

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

Fernanda Junqueira Fortes Cardoso

**Desenvolvimento de um esquema computacional para
testes comparativos entre imagens mamográficas digitais
reais e simuladas: investigando a utilização de um novo
simulador de mama para avaliação de esquemas do tipo
CAD.**

São Carlos

2024

Fernanda Junqueira Fortes Cardoso

Desenvolvimento de um esquema computacional para testes comparativos entre imagens mamográficas digitais reais e simuladas: investigando a utilização de um novo simulador de mama para avaliação de esquemas do tipo CAD.

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

Orientador: Prof. Dr. Homero Schiabel

**São Carlos
2024**

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).

C268d Cardoso, Fernanda Junqueira Fortes
Desenvolvimento de um esquema computacional para
testes comparativos entre imagens mamográficas digitais
reais e simuladas: investigando a utilização de um novo
simulador de mama para avaliação de esquemas do tipo
CAD. / Fernanda Junqueira Fortes Cardoso; orientador
Homero Schiabel. São Carlos, .

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

1. Imagens digitais de mamografia. 2. Phantom
mamográfico. 3. Percepção visual em mamografia. I.
Título.

FOLHA DE APROVAÇÃO

Nome: Fernanda Junqueira Fortes Cardoso

Título: "Desenvolvimento de um esquema computacional para testes comparativos entre imagens mamográficas digitais reais e simuladas: investigando a utilização de um novo simulador de mama para avaliação de esquemas do tipo CAD."

Trabalho de Conclusão de Curso defendido e aprovado
em 22/03/2024,

com NOTA 10,0 (Dez, zero), pela Comissão Julgadora:

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

Profa. Associada Ana Claudia Patrocínio - UFU

Mestre Luciana Buffa Verçosa - Médica Radiologista Clínica Atom e
HU-UFSCar

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

AGRADECIMENTOS

Ao Prof. Dr. Homero Schiabel pela sua orientação dedicada e valiosa ao longo deste trabalho, contribuindo significativamente para o meu crescimento acadêmico e profissional.

À Dra. Luciana B. Verçosa, ao Oswaldo Jorge e à Prof. Eny M. Ruberti por sua assistência técnica e sugestões ao longo do processo de desenvolvimento deste trabalho.

À Gerência de Ensino e Pesquisa do Hospital Universitário da Universidade Federal de São Carlos (SP, Brasil) - Empresa Brasileira de Serviços Hospitalares (EBSERH) e ao Hospital das Clínicas da USP pelo auxílio na obtenção dos dados necessários ao desenvolvimento desta pesquisa.

Ao Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos por proporcionar a realização desta graduação.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo apoio financeiro e aos colaboradores que gentilmente forneceram avaliações sobre as comparações das imagens.

RESUMO

A mamografia de rastreamento desempenha um papel fundamental na detecção do câncer de mama, motivando o desenvolvimento de uma grande quantidade de pesquisas sobre as técnicas de detecção auxiliadas por computador, a fim de que se aumente a precisão do diagnóstico e se reduza o número de mamografias mal interpretadas. Contudo, a avaliação da eficácia dessas técnicas requer um banco de dados robusto de mamografias. Infelizmente, a obtenção de tais conjuntos de dados, que abrangem relatórios radiológicos e de biópsia, achados diversos e imagens de alta qualidade, costuma ser prejudicada por restrições impostas por centros de radiologia, hospitais ou equipes de pesquisa. Assim, para enfrentar estes desafios propõe-se, neste trabalho, uma nova abordagem: desenvolver um banco de dados de imagens mamográficas com base em imagens geradas por um simulador de mama desenvolvido anteriormente no grupo de pesquisa LAPIMO. Portanto, o objetivo principal deste trabalho é realizar uma análise comparativa entre imagens digitais derivadas desse modelo e de imagens reais de mama para avaliar o grau de confusão da percepção visual entre esses dois tipos de imagem. Essa investigação ajudará a determinar a viabilidade do uso de imagens de exposições desse *phantom* na criação de um banco de dados para avaliar o desempenho de técnicas usadas em esquemas de CAD em mamografia digital. Sendo assim, para realizar a investigação pretendida, foi desenvolvido um *software* para gerenciar vários testes comparativos envolvendo esses dois tipos de imagens obtidas de diferentes equipamentos de mamografia digital. Os resultados indicam que os grupos de observadores voluntários alcançaram taxas de precisão na faixa de 55% a 65%, em média, ao classificar as duas categorias (imagens obtidas por meio da exposição ao simulador e imagens reais da mama). Esses resultados sugerem um grau notável de confusão entre os dois tipos de imagens, especialmente quando as exposições de raios-X foram realizadas no mesmo equipamento. Logo, a avaliação anterior das propriedades físicas do simulador de mama, combinada com os resultados obtidos aqui, apontam para uma maneira promissora de incorporar imagens das exposições do simulador de mama na composição de um banco de dados de imagens para avaliar os esquemas de diagnóstico auxiliado por computador.

Palavras-chave: Imagens digitais de mamografia. *Phantom* mamográfico. Percepção visual em mamografia

ABSTRACT

The screening mammography plays a crucial role in breast cancer detection, motivating extensive research on computer-aided detection techniques to enhance diagnostic accuracy and reduce the incidence of misinterpreted mammograms. However, evaluating the effectiveness of these techniques necessitates a robust database of mammograms. Unfortunately, obtaining such datasets, encompassing radiological and biopsy reports, diverse findings, and high-quality images, is often hindered by constraints imposed by radiology centers, hospitals, or research teams. To address these challenges, this work proposes a novel approach: developing a mammographic image database based on images generated by a breast simulator previously developed in the LAPIMO research group. Thus, the primary objective of this study is to conduct a comparative analysis between digital images derived from this model and real breast images to assess the level of visual perception confusion between these two image types. This investigation will help determine the feasibility of using images from this phantom in creating a database to evaluate the performance of techniques used in computer-aided diagnosis (CAD) schemes in digital mammography. To conduct the intended investigation, a software was developed to manage various comparative tests involving these two types of images obtained from different digital mammography equipment. The results indicate that groups of volunteer observers achieved accuracy rates in the range of 55% to 65%, on average, when classifying the two categories (images obtained through exposure to the simulator and real breast images). These findings suggest a notable degree of confusion between the two image types, especially when X-ray exposures were performed on the same equipment. Therefore, the previous assessment of the physical properties of the breast simulator, combined with the results obtained here, points to a promising way to incorporate images from the simulator's exposures in the composition of a image database for evaluating computer-aided diagnostic schemes.

Keywords: Digital mammography images. Phantom mammography. Visual perception in mammography.

LISTA DE FIGURAS

Figura 1 – Distribuição da incidência de câncer entre mulheres no Brasil	19
Figura 2 – Estrutura do <i>phantom</i> antropomórfico	28
Figura 3 – Aparência radiográfica do <i>phantom</i> antropomórfico	29
Figura 4 – Estrutura do <i>phantom</i> mamográfico CIRS BR3D	29
Figura 5 – Exemplos de camadas do <i>phantom</i> de mama estruturado	30
Figura 6 – (a) Exemplos de algumas camadas do <i>phantom</i> e (b) dessas camadas sob compressão posicionadas no equipamento mamográfico no hospital para aquisição das imagens do <i>phantom</i>	33
Figura 7 – Exemplos de imagens usadas para o recorte das ROIs para compor o banco de dados do <i>software</i> : (a) mamografia de uma mama real, (b) uma imagem de uma configuração específica do <i>phantom</i> de mama. Em ambas as imagens, um quadrado é ilustrado mostrando as respectivas ROIs selecionadas.	34
Figura 8 – Exemplo de imagens para compor o banco de imagens 1 (baixo contraste): (a) imagem do <i>phantom</i> e (b) imagem de uma mama real.	35
Figura 9 – Exemplo de imagens para compor o banco de imagens 2 (alto contraste): (a) imagem do <i>phantom</i> e (b) imagem de uma mama real	36
Figura 10 – Interface gráfica do experimento 1	38
Figura 11 – Interface gráfica do experimento 2	38
Figura 12 – Interface gráfica do experimento 3	38
Figura 13 – Histogramas referentes as ROIs de imagens reais (direita) e do <i>phantom</i> (esquerda) exibidas no experimento 1.	42
Figura 14 – Histogramas referentes as ROIs de imagens reais (direita) e do <i>phantom</i> (esquerda) exibidas no experimento 2.	42
Figura 15 – Histogramas referentes as ROIs de imagens reais (direita) e do <i>phantom</i> (esquerda) exibidas no experimento 3.	43

LISTA DE TABELAS

Tabela 1 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para a primeira etapa do projeto	44
Tabela 2 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 1	45
Tabela 3 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 2	45
Tabela 4 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 1)	46
Tabela 5 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 2)	46
Tabela 6 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 1 completo	49
Tabela 7 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 1 completo)	50
Tabela 8 – Valores de distância de Bhattacharyya, correlação e interseção para cada banco de dados de imagens	51

LISTA DE ABREVIATURAS E SIGLAS

BI-RADS	<i>Breast Imaging Reporting and Data System</i>
CAD	<i>Computer-Aided Diagnosis</i> (Diagnóstico auxiliado por computador)
CADe	<i>Computer-Aided Detection</i> (Detecção auxiliada por computador)
CADx	<i>Computer-Aided Diagnosis</i> (Diagnóstico auxiliado por computador)
CC	Craniocaudal
CIRS	<i>Computerized Imaging Reference Systems</i>
DDSM	<i>Digital Database for Screening Mammography</i>
FN	Falso Negativo
FP	Falso Positivo
INCA	Instituto Nacional de Câncer
LAPIMO	Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas
MIAS	<i>Mammographic Image Analysis Society Digital Mammogram Database</i>
MLO	Médio-Lateral Oblíqua
OMS	Organização Mundial da Saúde
ROI	<i>Region of Interest</i> (Região de Interesse)
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Objetivos	23
1.1.1	Objetivos específicos	24
2	DESENVOLVIMENTO	25
2.1	Revisão Bibliográfica	25
2.1.1	Bancos de imagens mamográficas	25
2.1.1.1	MIAS (<i>Mammographic Image Analysis Society Digital Mammogram Database</i>)	26
2.1.1.2	DDSM (<i>Digital Database for Screening Mammography</i>)	26
2.1.2	<i>Phantoms</i> mamográficos	27
2.1.2.1	<i>Phantom</i> de mama antropomórfico Rachel	28
2.1.2.2	<i>Phantom</i> mamográfico CIRS BR3D	29
2.1.2.3	<i>Phantom</i> de mama estruturado (LAPIMO-EESC/USP)	30
2.1.3	Esquemas Computacionais de auxílio ao diagnóstico (CAD)	31
2.1.3.1	Avaliação de esquemas CAD	32
2.2	Materiais e Métodos	33
2.2.1	Aquisição de Imagens	33
2.2.2	Software de gerenciamento de testes comparativos	36
2.2.3	Análise quantitativa das características dos conjuntos de imagens	39
2.3	Resultados	41
2.3.1	Análise de robustez do software de gerenciamento de testes	41
2.3.2	Análise estatística dos resultados obtidos pelo software de gerenciamento de testes	43
2.4	Discussão	46
2.4.1	Análise estatística de percepção visual	46
2.4.2	Análise quantitativa das características das imagens nos bancos de imagens	50
3	CONCLUSÃO	53
3.1	Desenvolvimentos futuros	54
	REFERÊNCIAS	55

1 INTRODUÇÃO

Segundo dados estatísticos da Organização Mundial da Saúde (OMS), sabe-se que o câncer é a segunda principal causa de mortes no cenário global. No ano de 2018, estima-se que aproximadamente 9,6 milhões de mortes tenham sido causadas por esta doença. (WORLD HEALTH ORGANIZATION, CANCER, 2023)

O câncer mama é a forma desta doença mais prevalente entre a população feminina, sendo responsável por 25% dos casos de câncer diagnosticados globalmente por ano, se colocando como uma das principais causas de morte de mulheres tanto em países desenvolvidos como em desenvolvimento, caso do Brasil (BARCELOS BRUNO PEREIRA NUNES; FACCHINI, 2018). No país, segundo o Instituto Nacional de Câncer (INCA) a taxa de mortalidade desta doença foi de 11,71 óbitos a cada 100.000 mulheres no ano de 2021, representando 16,1% do total de óbitos causados por câncer (INCA, 2019), sendo válido ressaltar que na estimativa de novos casos de câncer para os próximos anos, o câncer de mama se coloca como a principal preocupação entre as mulheres, conforme apresentado na Figura 1.

Figura 1 – Distribuição da incidência de câncer entre mulheres no Brasil

Incidência de Câncer no Brasil



Localização primária	Casos	%
Mama feminina	73.610	30,1%
Côlon e Reto	23.660	9,7%
Colo do útero	17.010	7,0%
Traqueia, Brônquio e Pulmão	14.540	6,0%
Glândula Tireoide	14.160	5,8%
Estômago	8.140	3,3%
Corpo do útero	7.840	3,2%
Ovário	7.310	3,0%
Pâncreas	5.690	2,3%
Linfoma não Hodgkin	5.620	2,3%

Fonte: INCA, 2022

Fonte: (ONCOGUIA, 2024)

Dessa forma, tendo em vista a magnitude do impacto global associado a esta doença, tem-se que a principal forma de reversão deste cenário se encontra na detecção do câncer em seus estágios iniciais, pois o pilar para o controle do câncer de mama se baseia na prevenção, na detecção precoce e no tratamento. Estudos demonstram que a detecção precoce da doença é primordial para a obtenção de um prognóstico positivo e de um tratamento eficaz, pois observa-se que pacientes que possuem lesões, tumores menores

no momento do diagnóstico possuem uma taxa de sobrevivência significativamente maior, além de proporcionar a obtenção de tratamentos menos agressivos e invasivos (BHUSHAN; GONSALVES; MENON, 2021), de forma que a busca por ferramentas de imagens que possibilitem essa detecção precoce e a avaliação precisa de lesões é de suma importância.

