

**UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS**

Pedro Henrique Leal Silva

**SEGMENTAÇÃO DE IMAGENS MAMOGRÁFICAS
PARA IDENTIFICAÇÃO DE MICROCALCIFICAÇÕES
ATRAVÉS DO ALGORITMO FIREFLY**

São Carlos

2025

Pedro Henrique Leal Silva

**SEGMENTAÇÃO DE IMAGENS MAMOGRÁFICAS
PARA IDENTIFICAÇÃO DE MICROCALCIFICAÇÕES
ATRAVÉS DO ALGORITMO FIREFLY**

Monografia apresentada ao Curso de Engenharia Elétrica - Ênfase em Eletrônica, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Engenheira Eletricista.

Área de concentração: Engenharia Elétrica

Orientador: Prof. Homero Schiabel

São Carlos

2025

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

SS586s Silva, Pedro Henrique Leal
SEGMENTAÇÃO DE IMAGENS MAMOGRÁFICAS PARA
IDENTIFICAÇÃO DE MICROCALCIFICAÇÕES ATRAVÉS DO
ALGORITMO FIREFLY / Pedro Henrique Leal Silva;
orientador Homero Schiabel. São Carlos, 2025.

Monografia (Graduação em Engenharia Elétrica com
ênfase em Eletrônica) -- Escola de Engenharia de São
Carlos da Universidade de São Paulo, 2025.

1. Câncer de mama. 2. Mamografia. 3.
Microcalcificações. 4. Algoritmo Firefly. 5. Imagens
Médicas. I. Título.

FOLHA DE APROVAÇÃO

Nome: Pedro Henrique Leal Silva

Título: "Segmentação de imagens mamográficas para identificação de microcalcificações através do Algoritmo Firefly"

Trabalho de Conclusão de Curso defendido e aprovado
em 27 / 11 / 25,

com NOTA 8,3 (oito, três), pela Comissão
Julgadora:

Prof. Associado Homero Schiabel - Orientador - SEL/EESC/USP

Profa. Associada Ana Cláudia Patrocínio - UFU

Prof. Dr. Ricardo Augusto Souza Fernandes - SEL/EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Professor Associado José Carlos de Melo Vieira Júnior

Este trabalho é dedicado às milhares de mulheres — mães, filhas e esposas — que, com coragem e dignidade, enfrentaram o câncer de mama e, mesmo partindo, deixaram um legado de força, amor e esperança.

AGRADECIMENTOS

À minha querida mãe, Cláudia Lopes Leal Silva, cujo amor, dedicação e exemplo moldaram o homem que me tornei;

Ao meu querido pai, Elpídio Pereira Silva, por todo apoio ao longo da minha formação, não apenas como engenheiro, mas também como cidadão ético e responsável comprometido com seus valores;

Aos meus grandes mestres, que com tanta generosidade transmitiram conhecimento e tornaram possível a realização desse sonho, e um especial, ao meu coordenador Homero Schiabel pela dedicação, sensibilidade e presença ao longo de todo o projeto;

À Escola de Engenharia de São Carlos, por fornecer uma infraestrutura de aprendizado e servir como um espaço multidisciplinar de debates acadêmicos, sociais e humanitários;

Aos meus amigos e colegas com os quais pude compartilhar vivências e experiências ao longo da minha passagem;

Com dedicação e carinho, Pedro Leal.

“Ninguém ignora tudo. Ninguém sabe tudo. Todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa. Por isso aprendemos sempre”

Paulo Freire

RESUMO

LEAL SILVA, P.H. **SEGMENTAÇÃO DE IMAGENS MAMOGRÁFICAS PARA IDENTIFICAÇÃO DE MICROCALSIFICACOES ATRAVÉS DO ALGORITMO FIREFLY**. 2025. 64 p. Monograph - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

O câncer de mama é a principal causa de morte relacionada ao câncer entre as mulheres em todo o mundo, tornando a detecção precoce crucial para melhorar as taxas de sobrevivência. As microcalcificações visíveis nas mamografias são indicadores-chave de potencial malignidade, mas a dificuldade de identificá-las corretamente em tecido mamário denso representa um desafio significativo para os radiologistas. Este estudo desenvolve um novo método automatizado de Diagnóstico Assistido por Computador (CAD) para a detecção precisa dessas microcalcificações. Este método consiste em um pré-processamento de imagem em várias etapas para aumento do contraste e filtragem de ruído, seguido pela localização e detecção pelo Firefly, um algoritmo meta-heurístico bioinspirado de otimização. Dois modelos foram desenvolvidos, um básico e outro otimizado, e sua eficácia foi testada usando mamografias digitais anônimas com presença ou ausência clinicamente verificada de microcalcificações. Os resultados demonstram que o modelo otimizado alcançou um desempenho de sensibilidade em 85% e especificidade em 75% na classificação de imagens quanto a presença de microcalcificações. A robustez do modelo foi ainda mais validada pela detecção correta da localidade da microcalcificações dentro da em mamografias de casos positivos, indicando uma taxa de acerto global superior a 83%. Os resultados demonstram que a estrutura proposta é uma estratégia viável e eficaz para a detecção de microcalcificações por CAD. Em trabalhos futuros, evoluções relacionadas a identificação e redução de efeitos de borda na imagem pré-processada poderão aumentar significativamente a precisão na segmentação de áreas suspeitas, e assim, aumentar a especificidade. Essa abordagem tem um potencial para se tornar uma ferramenta confiável para apoiar radiologistas, melhorar a precisão do diagnóstico e, em última análise, contribuir para a detecção precoce do câncer de mama.

Palavras-chave: Câncer de mama. Mamografia. Microcalcificações. Algoritmo Firefly. Pré-processamento de Imagens Médicas.

ABSTRACT

LEAL SILVA, P.H. **SEGMENTATION OF MAMMOGRAPHIC IMAGES FOR THE IDENTIFICATION OF MICROCALCIFICATIONS USING THE FIREFLY ALGORITHM.** 2025. 64 p. - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2025.

Breast cancer is the leading cause of cancer-related death among women worldwide, making early detection crucial to improving survival rates. Microcalcifications visible on mammograms are key indicators of potential malignancy, but the difficulty of correctly identifying them in dense breast tissue poses a significant challenge for radiologists. This study develops a new automated Computer-Aided Diagnosis (CAD) method for the accurate detection of these microcalcifications. This method consists of multi-step image preprocessing for contrast enhancement and noise filtering, followed by localization and detection using the Firefly algorithm, a bio-inspired metaheuristic optimization algorithm. Two models were developed, one basic and one optimized, and their effectiveness was tested using anonymous direct radiography mammograms with clinically verified presence or absence of microcalcifications. The results show that the optimized model achieved a sensitivity of 85% and specificity of 75% in classifying images for the presence of microcalcifications. The robustness of the model was further validated by the correct detection of the location of microcalcifications within mammograms of positive cases, indicating an overall accuracy rate of over 83%. The high level of accuracy shows that the proposed structure is a viable and effective strategy for the detection of microcalcifications by CAD. In future work, developments related to the identification and reduction of edge effects in the pre-processed image could significantly increase the accuracy in the segmentation of suspicious areas, and thus increase specificity. This approach has significant potential to become a reliable tool to support radiologists, improve diagnostic accuracy, and ultimately contribute to the early detection of breast cancer.

Keywords:Breast cancer. Mammography. Microcalcifications. Firefly algorithm. Preprocessing of Medical Images.

LISTA DE FIGURAS

Figura 1 – Relação entre o coeficiente de atenuação para amostras de tecido normal, tecido adiposo, tecido normal granular, fibroadenoma (neoplasia benigna) e a média para diferentes tipos de carcinomas (neoplasia maligna). Imagem extraída de (Tomal, 2007)	28
Figura 2 – Modelo de um mamógrafo convencional. Imagem extraída de (Angelo, 2007)	29
Figura 3 – Exemplos de marcações realizadas por sistemas CAD em mamografias na posição cranio-caudal da mama direita e esquerda. Imagem extraída de (Porto, 2010).	30
Figura 4 – Representação gráfica do ambiente DICOM-PACS em ambiente hospitalar. Representação da integração entre as etapas de realização do exame, análise na estação de trabalho do médico operador e o servidor de armazenamento das imagens. Imagem extraída de (Martins, 2006)	33
Figura 5 – Imagem com exemplos de observação da sequencia de Fibonacci na natureza, arquitetura, microbiologia e Botânica. Retirada do artigo "A Matemática na Natureza"(Martins, 2014)	34
Figura 6 – Exemplo de Termogramas em posição (a) Frontal (b) Lateral Externa. Retirada de (Silva, 2010)	35
Figura 7 – Comparação entre a imagem original e três modelos de imagem com: a) três, b) cinco e c) sete agrupamentos (ou classes) no Firefly.	38
Figura 8 – Mamografia craniocaudal esquerda original da qual as imagens foram extraídas.	40
Figura 9 – Corte centrado em torno da microcalcificação antes do pré-processamento.	41
Figura 10 – Corte centrado na microcalcificação após a fase de aprimoramento do contraste ajustado.	42
Figura 11 – Corte centrado na microcalcificação após a etapa de remoção do fundo.	43
Figura 12 – Corte centrado na microcalcificação após a etapa de filtragem.	44
Figura 13 – Corte centrado na microcalcificação após a fase de <i>Thresholding</i>	44
Figura 14 – Recapitulação das fases de pré-processamento.	45
Figura 15 – Resultado do algoritmo de segmentação Firefly. Do lado esquerdo, um uma imagem mamográfica antes do processamento. Do lado direito, a mesma imagem pós-processada, no qual cada agrupamento é representado por uma tonalidade.	46

Figura 16 – A circularidade é um elemento chave na distinção entre estrias de alta intensidade e conjunto de microcalcificações. No lado direito da Figura, um conjunto de microcalcificações classificadas dentro do conjunto de interesse pelo algoritmo Firefly. No lado esquerdo, fibras em formatos de estrias também classificadas dentro do mesmo conjunto, mas em formatos alongado e não-circular.	48
Figura 17 – Comparação na classificação de regiões de interesse para dois conjunto de parâmetros dentro do algoritmo Firefly. Destaque para maior especificidade do modelo B frente ao modelo A.	49
Figura 18 – Evolução da função objetivo ao longo das iterações do algoritmo. A análise da função permite refinar o modelo e determinar a melhor relação entre custo computacional e qualidade da solução em relação ao número de iterações necessárias	50
Figura 19 – Exemplo de imagem mamográfica sem recorrência de microcalcificações após o processamento pelo algoritmo Firefly	51

LISTA DE TABELAS

Tabela 1 – Parâmetros principais e seus papéis no Algoritmo dos Vagalumes. . . .	37
Tabela 2 – Parâmetros dos Modelos no Algoritmo dos Vagalumes.	49
Tabela 3 – Resultados do Algoritmo Firefly	52
Tabela 4 – Ficha das imagens com microcalcificações	63
Tabela 5 – Ficha das imagens sem microcalcificações	64

LISTA DE ABREVIATURAS E SIGLAS

BI-RADS	Breast Imaging Reporting and Data System ou Sistema de Relatórios e Dados de Imagem Mamária
CAD	Diagnóstico Assistido por Computador
CC	Crânio-cauda
MLO	Médio-lateral oblíqua
CR	Radiografia Computadorizada,
DICOM	Digital Imaging and Communications in Medicine ou Imagem digital e comunicações na medicina
PACS	Picture Archiving and Communication System ou Sistema de Arquivamento e Comunicação de Imagens
CT	Tomografia Computadorizada
MR	Ressonância Magnética
US	Ultrassonografia
ACO	Otimização por Colônia de Formigas
PSO	Otimização por Enxame de Partículas
FA	Otimização por Enxame de Vagalumes ou Fireflys
FFDM	Mamografia digital de campo total
AHE	Equalização adaptativa do histograma
CLAHE	Equalização adaptativa do histograma com contraste limitado
GHE	Equalização global do histograma
CVM	Calcificações vasculares mamária

SUMÁRIO

1	INTRODUÇÃO	23
1.1	Motivações	23
1.2	Objetivos	24
1.3	Estrutura da Monografia	24
2	AVALIAÇÃO DE IMAGENS MAMOGRÁFICAS	27
2.1	Mamógrafo: Do convencional ao digital	27
2.2	Diagnóstico Assistido por Computador: Evoluções e desafios	30
2.3	Protocolo DICOM: Padronização e compartilhamento.	31
2.4	Aplicação de algoritmos bio-inspirados em imagens médicas	33
2.5	Firefly: Segmentação por otimização	36
3	MATERIAIS E MÉTODOS	39
3.1	INBreast	39
3.2	Pré-Processamento	39
3.2.1	Conjunto de dados e resolução da imagem	39
3.2.2	Pré-processamento	39
3.2.2.1	Aumento do contraste	41
3.2.2.2	Remoção do fundo	42
3.2.2.3	Filtragem	43
3.2.2.4	Thresholding	44
3.3	Processamento	45
4	RESULTADOS E DISCUSSÃO	49
5	CONCLUSÃO	55
	REFERÊNCIAS	57
	APÊNDICES	61
	APÊNDICE A – SELEÇÃO DE IMAGENS E RESULTADOS	63

1 INTRODUÇÃO

1.1 Motivações

O câncer de mama é a forma mais comum de câncer em mulheres, com mais de 20 milhões de novos casos registrados em 2022, e a principal causa de morte relacionada ao câncer, com mais de 660 mil mortes em todo o mundo (Li *et al.*, 2025). No Brasil, o câncer de mama também apresenta elevada magnitude, causando alta mortalidade, redução da qualidade de vida e elevados custos para os serviços de saúde públicos e privados (Silva *et al.*, 2021).

As mamografias são as principais ferramentas de rastreamento, e o diagnóstico precoce pode levar a taxas de sobrevivência significativamente melhores (Shi *et al.*, 2025)(Marcon *et al.*, 2024). A detecção precoce do câncer de mama através da mamografia de rastreamento é uma prática comprovadamente eficaz e capaz de reduzir a mortalidade associada a essa doença em até 22% (Geras; Mann; Moy, 2019).

Em imagens mamográficas, existem diversos parâmetros que podem ser analisados para a detecção de patologias, dentre elas a presença de microcalcificações. Microcalcificações são pequenos depósitos de tecido calcificado que geralmente aparecem nas mamografias como pequenos pontos claros e são considerados indicações importantes de uma possível malignidade (Logullo *et al.*, 2022). Elas podem ser difíceis de reconhecer nas imagens devido ao seu tamanho, à falta de características morfológicas distintas e ao baixo contraste com o tecido mamário denso ao redor. (Tot *et al.*, 2021). A análise de sua distribuição ao longo do tecido mamário é um componente crítico da avaliação do Sistema de Relatórios e Dados de Imagem Mamária (BI-RADS), que padroniza os relatórios e orienta as decisões clínicas, incluindo a necessidade de uma biópsia (Kim *et al.*, 2022)(Logullo *et al.*, 2022)(Burnside *et al.*, 2007).

