa, comparação de temperatura de mamas contralaterais, e as margens/bordas. Foram encontrados resultados com especificidade de até 98,9%.

Lessa e Marengoni (2016) desenvolveram uma metodologia que classifica as imagens termográficas em sem lesão e com lesão. A base de dados utilizada na pesquisa foi composta por imagens de 47 pacientes diferentes, portanto, imagens de 94 mamas foram consideradas, sendo elas 48 mamas normais e 46 mamas com anomalias. As características consideradas foram: média, mediana, variância, desvio padrão, assimetria, curtose, entropia e intervalo. Para a classificação, foram utilizadas duas propostas de Redes Neurais Artificiais. A primeira abordagem testada foi uma RNA que resolve problemas lineares e a segunda que resolve problemas não lineares. A primeira proposta não apresentou resultados consideráveis, enquanto que a segunda abordagem apresentou resultados promissores, com 87% de sensibilidade, 83% de especificidade e 85% de acurácia.

Silva-Neto (2016) propôs uma metodologia computacional para auxiliar na detecção de massas de mamas densas e não densas, utilizando a mamografia. Esta pesquisa foi dividida em seis etapas. A primeira fase é composta pela aquisição de imagens, adquiridas a partir da *Digital Database for Screening Mammography* (DDSM). Na segunda fase, foi realizado o pré-processamento das imagens, com o objetivo de eliminar e realçar as estruturas presentes nas mesmas. Na terceira fase, realizou-se a segmentação utilizando a *Particle Swarm Optimization* (PSO) para encontrar as regiões de interesses (ROIs) candidatas à massas. A quarta fase, foi caracterizada pela redução de falsos positivos, tendo como objetivo remover ROIs indesejáveis. Na quinta fase, foram extraídas as características de textura, utilizando os índices de diversidade funcional (FD). Por fim, na sexta fase, foi utilizado o classificador SVM para testar a metodologia proposta. O melhor resultado para mama não densa foi de acurácia de 93,52% e o melhor resultado para mama densa foi de acurácia de 94,82%.

No estudo de Borchartt et al. (2013), a metodologia proposta buscou distinguir a presença ou ausência de lesões mamárias em imagens termográficas, sem o objetivo de identificar se a lesão era benigna ou maligna. A base de dados utilizada foi constituída de imagens termográficas de 69 pacientes, sendo 19 pacientes saudáveis e 50 com alguma lesão mamária. O pré-processamento de imagem foi feito através do *Software Development Kit* (SDK) da Flir. A extração das regiões de interesse foi realizada de forma automática através do aplicativo *Groud Truth Maker*. Para extração de características, foram utilizadas

diversas técnicas diferentes, sendo elas: dimensão fractal, características extraídas dos histogramas, características extraídas de medidas estatísticas, e características extraídas de métodos de geoestatística; as características obtidas foram também combinadas entre si para fins de comparação. A classificação foi realizada através do classificador SVM e otimizada por algoritmos genéticos. Nesta pesquisa, foi alcançada uma acurácia de 88%, especificidade de 79% e sensibilidade de 92%.

Tomando como base estes trabalhos na literatura, percebe-se que há poucos trabalhos que propõem-se classificar uma base de imagens termográficas de mama em quatros classes diferentes (normais, cisto, benignas e malignas), como é o objetivo deste trabalho. A maioria dos artigos adota classificação binária, seja ela com ausência ou com presença do câncer (no caso da ausência, tanto pacientes normais quanto com alterações benignas pertenceriam a esta classe), seja com alteração (benigna ou maligna) ou normais.

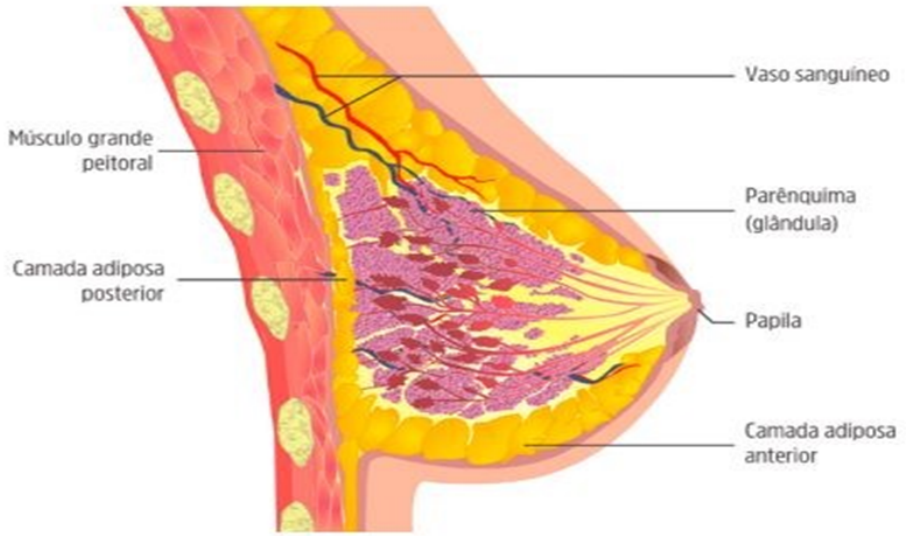
3 Referencial Teórico

Para o entendimento deste trabalho é necessário o conhecimento de alguns conceitos relacionados ao câncer de mama, a métodos utilizados para detecção da doença, à termografia infravermelha e a técnicas computacionais. Estes conceitos serão apresentados neste capítulo.

3.1 Anatomia e fisiologia da mama

A mama é uma glândula sudorípara modificada, formada por parte glandular, gordura e elementos fibrosos. A glândula, também chamada de parênquima, é composta por ductos e lobos. A parte gordurosa envolve toda a mama e é dividida em camada adiposa anterior e camada adiposa posterior. A porção fibrosa sustenta a mama e, para isso, circunda e atravessa a glândula (BRASIL, 2018c). A Figura 2 apresenta a distribuição dos elementos que compõem a mama.

Figura 2 – Distribuição dos elementos que compõem a mama, visão lateral



Fonte: BRASIL (2018c)

De acordo com BRASIL (2018c), as mulheres mais jovens possuem mamas mais densas e firmes como consequência da maior quantidade de tecido glandular. Na menopausa, o tecido mamário atrofia e é substituído por tecido gorduroso, até ser formado principalmente por gordura e alguns resquícios de tecido glandular na fase pós-menopausa. A principal função desta glândula é a produção do leite para a amamentação, mas tem tam-

bém grande importância psicológica para a mulher, possuindo papel essencial na formação de sua autoestima e autoimagem.

Na infância, as meninas apresentam discreta elevação na região mamária, decorrente da presença de tecido mamário rudimentar. Na puberdade, a hipófise, uma glândula localizada no cérebro, produz os hormônios folículo-estimulante e luteinizante, que controlam a produção hormonal de estrogênios pelos ovários. Com isso, as mamas iniciam seu desenvolvimento com a multiplicação dos ácinos e lóbulos. A progesterona que passa a ser produzida quando os ciclos menstruais se tornam ovulatórios, depende da atuação prévia do estrogênio, é diferenciadora da árvore ducto-lobular mamária (BRASIL, 2002).

Na vida adulta, o estímulo cíclico de estrogênios e progesterona fazem com que as mamas fiquem mais túrgidas no período pré-menstrual, por retenção de líquido. A ação da progesterona, na segunda fase do ciclo, leva a uma retenção de líquidos no organismo, mais acentuadamente nas mamas, provocando nelas aumento de volume, endurecimento e dor. Depois da menopausa, devido à carência hormonal, ocorre atrofia glandular e tendência à substituição do tecido parenquimatoso por gordura (BRASIL, 2002).

No período de amamentação é quando ocorre o perfeito funcionamento das mamas. O hormônio ocitocina, produzido na hipófise, é estimulado com a sucção da criança no momento do aleitamento (BRASIL, 2002). A saída do leite é decorrente da contração das células mioepiteliais, que circundam os ácinos. A mulher que nunca amamentou jamais atingirá a maturidade funcional mamária (BRASIL, 2002; FRANCO, 1997; HARRIS et al., 1996).

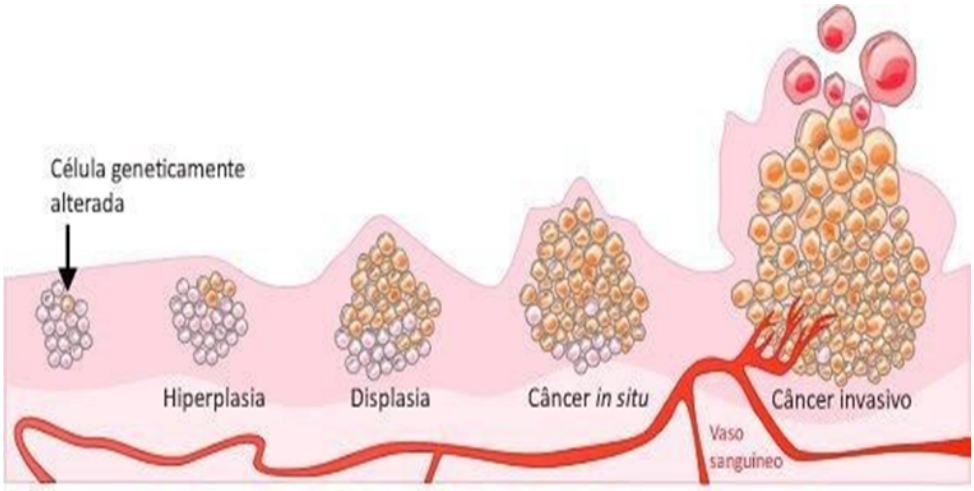
3.2 Câncer de mama

O câncer é um desenvolvimento patológico que é iniciado quando células normais são alteradas por mutação genética do DNA celular. Essas células apresentam propriedades invasivas, infiltrando-se nos tecidos adjacentes, e possuem acesso a vasos sanguíneos e linfáticos, através dos quais migram para outras regiões distante no corpo, esse episódio é chamado de metástase. Essa anormalidade celular forma cópias e prolifera de maneira irregular, ignorando os sinais de regulação do crescimento no ambiente vizinho à célula (SMELTZER; BARE, 2015).

Dentro do período de vida, diversos tecidos orgânicos habitualmente sofrem períodos de crescimento rápido, ou proliferativo, que devem ser diferenciados da atividade de crescimento maligno (SMELTZER; BARE, 2015). A multiplicação celular pode ser controlada ou não controlada. No crescimento controlado, tem-se um aumento localizado e autolimitado do número de células de tecidos normais que formam o organismo, provocado por estímulos fisiológicos ou patológicos. Neste processo, as células são normais ou com mínimas modificações no seu aspecto e funcionalidade, podendo ser iguais ou diferentes do tecido

onde se instalam. O efeito é reversível após o término dos estímulos que o provocaram. A hiperplasia, a metaplasia, a displasia e a neoplasia são exemplos desse tipo de crescimento celular (BRASIL, 2011). A Figura 3 apresenta os tipos de crescimento celular.

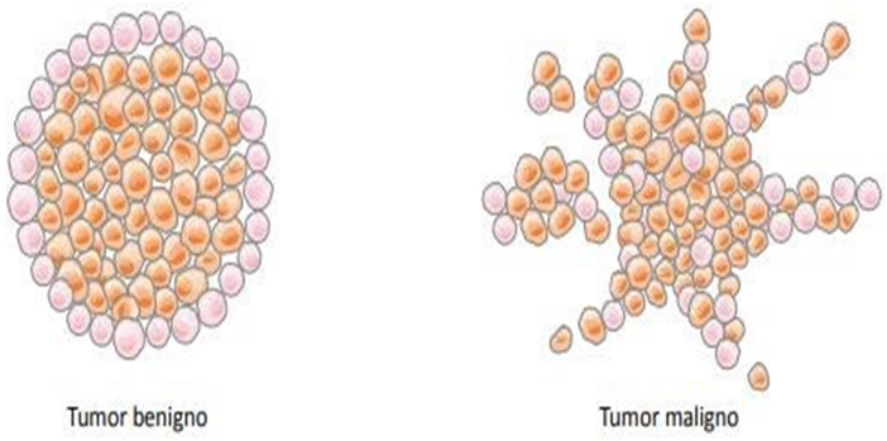
Figura 3 – Tipos de crescimento celular



Fonte: BRASIL (2011)

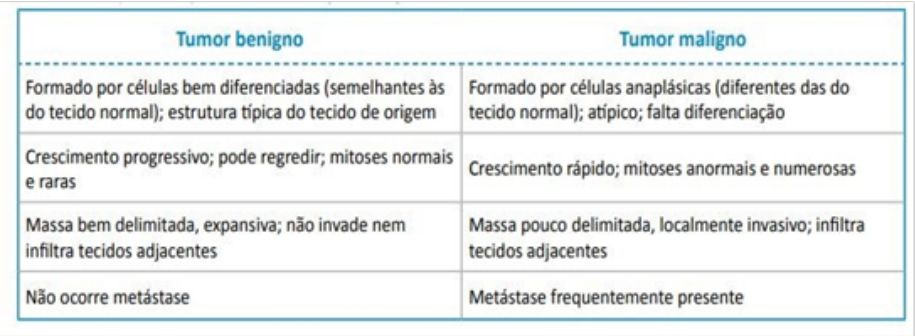
Os tumores podem ser benignos ou malignos (ver Figura 4). Os tumores benignos têm seu desenvolvimento de maneira ordenada, geralmente demorado, expansivo e apresentam bordas bem definidas. Mesmo não invadindo os tecidos vizinhos, podem comprimir os órgãos e tecidos adjacentes. As neoplasias malignas apresentam um maior grau de independência e são capazes de invadir tecidos vizinhos e provocar metástases, podendo ser resistentes ao tratamento e causar a morte do hospedeiro (BRASIL, 2011). Portanto, o câncer pode ser definido como uma neoplasia maligna. Existem algumas características que diferenciam as células benignas das malignas (SMELTZER; BARE, 2015). Essas diferenças são resumidas no quadro da Figura 5.