O principal método utilizado para o rastreamento do câncer de mama é através da mamografia de rastreamento, sendo considerada como a forma mais eficaz na identificação precoce de alterações nas mamas, incluindo pequenas lesões ou aquelas que não são superficiais, que acabam passando por despercebidas no exame clínico, muitas vezes por não serem palpáveis, o que poderia resultar em detecções tardias. Portanto, a mamografia de rastreamento é uma ferramenta de imagem que comprovadamente contribui para a diminuição da taxa de mortalidade causada pelo câncer de mama.(INCA, 2019)

No Brasil, a recomendação do Sistema Único de Saúde (SUS) para a realização do rastreamento do câncer de mama é de que o exame mamográfico seja realizado em mulheres que se encontrem na faixa etária de 50 a 69 anos e com um intervalo máximo de dois anos, além da realização do exame clínico anual das mamas para mulheres em faixa etária inferior (INSTITUTO NACIONAL DE CANCER - INCA, 2023), já que o câncer de mama normalmente é encontrado em mulheres com mais de 35 anos e a probabilidade da doença aumenta de forma conjunta com o aumento da idade.

Outros exames são frequentemente utilizados como complementares ao exame mamográfico, com destaque para o ultrassom de mama, cujo principal objetivo é analisar alterações em mamas densas e áreas suspeitas não visualizadas na mamografia, a fim de que se reduza o risco de realização de biópsias desnecessárias, e tendo como principais vantagens seu baixo custo e a não exposição do paciente a radiação. Entretanto uso do exame de ultrassom, neste contexto, não deve ser utilizado como exame de rastreio, pois alguns fatores impõem limitações a esta ferramenta como a geração de imagens ruidosas que dificultam a análise do exame sem o auxílio da mamografia e por não detectar alguns sinais precoces de câncer. (BHUSHAN; GONSALVES; MENON, 2021).

Neste contexto, nota-se o papel de padrão-ouro exercido pela mamografia no que se refere a exames de imagens associados ao câncer de mama, sendo, portanto, crucial a garantia da acurácia final do resultado desses exames, pois a redução de taxas de biópsias desnecessárias é fundamental para a redução do estresse da paciente bem como os custos de tratamento.

Todavia, o diagnóstico de lesões em mamografias não é uma tarefa fácil, já que a imagem mamográfica não possui muito contraste, resultando em uma similaridade de intensidade entre a lesão e os tecidos do parênquima mamário que se encontram próximos a esta massa, o que torna a análise do exame mais subjetiva, trazendo uma maior variabilidade para o diagnóstico final do exame (MAHMOOD *et al.*, 2022). Além dessas dificuldades e da divergência de opiniões entre especialistas, pode-se ressaltar

alguns outros fatores que podem afetar a acurácia final do diagnóstico, como o fato do radiologista realizar um grande número de análises por dia de forma manual, tornando o processo mais demorado, complexo e influenciado pelas condições do ambiente de análise, o que resulta em uma maior chance do diagnóstico ser afetado por um estado de fadiga mental e visual do radiologista, de forma, que esse mesmo especialista possa apresentar diferentes interpretações sobre uma mesma imagem, se analisada em momentos diferentes. (MAHMOOD *et al.*, 2022) (BRADY, 2017)

Portanto, estudos demonstram que os principais erros que ocorrem em radiologia correspondem às categorias de erros de busca, de detecção e de interpretação, ou seja, erros que ocorrem quando o radiologista não reconhece uma lesão mesmo que esta esteja no campo de visão ou em casos em que a lesão é identificada, porém classificada de forma errônea, ressaltando que a acurácia do exame neste tipo de análise sofre com dificuldades impostas pelo limite da percepção humana. (CALAS; GUTFILEN; PEREIRA, 2012)

Estudos demonstram que devido a fatores como estes apresentados anteriormente, cerca de 10% a 30% de lesões mamárias correspondentes a casos de câncer não são identificadas como tais, e cerca de 80% das mulheres que são chamadas para a realização de exames complementares possuem resultados normais (EKPO; ALAKHRAS; BRENNAN, 2018).

Nesse sentido, a fim de que seja possível minimizar o número de erros de diagnósticos e melhorar a análise mamográfica, uma estratégia comumente empregada é o uso da dupla leitura do exame mamográfico por dois radiologistas, onde estudos demonstram que 7,7% casos adicionais de câncer foram detectados por meio da segunda leitura (GEORGIAN-SMITH *et al.*, 2007). Entretanto, em grande parte dos casos isto não é possível tanto por problemas associados à organização desse processo como falta de recursos financeiros de algumas instituições de saúde. (CALAS; GUTFILEN; PEREIRA, 2012)

Com o intuito de propor uma alternativa à limitação apresentada anteriormente para o aumento da acurácia dos exames mamográficos vem surgindo, nos últimos anos, o desenvolvimento de esquemas computacionais de auxílio à detecção/diagnóstico (CADe e CADx), assim como de muitas outras técnicas computacionais, incluindo aquelas no campo da inteligência artificial. Em relação às técnicas associadas ao campo de inteligência artificial pode-se citar o uso de um processo capaz de extrair um grande número de características de imagens digitais, disponibilizando-os em bancos de dados para que seja possível extrair esses dados para a geração de hipóteses ou realização de testes, auxiliando radiologistas (GILLIES; KINAHAN; HRICAK, 2016), porém o foco deste trabalho estará mais voltado a aplicação dos esquemas do tipo CADe e CADx.

Os esquemas do tipo CAD, segundo estudos, para a detecção de câncer de mama, tem um índice de detecção similar à utilização de um segundo radiologista como uma segunda opinião (GEORGIAN-SMITH *et al.*, 2007). E esses esquemas podem ser divididos

entre CADe (*Computer aided detection*) e CADx (*Computer aided diagnosis*), de forma que enquanto o primeiro auxilia o radiologista na detecção de lesões ou anomalias, destacando as áreas suspeitas ou anormais nas imagens, o segundo vai além da detecção e envolve a interpretação e a classificação dos achados, por exemplo avaliando formato, contorno e textura de nódulos.

Ou seja, os esquemas do tipo CAD representam técnicas que empregam o reconhecimento de padrões com o objetivo de identificar características suspeitas em uma imagem, destacando-as, marcando-as e classificando-as. Esses algoritmos são elaborados por meio do treinamento do sistema com um conjunto de imagens, envolvendo etapas como pré-processamento, segmentação, extração de características e classificação. (KATZEN; DODELZON, 2018)

Nesse sentido, se faz de suma importância a garantia de credibilidade e de um bom funcionamento destas ferramentas de auxílio ao diagnóstico, já que, uma vez utilizadas para a análise de imagens elas poderão influenciar a opinião final do médico no estabelecimento de um diagnóstico. Assim, para a garantia dos aspectos descritos anteriormente, a eficácia de todas as técnicas envolvidas em esquemas computacionais deve ser avaliada a partir de uma grande conjunto de imagens mamográficas que contenham uma variedade tanto casos normais como suspeitos (de preferência com sinais clínicos relevantes), que apresentem informações precisas sobre lesões e relatórios médicos.

No entanto, a dificuldade se encontra na disponibilização desses grandes conjuntos de dados, em uma quantidade estatisticamente significativa, principalmente devido a restrições de diferentes centros de radiologia e hospitais. Além disso, os bancos de dados mamográficos desenvolvidos por pesquisadores da área são comumente de uso restrito pela equipe envolvida nos testes (WARREN *et al.*, 2007) (MOREIRA *et al.*, 2012), e, em muitos casos não possuem a versatilidade necessária para testar técnicas genéricas de processamento, tanto em termos de pacientes/imagens características (casos documentados com a presença de lesões malignas, benignas ou de casos normais) e variedade de equipamentos. Na verdade, os poucos bancos de imagens mamográficas abertos, como o MIAS (SUCKLING, 1994) e o DDSM (HEATH *et al.*, 2000), por exemplo, possuem arquivos de mamografias e filmes digitalizados, muito distante do atual estado da arte dos equipamentos de mamografia digital direta.

Como alternativa, conjuntos de imagens baseados em exposições de objetos simuladores de mama (*phantoms*), especialmente e desenvolvidos para simular mamas reais em vários aspectos, assim como *phantoms* virtuais (SOUSA *et al.*, 2017) (BAKIC *et al.*, 2018), podem constituir um recurso útil e interessante para substituir conjunto de dados de imagens mamográficas de mamas reais (SCHIABEL; MATHEUS; CARDOSO, 2022). Vários motivos podem ser elencados como vantagens para a utilização de um *phantom* de mama: variedade de casos e equipamentos de imagem, variedade de lesões simuladas

e, principalmente disponibilidade de todas as informações sobre cada imagem, como a existência e localização de nódulos e microcalcificações, características de densidade e espessura, doses de exposição. Além disso, o uso de um banco de imagens baseado na exposição do *phantom* de mama exclui a necessidade de laudos médicos e/ou de biópsia para a obtenção dos dados necessários para a realização de avaliação dos esquemas computacionais. Como requisito principal, no entanto, a semelhança entre as imagens da mama real típica e do *phantom* nas regiões internas, que mostram onde as lesões simuladas são circundadas pela estrutura que simula o parênquima mamário (fundo da imagem) é muito importante para a garantia do sucesso do uso da imagem do *phantom* em testes de técnicas de processamento digital de esquemas CAD.

Um *phantom* de mama estruturado com distribuição aleatória de elementos de interesse foi desenvolvido anteriormente (SOUZA; MATHEUS; SCHIABEL, 2018) com a intenção de ser uma ferramenta capaz de estabelecer um banco de dados mamográficos adequado para ser usado para testar a eficácia das técnicas de processamento de imagens nesta área. A sua principal característica é a possibilidade de obtenção de imagens de estruturas simuladas distribuídas aleatoriamente sobre camadas que simulam o parênquima mamário também distribuído de forma aleatória, o que permite simular não apenas estruturas, mas também diferentes densidades. Dessa forma, uma das características mais importantes a serem avaliadas nessas imagens provenientes do *phantom* de mama para o desenvolvimento do banco de dados é o nível de similaridade com mamografias reais em termos de regiões internas de interesse.

À vista disso, este trabalho está voltado à investigação do nível de similaridade entre as imagens mamográficas de mamas reais e do *phantom* para a constituição de um novo banco de imagens, associado a avaliação de um esquema CADx previamente desenvolvido (MATHEUS, 2015).

1.1 Objetivos

Com base nas informações expostas, pode-se citar como o principal objetivo deste projeto, para que seja possível validar as técnicas desenvolvidas anteriormente para um esquema CADx (MATHEUS, 2015), a realização de comparações entre imagens digitais obtidas do *phantom* de mama e exposições de mamas reais em equipamento do tipo FFDM, de forma que seja possível investigar o nível de confusão entre esses dois tipos de imagens em termos de percepção visual, a fim de verificar se o *phantom* pode ser utilizado no desenvolvimento de um banco de dados versátil para testar o desempenho de técnicas de esquema CAD.

1.1.1 Objetivos específicos

- Elaborar e desenvolver um *software* de gerenciamento de testes, que realizará diferentes tipos de avaliação, com base na análise visual das imagens mamográficas pelo observador voluntário.
- Estabelecer uma sequência de experimentos, gerenciados pelo *software*, que permitam ao usuário comparar visualmente imagens simuladas com mamografias reais.
- Realizar análise estatística dos resultados obtidos pelo *software* desenvolvido e identificar a taxa de sucesso, que corresponde à similaridade das imagens do *phantom* com as reais.
- A partir dos resultados criar uma base de imagens para o desenvolvimento dos esquemas CAD em substituição às bases de imagens reais.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

O avanço constante da tecnologia na área de imagem médica, em particular na mamografia, tem desempenhado um papel crucial no diagnóstico precoce e no tratamento eficaz do câncer de mama, sobretudo através da inclusão de esquemas de detecção auxiliada por computador na rotina de médicos e radiologistas. No entanto, a eficácia dos sistemas de diagnóstico automatizado depende intrinsecamente da qualidade e diversidade dos bancos de imagens mamográficas utilizados para treinamento e validação.

Neste contexto, a presente revisão bibliográfica se propõe a explorar as características dos atuais bancos de imagens mamográficas mais utilizados na área, destacando desafios cruciais e suas limitações que afetam a generalização e desempenho desses sistemas. Além disso, também serão abordados o atual estado da arte dos principais *phantoms* mamográficos, estabelecendo-se uma investigação acerca da viabilidade do uso do mesmos como uma alternativa aos bancos de imagens mamográficas reais, bem como será abordado características acerca do funcionamento dos esquemas CAD, como ocorre a avaliação e validação dos mesmos.

2.1.1 Bancos de imagens mamográficas

Conforme descrito anteriormente, sabe-se que para o desenvolvimento, testes e aplicação de melhorias em esquemas computacionais de auxílio ao diagnóstico (CADe e CADx) os pesquisadores necessitam de uma grande quantidade de imagens mamográficas, que contenham alguns requisitos essenciais para o desenvolvimento das pesquisas. Como tais requisitos pode-se citar a variedade de casos, ou seja, a presença de casos normais e de diferentes tipos de achados, bem como de diferentes tipos de densidades mamárias, além de imagens normais com estruturas enganosas, que podem ser utilizadas a fim de tornar os esquemas computacionais mais robustos; a presença de laudos de confirmação de biópsia com localização e limites da lesão na imagem; informações adicionais, como aquelas referentes a densidade da mama e a classificação BI-RADS; e o acesso ao banco de imagens, que deveria ser de fácil disponibilidade.