Apesar dos benefícios evidentes, o exame mamográfico ainda enfrenta desafios seja pela limitada sensibilidade em casos de tecido mamário denso ou seja pela ocorrência de erros perceptivos e interpretativos por parte do médico radiologista responsável pela análise (Guerreiro *et al.*, 2024). O diagnóstico assistido por computador (CAD) surge então como uma ferramenta poderosa para auxiliar os radiologistas na interpretação de imagens médicas. Algoritmos foram treinados para detectar microcalcificações em phantoms e mamografias reais. No entanto, alcançar alta precisão de detecção e, ao mesmo tempo, minimizar falsos positivos continua sendo um desafio significativo. (Cheng *et al.*, 2003a)(Wang; Yang, 2018)

Na literatura, diversos algoritmos de análise de imagens mamográficas vêm sendo desenvolvidos. Modelos baseados na integração entre ferramentas de inteligência artifi-

cial e análise humana tem-se demonstrado um caminho promissor, agilizando processos, aumentando a assertividade do diagnóstico, diminuindo o tempo de atendimento e aumentando a qualidade geral dos serviços prestados (Guerreiro *et al.*, 2024). Entretanto ainda existem desafios inerentes associados a vieses nas bases de treinamento, alta exigência de processamento computacional e disponibilidade de bases de imagens digitalizadas com alta qualidade para treinamento de modelos o que dificulta a sua aplicabilidade (Miranda; Alves; Miranda, 2025).

Nesse cenário, surgem outras ferramentas para detecção e classificação de imagens mamográficas, e mais precisamente, para identificação de microcalcificações como uma das etapas de triagem e classificação de probabilidade de risco. Este estudo propõe um novo método de detecção que combina poderosas técnicas de processamento automatizado de imagens e o Firefly, um algoritmo inspirado na natureza para localizar microcalcificações em imagens mamográficas digitais.

1.2 Objetivos

Os objetivos específicos deste Trabalho de Conclusão de Curso são:

a) desenvolver e ajustar um *pipeline* de pré-processamento de imagens mamográficas para melhorar o contraste em torno das microcalcificações e remover dados desnecessários, incluindo ruído, da imagem original;

b) implementar o algoritmo Firefly para segmentação das imagens mamográficas pré-processadas em zonas de interesse;

d) desenvolver um algoritmo para isolar regiões com microcalcificações para uma visualização clara pelo radiologista;

De maneira geral, o presente projeto visa propor novas ferramentas de auxílio computacional baseado em algoritmos bio-inspirados para atuar no diagnóstico e identificação do câncer de mama em estágios iniciais, aumentando assim as chances de tratamento e cura.

1.3 Estrutura da Monografia

Esse trabalho foi organizado em cinco capítulos.

No primeiro, Avaliação de Imagens Médicas, serão apresentadas as bases teóricas relacionadas à evolução das técnicas de avaliação de imagens mamográficas, seguindo uma linha temporal, desde o mamógrafo convencional até o digital, passando pelas revoluções dentro da análise de técnicas de diagnóstico auxiliado pelo computador. Além disso, serão abordados temas como o modelo de padronização de imagens médicas pelo protocolo DICOM. Por fim, serão introduzidos os princípios teóricos e matemáticos do algoritmo

Firefly, assim como suas motivações e potenciais aplicações.

O capítulo seguinte será dedicado aos materiais e métodos utilizados ao longo da execução desse projeto. Nele, será apresentada a Base *Inbreast*, das quais foram retiradas as amostras estudadas, assim como os métodos de pré-processamento, passando por técnicas de aumento de contraste, remoção de fundo, filtragem e *Thresholding*. Em seguida, será descrito o processo de processamento através do algoritmo Firefly.

Posteriormente, o capítulo Resultados e Discussões será dedicado à análise dos resultados obtidos à luz de critérios como eficiência computacional, sensibilidade e especificidade do modelo encontrado.

Por fim, na seção de conclusão serão apresentados os principais pontos do projeto, assim como oportunidades de continuidade em pesquisas futuras. Em seguida, serão apresentadas as referências bibliográficas utilizadas ao longo deste trabalho.

2 AVALIAÇÃO DE IMAGENS MAMOGRÁFICAS

A avaliação de imagens mamográficas se tornou um ramo proeminente dentro da medicina moderna. Paralelamente ao desenvolvimento de áreas correlatas à saúde, a engenharia vem sendo vista como um setor central no desenvolvimento de tecnologias capazes de acompanhar as novas demandas do setor médico.

Dentre os diversos modelos de imagens médicas, incluindo desde tomografias computadorizadas até ressonâncias magnéticas, a mamografia ganha um significativo destaque dentro da literatura biomédica.

2.1 Mamógrafo: Do convencional ao digital

Mamografias são tipos particulares de radiografias aplicadas à região da mama. O equipamento utilizado para gerar essas imagens é então chamado de mamógrafo. Usualmente, um exame de mamografia é feito em duas posições, também chamadas de incidências para cada mama: crânio-caudal (CC) e médio-lateral oblíqua (MLO).

Um dos grandes desafios desse tipo de exame é a própria composição dos tecidos da mama. O tecido glandular, tecido fibroso e a gordura possuem índices de atenuação muito próximos, o que resulta em baixo contraste (Rodrigues, 2008). A Figura 1 ilustra a comparação entre o coeficiente de atenuação para amostras de tecido normal, tecido adiposo, tecido normal granular, fibroadenoma (neoplasia benigna) e a média para diferentes tipos de carcinomas (neoplasia maligna) em função da energia característica (Tomal, 2007). Como demonstrado pela proximidade entre as curvas da Figura 1, na prática, a diferença entre tecidos normais e enfermos acaba sendo muito sutil, exigindo a implementação de técnicas mais apuradas para aumentar a resolução de contraste nas imagens mamográficas.

Inicialmente, os mamógrafos geravam imagens essencialmente em filme. O modelo da Figura 2 ilustra o funcionamento desses equipamentos (Angelo, 2007). Dentro do tubo mamográfico, um alvo em molibdênio é atingido por um feixe de elétrons que desaceleram e liberam raios-X. Por sua vez, esses raios são filtrados na saída do tubo evitando que feixes de baixa energia aumentem a dose de radiação a níveis não seguros. Em paralelo, os feixes muito mais energéticos que 20 keV também são filtrados, o que acaba selecionando um espectro bem característico de radiação a ser liberado ao paciente (Porto, 2010).

Essa radiação atravessa um mecanismo de compressão, diretamente em contato com a mama, responsável por homogeneizar, ao máximo possível, a densidade a ser radiografada, tornando a mama menos espessa e mais uniforme, o que permite um menor espalhamento da radiação, e conseqüentemente uma diminuição do ruído na imagem. Em seguida, a grade mamográfica é a responsável por filtrar a radiação secundária gerada pela interação

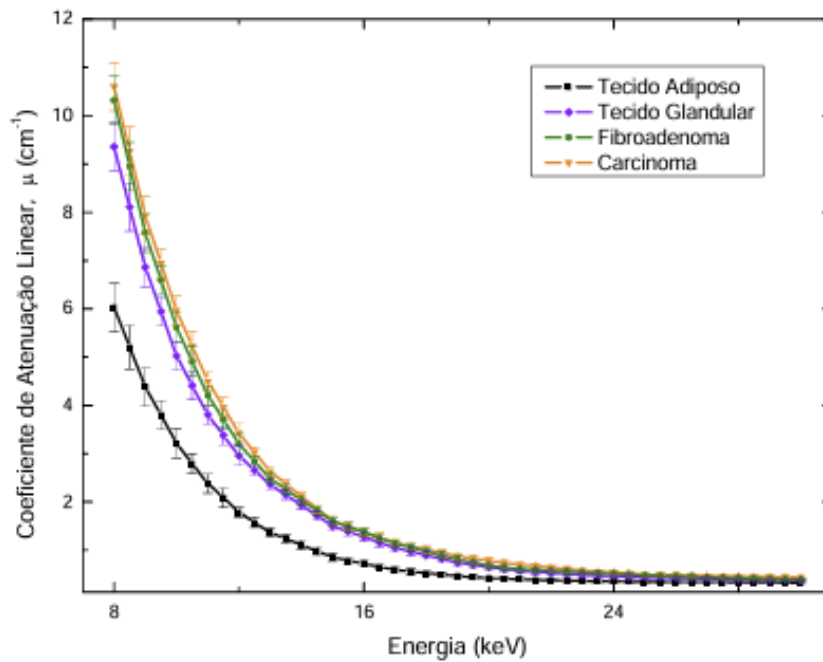


Figura 1 – Relação entre o coeficiente de atenuação para amostras de tecido normal, tecido adiposo, tecido normal granular, fibroadenoma (neoplasia benigna) e a média para diferentes tipos de carcinomas (neoplasia maligna). Imagem extraída de (Tomal, 2007)

entre a radiação primária e o tecido, que poderia gerar distorções na imagem final. Por fim, os raios encontram o cassete mamográfico, um sistema filme-écran, capaz de absorver raios X e transformar em luz visível a ser utilizada para sensibilizar o filme e formar uma imagem. Após uma série de processos físico-químicos, o filme é capaz de revelar a imagem mamográfica de interesse (Porto, 2010).

Existem dois métodos para digitalização de uma imagem mamográfica. O primeiro envolve a digitalização indireta do filme mamográfico por um *scanner*. A segunda é a substituição do sistema filme-écran por uma placa de material fosfórico ligada a um sistema *Computed Radiography*, ou CR. Nesse segundo caso, essa placa armazena os feixes de raios X que em seguida são lidos por um sistema a laser. A reflexão desse laser é por sua vez convertida em sinal digital, o que gera a imagem de interesse. A placa pode, posteriormente, ser apagada utilizando luz fluorescente e reutilizada para geração de novas imagens (Souza, 2012).

No final da década de 80, novos equipamentos surgiram com o objetivo de gerar imagens digitais diretamente a partir da exposição de semicondutores à radiação, sem que haja a necessidade da intermediação de uma placa. Nesse caso, a radiação é convertida em sinais elétricos. Esse método evita o espalhamento da informação e aumenta significativamente o contraste e a nitidez das imagens geradas (Freitas *et al.*, 2006).

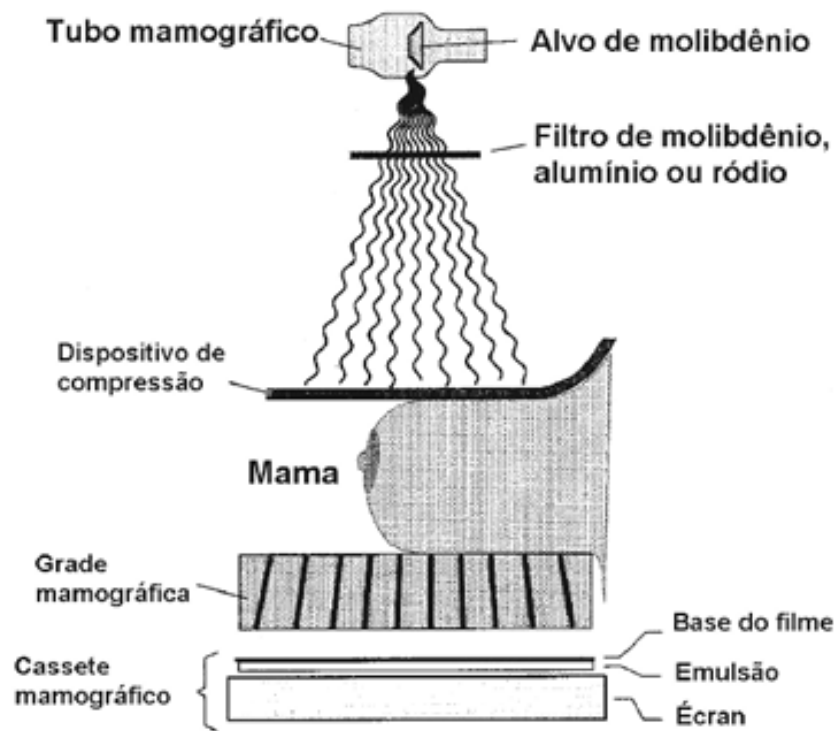


Figura 2 – Modelo de um mamógrafo convencional. Imagem extraída de (Angelo, 2007)

Apesar dos avanços significativos na engenharia dos equipamentos de obtenção de imagens mamográficas, ainda existem desafios relacionados especialmente à fase seguinte à obtenção do exame: a sua análise. Mais especificamente, na metodologia de avaliação de imagens alguns desafios se impõem, dentre eles: i) a incerteza relativa ao diagnóstico e classificação da patologia (em função, por exemplo de uma variante anatômica), ii) a fadiga ocular do médico radiologista, devida à elevada demanda de análise, e questões ambientais, como a iluminação, que também pode afetar a visualização das estruturas anatômicas, iii) a falta de qualidade das imagens (por ruído ou por mau posicionamento da paciente), além da iv) falta de estudos comparativos ou em alguns casos, até mesmo da v) falta de experiência do médico radiologista. (Potnis *et al.*, 2022) (Delija; Mihanović, 2024) (Perre; Freire, 2021)

Para além da dupla validação por especialistas, o que aumenta consideravelmente os custos e o tempo das análises, outras técnicas têm sido desenvolvidas para auxiliar o processo de tomada de decisão. Dentre elas, estão os sistemas de *Computer Aided Diagnostis*, ou Diagnósticos Assistidos por Computador (CAD), concebidos com o objetivo de auxiliar uma análise humana a partir de técnicas de processamento de imagem e discriminação de regiões com uma elevada precisão e uma admissível quantidade de falsos-positivos.

2.2 Diagnóstico Assistido por Computador: Evoluções e desafios

Os primeiros modelos de Diagnóstico Assistido por Computador (*CAD*) foram desenvolvidos na década de 1980 para auxiliar na identificação precoce de doenças, especialmente tumores de mama e pulmão e eram baseados no processamento de filmes radiológicos digitalizados (Gillies; Kinahan; Hricak, 2016)(Santos *et al.*, 2014).

O objetivo do CAD é aumentar a acurácia do diagnóstico através de um mecanismo de sugestão de resposta, (ou "segunda opinião") como um método auxiliar, quantitativo, não operador-dependente no diagnóstico, segmentação e/ou classificação de estruturas anatômicas em imagens médicas. Para além de aumentar a assertividade dos diagnósticos, o CAD também se propõe a reduzir o tempo de análise humana, garantindo assim uma maior eficiência operacional e redução de custos (Jr; Azevedo-Marques; Oliveira, 2017). A Figura 3 mostra exemplos de marcações realizadas por sistemas CAD em mamografias na posição crânio-caudal da mama direita e esquerda (Porto, 2010).

