

Aplicação de Detecção de Objetos na Avaliação de Qualidade de Imagem em Mamografia

Roger Ferreira da Silva

*Centro de Desenvolvimento da Tecnologia Nuclear, CDTN
Programa de Pós-Graduação em Engenharia Elétrica - UFMG
Belo Horizonte, Brasil
roger.silva@cdtn.br*

Laila F. Moreira de Almeida

*Centro de Desenvolvimento da Tecnologia Nuclear, CDTN
Belo Horizonte, Brasil
laila.almeida@cdtn.br*

Maria do Socorro Nogueira

*Centro de Desenvolvimento da Tecnologia Nuclear, CDTN
Belo Horizonte, Brasil
mnogue@cdtn.br*

Frederico Gadelha Guimarães

*Programa de Pós-Graduação em Engenharia Elétrica - UFMG
Belo Horizonte, Brasil
fredericoguimaraes@ufmg.br*

Resumo—O câncer de mama é o segundo mais incidente no mundo, com 2,3 milhões de novos casos em 2022, sendo o de maior incidência e mortalidade entre as mulheres. A mamografia tem papel fundamental na detecção precoce e diminuição da mortalidade do câncer de mama e deve estar dentro de padrões rigorosos para permitir o diagnóstico apropriado. Entre os testes regulares e obrigatórios para a garantia da qualidade do exame, está a avaliação visual da imagem radiográfica do fantoma (simulador) de mama, a qual deve demonstrar a capacidade do sistema de mamografia em visibilizar as estruturas de interesse. Como a avaliação é feita por humanos, há um evidente fator subjetivo envolvido, que pode afetar sua precisão. O presente trabalho propõe a aplicação de redes neurais convolucionais de detecção de objetos (RNC-DO) para a automação do processo de avaliação de imagens de fantasmas de mamografia do Colégio Brasileiro de Radiologia (CBR), utilizando 815 imagens escaneadas do Programa de Controle de Qualidade em Mamografia de Minas Gerais (PECQMamo-MG). O algoritmo YOLOv7, tamanhos de imagem 1280x640 e 1600x800 e formatos de imagem JPG e PNG são utilizados para obtenção de modelos de detecção de fibras, massas, microcalcificações e grades. Os resultados mostraram o melhor desempenho do uso de imagens JPG 1280x640 (mAP de teste 0.896); a precisão da detecção de cada tipo de estrutura ficou acima de 96%, chegando a 100% para grade; nenhuma estrutura foi erroneamente predita como uma estrutura diferente.

Palavras-chave—qualidade de mamografia; fantoma; redes neurais; detecção de objetos.

I. INTRODUÇÃO

O câncer é um grande problema de saúde pública, figurando como uma das principais causas de morte na população mundial. O câncer de mama é o de maior incidência e mortalidade entre as mulheres [1]. No Brasil, o câncer de mama vem após o de pele não melanoma em número de novos casos e é o câncer de maior mortalidade [2].

A mamografia tem papel fundamental na detecção precoce e diminuição da mortalidade do câncer de mama, e deve estar dentro de padrões rigorosos, permitindo uma imagem com

qualidade diagnóstica. Existem várias diretrizes para a garantia da qualidade desse exame, como as do Colégio Americano de Radiologia (ACR) [3], [4], da Comissão Europeia [5] e da Agência Internacional de Energia Atômica (IAEA) [6], [7]. Entre os testes requisitados, está o uso de simuladores de mama (fantomas) para a avaliação do desempenho do equipamento na geração da imagem.

O Programa Nacional de Qualidade em Mamografia (PNQM), atribuído ao Colégio Brasileiro de Radiologia e Diagnóstico por Imagem (CBR) e ao Instituto Nacional de Câncer (INCA), trabalha monitorando os serviços de mamografia quanto à qualidade da imagem, do laudo e da dose de radiação [8], [9]. Em Minas Gerais, o Programa Estadual de Controle de Qualidade em Mamografia (PECQMamo-MG) monitora a qualidade dos mamógrafos dos serviços públicos e privados do estado por meio da avaliação da imagem radiográfica do fantoma. O sistema Atalanta [10], desenvolvido em conjunto com o Centro de Desenvolvimento da Tecnologia Nuclear (CDTN), é utilizado para o gerenciamento e divulgação das informações do PECQMamo-MG [11]–[13] e é a fonte de dados do presente trabalho.