Dessa forma, tendo em vista a crucialidade da existência e disponibilização de banco de imagens completos para o desenvolvimento de pesquisas na área de câncer de mama, a seguir serão apresentados alguns aspectos relativos aos dois bancos de imagens mais comumente utilizados por este ramo de pesquisa: o MIAS (*Mammographic Image Analysis Society Digital Mammogram Database*) e o DDSM (*Digital Database for Screening Mammography*).

2.1.1.1 MIAS (*Mammographic Image Analysis Society Digital Mammogram Database*)

O banco de imagens mamográficas MIAS (*Mammographic Image Analysis Society Digital Mammogram Database*) é conhecido por ser o primeiro banco de imagens mamográficas a fornecer um conjunto inicial de imagens mamográficas digitais, estabelecendo-se como uma referência na comunidade de pesquisa e sendo, ainda, amplamente utilizado.

Este banco de imagens possui 161 casos, que englobam um total de 322 imagens MLO digitalizadas, contendo todos os tipos de achados, ou seja, incluindo tantos casos benignos como malignos e casos normais, entretanto, neste banco de imagens é possível observar um número elevado de casos benignos quando comparado a casos malignos, dessa forma, o MIAS pode não capturar adequadamente a diversidade de casos encontrados na prática clínica diária. Essa limitação compromete a capacidade de desenvolver e avaliar algoritmos robustos que lidem com uma ampla gama de condições mamárias (SUCKLING, 1994).

Além disso, o MIAS também possui informações relativas a densidade de mama e suas anotações relativas às lesões presentes nas imagens consistem em um centro e um raio de um círculo ao redor da área de interesse, o que nem sempre é o suficiente para todos os tipos de estudos (SUCKLING, 1994). Ademais, também pode-se citar como uma outra limitação relativa ao MIAS é a resolução na qual as imagens foram digitalizadas, já que isto impede o uso do banco de forma adequada para pesquisas associadas à detecção de microcalcificações.

Entretanto, uma de suas principais limitações associada a sua aplicação nos dias atuais está associada a evolução da tecnologia de aquisição de imagens mamográficas desde a criação do banco de imagens, ou seja, as imagens digitalizadas estão muito distantes das características dos atuais equipamentos de mamografia digital direta, fazendo com que a sua utilização não seja a ideal, no cenário atual, para o desenvolvimento de novas técnicas de auxílio ao diagnóstico.

2.1.1.2 DDSM (*Digital Database for Screening Mammography*)

O banco de imagens DDSM (*Digital Database for Screening Mammography*) é um dos bancos de imagens mais utilizados e que aborda algumas das limitações do MIAS ao oferecer um conjunto de dados mais abrangente e representativo. Com um total de 10.480 imagens que incluem ambos os seios, em diversas projeções (MLO e CC), e uma variedade de casos benignos, malignos e normais, com informações sobre a densidade da mama e a classificação BI-RADS (HEATH *et al.*, 2000), aspectos que fazem do DDSM um banco de imagens que fornece uma visão mais holística das condições encontradas na prática clínica.

Entretanto, o DDSM não está isento de limitações, já que apesar de oferecer um conjunto de dados mais abrangentes e representativo se comparado ao MIAS e as

demarcações das lesões nas imagens incluírem o limite em nível de pixel dos achados, as imagens não são adequadas para estudos associados a validação de algoritmos de segmentação por não possuírem uma precisão boa o suficiente na demarcação das lesões (SAMPAT *et al.*, 2008). Além disso, da mesma forma que o MIAS, esse banco de imagem utiliza imagens digitalizadas, o que distancia o uso desta base de imagens das atuais características presentes em imagens provenientes de equipamentos de mamografia digital direta.

Ambos os bancos de imagens desempenharam um papel crucial no avanço da pesquisa em mamografia e no desenvolvimento de algoritmos de auxílio ao diagnóstico. No entanto, suas limitações ressaltam a necessidade de abordagens mais abrangentes na criação de conjuntos de dados representativos, de livre acesso à comunidade científica e que reflitam as atuais tecnologias utilizadas na área de mamografia. Portanto, a combinação de múltiplos bancos de imagens, juntamente com técnicas de aumento de dados, é fundamental para superar as lacunas existentes e melhorar a robustez dos modelos de esquemas computacionais de auxílio ao diagnóstico na área, ressaltando, neste sentido, conforme dito anteriormente, o uso de *phantoms* de mama como um recurso interessante na obtenção de bancos de imagens mais versáteis.

2.1.2 *Phantoms* mamográficos

Um *phantom* ou objeto simulador, no contexto de radioterapia ou de imagens médicas, pode ser definido como um objeto projetado para simular características específicas do corpo humano, sendo constituídos de materiais que permitam a simulação das propriedades físicas dos tecidos humanos, tais como materiais gelatinosos (parafina, ceras) e polímeros (resinas, acrílicos, polietileno). Além disso, dentre os principais fins para os quais se utiliza um *phantom* de mama pode-se citar o seu uso em testes e calibração de equipamentos de imagem, garantia de qualidade, pesquisa, treinamento e medições de doses, sendo que, na área de radiologia, os *phantoms* são classificados entre *phantoms* de calibração e *phantoms* de imagem (SILVA, 2012).

Os *phantoms* de calibração são objetos simuladores utilizados no controle de qualidade, a fim de permitir a avaliação sistemática da qualidade da imagem gerada pelos equipamentos de diagnóstico e da calibração dos sistemas de aquisição de imagens. Já o uso de *phantoms* de imagem está voltado à avaliação da qualidade da imagem, pois eles são projetados para simular condições específicas que possam ser encontradas em exames clínicos reais, reproduzindo anomalias, patologias e variações anatômicas. Portanto, tendo em vista o que são os *phantoms*, a seguir serão apresentados alguns tipos de *phantoms* mamográficos presentes na literatura bem como características e justificativas acerca da escolha do *phantom* mamográfico utilizado para o desenvolvimento deste projeto.

2.1.2.1 *Phantom* de mama antropomórfico Rachel

Um tipo de *phantom* mamográfico realista, comumente utilizado na área é o *phantom* de mama antropomórfico "Rachel" (CALDWELL; YAFFE, 1990), que é capaz de fornecer imagens com detalhes e com contraste correspondente a uma mamografia real, sendo utilizado para avaliação dos efeitos de exposição do mesmo a diferentes equipamentos e a parâmetros de exposição variados.

Este *phantom* é constituído por três partes, conforme apresentado nas Figuras 2 e 3. O material de base BR-12 (CALDWELL; YAFFE, 1990) foi moldado para proporcionar variações de espessuras para a obtenção de diferenças na densidade óptica radiográfica, formando um tecido equivalente à mama. A camada de filme utilizada é responsável por produzir uma estrutura fina e detalhada da imagem, já que este filme corresponde a uma mamografia real, que foi processada digitalmente, e cujo filme de raios-X foi intensificado por mercúrio. Por fim, a caixa externa do *phantom* se encaixa entre a placa de compressão e a mesa do equipamento mamográfico, auxiliando, dessa forma, na obtenção de um posicionamento consistente com uma mamografia de mama real (CALDWELL; YAFFE, 1990).

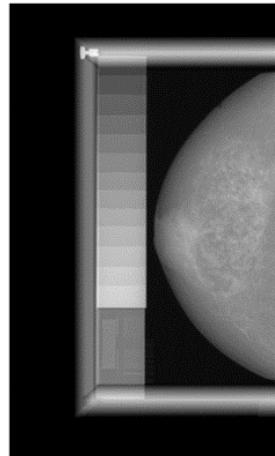
Figura 2 – Estrutura do *phantom* antropomórfico



Fonte: (CALDWELL; YAFFE, 1990)

Entretanto, apesar desse *phantom* antropomórfico ser um objeto simulador de mama realista, ele não se mostra como adequado para a utilização neste projeto, pois por conta de sua estrutura fixa não é possível a inserção de nenhum tipo de volume em seu interior para que se simule qualquer tipo de lesão, o que é essencial quando se trata da obtenção de um banco de imagens para o desenvolvimento de esquemas computacionais de auxílio ao diagnóstico.

Figura 3 – Aparência radiográfica do *phantom* antropomórfico



Fonte: (CALDWELL; YAFFE, 1990)

2.1.2.2 *Phantom* mamográfico CIRS BR3D

Outro tipo de *phantom* utilizado na área é o CIRS BR3D, que se destaca por sua capacidade de simular projeções tridimensionais da mama, representando de forma precisa e detalhada as características anatômicas do tecido mamário.

Esse *phantom* é constituído por seis placas heterogêneas, conforme apresentado na Figura 4, onde cada placa é composta por dois materiais que buscam imitar 100% dos tecidos adiposos e glandulares da mama, assim como cada placa possui um padrão único elas podem ser rearranjadas para gerar diferentes fundos de imagem (COMPUTERIZED IMAGING REFERENCE SYSTEMS, 2023). Contudo, apenas uma das seis placas contém alguns tipos de massas e microcalcificações em diferentes profundidades (COMPUTERIZED IMAGING REFERENCE SYSTEMS, 2023), o que não permite a representação de uma quantidade suficiente de lesões e de seu posicionamento interno na mama.

Figura 4 – Estrutura do *phantom* mamográfico CIRS BR3D



Fonte: (COMPUTERIZED IMAGING REFERENCE SYSTEMS, 2023)

Portanto, essa limitação é especialmente pertinente ao construir um conjunto de

dados de casos suspeitos, pois trata-se de um componente vital para avaliar os esquemas CADe/CADx, conforme mencionado anteriormente.

2.1.2.3 *Phantom* de mama estruturado (LAPIMO-EESC/USP)

Outro tipo de *phantom* a ser abordado consiste em um *phantom* de mama estruturado para aplicações de mamografia 2D, desenvolvido no Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO) da Escola de Engenharia de São Carlos, Universidade de São Paulo (EESC-USP).

O *phantom* foi desenvolvido de forma que seu *design* permitisse mudanças de suas características internas de diferentes formas. Assim, para representar a variedade de densidades de mamas observadas em casos clínicos o *phantom* foi desenvolvido de forma que este tivesse diferentes camadas, sendo 10 camadas compostas de gel de parafina com filme de PVC submerso em uma distribuição não uniforme, e outras 4 compostas apenas por parafina em gel, de forma que a distribuição das camadas principais simula regiões mais ou menos densas dependendo da concentração do material (SOUSA; MATHEUS; SCHIABEL, 2018).

Para a simulação dos nódulos foram utilizados dois modelos tridimensionais impressos, que representassem tanto lesões circunscritas como espiculadas, além disso os nódulos também podem ser inseridos na imagem obtida pela exposição do *phantom* de mama aos raios-X computacionalmente, de forma que a intensidade da lesão possa de ser ajustada de uma maneira controlada. E para a simulação de microcalcificações utilizou-se partículas de hidroxiapatita que foram agrupadas de maneira a simular *clusters* (SOUSA; MATHEUS; SCHIABEL, 2018). Logo, todas as estruturas simuladas podem ser fisicamente inseridas entre as camadas do *phantom* antes da sua exposição ao equipamento mamográfico. A Figura 5 ilustra algumas camadas do *phantom* mamográfico desenvolvido.

Figura 5 – Exemplos de camadas do *phantom* de mama estruturado



Fonte: (SOUSA; MATHEUS; SCHIABEL, 2018)

A partir da descrição apresentada para este *phantom* percebe-se que o mesmo

possui a versatilidade essencial para o desenvolvimento do banco de imagens para o teste e desenvolvimento de técnicas associadas a esquemas computacionais de auxílio ao diagnóstico, já que permite que as camadas do *phantom* sejam arranjadas de inúmeras formas, possibilitando a obtenção de diferentes densidades nas imagens, e que as lesões também sejam alocadas de formas diversas, o que é capaz de gerar uma variedade de casos que se aproxima muito da variabilidade clínica, fazendo com que este seja, portanto, o *phantom* de mama escolhido para o desenvolvimento do presente projeto.

2.1.3 Esquemas Computacionais de auxílio ao diagnóstico (CAD)

O *Computer-Aided Diagnosis* (CAD), definido como sistemas que auxiliam médicos e radiologistas na interpretação de imagens, começou a ser investigado por volta de 1960, que corresponde a época do surgimento de pesquisas voltadas ao uso de análises computadorizadas em radiografia. Entretanto, apenas em 1998 o diagnóstico auxiliado por computador teve o seu uso clínico aprovado na área da mamografia (MASUD *et al.*, 2019).

Resumidamente, os esquemas computacionais de auxílio ao diagnóstico funcionam da seguinte forma: a entrada consiste em uma imagem digital, obtida por equipamentos mamográficos digitais ou pela digitalização de um filme. A partir dessa imagem é feito um pré processamento, em que são removidos o fundo da imagem, além do músculo peitoral e de anotações presentes. A próxima etapa é a detecção da lesão na imagem, quando são empregadas, por exemplo, técnicas de detecção baseada em borda e em *deep learning* (HASSAN; HAMAD; MAHAR, 2022).

Após a etapa de detecção ocorre a segmentação da lesão detectada, a qual pode utilizar um limiar de nível de cinza local ou global (HASSAN; HAMAD; MAHAR, 2022), por exemplo, já que, geralmente, uma lesão correspondente a um câncer é mais brilhante que o tecido circundante ou através do uso de uma rede neural convolucional. A realização desta etapa demarca o limite entre os esquemas CADe e CADx, já que o primeiro busca apenas disponibilizar a localização e detecção de massas e anormalidades presentes na imagem, fazendo com que a interpretação dos achados sejam feitas apenas pelos médicos e radiologistas, enquanto o CADx vai além e também disponibiliza uma classificação dos achados, funcionando como uma segunda opinião.