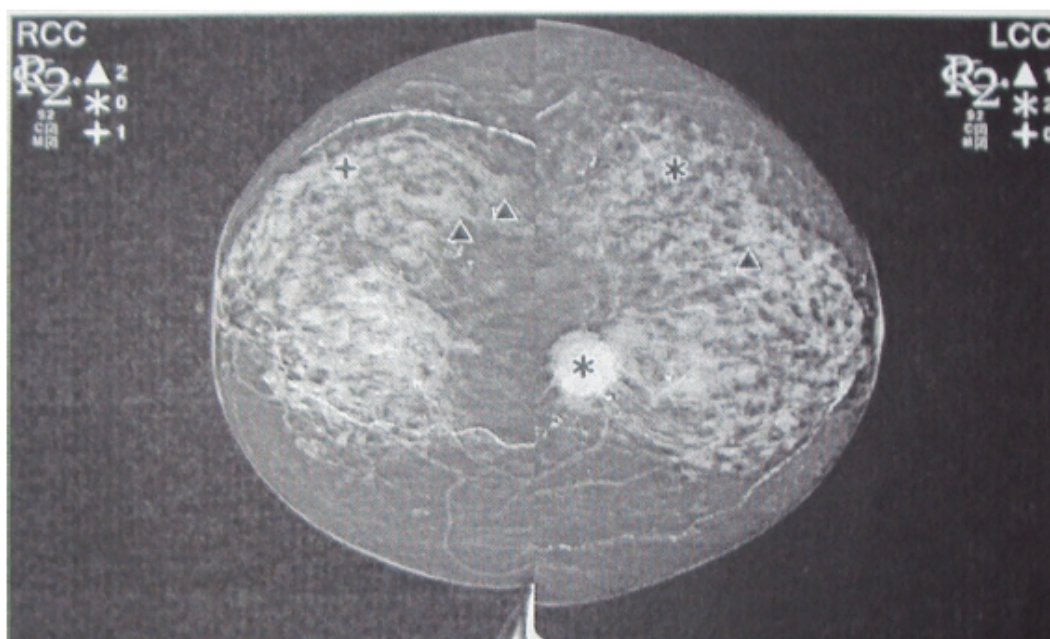


Figura 3 – Exemplos de marcações realizadas por sistemas CAD em mamografias na posição cranio-caudal da mama direita e esquerda. Imagem extraída de (Porto, 2010).

No canto superior esquerdo temos informações quanto ao tipo da imagem da mama direita, enquanto que no canto superior esquerdo, temos a mesma informação para a mama esquerda. Verifica-se ainda nas marcações realizadas pelo sistema, no caso direito, duas regiões com microcalcificações e uma região com microcalcificações e massa ao mesmo tempo. Na esquerda, uma região com microcalcificações e outra com massa densa, conforme descrição apresentada no manual do equipamento (Porto, 2010).

Entretanto, apesar dos benefícios teóricos, a utilização de modelos de CAD tra-

dicionais em centros de pesquisa e programas de rastreamento do câncer de mama ainda é questionada (Fazal *et al.*, 2018). Segundo a análise de Kholi e Saurabh, realizada em 2017 em diversos centros de pesquisa nos Estados Unidos, para cada verdadeiro-positivo detectado através do recurso, existem ainda mais falsos-positivos que precisam ser ignorados pelos radiologistas, o que gera desconfianças práticas sobre a confiabilidade dos sistemas e dúvidas no momento de interpretação das imagens, com efeitos práticos na taxa de repetição de exames e biópsias, e consequente aumento de custos. (Kohli; Jha, 2018)

O tamanho de lesões pode ser ainda um agravante para a assertividade de ferramentas de auxílio computacional. Especialmente em ferramentas que se propõem a identificar pequenas lesões (inferiores a 3 cm), estruturas mais evidentes podem ser ignoradas. Estudos mostram que os melhores desempenhos de CAD ocorreram para as lesões medindo entre 1 e 3 cm. Os menores valores de sensibilidade foram obtidos para estruturas com raios de 4 cm (taxa de detecção = 52,9%), ao passo que lesões pequenas (< 10 mm) e maiores (> 30 mm) foram detectadas com sensibilidade intermediária (taxa de detecção = 83%). As lesões de tamanho entre 10 e 30 mm foram todas detectadas. (Brem *et al.*, 2005).

Entretanto, no que tange à identificação de regiões de interesse quanto à presença de microcalcificações, as técnicas atuais de CAD têm-se mostrado amplamente eficazes. Estudos recentes mostram ferramentas capazes de atingir uma precisão superior a 88,6% com base no método de Otsu e filtros morfológicos. (Duarte *et al.*, 2011) (Calas; Gutfilen; Pereira, 2012)

O desenvolvimento de novas técnicas de processamento de imagens aliado à inteligência artificial, tem impulsionado o desenvolvimento de sistemas de detecção assistida por computador nos últimos anos. Atualmente, o campo de pesquisa em ferramentas computacionais de análise de imagens médicas está em franca expansão com a associação com técnicas de redes neurais convolucionais. Pesquisas em detecção automática de nódulos pulmonares e estimativa do risco de malignidade, detecção e quantificação de enfisema e chance de invasão local do câncer pulmonar têm ganhado espaço dentro da literatura biomédica. (Tao *et al.*, 2019) (Nam *et al.*, 2019)

2.3 Protocolo DICOM: Padronização e compartilhamento.

Atualmente existe uma grande demanda por métodos de processamento e armazenamento de dados. Esse cenário não é diferente no caso de imagens médicas. Para além das informações visuais geradas pelo exame, os protocolos médicos usualmente contêm o registro eletrônico do paciente, informações alfa-numéricas relevantes como idade, gênero, ou ainda análises prévias ou informações técnicas sobre o equipamento no qual o exame foi realizado.

Avanços no desenvolvimento de equipamentos que geram imagens médicas digitais

aliados à necessidade de padronização na comunicação entre diferentes sistemas, hospitais e clínicas em diferentes regiões do planeta guiaram a busca por protocolos eficientes de armazenamento e troca de informações.

O padrão DICOM foi então criado, em 1983, a partir de uma parceria entre o Colégio Americano de Radiologia e a Associação Americana de Equipamentos Elétricos. O seu objetivo era permitir o desenvolvimento de sistemas de comunicação e arquivamento de imagens médicas (ou *Picture Archiving and Communication System - PACS*) de modo padronizado entre vários fabricantes de equipamentos, garantindo harmonia entre as fases de aquisição, transmissão, distribuição, armazenamento e apresentação de imagens médicas em ambientes hospitalares (Dellani, 2001).

A busca pela otimização de custos aliada à redução do tempo de processamento de imagens impulsionou o desenvolvimento e implementação dos PACS em grande escala, facilitando assim a comunicação entre diferentes sistemas (Caritá *et al.*, 2006). Na prática, esse modelo permite que um hospital acesse imagens geradas em qualquer outro lugar do mundo e até mesmo acesse exames gerados por equipamentos diferentes através de um clique.

O padrão DICOM integra os diversos tipos de imagens geradas por diferentes equipamentos (como CT, MR, X-ray, US), a estações e servidores de trabalho do médico e a servidores de arquivos, responsáveis pelo armazenamento tanto das imagens quanto das informações alfa-numéricas do paciente, do estudo ou até mesmo de uma análise dos resultados obtidos (Dellani, 2001). A Figura 4 demonstra graficamente a conexão entre o ambiente DICOM-PACS (Martins, 2006).

Esse tipo de padronização foi essencial para a consolidação do modelo de atendimento entre diferentes centros médicos. Hospitais ao redor do mundo começaram a compartilhar imagens que poderiam ser acessadas por profissionais autorizados a qualquer momento através da rede de computadores, evitando assim a necessidade de repetição de exames já realizados.

A terceira versão do protocolo DICOM foi criada em 1993 e permanece até os dias atuais como a sua versão definitiva. Ela é chamada de ACR-NEMA DICOM ou DICOM-3.0 e pode ser aplicada a um ambiente em rede e à troca de mídia *off-line*, definindo o formato do arquivo e dando suporte a diversas modalidades de imagem, como tomografia computadorizada, ressonância magnética e ultrassonografia (Dellani, 2001). Grandes empresas, como a *PHILIPS*, *MedImage Inc*, *AGFA* e *RealTimeImage*, possuem uma extensa literatura de artigos que descrevem a conformidade dos seus equipamentos com o padrão DICOM 3.0. (Martins, 2006)

O desenvolvimento de um modelo de padronização entre a obtenção, armazenamento e leitura de imagens médicas foi essencial para possibilitar a posterior criação de banco

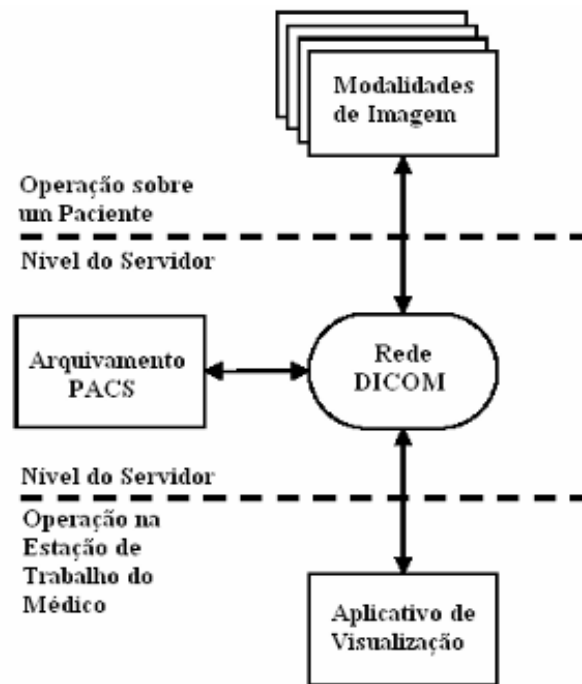


Figura 4 – Representação gráfica do ambiente DICOM-PACS em ambiente hospitalar. Representação da integração entre as etapas de realização do exame, análise na estação de trabalho do médico operador e o servidor de armazenamento das imagens. Imagem extraída de (Martins, 2006)

de dados comuns e ferramentas de análise capazes de treinar algoritmos inteligentes para auxiliar os profissionais no diagnóstico de doenças.

2.4 Aplicação de algoritmos bio-inspirados em imagens médicas

Com o desenvolvimento do campo de processamento de imagens, diversos métodos da engenharia têm sido desenvolvidos com base no comportamento fisiológico de animais ou até mesmo da organização de insetos, como abelhas e vaga-lumes, em comunidades.

A inspiração na natureza para criação de ferramentas de engenharia não é uma prática recente. Sequências numéricas, como a de Fibonacci, já vêm encantando pesquisadores e matemáticos há séculos, e a sua relação com fenômenos naturais, estrutura de organismos vivos, fenômenos sociológicos, estatísticos ou epidêmicos não tem passado despercebida (Martins, 2014). A Figura 5 ilustra alguns exemplos da observação dessa sequência em alguns campos do conhecimento como na microbiologia, na botânica e na arquitetura. A natureza tem inspirado grande parte das grandes descobertas e tecnologias da humanidade, e no campo de análise de imagens médicas, isso não é diferente.

Diversas pesquisas têm sido desenvolvidas no sentido de aplicar algoritmos meta-heurísticos na análise de imagens médicas. Dentre eles, destacam-se estudos sobre a

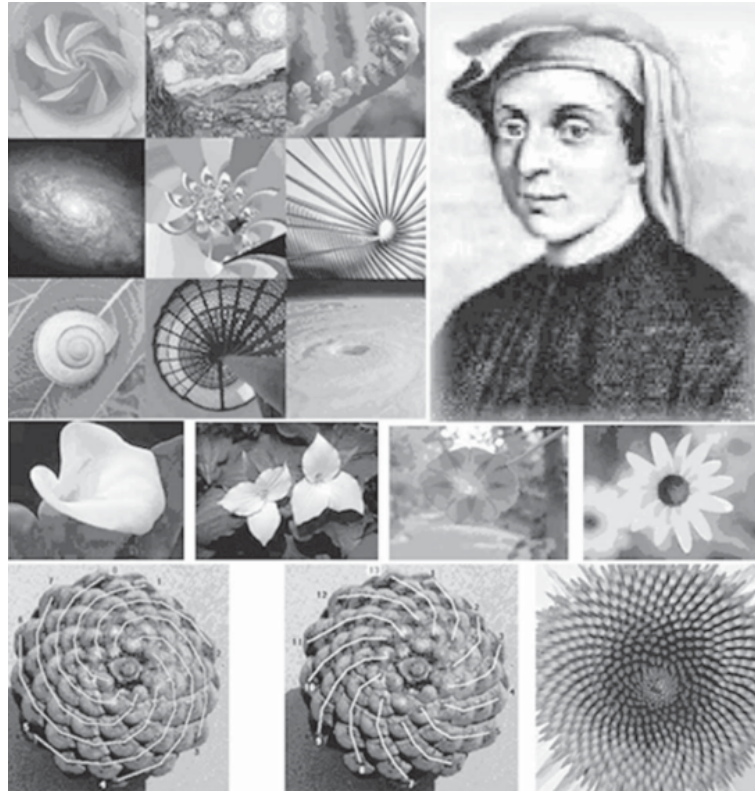


Figura 5 – Imagem com exemplos de observação da sequência de Fibonacci na natureza, arquitetura, microbiologia e Botânica. Retirada do artigo "A Matemática na Natureza"(Martins, 2014)

aplicação de técnicas estocásticas de otimização aplicadas à detecção, localização e classificação de tumores (Figueiredo; Fernandes; Guimaraes, 2018). Modelos recentes têm utilizado imagens termográficas, obtidas a partir da radiação infravermelha emitida pela mama, para medir a temperatura na superfície da pele e identificar áreas suspeitas (Rocha *et al.*, 2024). Esses algoritmos baseados em imagens termográficas têm se destacado devido ao seu custo reduzido e a natureza não invasiva em comparação com outras técnicas de diagnósticos. A Figura 6 ilustra um exemplo de imagens termográficas obtidas em posições frontais e laterais externas.

A partir das imagens termográficas, modelos de Algoritmo Genético (AG) foram utilizados para classificar de mamas quanto a presença de tumores, além de determinar variáveis desconhecidas como as propriedades do tumor, sua forma e localização sem medições diretas. A Acurácia no acerto da classificação foi medida com base no sistema TNM que descreve o estadiamento do câncer base em três critérios: tamanho e invasão do tumor primário, disseminação para os linfonodos regionais e presença de metástase à distância (Rocha *et al.*, 2024).

