Revista Multidisciplinar do Nordeste Mineiro, v.12, 2024
ISSN 2178-6925

O USO DE MACHINE LEARNING NO DIAGNÓSTICO PRECOCE DE CÂNCER DE MAMA POR MEIO DE EXAMES DE IMAGEM: UMA REVISÃO BIBLIOGRÁFICA

THE USE OF MACHINE LEARNING IN THE EARLY DIAGNOSIS OF BREAST CANCER THROUGH IMAGING EXAMS: A LITERATURE REVIEW

Sandro Junior Santos da Silva

Ciência da computação, UNEMAT – Universidade do Estado de Mato Grosso.

Campus Universitário de Barra do Bugres, Brasil

E-mail: sandro.silva1@unemat.br

Rodrigo Fernando Shimazu

Ciência da computação, UNEMAT – Universidade do Estado de Mato Grosso.

Campus Universitário de Barra do Bugres, Brasil

E-mail: shimazu@unemat.br

Raquel da Silva Vieira Coelho

Ciência da computação, UNEMAT – Universidade do Estado de Mato Grosso.

E-mail: raquelcoelho@unemat.br

Resumo

Este trabalho analisa o uso de *Machine Learning* (ML) no diagnóstico precoce do câncer de mama, com foco em sua aplicação em exames de imagem. A problemática reside na necessidade de melhorar a precisão e a eficiência dos diagnósticos, uma vez que o câncer de mama, sendo uma das principais causas de morte entre mulheres, exige intervenções precoces para aumentar as taxas de sobrevivência. O objetivo central deste estudo é avaliar a eficácia dos algoritmos de ML na detecção de padrões sutis em imagens mamográficas, superando as limitações dos métodos tradicionais, como a mamografia, que apresentam altas taxas de falsos positivos. Para alcançar esse objetivo, foi realizada uma revisão bibliográfica que incluiu a análise de artigos científicos e estudos relevantes na base de dado Google Scholar. A pesquisa evidenciou que técnicas como as *Convolutional Neural Network* (CNNs) demonstram capacidade superior em identificar lesões malignas com maior precisão. Os resultados mostram que a aplicação de ML pode transformar a prática clínica, permitindo uma triagem mais eficiente e um diagnóstico mais rápido

Palavras-chave: *Machine Learning*; câncer de mama; diagnóstico precoce; exames de imagem; *Convolutional Neural Network*.

Abstract

This work analyzes the use of Machine Learning (ML) in the early diagnosis of breast cancer, focusing on its application in imaging exams. The issue lies in the need to improve the accuracy and efficiency of diagnoses, as breast cancer, being one of the leading causes of death among women, requires early interventions to increase survival rates. The central objective of this study is to evaluate the effectiveness of ML algorithms in detecting subtle patterns in mammographic images, overcoming the limitations of traditional methods, such as mammography, which have high false-positive rates. To achieve this objective, a literature review was conducted, including the analysis of scientific articles and relevant studies in the Google Scholar database. The research highlighted that techniques such as Convolutional Neural Networks (CNNs) demonstrate superior ability in identifying malignant lesions with greater accuracy. The results show that the application of ML can transform clinical practice, enabling more efficient screening and faster diagnoses.

Keywords: Machine Learning; breast cancer; early diagnosis; imaging exams; Convolutional Neural Networks.

1. Introdução

O câncer de mama é uma das principais preocupações de saúde pública global, sendo o tipo de câncer mais frequentemente diagnosticado entre mulheres e a segunda principal causa de morte por câncer no sexo feminino (IARC, 2020). Em 2020, estima-se que mais de 2,3 milhões de novos casos foram registrados em todo o mundo, representando cerca de 24,5% de todos os casos de câncer diagnosticados em mulheres. A taxa de mortalidade também é significativa, com aproximadamente 685.000 mortes atribuídas a essa condição no mesmo ano (IARC, 2020).