Logo, esquemas computacionais mais completos ainda incluem as etapas de extração de características e classificações, na qual incluem a obtenção de características como textura, BI-RADS e características morfológicas. Na classificação utilizam-se técnicas convencionais de aprendizado de máquina e de aprendizado profundo (*deep learning*). E por fim, o sistema retorna as imagens com a demarcação das lesões e classificação das mesmas como suspeitas ou não.

2.1.3.1 Avaliação de esquemas CAD

Para a avaliação de esquemas de auxílio ao diagnóstico existem, basicamente, três fatores que são tido como primordiais para a verificação de performance do mesmo, sendo eles a sensibilidade, a especificidade e a natureza da base de dados (VYBORNY; GIGER; NISHIKAWA, 2000).

A sensibilidade (S) pode ser definida como a porcentagem de verdadeiros-positivos (VP) que foram classificados corretamente como genuínos, ou seja, são detecções positivas corretamente classificadas. Essa taxa é calculada pela razão entre a quantidade de verdadeiros-positivos e a soma dos casos de falsos-negativos (FN) e de verdadeiros positivos, que correspondem ao total de elementos genuínos (TG).

$$S = \frac{VP}{TG} = \frac{VP}{FN + VP}$$

A especificidade (E) é definida como sendo a proporção de verdadeiros negativos (VN) que foram classificados corretamente como impostores, ou seja, são detecções negativas corretamente classificadas. Essa taxa é calculada pela razão entre a quantidade de verdadeiros-negativos e a soma a soma dos casos de falsos-positivos (FP) e de verdadeiros-negativos, que correspondem ao total de elementos impostores (TI).

$$E = \frac{VN}{TI} = \frac{VN}{VN + FP}$$

E a natureza dos bancos de dados se refere ao conjunto do banco de imagens que foi utilizado para a avaliação do esquema computacional, o que inclui tanto as características físicas das imagens, com o nível de contraste, resolução espacial e tamanho das imagens, por exemplo, como a disponibilidade das informações presentes nas imagens, tais como a densidade da mama, os tipos de achados presentes nas imagens, informações de biópsia e laudos médicos, o que conforme já foi citado anteriormente é difícil de ser encontrado com acesso livre a todos pesquisadores da área. Logo, a natureza dos bancos de dados é um fator crucial para a avaliação e validação dos esquemas CAD, já que pequenas diferenças nos fatores citados podem impactar fortemente o resultado final.

Portanto, mais uma vez, nota-se a importância do acesso a um banco de imagens mamográficas versátil e com uma grande variedade de casos e equipamentos, destacando-se a relevância do impacto que o desenvolvimento de um banco de imagens baseados em exposições de um *phantom* de mama pode proporcionar para um melhor desenvolvimento dos esquemas computacionais de auxílio ao diagnóstico, justificando o principal escopo a ser abordado neste projeto.

2.2 Materiais e Métodos

Nesta seção, serão detalhadas as principais etapas realizadas no decorrer deste projeto. Assim, serão apresentados aspectos relativos ao processo de aquisição das imagens do *phantom* de mama, bem como, das imagens reais para compor os dois bancos de imagens a serem utilizados na primeira etapa deste projeto, além de aspectos relativos ao desenvolvimento do *software* de gerenciamento de testes comparativos entre esses dois conjuntos de imagens, como informações a respeito de seu funcionamento e das propostas dos experimentos a serem executados, permitindo, então uma maior compreensão acerca da pretendida análise de investigação do nível de confusão, em termos de percepção visual, entre os dois conjuntos de imagens citados.

2.2.1 Aquisição de Imagens

O *phantom* de mama utilizado nesta investigação corresponde ao *phantom* de mama estruturado desenvolvido no Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO), cujas características foram descritas anteriormente. Dessa forma, para a obtenção das imagens, foram fisicamente inseridas as estruturas correspondentes a simulação dos nódulos redondos e espiculados entre as camadas do *phantom* mamográfico em diferentes configurações, a fim de que fosse possível a simulação de diferentes densidades de mamas e casos. A Figura 6 ilustra algumas das camadas do *phantom* assim como a estrutura fixa na placa de compressão para a exposição no equipamento mamográfico e a Figura 7 ilustra exemplos de imagens utilizadas para o recorte das regiões de interesse (ROIs).

Figura 6 – (a) Exemplos de algumas camadas do *phantom* e (b) dessas camadas sob compressão posicionadas no equipamento mamográfico no hospital para aquisição das imagens do *phantom*.



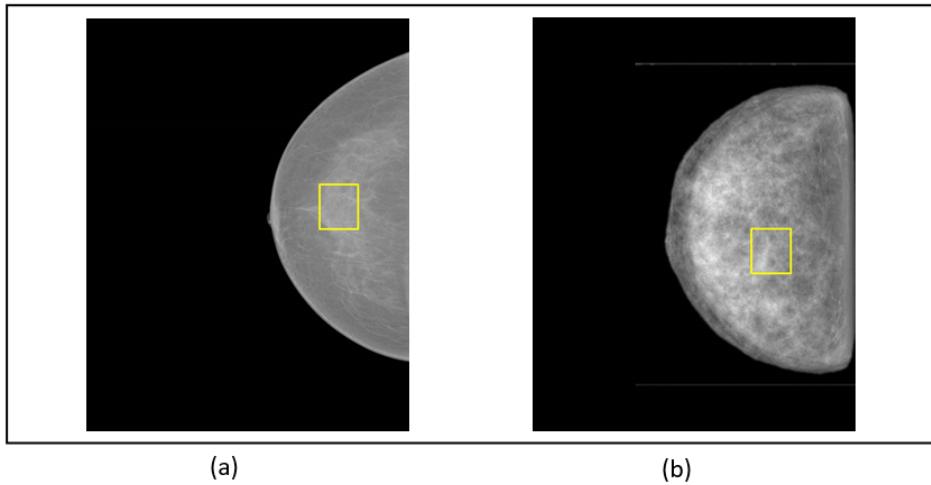
(a)



(b)

Fonte: Elaborada pelo autor

Figura 7 – Exemplos de imagens usadas para o recorte das ROIs para compor o banco de dados do *software*: (a) mamografia de uma mama real, (b) uma imagem de uma configuração específica do *phantom* de mama. Em ambas as imagens, um quadrado é ilustrado mostrando as respectivas ROIs selecionadas.



Fonte: Elaborada pelo autor

O *software* foi desenvolvido para permitir a realização de três experimentos, os quais serão descritos em detalhes na próxima seção, de análise visual de um conjunto de imagens mamográficas digitais por um grupo de observadores colaboradores. O principal objetivo é de se realizar uma comparação das regiões de interesse (ROIs) dos dois tipos de imagens (de mamas reais e do *phantom*) para identificar o quanto as imagens simuladas se aproximam visualmente das reais, uma vez que a maioria das técnicas de processamento envolvidas em esquemas CAD atuam sobre ROIs internas à estrutura mamária em busca de sinais de interesse clínico. Assim, a partir das imagens mamográficas completas, foram selecionadas ROIs que caracterizam áreas internas da mama (real ou simulada) em que o fundo estava essencialmente presente, caracterizando a estrutura do parênquima mamário. Essas ROIs foram, então, estabelecidas como o conjunto de imagens efetivas a serem exibidas pelo *software* durante os experimentos.

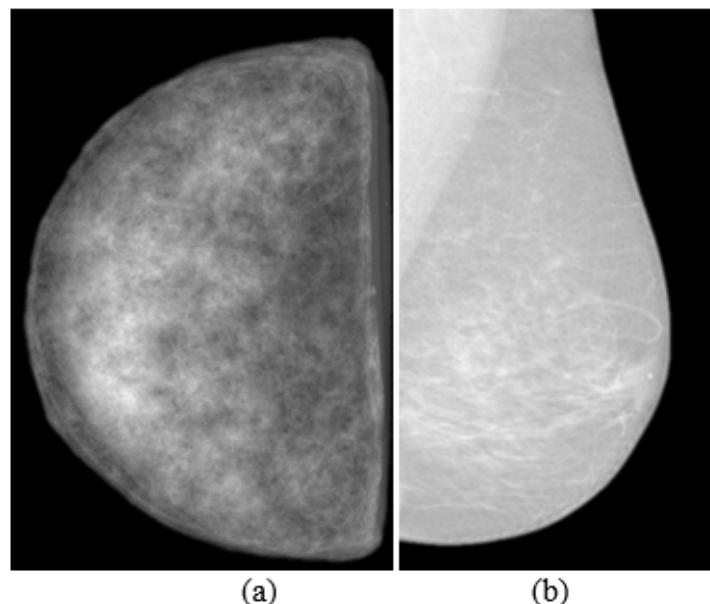
Dessa forma, no *software* desenvolvido para gerenciar as avaliações comparativas, foram criados dois bancos de imagens para a realização dos testes. O processo de seleção de imagens para esses bancos de imagens seguiu a seguinte abordagem: para cada banco de imagens, um conjunto de imagens incluiu regiões de interesse (ROIs) extraídas de mamografias reais, enquanto outro conjunto foi formado pela seleção de ROIs de imagens geradas utilizando várias configurações do *phantom* exposto aos raios-X em equipamentos de mamografia digital de campo total (FFDM - *Full Field Digital Mammography*).

As imagens do *phantom* utilizadas neste estudo foram adquiridas através de dois equipamentos FFDM diferentes (de fabricantes diferentes) instalados em hospitais públicos do estado de São Paulo. Esses sistemas de mamografia digital diferem em termos de material de ânodo (molibdênio/ródio ou tungstênio), material do filtro (molibdênio ou ródio) e

sistema de aquisição direta de imagens digitais. Todas as exposições foram realizadas usando a configuração de controle automático de exposição do respectivo sistema mamográfico (autocontraste ou equivalente). Em seguida, uma variedade de imagens mamográficas de mamas reais, obtidas anteriormente usando o mesmo equipamento e condições de exposição semelhantes (autocontraste ou equivalente) foi selecionada para o estudo.

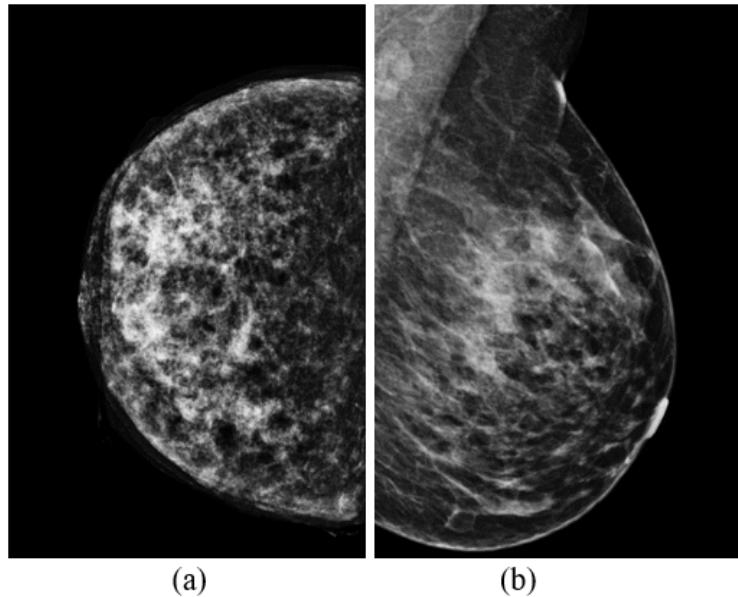
A seleção para compor os dois bancos de imagens foi baseada nas semelhanças visuais entre eles, principalmente em termos de suas características básicas de contraste conforme apresentado nas Figuras 8 e 9. Nessas figuras podem ser observados dois tipos de contraste nas imagens, que são característicos do processo de aquisição de cada equipamento utilizado. Assim, o primeiro banco de imagens para esses testes comparativos foi composto por imagens de menor contraste, adquiridas em equipamentos da *GE Senograph Essential*, conforme apresentado na Figura 8, enquanto o segundo banco de imagens foi composto por imagens de maior contraste, adquiridas por equipamentos da *Hologic Dimensions* como apresentado na Figura 9.

Figura 8 – Exemplo de imagens para compor o banco de imagens 1 (baixo contraste): (a) imagem do *phantom* e (b) imagem de uma mama real.



Fonte: Elaborada pelo autor

Figura 9 – Exemplo de imagens para compor o banco de imagens 2 (alto contraste): (a) imagem do *phantom* e (b) imagem de uma mama real



Fonte: Elaborada pelo autor

2.2.2 *Software* de gerenciamento de testes comparativos

O objetivo dos experimentos realizados pelo *software* é avaliar o quanto o observador do teste foi (ou não) capaz de identificar visualmente a diferença entre as imagens reais e simuladas, apontando se a ROI exibida se refere a um ou outro tipo de estrutura. Assim, foram elaborados três tipos de experimentos a serem executados a partir de um esquema computacional desenvolvido com base na linguagem de programação *Python* e em suas ferramentas de criação de interface gráficas (*Graphical User Interface (GUI)*) como o *PyQt5* (FOUNDATION, 2023) .