Figura 4 – Diferença entre os tipos de tumores



Fonte: BRASIL (2011)

Figura 5 – Diferença entre os tipos de tumores



Tumor benigno	Tumor maligno
Formado por células bem diferenciadas (semelhantes às do tecido normal); estrutura típica do tecido de origem	Formado por células anaplásicas (diferentes das do tecido normal); atípico; falta diferenciação
Crescimento progressivo; pode regredir; mitoses normais e raras	Crescimento rápido; mitoses anormais e numerosas
Massa bem delimitada, expansiva; não invade nem infiltra tecidos adjacentes	Massa pouco delimitada, localmente invasivo; infiltra tecidos adjacentes
Não ocorre metástase	Metástase frequentemente presente

Fonte: BRASIL (2011)

A carcinogênese é definida como o processo de formação do câncer, e normalmente, desenvolve-se de forma lenta, sendo capaz de levar vários anos para que uma célula cancerosa se multiplique e dê origem a um tumor visível. Os resultados cumulativos de diversos agentes cancerígenos são os responsáveis pelo início, promoção, progressão e inibição do tumor. A carcinogênese é caracterizada pela exposição frequente a esses agentes e também pela interação entre eles. Devem ser consideradas, no entanto, as características individuais, que favorecem ou dificultam a instalação do dano celular (BRASIL, 2018c; BRASIL, 2018a). Esse desenvolvimento é composto por três estágios:

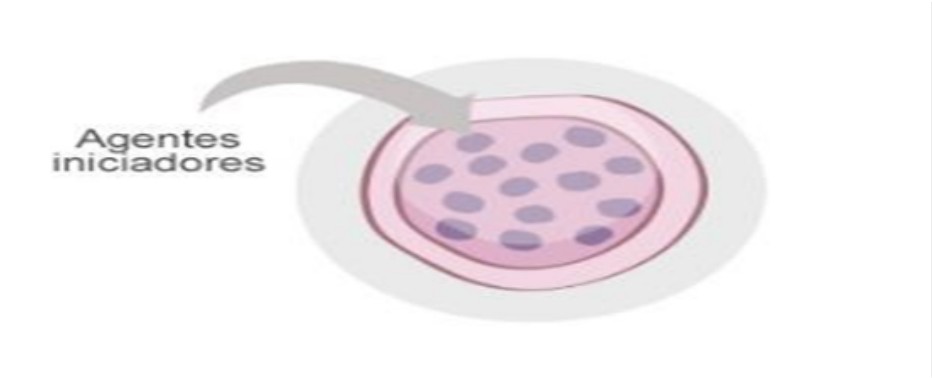
Estágio de iniciação: o DNA é atingindo frequentemente por carcinógenos, que ocasionam alterações em alguns de seus genes. Nessa etapa, as células se modificam geneticamente, entretanto ainda não é possível identificar um tumor clinicamente. Elas encontram-se aptas, ou seja, iniciadas para a ação de um segundo grupo de agentes que atuará no próximo estágio (BRASIL, 2018c). Ver Figura 6.

Estágio de promoção: as células que sofreram o processo de iniciação, são atingidas pela ação de novos agentes cancerígenos, classificados como oncopromotores. A célula iniciada é modificada para célula maligna, de maneira lenta e gradativa. Para que ocorra essa transformação é necessário um longo e continuado contato com o agente cancerígeno promotor. A interrupção do contato com agentes promotores muitas vezes encerra o processo nesse estágio (BRASIL, 2018c). Ver Figura 7.

Estágio de progressão: esse estágio é definido pela proliferação descontrolada e irreversível das células modificadas. Nessa etapa, o câncer já está introduzido, evoluindo até o aparecimento das primeiras manifestações clínicas da doença (BRASIL, 2018c). Ver Figura 8.

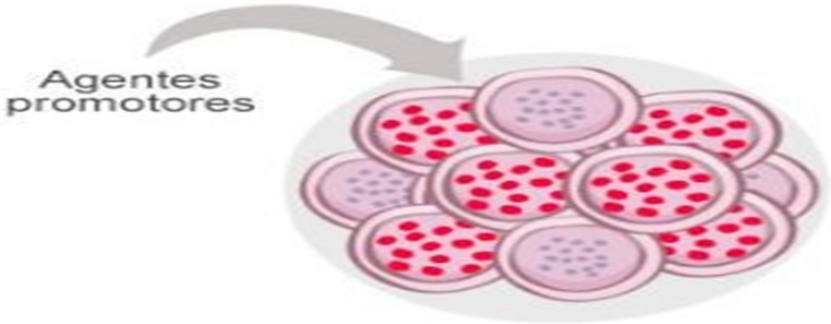
É importante destacar que a neoplasia maligna desenvolve um processo chamado angiogênese. Essa alteração pode ser definida pelo aparecimento de novos capilares, a partir do tecido do hospedeiro, pela liberação de fatores de crescimento e enzimas, como o

Figura 6 – Estágio de iniciação: agentes iniciadores



Fonte: BRASIL (2011)

Figura 7 – Estágio de promoção: agentes promotores



Fonte: BRASIL (2011)

Figura 8 – Estágio de progressão

Fonte: BRASIL (2011)

fator de crescimento do endotélio vascular. Essas proteínas promovem, de forma acelerada, a criação de novos vasos sanguíneos, o que auxilia as células malignas a obter os nutrientes e oxigênio necessários (SMELTZER; BARE, 2015). Uma das principais doenças que podem acometer a mama é o câncer de mama, também chamado de carcinoma (BORCHARTT et al., 2013). São identificados mais de 20 tipos dessa patologia, entre eles estão o carcinoma ductal, o carcinoma e o câncer inflamatório (KANDLIKAR et al., 2017).

3.3 Métodos de diagnóstico do câncer de mama

Além do exame clínico, os exames de imagens podem ser recomendados para a detecção do câncer de mama. Os principais métodos utilizados são: mamografia, ultrassonografia e ressonância magnética. Porém, a confirmação diagnóstica só é validada por meio de biópsia, procedimento que consiste na retirada de um fragmento do nódulo ou da lesão suspeita através de punções ou de uma pequena cirurgia. O material retirado é analisado pelo patologista para a definição do diagnóstico (BRASIL, 2018a).

A ultrassonografia de mama é empregada especialmente em pacientes com mamas radiologicamente densas, nas quais a quantidade de tecido glandular pode ofuscar nódulos ou outras alterações. Nesse caso, ao se diagnosticar, através da mamografia, assimetrias de densidade e determinados tipos de nódulos pode ser necessária uma ultrassonografia de mama (LABADESSA, 2019).

Segundo a *American Cancer Society*, a ressonância magnética também pode ser utilizada no diagnóstico do câncer de mama, especialmente em mulheres que já foram diagnosticadas com a doença, para definir com mais exatidão o tamanho do tumor e a existência de outros tumores na mama (RIEBER et al., 1997). Essa técnica é recomendada juntamente com a mamografia anual para diagnóstico do câncer de mama em mulheres com alto risco da doença. Porém, não é indicada como um exame de rastreamento de forma isolada porque pode perder alguns tipos de câncer que poderiam ser diagnosticados através da mamografia. Outra desvantagem desse método, é o alto custo.

Atualmente, a mamografia é o procedimento padrão para o diagnóstico de câncer de mama (KANDLIKAR et al., 2017), porém este exame apresenta algumas limitações, tais como a deficiência na detecção da doença no caso de mamas densas, que são mamas formadas principalmente por tecido glandular, o que é o caso da maioria das pacientes jovens (BORCHARTT et al., 2013; KANDLIKAR et al., 2017); os altos índices de falsos positivos e a exposição da paciente a radiação ionizante, podendo inclusive aumentar as chances de desenvolver a doença. Dessa forma, percebe-se a necessidade de outras metodologias que deem suporte ao diagnóstico da doença em questão (LELES et al., 2015).

3.4 Algoritmos Genéticos

Os algoritmos genéticos são uma espécie particular de algoritmos evolutivos que utilizam estratégias inspiradas pelo princípio do Darwinismo de seleção natural e reprodução genética. De acordo com a teoria de Darwin, o processo de seleção permite maior durabilidade aos indivíduos mais aptos, ou seja, maior probabilidade de reprodução. Estes indivíduos têm maior chance de perpetuar os seus códigos genéticos para as próximas gerações. Os códigos genéticos constituem a identidade de cada indivíduo e são representados nos cromossomos (EBERHART; SHI, 2007). Esses algoritmos foram propostos por John H. Holland (1977) e proporcionam explorar uma ampla gama de potenciais soluções, utilizando-se de aleatoriedade, cruzamento entre estas, e mutação de parâmetros para melhor explorar o espaço de estados de um problema (HOLLAND, 1992). Tratando-se de um algoritmo de otimização, possui um amplo campo de aplicação em diferentes áreas (EBERHART; SHI, 2007; RIBEIRO et al., 2014a; RIBEIRO et al., 2014b; FEITOSA et al., 2014b; BARBOSA et al., 2017; FEITOSA et al., 2014a).

Os AGs surgiram da necessidade de problemas computacionais que precisam de soluções capazes de se adaptar ao seu meio, observando mudanças na caracterização do problema, e, então, modificando sua reação (MITCHELL, 1998). É admissível realizar um comparativo entre a evolução de organismos vivos, e a atividade dos Algoritmos Genéticos. Os cromossomos fazem parte de todas as células de organismos vivos, esses servem como um guia de como cada célula deve se desenvolver dentro deste organismo. Cromossomos são formados por genes, que codificam características específicas dessas células e, consequentemente, do organismo (MITCHELL, 1998). Para um AG, organismos ou indivíduos representam possíveis soluções para o problema apresentado. Comumente, cada indivíduo de um AG dispõe de apenas um cromossomo. Cada um desses cromossomos é codificado como um conjunto de genes, no qual cada gene representa um parâmetro ou variável da solução (EBERHART; SHI, 2007).

Os indivíduos reproduzem-se e transferem suas particularidades para sua prole, através de uma recombinação de seus cromossomos, fazendo com que seu descendente possua genes (características) de ambos indivíduos que o geraram. As mutações ocasionam mudanças neste gene, frequentemente resultantes de falhas no modo de replicação e recombinação dos cromossomos, habitualmente ampliando a diversidade genética de uma população e, possivelmente, favorecendo a evolução (GRIFFITHS, 2012). No AG, a reprodução entre indivíduos permite a criação de novas soluções que possuam características de ambos reprodutores. Por fim, organismos mais adaptados ao seu meio dispõem de maior probabilidade de passar seus genes adiante, e auxiliar na evolução de uma espécie. Da mesma maneira, soluções mais próximas do objetivo desejado passam suas características para gerações seguintes, transferindo o problema a uma resposta (EBERHART; SHI, 2007; RIBEIRO et al., 2014a; RIBEIRO et al., 2014b; FEITOSA et al., 2014b; BARBOSA et al.,

2017; FEITOSA et al., 2014a).

Em resumo, o conceito fundamental do funcionamento dos AGs é o de tratar as prováveis soluções do problema como indivíduos de uma população, que irá evoluir a cada geração. A seguir, serão apresentados os principais passos que compõe os algoritmos genéticos de acordo com Linden (2006): População Inicial, Avaliação, Seleção, Crossover, Mutação, Avaliação de Indivíduo Filho, Geração de Nova População e Critério de Parada.

População inicial: Frequentemente, o início da população se dá por meio de sorteios de indivíduos. Pode ocorrer de ter indivíduos repetidos, porém não é muito comum devido ao fato do AG ser utilizado em um espaço de busca muito grande. Geralmente o espaço de busca é dividido em partes e é realizado o sorteio de indivíduos de cada parte, para proporcionar a maior diversidade possível à população. Desta forma, quando esses indivíduos estiverem sujeitos ao Crossover e a mutação, mais diversificados serão os indivíduos gerados e, certamente, com uma melhor qualidade de resposta ao problema (LINDEN, 2006).

Avaliação (*Fitness*): Posteriormente a criação da população inicial, todos os indivíduos da população são submetidos a uma avaliação. A função de avaliação define a eficiência do indivíduo para o problema em questão. Para que a função possua uma boa solução, é necessário que sejam introduzidas nela o máximo de informações possíveis acerca do problema (LINDEN, 2006).

Seleção: Após a avaliação, normalmente os indivíduos (pais) que irão dar origem as próximas gerações (filhos) são selecionados. A seleção segue o conceito da seleção natural, na qual os indivíduos mais aptos são aqueles com melhores capacidades de reprodução. No AG os indivíduos com um valor de avaliação melhor, serão selecionados. Porém, indivíduos com menor aptidão podem participar do processo de seleção para gerar indivíduos filhos, da mesma maneira que acontece na natureza (LINDEN, 2006).

A etapa de seleção de pais para a geração de indivíduos filhos é essencial no processo do AG. As escolhas de indivíduos poderão modificar o resultado final. Os métodos mais comuns para fazer a seleção dos indivíduos são os de roleta e torneio, descritos a seguir (LINDEN, 2006):

Torneio: É um método feito através do sorteio de vários indivíduos que vão concorrer entre si. Para tal, determina-se uma porcentagem da população de indivíduos total que serão submetidos à seleção. Posteriormente, o tour é definido, que é a quantidade de indivíduos sorteados que competirão. Sorteia-se os indivíduos e o que possuir a melhor aptidão no seu tour, passa para as etapas seguintes do AG. O benefício deste método é o privilégio concedido a todos os indivíduos, independentemente de

suas características. A vantagem dos indivíduos com melhores aptidões é levada em consideração exclusivamente na etapa da competição, onde o melhor é selecionado. A chance de ser sorteado é mesma entre os indivíduos da população (LINDEN, 2006).

Roleta: É um método que beneficia os indivíduos com melhores aptidões, entretanto os indivíduos com capacidades inferiores também podem fazer parte do processo, porém com menor probabilidade. A técnica opera da seguinte maneira: é analisada a porcentagem da avaliação de cada indivíduo em relação ao todo. Essa mesma porcentagem pode ser usada para estabelecer a quantidade de casas que esse indivíduo vai ter numa roleta, variando de 0 a 360. Após determinar as porcentagens de cada indivíduo, é estipulado os limites que cada um vai ocupar na roleta. Feito isso, sorteia-se um número na roleta, que pode ser um número entre 0 e 100 (correspondente a porcentagem de cada indivíduo) ou de 0 a 360 (correspondente aos graus da roleta). O número sorteado representará um dos indivíduos (LINDEN, 2006).

Como exemplo, pode-se observar na Tabela 1, em que os indivíduos possuem suas representações por binários, e vão ocupar um pedaço da roleta através da porcentagem que suas avaliações representam para o total.

Tabela 1 – Representação do método de seleção por roleta

Indivíduos	Avaliação	Pedaço da Roleta
0001	1	1.61
0011	9	14.51
0100	16	25.81
0110	36	58.07
Total	62	100

Fonte: A Autora

Por se tratar de um método que beneficia os indivíduos com melhores aptidões e também possibilita que indivíduos menos aptos participem do processo de geração de filhos, o método da roleta é o mais utilizado (LINDEN, 2006). Em seguida, os operadores genéticos são analisados:

***Crossover*:** Após eleger os indivíduos pais, alguns indivíduos passarão pelo Crossover. Este é um dos principais métodos para a contribuição de novos indivíduos. Também chamado de cruzamento ou recombinação, este procedimento é responsável por desempenhar a troca de dados entre os indivíduos, de modo que novos indivíduos sejam gerados a partir dessa recombinação de informações (MATOS, 2011).