A detecção precoce do câncer de mama é crucial para melhorar a eficácia dos tratamentos e as taxas de sobrevivência. Embora a mamografia continue sendo uma ferramenta eficaz, ela enfrenta desafios em termos de precisão e tempo de análise. Nesse contexto, técnicas de ML, especialmente as *Convolutional Neural Networks* (CNNs), estão se destacando como uma solução promissora. Estudos indicam que algoritmos baseados em CNNs podem identificar padrões sutis em imagens médicas que muitas vezes são difíceis de detectar com métodos convencionais. Essas técnicas de aprendizado profundo são capazes de processar grandes volumes de dados e melhorar a precisão na detecção de câncer de mama, superando muitas vezes a eficácia dos métodos tradicionais de diagnóstico (ABDELHAFIZ *et al.,* 2019). A combinação de imagens obtidas por mamografia com algoritmos de ML pode resultar em uma análise mais rápida e precisa, possibilitando intervenções mais precoces e melhores resultados para os pacientes.

O uso de ML não apenas pode melhorar a precisão dos diagnósticos, mas também pode acelerar o processo de análise, permitindo uma triagem mais rápida e uma resposta mais eficiente. Isso é crucial em um cenário onde o tempo é um fator decisivo para o sucesso do tratamento. Adicionalmente, a automação proporcionada pelo ML pode reduzir a carga de trabalho dos radiologistas e aumentar a capacidade de diagnóstico, especialmente em contextos com recursos limitados.

Portanto, este projeto visa investigar a aplicação de técnicas de ML no diagnóstico precoce do câncer de mama, com o objetivo de demonstrar como essas tecnologias podem ser integradas aos métodos atuais para oferecer uma solução mais eficaz e eficiente no combate a essa doença.

2. Revisão da Literatura

2.1 Câncer de Mama e a Importância da Detecção Precoce

O câncer de mama é um dos principais tipos de câncer que afetam as mulheres, e a detecção precoce é essencial para o sucesso do tratamento e para a redução das taxas de mortalidade. Segundo o Instituto Nacional de Câncer (INCA), "quanto mais cedo o câncer de mama é diagnosticado, maiores são as chances de cura e menores as taxas de mortalidade associadas" (INCA, 2023). Identificar o câncer nos estágios iniciais permite que tratamentos menos agressivos sejam eficazes, o que impacta positivamente na qualidade de vida das pacientes.

A mamografia é o principal exame recomendado para o rastreamento precoce do câncer de mama, especialmente em mulheres a partir dos 40 anos. De acordo com a Sociedade Brasileira de Mastologia (SBM), a mamografia pode reduzir a mortalidade pela doença em até 30% entre as mulheres que realizam o exame periodicamente (SBM, 2023). Esse exame consegue identificar lesões antes que se tornem palpáveis, possibilitando intervenções mais rápidas e com maiores chances de sucesso.

Além da mamografia, a ultrassonografia é frequentemente utilizada como exame complementar, especialmente em mulheres com mamas densas, nas quais a mamografia pode apresentar limitações. Segundo estudos, a combinação de

métodos de imagem para a detecção do câncer de mama aumenta significativamente a sensibilidade da detecção em comparação ao uso de um único método (Berg *et al.,* 2008). Isso destaca a importância de uma abordagem personalizada para o diagnóstico precoce.

Campanhas de conscientização, como o outubro Rosa, também são fundamentais para incentivar a prática do rastreamento. Essas iniciativas promovem a informação sobre a importância da detecção precoce e estimulam as mulheres a buscarem exames regulares. Como observa a Organização Mundial da Saúde (OMS), “a educação e a conscientização populacional são passos cruciais para o aumento do diagnóstico precoce, sobretudo em países em desenvolvimento, onde o acesso aos serviços de saúde pode ser limitado” (OMS, 2022). Assim, a detecção precoce do câncer de mama não apenas salva vidas, mas também diminui a necessidade de tratamentos agressivos, o que representa um benefício para as pacientes e para o sistema de saúde como um todo.

2.2 Métodos Tradicionais de Diagnóstico

Os métodos tradicionais de diagnóstico do câncer de mama desempenham um papel crucial no rastreamento e tratamento eficaz da doença, especialmente por sua capacidade de identificar tumores em estágios iniciais. Entre esses métodos, a mamografia é a principal ferramenta de rastreamento e o exame mais amplamente utilizado para a detecção precoce de lesões na mama. A mamografia é capaz de capturar imagens detalhadas das mamas usando baixa dose de raios-X, identificando calcificações e massas suspeitas que podem indicar a presença de câncer. Esse exame é recomendado, principalmente, para mulheres acima dos 40 anos e tem demonstrado eficácia na redução da mortalidade associada ao câncer de mama ao detectar lesões antes de se tornarem palpáveis (INCA, 2023).