O primeiro experimento consiste em exibir aleatoriamente um conjunto de imagens correspondentes a ROIs selecionadas de mamografias digitais reais e de exposições do *phantom* ao mesmo sistema digital de mamografia. Nesse processo, apenas uma imagem é apresentada por vez ao observador, que é solicitado a indicar a sua classificação visual marcando uma caixa no menu do *software* para informar se considera aquela ROI como sendo de uma mamografia real ou não. O segundo experimento é realizado como um estudo de escolha forçada de duas alternativas, no qual, novamente, um conjunto aleatório de imagens é exibido, mas com duas ROIs exibidas simultaneamente lado a lado (sendo uma de uma mamografia real e outra do objeto simulador de mama). Nessa etapa, o observador deve determinar qual delas seria a ROI da mamografia real, o que significa que o *software* considera automaticamente a ROI não escolhida como simulada. O terceiro experimento exibe na tela um conjunto de 9 ROIs misturadas aleatoriamente (entre imagens reais e simuladas) de cada vez. Em seguida, o observador deve marcar no menu do *software* as caixas correspondentes às ROIs consideradas a partir de imagens reais. Agora,

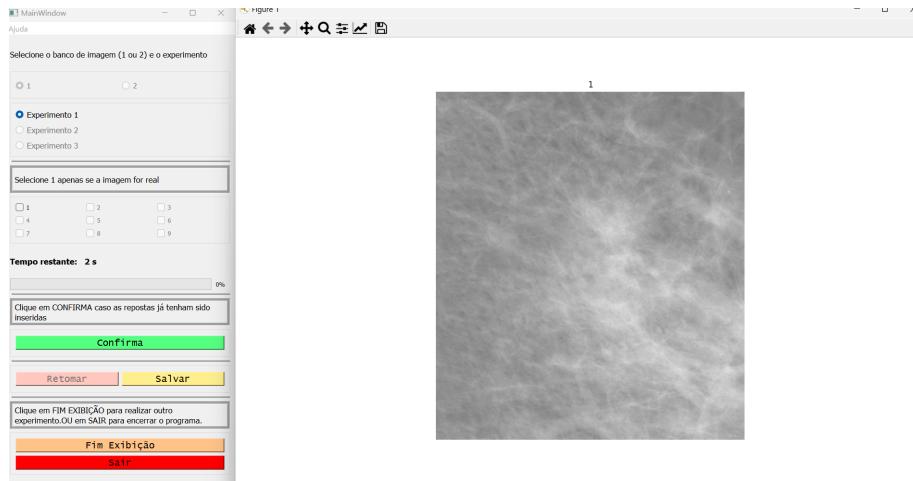
diferentemente do experimento anterior, o número de ROIs exibidas de cada categoria é aleatório, aumentando a variabilidade e a dificuldade do participante em determinar as escolhas certas.

O *software* foi projetado de forma que um determinado número de telas seja exibido para cada experimento. No primeiro, são exibidas 15 telas - ou seja, 15 imagens - para que o participante indique se corresponde a uma mama real ou ao *phantom*. No segundo, há mais 15 telas com pares de imagens, sempre uma de uma mama real e outra do *phantom* para que o observador indique novamente no menu qual é a real em cada tela. Além disso, o terceiro experimento apresenta um total de 6 telas com 9 ROIs em cada uma, e para cada tela exibida, as ROIs consideradas de imagens reais de mamas devem ser indicadas nas caixas apropriadas.

Assim, é importante ressaltar que o banco de dados de imagens interno do *software* armazenou um total de 397 ROIs das quais 218 pertencem ao banco de imagens 1 (onde há 119 imagens selecionadas de mamografias reais e 99 de imagens do *phantom*) e 179 pertencem ao banco de imagens 2 (onde há 83 imagens selecionadas de mamografias reais e 96 de imagens do *phantom*), sendo que todas as imagens foram adquiridas nos sistemas de mamografia já mencionados. Além disso, o esquema computacional extrai automaticamente e de forma aleatória as imagens desse banco de dados interno para exibi-las em cada experimento.

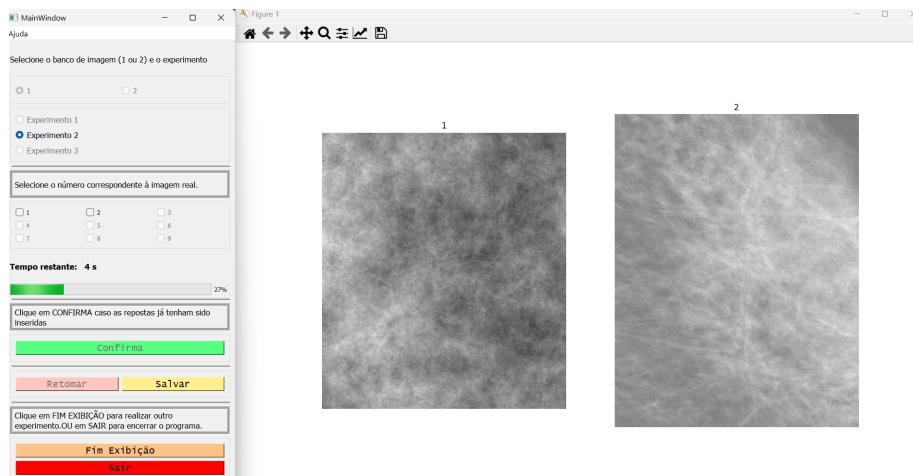
Outra característica a ser destacada em relação aos experimentos é que foi estabelecido um tempo máximo para que o observador visualizasse as respectivas telas de imagem e fizesse a indicação conforme descrito anteriormente. Assim, para o primeiro experimento, cada tela (de uma imagem) é mostrada por 5 s, após os quais uma nova tela é exibida. Esse tempo máximo para o caso dos experimentos 2 e 3, em que uma maior quantidade de imagens são exibidas em cada tela (um par de imagens para o experimento 2 e 9 imagens para o experimento 3), são de 7 s e 25 s, respectivamente. Esse "cronômetro" permite acelerar a execução do conjunto de testes, bem como diferenciar a precisão das respostas de acordo com a maior ou menor experiência do participante nessa análise. O esquema desenvolvido tem uma interface gráfica que exibe os comandos para selecionar o experimento e inserir as respostas marcando a caixa correspondente a cada ROI. Os comandos no menu também permitem salvar o andamento no experimento para retomada posterior, além de uma interface secundária utilizada para a exibição das imagens. As figuras 10, 11 e 12 apresentam a interface do programa para cada um dos experimentos descritos.

Figura 10 – Interface gráfica do experimento 1



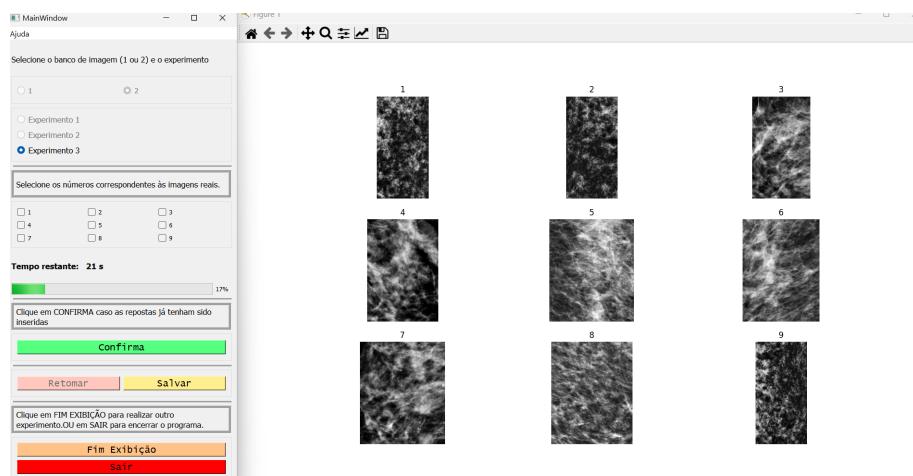
Fonte: Elaborada pelo autor

Figura 11 – Interface gráfica do experimento 2



Fonte: Elaborada pelo autor

Figura 12 – Interface gráfica do experimento 3



Fonte: Elaborada pelo autor

A seleção das ROIs para compor o banco de dados foi realizada com base na busca de características análogas internas das imagens, ou seja, descartando regiões com evidências flagrantes de sinais clássicos que evidenciassem alguma estrutura anatômica, como ossos, vértebras, microcalcificações, mamilos ou bordas da pele. Na verdade, o objetivo era comparar o plano de fundo em relação às imagens da mama real e do simulador de mama. Além disso, o principal objetivo de usar esse *phantom* é gerar um número muito grande de imagens de diferentes características (em termos de densidade, espessura, fundo, etc.), bem como com diferentes tipos de estruturas internas simuladas (como microcalcificações e nódulos) localizadas aleatoriamente.

Depois de devidamente validado, este banco de dados de mamografia será aplicado na próxima etapa do projeto, além de constituir como uma ferramenta fundamental para testar o desempenho das técnicas de processamento digital nesse campo. Portanto, para atingir esse objetivo, o presente estudo procura avaliar o nível de confusão entre o *phantom* e as imagens reais da mama de acordo com sua percepção visual, o que pode indicar até que ponto essa imagem do *phantom* possibilitará o desenvolvimento desse banco de dados.

O desenvolvimento do *software* de gerenciamento de testes foi projetado para permitir que o observador realize a avaliação comparativa por conta própria sem precisar de ajuda externa pra realizar os experimentos propostos, já que eles são autoexplicativos e não consomem muito tempo do usuário, pois leva apenas cerca de 5 minutos para realizar os três experimentos. Além disso, quando o observador marca uma caixa com a opção do tipo de imagem que está sendo avaliada, o programa regista automaticamente a resposta quando o cronômetro chega a zero. Os arquivos de texto são gerados ao final de cada experimento concluído com as informações fornecidas pelo observador, de modo que esses são os arquivos a serem avaliados para o levantamento estatístico das respostas do observador.

2.2.3 Análise quantitativa das características dos conjuntos de imagens

Para complementar a análise baseada na percepção visual realizada pelo *software* proposto, também foram estabelecidas avaliações quantitativas, focando diretamente nas características das imagens em ambos os bancos de dados. Isso envolveu a obtenção de métricas de similaridade entre os conjuntos de imagens. Ao selecionar essas métricas, procurou-se escolher as mais adequadas aos objetivos dos experimentos realizados com o *software* desenvolvido, ou seja, comparar ROIs selecionadas de mamas reais e do *phantom* de mama. Assim, foram escolhidas métricas que permitissem uma análise global dos dois conjuntos de dados, o que é possível por meio do uso de métodos baseados em histogramas. Os histogramas permitem a inferência de informações importantes sobre brilho e contraste, servindo como ferramentas amplamente aplicadas na classificação e no agrupamento de padrões no campo do processamento de imagens e da visão computacional.

(CHA; SRIHARI, 2002)

Portanto, foram escolhidos métodos para estabelecer diferentes comparações entre os histogramas de ROIs das mamografias reais e os das imagens do *phantom* de mama para cada banco de imagem individualmente: (1) distância de *Bhattacharyya*(KASHYAP, 2019), (2) correlação de *Pearson*(MUKAKA, 2012) e (3) interseção(LEE; XIN; WESTLAND, 2005). A primeira métrica, também conhecida como *B-distance*, pode ser definida como uma medida de similaridade usada para comparar e quantificar a semelhança entre duas distribuições, ou entre duas imagens ou conjuntos de imagens, fornecendo um valor numérico no intervalo de 0 a 1, que indica as dissimilaridades entre as distribuições de probabilidade dos valores de pixel(KASHYAP, 2019). Nesse caso, o valor 0 representa o cenário ideal em que não há diferenças entre as distribuições. Essa métrica é descrita pelas Equações (2.1) e (2.2):

$$D_{BC}(p_i, p'_i) = -\ln[\rho(p_i, p'_i)] \quad (2.1)$$

$$\rho(p_i, p'_i) = \sum_i^k \sqrt{p_i p'_i} \quad (2.2)$$

Onde $D_{BC}(p_i, p'_i)$ é a distância de *Bhattacharyya* e p_i e p'_i correspondem às distribuições de probabilidades. (KASHYAP, 2019)

A segunda métrica utilizada, descrita pela equação (2.3), permite obter uma análise diferente ao quantificar uma tendência linear da relação entre os valores de intensidade de pixel entre dois conjuntos de imagens, de modo que seu valor numérico seja restrito em um intervalo de -1 a 1 - sendo o valor 1 um indicativo de uma forte correlação positiva, o valor 0 indicando nenhuma correlação e o valor -1 indicando uma forte correlação negativa, mostrando uma relação linear inversa(MUKAKA, 2012).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (2.3)$$

Onde r corresponde ao coeficiente de correlação e x_i e y_i são os valores das variáveis investigadas.

Por fim, como uma métrica complementar às outras, a interseção, descrita na equação (4), permite a avaliação da sobreposição ou semelhanças entre as distribuições de dois conjuntos de imagens. Um valor de interseção mais alto indica uma maior semelhança entre as imagens comparadas (LEE; XIN; WESTLAND, 2005).

$$H(I) \cap H(I') = \sum_{j=1}^n \min(H_j(I), H_j(I')) \quad (2.4)$$

Onde $H(I)$ e $H(I')$ são os histogramas das imagens.

2.3 Resultados

Nas subseções a seguir serão apresentados os resultados que destacam o escopo abrangente do presente projeto, onde são analisadas características técnicas que verificam se o funcionamento do *software* cumpre com a sua funcionalidade principal de execução da avaliação comparativa proposta, sendo que para isso foi feita uma investigação da robustez do mesmo no que se refere a exibição das imagens de forma completamente aleatória, o que é essencial para a garantia de que o usuário colaborador não estabeleça nenhum tipo de memória ao longo da execução dos experimentos.