O Algoritmo Genético (AG) é um método estocástico de otimização inspirado na evolução natural que busca alternativas para problemas complexos através de uma população de soluções candidatas (Holland, 1975). Conceitos como cruzamento (ou *crosso-*

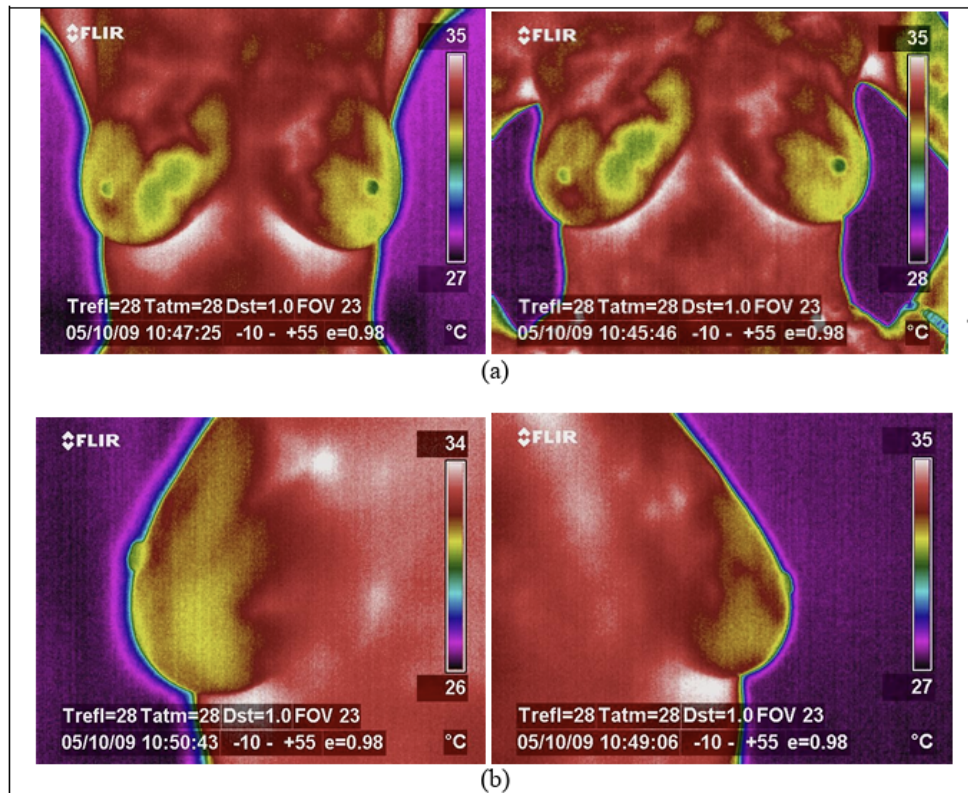


Figura 6 – Exemplo de Termogramas em posição (a) Frontal (b) Lateral Externa. Retirada de (Silva, 2010)

ver) e mutação são utilizados para explorar iterativamente o espaço de soluções possíveis, favorecendo aquelas com maiores aptidões ou adequações ao problema (Eberhart; Kennedy, 1995).

Outros estudos, por sua vez, têm utilizado redes neurais convolucionais para classificar imagens termográficas estáticas da mama (Gonçalves *et al.*, 2022). Nesse caso, obter uma arquitetura com os hiperparâmetros adequados é essencial para o bom desempenho da rede. Algoritmos bioinspirados de otimização por enxame de partícula (PSO) foram utilizados para otimizar esses hiperparâmetros. O PSO é uma técnica inspirada em comportamentos sociais como os de bandos de pássaros e cardumes de peixes (Kennedy; Eberhart, 1995). A evolução do modelo ocorre através de dois mecanismos: da competição, ou seja do aprendizado individual, e da cooperação, baseado no aprendizado do grupo (Serapião, 2009).

Uma diferença conceitual importante entre o Otimização por Enxame de Partícula (PSO) e o Algoritmo Genético (AG) é que enquanto a motivação nestes é a sobrevivência dos melhores indivíduos, os quais devem competir entre si para sobreviver e gerar uma prole, no PSO, a força motriz é a interação social e a troca de conhecimento sobre o espaço de busca (Engelbrecht, 2005). De todo modo, é válido afirmar que modelos metaheurísticos têm ganhado força na literatura biomédica, seja como ferramentas para otimização de

parâmetros seja na interpretação direta de imagens. Nesse contexto, outros algoritmos surgem, como Firefly, um modelo também bio-inspirado, relacionado ao comportamento reprodutivo de comunidades de vaga-lumes.

2.5 Firefly: Segmentação por otimização

O Firefly, proposto por Xin-She Yang (2008), foi o algoritmo de otimização metaheurístico escolhido para classificação de imagens mamográficas ao longo desse projeto. O modelo é baseado no padrão de luminescência característico de um conjunto de vaga-lumes e nas propriedades de comunicação globais entre as partículas do enxame.

Na literatura, ele tem sido explorado principalmente na segmentação de imagens cerebrais e pulmonares, pela sua conhecida capacidade de exploração e busca de centros de agrupamentos iniciais ideais para posterior aplicação de outros algoritmos, como o Fuzzy c-means (FCM), e da inteligência do enxame na identificação das intensidades ideais de agrupamentos.

O método Firefly pode ser descrito por metáforas matemáticas ao comportamento biológico dos vaga-lumes, nas quais:

- i) Cada vaga-lume (solução candidata) tem um brilho I proporcional à qualidade da solução.
- ii) A atratividade entre dois vaga-lumes decai segundo a equação:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (2.1)$$

Onde r é a distância entre eles e β_0 é a atratividade base. O coeficiente γ controla a absorção da luz.

- iii) Para quaisquer dois vaga-lumes, o menos brilhante vai avançar em direção ao mais brilhante, e caso não exista algum elemento mais brilhante em relação aos demais, ele será movido aleatoriamente.

A função que descreve o movimento de um vaga-lume i que se desloca na direção de um vaga-lume j mais brilhante, é dada por:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \beta(r_{ij}) (\mathbf{x}_j^t - \mathbf{x}_i^t) + \alpha \varepsilon_i^t \quad (2.2)$$

com α sendo a amplitude do termo aleatório ε .

O número máximo de iterações é fornecido como parâmetro de entrada do algoritmo e pode ser definido de modo empírico em função da natureza do problema, o orçamento computacional disponível e a taxa de convergência desejada.

A inicialização da população de *fireflies* é representada por um vetor de valores reais dado por uma distribuição uniforme com base nos níveis de cinza mínimos e máximos da imagem. Os parâmetros de entrada do algoritmo, por sua vez, são usualmente definidos empiricamente. Os principais parâmetros, assim como sua interpretação, são apresentados na Tabela 1.

Tabela 1 – Parâmetros principais e seus papéis no Algoritmo dos Vagalumes.

Parâmetro	Interpretação
MaxIt	Número máximo de iterações (limite de gerações).
nPop	Tamanho da população de vagalumes (diversidade \times custo computacional).
γ (gamma)	Coefficiente de absorção da luz; regula o decaimento da atratividade com a distância.
β_0 (beta0)	Atratividade máxima quando a distância entre dois vagalumes é zero.
α	Amplitude do termo aleatório que promove exploração estocástica.
alpha_damp	Fator de amortecimento aplicado a α a cada iteração, favorecendo convergência.
delta	Escala absoluta do ruído, proporcional à faixa de intensidades da imagem.
m	Expoente no termo de decaimento; valores maiores reduzem o alcance da atração.
k	Número de centroids/limiares (clusters) na segmentação.
dmax	Distância máxima usada para normalizar r_{ij} em $[0, 1]$.

Dentro do Firefly, cada um dos parâmetros deve ser determinado experimentalmente. O número maior de iterações, juntamente com o tamanho da população de vagalumes são os principais responsáveis pela definição do custo computacional. Quanto maior o número de vagalumes e quanto mais iterações o sistema fizer, maior é o tempo de processamento e precisão do modelo. Em geral, os modelos atingem uma região de saturação na qual a curva da função de custo pelo limite de geração se estagna, o que define, na prática, o número ótimo de iterações que o sistema deve operar.

Como saída, obtém-se uma imagem em que cada pixel é catalogado dentro de uma das classes pré-definidas. Na visualização, para cada uma dessas classes é atribuída uma tonalidade, de modo que regiões da imagem categorizadas dentro de uma mesma classe, possuem uma mesma cor. A Figura 7 ilustra o a comparação entre o resultado da aplicação da técnica Firefly e a imagem original em três casos: com três, com cinco e sete agrupamentos (ou classes).

O algoritmo do Firefly pode então ser usado para segmentar imagens em regiões de interesse (ROIs). Cada solução, ou seja, agrupamento ideal é melhorado iterativamente, de modo a convergir para a solução esperada. No caso da identificação de microcalcificações, agrupamentos com alta intensidade média de tons de cinza e baixa concentração espacial podem ser bons candidatos.

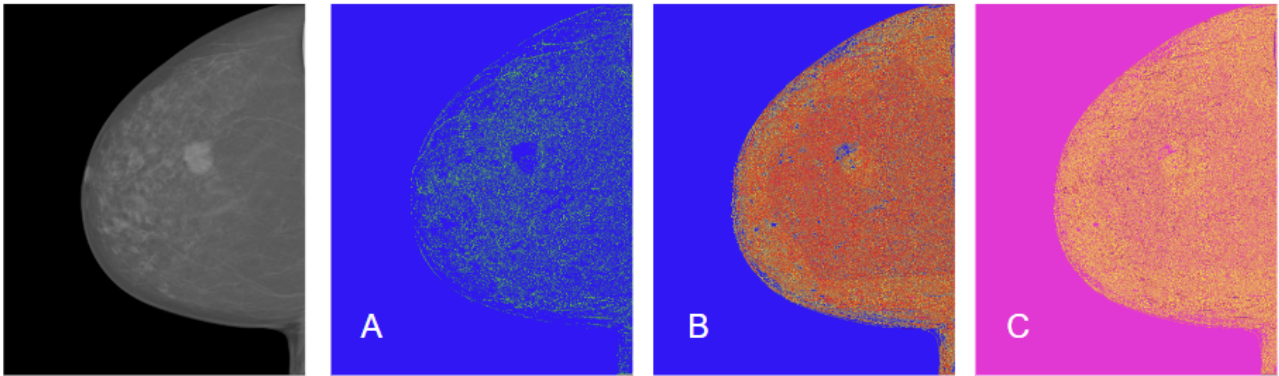


Figura 7 – Comparação entre a imagem original e três modelos de imagem com: a) três, b) cinco e c) sete agrupamentos (ou classes) no Firefly.

Dentre as principais vantagens desse algoritmo estão: a) uma melhor exploração do espaço de soluções quanto comparado a outros métodos determinísticos, devido especialmente ao termo aleatório e a atração modulada pela distância, que lhe permitem escapar mais facilmente de mínimos locais; b) uma boa relação com funções ruidosas, descontínuas ou com muitos picos- comuns em imagens médicas; c) a facilidade de se combinar com outros métodos, como K-means ou ainda Fuzzy C-Means (FA-FCM), para buscar soluções mais estáveis e precisas (Crawford *et al.*, 2014).

A grande desvantagem desse tipo de algoritmo é sua elevada sensibilidade aos parâmetros, em especial ao termo de aleatoriedade, atração inicial e ao coeficiente de absorção de luz. Uma má configuração pode ter impactos em uma convergência mais lenta, uma estagnação ou excesso de aleatoriedade (Crawford *et al.*, 2014).

Pesquisas recentes aplicando o algoritmo Firefly na análise de imagens mamográficas alcançaram resultados entre 20% até 91% de acurácia na identificação de microcalcificações em mamadas densas e entre 25% e 95% em mamas não densas, em função da definição de parâmetros como o número de Fireflies, o número máximo de iterações, o coeficiente de absorção e o fator de atratividade (Cruz *et al.*, 2016). Esse resultado demonstra a viabilidade do método e representa a abertura de uma promissora avenida para exploração do algoritmo em novos conjuntos de imagens, novas técnicas de pré-processamento, novos parâmetros e métodos de filtragem.

3 MATERIAIS E MÉTODOS

3.1 INBreast

As imagens utilizadas nesse estudo foram retiradas da base *Inbreast*, desenvolvida em Portugal entre abril de 2008 e março de 2010, com o objetivo de apoiar pesquisas em detecção e diagnóstico auxiliados por computador (CAD) no câncer de mama. O artigo “*INbreast: Toward a Full-field Digital Mammographic Database*” detalha a construção da base, composta por 115 casos (410 imagens) obtidas com mamografia digital de campo total (FFDM) com detector de selênio amorfo. A base inclui diferentes tipos de achados (massas, calcificações, assimetrias, distorções e casos normais) com descrições, marcações e breves comentários médicos feito por especialistas (um radiologista e revisados por outro) através do uso do software OsiriX (PACS open-source).

3.2 Pré-Processamento

Assim como os olhos humanos treinados, os algoritmos de aprendizado de máquina podem se beneficiar muito com técnicas de processamento de imagem, como o aprimoramento de contraste. Neste trabalho, essas técnicas são chamadas de pré-processamento, reservando o título de processamento para o funcionamento dos principais algoritmos de detecção baseado na técnica *Firefly*.

A etapa de pré-processamento é fundamental e tem impactos diretos nos resultados das etapas subsequentes. Ao longo do projeto, foram realizadas diversas fases com o objetivo de aumentar o contraste, remover o fundo, filtrar ruídos indesejados e equalizar as imagens mamográficas para posterior processamento.

3.2.1 Conjunto de dados e resolução da imagem

Todas as imagens utilizadas neste estudo são anônimas e foram disponibilizadas pelo Conjunto de Dados *InBreast* para esta pesquisa. As mamografias consistem em imagens com resolução espacial de 0,07 mm, e resolução de contraste de 16 bits.

3.2.2 Pré-processamento

As mamografias brutas não são ideais para análise médica sem o processamento adequado. Os receptores de luz humanos não são lineares, como a maioria dos receptores para frequências amplas, e isso resulta em uma interação complexa entre o nível real de brilho de um segmento exibido em filme, impressão ou digitalmente e o brilho percebido, o que pode levar a diagnósticos errados. O protocolo DICOM 14 visa padronizar a experiência de visualização entre radiologistas em favor de uma curva de visualização de brilho não

linear que aproxima ou separa os valores de brilho para que possam ser melhor discernidos a olho nu por especialistas.

