

ESCOLA DE ENFERMAGEM NOVA ESPERANÇA LTDA
FACULDADE DE ENFERMAGEM NOVA ESPERANÇA - FACENE

GUSTAVO MARCOS FERREIRA DA SILVA

**CARACTERIZAÇÃO DE NÓDULOS MAMÁRIOS POR ANÁLISE RADIÔMICA
DE MAMOGRAFIAS**

JOÃO
PESSOA 2025

GUSTAVO MARCOS FERREIRA DA SILVA

**CARACTERIZAÇÃO DE NÓDULOS MAMÁRIOS POR ANÁLISE RADIÔMICA
DE MAMOGRAFIAS**

Trabalho de conclusão de curso apresentado
à Faculdade de Enfermagem Nova Esperança
- FACENE, como exigência para obtenção
do título de Tecnólogo em Radiologia.

ORIENTADOR: Prof. Dr. Alex Cristóvão
Holanda de Oliveira

JOÃO
PESSOA 2025

S581c

Silva, Gustavo Marcos Ferreira da

Caracterização de nódulos mamários por análise radiômica de mamografias / Gustavo Marcos Ferreira da Silva. – João Pessoa, 2025.

16f.; il.

Orientador: Prof.º D.º Alex Cristóvão Holanda de Oliveira.

Trabalho de Conclusão de Curso (Tecnólogo em Radiologia) – Faculdade Nova Esperança - FACENE

1. Algoritmos de Aprendizado de Máquina. 2. Diagnóstico por Imagem. 3. Câncer de Mama. 4. Processamento de Imagem. 5. Visão Computacional. I. Título.

CDU: 615.849:618.19

GUSTAVO MARCOS FERREIRA DA SILVA

**CARACTERIZAÇÃO DE NÓDULOS MAMÁRIOS POR ANÁLISE RADIÔMICA DE
MAMOGRAFIAS**

Trabalho de conclusão de curso apresentado pelo aluno Gustavo Marcos Ferreira da Silva, do curso de Tecnologia em Radiologia da Faculdade de Enfermagem Nova Esperança – FACENE, tendo obtido o conceito de _____, conforme a apreciação da banca examinadora constituída pelos professores:

Aprovado em: _____ de _____ de _____

BANCA EXAMINADORA

Prof. Dr. Alex Cristóvão Holanda de Oliveira
Faculdade de Enfermagem Nova Esperança – FACENE

Prof. Arthur da Nóbrega Carreiro
Faculdade de Enfermagem Nova Esperança – FACENE

Prof.^a Débora Teresa da Rocha Gomes Ferreira de Almeida
Faculdade de Enfermagem Nova Esperança – FACENE

CARACTERIZAÇÃO DE NÓDULOS MAMÁRIOS POR ANÁLISE RADIÔMICA DE MAMOGRAFIAS

SILVA, G. M. F.; OLIVEIRA, A. C. H.

RESUMO

A mamografia é um exame essencial para a detecção precoce do câncer de mama, embora apresente limitações, como taxas elevadas de falsos positivos e negativos. Com o avanço da inteligência artificial, a análise radiômica tem sido empregada para aprimorar a caracterização de nódulos mamários, permitindo a extração de dados quantitativos de imagens médicas. Este estudo teve como objetivo implementar um método de caracterização de nódulos mamários por meio da análise radiômica de mamografias, utilizando a linguagem de programação Python. Foram utilizadas imagens do banco público CBIS-DDSM, segmentadas com o uso de máscaras binárias para isolar as regiões dos nódulos. Foram extraídas características geométricas e texturais, como dimensão fractal, contraste, homogeneidade, correlação e entropia, empregando-se a matriz de coocorrência de níveis de cinza e a análise fractal. A correlação entre essas características e variáveis clínicas foi avaliada por meio do algoritmo Random Forest, com divisão dos dados em conjuntos de treino e teste. Os resultados indicaram que a dimensão fractal apresentou maior relevância na classificação, seguida por características texturais como desvio-padrão e correlação. No entanto, o modelo apresentou acurácia limitada, variando entre 22% e 39%, com desempenho inferior nas classes menos representadas, evidenciando o impacto do desbalanceamento do conjunto de dados. Conclui-se que a análise radiômica contribui para a caracterização de nódulos mamários; porém, melhorias metodológicas, como o balanceamento das classes e o uso de modelos mais sofisticados, são necessárias para aprimorar a acurácia e a aplicabilidade clínica dos resultados.