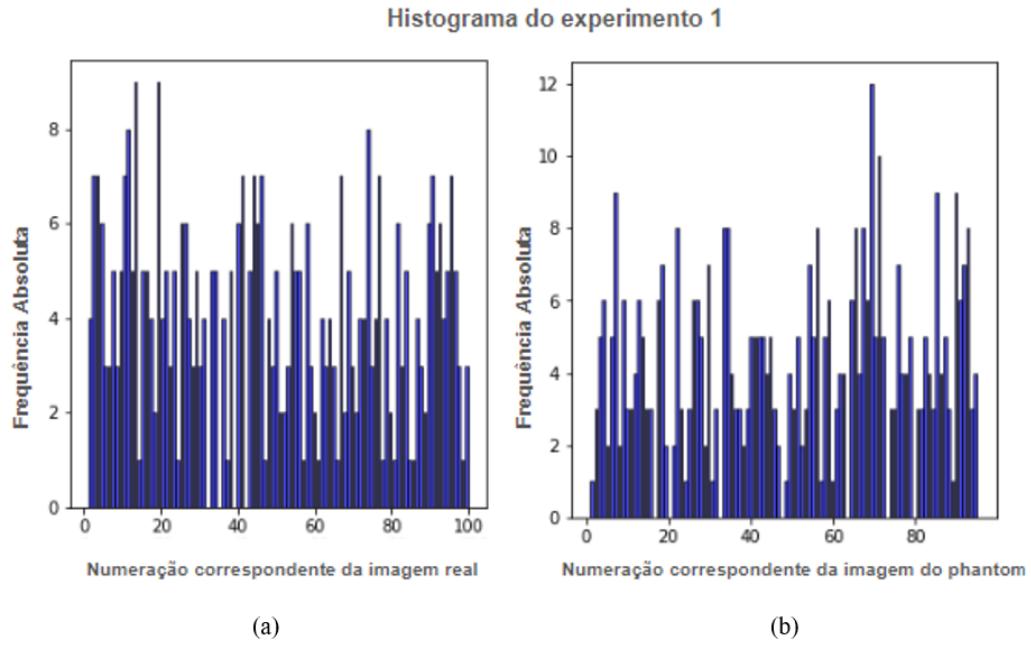
Além disso, também serão apresentadas análises estatísticas relativas aos resultados obtidos pelo uso do *software*, o que é essencial para a compreensão da viabilidade do uso de um banco de imagens baseado na exposição do *phantom* de mama para o desenvolvimento de técnicas de esquemas CAD.

2.3.1 Análise de robustez do *software* de gerenciamento de testes

Uma análise relativa à robustez do *software* de gerenciamento de testes desenvolvido foi realizada em relação à exibição de ROIs em cada sequência de experimentos. O estudo se preocupou em verificar se o procedimento de exibição de imagens reais e do *phantom* garante: (a) o uso de todas as imagens correspondentes às ROIs armazenadas no banco de dados interno e (b) de forma aleatória. Ou seja, se o processo interno do programa garantiu que a mesma imagem não fosse exibida excessivamente ao longo das avaliações feitas pelos observadores voluntários - ao contrário de outras imagens que nunca foram exibidas. Isso é importante para garantir a real aleatoriedade da exibição dos casos armazenados, evitando, por exemplo, que o observador reconheça algumas ROIs já classificadas em um experimento anterior ao repetir novamente a sequência de experimentos.

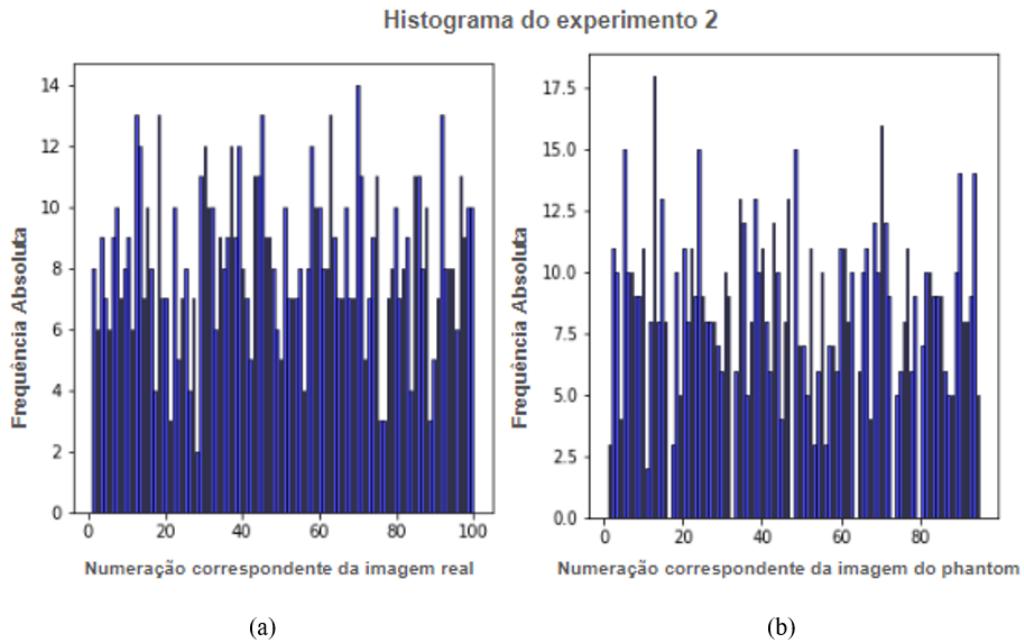
Assim, para verificar a aleatoriedade esperada, foram determinados histogramas do procedimento para os testes realizados. Dois histogramas foram plotados para cada um dos três experimentos: um deles mostra a frequência de exibição das ROIs correspondentes às mamografias reais, enquanto o outro mostra a mesma frequência, porém para ROIs obtidas das imagens do *phantom*. Além disso, é importante ressaltar que considerou-se a simulação de 60 avaliações para cada um dos três experimentos realizados pelos observadores voluntários para determinar a distribuição das imagens exibidas em todas as avaliações. As Figura 13, 14 e 15 ilustram esses histogramas.

Figura 13 – Histogramas referentes as ROIs de imagens reais (direita) e do *phantom* (esquerda) exibidas no experimento 1.



Fonte: Elaborada pelo autor

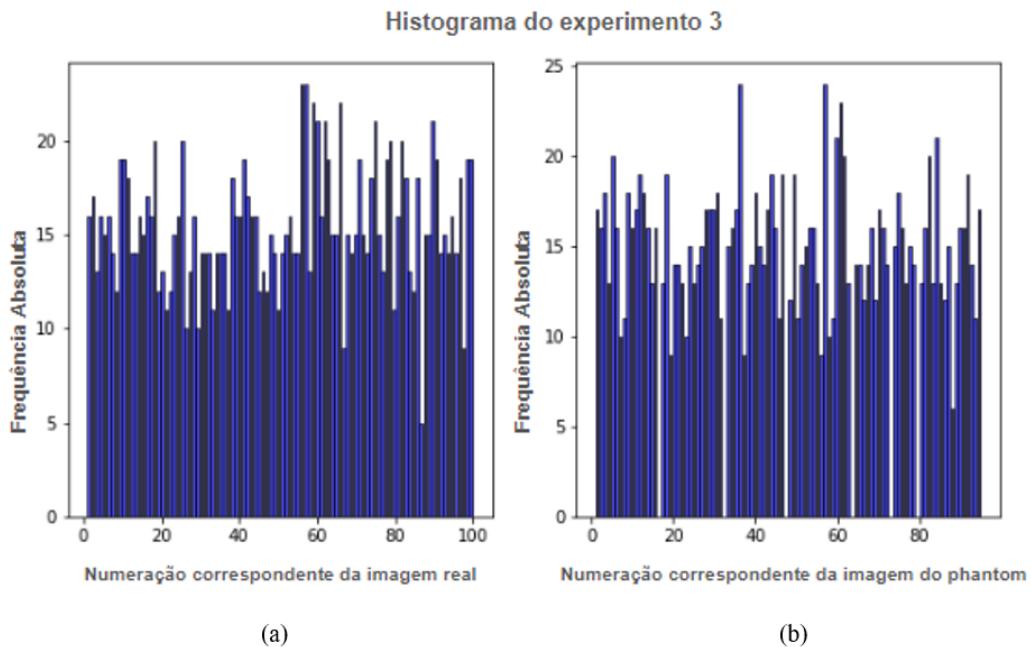
Figura 14 – Histogramas referentes as ROIs de imagens reais (direita) e do *phantom* (esquerda) exibidas no experimento 2.



Fonte: Elaborada pelo autor

Observando as Figuras 13, 14 e 15, pode-se inferir que, embora existam alguns casos em que as imagens são exibidas com mais frequência do que em outros (principalmente nos experimentos 1 e 2), os histogramas obtidos não estão muito distantes do caso ideal uniforme - quando as imagens do banco de dados interno teriam sido exibidas em ordem

Figura 15 – Histogramas referentes as ROIs de imagens reais (direita) e do *phantom* (esquerda) exibidas no experimento 3.



Fonte: Elaborada pelo autor

completamente aleatória e com aproximadamente a mesma frequência. Portanto, com base nesses resultados, é possível garantir que o *software* de gerenciamento de testes seja consistente com relação à aleatoriedade das imagens exibidas no banco de dados disponível.

2.3.2 Análise estatística dos resultados obtidos pelo *software* de gerenciamento de testes

Um grupo de observadores colaboradores foi criado para a execução dos experimentos propostos pelo *software*. Assim, para detalhar melhor os resultados da avaliação com base nas opiniões registradas, elas foram categorizadas em três subgrupos:

- (a) Experiente - composto por observadores com ampla experiência, que abrange vários anos, na análise visual de imagens mamográficas;
- (b) Intermediário - composto por observadores com alguma experiência e familiaridade com a visualização de imagens mamográficas, mas que não se envolvem rotineiramente na análise visual de mamografias, ou aqueles com experiência mais limitada nesse campo;
- (c) Novato (pouca experiência) - refere-se a observadores com exposição mais recente aos procedimentos de análise desse tipo de imagem (como, por exemplo, estudantes que realizam pesquisas nessa área).

Em uma primeira etapa do desenvolvimento deste trabalho utilizaram-se exclusivamente imagens com características semelhantes ao primeiro banco de imagens, descrito

anteriormente, sendo realizadas 60 avaliações, levando em conta todos os subgrupos de observadores, de forma que os resultados são apresentados na Tabela 1. Nesta avaliação os observadores experientes registraram taxas de precisão de 60,9%, 58,9% e 64,0% para os experimentos 1, 2 e 3, respectivamente. Por outro lado, os observadores do subgrupo intermediário apresentaram taxas de precisão de apenas 48,8%, 54,6% e 54,0% para os mesmos experimentos respectivos, indicando uma clara confusão entre as imagens do *phantom* e as reais. Por fim, para observadores com pouca experiência, as taxas de precisão nesta avaliação variaram de 57,6% a 47,0%. Sem dúvida, esses resultados demonstraram um desafio considerável para os observadores na classificação correta das imagens, pois até mesmo observadores experientes apresentaram uma taxa de erro entre 35% e 40% em todos os experimentos.

Tabela 1 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para a primeira etapa do projeto

Avaliador	Experimento	Porcentagem de Acertos (%)	Desvio Padrão (%)
Experiente	1	60,9	(±2,9)
	2	58,9	(±2,7)
	3	64,0	(±8,0)
Intermediário	1	48,0	(±2,5)
	2	54,6	(±2,5)
	3	54,0	(±7,1)
Novato	1	57,6	(±2,6)
	2	48,3	(±1,2)
	3	47,0	(±5,9)

Fonte: Elaborada pelo autor.

Assim, para a continuação da avaliação, foi introduzido um ajuste no processo experimental, que consistiu na incorporação do segundo banco de imagens juntamente com uma nova alteração do primeiro banco de imagens. A justificativa por trás dessa mudança é a variação significativa de contraste nas imagens de mamografia produzidas por diferentes equipamentos, dependendo do modelo do sistema de aquisição de imagens digitais utilizado. Consequentemente, nesse estágio, 29 avaliações foram realizadas em imagens do primeiro banco de dados, abrangendo todos os subgrupos de observadores.

O levantamento estatístico referente à análise dos relatórios de resposta fornecidos pelos observadores colaboradores, para os conjuntos de imagens relacionados a ambos os bancos de dados, mostra a porcentagem média de respostas corretas para cada grupo de observadores e também separadamente para cada um dos três experimentos, bem como o desvio padrão, conforme exibido nas Tabelas 2 e 3. Esses dados estatísticos correspondem ao número total de respostas dadas por cada categoria de observadores (conforme descrito anteriormente): 16 para o subgrupo A (Experiente), 7 para o subgrupo B (Intermediário)

e 9 para o subgrupo C (Novato).

Tabela 2 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 1

Avaliador	Experimento	Porcentagem de Acertos (%)	Desvio Padrão (%)
(A) Experiente	1	59,0	(±1,9)
	2	73,3	(±2,4)
	3	58,1	(±6,4)
(B) Intermediário	1	65,0	(±3,2)
	2	75,0	(±3,2)
	3	58,3	(±9,8)
(C) Novato	1	48,0	(±3,1)
	2	42,7	(±4,4)
	3	53,7	(±11,1)

Fonte: Elaborada pelo autor.

Tabela 3 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 2

Avaliador	Experimento	Porcentagem de Acertos (%)	Desvio Padrão (%)
(A) Experiente	1	67,5	(±2,9)
	2	82,6	(±2,4)
	3	75,2	(±7,8)
(B) Intermediário	1	66,7	(±2,1)
	2	84,4	(±2,3)
	3	70,4	(±8,0)
(C) Novato	1	52,0	(±2,7)
	2	48,0	(±4,3)
	3	55,9	(±14,1)

Fonte: Elaborada pelo autor.