De forma geral, o fantoma de mamografia é um objeto que simula as características radiográficas dos tecidos da mama e suas estruturas. O fantoma de mamografia especificado pelo Colégio Brasileiro de Radiologia e Diagnóstico por Imagem (CBR) [14] é o mais utilizado em Minas Gerais e possui estruturas simuladoras de fibras, massas e microcalcificações, em diferentes tamanhos, além de grades com diferentes espaçamentos. O fantoma CBR é mostrado nas Figs. 1 e 2.

Parte da avaliação das imagens radiográficas desses fantasmas consiste em identificar a visibilidade de cada estrutura, o que deve ser feito em condições controladas e por profissionais treinados das instalações de mamografia e dos órgãos reguladores. No PECQMamo-MG, os 496 serviços de mamografia cadastrados enviam semestralmente uma imagem radiográfica física (filme) do fantoma para a análise, o que não isenta os serviços da obrigatoriedade do teste na periodicidade diária

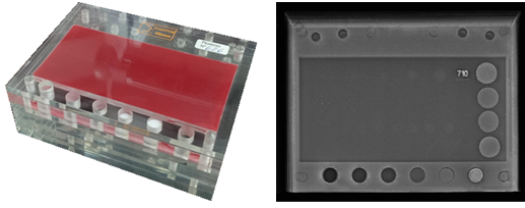


Fig. 1: Objeto (esquerda) e imagem radiográfica (direita) do fantoma de mamografia CBR. Fonte: Elaborado pelo autor

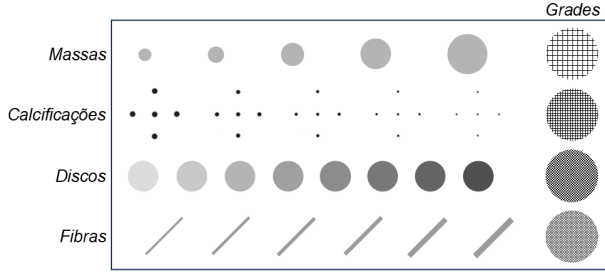


Fig. 2: Esquema da região de cera (área em vermelho) do fantoma de mamografia CBR com indicação das estruturas simuladoras. Fonte: Elaborado pelo autor

exigida pela ANVISA [11].

Como a avaliação é feita por humanos, há um evidente fator subjetivo envolvido que pode afetar sua precisão. Além disso, a análise é demorada, exige treinamento e é afetada por fatores como condições de visualização, nível de experiência e fadiga ocular. A automação do processo por meio de visão computacional o tornaria mais objetivo e padronizado, reduziria a carga de trabalho dos avaliadores e ajudaria os serviços de mamografia na sua autoavaliação. O sistema poderia, ainda, ser usado para treinamento.

Embora técnicas de aprendizado profundo (redes neurais artificiais) sejam uma abordagem comum para detectar anomalias em mamografias [15]–[17], o número de trabalhos que as aplicam para avaliação de qualidade a partir de imagens de fantasmas de mamografia ainda é limitado. Especificamente para o fantoma CBR, não há trabalhos publicados que tenham usado aprendizado profundo na avaliação de sua imagem; aplicam apenas técnicas tradicionais de processamento digital de imagem e extração de atributos [18]–[21].

Redes neurais convolucionais de detecção de objetos (RNC-DO) permitem uma abordagem em que a localização e a classificação dos objetos de interesse são feitas pelo mesmo algoritmo. A técnica de detecção de objetos aplicada à avaliação de imagens dos fantasmas emula o processo visual realizado pelos especialistas, em que o parecer de qualidade é feito pela contagem das estruturas detectadas na imagem. Com a disponibilização de dados de qualidade em quantidade suficiente, é possível criar um modelo preditivo robusto em relação às variabilidades da prática clínica.