PALAVRAS-CHAVE: Algoritmos de Aprendizado de Máquina; Diagnóstico por Imagem; Câncer de Mama; Processamento de Imagem; Visão Computacional.

CHARACTERIZATION OF BREAST NODULES BY RADIOMIC ANALYSIS OF MAMMOGRAPHS

SILVA, G. M. F.; OLIVEIRA, A. C. H.

ABSTRACT

Mammography is an essential imaging method for the early detection of breast cancer, although it presents limitations such as high rates of false positives and false negatives. With advances in artificial intelligence, radiomic analysis has been used to improve the characterization of breast nodules by extracting quantitative data from medical images. This study aimed to implement a method for characterizing breast nodules through radiomic analysis of mammograms using the Python programming language. Images from the public CBIS-DDSM dataset were used, segmented with binary masks to isolate the nodule's regions. Geometric and textural features were extracted, including fractal dimension, contrast, homogeneity, correlation, and entropy, using the gray-level co-occurrence matrix and fractal analysis. The correlation between these features and clinical variables was evaluated using the Random Forest algorithm, with data divided into training and test sets. Results indicated that fractal dimension was the most relevant feature for classification, followed by textural attributes such as standard deviation and correlation. However, the model showed limited accuracy, ranging from 22% to 39%, with lower performance in underrepresented classes, highlighting the impact of dataset imbalance. It is concluded that radiomic analysis contributes to the characterization of breast nodules; nevertheless, methodological improvements, such as class balancing and the use of more advanced models, are necessary to enhance the accuracy and clinical applicability of the results.

KEYWORDS: Machine Learning Algorithms. Diagnostic Imaging. Breast Cancer. Image Processing.

INTRODUÇÃO

A mamografia configura-se como um exame radiológico indispensável para a detecção precoce do câncer de mama, permitindo a identificação de lesões em estágios iniciais, o que contribui para o aumento das taxas de sobrevivência¹. Dessa forma, programas de rastreamento têm sido instituídos para favorecer diagnósticos mais precisos e tratamentos eficazes, bem como o monitoramento contínuo da evolução da doença².

Com o avanço das tecnologias, a inteligência artificial (IA) tem emergido como uma ferramenta promissora na interpretação das mamografias. Os algoritmos de IA auxiliam na identificação de anomalias e na redução da carga de trabalho dos radiologistas, possibilitando uma caracterização mais abrangente das lesões quando integrados à análise radiômica^{3,4}.

A radiômica consiste na extração e análise computacional de dados quantitativos provenientes de imagens médicas, associados a informações clínicas e laboratoriais, ampliando a eficácia dos biomarcadores e dos modelos de decisão. Essa abordagem possibilita a avaliação da complexidade espacial do tumor e a identificação de sua heterogeneidade intratumoral, fatores frequentemente relacionados à progressão da doença e à resistência terapêutica⁵.

Apesar de sua relevância, a mamografia apresenta limitações, tais como as taxas de falsos negativos, sobretudo, de falsos positivos, os quais podem conduzir às intervenções desnecessárias e impactar negativamente a saúde física e mental das pacientes⁶. Estudos indicam que as taxas de falsos positivos são mais elevadas entre mulheres de 40 (quarenta) a 49 (quarenta e nove) anos, tendendo a diminuir com o avanço da idade, enquanto os falsos negativos apresentam tendência ao aumento, sem alcançar significância estatística⁷. Essa variabilidade ressalta a necessidade de uma abordagem personalizada no rastreamento mamográfico.

O diagnóstico precoce do câncer de mama invasivo demanda a utilização conjunta de exames radiológicos, avaliação clínica e biópsia. Em paralelo, o desenvolvimento dos métodos de IA tem permitido a extração de novas informações por meio da integração de grandes bases de dados com perfis genotípicos, o que se revela essencial para a detecção da doença, a avaliação da resposta terapêutica e o monitoramento de sua progressão^{8,9}. Assim, a aplicação de algoritmos avançados na análise de mamografias tem proporcionado maior acurácia na identificação de lesões malignas, além de otimizar o fluxo de trabalho dos profissionais de saúde^{10,4}.