Dentro da faixa de flexibilidade permitida, diferentes empresas orientam seu processamento de imagem alterando ligeiramente as curvas para diferentes equipamentos, o que aumenta o contraste em torno dos espectros de brilho de interesse. Às vezes, isso reduz os contrastes no brilho mais alto, dificultando a percepção de elementos pequenos e brilhantes, como microcalcificações. A Figura 8 é uma mamografia craniocaudal esquerda com microcalcificações no segundo quadrante, perto do eixo horizontal médio.

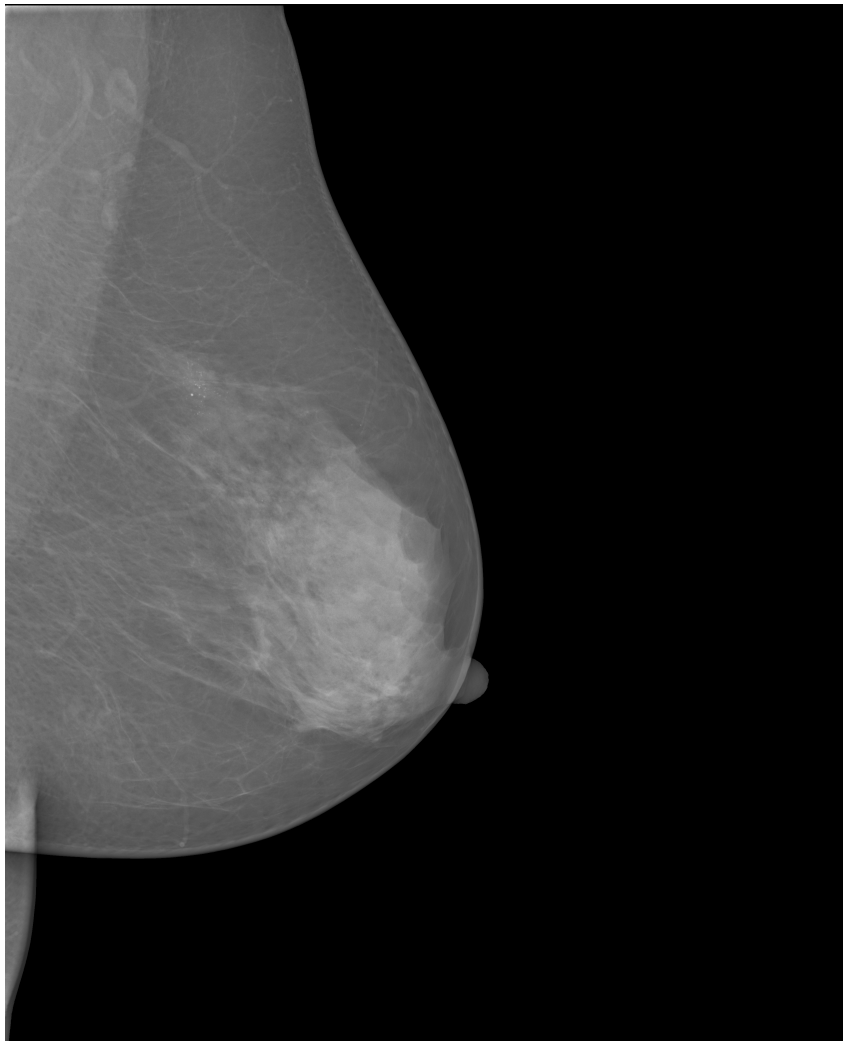


Figura 8 – Mamografia craniocaudal esquerda original da qual as imagens foram extraídas.

As imagens subsequentes são um recorte desta imagem original em diferentes estágios de processamento. A Figura 9 é a primeira delas, que expõe as microcalcificações antes de qualquer pré-processamento.

O pré-processamento de imagens é feito por meio de quatro etapas consecutivas: (i) Aumento do contraste; (ii) Subtração do fundo; (iii) Filtragem; (iv) *Thresholding*. O objetivo dessas técnicas é reduzir a quantidade de dados que são alimentados no algoritmo

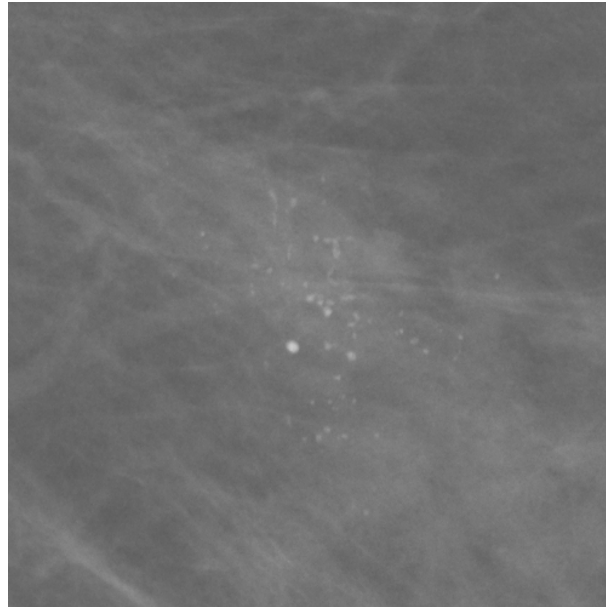


Figura 9 – Corte centrado em torno da microcalcificação antes do pré-processamento.

de treinamento, bem como aumentar sua qualidade para minimizar falsos positivos.

Torna-se relevante ressaltar, que a ordem das etapas, assim como a busca pelos melhores modelos de filtros foram feitas de maneira iterativa, combinando métodos de pré-processamento comuns na literatura biomédica e bens explorados por autores como Ted C. Wang, Nicolaos B. Karayiannis e Heng-Da Cheng, com outros métodos menos usuais, mas com alto potencial, e adaptando-os para os objetivos específicos deste trabalho (Wang; Karayiannis, 1998),(Cheng *et al.*, 2003b).

3.2.2.1 Aumento do contraste

O aumento do contraste é como se denomina a adaptação da curva de contraste. Foi utilizada uma técnica de equalização adaptativa do histograma com contraste limitado (CLAHE) devido às suas vantagens em relação à equalização global do histograma (GHE) e à equalização adaptativa do histograma (AHE). A GHE é uma técnica mais simples, que tenta expor a maior parte da imagem a um espectro de contraste mais amplo. Embora esse algoritmo se beneficie de um processamento mais simples, ele pode causar efeitos indesejáveis em imagens de raio-X, e em mamografias em específico, pois muitas vezes dá espectro de contraste ao espaço vazio do fundo devido à proporção da área em comparação com as informações realmente importantes dentro da imagem.

A AHE é uma técnica semelhante, que tenta contornar esse problema segmentando a imagem em várias zonas e aplicando a GHE individualmente dentro de cada zona. Embora isso contorne o principal problema do GHE, permite o aumento do contraste do ruído em segmentos relativamente uniformes, o que é conhecido por ser um problema para mamas densas.

CLAHE é a soma das duas técnicas anteriores com um limite (clip-limit) para evitar que uma largura de banda de contraste excessiva seja usada em pequenas flutuações de intensidade bruta. A Figura 10 mostra a região ao redor da microcalcificação após o aumento do contraste através do CLAHE.

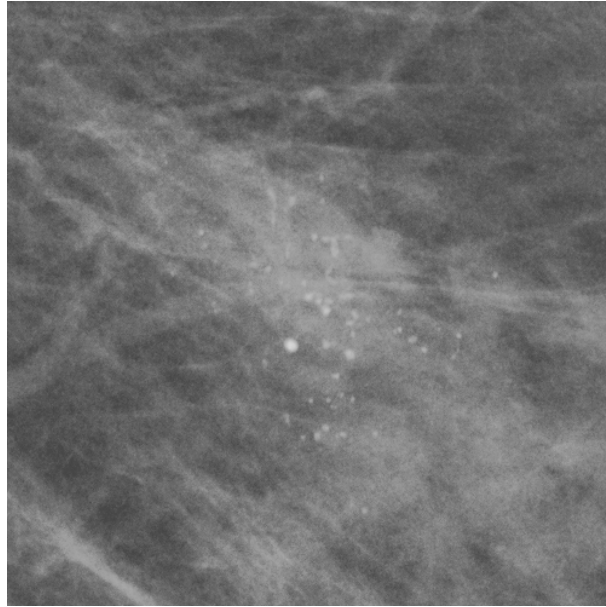


Figura 10 – Corte centrado na microcalcificação após a fase de aprimoramento do contraste ajustado.

3.2.2.2 Remoção do fundo

A remoção do fundo é feita instintivamente pelos olhos de um profissional treinado, que ignora as partes da imagem fora da mama. Para um algoritmo, no entanto, toda a imagem radiográfica é um dado útil, e ele não consegue diferenciar adequadamente entre a mama e o fundo.

Uma estratégia poderosa para remover flutuações indesejadas do fundo na radiografia é subtrair da imagem original seus próprios componentes de baixa frequência espacial. A frequência espacial é a relação entre o brilho do pixel e a distância: frequências mais altas são encontradas onde pixels pequenos têm brilho muito maior do que seus arredores, enquanto a frequência mais baixa possível é vista em grandes áreas com densidade uniforme, independentemente da intensidade alta ou baixa.

Uma maneira clássica de tornar a subtração possível é passar uma cópia da imagem original por um filtro passa-baixa, como um filtro gaussiano, e subtrair a imagem resultante da original. Também é importante garantir que não haja valores de intensidade negativos, o que poderia ser possível em fundos muito uniformes, como espaços vazios em radiografias. A remoção do fundo é evidente na Figura 11, que mostra que não remove grande parte da imagem ao redor da microcalcificação, mas sim nas partes vazias da mamografia.

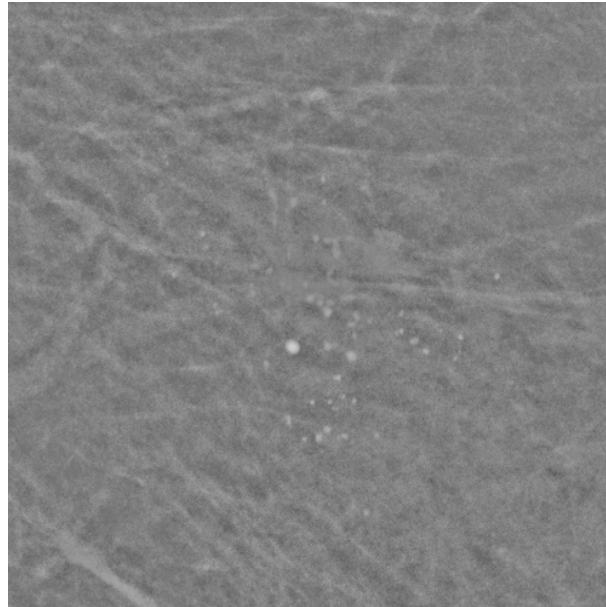


Figura 11 – Corte centrado na microcalcificação após a etapa de remoção do fundo.

3.2.2.3 Filtragem

Um filtro Sobel-Gauss-Sobel possui todas as propriedades de detecção de bordas de um filtro Sobel padrão, mas também oferece redução de ruído aprimorada e permanência de continuidade em contrastes não uniformes, que são duas características importantes em imagens médicas.

Esse filtro de três estágios consiste em três filtros individuais aplicados sucessivamente a uma imagem original. O primeiro filtro é aplicado diretamente à imagem original: um filtro Sobel é um filtro diferenciador, que atua como um “detector de bordas”. Ele procura sinais com grande inclinação, ou seja, altas frequências, e os isola. O resultado é um mapa de gradiente de intensidade e direção sobre a imagem original, que representa regiões com alto contraste localizado, ou seja, bordas vistas entre pixels brilhantes e opacos. Esse mapa é conectado a um filtro gaussiano passa-baixa, que atenua os componentes de alta frequência, resultando em uma imagem “borrada”. Isso é importante por dois motivos: a) Remove ruídos indesejados detectados como bordas no filtro anterior, como pontos de intensidade média cercados por todos os lados por intensidades ultra baixas e planas. B) Reconecta bordas finas, mas longas, muitas vezes não perfeitamente contínuas, comuns em imagens radiográficas reais.

A imagem resultante é então executada novamente através de um segundo filtro Sobel, que deve resultar em uma versão mais precisa da primeira. É aqui que a ação começa: a maior parte do fundo, que é quase monótono, é filtrada pelo filtro passa-alta, e a microcalcificação se destaca, como mostrado na Figura 12.

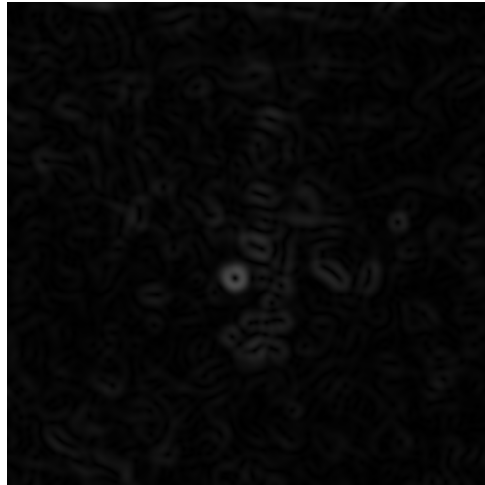


Figura 12 – Corte centrado na microcalcificação após a etapa de filtragem.

3.2.2.4 Thresholding

Esta etapa final do pré-processamento não é um filtro, mas uma solução baseada em dados para remover o ruído da imagem, limitando o tamanho dos dados que serão oferecidos para treinamento e aumentando sua qualidade.

Um histograma da imagem resultante da etapa anterior organiza os pontos por intensidade e frequência de exibição. Um limite alto de brilho é definido como uma porcentagem de todos os pixels, e aqueles que não ultrapassam o valor exigido têm seu brilho definido como zero, eliminando-os da imagem, o que efetivamente reduz a imagem apenas às áreas mais importantes para identificação de microcalcificações, no caso as regiões de maior intensidade de tom de cinza. A Figura 13 ilustra o final do pré-processamento: a microcalcificação foi isolada do tecido circundante.

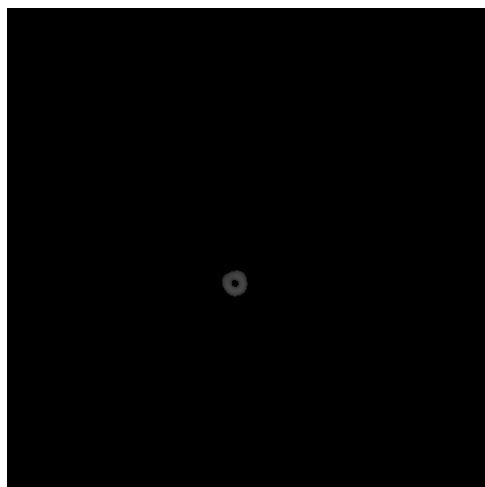


Figura 13 – Corte centrado na microcalcificação após a fase de *Thresholding*.