Em complemento as Tabelas 4 e 5 mostram a mesma análise, porém, agora, do ponto de vista das taxas de sensibilidade, especificidade e precisão para cada um dos experimentos e subgrupos de observadores (considerando o "sinal"verdadeiro como uma imagem de uma mama real e o "sinal"falso como a imagem obtida pela exposição do *phantom* de mama). A primeira taxa (sensibilidade) corresponde à proporção de sinais verdadeiros classificados corretamente, dada pela razão entre os verdadeiros positivos e o total de sinais verdadeiros (as imagens mamográficas reais). A segunda taxa (especificidade) corresponde ao número de sinais falsos classificados corretamente como tais, o que é dado pela proporção de verdadeiros negativos em relação ao total de sinais falsos (as imagens de mamografia do *phantom* de mama). Por fim, a taxa de "precisão"corresponde à proporção

de classificações corretas (a soma de verdadeiros positivos e verdadeiros negativos) em relação ao número total de sinais.

Tabela 4 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 1)

Avaliador	Experimento	Sensibilidade (%)	Especificidade (%)	Precisão(%)
(A) Experiente	1	46,2	73,6	59,0
	2	75,0	77,5	76,3
	3	46,7	72,4	58,1
(B) Intermediário	1	54,1	82,6	65,0
	2	75,0	75,0	75,0
	3	46,8	70,5	58,3
(C) Novato	1	52,2	41,4	48,0
	2	40,0	45,3	42,7
	3	50,8	57,7	53,7

Fonte: Elaborada pelo autor.

Tabela 5 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 2)

Avaliador	Experimento	Sensibilidade (%)	Especificidade (%)	Precisão(%)
(A) Experiente	1	62,9	74,0	68,2
	2	83,7	83,7	83,7
	3	71,0	79,3	75,2
(B) Intermediário	1	54,3	77,5	66,7
	2	87,6	48,9	50,0
	3	67,5	79,1	74,1
(C) Novato	1	73,3	51,3	60,9
	2	45,3	53,3	49,3
	3	45,1	62,4	55,2

Fonte: Elaborada pelo autor.

2.4 Discussão

2.4.1 Análise estatística de percepção visual

A partir dos dados apresentados na Tabela 2, é possível fazer algumas observações importantes. Por exemplo, ao considerar tanto os observadores com pouca experiência (subgrupo C) quanto os experientes (subgrupo A), fica evidente que eles possuem dificuldades no que se refere à distinção com precisão das categorias das imagens, principalmente ao avaliar casos individuais, como ocorre no experimento 1.

Além disso, uma porcentagem não negligenciável de erros pode ser observada em todos experimentos e subgrupos de observadores. Com relação à categoria de observadores experientes (subgrupo A), é válido ressaltar uma taxa de erro superior a 40% nos casos dos Experimentos 1 e 3. Tal fato indica claramente que mesmo esse tipo de observador foi confundido um número significativo de vezes entre as ROIs extraídas de imagens do *phantom* de mama e das mamas reais. Isso aponta inequivocamente para a validade do *phantom* de mama do ponto de vista da percepção visual como um simulador de mama eficaz a ser utilizado em testes de processamento de imagens destinados a detectar ou interpretar sinais clinicamente relevantes em mamografia digital.

Os resultados também possibilitam identificar o nível de influência que cada tipo de experimento tem sobre essas taxas de precisão e erro. Por exemplo: para os observadores dos subgrupos (A) e (B), observa-se um aumento nas classificações corretas das imagens no Experimento 2. Entretanto, para os Experimentos 1 e 3 as taxas de precisão e erro permanecem praticamente inalteradas. Em termos práticos, isso significa que as informações fornecidas aos observadores no Experimento 2, ou seja, que uma imagem corresponde necessariamente a uma mamografia real e a outra não, levam a um aumento no nível de discernimento para categorizações de imagens.

Ademais, levando em conta os dados da Tabela 2 e os dados de precisão da Tabela 4, um aspecto pode ser destacado para explicar o motivo do subgrupo (C) não seguir o mesmo padrão discutido anteriormente, já que, neste caso, o aumento da taxa de precisão foi registrado no Experimento 3. Assim, foi demonstrado anteriormente que o *software* desenvolvido garante a aleatoriedade na exibição das imagens, permitindo que o mesmo experimento seja realizado várias vezes pelo mesmo observador, resultando em diferentes taxas de precisão/erro. A observação de um aumento na taxa de precisão para o mesmo participante após a repetição do mesmo tipo de avaliação, no caso do Experimento 3, sugere alguma memorização involuntária das imagens exibidas. Esse aspecto é, provavelmente, o principal fator que explica os resultados do subgrupo (C), pois ele é composto por um número menor de avaliadores diferentes que realizaram os testes de avaliação.

De fato, o Experimento 3 apresenta um número maior de ROIs por tela (um total de 54 imagens exibidas durante cada tentativa de avaliação comparativa), o que pode fazer com que o observador "lembre-se" de algumas imagens, mesmo que não tenha conhecimento de suas próprias taxas de precisão anteriores. Essa memorização involuntária do observador ao realizar o mesmo experimento várias vezes foi detectada e também se estendeu aos outros subgrupos; no entanto, parece não ter tido influência significativa nos resultados, pois poucos observadores realizaram o mesmo experimento mais de três vezes.

Em analogia à análise realizada em relação aos testes envolvendo o banco de imagens 1 - com ROIs de menor contraste-, alguns aspectos podem ser destacados em relação às opiniões dos observadores sobre comparações realizadas entre as ROIs do banco de imagens

2, conforme resumido na Tabela 3. Como esse conjunto de dados é composto por imagens de maior contraste, os resultados revelaram um aumento na porcentagem de identificação correta registrada nos testes para conjuntos de colaboradores com mais experiência em análise de imagens mamográficas (ou seja, os grupos experientes e intermediários). No entanto, uma taxa de erro razoável (superior a 32%) ainda é evidente nos resultados dos testes mesmo para esses observadores, principalmente no Experimento 1, que envolve a análise de imagens isoladas. Além disso, é possível notar que as taxas de precisão do subgrupo (C) permaneceram consistentemente baixas para esse conjunto de dados de outras imagens, assim como ocorreu na análise do banco de dados 1.

Outro resultado que pode ser inferido da Tabela 3 é a influência que cada tipo de experimento teve sobre as taxas de precisão dos subgrupos de observadores. Logo, de forma semelhante ao primeiro banco de imagens, o nível de discernimento correto foi maior no caso do Experimento 2 para os subgrupos (A) e (B). No entanto, o interessante é que também houve um aumento nos índices de precisão no Experimento 3 (com exibição de várias ROIs por tela). Isso provavelmente ocorreu devido à facilidade de estabelecer padrões nas imagens ao analisar vários casos simultaneamente, graças ao melhor contraste nesse segundo conjunto de dados em comparação com o banco de imagens 1.

Agora, considerando a análise comparativa das tabelas 4 e 5, dois aspectos devem ser inicialmente enfatizados:

- (a) As taxas de especificidade (para os subgrupos A e B) permanecem consistentemente mais altas do que as taxas de sensibilidade para ambos os conjuntos de imagens;
- (b) Entretanto, as taxas de sensibilidade (para todos os subgrupos) foram baixas, especialmente no caso da Tabela 4 (banco de imagens 1, menor contraste), onde permaneceram em torno de 50% ou menos nos experimentos 1 e 3 (apenas no experimento 2 que a taxa atingiu 75%, conforme discutido anteriormente). Mesmo na Tabela 5, com as imagens de melhor contraste as taxas de sensibilidade permaneceram na faixa de 60% a 70% (exceto no Experimento 2, mais uma vez).

Isso significa que, em primeiro lugar, houve uma tendência de identificar corretamente as imagens do simulador de mama como pertencentes a essa classe, conforme foi destacado em (a). Contudo, a baixa taxa de sensibilidade obtida em todos experimentos e para todos os subgrupos de avaliadores indica que um número significativo de imagens reais não foi identificado corretamente. Portanto, ao analisar especificamente este parâmetro, ou seja, a baixa sensibilidade geral nos dados registrados nessas tabelas, é possível reconhecer que, do ponto de vista da percepção visual, houve confusão entre imagens reais e simuladas para todos os grupos de observadores colaboradores.

Além disso, os resultados das Tabelas 4 e 5 são obviamente um reflexo dos resultados apresentados nas Tabelas 2 e 3, respectivamente. Também é perceptível que todas as três taxas (sensibilidade, especificidade e precisão) aumentam no caso do Experimento 2, especialmente para os subgrupos (A) e (B). Embora esses índices estejam diretamente correlacionados com os dados de erro e de dados corretos anteriores, eles oferecem uma visão mais clara sobre qual tipo de imagem (da mamografia real ou da exposição do phantom) foi classificada com mais precisão pelos participantes. Particularmente no caso do Experimento 2, esses valores mais altos indicam que, quando os participantes estavam cientes de que uma das imagens correspondia a uma mamografia real e a outra não, houve discernimento mais eficaz entre os dois tipos de imagens, mesmo para o conjunto de dados com menor contraste.

Também é interessante analisar as taxas de precisão nessas duas tabelas. Assim, fica bastante claro que os dados da tabela 3, por exemplo, mostram um nível razoavelmente alto de confusão para todos os subgrupos de participantes ao analisar uma única imagem (Exp. 1) ou várias ROIs (Exp. 3). De fato, os experimentos não conseguiram atingir nem mesmo uma taxa de precisão de 60% nesses casos. E mesmo no caso de imagens de alto contraste (Tabela 5), essa taxa foi limitada a 66 a 75% para os subgrupos A e B - para esse subgrupo em particular, a precisão de apenas 50% no caso do Experimento 2 é digna de nota.

Além das análises já apresentadas, também é possível estabelecer uma discussão ao considerar o banco de dados de imagens completo 1. Essa avaliação combina os resultados obtidos com a realização de experimentos usando ROIs que compunham o banco de dados de imagens da primeira etapa do estudo e a atual. Os resultados obtidos para o conjunto de dados de teste geral estão resumidos nas Tabelas 6 e 7.

Tabela 6 – Resultados dos experimentos com observadores colaborativos (categorizados nos 3 subgrupos) para o banco de imagens 1 completo

Avaliador	Experimento	Porcentagem de Acertos (%)	Desvio Padrão (%)
(A) Experiente	1	63,6	(±2,72)
	2	66,9	(±2,72)
	3	64,7	(±7,97)
(B) Intermediário	1	59,2	(±2,72)
	2	67,2	(±2,79)
	3	60,6	(±8,05)
(C) Novato	1	48,9	(±2,91)
	2	47,4	(±3,48)
	3	56,5	(±12,35)

Fonte: Elaborada pelo autor.

Portanto, ao comparar os resultados do banco de imagens 1 completo com os dois

Tabela 7 – Taxas de sensibilidade, especificidade e precisão para cada subgrupo de observadores e para cada experimento (banco de imagens 1 completo)

Avaliador	Experimento	Sensibilidade (%)	Especificidade (%)	Precisão(%)
(A) Experiente	1	56,4	68,8	62,7
	2	62,8	71,9	67,4
	3	59,6	71,6	65,2
(B) Intermediário	1	49,0	68,4	57,9
	2	60,5	75,1	67,8
	3	54,2	73,4	63,2
(C) Novato	1	51,6	42,9	49,6
	2	35,2	51,1	47,4
	3	42,4	65,7	56,5

Fonte: Elaborada pelo autor.

bancos de imagens separados- com características de contraste semelhantes -, fica evidente que, apesar de algumas melhorias percentuais em comparação com a primeira etapa deste estudo, há uma clara confusão visual entre as imagens do *phantom* e as mamografias reais. Essa confusão se reflete nas taxas de erro, especialmente entre os experientes em mamografia, que consistentemente excederam 35% em todos os experimentos. Além disso, os valores de sensibilidade mais baixos indicam que um número significativo de imagens reais foi classificado incorretamente, enfatizando que, de fato, houve confusão entre as categorias de imagens.

Dessa forma, esses fatores, juntamente com as taxas consistentemente baixas em quase todos os experimentos para o subgrupo (C), indicam que o procedimento de tentar estabelecer um tipo de *phantom* capaz de simular consistentemente a estrutura de distribuição interna aleatória do parênquima mamário em imagens mamográficas digitais foi bem direcionado.

2.4.2 Análise quantitativa das características das imagens nos bancos de imagens

Para ampliar a análise de percepção visual explorada na seção anterior, explorou-se adicionalmente as características de similaridade se ambos conjuntos de imagens por meio das métricas consideradas na Seção 2.2.3. Essa análise quantitativa tem como objetivo complementar não apenas os resultados da avaliação baseada na percepção visual anterior, mas também os resultados de um estudo anterior (SOUZA; MATHEUS; SCHIABEL, 2018), que detalhou o desenvolvimento do *phantom* de mama e suas imagens correspondentes a partir de perspectivas físicas e quantitativas, incluindo análises de textura em imagens adquiridas de equipamentos de mamografia DR. Além disso, uma avaliação realizada em um trabalho anterior (SOUZA *et al.*, 2016) - com foco na influência do uso de imagens desse *phantom* de mama na detecção de microcalcificações por especialistas - destacou a obtenção

de taxas de identificação semelhantes quando comparadas às de imagens mamográficas reais. Portanto, a avaliação de métricas descrita aqui não apenas aprimora as conclusões tiradas desses estudos anteriores (SOUSA; MATHEUS; SCHIABEL, 2018), (SOUSA *et al.*, 2016), mas também permite uma análise abrangente da viabilidade do uso de bancos de dados de imagens gerados a partir das exposições desse simulador de mama (SOUSA; MATHEUS; SCHIABEL, 2018) para avaliar a eficácia das técnicas de esquema CAD.

A Tabela 8 apresenta dados sobre a avaliação quantitativa dos dois bancos de dados de imagens utilizados, mostrando os valores médios e os desvios padrão correspondentes dos métodos de avaliação de similaridade, incluindo a distância de *Bhattacharyya* (KASHYAP, 2019), a correlação (MUKAKA, 2012) e a interseção (LEE; XIN; WESTLAND, 2005). Esses valores foram obtidos pela comparação de cada imagem mamográfica real de um banco de dados de imagens com todas as imagens mamográficas simuladas do mesmo banco de dados de imagens, usando histogramas normalizados.