Mutação: Posteriormente, é realizado o processo de mutação, no qual um ou mais indivíduos dos que foram selecionados são alterados. Esse parâmetro é usado para proporcionar ainda mais a heterogeneidade de indivíduos na população. A mutação

consiste em efetuar uma modificação na constituição do indivíduo. Portanto, a finalidade da mutação é garantir que os indivíduos no final de uma geração não sejam os mesmos que iniciaram o processo (POZO, 2019).

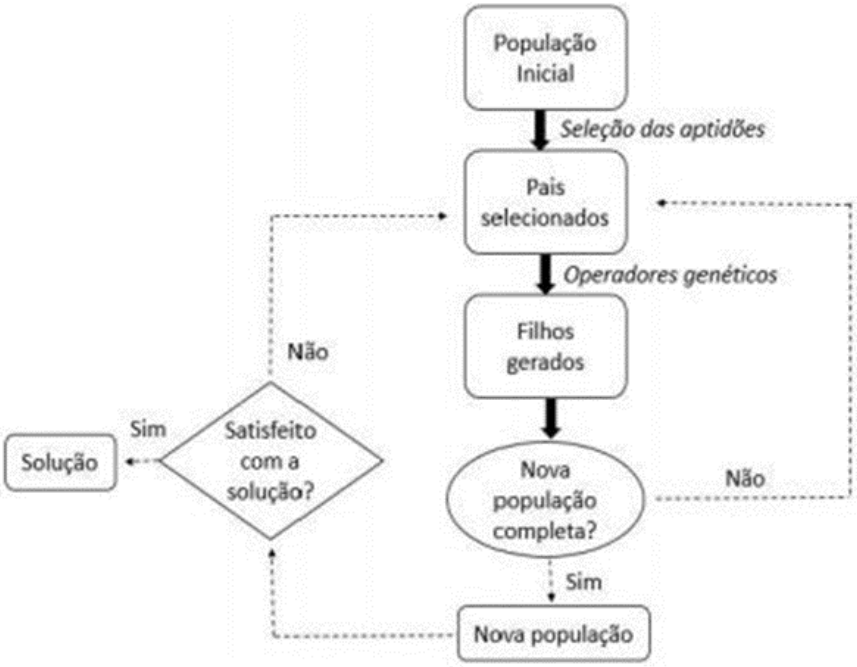
Posteriormente às etapas de Crossover e mutação, novos indivíduos são gerados, e logo há a necessidade de submetê-los à função de avaliação (LINDEN, 2006). E por último, são analisados a nova população gerada e o critério de parada

Geração da Nova População: Depois de avaliar os indivíduos filhos, estes são reunidos com os indivíduos da população inicial e uma regra é empregada com a finalidade de selecionar uma quantidade de indivíduos para a próxima geração. Geralmente são selecionados os indivíduos que possuem melhor avaliação (LINDEN, 2006).

Critério de Parada: É definida a quantidade de gerações máximas permitidas e o processo continua enquanto a quantidade de gerações estipulada não tiver sido alcançada, ou enquanto não atingir o indivíduo que satisfaça o problema (LINDEN, 2006).

A Figura 9 mostra o fluxograma do AG.

Figura 9 – Fluxograma do Algoritmo Genético



Fonte: Franciscani e Queiroz (2011)

3.5 Otimização por Enxame de Partículas

Em 1995, Kennedy e Eberhart criaram o método de otimização por enxame de partículas (PSO) (EBERHART; SHI, 2007). A técnica surgiu baseada no comportamento social dos pássaros a procura de alimento ou de um local para construção do ninho. Nessa busca, todo indivíduo (partícula) pode ganhar com as experiências dos integrantes do grupo (enxame) (EBERHART; SHI, 2007; RIZZI et al., 2016). Então, a partir disso foi apresentado um algoritmo de otimização que tem benefícios como: simples implementação computacional, uso mínimo de memória, pouca velocidade de processamento e o processo de busca é racionalizado pelo constante aprendizado das partículas (EBERHART; SHI, 2007).

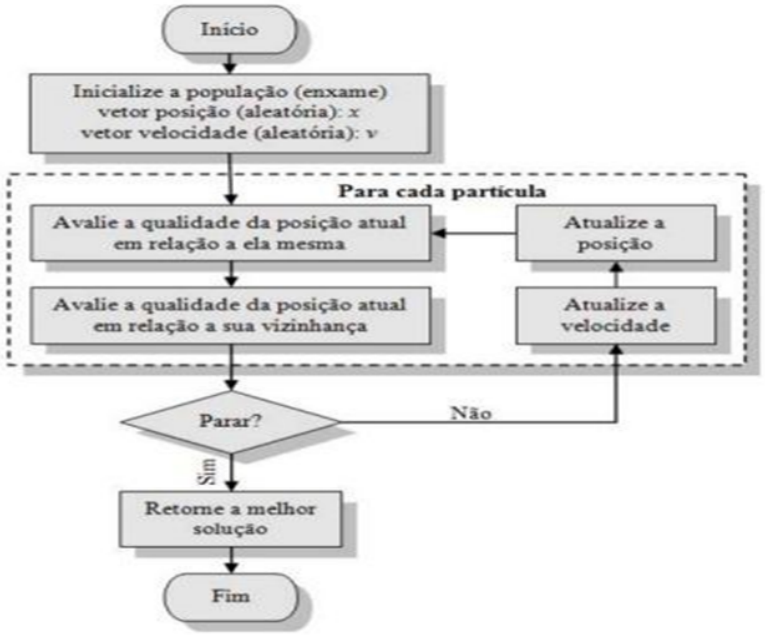
De acordo com Merwe e Engelbrecht (2003), apresentado um problema, o PSO mantém um grupo de partículas na qual cada uma constitui uma solução potencial para o mesmo e está relacionada a uma posição em um espaço de busca multidimensional. Inicialmente é gerado um conjunto aleatório de partículas e a este é atribuído uma velocidade, ou seja, cada indivíduo possui uma posição e uma velocidade (ENGELBRECHT, 2005). Cada partícula deste grupo é deslocada através do espaço de busca do problema por duas forças. Uma os atrai, com uma magnitude aleatória, para a melhor localização já encontrada por ele próprio (pbest) e outra para a melhor localização encontrada entre alguns ou todos os indivíduos do enxame (gbest) (CASTRO; TSUZUKI, 2007). Para cada iteração, a velocidade da partícula é modificada e um novo posicionamento é encontrado pela soma de sua posição atual e a nova velocidade (ENGELBRECHT, 2005), até todo o enxame convergir obtendo o melhor resultado (NASCIMENTO et al., 2012). A Figura 10 mostra o fluxograma do mecanismo do PSO.

3.6 Reconhecimento de Padrões de Imagens

Desde a década de 70, inúmeros métodos de recuperação de imagens vêm sendo desenvolvidos, especialmente pelas áreas de gerenciamento de banco de dados (indexação textual) e visão computacional (RUI; HUANG; CHANG, 1999). Com o crescimento dos bancos de dados de imagens, decorrente principalmente das inovações tecnológicas, é necessário algum tipo de mecanismo de indexação para posterior recuperação das imagens armazenadas. Porém, desempenhar esta indexação manualmente é uma tarefa propensa a interpretações subjetivas e sensível a erros (ANTANI; KASTURI; JAIN, 2002). As técnicas automáticas de indexação e recuperação com base em algum tipo de característica (cor, forma, textura ou região de uma imagem) são importantes neste contexto, uma vez que podem diminuir a intervenção humana, proporcionando maior eficácia e, em muitos casos, uma relevante diminuição da margem de erro (LIU; ZHOU, 2003).

O método reconhecimento de padrões busca realizar a descrição com base nas

Figura 10 – Fluxograma do PSO



Fonte: Franciscani e Queiroz (2011)

características próprias da imagem por meios computacionais, evitando desta forma a subjetividade provocada por um operador humano. Portanto, todas as imagens armazenadas possuiriam o mesmo nível de detalhamento (ERPEN, 2004). De acordo com Marques (1999), um sistema para reconhecimento de padrões pode ser dividido em 3 etapas. Estas fases são definidas na Figura 11.

O método de classificação pode ser dividido em supervisionado e não-supervisionado. A classificação supervisionada ocorre quando o classificador considera classes pré-definidas e uma fase de treinamento é realizada antes da classificação para que os parâmetros que caracterizam cada classe sejam obtidos. Na classificação não-supervisionada não se desfruta de conhecimento prévio na aplicação do algoritmo de classificação (PEDRINI; SCHWARTZ, 2008).

Atualmente, as técnicas automáticas de indexação de imagens podem ser restauradas a partir de um banco de dados mediante elementos gráficos como cor, textura, forma, entre outros. A fase de extração de características é executada utilizando as mais diversas abordagens. As mais habituais são aquelas que usam análise de textura e forma, podendo ser de modo individual ou combinado (PEDRINI; SCHWARTZ, 2008; OLIVEIRA et al., 2019; SOUZA et al., 2019).

A análise de textura foi desenvolvida na década de 1970 como método de avaliação e classificação de imagens (HARALICK; SHANMUGAM; DINSTEIN, 1973; HARALICK, 1979). É uma maneira de descrever a distribuição espacial de intensidades, o que a torna útil na classificação de regiões similares em imagens diferentes (MATERKA, 2004). Na observação

Figura 11 – Fases de um sistema para reconhecimento de padrões

Etapas	Características
Representação de dados e mensuração	• Representa dados de entrada que podem ser mensurados a partir do objeto a ser estudado; • Descreve os padrões característicos do objeto, possibilitando sua posterior classificação em uma determinada classe; • O vetor que define perfeitamente um objeto seria de dimensionalidade infinita.
Extração de características	• Retira características individuais e atributos do objeto; • Etapa em que é objetivado os fenômenos que se pretendem classificar; • Redução da dimensionalidade do vetor padrão, sem provocar danos a informação inerente a classificação, visando redução do esforço computacional.
Classificação do objeto em estudo	• Estabelece procedimentos que permitam a identificação e classificação do objeto em uma determinada classe de objetos.

Fonte: Marques (1999)

de imagens médicas, os descritores de textura foram adotados para análise de imagens ultrassonográficas do fígado (LERSKI et al., 1979) e do coração (SKORTON et al., 1983) no final dos anos 1970 e início dos anos 1980, conquistando popularidade nos anos 1990 e 2000 em muitas aplicações de imagens médicas, incluindo oncologia (BRYNOLFSSON et al., 2017). A utilização da textura permite a descrição da heterogeneidade tecidual, uma propriedade que se acredita influenciar o resultado do tratamento do câncer (O'CONNOR et al., 2015). Haralick é um método comum para representar textura da imagem, uma vez que é simples de implementar e resulta em um conjunto de descritores de textura interpretáveis (HARALICK; SHANMUGAM; DINSTEIN, 1973; HARALICK, 1979).

As características de forma têm grande relevância no contexto de diagnóstico de nódulos cancerígenos. Isto ocorre porque, na maioria dos casos, o formato de um nódulo é um dos principais fatores para definir benignidade ou malignidade (OLIVEIRA et al., 2019; SOUZA et al., 2019). Os descritores de forma Zernike são uma classe de momentos

ortogonais que vem sendo aplicada na área de representação de imagens (HSE; NEWTON, 2004). Este método dispõe de dois princípios: repetição e ordem, os quais relacionam-se à capacidade dos momentos representarem detalhes nas imagens. Portanto, diante da sua capacidade de prover informações relevantes relativas às formas presentes nas imagens, este descritor é muito utilizado em trabalhos de reconhecimento e classificação envolvendo imagens digitais (PEDRINI; SCHWARTZ, 2008; SINGH; MITTAL; WALIA, 2011; KHOTANZAD; HONG, 1990; TAHMASBI; SAKI; SHOKOUHI, 2011).

3.7 Seleção de Atributos

Dentre as técnicas empregadas em mineração de dados, destacam-se a criação e a seleção de atributos (KOHAVI; JOHN et al., 1997). A maioria dos estudos referentes a este método tem a finalidade de melhorar o processo de Aprendizado de Máquina (AM) quanto à sua acurácia (KOLLER; SAHAMI, 1996). A seleção de atributos pode ser entendida como um problema de otimização, onde é executado um processo de busca no espaço e é selecionado o subconjunto ótimo, ou que apresente melhor desempenho. Para isso, é fundamental determinar qual a estratégia de busca, a que frequentemente é utilizada uma meta-heurística (ALBONICO, 2017).

Intuitivamente, pode se acreditar que quanto maior o número de atributos em um conjunto de dados, maior o poder discriminatório e consequentemente maior a acurácia do classificador. Entretanto, na prática, esse comportamento não é verdadeiro (KOLLER; SAHAMI, 1996). Uma grande quantidade de atributos insignificantes pode fazer com que os algoritmos de aprendizagem tenham dificuldade em extrair informações que sejam realmente necessárias e relevantes para classificação. Além do mais, diversas pesquisas nessa área apontam que um grande número de atributos desnecessários pode introduzir ruídos aos dados, confundindo o algoritmo de aprendizagem e ocasionando erros na classificação (LIU; MOTODA, 1998).