A Figura 14 recapitula e compara os resultados das principais etapas de pré-

processamento realizadas a as imagens resultantes em cada uma das fases.

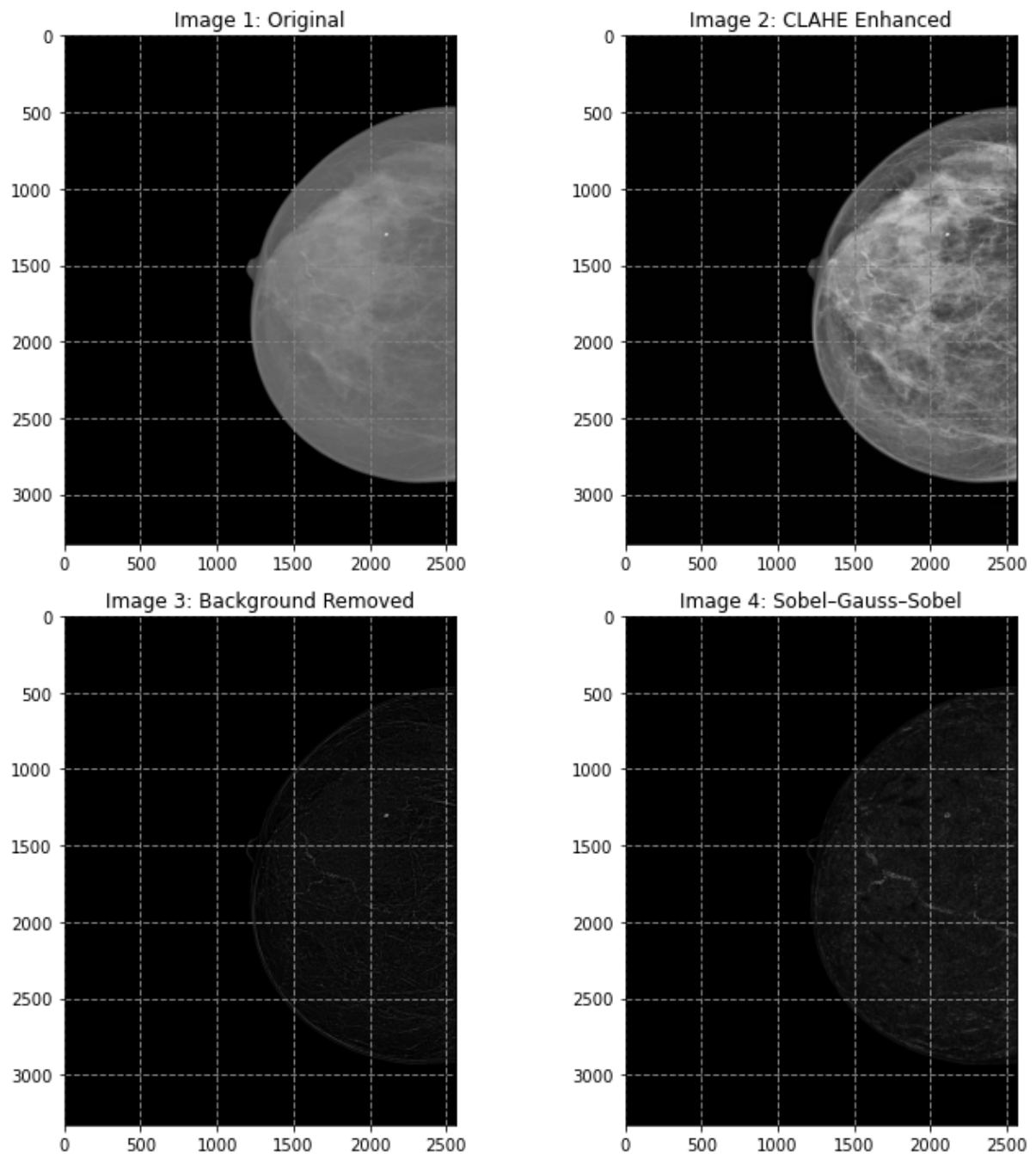


Figura 14 – Recapitulação das fases de pré-processamento.

3.3 Processamento

Após a fase de pré-processamento, as imagens estão aptas para a fase de processamento através do algoritmo Firefly. Como discutido na seção anterior o objetivo dessa etapa é segmentar regiões de interesse com potenciais aglomerados de microcalcificações para posterior análise médica.

O algoritmo foi compilado em linguagem Python e executado em um computador com processador Intel(R) Core(TM) de décima segunda geração, i7-1260P, 16 GB de memória RAM e 2.10 GHz. Dentre as bibliotecas usadas, destacam-se *numpy*, para computação científica, *matplotlib.pyplot*, para criação de gráficos, *pydicom* para leitura de imagens no formato DICOM e *OpenCV*, especializada em visão computacional e processamento de imagens.

Como resultado, obtiveram imagens re-processadas através de uma colorização, na qual cada cor representa um conjunto (ou *cluster*) identificado pelo algoritmo. O número de agrupamentos a serem identificados é uma das variáveis de entrada do algoritmo e dita o seu comportamento de otimização. A Figura 15 mostra um exemplo desse processo.

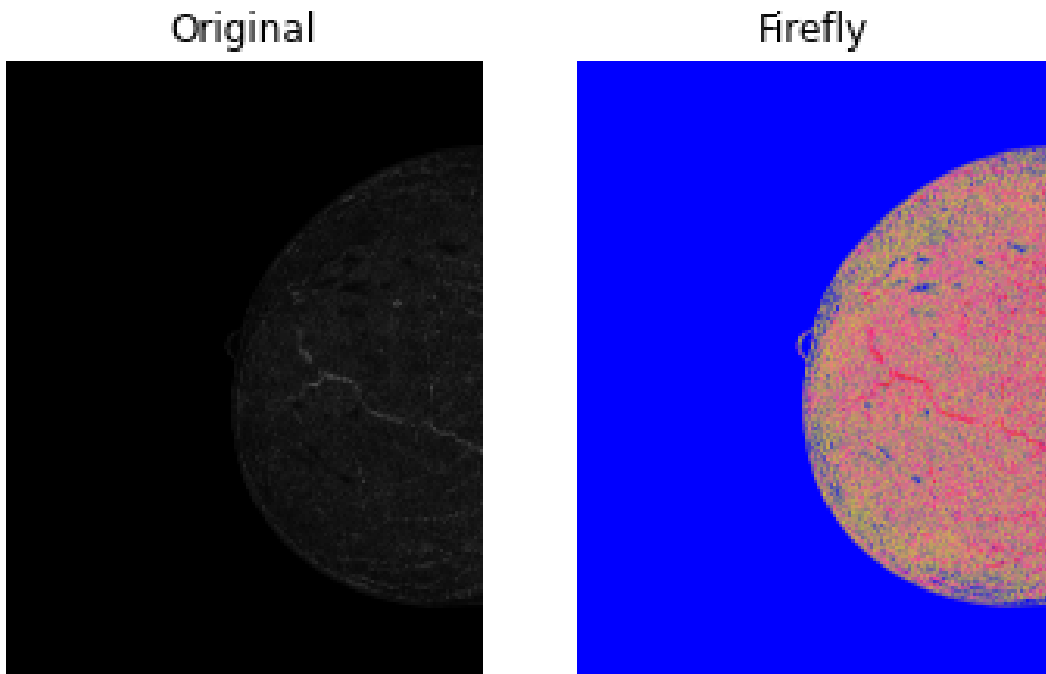


Figura 15 – Resultado do algoritmo de segmentação Firefly. Do lado esquerdo, uma imagem mamográfica antes do processamento. Do lado direito, a mesma imagem pós-processada, no qual cada agrupamento é representado por uma tonalidade.

O processo começa com a conversão da imagem bidimensional em um vetor unidimensional, pois o algoritmo funcionará otimizando a atribuição de cada pixel a um cluster específico dentro da imagem, com base em sua intensidade na escala de cinza. Após essa preparação, o algoritmo inicializa aleatoriamente sua população de vagalumes. Cada vagalume representa uma solução candidata — neste caso, um conjunto de centróides que definem os clusters usados para segmentar a imagem.

Em seguida, a função de custo é ativada para calcular as distâncias dos pixels aos centróides mais próximos. Essa função associa cada pixel ao seu cluster correspondente,

calcula as distâncias e retorna o custo total da solução intermediária. Após a avaliação inicial, o melhor vagalume é armazenado como uma solução candidata.

Então, começa a fase iterativa do algoritmo: em cada iteração, cada vagalume é comparado com todos os outros. Se outro elemento tiver um custo menor (ou seja, for “mais brilhante”), o vagalume atual se move em sua direção. Esse movimento é calculado combinando três componentes: a posição atual, um termo atrativo proporcional à diferença entre as soluções e um termo aleatório que promove a exploração do espaço de pesquisa. A força de atração diminui com a distância entre os vagalumes e é modulada pelos parâmetros do coeficiente de absorção de luz e atratividade máxima.

O termo aleatório permite que o algoritmo explore novas regiões, enquanto seu amortecimento gradual reduz o comportamento aleatório ao longo das iterações, favorecendo a convergência. A nova posição gerada é avaliada e, se apresentar um custo melhor, o vagalume é atualizado. Se essa solução for a melhor encontrada até o momento, ela também é registrada como a melhor solução global.

Após calcular todos os novos valores da iteração, o código combina a população antiga com a nova, classifica todas as soluções por custo e mantém apenas as melhores. Esse mecanismo garante que soluções promissoras não sejam descartadas ao longo do processo. O parâmetro de amplitude do termo aleatório é reduzido progressivamente, limitando a aleatoriedade da busca à medida que o algoritmo se aproxima de uma solução estável. A cada iteração, o custo da melhor solução é armazenado, permitindo a visualização posterior da curva de convergência.

Quando o número máximo de iterações é atingido, o algoritmo considera que encontrou sua melhor configuração de centróides. Na sequência, os índices dos agrupamentos para cada pixel são recuperados e reorganizados na forma da imagem original. Esses índices constituem o mapa de segmentação final. O código então aplica cores personalizadas a cada cluster, gerando uma imagem segmentada colorida.

Por fim, a imagem é repartida em regiões de análise, e em cada uma delas, os clusters são avaliados com base na intensidade média do nível de cinza do pixel original a fim de identificar possíveis microcalcificações. Entretanto, o objetivo final é identificar aglomerados de microcalcificações, e não apenas amostras isoladas com pouca relevância médica. Por esse motivo, o algoritmo calcula a porcentagem de área ocupada pelo potencial conjunto de microcalcificações e compara com um limite pré-definido (obtido empiricamente) a fim de separar regiões de interesse de ruídos.

Empiricamente, notou-se que o modelo, por si só, ainda não era capaz de distinguir microcalcificações de fibras mamárias, ambas estruturas de alto contraste e alta intensidade em tons de cinza: em 73% dos casos analisados, o algoritmo classificou no mesmo agrupamento ambas as estruturas.

Para contornar esse problema, adotou-se uma segunda camada de análise baseada no formato das amostras. A circularidade foi definida então, como um parâmetro de distinção entre fibras, usualmente em formato alongado e retilíneo, e microcalcificações, em formato de anéis circulares. A Figura 16 ilustra um exemplo da aplicação do modelo de distinção entre estrias e microcalcificações. Desse modo, apenas regiões com elevados índices de circularidade são selecionadas como áreas de interesse.

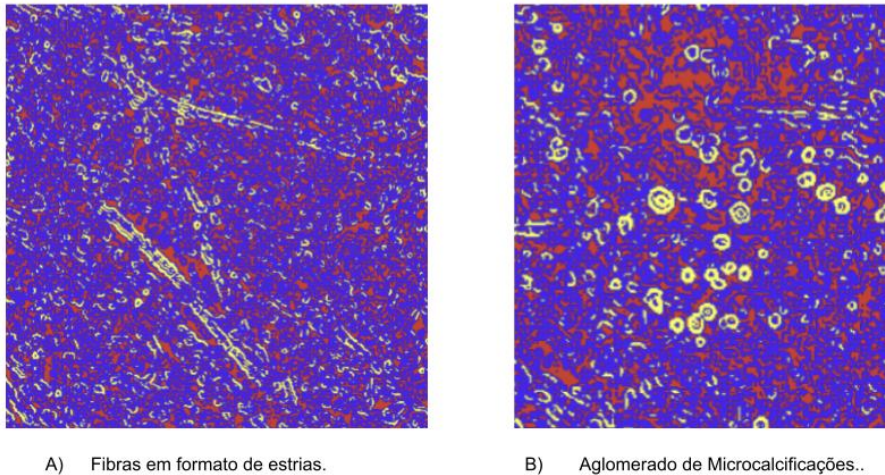


Figura 16 – A circularidade é um elemento chave na distinção entre estrias de alta intensidade e conjunto de microcalcificações. No lado direito da Figura, um conjunto de microcalcificações classificadas dentro do conjunto de interesse pelo algoritmo Firefly. No lado esquerdo, fibras em formatos de estrias também classificadas dentro do mesmo conjunto, mas em formatos alongado e não-circular.

Ao final foram analisadas 40 amostras, 20 com microcalcificações apontadas pelo relatório médico e 20 sem anomalias. Como saída, o algoritmo identifica regiões retangulares com potenciais aglomerados de microcalcificações. Uma máscara verde é inserida no perímetro da área de interesse sobre a imagem pós-processada para auxiliar na identificação, detecção e posterior diagnóstico pelo médico responsável.

4 RESULTADOS E DISCUSSÃO

Dois conjuntos de parâmetros foram testados, o modelo A e o modelo B em uma mesma imagem mamográfica. A diferença entre cada um dos modelos está na escolha dos parâmetros chaves. A Tabela 2 traz alguns valores empiricamente definidos para cada um deles. Ainda que o coeficiente de absorção de luz e o valor base do coeficiente de atração sejam iguais, o número máximo de iterações, assim como o número de vagalumes, foram significativamente maiores no Modelo B do que no Modelo A.

Tabela 2 – Parâmetros dos Modelos no Algoritmo dos Vagalumes.

# Parâmetros	Modelo A	Modelo B
Número máximo de iterações	5	20
Número de vaga-lumes	5	20
Coefficiente de absorção de luz	1	1
Valor base do coeficiente de atração	0,2	0,2

Como resultado, na imagem referência da Figura 17, o modelo A identificou um total de 61 regiões de interesse, das quais, apenas uma se trata, de fato, de microcalcificações. O modelo B, por sua vez, foi capaz de identificar apenas duas zonas suspeitas. Uma zona na parte superior esquerda da imagem, sem presença de microcalcificações, e outra próxima ao centro da mama, com microcalcificações.