Neste contexto, o presente trabalho pretende aplicar RNC-DO, mais especificamente YOLOv7, na avaliação de qualidade de imagens de fantasmas CBR, por meio da detecção das estru-

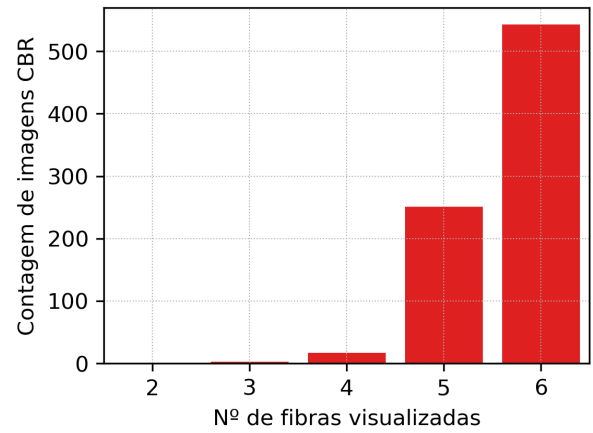


Fig. 3: Distribuição de número de fibras visualizadas nas imagens de fantasmas de mamografia CBR. Fonte: Elaborado pelo autor

turas das estruturas fibras, massas, microcalcificações e grades. São utilizados diferentes formatos e tamanhos e imagem para se comparar o desempenho para cada configuração de imagem.

II. MÉTODO

Para o treinamento e teste dos modelos de RNC-DO YOLOv7 para detecção das estruturas fibra, microcalcificação, massa e grade, é utilizado banco de imagens da prática clínica (PECQMamo-MG), submetido à avaliação visual de especialistas para a identificação das estruturas e anotado com caixas delimitadoras.

A. Banco de imagens

O banco de imagens foi obtido a partir do escaneamento dos filmes radiográficos de fantasmas de mama CBR, obtidos em 287 mamógrafos de 53 diferentes modelos, e localizados em 106 municípios de Minas Gerais. Os mamógrafos são de tecnologia digital (DR), computadorizada (CR) e analógica. O banco é bastante abrangente em relação aos equipamentos que deram origem às imagens, sendo essa representatividade importante para o aprendizado de máquina.

O escaneamento foi realizado em alta qualidade e utilizando um *scanner* dedicado. Após o escaneamento, as imagens foram submetidas a pré-processamento, em que a área de cera do fantoma (área onde se encontram as estruturas de interesse) é alinhada e extraída da imagem, resultando em um banco final de 815 imagens.

Nas Figs. 3, 4 e 5 são mostrados os números de imagens de acordo com a quantidade de fibras, microcalcificações e massas, respectivamente, identificadas na avaliação visual. Em todas as imagens de fantoma CBR estão presentes quatro grades (desconsiderando definição visual). Embora o banco seja relativamente pequeno, cada imagem possui várias estruturas identificadas, o que contribui para o cômputo geral de anotações.

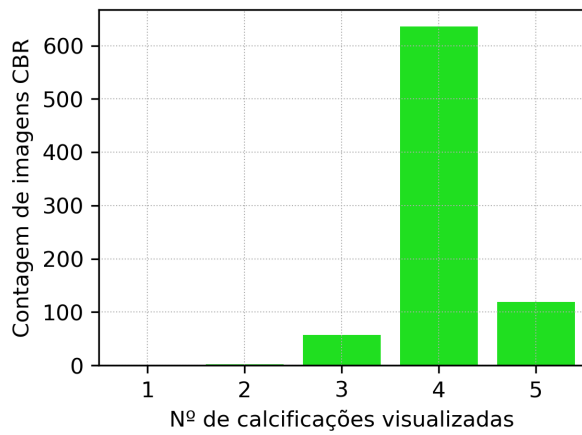


Fig. 4: Distribuição de número de microcalcificações visualizadas nas imagens de fantoma de mamografia CBR. Fonte: Elaborado pelo autor

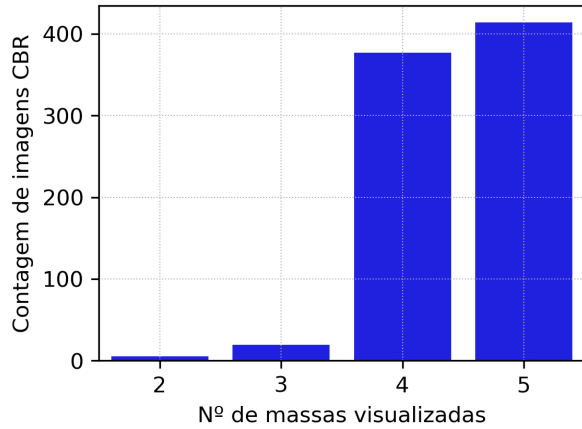


Fig. 5: Distribuição de número de massas visualizadas nas imagens de fantoma de mamografia CBR. Fonte: Elaborado pelo autor