Apesar de sua eficácia, a mamografia pode apresentar limitações, como a baixa sensibilidade em mamas densas e a alta taxa de falsos positivos, o que pode levar a tratamentos desnecessários e ansiedade para as pacientes. Estudos indicam que a sensibilidade da mamografia é reduzida em mulheres com mamas densas, onde a presença de tecido denso pode obscurecer as imagens,

dificultando a identificação de tumores e aumentando a chance de resultados falso-positivos (Boyd *et al.,* 2007). Nessas situações, exames complementares, como a ultrassonografia, são recomendados para uma avaliação mais precisa. A ultrassonografia mamária, utilizada como exame complementar, é especialmente eficaz em mulheres com mamas densas. Esse método utiliza ondas de ultrassom para criar imagens dos tecidos da mama, ajudando a diferenciar entre cistos e tumores sólidos. Segundo Berg *et al.* (2008), a combinação da ultrassonografia com a mamografia pode aumentar a sensibilidade da detecção do câncer, especialmente em mulheres com risco elevado da doença.

Outro método tradicional é a Ressonância Magnética (RM), indicada especialmente para mulheres com alto risco de câncer de mama, Mann *et al*. (2019), destaca os avanços no uso da RM para a detecção precoce do câncer de mama, com atenção especial a mulheres portadoras de mutações genéticas BRCA1/BRCA2[1]. Essas mutações estão associadas a um risco significativamente elevado de desenvolver câncer de mama ao longo da vida, tornando essencial o monitoramento rigoroso dessas pacientes. A RM tem se mostrado uma ferramenta crítica para a triagem desse grupo, devido à sua alta sensibilidade na detecção de lesões malignas, especialmente em pacientes com mamas densas, onde outras modalidades, como a mamografia, apresentam limitações. O estudo enfatiza que, em mulheres com BRCA1/BRCA2, a RM é capaz de identificar tumores em estágios iniciais, quando ainda são mais tratáveis, reduzindo significativamente a mortalidade associada (Mann *et al.,* 2019). Além disso, a RM auxilia no planejamento cirúrgico e na avaliação da extensão da doença em pacientes já diagnosticadas, oferecendo informações detalhadas que contribuem para decisões terapêuticas mais precisas. Apesar de sua eficácia, os autores também mencionam desafios, como os custos elevados, taxa de falso positivo e a necessidade de interpretação especializada.

[1] BRCA é a abreviatura de *Breast Cancer* gene 1 ou 2, são genes que suprimem os tumores, produzindo proteínas que reparam os danos no DNA e mantêm o crescimento celular estável. As mutações genéticas podem ser herdadas ou ocorrer por acaso, que é categorizado como uma mutação somática. Quando um dos genes é mutado, sua função principal é alterada (ou interrompida), o que pode levar a danos no DNA não reparados ou outras alterações, aumentando o risco de desenvolver câncer.

2.3 *Machine Learning* na Análise de Imagens de Câncer de Mama

A mamografia é o principal método de rastreamento para o câncer de mama, mas apresenta desafios, especialmente em mulheres com mamas densas, onde a detecção de lesões pode ser mais difícil devido à baixa sensibilidade do exame em tais casos (Lehman *et al.*, 2015). Embora o uso de técnicas de ML tenha mostrado avanços significativos na melhoria da acurácia diagnóstica, o *Deep Learning* (DL) tem se destacado como uma subárea de ML com grande potencial para superar essas limitações, especialmente em tarefas complexas como a análise de imagens mamográficas.

Em estudos sobre a aplicação de DL para análise de imagens mamográficas, como o de McKinney *et al.* (2020), o uso de CNNs mostrou-se promissor na superação das dificuldades encontradas em mamografias de mulheres com mamas densas. McKinney *et al.* (2020) desenvolveram um modelo de DL treinado com grandes volumes de dados de mamografias, alcançando uma taxa de precisão que, em certos casos, superava a de radiologistas experientes. Esse avanço sugere que o DL pode atuar como uma "segunda opinião," auxiliando radiologistas na identificação de lesões sutis que poderiam passar despercebidas.