Ademais, a combinação dos métodos de IA com a análise radiômica permite a extração de características quantitativas que superam as limitações das avaliações qualitativas tradicionais, contribuindo para diagnósticos mais precisos, bem como para a avaliação do subtipo tumoral e da resposta terapêutica ^(1^o). A radiômica, ao decodificar informações clínicas essenciais, se revela eficaz na distinção entre lesões malignas e benignas, na classificação dos tipos e graus do câncer de mama, bem como na previsão do risco de recorrência, reforçando sua aplicabilidade na medicina personalizada ^(11,12). O fluxo dessa abordagem compreende a segmentação das regiões de interesse, a extração de descritores numéricos, tais como medidas de textura, atributos de forma e análise dos níveis de cinza, e a aplicação de análises estatísticas ou de machine learning, correlacionando essas características com dados clínicos, perfis genômicos, biomarcadores e informações histológicas ^(9,13).

O objetivo principal deste trabalho é implementar um método de caracterização de nódulos mamários por análise radiômica de mamografias.

MATERIAIS E MÉTODO

Este projeto trata-se de um estudo descritivo experimental com abordagem quantitativa, cujo objetivo é implementar um método de caracterização de nódulos mamários por análise radiômica de mamografias em linguagem de programação Python. A escolha da linguagem de programação Python justifica-se por sua simplicidade e versatilidade, características que a torna amplamente empregada nas áreas de ciência de dados e análise de imagens.

As imagens que foram utilizadas para as análises serão obtidas no “banco de imagens médicas para pesquisa do câncer” (<https://www.cancerimagingarchive.net/>). Essas imagens fazem parte de um banco de imagens mamográficas contendo nódulo e/ou microcalcificações: <https://www.cancerimagingarchive.net/collection/cbis-ddsm/>. Em conjunto com as imagens, há variáveis clínicas relacionadas a esses achados. Para este trabalho, foram consideradas as imagens com nódulos.

Para este estudo, foram analisadas 177 (cento e setenta e sete) imagens de mamografias provenientes do banco público CBIS-DDSM, contendo nódulos previamente identificados. A segmentação foi realizada com máscaras binárias para isolar as regiões de interesse, garantindo precisão na extração de características geométricas e texturais.

Na Figura 1, são apresentadas uma das mamografias disponíveis nesse banco de

imagens (a) e uma imagem mostrando a região onde está localizado o nódulo presente na mamografia (b). Esse tipo de imagem binária é chamada de máscara, na qual os pixels da região de interesse são todos iguais a 255 (branco) e os pixels restantes são iguais a 0 (preto).

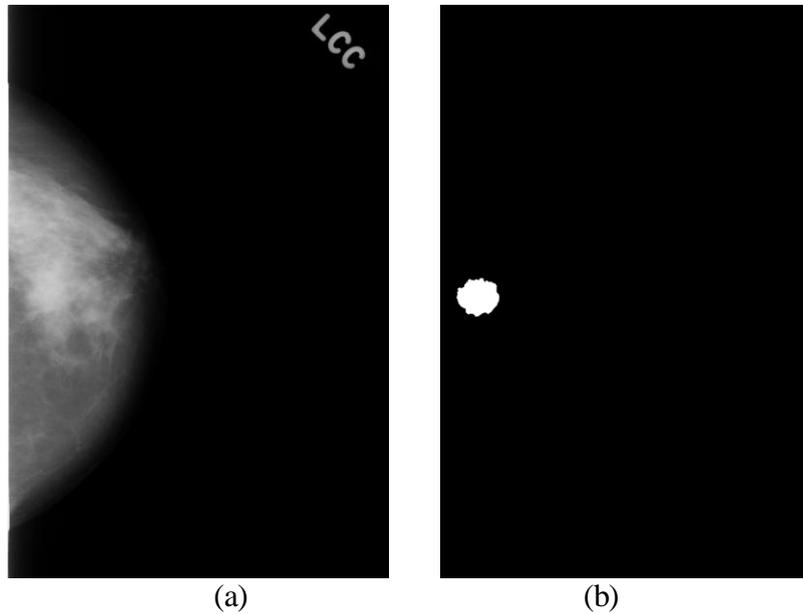


Figura 1 - Imagem mamográfica (a) e sua correspondente máscara (b) mostrando a localização de um nódulo.