Apesar de essa informação não estar registrada no banco do Atalanta, é importante destacar a grande diversidade de exemplares de fantasmas CBR que foram usados para a obtenção das imagens. Como os fantasmas podem ter pequenas variações construtivas, é importante que essa variabilidade esteja presente no banco de treinamento dos modelos de detecção de objetos.

Para o desenvolvimento dos modelos, o banco de dados foi dividido em conjunto de treinamento (85%) e teste (15%) de forma aleatória e mantendo a proporção original de estruturas. As imagens foram convertidas para os formatos PNG e JPG, e redimensionadas para os tamanhos 1280x640 e 1600x800, para se verificar a melhor configuração de imagem.

B. Avaliação humana das imagens (ground truth)

A avaliação visual de acordo com a metodologia do PECQMamo-MG [11] foi utilizada para a contagem das estruturas presentes em cada imagem, exceto para as quatro

grades, que foram contadas independentemente da sua definição visual. A avaliação da imagem física foi feita por profissional treinado, em condições controladas. Em caso de reprovação, a imagem foi reavaliada por um segundo profissional de maior experiência.

Fibra e massa devem ter pelos menos 80% de sua área visível para serem consideradas na contagem. Um grupo de microcalcificações deve estar visível por completo para ser considerado.

Após a extração da área de cera das imagens de fantoma, elas foram anotadas com caixas delimitadoras utilizando como *ground truth* as estruturas identificadas na avaliação visual.

C. Obtenção dos modelos de detecção

Para a detecção das estruturas fibras, microcalcificações, massas e grades nas imagens de fantoma CBR, foram treinados modelos de detecção de objetos a partir da família de algoritmos YOLOv7. Mais especificamente, foram usadas a arquitetura: YOLOv7, com *backbone* v7-L e *neck* FPN.

O formato da imagem (JPG e PNG) e o tamanho da imagem (1280x640 e 1600x800) foram considerados como hiperparâmetros, com o intuito de comparar sua influência no desempenho final da detecção das estruturas. Outros hiperparâmetros (número de *workers*, tamanho do *batch*, número de épocas e taxa de aprendizagem) foram determinados por testes preliminares que não serão mostrados aqui. O número de épocas de treinamento foi 55 e foi utilizado o recurso *save the best*, em que o modelo com o maior mAP de validação é selecionado ao final do número de épocas.

Foi realizado *data augmentation* durante o treinamento com as seguintes estratégias: redimensionamento aleatório da imagem entre 90% e 110% do tamanho original e espelhamento horizontal com probabilidade de 50%. A estratégia de mosaico é padrão do YOLOv7 e também foi utilizada.

O *framework* MMYolo [22], baseado em Pytorch, disponibiliza modelos treinados utilizando o banco de imagens COCO 2017. Usando como ponto de partida modelo pré-treinado, foi realizado o ajuste fino por meio do conjunto de treinamento do banco de imagens, utilizando validação cruzada *k-fold* com $k=5$. Assim, o total de modelos treinados foi 20. O TensorBoard foi utilizado para acompanhar o treinamento dos modelos.

A melhor combinação de hiperparâmetros foi escolhida após a validação cruzada e, então, o modelo com o maior mAP de validação entre os cinco obtidos para aquela combinação foi selecionado.

Após isso, foram plotadas matrizes de confusão para o modelo selecionado, aplicando limiar de interseção sobre união (IoU) de 0.5 e limiares de confiança variando de 0.1 a 0.9. O limiar de confiança que resultou no melhor balanço entre falsos positivos e falsos negativos foi selecionado. Por fim, o modelo foi testado com o conjunto de teste do banco de imagens e a matriz de confusão gerada.