Além disso, os sistemas de DL têm a possibilidade de se tornar mais eficazes à medida que são continuamente treinados com novos dados. A integração desses modelos em fluxos de trabalho clínicos poderia não apenas aumentar a taxa de detecção, mas também reduzir a carga sobre os radiologistas, permitindo que eles se concentrem em casos mais complexos e potencialmente graves. No entanto, é importante notar que técnicas tradicionais de aprendizado de máquina, como *Support Vector Machines* (SVM) e *Linear Discriminant Analysis* (LDA), continuam a ser muito úteis, especialmente quando há dados limitados ou quando a interpretabilidade do modelo é crucial.

No estudo de Egwom *et al.* (2022), foi proposto um modelo híbrido que combina LDA para extração de características e SVM para classificação de câncer de mama. Este modelo foi desenvolvido para melhorar a precisão da previsão, especialmente em cenários com dados limitados. A pesquisa destacou a importância da extração de características, usando LDA para gerar um novo

conjunto de dados a partir das características extraídas, o que aumentou o desempenho do modelo em comparação com outras abordagens existentes. Além disso, os resultados mostraram que o uso de medianas para calcular valores ausentes na base de dados também ajudou a melhorar a precisão da previsão. A escolha de usar medianas, como mencionado no estudo de Egwom *et al.* (2022), indica que essa abordagem pode ser benéfica para tratar dados faltantes em previsões de câncer de mama, contribuindo para um desempenho mais robusto dos modelos de classificação.

No campo do diagnóstico de câncer, as redes de DL têm demonstrado um potencial significativo para melhorar a acurácia e a eficiência dos processos de detecção precoce, especialmente no câncer de mama. Diversos estudos têm se concentrado na aplicação dessas técnicas em imagens médicas para auxiliar na análise de mamografias, ultrassons e ressonâncias magnéticas. O estudo de Lehman *et al.* (2022) discute como o DL pode ser integrado aos sistemas clínicos para ajudar na detecção precoce de câncer de mama, melhorando a precisão dos diagnósticos e reduzindo as taxas de falsos negativos. Eles destacam que, quando combinadas com a interpretação humana, essas redes neurais podem identificar padrões sutis que podem passar despercebidos por médicos, oferecendo uma ferramenta poderosa para a prática clínica.

O estudo realizado por Arefan *et al.* (2019) aborda o impacto do DL na melhoria da visualização de imagens médicas. O estudo revela que, por meio de redes neurais profundas treinadas em grandes bases de dados, é possível detectar com maior precisão tumores e outras anomalias nas imagens, reduzindo o tempo de análise e melhorando a acurácia no diagnóstico. A eficiência desses modelos pode acelerar significativamente o processo de avaliação, permitindo que os médicos se concentrem em aspectos mais críticos do tratamento.

Yala *et al.* (2019), em sua pesquisa, exploraram o uso do DL para a detecção de câncer de mama, mostrando como essas técnicas podem ser aplicadas na prática clínica. Eles argumentam que o DL tem a capacidade de transformar os cuidados com a saúde, automatizando e otimizando o processo de diagnóstico, mas ressaltam que a colaboração com profissionais médicos continua sendo essencial para validar e interpretar os resultados gerados pelos modelos.

De acordo com Houssein *et al.* (2021), o uso de técnicas de ML e DL na área de diagnóstico médico tem mostrado um grande potencial, especialmente quando aplicado à detecção precoce de câncer de mama. A pesquisa de Houssein *et al.* (2021), destaca como essas tecnologias, quando treinadas com conjuntos de dados de imagens médicas, podem não apenas melhorar a precisão do diagnóstico, mas também reduzir o tempo necessário para a análise das imagens, o que é crucial para um diagnóstico mais rápido e preciso. O estudo explora várias abordagens de redes neurais, como CNNs e técnicas de aprendizado supervisionado, demonstrando a eficácia de tais métodos em comparação com métodos tradicionais.

Além disso, Houssein *et al.* (2021) complementam esses achados, explorando como a combinação de diferentes modelos de aprendizado de máquina pode levar a uma melhoria significativa nos resultados de classificação de imagens. A pesquisa destaca que a integração de métodos como redes neurais profundas e algoritmos genéticos pode resultar em um modelo mais robusto, capaz de identificar padrões complexos em imagens de mamografias. A análise multifacetada apresentada por Houssein *et al.* (2021) também discute como a automação do processo de diagnóstico pode aliviar a carga dos profissionais de saúde, permitindo-lhes focar em casos mais complexos e urgentes. O estudo aborda os desafios que ainda existem para a aplicação em larga escala dessas técnicas, incluindo a necessidade de grandes volumes de dados e a superação de questões éticas relacionadas ao uso de inteligência artificial na saúde. Esse trabalho coletivo reflete os avanços significativos na área, mas também sublinha a necessidade de continuar o desenvolvimento e a validação dessas técnicas para garantir sua implementação bem-sucedida e ética no sistema de saúde global.