Tabela 8 – Valores de distância de Bhattacharyya, correlação e interseção para cada banco de dados de imagens

Banco de Imagens	Distância de Bhattacharyya	Correlação	Intersecção
Banco 1	$0,371 \pm 0,163$	$0,751 \pm 0,168$	$32,400 \pm 9,436$
Banco 2	$0,272 \pm 0,087$	$0,665 \pm 0,157$	$59,980 \pm 21,700$

Fonte: Elaborada pelo autor.

Ao analisar os resultados da Tabela 8 para o primeiro banco de dados de imagens, pode-se inferir que o valor da correlação (0,751) está próximo de 1 (que corresponde ao valor de um caso ideal de correlação perfeita entre histogramas). Isso indica que, normalmente, as imagens analisadas têm um nível razoavelmente bom de similaridade, em torno de 75%, o que é esperado, pois as ROIs analisadas não são idênticas, mas se referem à mesma característica visual do parênquima mamário, resultando em diferentes contrastes. O objetivo é justamente possibilitar um banco de dados de imagens versátil, que inclua a simulação de diferentes densidades de mama. Com relação ao baixo valor (0,37) obtido para o cálculo da distância de *Bhattacharyya*: isso indica, de acordo com a definição da métrica (KASHYAP, 2019), que há uma sobreposição entre as distribuições investigadas na comparação proposta. Como essa métrica fornece a distância entre as sobreposições, o caso ideal corresponde à obtenção de um índice com valor nulo (ou seja, sobreposição total). Além disso, como a curva apresenta um comportamento semelhante ao logaritmo natural negativo, o valor obtido está próximo do caso ideal de similaridade (ou ausência de dissimilaridade) entre os conjuntos de imagens. Por fim, o valor da interseção serve para complementar os resultados mencionados acima. Esse índice quantifica as interseções entre os conjuntos avaliados sem um valor máximo predefinido. Portanto, quando examinado

juntamente com as métricas anteriores, ele acentua a presença de semelhanças entre as ROIs escolhidas para este estudo.

Na análise das mesmas métricas para o segundo conjunto de dados de imagens, observa-se mais uma vez que os valores de correlação e distância de *Bhattacharyya* são consideravelmente próximos aos obtidos para o primeiro conjunto de imagens, especialmente a *B-distance*, que é ainda menor (0,27 em comparação com 0,37). Isso sugere que as distâncias entre as distribuições estudadas não são grandes, o que é aumentado por um valor de interseção mais alto (aproximadamente 76% maior do que no primeiro conjunto), destacando um perfil comparável em características visuais (em termos de *pixels* e intensidades de escala de cinza) entre os conjuntos de imagens reais e *phantom*, enfatizando ainda mais as conclusões tiradas da análise de percepção visual discutida anteriormente para esse conjunto de dados.

3 CONCLUSÃO

Com relação à investigação para validar esse *phantom* de mama em termos de características radiográficas visuais, o *software* desenvolvido provou ser uma ferramenta útil para coletar dados relacionados à percepção visual dos grupos de observadores considerados, cumprindo o seu objetivo e sua funcionalidade inicial de execução da avaliação comparativa proposta. Ele fornece uma visualização altamente aleatória de ROIs em diferentes experimentos. Essa aleatoriedade é importante, pois impede o estabelecimento da memória do observador para as imagens apresentadas em cada experimento. No entanto, ela também pode ser vista como uma desvantagem ao analisar as imagens do banco de imagens 2, pois o contraste mais alto nessas imagens não garantiu que todas as ROIs das duas categorias diferentes tivessem contraste semelhante, ao contrário do banco de imagens 1. Como resultado, a aleatoriedade, especialmente no Experimento 2, pode levar à exibição de imagens que não compartilham características semelhantes, o que explica a alta porcentagem de precisão obtida nesse caso.

Além disso, as taxas de precisão que variam de 55% a 65%, em média, indicam que as imagens obtidas a partir da exposição desse *phantom* de mama são frequentemente confundidas com as reais com base na percepção visual da maioria dos participantes, especialmente quando as exposições de raios-X foram feitas usando o mesmo equipamento. Combinando com a validação prévia das propriedades físicas do *phantom* (SOUZA; MATHEUS; SCHIABEL, 2018) e também com os resultados da análise quantitativa das imagens utilizando as métricas aqui consideradas, isso sugere uma direção promissora para seu uso em um futuro banco de dados de imagens para avaliação de algoritmos de diagnóstico auxiliado por computador em mamografia.

Os resultados desses testes refletem uma avaliação comparativa entre imagens obtidas do mesmo equipamento ou de equipamentos cujas imagens resultantes são visualmente semelhantes. Entretanto, se todo o conjunto de imagens tivesse um nível de contraste semelhante independentemente do tipo de equipamento de mamografia digital utilizado, as conclusões do parágrafo anterior certamente seriam mais evidentes. No entanto, devido às características intrínsecas, incluindo os métodos de processamento de imagem empregados por cada fabricante desses sistemas FFDM atuais na exibição das imagens "*for presentation*"(para avaliação médica), essa uniformidade de contraste não está presente nessas imagens. Portanto, novas pesquisas estão em andamento para determinar as curvas características de diferentes sistemas de aquisição de imagens digitais em mamografia, com o objetivo de desenvolver uma ferramenta computacional, entre outras finalidades, capaz de ajustar o contraste em todas essas imagens. Se essa meta for alcançada, a avaliação comparativa realizada neste estudo provavelmente produzirá resultados mais amplos, pois

as imagens presumivelmente teriam características mais independentes do equipamento usado para sua aquisição.

3.1 Desenvolvimentos futuros

- Avaliação do esquema CADx guiado já desenvolvido no grupo de pesquisa LAPIMO para imagens digitalizadas, mas agora para imagens digitais com base no novo banco de imagens do *phantom* que foi validado;
- Utilizar conjuntos de imagens avaliados pelo CADx para "confrontar" a opinião de médicos sobre sinais demarcados neste mesmo conjunto de imagens do *phantom* para verificar o nível de influência que o esquema CAD produz na opinião médica;
- Levantamento de curvas características de diferentes equipamentos mamográficos digitais para o desenvolvimento de uma ferramenta computacional capaz de uniformizar o contraste presente em todas as imagens independentemente do equipamento de aquisição utilizado. E a partir do desenvolvimento dessa ferramenta, realizar um aprimoramento do esquema CADx guiado existente no grupo de pesquisa LAPIMO, a fim de buscar uma maior taxa de eficácia da ferramenta em questão, correspondendo ao atual estado da arte dos equipamentos mamográficos;

REFERÊNCIAS

- BAKIC, P. R. *et al.* Optimized simulation of breast anatomy for virtual clinical trials. In: SPIE. **14th International Workshop on Breast Imaging (IWBI 2018)**. [S.l.: s.n.], 2018. v. 10718, p. 507–516.
- BARCELOS BRUNO PEREIRA NUNES, S. M. S. D. E. T. R. d. C. D. L. M. N. C. T. R. R. M. R. B.; FACCHINI, L. A. Utilization of breast cancer screening in brazil: An external assessment of primary health care access and quality improvement program. **Health Systems & Reform**, Taylor Francis, v. 4, n. 1, p. 42–55, 2018. Disponível em: <https://doi.org/10.1080/23288604.2017.1405770>.
- BHUSHAN, A.; GONSALVES, A.; MENON, J. U. Current state of breast cancer diagnosis, treatment, and theranostics. **Pharmaceutics**, MDPI, v. 13, n. 5, p. 723, 2021.
- BRADY, A. P. Error and discrepancy in radiology: inevitable or avoidable? **Insights into imaging**, Springer, v. 8, p. 171–182, 2017.
- 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.
- CALDWELL, C. B.; YAFFE, M. J. Development of an anthropomorphic breast phantom. **Medical physics**, Wiley Online Library, v. 17, n. 2, p. 273–280, 1990.
- CHA, S.-H.; SRIHARI, S. N. On measuring the distance between histograms. **Pattern Recognition**, Elsevier, v. 35, n. 6, p. 1355–1370, 2002.
- COMPUTERIZED IMAGING REFERENCE SYSTEMS. **BR3D BREAST IMAGING PHANTOM**. 2023. Disponível em: <https://www.cirsinc.com/products/mammography/br3d-breast-imaging-phantom/>. Acesso em: 25 novembro 2023.
- EKPO, E. U.; ALAKHRAS, M.; BRENNAN, P. Errors in mammography cannot be solved through technology alone. **Asian Pacific journal of cancer prevention: APJCP**, Shahid Beheshti University of Medical Sciences, v. 19, n. 2, p. 291, 2018.
- FUNDATION, P. S. **Python Language Site: Documentation**. 2023. Disponível em: <https://www.python.org/doc/>.
- GEORGIAN-SMITH, D. *et al.* Blinded comparison of computer-aided detection with human second reading in screening mammography. **American Journal of Roentgenology**, Am Roentgen Ray Soc, v. 189, n. 5, p. 1135–1141, 2007.
- 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.
- HASSAN, N. M.; HAMAD, S.; MAHAR, K. Mammogram breast cancer cad systems for mass detection and classification: a review. **Multimedia Tools and Applications**, Springer, v. 81, n. 14, p. 20043–20075, 2022.
- HEATH, M. *et al.* The digital database for screening mammography. In: **5th International Workshop on digital mammography**. [S.l.: s.n.], 2000. p. 212–218.

INCA. **Atualização em mamografia para técnicos em radiologia.** [S.l.: s.n.], 2019. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/2a_edicao_atualizacao_em_mamografia_para_tecnicos_em_radiologia_2019.pdf.

INSTITUTO NACIONAL DE CANCER - INCA. **Câncer de mama.** 2023. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/mama>. Acesso em: 21 outubro 2023.

KASHYAP, R. The perfect marriage and much more: Combining dimension reduction, distance measures and covariance. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 536, p. 120938, 2019.

KATZEN, J.; DODELZON, K. A review of computer aided detection in mammography. **Clinical imaging**, Elsevier, v. 52, p. 305–309, 2018.

LEE, S.-M.; XIN, J. H.; WESTLAND, S. Evaluation of image similarity by histogram intersection. **Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur**, Wiley Online Library, v. 30, n. 4, p. 265–274, 2005.

MAHMOOD, T. *et al.* Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. **Plos one**, Public Library of Science San Francisco, CA USA, v. 17, n. 1, p. e0263126, 2022.

MASUD, R. *et al.* Computer-aided detection for breast cancer screening in clinical settings: scoping review. **JMIR medical informatics**, JMIR Publications Inc., Toronto, Canada, v. 7, n. 3, p. e12660, 2019.

MATHEUS, B. R. N. **Sistema JAVA para gerenciamento de esquema CADx em mamografia.** 2015. Tese (Doutorado) — Universidade de São Paulo, 2015.

MOREIRA, I. C. *et al.* Inbreast: toward a full-field digital mammographic database. **Academic radiology**, Elsevier, v. 19, n. 2, p. 236–248, 2012.

MUKAKA, M. M. A guide to appropriate use of correlation coefficient in medical research. **Malawi medical journal**, v. 24, n. 3, p. 69–71, 2012.

ONCOGUIA. **Estimativas de câncer no Brasil.** 2024. Disponível em: <https://www.oncoguia.org.br/conteudo/cancer-em-numeros/16792/1343/#:~:text=N%C3%A3o%20do%20c%C3%A2ncer%20no%20Brasil,3%25%20do%20total%20de%20casos>. Acesso em: 22 março 2024.

SAMPAT, M. P. *et al.* A model-based framework for the detection of spiculated masses on mammography a. **Medical physics**, Wiley Online Library, v. 35, n. 5, p. 2110–2123, 2008.

SCHIABEL, H.; MATHEUS, B. R. N.; CARDOSO, F. J. F. “features of a free mammography database to manage a random breast phantom images dataset intended for testing cad schemes. **Int. J. of Health Science**, Atena Editora, n. 65, p. 1–12, 2022.

SILVA, E. S. d. **Simulador radiográfico antropomórfico de mama humana.** 2012. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2012.

- SOUSA, M.; MATHEUS, B.; SCHIABEL, H. Development of a structured breast phantom for evaluating cade/dx schemes applied on 2d mammography. **Biomedical Physics & Engineering Express**, IOP Publishing, v. 4, n. 4, p. 045018, 2018.
- SOUSA, M. A. *et al.* Observer study to evaluate the simulation of mammographic calcification clusters. *In: SPIE. Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment.* [S.l.: s.n.], 2016. v. 9787, p. 360–368.
- SOUSA, M. A. *et al.* A comparison of methods to evaluate gray scale response of tomosynthesis systems using a software breast phantom. *In: SPIE. Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment.* [S.l.: s.n.], 2017. v. 10136, p. 120–130.
- SUCKLING, J. The mammographic images analysis society digital mammogram database. *In: Excerpta Medica. International Congress Series*, 1994. [S.l.: s.n.], 1994. v. 1069, p. 375–378.
- VYBORNY, C. J.; GIGER, M. L.; NISHIKAWA, R. M. Computer-aided detection and diagnosis of breast cancer. **Radiologic Clinics of North America**, Elsevier, v. 38, n. 4, p. 725–740, 2000.
- WARREN, R. *et al.* Mammogrid—a prototype distributed mammographic database for europe. **Clinical radiology**, Elsevier, v. 62, n. 11, p. 1044–1051, 2007.
- WORLD HEALTH ORGANIZATION, CANCER. **World Health Organization**. 2023. Disponível em: https://www.who.int/health-topics/cancer#tab=tab_1. Acesso em: 15 outubro 2023.