O código para treinamento, validação e teste dos modelos foi escrito e executado na ferramenta de processamento em nuvem Google Colab, utilizando ambiente de execução com

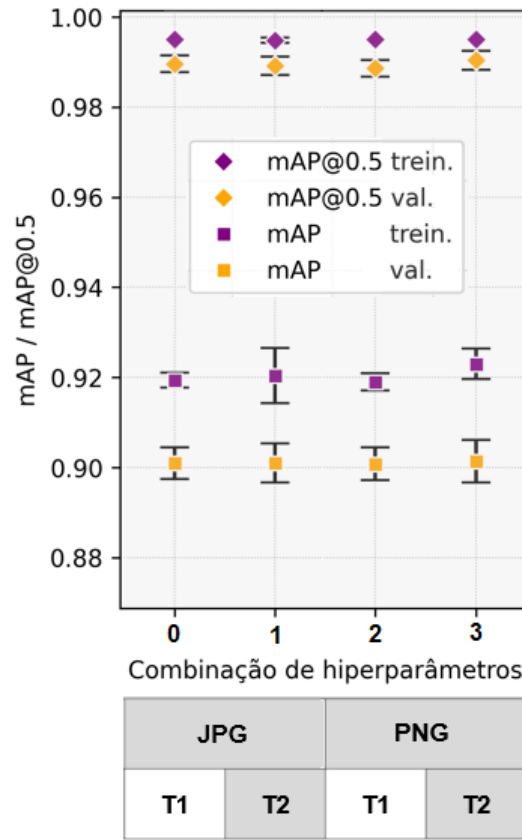


Fig. 6: Média e intervalo de confiança (95%) de mAP e mAP@0.5 de treinamento e validação dos modelos de detecção do fantoma CBR submetidos a validação cruzada k -fold com $k=5$ utilizando as combinação de hiperparâmetros formato imagem (JPG e PNG) e tamanho imagem (1280x640 e 1600x800). O tamanho 1280x640 é identificado como T1 e o tamanho 1600x800 é identificado como T2. Fonte: Elaborado pelo autor

12 CPUs, 90 GB de RAM e GPU NVIDIA A100 com 40 GB de VRAM.

III. RESULTADOS E DISCUSSÃO

Na Fig. 6, são comparados os desempenhos dos modelos treinados com validação cruzada utilizando o conjunto de treinamento e combinação de hiperparâmetros formato imagem e tamanho imagem. São apresentadas as médias (k -fold com $k=5$) e intervalo de confiança (95%) das métricas mAP e mAP@0.5 para treinamento e validação.

Podemos observar que, em geral, não houve diferença significativa entre o uso de imagens 1280x640 ou 1600x800, para as métricas de validação. O mesmo se observa na comparação entre o uso de imagens JPG e PNG. Dessa forma, foi escolhida a combinação de hiperparâmetros 0 (imagem JPG 1280x640), pois resultou em modelos com melhor balanço entre desempenho e uso de recursos computacionais (imagens com menor tamanho em memória). Como na validação cruzada haviam sido obtidos cinco modelos com esses hiperparâmetros, foi

TABELA I: Métricas de desempenho do detector de estruturas CBR. *Average precision* (AP) e *Average precision* com IoU de 50% (AP@0.5) na detecção de fibra, microcalcificação, massa e grade, e médias (mAP e mAP@0.5), para conjuntos de dados de treinamento, validação e teste

MÉTRICA		TREINAMENTO	VALIDAÇÃO	TESTE
		PECQMamo train-80%	PECQMamo train-20%	PECQMamo teste
AP	fibra	0,864	0,833	0,821
	microcalc.	0,994	0,992	0,993
	massa	0,823	0,800	0,770
	grade	1,000	0,999	1,000
AP@0.5	fibra	0,990	0,987	0,987
	microcalc.	1,000	0,999	0,999
	massa	0,990	0,982	0,964
	grade	1,000	1,000	1,000
mAP		0,920	0,906	0,896
mAP@0.5		0,995	0,992	0,988

selecionado, então, o modelo com maior mAP de validação (0.906). Seu mAP@0.5 de validação foi 0.992.

Após isso, por meio da comparação das matrizes de confusão em vários limiares de confiança, para escolha do melhor balanço entre falsos positivos e falsos negativos, foi escolhido o limiar de confiança 0.6 para o modelo final.