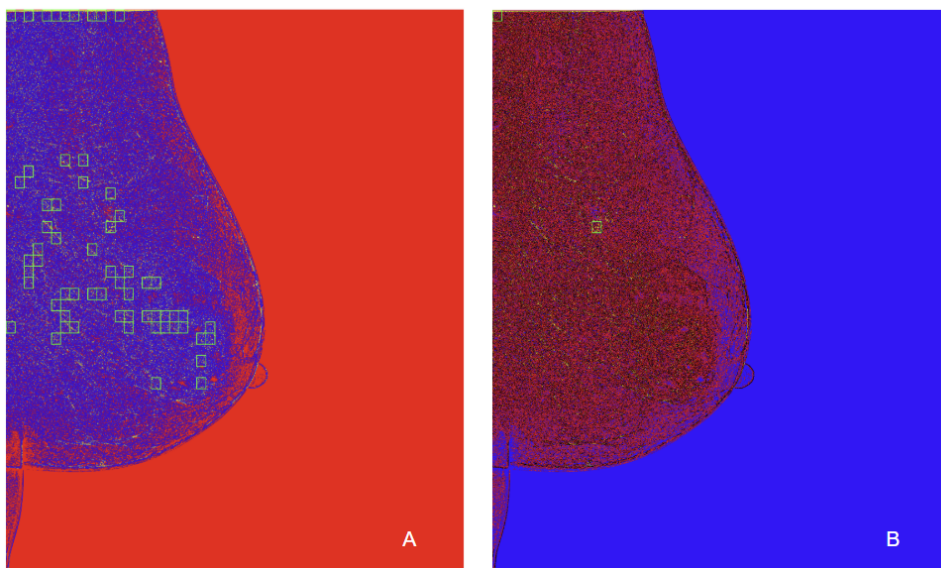


Figura 17 – Comparação na classificação de regiões de interesse para dois conjunto de parâmetros dentro do algoritmo Firefly. Destaque para maior especificidade do modelo B frente ao modelo A.

Deve-se destacar que grande parte da região superior da imagem, na fronteira da mama, sofre com efeitos de borda gerando ruídos de alta intensidade que, em certa medida podem ser facilmente lidos como aglomerados de interesse. Como medida corretiva, optou-se por excluir regiões periféricas da imagem, destacando apenas zonas onde há maior chance de formação de microcalcificações e menor interferência de ruídos externos.

Feitas tais considerações, pode-se afirmar que o modelo B mantém a elevada sensibilidade do modelo A, ao identificar a região de interesse, mas é significativamente mais preciso, pois não retorna falsos positivos e portanto, foi escolhido como referência para categorização das demais imagens.

A curva da Figura 18 mostra a evolução do melhor valor de custo (função objetivo ou melhor resultado) para o modelo B ao longo de um conjunto de iterações. A análise da função custo permite refinar o algoritmo e determinar a melhor relação entre custo computacional e qualidade da solução (ponto imediatamente anterior a saturação da curva).

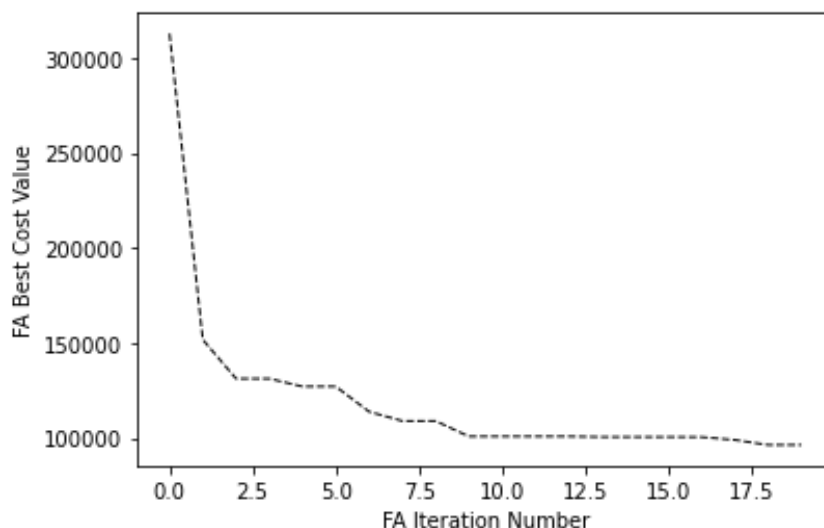


Figura 18 – Evolução da função objetivo ao longo das iterações do algoritmo. A análise da função permite refinar o modelo e determinar a melhor relação entre custo computacional e qualidade da solução em relação ao número de iterações necessárias

Por fim, utilizando o modelo B, foi feita uma análise sobre uma nova imagem, dessa vez sem recorrência de microcalcificações nos relatórios médicos, a fim de se constatar a capacidade de identificação de verdadeiros negativos pelo algoritmo. O resultado é mostrado na imagem da Figura 19. Nesse caso, verifica-se que o modelo B acerta ao não identificar nenhuma região de interesse com potenciais aglomerados de microcalcificações.

Para além de identificar microcalcificações no tecido da mama, algumas imagens estudadas também possuíam calcificações vasculares mamárias (CVM). As CVM são denominadas calcificações de Mönckeberg e são tipicamente alterações benignas que não

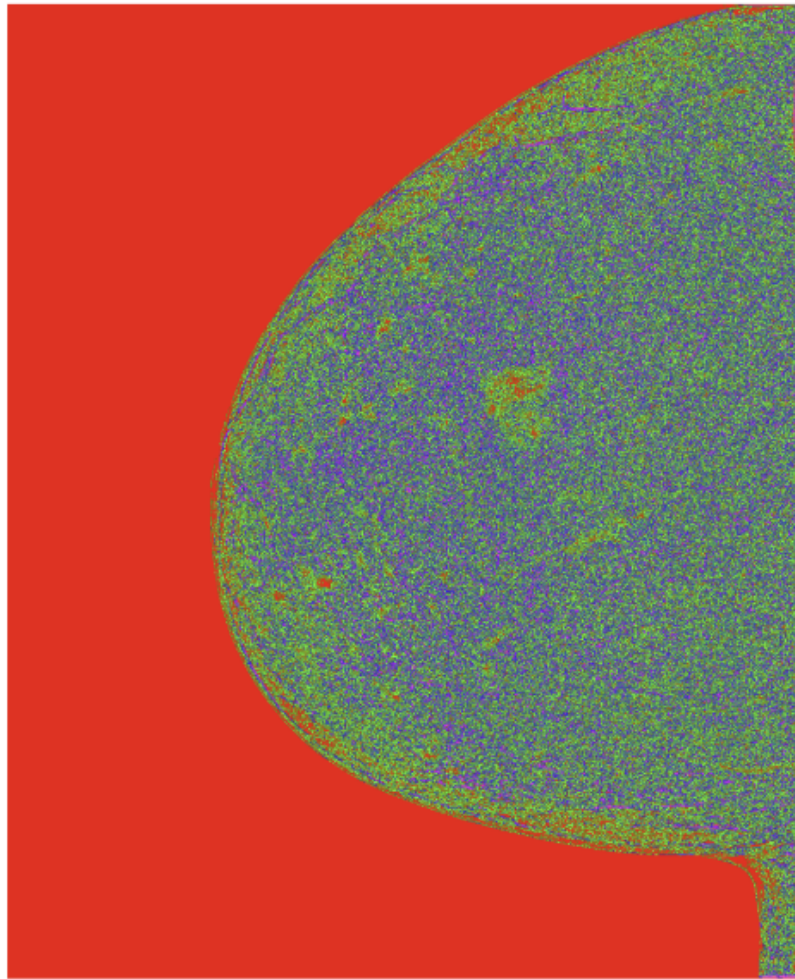


Figura 19 – Exemplo de imagem mamográfica sem recorrência de microcalcificações após o processamento pelo algoritmo Firefly

estão associadas a um maior risco de desenvolvimento de neoplasias malignas (Freitas-Júnior; Murta; Oliveira, 2009). Ainda que não esteja diretamente relacionada ao câncer de mama, diversos trabalhos na literatura biomédica têm apontado uma associação entre as CVM a outras entidades clínicas, incluindo o diabetes, a hipertensão arterial sistêmica, as coronariopatias, a insuficiência renal, a neuropatia autonômica e a hipervitaminose D (Moe *et al.*, 2002),(Maas *et al.*, 2004),(Çetin *et al.*, 2004), o que ressalta a importância da sua identificação no momento do exame mamográfico.

A categorização entre microcalcificações e CVMs seria essencialmente papel do médico responsável, uma vez que o algoritmo, nessa fase de desenvolvimento, ainda não seria capaz de apontar a diferença entre as duas estruturas. Uma proposta de evolução futura desse estudo seria justamente incluir análises mais profundas de formato, não apenas dentro de uma região suspeita, mas entre regiões, a fim de identificar padrões de alinhamento e apontar para a ocorrência de calcificações vasculares mamárias.

Ao todo, quarenta imagens foram analisadas, 20 com estruturas suspeitas (mi-

crocalcificações ou CVMs) e vinte sem nenhuma anomalia. Inicialmente, analisou-se se o sistema é capaz de identificar uma ou mais estruturas suspeitas em cada uma das imagens, ignorando a sua posição ou se de fato representaria microcalcificações descritas pelo relatório médico. Nesse caso, dentre as imagens com estruturas suspeitas, o modelo identificou 85% das imagens. Em apenas 15% das mamografias com zonas suspeitas, o algoritmo não foi capaz de identificar nenhuma região suspeita.

As Tabelas 4 e 5 no Apêndice A sintetizam as principais informações sobre cada uma das imagens analisadas, tais como: a) o lado da mama (direito ou esquerdo), b) a incidências do exames de imagem, c) a classificação ACR com Categoria de Densidade Mamária, d) a Categoria de Avaliação Final (*ACR BI-RADS® Assessment Category*), assim como e) eventuais comentários do médico responsável e por fim, f) a saída do algoritmo (se ao menos um aglomerado de microcalcificações foi identificado).

Para as imagens sem microcalcificações, foram identificadas erroneamente zonas de interesse em 25% das mamografias. Em 75% destas, portanto, o método acompanhou o relatório médico e não identificou nenhum alerta. A Tabela 3 sintetiza esses resultados.

Tabela 3 – Resultados do Algoritmo Firefly

Resultado	Verdadeiro	Falso
Positivo	85%	15%
Negativo	75%	25%

Nesse caso, pode-se concluir que o modelo apresentou uma sensibilidade de 85% e uma especificidade de 75%. Dentre as limitações do modelo, o baixo número de amostras analisadas, apenas 40 imagens: 20 com microcalcificações descritas no relatório médico e 20 sem regiões de interesse, pode ter um impacto sensível nos indicadores apresentados. De todo modo, dentre as imagens com microcalcificações, o sistema identificou em média 1,76 regiões de interesse por imagem, entretanto, nem todas as regiões identificadas, de fato, merecem atenção médica.

Em alguns casos, as bordas das imagens (superiores e inferiores) ou até mesmo alterações fibrocísticas acabaram sendo erroneamente classificadas como áreas de interesse. Em geral, dentre as amostras com microcalcificações identificadas e validadas pelo relatório médico, cerca de 83% das amostras eram de fato aglomerados de microcalcificações, enquanto que 17% das regiões identificadas não possuíam estruturas de interesse. Destas, 12% se localizavam em regiões de borda e apenas 5% em regiões centrais da mamografia.

Portanto, uma proposição de melhoria e continuidade desse estudo seria justamente excluir regiões de borda, com baixa probabilidade de associação com microcalcificações, buscando assim reduzir significativamente a taxa de falsos positivos e melhorar a especificidade do modelo proposto. De todo modo, o algoritmo Firefly se mostrou uma ferramenta promissora dentro de técnicas de auxílio computacional ao diagnóstico médico, com efeitos

sensíveis no ganho de produtividade, redução de custos operacionais e otimização do trabalho médico.

5 CONCLUSÃO

Este estudo teve como objetivo desenvolver e avaliar uma nova estratégia para detectar microcalcificações em mamografias, combinando técnicas automatizadas de pré-processamento com o algoritmo Firefly, inspirado na natureza. Os objetivos principais foram criar uma metodologia eficaz de aprimoramento de imagens, avaliar o impacto do ajuste de parâmetros no desempenho do modelo, implementar um algoritmo ajustado para detecção e desenvolver um método para isolar claramente as regiões de interesse para os radiologistas.

O modelo final desenvolvido atingiu os objetivos. Ao comparar dois modelos com configurações de parâmetros distintos, este trabalho demonstrou que o desempenho da estrutura proposta é altamente dependente do ajuste dos parâmetros, fazendo a diferença entre um modelo excelente e um modelo com baixo desempenho. A configuração com melhor desempenho, o Modelo B, alcançou resultados positivos nas imagens de teste, com uma taxa de sensibilidade de 85%, e de especificidade em 75%, concordando com a avaliação do radiologista. Além disso, o modelo se mostrou capaz de identificar a região exata da microcalcificações em 83% das zonas suspeitas. Para aprimorar ainda mais esse resultados, pesquisas futuras devem alterar os parâmetros definidos empiricamente nesse estudo, aumentar a base de amostras e eliminar regiões de borda, com alta concentração de falsos-positivos pela ocorrência de pontos de alta frequência.

O desempenho é promissor e sua precisão é comparável e, em alguns casos, superior à de outras metodologias relatadas na literatura recente. Ainda assim, o escopo das descobertas deste estudo é limitado pelo tamanho do conjunto de dados. Portanto, trabalhos futuros são necessários para validar a eficácia do algoritmo em um repositório grande e diversificado de mamografias clínicas para estabelecer seu poder estatístico.

O método proposto representa, assim, uma estratégia promissora para a detecção de microcalcificações por *CAD*. Com validação adicional, essa abordagem tem o potencial de se tornar uma ferramenta inestimável no diagnóstico assistido por computador, auxiliando os radiologistas na triagem do câncer de mama.