Para a análise proposta, foram consideradas quatro variáveis clínicas: assessment, mass margins, mass shape e pathology. A variável assessment corresponde à avaliação baseada nos critérios do BI-RADS, com classificação de 0 a 5, indicando o nível de suspeita da lesão. A variável mass shape refere-se à forma da massa, sendo categorizada conforme os descritores do BI-RADS como redonda, oval ou irregular. A variável mass margins descreve as margens da massa, classificadas como circunscrita, obscurecida, microlobulada ou espiculada, de acordo com critérios radiológicos. Por fim, a variável pathology identifica a natureza da anormalidade, podendo ser benigna ou maligna. As classes associadas a essas variáveis foram definidas conforme padrões clínicos, garantindo que a segmentação dos dados respeite os critérios utilizados na prática médica.

A variável mass shape foi composta pelas seguintes classes: (1) irregular; (2) redonda; (3) irregular com distorção arquitetural; (4) lobulada; (5) oval; (6) linfonodo; (7) distorção arquitetural; (8) tecido mamário assimétrico; (9) irregular com tecido mamário assimétrico; (10) oval lobulada; (11) lobulada irregular; (12) lobulada linfonodo e (13) densidade assimétrica focal. Para a variável pathology, as classes foram: (0) para alterações

benignas ou benignas sem necessidade de retorno e (1) alterações malignas. A variável mass margins apresentou as seguintes categorias: (0) não aplicável (N/A); (1) espiculada; (2) circunscrita; (3) obscurecida; (4) mal definida; (5) microlobulada; (6) obscurecida circunscrita; (7) mal definida espiculada; (8) circunscrita espiculada; (9) circunscrita obscurecida mal definida; (10) obscurecida mal definida espiculada; (11) circunscrita mal definida; (12) obscurecida mal definida; (13) circunscrita microlobulada; (14) microlobulada mal definida e (15) circunscrita microlobulada mal definida. Essa organização foi feita para garantir clareza e robustez ao processo de classificação, possibilitando que as variáveis sejam compreendidas de forma estruturada no contexto do estudo.

Para realizar a análise radiômica, foi necessário segmentar as imagens de modo a detectar as regiões de interesse. A segmentação das imagens mamográficas selecionadas ocorreu por meio da multiplicação de cada mamografia pela sua correspondente máscara. Nesse caso, os valores dos pixels das regiões de interesse nas máscaras devem ser substituídos de 255 por 1. Desse modo, os pixels das regiões dos nódulos nas mamografias permanecem iguais, e os valores pixels restantes serão iguais a 0, isolando os nódulos.

As características radiômicas extraídas das imagens segmentadas incluem Dimensão Fractal (DF), área, perímetro, média, desvio-padrão, contraste, homogeneidade, energia, correlação e entropia. A análise dessas características foi realizada por meio da Matriz de Coocorrência de Níveis de Cinza (GLCM - Gray-Level Co-Occurrence Matrix) e da DF, utilizando as bibliotecas NumPy, Pydicom e Skimage.

A GLCM é empregada na análise de textura para quantificar padrões de intensidade e a relação entre os pixels da imagem, sendo essencial para identificar características relevantes em nódulos mamários. A DF complementa essa análise ao fornecer informações sobre a complexidade e a irregularidade das formas, contribuindo para uma caracterização mais precisa dos nódulos.

A análise da correlação entre as características radiômicas e as variáveis clínicas foi realizada utilizando uma análise preditiva por IA, especificamente, o algoritmo Random Forest (RF). Esse método opera por meio da construção de múltiplas árvores de decisão, em que cada uma é treinada com um subconjunto aleatório dos dados de treinamento. Durante a fase de treinamento, para cada divisão dentro de uma árvore, uma amostra aleatória de características é selecionada. Após o treinamento, as previsões geradas por todas as árvores foram combinadas, utilizando votação para as tarefas de classificação e média para as tarefas de

regressão, resultando em estimativas mais precisas e robustas. Além disso, o RF apresenta a capacidade de lidar com dados desbalanceados e variáveis com valores ausentes, tornando-se uma ferramenta versátil para diversas aplicações de aprendizado de máquina. O modelo também permite a avaliação da importância das características utilizadas, auxiliando na identificação das variáveis que exercem maior impacto sobre as previsões, bem como na interpretação dos resultados e na seleção de atributos relevantes. Graças à sua combinação de aleatoriedade e agregação de resultados, o RF é amplamente reconhecido como um dos métodos mais eficazes na análise preditiva (21).