Existem outros elementos a serem analisados no âmbito de AM (CASTRO et al., 2004). De acordo com Michie et al. (1994), são conectadas quatro propriedades associadas ao processo de AM, sendo que três delas podem ser diretamente otimizadas com o uso de métodos de seleção do conjunto de atributos. São elas:

Precisão: a eficácia de classificação de instâncias não observadas é uma das principais motivações para o processo de seleção de atributos, uma vez que é possível remover atributos irrelevantes e redundantes, o que normalmente leva a um aumento da precisão do método de aprendizado;

Velocidade: uma maior rapidez na classificação de novas instâncias pode ser decisória em muitos casos. Esta propriedade é bastante otimizada com o método de seleção de

atributos, pois, com a redução da quantidade de atributos, o conceito induzido tende a ser menos complexo, além da diminuição do volume de dados, reduzindo a carga computacional exigida;

Tempo de aprendizado: a velocidade do aprendizado é fundamental em casos onde o ambiente de aprendizado é alterado frequentemente. Esta característica nem sempre é otimizada já que um tempo adicional (relativo à redução do número de atributos) deve ser levado em conta;

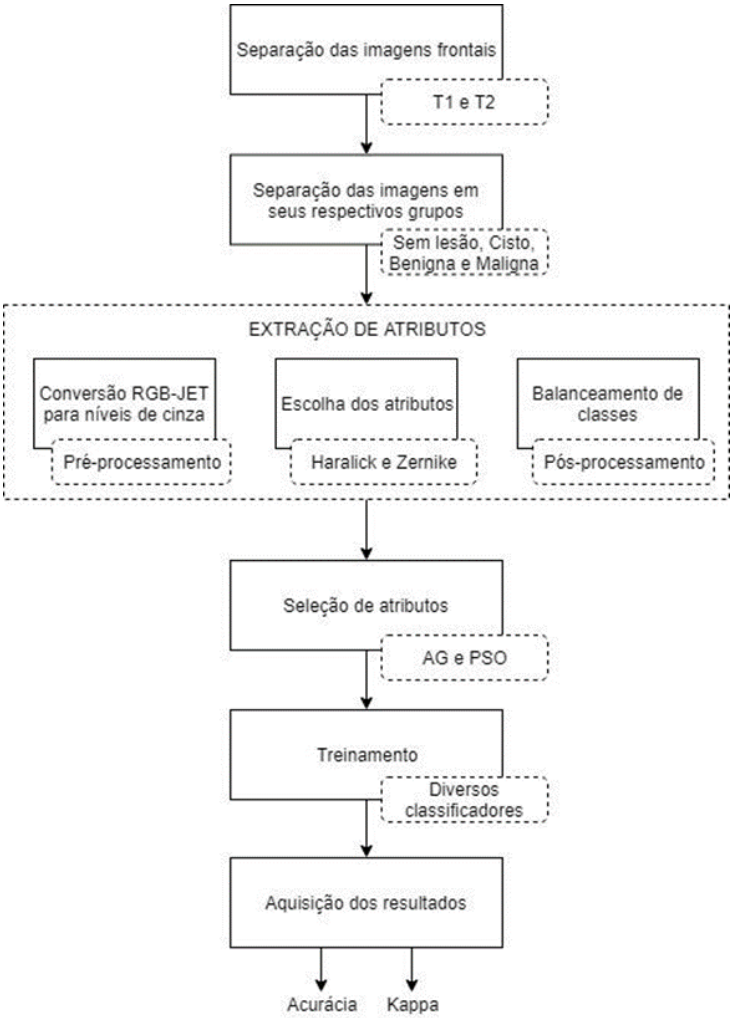
Compreensibilidade: em vários âmbitos do conhecimento é importante o entendimento dos porquês de um conceito induzido (para obter confiança do usuário). A compreensão geralmente é melhorada com a redução dos atributos e a consequente simplificação do conceito induzido.

Portanto, o processo de seleção de atributos é particularmente importante em situações em que instâncias são descritas usando um número grande de atributos e se desconhece a relevância de cada um desses atributos na representação do conceito (CASTRO et al., 2004). Foi visto que os métodos de seleção podem reduzir ou evitar o problema conhecido como “maldição da dimensionalidade” em muitas situações de aprendizado (KOLLER; SAHAMI, 1996).

4 Metodologia

Neste capítulo serão apresentados o banco de dados em conjunto com a descrição das etapas do método de seleção de atributos com os Algoritmos Genéticos e Enxame de Partículas, englobando desde o pré-processamento das imagens até a avaliação final dos classificadores utilizados. O fluxograma da Figura 12 ilustra as etapas da metodologia adotada.

Figura 12 – Etapas da metodologia



Fonte: A Autora

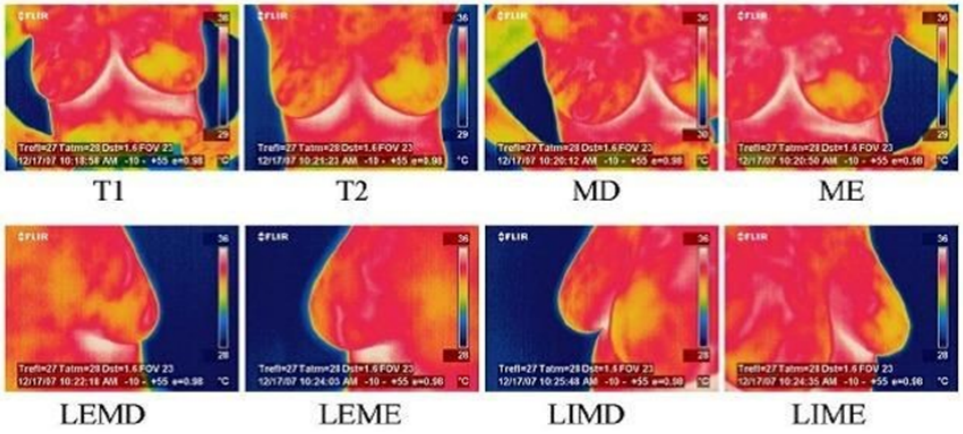
4.1 Base de dados

O banco de dados utilizado neste trabalho é constituído por imagens termográficas obtidas por meio de uma câmera de infravermelho do modelo FLIR S45. Estas imagens foram fornecidas pelo grupo de pesquisa em termografia de mama do Departamento de

Engenharia Mecânica da UFPE, sendo adquiridas de pacientes voluntárias do Ambulatório de Mastologia do Hospital das Clínicas (HC) da Universidade Federal de Pernambuco (UFPE), em um período entre 2005 e 2014. As pacientes analisadas no presente trabalho já tinham um diagnóstico concluído, realizado por exames clínicos convencionais, mamografia ou ultrassonografia, bem como a biópsia confirmando seu diagnóstico, quando necessário esse exame. O projeto teve a aprovação do Comitê de Ética da UFPE e foi registrado no Ministério da Saúde sob CEP/CCS/UFPE Nº 279/05, de novembro de 2005, e as pacientes assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), permitindo e confirmando total consciência na finalidade do projeto (VILA-NOVA, 2017).

Para cada paciente foram adquiridas imagens da região da mama em oito posições distintas: duas imagens frontais de ambas as mamas (T1 e T2) e três imagens de cada mama isoladamente, direita e esquerda, em ângulos distintos, sendo eles frontal (MD e ME), lateral externa (LEMD e LEME) e lateral interna (LIMD e LIME) (SANTANA et al., 2018). Essas posições são mostradas na Figura 13.

Figura 13 – Posições de aquisição das imagens por paciente



Fonte: Santana et al. (2018)

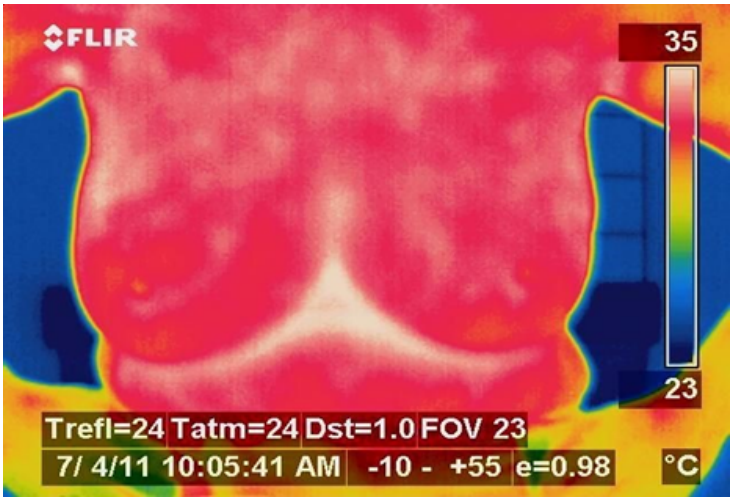
Nesta pesquisa foi utilizada uma amostra de 336 imagens termográficas, das quais, 121 pacientes foram diagnosticadas com tumor benigno, 76 com tumor maligno, 73 com cisto e 66 pacientes sem nenhuma anomalia mamária. Além disso, os experimentos foram realizados apenas com as imagens frontais T1 e T2, pois considera-se que essas condições favorecem a identificação da região de interesse.

As Figuras 14, 15, 16 e 17 contém quatro imagens da base de dados das classes sem lesão, benigno, maligno e cisto, respectivamente.

4.2 Processamento e segmentação das imagens

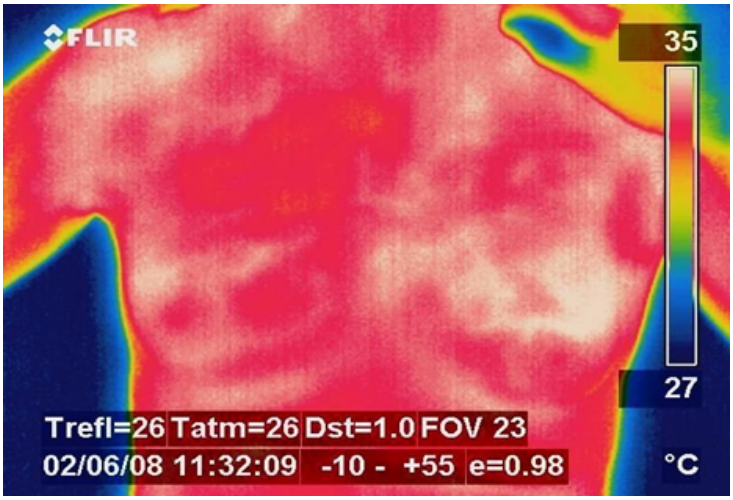
Esta etapa é utilizada para favorecer a visualização da imagem, aumentando o contraste, assim como diminuir ruídos associados à aquisição da imagem, auxiliando em

Figura 14 – Imagem termográfica de paciente sem lesão



Fonte: Santana et al. (2018)

Figura 15 – Imagem termográfica de paciente com lesão benigna



Fonte: Santana et al. (2018)

sua interpretação pelo especialista (ALBONICO, 2017).

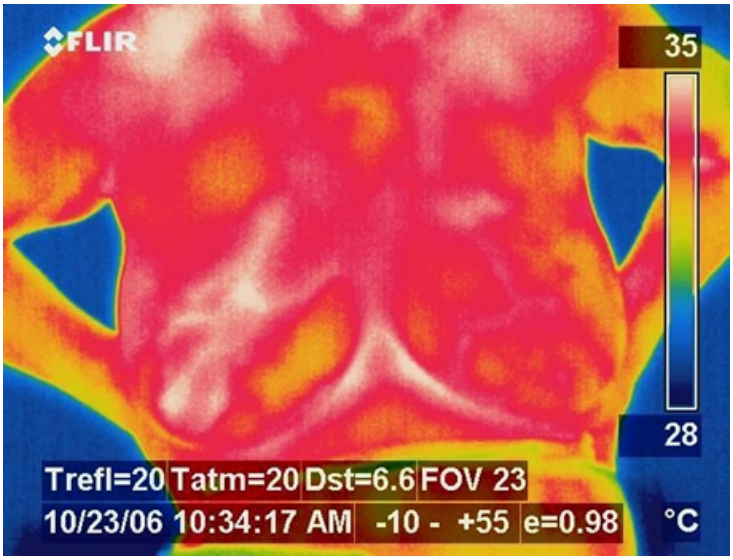
Inicialmente as imagens foram submetidas a uma etapa de pré-processamento de conversão de RGB-JET para níveis de cinza, com os tons mais claros indicando temperaturas mais altas (SANTANA et al., 2018).

4.3 Extração de atributos

Para a extração de atributos foram aplicadas técnicas do ponto de vista computacional para retirar atributos das imagens. Estes atributos são utilizados para classificar as lesões e, portanto, devem representar a natureza do nódulo. Normalmente são atributos morfológicos, relacionados à forma da imagem ou atributos de textura (ALBONICO, 2017).

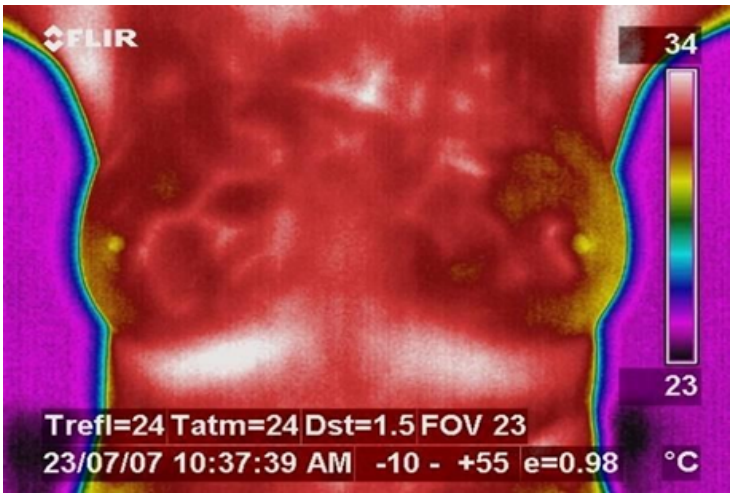
Foi proposta a utilização de momentos de Haralick e momentos de Zernike como

Figura 16 – Imagem termográfica de paciente com lesão maligna



Fonte: Santana et al. (2018)

Figura 17 – Imagem termográfica de paciente com lesão cística



Fonte: Santana et al. (2018)

atributos para a representação do conjunto de dados (SANTANA et al., 2018). Os momentos de Haralick fornecem informações relacionadas à textura, a partir da avaliação da probabilidade de ocorrência das combinações entre os níveis de cinza da imagem (OLIVEIRA et al., 2012). Já os de Zernike foram utilizados no reconhecimento de padrões relacionados à forma (FELIPE; OLIOTI; TRAINA, 2005).

Como exposto anteriormente, o número de imagens no banco de dados utilizado é distinto para cada classe, fato esse que pode provocar um resultado tendencioso durante o treinamento, isto é, as lesões poderiam ser mais comumente classificadas como da classe que possui mais representantes; para evitar este problema, foi realizado o balanceamento das classes (SANTANA et al., 2018).

4.4 Seleção de atributos

Nem todos os atributos extraídos na etapa anterior são importantes para diferenciar as lesões mamárias. A finalidade desta etapa é retirar atributos que apresentam informações redundantes ou irrelevantes, visando aumentar o desempenho do classificador utilizado (ALBONICO, 2017).

Para o desenvolvimento desta fase foi utilizado o software Weka, na versão 3.8. A escolha foi motivada por ser uma ferramenta de código aberto para mineração de dados, de fácil acesso e interface amigável, que agrega um conjunto de algoritmos de classificação, regras de associação, regressão, pré-processamento, todos implementados em Java (WITTEN; FRANK; HALL, 2005; WITTEN; FRANK; HALL, 2011). Entretanto, o Weka utiliza como arquivo padrão para as tarefas de mineração o formato ARFF - Formato de Arquivo de Relação de Atributos (CLESIO, 2012). Desta forma, os atributos extraídos foram organizados em formato ARFF, no qual foi contabilizado um total de 169 atributos e 968 instâncias.