As métricas do modelo final, que será chamado aqui de **detector de estruturas CBR**, são apresentadas nas Tab. I. Podemos observar que não houve grandes diferenças de performance na comparação das métricas entre treinamento, validação e teste, indicando que não houve sobreajuste dos modelos aos dados de treinamento.

Nota-se, comparando as métricas de avaliação AP e AP@0.5 entre as estruturas, que a massa teve a pior performance na sua detecção, seguida pela estrutura fibra. A performance mais baixa na detecção de massas e fibras pode ser explicada por seu baixo contraste e, principalmente no caso da massa, por terem o contorno pouco definido, o que as torna mais difíceis de se detectar inclusive na avaliação visual.

Ainda pela Tab. I, observamos que a detecção de microcalcificação teve resultados excelentes, com métricas de teste maiores que 0.97, apesar das variações esperadas para essa estrutura entre exemplares de fantoma.

A matriz de confusão do detector de estruturas CBR, obtida a partir do conjunto de teste, é mostrada nas Fig. 7. A matriz de confusão permite uma melhor avaliação do uso final do modelo, pois já considera o limiar de IoU e o limiar de confiança escolhidos, além de oferecer uma visão mais detalhada das predições.

Observa-se nas Fig. 7 que houve falsos positivos (última linha) e falsos negativos (última coluna) nas predições (detecções) de todas as estruturas, exceto grade, para a qual se obteve 100% de acerto. Entre as outras estruturas, a precisão (proporção de detecções realmente corretas) variou entre 96% e 99%. Houve maior número de falsos negativos para fibra e massa.

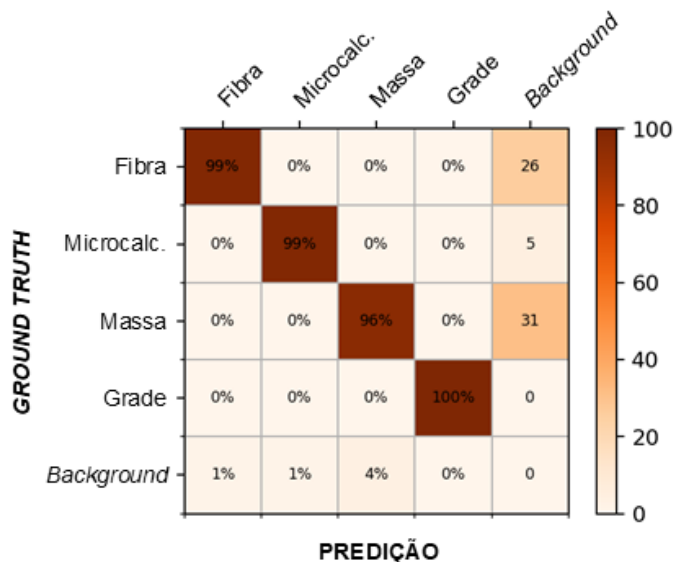


Fig. 7: Matriz de confusão do detector de estruturas CBR para conjunto de teste. São mostrados os números de detecções normalizados por coluna, exceto para a última coluna (*background*), na qual são apresentados os valores brutos. Fonte: Elaborado pelo autor

Por fim, nota-se pelas matrizes que nenhuma estrutura foi erroneamente detectada como uma estrutura diferente.

IV. CONCLUSÕES

Este trabalho propõe a aplicação de Rede Neural Convucional de Detecção de Objetos para a detecção das estruturas fibra, massa, microcalcificação e grade em imagens radiográficas de fantasmas de mama do Colégio Brasileiro de Radiologia e Diagnóstico por Imagem (CBR). Resultados com o uso de 815 imagens escaneadas de vários serviços de mamografia de Minas Gerais (PECQMamo-MG) mostraram o potencial do sistema para auxiliar na avaliação de qualidade de imagem em mamografia. A precisão da detecção de cada tipo de estrutura ficou acima de 96%, chegando a 100% para grade.

Trabalhos futuros incluirão testes utilizando imagens originalmente digitais, utilização de algoritmos de estado da arte para a obtenção de novos modelos de detecção e comparação de performance, e a integração do modelo de detecção desenvolvido ao sistema Atalanta, para utilização e teste pelo PECQMamo-MG.