O código para execução do RF foi escrito em Python, utilizando as bibliotecas Pandas, Numpy e Sklearn. Para garantir uma avaliação precisa do modelo, o conjunto de dados foi dividido em treino (70%) e teste (30%). O treinamento foi realizado com 100 árvores de decisão, garantindo uma abordagem robusta para a classificação. O resultado foi avaliado utilizando as seguintes métricas: acurácia, que mede a taxa de previsões corretas; precisão, que indica a proporção de previsões corretas entre todas as classificações positivas feitas pelo modelo; e importância, que quantifica a influência das características radiômicas na classificação das variáveis clínicas analisadas. Essas métricas foram armazenadas em arquivos TXT (Plain Text File) e CSV (Comma-Separated Values), permitindo sua análise no Libreoffice Calc.

Essa abordagem baseada em aprendizado de máquina contribuiu para uma compreensão mais aprofundada da influência das características radiômicas na classificação dos nódulos mamários, possibilitando a identificação de padrões relevantes de forma automatizada.

RESULTADOS E DISCUSSÃO

Este estudo avaliou de que forma as características geométricas e texturais contribuem para a classificação de nódulos mamários em mamografias, utilizando o modelo RF. A Tabela 1 apresenta a importância relativa dessas características na classificação dos nódulos mamários, de acordo com as variáveis clínicas consideradas.

Tabela 1 - Importância relativa das características geométricas e texturais na classificação dos nódulos mamários, de acordo com as variáveis clínicas.

VARIÁVEIS	ASSESSMENT	MASS MARGINS	MASS SHAPE	PATHOLOGY
Dimensão Fractal	0,16354	0,15756	0,13402	0,15203
Correlação	0,12983	0,10789	0,12966	0,11507
Desvio-Padrão	0,12239	0,14287	0,1444	0,14869
Área	0,11751	0,09965	0,11855	0,0958
Perímetro	0,11256	0,13174	0,11791	0,0945
Contraste	0,10684	0,10999	0,12041	0,11015
Entropia	0,08935	0,09391	0,08605	0,09135
Homogeneidade	0,08397	0,07873	0,07992	0,0992
Energia	0,07401	0,07766	0,06907	0,09321

A DF se destacou como uma das variáveis mais relevantes em praticamente todas as classificações, apresentando um peso aproximado de 0,16 no modelo. Esse achado está alinhado com os estudos de Raguso et al. ⁽⁵⁾ e Pezeshki et al. ⁽⁴⁾, os quais demonstram que a DF é uma métrica eficaz na quantificação da complexidade das bordas de lesões mamárias, sendo tipicamente mais elevada em tumores malignos. Conforme apontado por Pezeshki et al. ⁽⁴⁾, essa métrica permite diferenciar contornos mais suaves, característicos de lesões benignas, dos contornos irregulares, comuns em tumores malignos, o que explica sua relevância neste estudo.

Além da DF, variáveis texturais desvio-padrão, correlação e contraste, também apresentaram impacto significativo na classificação dos nódulos. Essas variáveis fornecem informações sobre a heterogeneidade e a estrutura interna das lesões, elementos essenciais na distinção entre tecidos saudáveis e patológicos. Conforme observado por Vujasinovic et al. ⁽⁶⁾, essas características possuem valor prognóstico e podem se correlacionar com o risco de metástase, reforçando sua importância para modelos preditivos. Em contrapartida, características como energia e homogeneidade tiveram impacto reduzido, possivelmente devido à limitação na captura de variações estruturais mais sutis.

Apesar da relevância das variáveis utilizadas, o modelo RF apresentou um desempenho relativamente limitado, com acurácia variando entre 22% e 39%, dependendo

das variáveis clínicas consideradas. Esse resultado pode ser atribuído, em grande parte, ao desbalanceamento das classes no conjunto de dados com variáveis de baixa importância.