Os testes foram realizados de forma empírica, e exaustivamente, modificando os parâmetros de ambos algoritmos, a fim de analisar alterações no seu comportamento. As Tabelas 2 e 3 mostram os parâmetros para os AG e PSO, respectivamente.

Tabela 2 – Parâmetros dos Algoritmos Genéticos

Parâmetros	Valores
Geração	10 a 100
População	10 a 100
Taxa de Cruzamento	0.1 a 0.9
Taxa de Mutação	0.05 e 0.1
Operador de Seleção	Roleta

Fonte: A Autora

Tabela 3 – Parâmetros do Algoritmo de Otimização por Enxame de Partículas

Parâmetros	Valores
Peso individual	0.34
Peso de inércia	0.33
Peso social	0.33
Iterações	20/50/100/150/200
População	10 a 100

Fonte: A Autora

Como citado anteriormente, os parâmetros foram definidos empiricamente devido

não existir na literatura um padrão estabelecido, pois estes variam a cada caso. Entretanto, foram tomados, com base em estudos, valores mínimos e máximos a serem experimentados.

O operador de mutação fornece ao algoritmo um comportamento exploratório, já que o estimula a buscar novos pontos no espaço de busca. Então, se um algoritmo genético fosse desenvolvido baseando-se apenas em seleção e cruzamento, o sistema iria convergir prematuramente, já que o operador de cruzamento gera novos indivíduos de forma muito limitada após algumas gerações. Por isso, a mutação é fundamental para conservar a diversidade e renovar o material genético. Devido a mutação alterar a estrutura do cromossomo criando indivíduos com propriedades diferentes daquelas encontradas na maior parte da população, este parâmetro evita que o modelo fique preso a um ótimo local. Porém, por ser uma alteração bastante agressiva e, até certo ponto, imprevisível quanto aos resultados, a taxa de aplicação deste operador deve ser baixa (DIAS, 2006).

Em relação aos parâmetros da geração e da população, este foi o limite estabelecido devido à realização de testes aleatórios, visto que valores maiores não apresentaram resultados significativos. O operador de seleção adotado foi a roleta, por ter se mostrado mais efetivo em testes realizados aleatoriamente.

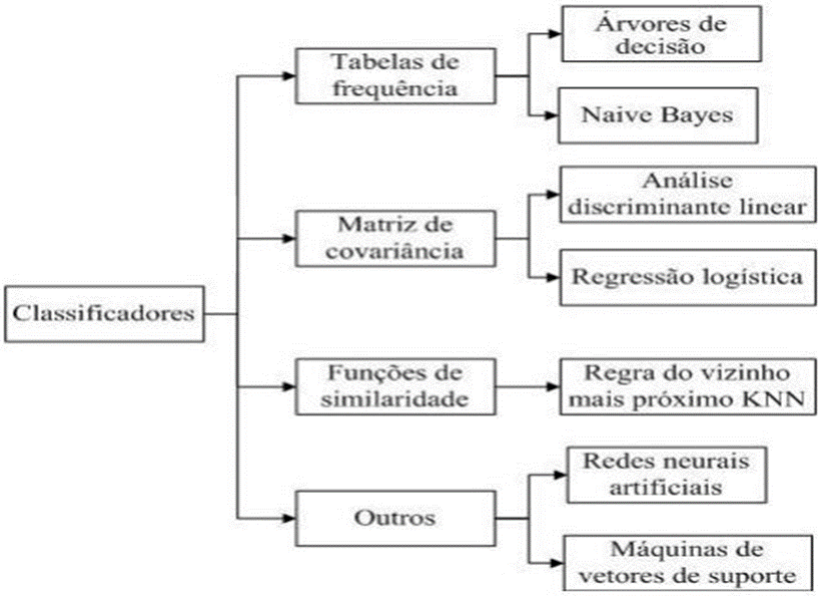
Quanto aos parâmetros do algoritmo PSO, os valores dos pesos não foram alterados devido a própria ferramenta Weka não permitir modificações, possibilitando alterar apenas os valores das iterações e população.

4.5 Classificação

Depois de selecionados os subconjuntos de atributos, os classificadores devem ser avaliados quanto ao desempenho, utilizando-se como medida a acurácia, o índice kappa e a matriz de confusão. Diversos classificadores podem ser utilizados para testar os conjuntos de atributos selecionados por algoritmos. A Figura 18 demonstra alguns destes classificadores clássicos, já consolidados e testados em diversos problemas de reconhecimento de padrões (TAN; STEINBACH; KUMAR, 2006).

Em ferramentas automáticas que executam a análise de imagens de lesões mamárias, foi observada uma maior utilização de quatro tipos de classificadores, sendo eles: análise discriminante linear, árvores de decisão, máquinas de vetores de suporte e redes neurais artificiais (ALBONICO, 2017). De acordo com Cheng et al. (2010), estas ferramentas são utilizadas principalmente devido à grande simplicidade de implementação e na obtenção de bons resultados. Entretanto, o desempenho dos classificadores está conectado principalmente com a natureza do banco de dados utilizados e não tão conectadas com os princípios das técnicas de classificação (EBERHART; SHI, 2007). A Figura 19 resume as principais vantagens e desvantagens destes classificadores (ALBONICO, 2017).

Figura 18 – Principais grupos de classificadores



Fonte: Santana et al. (2018)

Figura 19 – Principais vantagens e desvantagens dos classificadores

Classificador	Vantagens	Desvantagens
Análise discriminante linear	Método simples e rápido	Baixo desempenho em um conjunto de dados complexo
Árvores de decisão	Método simples e rápido	Desempenho depende do critério adotado para ramificação da árvore
Máquina de vetores de suporte	Alta capacidade de abstração; Fornece uma única solução otimizada;	Funciona apenas como método de aprendizado supervisionado;
Redes neurais artificiais	Alta capacidade de abstração; Pode ser utilizada como método de aprendizado supervisionado e não supervisionado;	Alto tempo de treinamento; Múltiplas soluções;

Fonte: Albonico (2017)

Diante do exposto, os classificadores utilizados nesta pesquisa foram: *NaiveBayes*, *BayesNet*, uma rede neural do tipo *Multilayer Perceptron* (MLP), máquinas de vetor de suporte (SVM) e árvores de decisão. Todos os classificadores citados foram utilizados na forma padrão que o software apresenta, exceto o MLP e o SVM. Para o MLP foram realizados testes sem camadas escondidas, com uma camada escondida e duas camadas escondidas, na qual cada camada possuía 100 neurônios. Com relação ao método SVM, o expoente da função kernel foi variado do 1 ao 9 e também foram realizados testes com kernel RBF. Por fim, as árvores de decisão utilizadas foram o J48, Random Tree e Random Forest. Nos experimentos realizados, os algoritmos foram executados 30 vezes nos subconjuntos selecionados, usando a técnica de validação cruzada com 10 folds.

5 Resultados e Discussão

Neste capítulo são exibidos os resultados utilizando os métodos descritos no capítulo anterior, nas etapas de seleção de atributos, classificação e avaliação. As etapas relacionadas com o pré-processamento, a segmentação e a extração de atributos foram realizadas em trabalhos passados.

Tendo em vista que foi realizada uma quantidade expressiva de testes, neste trabalho serão apresentados apenas os cincos resultados mais relevantes de cada grupo de teste. Serão expostos no Apêndice A os resultados descritos aqui, com todos os classificadores utilizados. Todos os testes feitos serão disponibilizados em um documento externo, na forma de um relatório técnico.

As Tabelas 4, 5 e 6 apresentam os resultados descritos acima com todos atributos e subconjuntos selecionados com AG e PSO, respectivamente. A melhor acurácia e o melhor índice kappa foram tomados como base para a seleção dos resultados. Acurácia é uma medida de avaliação do desempenho de um modelo de classificação, que mede a taxa de acerto global, ou seja, o número de classificações corretas dividido pelo número total de instâncias a serem classificadas (OLIVEIRA-JÚNIOR et al., 2017). O coeficiente de concordância de Kappa é utilizado para descrever a concordância entre duas ou mais classes quando realizam uma avaliação de uma mesma amostra (LANDIS; KOCH, 1977).

Tabela 4 – Classificação com todos os atributos

Classificador	Kernel	Acurácia (%)	Kappa	N° de Atributos
SVM	4	91.115	0,881	169
SVM	3	90.809	0,877	169
SVM	2	89.979	0,866	169
SVM	9	84.607	0,794	169
SVM	5	84.297	0,790	169

Fonte: A Autora

Diante do exposto, nota-se uma redução significativa no número de atributos, enquanto que não houve uma redução expressiva na acurácia, permanecendo em níveis similares aos obtidos com o uso de todos os atributos. O algoritmo genético se mostrou um pouco mais eficiente quando comparado ao PSO. Este fato não foi observado no estudo de Tavares, Nedjah e Mourelle (2015), no qual o PSO obteve melhor desempenho. Entretanto, na pesquisa de (SILVA-NETO, 2016), realizada apenas com o PSO para detecção de massas em imagens mamográficas, afirmou-se que o algoritmo obteve um desempenho

Tabela 5 – Classificação com subconjuntos de atributos selecionados com AG

Classificador	Kernel	Acurácia (%)	Kappa	N° de Atributos
SVM	5	87.082	0,827	57
SVM	5	86,883	0,825	66
SVM	4	86.359	0.818	57
SVM	4	85,954	0,825	73
SVM	5	85,848	0,811	81

Fonte: A Autora

Tabela 6 – Classificação com subconjuntos de atributos selecionados com PSO

Classificador	Kernel	Acurácia (%)	Kappa	N° de Atributos
SVM	5	86,157	0,815	60
SVM	4	85,950	0,812	60
SVM	6	85,743	0,809	60
SVM	5	84,504	0,793	56
SVM	5	84,297	0,790	75

Fonte: A Autora

positivo.

Como foi dito anteriormente, pode-se acreditar que quanto maior o número de atributos em um conjunto de dados, maior o poder discriminatório e, consequentemente, maior a acurácia do classificador, entretanto na prática esse comportamento não é verdadeiro (KOLLER; SAHAMI, 1996). Os testes mostraram uma acurácia maior em subconjuntos com menos atributos. Uma grande quantidade de atributos insignificantes pode fazer com que os algoritmos de aprendizagem tenham dificuldade em extrair informações que sejam realmente necessárias e relevantes para classificação. Além do mais, diversas pesquisas nessa área apontam que um número de atributos desnecessários pode introduzir ruídos nos dados, confundindo o algoritmo de aprendizagem e ocasionando erros na classificação (LIU; MOTODA, 1998).

O valor do coeficiente de concordância de Kappa pode variar de 0 a 1. Quanto mais próximo de 1 for seu valor, maior será a concordância entre as classes e quanto mais próximo de zero, maior é o indicativo de que a concordância é aleatória. De acordo com Landis e Koch (1977), quando o Kappa varia entre 0,61 e 1, significa que a concordância é forte. Em conformidade com a acurácia e o índice kappa, os resultados de um classificador podem ser representados por uma matriz de confusão, a qual mostra diretamente a quantidade de predições corretas e incorretas que o classificador executou (FAWCETT, 2006).

Nas Tabelas 7 e 8 estão apresentados os valores utilizados nos parâmetros do AG e PSO, respectivamente.

Tabela 7 – Parâmetros utilizados no AG

Acurácia (%)	Geração	População	Crossover	Mutação
87.082	50	60	0,5	0,05
86.883	50	30	0,3	0,05
86.359	50	60	0,5	0,05
85.954	50	20	0,3	0,05
85.848	50	30	0,1	0,05

Fonte: A Autora

Tabela 8 – Parâmetros utilizados no PSO

Acurácia (%)	Iteração	População	Mutação
86,157	200/150	20	0,05
85,950	200/150	20	0,05
85,743	200/150	20	0,05
84,504	200/150	20	0,05
84,297	200/100	10	0,05

Fonte: A Autora

No algoritmo genético, os testes realizados com a taxa de mutação de 0,05 obtiveram maior relevância do que com taxa de mutação de 0.1. Esse resultado está de acordo com a literatura, visto que Chambers (2019) aborda em seu estudo que deve ser evitada uma taxa de mutação muito alta, uma vez que esta pode tornar a busca essencialmente aleatória, prejudicando fortemente a convergência para uma solução ótima. Portanto, taxas abaixo de 0,1 são mais indicadas.

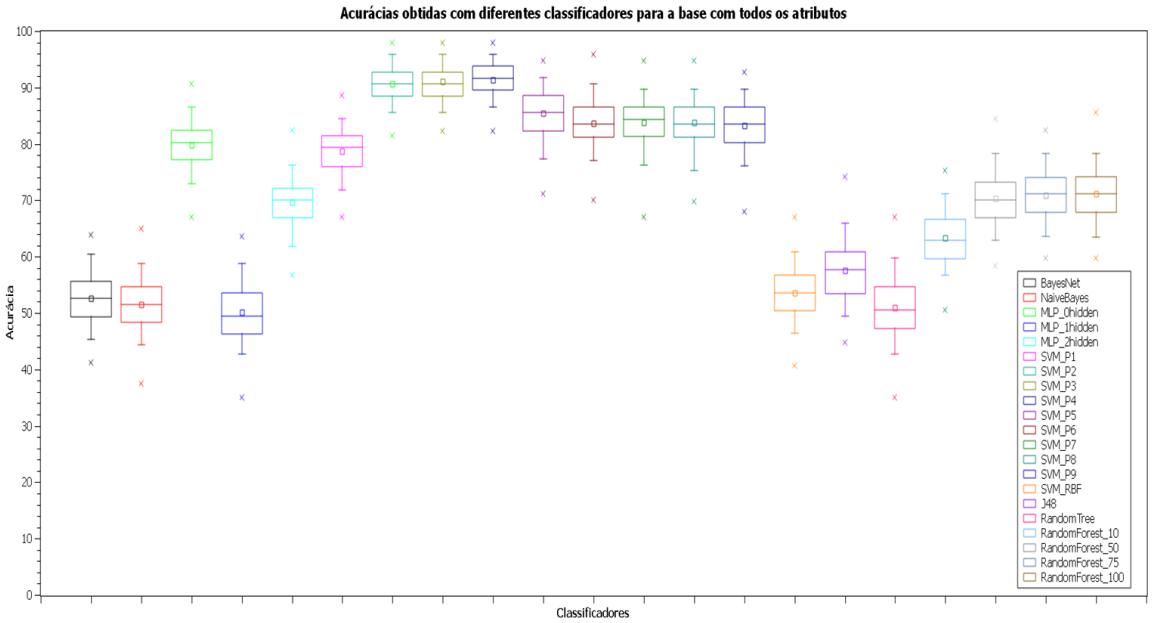
O valor das gerações neste artigo também corrobora com a literatura, mostrando que o valor de geração muito baixo causa uma queda no desempenho e um valor alto faz necessário um tempo maior de processamento, mas fornece uma melhor cobertura do domínio do problema, evitando a convergência para soluções locais. Portanto, deve-se buscar um ponto de equilíbrio no que diz respeito ao tamanho escolhido para geração (CHAMBERS, 2019).