REFERÊNCIAS

- [1] International Agency for Research on Cancer, "Cancer Today," 2024.
- [2] Instituto Nacional de Câncer, "Estatísticas de câncer," 2023b.
- [3] American College of Radiology, *1999 Mammography Quality Control Manual*. Reston: ACR, 1999.
- [4] American College of Radiology, *2018 Digital Mammography Quality Control Manual*. Reston: ACR, 2 ed., 2020.
- [5] European Commission, *European guidelines for quality assurance in breast cancer screening and diagnosis*. Publications Office, 4, supplements ed., 2013.

- [6] International Atomic Energy Agency, *Quality Assurance Programme for Screen Film Mammography*. No. 2 in Human Health Series, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2009.
- [7] International Atomic Energy Agency, *Quality Assurance Programme for Digital Mammography*. No. 17 in Human Health Series, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2011.
- [8] S. M. da Silva, J. E. Peixoto, F. E. Aduan, L. A. Urban, L. V. Travassos, E. O. Canella, S. F. Rego, A. C. Campos, A. M. Araújo, M. B. Schaefer, A. L. Kefalas, J. L. Francisco, N. M. Maranhão, R. P. dos Santos, S. di Pace Bauab, J. M. Kalaf, L. F. Chala, H. L. Couto, R. R. Jakubiak, and G. W. Tinoco, "Two national mammography quality certification programs in Brazil: Framework and main outcomes between 2017 and 2021," *Journal of Cancer Policy*, vol. 38, p. 100437, dec 2023.
- [9] Brasil, "Portaria de Consolidação nº 5, de 28 de setembro de 2017, do Ministério da Saúde," *Diário Oficial da União*, 2017.
- [10] Centro de Desenvolvimento da Tecnologia Nuclear, "Atalanta," 2020.
- [11] F. A. d. R. Silva and L. d. A. Vieira, *Instrutivo do Programa Estadual de Controle de Qualidade em Mamografia*. Belo Horizonte: Secretaria de Estado de Saúde de Minas Gerais, 2017.
- [12] Minas Gerais, "Resolução nº 5635, de 3 de março de 2017, da Secretaria de Estado de Saúde," 2017.
- [13] Minas Gerais, "Resolução nº 6.711, de 17 de abril de 2019, da Secretaria de Estado de Saúde," 2019.
- [14] M. Oliveira, M. Nogueira, E. Guedes, M. Andrade, J. Peixoto, G. Joana, and J. Castro, "Average glandular dose and phantom image quality in mammography," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 580, pp. 574–577, sep 2007.
- [15] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1–16, mar 2019.
- [16] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC Bioinformatics*, vol. 20, p. 281, jun 2019.
- [17] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, "Convolutional neural networks for breast cancer detection in mammography: A survey," *Computers in Biology and Medicine*, vol. 131, p. 104248, apr 2021.
- [18] P. d. C. Santana, "Desenvolvimento de metodologia para avaliação automatizada da qualidade de imagens digitalizadas em mamografia," mestrado em ciência e tecnologia das radiações, minerais e materiais, Centro de Desenvolvimento da Tecnologia Nuclear, Belo Horizonte, 2010.
- [19] B. Barufaldi, S. S. De Oliveira, L. V. Batista, H. Schiabel, and M. S. C. Almeida, "A COMPUTER AIDED DETECTION SYSTEM FOR MICROCALCIFICATIONS IN BREAST PHANTOM IMAGES," in *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pp. 414–418, SciTePress - Science and Technology Publications, 2012.
- [20] C. M. V. Barragan, *Desenvolvimento de algoritmos para avaliação da imagem do Simulador Phantom mama para o controle de qualidade em mamografia*. Doutorado em ciências e técnicas nucleares, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.
- [21] C. M. Viloria, L. Paixão, P. Santana, F. Leyton, and M. d. S. Nogueira, "Validação de um programa semiautomática para a avaliação de microcalcificações em um phantom de mama," *Scientia Plena*, vol. 12, pp. 1–7, jul 2016.
- [22] M. Contributors, "MMYOLO: OpenMMLab YOLO series toolbox and benchmark," 2022.