No caso de Assessment, a acurácia média foi de 39%, demonstrando uma performance moderada e indicando espaço para melhorias. As classes 0, 1 e 2 não obtiveram previsões corretas, com precisão igual a 0%, o que pode ser atribuído à baixa representatividade dessas classes no conjunto de dados. Em contraste, a melhor performance foi observada na classe 4, com precisão de 50%, seguida pela classe 5, com 38%, e pela classe 3, com 18%. Fica evidente que as classes com maior número de amostras (3, 4 e 5) obtiveram desempenho superior, enquanto as classes minoritárias enfrentaram maiores dificuldades de distinção.

Em relação à variável Mass Margins, algumas classes apresentaram precisão igual a 0%, como as classes 0, 7, 9, 11, 12 e 15, novamente, refletindo o impacto negativo do desbalanceamento no desempenho do modelo. A classe 1 obteve precisão moderada, com 29%, enquanto as classes 2 e 4 apresentaram desempenhos superiores, com 33% e 44% respectivamente. A classe 3 alcançou precisão de 17%, e a classe 5, apesar de contar com poucos exemplos, obteve precisão de 20%. A matriz de confusão revelou que as classes mais frequentes foram melhor representadas, enquanto as de menor ocorrência demonstraram significativa dificuldade de classificação.

Já em relação à variável Mass Shape, os resultados indicaram um desempenho geral abaixo do esperado, com uma acurácia de apenas 22%. A classe 1 apresentou o melhor resultado, com precisão de 26%, seguida pela classe 5, com precisão de 25%. Por outro lado, algumas classes, como 2, 3, 12 e 13, não tiveram quaisquer previsões corretas, apresentando precisão igual a 0%. Estes achados evidenciam as limitações do modelo em lidar com categorias menos representadas e o impacto adverso do desbalanceamento de classes.

Por fim, na variável Pathology, o modelo apresentou acurácia de 36%, revelando certa dificuldade em prever corretamente todas as classes de Pathology. No entanto, a classe 1 obteve precisão de 62%, enquanto a classe 0 obteve destaque com precisão de 53%, demonstrando que o modelo capturou adequadamente as características desta última categoria. Ainda assim, houve uma diferença considerável entre as previsões do modelo e os valores reais das classes menos representadas, o que afetou seu desempenho geral, consequência direta do desbalanceamento no conjunto de dados.

Observou-se que as classes mais representadas, como os grupos 3, 4 e 5 do Assessment, tiveram melhor desempenho na classificação, enquanto aquelas menos frequentes apresentaram índices inferiores de precisão. Esse padrão está de acordo com as análises de Buda et al. ⁽¹⁷⁾ e Qu et al. ⁽¹⁸⁾, as quais apontam a perda de performance em modelos supervisionados diante de distribuições desiguais entre categorias.

Diante dessa limitação, torna-se essencial adotar estratégias para mitigar o desbalanceamento das classes. O uso da técnica SMOTE (Synthetic Minority Oversampling Technique), conforme descrito por Nashine et al. ⁽¹⁹⁾, pode ajudar a equilibrar o conjunto de dados ao gerar amostras sintéticas das classes minoritárias, aprimorando a capacidade de aprendizado do modelo. De forma alternativa, a atribuição de pesos diferenciados às classes ou a adoção de funções de perda como a Focal Loss ⁽²⁰⁾, pode ser uma estratégia eficaz para enfatizar exemplos mais difíceis, o que é especialmente relevante quando se busca sensibilidade em classes clínicas menos prevalentes.

O desequilíbrio amostral, a ausência de validação cruzada com bases externas e o uso de um único algoritmo de aprendizado são as principais limitações deste estudo. Pesquisas futuras podem explorar abordagens mais sofisticadas, como redes neurais convolucionais com atenção a regiões de interesse, combinadas a técnicas de balanceamento dinâmico.