Em relação ao tamanho da população estudos mostram que população muito pequena oferece uma cobertura inferior do espaço de busca, causando uma queda no desempenho. Uma população suficientemente grande fornece uma melhor cobertura do domínio do problema e previne a convergência prematura para soluções locais. Entretanto, com uma grande população tornam-se necessários maiores recursos computacionais (CHAMBERS, 2019).

Entre todos os classificadores testados, o SVM obteve o melhor desempenho. Este fato entra em concordância com os estudos de Gonçalves (2017), Borchartt et al. (2013), Acharya et al. (2012). Para analisar a variação das acurácias com os diversos classificadores

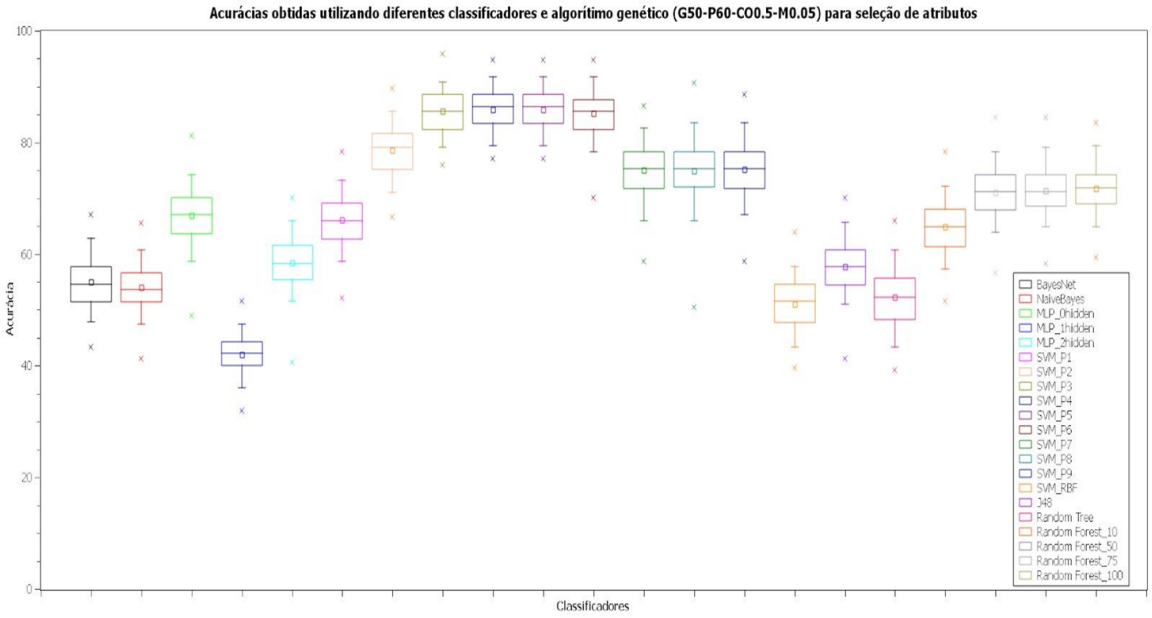
testados, foram gerados gráficos BloxPlots para avaliar a distribuição dos dados. Os gráficos das Figuras 20, 21 e 22 mostram os bloxplots das melhores classificações com todos os atributos, com os subconjuntos selecionados com AG e PSO, respectivamente.

Figura 20 – Bloxplots das classificações com todos os atributos



Fonte: A Autora

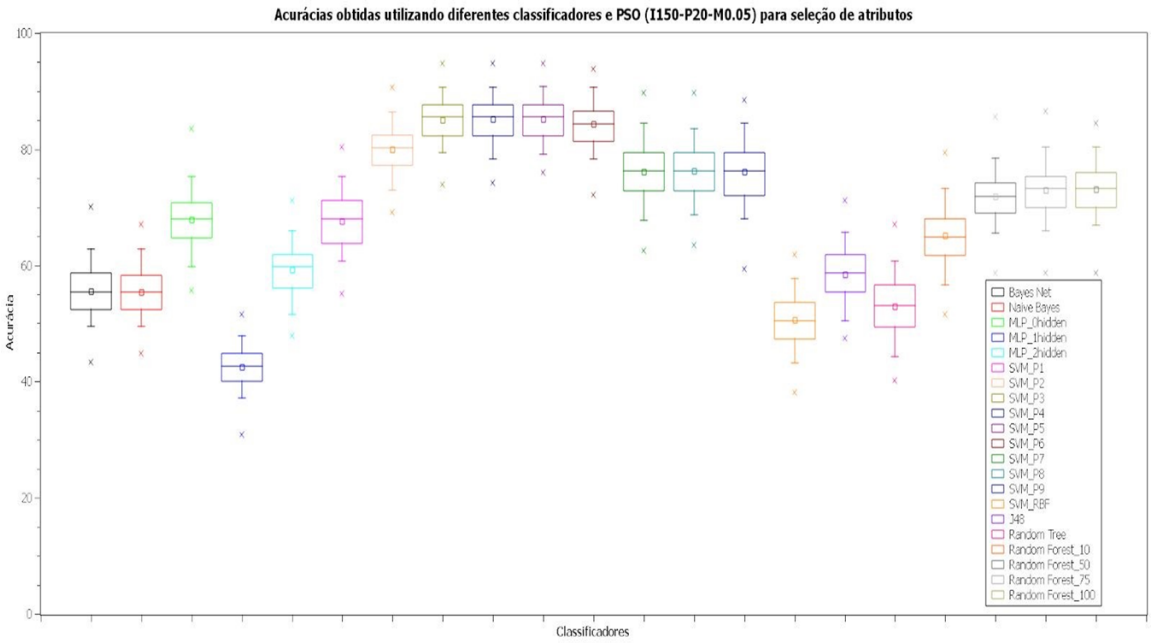
Figura 21 – Bloxplots das classificações dos subconjuntos selecionados com AG



Fonte: A Autora

Nas Figuras 23, 24 e 25 estão compiladas as matrizes de confusão dos melhores resultados de cada grupo. Ao lado direito da matriz, encontram-se as classes que foram

Figura 22 – Bloxplots das classificações dos subconjuntos selecionados com PSO



Fonte: A Autora

classificadas e cada uma delas é representada por uma letra. O lado esquerdo indica a classificação das instâncias propriamente dita.

Figura 23 – Matriz de confusão da classificação com todos atributos

```
  a   b   c   d   <-- classified as
225   8   6   3 |   a = Cisto
 23 187  17  15 |   b = Lesao_Benigna
 11  14 210   7 |   c = Lesao_Maligna
  6   7   8 221 |   d = Sem_Lesao
```

Fonte: A Autora

Figura 24 – Matriz de confusão da classificação com atributos selecionados com AG

```
  a   b   c   d   <-- classified as
222   8   6   6 |   a = Cisto
 24 186  19  13 |   b = Lesao_Benigna
 11  20 207   4 |   c = Lesao_Maligna
  9  11   3 219 |   d = Sem_Lesao
```

Fonte: A Autora

- Classe de Cisto: 232 instâncias foram classificadas corretamente como cisto, enquanto

Figura 25 – Matriz de confusão da classificação com atributos selecionados com PSO

```
  a   b   c   d   <-- classified as
222   8   6   6 |   a = Cisto
 24 186  19  13 |   b = Lesao_Benigna
 11  20 207   4 |   c = Lesao_Maligna
  9  11   3 219 |   d = Sem_Lesao
```

Fonte: A Autora

que 7 foram classificadas incorretamente como lesão benigna, 2 como lesão maligna e 1 como sem lesão;

- Classe de lesão benigna: 199 instâncias foram classificadas corretamente como lesão benigna, enquanto 25 foram classificadas incorretamente como cisto, 14 como lesão maligna e 4 sem lesão;
- Classe de lesão maligna: 225 instâncias foram classificadas corretamente como lesão maligna, enquanto 4 foram classificadas incorretamente como cisto, 9 como lesão benigna e 4 sem lesão;
- Classe sem lesão: 226 instâncias foram classificadas corretamente sem lesão, enquanto 8 foram classificadas incorretamente como cisto, 6 como lesão benigna, 2 como lesão maligna.
- Classe de Cisto: 225 instâncias foram classificadas corretamente como cisto, enquanto que 8 foram classificadas incorretamente como lesão benigna, 6 como lesão maligna e 3 como sem lesão;
- Classe de lesão benigna: 187 instâncias foram classificadas corretamente como lesão benigna, enquanto 23 foram classificadas incorretamente como cisto, 17 como lesão maligna e 15 sem lesão;
- Classe de lesão maligna: 210 instâncias foram classificadas corretamente como lesão maligna, enquanto 11 foram classificadas incorretamente como cisto, 14 como lesão benigna e 7 sem lesão;
- Classe sem lesão: 221 instâncias foram classificadas corretamente sem lesão, enquanto 6 foram classificadas incorretamente como cisto, 7 como lesão benigna, 8 como lesão maligna.
- Classe de Cisto: 222 instâncias foram classificadas corretamente como cisto, enquanto que 8 foram classificadas incorretamente como lesão benigna, 6 como lesão maligna e 6 como sem lesão;

- Classe de lesão benigna: 186 instâncias foram classificadas corretamente como lesão benigna, enquanto 24 foram classificadas incorretamente como cisto, 19 como lesão maligna e 13 sem lesão;
- Classe de lesão maligna: 207 instâncias foram classificadas corretamente como lesão maligna, enquanto 11 foram classificadas incorretamente como cisto, 20 como lesão benigna e 4 sem lesão;
- Classe sem lesão: 219 instâncias foram classificadas corretamente sem lesão, enquanto 9 foram classificadas incorretamente como cisto, 11 como lesão benigna, 3 como lesão maligna.

Na Tabela 9 são apresentadas as classificações corretas e incorretas das instâncias de cada grupo de teste.

Tabela 9 – Classificação das instâncias

Classificação das instâncias	Todos atributos	AG	PSO
Corretamente	882	843	834
Incorretamente	86	125	134

Fonte: A Autora

O banco de dados é composto por um total de 968 instâncias. Com a redução dos atributos, não houve uma redução significativa na correta classificação das instâncias, quando comparada com os testes realizados com todos os atributos.

6 Conclusão e Trabalhos Futuros

Com base nos resultados alcançados no presente trabalho, pode-se concluir que a termografia pode ser proposta como uma boa ferramenta no auxílio ao diagnóstico do câncer de mama. Ela mostra ser uma técnica simples, de fácil aplicação e de baixo custo que apresenta excelentes resultados quanto ao diagnóstico precoce do câncer de mama quando são usados classificadores estatísticos. Assim, coloca-se este procedimento em evidência quando comparados a outros procedimentos tradicionais de diagnóstico ao câncer de mama.

Em relação as técnicas de seleção de atributos, os resultados mostraram que a nossa abordagem foi positiva, a qual foi caracterizada por uma significativa redução na quantidade de atributos sem diminuição considerável na acurácia em relação a classificação com todos os atributos.

Este trabalho resultou numa publicação na revista Research on Biomedical Engineering, classificado como B1 no Qualis (SILVA et al., 2019).

Como trabalho futuro, é importante a utilização de outros algoritmos para selecionar atributos mais relevantes. É essencial também realizar um estudo específico para analisar as influências da presença da classe sem lesão nos resultados do classificador, e procurar entender e minimizar o efeito desta classe na diminuição do desempenho durante a classificação.

Referências

ACHARYA, U. R.; NG, E. Y.-K.; TAN, J.-H.; SREE, S. V. Thermography based breast cancer detection using texture features and support vector machine. *Journal of medical systems*, Springer, v. 36, n. 3, p. 1503–1510, 2012.

ALBONICO, G. A. M. *Seleção de atributos e classificação automática de lesões mamárias em imagens de ultrassom*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Elétrica e Computação, Universidade Estadual do Oeste do Paraná, 2017. Disponível em: <http://tede.unioeste.br/handle/tede/2998>.

ANDRADE, M. K.; SANTANA, M. A. de; MORENO, G.; OLIVEIRA, I.; SANTOS, J.; RODRIGUES, M. C. A.; SANTOS, W. P. dos. An EEG brain-computer interface to classify motor imagery signals. In: NAIK, G. R. (Ed.). *Biomedical Signal Processing*. [S.l.]: Springer, 2020. p. 83–98.

ANTANI, S.; KASTURI, R.; JAIN, R. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern recognition*, Elsevier, v. 35, n. 4, p. 945–965, 2002.

AZEVEDO, W. W.; LIMA, S. M.; FERNANDES, I. M.; ROCHA, A. D.; CORDEIRO, F. R.; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. Fuzzy morphological extreme learning machines to detect and classify masses in mammograms. In: IEEE. *2015 IEEE international conference on fuzzy systems (fuzz-IEEE)*. [S.l.], 2015. p. 1–8.

AZEVEDO, W. W.; LIMA, S. M.; FERNANDES, I. M.; ROCHA, A. D.; CORDEIRO, F. R.; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. Morphological extreme learning machines to detect and classify masses in mammograms. In: *The International Joint Conference on Neural Networks - IJCNN 2015*. [S.l.: s.n.], 2015.

BANDYOPADHYAY, S. K. Survey on segmentation methods for locating masses in a mammogram image. *International Journal of Computer Applications*, v. 9, n. 11, p. 25–28, 2010.

BARBOSA, V. A.; RIBEIRO, R. R.; FEITOSA, A. R.; SILVA, V. L.; ROCHA, A. D.; FREITAS, R. C.; SOUZA, R. E.; SANTOS, W. P. Reconstruction of electrical impedance tomography using fish school search, non-blind search, and genetic algorithm. *International Journal of Swarm Intelligence Research (IJSIR)*, IGI Global, v. 8, n. 2, p. 17–33, 2017.

BARBOSA, V. A. F.; SANTANA, M. A.; ANDRADE, M. K. S.; LIMA, R. C. F.; SANTOS, W. P. Deep-Wavelet Neural Networks for breast cancer early diagnosis using mammary termographies. In: DAS, H.; PRADHAN, C.; DEY, N. (Ed.). *Deep Learning for Data Analytics: Foundations, Biomedical Applications, and Challenges*. [S.l.]: Elsevier, 2020.