REFERÊNCIAS

- ANGELO, M. F. **Sistema de processamento de imagens mamográficas e auxílio ao diagnóstico via internet**. 2007. Tese (Tese (Doutorado em Engenharia Elétrica)) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2007.
- BREM, R. F. *et al.* A computer-aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size. **American Journal of Roentgenology**, American Roentgen Ray Society, v. 184, n. 3, p. 893–896, 2005.
- BURNSIDE, E. S. *et al.* Use of microcalcification descriptors in bi-rads 4th edition to stratify risk of malignancy. **Radiology**, v. 242, n. 2, p. 388–395, 2007. PMID: 17255409. Disponível em: <https://doi.org/10.1148/radiol.2422052130>.
- CALAS, M. J. G.; GUTFILEN, B.; PEREIRA, W. C. d. A. Cad e mamografia: por que usar esta ferramenta? **Radiologia Brasileira**, SciELO Brasil, v. 45, p. 46–52, 2012.
- CARITÁ, E. C. *et al.* Implantação de pacs com suporte à recuperação de imagens baseada em conteúdo em hospital universitário. *In: X Congresso Brasileiro de Informática em Saúde, Florianópolis–SC*. [S.l.: s.n.], 2006.
- ÇETIN, M. *et al.* Breast arterial calcifications associated with diabetes and hypertension. **Journal of diabetes and its complications**, Elsevier, v. 18, n. 6, p. 363–366, 2004.
- CHENG, H. *et al.* Computer-aided detection and classification of microcalcifications in mammograms: a survey. **Pattern Recognition**, v. 36, n. 12, p. 2967–2991, 2003. ISSN 0031-3203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320303001924>.
- CHENG, H.-D. *et al.* Computer-aided detection and classification of microcalcifications in mammograms: a survey. **Pattern recognition**, Elsevier, v. 36, n. 12, p. 2967–2991, 2003.
- CRAWFORD, B. *et al.* Binary firefly algorithm for the set covering problem. *In: IEEE. 2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*. [S.l.: s.n.], 2014. p. 1–5.
- CRUZ, Y. L. K. *et al.* Segmentação de microcalcificações em imagens mamográficas utilizando algoritmo firefly. **J. health inform**, p. 631–642, 2016.
- DELIJA, M.; MIHANOVIĆ, F. Swot analysis of the application of artificial intelligence in radiologic technology and radiology. **Radiological Jurnal**, v. 48, n. 2, 2024.
- DELLANI, P. R. **Desenvolvimento de um servidor de imagens médicas digitais no padrão DICOM**. 2001. Dissertação (Dissertação (Mestrado)) — Universidade Federal de Santa Catarina, Centro Tecnológico, Florianópolis, SC, 2001.
- DUARTE, M. *et al.* Automatic microcalcifications segmentation procedure based on otsu's method and morphological filters. *In: IEEE. 2011 Pan American Health Care Exchanges*. [S.l.: s.n.], 2011. p. 102–106.

EBERHART, R.; KENNEDY, J. A new optimizer using particle swarm theory. *In: IEEE. MHS'95. Proceedings of the sixth international symposium on micro machine and human science.* [S.l.: s.n.], 1995. p. 39–43.

ENGELBRECHT, A. P. Fundamentals of computational swarm intelligence. **(No Title)**, 2005.

FAZAL, M. I. *et al.* The past, present and future role of artificial intelligence in imaging. **European journal of radiology**, Elsevier, v. 105, p. 246–250, 2018.

FIGUEIREDO, A. A. A.; FERNANDES, H. C.; GUIMARAES, G. Experimental approach for breast cancer center estimation using infrared thermography. **Infrared Physics & Technology**, Elsevier, v. 95, p. 100–112, 2018.

FREITAS, A. G. d. *et al.* Mamografia digital: perspectiva atual e aplicações futuras. **Radiologia Brasileira**, SciELO Brasil, v. 39, p. 287–296, 2006.

FREITAS-JÚNIOR, R.; MURTA, E. F. C.; OLIVEIRA, E. L. C. **Significado clínico das calcificações vasculares na mamografia: devemos valorizá-las?** [S.l.: s.n.]: SciELO Brasil, 2009. 377–379 p.

GERAS, K. J.; MANN, R. M.; MOY, L. Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives. **Radiology**, Radiological Society of North America, v. 293, n. 2, p. 246–259, 2019.

GILLIES, R. J.; KINAHAN, P. E.; HRICAK, H. Radiomics: images are more than pictures, they are data. **Radiology**, Radiological Society of North America, v. 278, n. 2, p. 563–577, 2016.

GONÇALVES, C. B. *et al.* Algoritmos bioinspirados para otimização de cnns aplicados na detecção de câncer de mama utilizando imagens termográficas. Universidade Federal de Uberlândia, 2022.

GUERREIRO, A. A. P. *et al.* Integrando inteligência artificial à mamografia: uma abordagem complementar no diagnóstico do câncer de mama. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 10, n. 5, p. 479–485, 2024.

HOLLAND, J. H. Adaptation in natural and artificial systems. **University of Michigan Press google schola**, v. 2, p. 29–41, 1975.

JR, J. R. F.; AZEVEDO-MARQUES, P. M. de; OLIVEIRA, M. C. Selecting relevant 3d image features of margin sharpness and texture for lung nodule retrieval. **International journal of computer assisted radiology and surgery**, Springer, v. 12, n. 3, p. 509–517, 2017.

KENNEDY, J.; EBERHART, R. Particle swarm optimization. *In: IEEE. Proceedings of ICNN'95-international conference on neural networks.* [S.l.: s.n.], 1995. v. 4, p. 1942–1948.

KIM, S. *et al.* Microcalcifications, mammographic breast density, and risk of breast cancer: a cohort study. **Breast Cancer Research**, v. 24, n. 1, p. 96, 2022. Disponível em: <https://doi.org/10.1186/s13058-022-01594-0>.

KOHLI, A.; JHA, S. Why cad failed in mammography. **J Am Coll Radiol**, v. 15, n. 3 Pt B, p. 535–537, 2018.

LI, T. *et al.* Global status and attributable risk factors of breast, cervical, ovarian, and uterine cancers from 1990 to 2021. **Journal of Hematology Oncology**, v. 18, n. 1, p. 5, 2025. Disponível em: <https://doi.org/10.1186/s13045-025-01660-y>.

LOGULLO, A. F. *et al.* Breast microcalcifications: Past, present and future (Review). **Molecular and Clinical Oncology**, v. 16, n. 4, p. 81, apr 2022.

MAAS, A. H. *et al.* Prevalence and determinants of breast arterial calcium in women at high risk of cardiovascular disease. **The American journal of cardiology**, Elsevier, v. 94, n. 5, p. 655–659, 2004.

MARCON, M. *et al.* ESR Essentials: screening for breast cancer - general recommendations by EUSOBI. **European Radiology**, v. 34, p. 6348–6357, 2024. Disponível em: <https://doi.org/10.1007/s00330-024-10740-5>.

MARTINS, A. R. **Desenvolvimento de um sistema de gerenciamento do processamento de impressão de imagens médicas digitais utilizando o protocolo DICOM**. 2006. Dissertação (Dissertação (Mestrado em Engenharia Elétrica)) — Universidade de Brasília, Brasília, DF, 2006. 200 f.

MARTINS, M. d. C. A matemática na natureza. **Correio dos Açores**, Gráfica Açoreana, Lda., p. 17–17, 2014.

MIRANDA, A.; ALVES, D.; MIRANDA, J. Desafios da implementação de inteligência artificial na detecção de cancro de mama. **ROENTGEN-Revista Científica das Técnicas Radiológicas**, v. 6, n. 2, p. 34–40, 2025.

MOE, S. M. *et al.* Medial artery calcification in esrd patients is associated with deposition of bone matrix proteins. **Kidney international**, Elsevier, v. 61, n. 2, p. 638–647, 2002.

NAM, J. G. *et al.* Development and validation of deep learning–based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. **Radiology**, Radiological Society of North America, v. 290, n. 1, p. 218–228, 2019.

PERRE, A. C.; FREIRE, L. C. Avaliação da caracterização de lesões em mamografia com recurso a sistemas cad. **ROENTGEN-Revista Científica das Técnicas Radiológicas**, v. 2, n. 2, p. 41–49, 2021.

PORTO, F. N. **Análise de sensibilidade de um sistema CAD para mamografia digital**. 2010. Dissertação (Dissertação (Mestrado em Engenharia Elétrica)) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2010.

POTNIS, K. C. *et al.* Artificial intelligence in breast cancer screening: evaluation of fda device regulation and future recommendations. **JAMA internal medicine**, American Medical Association, v. 182, n. 12, p. 1306–1312, 2022.

ROCHA, J. P. A. *et al.* Metaheurísticas aplicadas à detecção, localização e classificação de tumores mamários compostos por múltiplos focos via termografia. *In*: SBC. **Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)**. [S.l.: s.n.], 2024. p. 61–66.

RODRIGUES, C. Sistemas cad em patologia mamária. **Faculdade de Engenharia da Universidade do Porto, Setembro, 2008.**

SANTOS, M. K. *et al.* Morphological computed tomography features of surgically resectable pulmonary squamous cell carcinomas: impact on prognosis and comparison with adenocarcinomas. **European Journal of Radiology**, Elsevier, v. 83, n. 7, p. 1275–1281, 2014.

SERAPIÃO, A. B. d. S. Fundamentos de otimização por inteligência de enxames: uma visão geral. **Sba: Controle & Automação Sociedade Brasileira de Automatica**, SciELO Brasil, v. 20, p. 271–304, 2009.

SHI, J. *et al.* The screening value of mammography for breast cancer: an overview of 28 systematic reviews with evidence mapping. **Journal of Cancer Research and Clinical Oncology**, v. 151, p. 102, 2025. Disponível em: <https://doi.org/10.1007/s00432-025-06122-z>.

SILVA, J. D. D. E. *et al.* Breast cancer mortality in young women in brazil. **Frontiers in Oncology**, Frontiers Media SA, v. 10, p. 569933, 2021.

SILVA, S. V. Reconstrução da geometria da mama a partir de imagens termográficas. **Universidade Federal Fluminense, Nitéroi, 2010.**

SOUZA, F. H. **Mamografia digital em comparação com mamografia convencional no rastreamento de câncer de mama no Brasil: revisão sistemática, custo da doença e análise de custo-efetividade no Sistema Único de Saúde.** 2012. Tese (Tese (Doutorado em Epidemiologia)) — Universidade Federal do Rio Grande do Sul, Faculdade de Medicina, Programa de Pós-Graduação em Epidemiologia, Porto Alegre, RS, 2012. Disponível em: Biblioteca Digital de Teses e Dissertações da UFRGS.

TAO, Q. *et al.* Deep learning-based method for fully automatic quantification of left ventricle function from cine mr images: a multivendor, multicenter study. **Radiology**, Radiological Society of North America, v. 290, n. 1, p. 81–88, 2019.

TOMAL, A. **Medidas experimentais dos coeficientes de atenuação de tecidos mamários e sua influência no contraste e dose mamográfica.** 2007. Tese (Doutorado) — Universidade de São Paulo, 2007.

TOT, T. *et al.* The clinical value of detecting microcalcifications on a mammogram. **Seminars in Cancer Biology**, v. 72, p. 165–174, 2021. ISSN 1044-579X. Precision Medicine in Breast Cancer. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1044579X19303566>.

WANG, J.; YANG, Y. A context-sensitive deep learning approach for microcalcification detection in mammograms. **Pattern Recognition**, v. 78, p. 12–22, 2018. ISSN 0031-3203. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0031320318300086>.

WANG, T. C.; KARAYIANNIS, N. B. Detection of microcalcifications in digital mammograms using wavelets. **IEEE transactions on medical imaging**, IEEE, v. 17, n. 4, p. 498–509, 1998.

APÊNDICES

APÊNDICE A – SELEÇÃO DE IMAGENS E RESULTADOS

As informações relativas às 40 imagens analisadas da base IBrest (20 imagens com microcalcificações e 20 sem microcalcificações), assim como o resultado do algoritmo Firefly são sintetizadas na tabela abaixo. Nela, cada linha representa uma amostra, com suas respectivas informações sobre: a) Lado da mama: onde *R* representa o lado direito e *L* o lado esquerdo; b) incidências do exames de imagem: onde Craniocaudal é representado por *CC* e mediolateral por *MLO*, c) a classificação ACR, com Categoria de Densidade Mamária e d) Categoria de Avaliação Final (*ACR BI-RADS Assessment Category*), assim como e) eventuais comentários do médico responsável, f) o resultado do algoritmo, onde *Identificada* significa que pelo menos uma região de interesse com suspeita de conter aglomerados de microcalcificações foi encontrada pelo algoritmo, enquanto que *Não Identificada* significa que nenhuma região de interesse foi encontrada.

Tabela 4 – Ficha das imagens **com microcalcificações**

Lado	Incidências	ACR	Bi-Rads	Comentários do médico	Resultado Firefly
R	CC	2	5	nódulo QSE + micros	Identificada
L	CC	2	2	nodulo + micros	Identificada
R	MLO	2	5	nódulo QSE + micros	Identificada
L	MLO	2	2	nódulo + micros	Identificada
L	MLO	3	2	micro	Identificada
R	MLO	3	4a	nodulo QSE + micros	Identificada
L	CC	3	2	micro	Identificada
R	CC	3	4a	nodulo QSE + micros	Não Identificada
L	MLO	1	4c	Nódulo + micros	Identificada
R	MLO	1	4c	Nódulo + micros	Identificada
L	CC	1	4c	nódulo +micros	Identificada
R	CC	1	4c	Nódulo + micros	Identificada
R	CC	3	2	micros + calcificação vasos	Identificada
L	MLO	3	5	micros	Não Identificada
R	MLO	3	2	micros + calcificação vasos	Não Identificada
L	CC	3	5	micros	Identificada
L	MLO	1	3	micros	Identificada
R	MLO	1	3	micros	Identificada
L	CC	1	3	micros	Identificada
R	MLO	3	2	micro	Identificada

Tabela 5 – Ficha das imagens **sem microcalcificações**

Lado	Incidências	ACR	Bi-Rads	Comentários do médico	Resultado Firefly
R	CC	4	1	normal	Identificada
R	MLO	4	1	normal	Não Identificada
L	CC	2	1	normal	Não Identificada
L	MLO	2	1	normal	Não Identificada
L	CC	1	1	normal	Não Identificada
L	MLO	1	1	normal	Identificada
L	MLO	1	1	normal	Não Identificada
L	CC	1	1	normal	Não Identificada
L	CC	1	1	normal	Identificada
L	MLO	1	1	normal	Não Identificada
R	CC	2	1	normal	Identificada
R	MLO	2	1	normal	Não Identificada
L	MLO	2	1	normal	Não Identificada
L	CC	2	1	normal	Não Identificada
R	MLO	1	1	normal	Não Identificada
R	CC	2	1	normal	Não Identificada
L	MLO	1	1	normal follow up	Identificada
R	MLO	1	1	normal follow up	Não Identificada
L	CC	1	1	normal follow up	Não Identificada
L	CC	1	1	normal follow up	Não Identificada