CONCLUSÃO

Os resultados demonstraram, ainda, que o modelo RF, embora robusto, apresentou desempenho limitado, com acurácia entre 22% e 39%, especialmente impactado pelo desbalanceamento das classes presentes no conjunto de dados. Dessa forma, o problema de pesquisa foi parcialmente respondido, evidenciando que a análise radiômica contribui para a caracterização de nódulos mamários, mas requer aprimoramentos metodológicos para alcançar maior acurácia e confiabilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Díaz O, Rodríguez-Ruíz A, Sechopoulos I. Artificial Intelligence for breast cancer detection: Technology, challenges, and prospects. *European journal of radiology*. 2024 Jun 1;175:111457–7.
2. Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. *Seminars in Nuclear Medicine* [Internet]. 2022 Mar 24; Available from: <https://www.sciencedirect.com/science/article/pii/S0001299822000174>.
3. Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M. AI-based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation. *Radiology*. 2021 Jul;300(1):57–65.
4. Crivelli P, Ledda RE, Parascandolo N, Fara A, Soro D, Conti M. A New Challenge for Radiologists: Radiomics in Breast Cancer. *BioMed Research International*. 2018 Oct 8;2018:1–10.
5. Ferreira Junior JR, Koenigkam-Santos M, Machado CVB, Faleiros MC, Correia NSC, Cipriano FEG, et al. Radiomic analysis of lung cancer for the assessment of patient prognosis and intratumor heterogeneity. *Radiologia Brasileira*. 2021 Apr;54(2):87–93.
6. Smith RA, Saslow D, Andrews Sawyer K, Burke W, Costanza ME, Evans WP, et al. American Cancer Society Guidelines for Breast Cancer Screening: Update 2003. *CA: A Cancer Journal for Clinicians*. 2003 May 1;53(3):141–69.
7. Nelson HD, O'Meara ES, Kerlikowske K, Balch S, Miglioretti D. Factors Associated With Rates of False-Positive and False-Negative Results From Digital Mammography Screening: An Analysis of Registry Data. *Annals of Internal Medicine*. 2016 Jan 12;164(4):226.
8. Marques da Silva AM, Patrocínio AC, Schiabel H. Processamento e Análise de Imagens Médicas. *Revista Brasileira de Física Médica*. 2019 Sep 1;13(1):34.
9. Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication. *The Breast*. 2020 Feb;49:74–80.

10. Pesapane F, De Marco P, Rapino A, Lombardo E, Nicosia L, Tantrige P, et al. How Radiomics Can Improve Breast Cancer Diagnosis and Treatment. *Journal of Clinical Medicine* [Internet]. 2023 Feb 9 [cited 2023 Aug 31];12(4):1372. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9963325/>.
11. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction. *Seminars in Cancer Biology*. 2021 Jul;72:238–50.
12. Mao N, Yin P, Yuan W, Zhang K, Song X, Xing D, et al. Mammography-based radiomics for predicting the risk of breast cancer recurrence: a multicenter study. 2021 Sep 14;94(1127).
13. Santos MK, Ferreira Júnior JR, Wada DT, Tenório APM, Barbosa MHN, Marques PM de A, et al. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiologia Brasileira* [Internet]. 2019 Dec 1;52(6):387–96. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-39842019000600011&tlng=en.
14. Pezeshki H, Rastgarpour M, Sharifi A, Yazdani S. Mass classification of mammograms using fractal dimensions and statistical features. *Multidimensional Systems and Signal Processing*. 2021 Jan 2;32(2):573–605.
15. Raguso G, Ancona A, Chieppa L, L'abbate S, Pepe ML, Mangieri F, et al. Application of fractal analysis to mammography. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference* [Internet]. 2010;2010:3182–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/21096599/>.
16. Vujasinović T, Pribic J, Kanjer K, Milošević NT, Tomasevic Z, Milovanovic Z, et al. Gray-Level Co-Occurrence Matrix Texture Analysis of Breast Tumor Images in Prognosis of Distant Metastasis Risk. 2015 Jun 1;21(3):646–54.
17. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. 2018 Oct;106:249–59.
18. Qu W, Balki I, Mendez M, Valen J, Levman J, Tyrrell PN. Assessing and mitigating the effects of class imbalance in machine learning with application to X-ray imaging. *International Journal of Computer Assisted Radiology and Surgery*. 2020 Sep 23;15(12):2041–8.
19. Nashine HK, Kadelburg Z, Radenović S. Couple common fixed point theorems for W^* -compatible mappings in ordered cone metric spaces. *Applied Mathematics and*

Computation [Internet]. 2012 Jan 1;218(9):5422–32. Available from: <https://www.researchgate.net/publication/256936092> Couple common fixed point theorems for W -compatible mappings in ordered cone metric spaces.

20. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. arXiv:1708.02002 [cs] [Internet]. 2018 Feb 7; Available from: <https://arxiv.org/abs/1708.02002>.

21. Salman HA, Kalakech A, Steiti A. Random Forest Algorithm Overview. Deleted Journal. 2024 Jun 8;2024:69–79.