BORCHARTT, T. B.; CONCI, A.; LIMA, R. C.; RESMINI, R.; SANCHEZ, A. Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, Elsevier, v. 93, n. 10, p. 2785–2803, 2013.

BRASIL. *Falando sobre câncer de mama*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2002. <http://www.saude.pb.gov.br/web_data/saude/cancer/aula11.pdf>. [Acesso em: 19 jan. 2019].

BRASIL. *ABC do Câncer: abordagens básicas para o controle do câncer*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2011. <http://bvsms.saude.gov.br/bvs/publicacoes/inca/abc_do_cancer_2ed.pdf>. [Acesso em: 01 dez. 2018].

BRASIL. *Atualização em mamografia para técnicos em radiologia*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2018. <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//atualizacao-em-mamografia-tecnicos-radiologia.pdf>. [Acesso em: 25 jan. 2019].

BRASIL. *Câncer de mama: detecção precoce*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2018. <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/deteccao_precoce>. [Acesso em: 17 jan. 2019].

BRASIL. *Como é o processo de carcinogênese*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2018. <http://www1.inca.gov.br/conteudo_view.asp?id=319>. [Acesso em: 20 dez. 2018].

BRASIL. *Determinantes Sociais e Riscos para a Saúde, Doenças Crônicas não transmissíveis e Saúde Mental*. [S.l.]: Organização Pan-Americana de Saúde, Organização Mundial de Saúde (Comp.), 2018. <https://www.paho.org/bra/index.php?option=com_content&view=article&id=5588:folha-informativa-cancer&Itemid=839>. [Acesso em: 30 jan. 2019].

BRASIL. *Estimativa 2018: Incidência de Câncer no Brasil*. [S.l.]: Instituto Nacional de Câncer, Ministério da Saúde, 2018. <https://www.inca.gov.br/publicacoes/livros/estimativa-2018-incidencia-de-cancer-no-brasil>. [Acesso em: 29 jan. 2019].

BRYNOLFSSON, P.; NILSSON, D.; TORHEIM, T.; ASKLUND, T.; KARLSSON, C. T.; TRYGG, J.; NYHOLM, T.; GARPEBRING, A. Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. *Scientific Reports*, Nature Publishing Group, v. 7, n. 1, p. 1–11, 2017.

CASTRO, E. G.; TSUZUKI, M. S. G. Simulation optimization using swarm intelligence as tool for cooperation strategy design in 3D predator-prey game. In: CHAN, F. T. S.; TIWARI, M. K. (Ed.). *Swarm intelligence: focus on ant and particle swarm optimization*. Vienna: InTech Education and Publishing, 2007.

CASTRO, P. A. D. de; SANTORO, D. M.; CAMARGO, H. A.; NICOLETTI, M. C. Improving a Pittsburgh learnt fuzzy rule base using feature subset selection. In: IEEE. *Fourth International Conference on Hybrid Intelligent Systems (HIS'04)*. [S.l.], 2004. p. 180–185.

CHAMBERS, L. D. *Practical handbook of genetic algorithms: complex coding systems*. [S.l.]: CRC press, 2019. v. 3.

CHENG, H.-D.; SHAN, J.; JU, W.; GUO, Y.; ZHANG, L. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, Elsevier, v. 43, n. 1, p. 299–317, 2010.

CLESIO, F. *Preparando arquivos para o Weka 2012*. 2012. <https://mineracaodedados.wordpress.com/2012/02/27/preparando-arquivos-para-o-weka>. [Acesso em: 01 fev. 2019].

COMMOWICK, O.; ISTACE, A.; KAIN, M.; LAURENT, B.; LERAY, F.; SIMON, M.; POP, S. C.; GIRARD, P.; AMELI, R.; FERRÉ, J.-C. et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific Reports*, Nature Publishing Group, v. 8, n. 1, p. 1–17, 2018.

CORDEIRO, F. R.; BEZERRA, K. F. P.; SANTOS, W. P. dos. Random walker with fuzzy initialization applied to segment masses in mammography images. In: IEEE. *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. [S.l.], 2017. p. 156–161.

CORDEIRO, F. R.; LIMA, S. M.; SILVA-FILHO, A. G.; SANTOS, W. P. dos. Segmentation of mammography by applying extreme learning machine in tumor detection. In: SPRINGER. *International Conference on Intelligent Data Engineering and Automated Learning*. [S.l.], 2012. p. 92–100.

CORDEIRO, F. R.; SANTOS, W. P.; SILVA-FILHO, A. G. An adaptive semi-supervised Fuzzy GrowCut algorithm to segment masses of regions of interest of mammographic images. *Applied Soft Computing*, Elsevier, v. 46, p. 613–628, 2016.

CORDEIRO, F. R.; SANTOS, W. P. dos; SILVA-FILHO, A. G. Segmentation of mammography by applying GrowCut for mass detection. *Studies in Health Technology and Informatics*, v. 192, p. 87–91, 2013.

CORDEIRO, F. R.; SANTOS, W. P. dos; SILVA-FILHO, A. G. Segmentation of mammography by applying GrowCut for mass detection. In: *14th World Congress on Medical and Health Informatics - MEDINFO 2013*. Copenhagen: [s.n.], 2013.

CORDEIRO, F. R.; SANTOS, W. P. dos; SILVA-FILHO, A. G. A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images. *Expert Systems with Applications*, Elsevier, v. 65, p. 116–126, 2016.

CORDEIRO, F. R.; SANTOS, W. P. dos; SILVA-FILHO, A. G. Analysis of supervised and semi-supervised GrowCut applied to segmentation of masses in mammography images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, v. 5, n. 4, p. 297–315, 2017.

DIAS, D. M. *Aplicação de algoritmos genéticos no scheduling automático e otimizado do petróleo em refinarias*. Rio de Janeiro: [s.n.], 2006. Relatório de Estudo Orientado, Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio.

DOURADO-NETO, H. M. *Segmentação e análise automática de termogramas: um método auxiliar na detecção do câncer de mama*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Pernambuco, Recife, 2014.

EBERHART, R.; SHI, Y. *Computational Intelligence: concepts to implementations*. [S.l.]: Morgan Kaufmann, 2007.

ENGELBRECHT, A. *Fundamentals of Computational Swarm Intelligence*. [S.l.]: John Wiley & Sons, 2005.

ERPEN, L. R. C. *Reconhecimento de padrões em imagens por descritores de forma*. Dissertação (Mestrado) — Programa de Pós-Graduação em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.

FEITOSA, A. R.; RIBEIRO, R. R.; BARBOSA, V. A.; SOUZA, R. E. de; SANTOS, W. P. dos. Reconstruction of electrical impedance tomography images using chaotic ring-topology particle swarm optimization and non-blind search. In: IEEE. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2014. p. 2618–2623.

FEITOSA, A. R.; RIBEIRO, R. R.; BARBOSA, V. A.; SOUZA, R. E. de; SANTOS, W. P. dos. Reconstruction of electrical impedance tomography images using particle swarm optimization, genetic algorithms and non-blind search. In: IEEE. *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*. [S.l.], 2014. p. 1–6.

FELIPE, J. C.; OLIOTI, J. B.; TRAINA, A. J. Discriminação de Aspectos Malignos em Massas Tumorais de Mamografias Usando Características de Forma das Imagens. In: *V Workshop de Informática Médica (WIM 2005), Porto Alegre, RS, Brazil*. [S.l.: s.n.], 2005.

FERREIRA, J. O.; OLIVEIRA, H. C. B.; MARTINEZ, M. Aplicação de uma metodologia computacional inteligente no diagnóstico de lesões cancerígenas. *Revista Brasileira de Inovação Tecnológica em Saúde*, v. 2, n. 1, p. 4–9, 2011.

FERREIRA, R. *Bates, Darwin, Wallace e a teoria da evolução*. [S.l.]: Editora Uinversidade de Brasilia, 1990.

FRANCISCANI, J. F.; QUEIROZ, D. M. Inteligência artificial: uma abordagem sobre algoritmos genéticos. *Revista Rumos da Pesquisa em Ciências Empresariais, Ciências do Estado e Tecnologia. Cadernos de Sistemas de Informação*, v. 2, n. 1, p. 172–189, 2011.

FRANCO, J. M. *Mastologia: Formação do Especialista*. Rio de Janeiro: Ateneu, 1997.

GAMBINO, O.; CONTI, V.; GALDINO, S.; VALENTI, C. F.; SANTOS, W. P. dos. Image Segmentation Techniques for Healthcare Systems. *Journal of Healthcare Engineering*, v. 2019, p. 1–2, 2019.

GONÇALVES, C. B. *Detecção de câncer de mama utilizando imagens termográficas*. Uberlândia: [s.n.], 2017. Trabalho de Conclusão de Curso, Ciência da Computação, Universidade Federal de Uberlândia.

HARALICK, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE*, IEEE, v. 67, n. 5, p. 786–804, 1979.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, n. 6, p. 610–621, 1973.

HARRIS, J. R.; LIPPMAN, M. E.; MORROW, M.; HELMAN, S. *Diseases of the Breast*. Philadelphia: Lippincott-Raven Publishers, 1996.

HOLLAND, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. [S.l.]: MIT press, 1992.

HSE, H.; NEWTON, A. R. Sketched symbol recognition using zernike moments. In: IEEE. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* [S.l.], 2004. v. 1, p. 367–370.

KANDLIKAR, S. G.; PEREZ-RAYA, I.; RAGHUPATHI, P. A.; GONZALEZ-HERNANDEZ, J.-L.; DABYDEEN, D.; MEDEIROS, L.; PHATAK, P. Infrared imaging technology for breast cancer detection: Current status, protocols and new directions. *International Journal of Heat and Mass Transfer*, Elsevier, v. 108, p. 2303–2320, 2017.

KEYSERLINGK, J.; AHLGREN, P.; YU, E.; BELLIVEAU, N. Infrared Imaging of the Breast: Initial Reappraisal Using High-Resolution Digital Technology in 100 Successive Cases of Stage I and II Breast Cancer. *The Breast Journal*, Wiley Online Library, v. 4, n. 4, p. 245–251, 1998.

KHOTANZAD, A.; HONG, Y. H. Invariant image recognition by Zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 12, n. 5, p. 489–497, 1990.

KOHAVI, R.; JOHN, G. H. et al. Wrappers for feature subset selection. *Artificial intelligence*, Elsevier Science, v. 97, n. 1-2, p. 273–324, 1997.

KOLLER, D.; SAHAMI, M. Toward optimal feature selection. In: MORGAN KAUFMANN PUBLISHERS INC. *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. [S.l.], 1996. p. 284–292.

KOOP, C. Health and health care for the 21st century: For all the people. *American Journal of Public Health*, American Public Health Association, v. 96, n. 12, p. 2090–2092, 2006.

LABADESSA, L. M. *Mamografia ou Ultrassom de Mama?* 2019. <http://www.documenta.com.br/noticias_show.php?id=94>. [Acesso em: 02 fev. 2019].

LANDIS, G.; KOCH, G. A medição do acordo de observador para dados categóricos. *Biometria*, v. 33, n. 1, p. 159–174, 1977.

LELES, A. C. Q.; GUIMARãES, G.; ARAúJO, A. C.; SOUZA, H. A. *Desenvolvimento de procedimento e análise de imagens térmicas para a identificação do câncer de mama*. Uberlândia: [s.n.], 2015. Universidade Federal de Uberlândia.

LERSKI, R.; BARNETT, E.; MORLEY, P.; MILLS, P.; WATKINSON, G.; MACSWEEN, R. Computer analysis of ultrasonic signals in diffuse liver disease. *Ultrasound in Medicine & Biology*, Elsevier, v. 5, n. 4, p. 341–343, 1979.

LESSA, V.; MARENGONI, M. Applying artificial neural network for the classification of breast cancer using infrared thermographic images. In: SPRINGER. *International Conference on Computer Vision and Graphics*. [S.l.], 2016. p. 429–438.

LIMA, S. M.; AZEVEDO, W. W.; CORDEIRO, F. R.; SILVA-FILHO, A. G.; SANTOS, W. P. Feature extraction employing fuzzy-morphological decomposition for detection and classification of mass on mammograms. In: *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society - EMBC 2015*. [S.l.: s.n.], 2015. v. 2015, p. 801–804.

LIMA, S. M. de; SILVA-FILHO, A. G. da; SANTOS, W. P. dos. A methodology for classification of lesions in mammographies using Zernike Moments, ELM and SVM Neural Networks in a multi-kernel approach. In: IEEE. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2014. p. 988–991.

LIMA, S. M. L.; FILHO, A. G. S.; SANTOS, W. P. Morphological decomposition to detect and classify lesions in mammograms. In: SANTOS, W. P. dos; SANTANA, M. A. de; SILVA, W. W. A. da (Ed.). *Understanding a Cancer Diagnosis*. 1. ed. New York: Nova Science, 2020. p. 27–64.

LINDEN, R. *Algoritmos genéticos: uma importante ferramenta da inteligência computacional*. [S.l.]: Brasport, 2006.

LIU, H.; MOTODA, H. *Feature Selection for Knowledge Discovery and Data Mining*. [S.l.]: Kluwer Academic Publishers, 1998.

LIU, Y.; ZHOU, X. A simple texture descriptor for texture retrieval. In: IEEE. *International Conference on Communication Technology Proceedings, 2003. ICCT 2003.* [S.l.], 2003. v. 2, p. 1662–1665.

MADHU, H.; KAKILETI, S. T.; VENKATARAMANI, K.; JABBIREDDY, S. Extraction of medically interpretable features for classification of malignancy in breast thermography. In: IEEE. *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [S.l.], 2016. p. 1062–1065.

MARQUES, J. *Reconhecimento de Padroes: Métodos Estatısticos e Neuronais, Ensino da Ciência e da Tecnologia*. Lisboa: IST Press, 1999.

MATERKA, A. Texture analysis methodologies for magnetic resonance imaging. *Dialogues in Clinical Neuroscience*, Les Laboratoires Servier, v. 6, n. 2, p. 243, 2004.

MATOS, R. D. *Utilização de Algoritmo Genético para Resolução do Problema de Geração de Horários*. 2011. <http://www2.dc.uel.br/nourau/document/?view=590>. [Acesso em: 02 fev. 2019].

MERWE, D. W. van der; ENGELBRECHT, A. P. Data clustering using particle swarm optimization. In: CIS-IEEE. *IEEE Congress on Evolutionary Computation*. Canberra, Australia, 2003. p. 185–191.

MICHIE, D.; SPIEGELHALTER, D. J.; TAYLOR, C. et al. Machine learning. *Neural and Statistical Classification*, Technometrics, v. 13, n. 1994, p. 1–298, 1994.

MITCHELL, M. *An introduction to genetic algorithms*. [S.l.]: MIT press, 1998.

NASCIMENTO, F. A. F.; DIAS, A. N.; FILHO, A. F.; ARCE, J. E.; MIRANDA, G. M. Uso da Meta-Heurística otimização por exame de partículas no planejamento Florestal. *Scientia Forestalis*, v. 40, n. 96, p. 557–565, 2012.

O'CONNOR, J. P. B.; ROSE, C. J.; WATERTON, J. C.; CARANO, R. A.; PARKER, C. J.; JACKSON, A. Imaging intratumor heterogeneity: Role in therapy response, resistance, and clinical outcome. *Clinical Cancer Research*, v. 21, p. 249–257, 2015.

OLIVEIRA-JÚNIOR, J. G.; NORONHA, R. V.; KAESTNER, C.; ALVES, A. Métodos de seleção de atributos aplicados na previsão da evasão de cursos de graduação. *Revista de Informática Aplicada*, v. 13, n. 2, p. 54–67, 2017.

OLIVEIRA, L. F.; NARLOCH, A. L. M.; KIST, D. M.; SOARES, M.; MENEGHELLO, G. E.; CAVALHEIRO, G. G. H.; TILLMANN, M. A. A. Extração de Características de Forma utilizando matriz de co-ocorrência e atributos de Haralick. In: *Workshop de Visão Computacional (WVC), Programa de Pós-graduação em Ciência e Tecnologia de Sementes, Universidade Federal de Pelotas*. [S.l.: s.n.], 2012.

OLIVEIRA, P. M. A.; SOUZA, G. M.; SANTANA, M. A.; SILVA, G. S. L. E.; SILVA, W. W. A.; SANTOS, W. P. Uso de classificadores na predição de lesões de mama a partir de imagens termográficas. In: *Simpósio de Inovação em Engenharia Biomédica - SABIO 2019*. Recife: BioTech Consultoria, 2019.

PEDRINI, H.; SCHWARTZ, W. R. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. São Paulo: Thomson Learning, 2008.

PEREIRA, J. M. S.; SANTANA, M. A.; LIMA, R. C. F.; SANTOS, W. P. Lesion detection in breast thermography using machine learning algorithms without previous segmentation. In: SANTOS, W. P. dos; SANTANA, M. A. de; SILVA, W. W. A. da (Ed.). *Understanding a Cancer Diagnosis*. 1. ed. New York: Nova Science, 2020. p. 81–94.

PEREIRA, J. M. S.; SANTANA, M. A.; LIMA, R. C. F.; LIMA, S. M. L.; SANTOS, W. P. Method for classification of breast lesions in thermographic images using ELM classifiers. In: SANTOS, W. P. dos; SANTANA, M. A. de; SILVA, W. W. A. da (Ed.). *Understanding a Cancer Diagnosis*. 1. ed. New York: Nova Science, 2020. p. 117–132.

PEREIRA, J. M. S.; SANTANA, M. A.; SILVA, W. W. A.; LIMA, R. C. F.; LIMA, S. M. L.; SANTOS, W. P. Dialectical optimization method as a feature selection tool for breast cancer diagnosis using thermographic images. In: SANTOS, W. P. dos; SANTANA, M. A. de; SILVA, W. W. A. da (Ed.). *Understanding a Cancer Diagnosis*. 1. ed. New York: Nova Science, 2020. p. 95–118.

PORTER, P. “Westernizing” women’s risks? Breast cancer in lower-income countries. *New England Journal of Medicine*, v. 358, p. 213–6, 2008.

POZO, A. *Grupo de Pesquisas em Computação Evolutiva Aurora*. [S.l.]: Departamento de Informática, Universidade Federal do Paraná, 2019. <http://www.inf.ufpr.br/aurora/tutoriais/Ceapostila.pdf>. [Acesso em: 22 jan. 2019].

RIBEIRO, R. R.; FEITOSA, A. R.; SOUZA, R. E. de; SANTOS, W. P. dos. A modified differential evolution algorithm for the reconstruction of electrical impedance tomography images. In: IEEE. *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*. [S.l.], 2014. p. 1–6.

RIBEIRO, R. R.; FEITOSA, A. R.; SOUZA, R. E. de; SANTOS, W. P. dos. Reconstruction of electrical impedance tomography images using genetic algorithms and non-blind search. In: IEEE. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. [S.l.], 2014. p. 153–156.

RIEBER, A.; MERKLE, E.; BÖHM, W.; BRAMBS, H.-J.; TOMCZAK, R. MRI of histologically confirmed mammary carcinoma: clinical relevance of diagnostic procedures for detection of multifocal or contralateral secondary carcinoma. *Journal of Computer Assisted Tomography*, LWW, v. 21, n. 5, p. 773–779, 1997.

RIZZI, M.; FURTADO, J. C.; COSTA, A. B.; GERBASE, A. E.; FERRãO, M. F. Método do Enxame de Partículas para otimização de modelos de regressão multivariada empregados na determinação de biodiesel em blendas biodiesel / óleo vegetal / diesel. *Revista Virtual de Química*, v. 8, n. 6, p. 1877–1892, 2016.

RUI, Y.; HUANG, T.; CHANG, S. Image retrieval: past, present, and future. *journal of Visual Communication and Image Representation*, v. 10, p. 1–23, 1999.

SANTANA, M. A.; PEREIRA, J. M. S.; LIMA, R. C. F.; SANTOS, W. P. Breast lesions classification in frontal thermographic images using intelligent systems and moments of Haralick and Zernike. In: SANTOS, W. P. dos; SANTANA, M. A. de; SILVA, W. W. A. da (Ed.). *Understanding a Cancer Diagnosis*. 1. ed. New York: Nova Science, 2020. p. 65–80.

SANTANA, M. A. d.; PEREIRA, J. M. S.; SILVA, F. L. d.; LIMA, N. M. d.; SOUSA, F. N. d.; ARRUDA, G. M. S. d.; LIMA, R. d. C. F. d.; SILVA, W. W. A. d.; SANTOS, W. P. d. Breast cancer diagnosis based on mammary thermography and extreme learning machines. *Research on Biomedical Engineering*, SciELO Brasil, v. 34, n. 1, p. 45–53, 2018.

SANTOS, J. Engenharia na Medicina. *Gazeta Médica*, v. 3, n. 2, 2018.

SANTOS, W. P.; ASSIS, F. M.; SOUZA, R. E. MRI Segmentation using Dialectical Optimization. In: EMBS-IEEE. *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Minneapolis, USA, 2009. p. 5752–5755.

SANTOS, W. P.; ASSIS, F. M.; SOUZA, R. E.; SANTOS-FILHO, P. B.; LIMA-NETO, F. B. Dialectical Multispectral Classification of Diffusion-Weighted Magnetic Resonance Images as an Alternative to Apparent Diffusion Coefficients Maps to Perform Anatomical Analysis. *Computerized Medical Imaging and Graphics*, v. 33, n. 6, p. 442–460, 2009.

SANTOS, W. P.; SOUZA, R. E.; SILVA, A. F. D.; FILHO, P. B. S. Evaluation of Alzheimer's disease by analysis of MR images using multilayer perceptrons and committee machines. *Computerized Medical Imaging and Graphics*, v. 32, n. 1, p. 17–21, 2008.

SILVA, A. L. R.; SANTANA, M. A. de; AZEVEDO, W. W.; BEZERRA, R. S.; BARBOSA, V. A.; LIMA, R. C. de; SANTOS, W. P. dos. Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Research on Biomedical Engineering*, Springer, v. 35, n. 3, p. 213–222, 2019.

SILVA-JÚNIOR, M. da; FREITAS, R. C. de; SANTOS, W. P. dos; SILVA, W. W. A. da; RODRIGUES, M. C. A.; CONDE, E. F. Q. Exploratory study of the effect of binaural beat stimulation on the EEG activity pattern in resting state using artificial neural networks. *Cognitive Systems Research*, Elsevier, v. 54, p. 1–20, 2019.

SILVA-JÚNIOR, M. da; FREITAS, R. C. de; SILVA, W. W. A. da; RODRIGUES, M. C. A.; CONDE, E. F. Q.; SANTOS, W. P. dos. Using Artificial Neural Networks on Multi-channel EEG Data to Detect the Effect of Binaural Stimuli in Resting State. In: NAIK, G. R. (Ed.). *Biomedical Signal Processing*. [S.l.]: Springer, 2020. p. 99–136.

SILVA-NETO, O. P. d. *Detecção automática de massas em imagens mamográficas usando Particle Swarm Optimization (PSO) e índice de diversidade funcional*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia de Eletricidade, 2016. Disponível em: <http://tedebc.ufma.br:8080/jspui/handle/tede/298>.

SILVA, W. W. A.; SANTANA, M. A.; FILHO, A. G. S.; LIMA, S. M. L.; SANTOS, W. P. Morphological Extreme Learning Machines applied to the detection and classification of mammary lesions. In: GANDHI, T. K.; BHATTACHARYYA, S.; DE, S.; KONAR, D.; DEY, S. (Ed.). *Advanced Machine Vision Paradigms for Medical Image Analysis*. London: Elsevier, 2020.

SINGH, C.; MITTAL, N.; WALIA, E. Face recognition using Zernike and complex Zernike moment features. *Pattern Recognition and Image Analysis*, Springer, v. 21, n. 1, p. 71–81, 2011.

SKORTON, D. J.; COLLINS, S.; WOSKOFF, S.; BEAN, J. A.; JR, H. M. Range-and azimuth-dependent variability of image texture in two-dimensional echocardiograms. *Circulation*, American Heart Association, v. 68, n. 4, p. 834–840, 1983.

SMELTZER, S. C.; BARE, B. G. *Brunner & Suddarth: tratado de enfermagem médico-cirúrgica*. 13. ed. Rio de Janeiro: Guanabara Koogan, 2015.

SOUZA, T. K. S.; ANDRADE, J. F. S.; ALMEIDA, M. B. J.; SANTANA, M. A.; SANTOS, W. P. Métodos computacionais aplicados ao diagnóstico de câncer de mama por termografia: uma revisão da literatura. In: *Simpósio de Inovação em Engenharia Biomédica - SABIO 2019*. Recife: BioTech Consultoria, 2019.

TAHMASBI, A.; SAKI, F.; SHOKOUHI, S. B. Classification of benign and malignant masses based on Zernike moments. *Computers in Biology and Medicine*, Elsevier, v. 41, n. 8, p. 726–735, 2011.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining*. [S.l.]: Pearson Education India, 2006.

TAVARES, Y. M.; NEDJAH, N.; MOURELLE, L. de M. Utilização de otimização por enxame de partículas e algoritmos genéticos em rastreamento de padrões. In: *12 Congresso Brasileiro de Inteligência Computacional*. [S.l.: s.n.], 2015.

VASCONCELOS, J. H. de; SANTOS, W. P. dos; LIMA, R. d. C. F. de. Analysis of methods of classification of breast thermographic images to determine their viability in the early breast cancer detection. *IEEE Latin America Transactions*, IEEE, v. 16, n. 6, p. 1631–1637, 2018.

VICTORA, C. G.; BARRETO, M. L.; LEAL, M. do C.; MONTEIRO, C. A.; SCHMIDT, M. I.; PAIM, J.; BASTOS, F. I.; ALMEIDA, C.; BAHIA, L.; TRAVASSOS, C. et al. Health conditions and health-policy innovations in Brazil: the way forward. *The Lancet*, Elsevier, v. 377, n. 9782, p. 2042–2053, 2011.

VILA-NOVA, R. L. *Uso de imagens termográficas de mama para análise de patologias através da comparação entre diversos classificadores estatísticos*. Dissertação (Mestrado) — Programa de Pós-Graduação em Engenharia Mecânica, Universidade Federal de Pernambuco, Recife, 2017.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical machine learning tools and techniques*. 2. ed. San Francisco: Morgan Kaufmann Publishers, 2005.

WITTEN, I. H.; FRANK, E.; HALL, M. A. *Data Mining: Practical machine learning tools and techniques*. 3. ed. San Francisco: Morgan Kaufmann Publishers, 2011.

APÊNDICE A – Testes com todos os classificadores dos melhores resultados

Tabela 10 – Classificação com todos os 169 atributos

Classificador	Acurácia (%)	Kappa
BayesNet	52,7893	0,3705
NaiveBayes**	52,7562	0,3567
0 camada	80,2686	0,7369
MLP 1 camada	86,2603	0,8168
2 camadas	84,0909	0,7879
P = 1	78,7190	0,7163
P = 2	89,9793	0,8664
P = 3	90,8058	0,8774
SVM P = 4*	91,1157	0,8815
P = 5	84,2975	0,7906
P = 6	82,6446	0,7686
P = 7	84,4008	0,7920
P = 8	84,2975	0,7906
P = 9	84,6074	0,7948
RBF	53,8223	0,3843
J48	58,2645	0,4435
RandomTree	53,8223	0,3843
RandomForest	72,7273	0,6364

Fonte: A Autora

Tabela 11 – Classificação com 57 atributos selecionados com AG

Classificador		Acurácia (%)	Kappa
BayesNet		55,6765	0,4089
NaiveBayes		53,8165	0,3841
MLP	0 camada	65,9020	0,5452
	1 camada	41,4218	0,2195
	2 camadas	56,0835	0,4142
SVM	P = 1	66,1114	0,5480
	P = 2	78,8198	0,7175
	P = 3	85,6389	0,8085
	P = 4	86,3595	0,8181
	P = 5	87,0822	0,8277
	P = 6	85,4306	0,8057
	P = 7	74,2654	0,6567
	P = 8	75,9213	0,6788
	P = 9	77,7870	0,7037
	RBF	50,0980	0,3347
J48		58,5738	0,4474
RandomTree		52,2830	0,3636
RandomForest		73,8606	0,6514

Fonte: A Autora

Tabela 12 – Classificação com 60 atributos selecionados com PSO

Classificador		Acurácia (%)	Kappa
BayesNet		55,2686	0,4036
NaiveBayes		55,2686	0,4036
MLP	0 camada	62,5310	0,4562
	1 camada	39,5398	0,3195
	2 camadas	53,1238	0,7162
SVM	P = 1	68,2851	0,5771
	P = 2	59,7107	0,4628
	P = 3	85,7438	0,8099
	P = 4	85,9504	0,8127
	P = 5	86,1570	0,8154
	P = 6	84,8140	0,7975
	P = 7	77,8926	0,7052
	P = 8	75,1033	0,6680
	P = 9	76,0331	0,6804
	RBF	50,7231	0,3430
J48		57,3347	0,4311
RandomTree		54,2355	0,3898
RandomForest		72,5207	0,6336

Fonte: A Autora