Segundo Jasti *et al.* (2022), o uso dessas técnicas pode superar limitações humanas, como a subjetividade na análise de imagens, aumentando a precisão na identificação de anomalias. Os métodos propostos utilizam dados de mamografias, ultrassonografias e ressonâncias magnéticas, analisando padrões complexos que seriam difíceis de detectar manualmente. Além disso, a aplicação dessas tecnologias não se limita à detecção de câncer, mas também contribui para a classificação de lesões como benignas ou malignas, reduzindo significativamente

os falsos positivos e falsos negativos. Os algoritmos de ML, combinados com ferramentas de processamento de imagem, podem analisar imagens médicas de maneira eficiente, oferecendo diagnósticos mais rápidos e precisos em relação aos métodos tradicionais.

Outro ponto abordado por Jasti *et al.* (2022) é a integração dessas ferramentas em sistemas de suporte à decisão clínica, ajudando os médicos a tomarem decisões mais precisas, oferecendo análises detalhadas que facilitam o diagnóstico. Embora promissoras, as técnicas baseadas em ML e DL apresentam desafios, como a necessidade de grandes conjuntos de dados para treinamento, o que pode limitar sua implementação em larga escala. As técnicas computacionais baseadas em ML e processamento de imagem têm um impacto significativo no diagnóstico do câncer de mama, aumentando a precisão e a velocidade da análise, permitindo intervenções mais rápidas e eficazes. No entanto, a implementação dessas tecnologias requer atenção à qualidade dos dados e à validação clínica rigorosa, pontos essenciais para garantir seu sucesso em cenários médicos reais.

2.4 Aplicações Práticas e Benefícios

A crescente aplicação de técnicas de ML na análise de imagens de câncer de mama tem promovido avanços significativos na detecção precoce e no diagnóstico da doença. Esses métodos inovadores não apenas melhoram a precisão dos diagnósticos, mas também contribuem para a redução de falsos positivos e otimizam o fluxo de trabalho nos serviços de saúde, transformando a maneira como os profissionais da saúde abordam o câncer de mama.

Para mulheres em grupos de alto risco, o ML oferece uma abordagem personalizada de monitoramento. Um estudo conduzido por Yala *et al.* (2019) revelou que o uso de ML pode prever com acurácia o risco de câncer com base em uma combinação de dados demográficos e imagens de mamografia. O modelo desenvolvido obteve uma taxa de acurácia de 81%, facilitando a criação de planos de monitoramento específicos para cada paciente, o que pode envolver exames mais frequentes e técnicas de imagem complementares, como ultrassonografia ou ressonância magnética (Yala *et al.,* 2019).

Além dos benefícios operacionais, a aplicação de ML na triagem e

diagnóstico do câncer de mama pode também impactar positivamente a experiência e qualidade de vida das pacientes. Com algoritmos de DL, a redução da taxa de falsos positivos significa menos mulheres passando por biópsias e procedimentos invasivos desnecessários. Um estudo de McKinney *et al.* (2020), demonstrou que o uso de ML em análises de mamografias pode reduzir em até 5,7% o número de falsos positivos, diminuindo a ansiedade das pacientes e os custos associados a tratamentos não necessários.

Chaurasia e Pal (2020) discutem como os avanços em técnicas de ML, como *Artificial Neural Networks* (ANN), SVM e *Decision Trees*, podem melhorar a precisão dos diagnósticos. Eles enfatizam que, ao aplicar esses modelos a grandes volumes de dados, como os disponíveis em bases como o *Wisconsin Breast Cancer Dataset* (WBCD), é possível identificar padrões complexos que podem não ser detectados por métodos tradicionais. O uso de algoritmos de ML oferece uma solução potencial para reduzir erros de diagnóstico e melhorar a detecção precoce, permitindo uma resposta mais rápida e eficiente ao câncer de mama, o que é crucial para a redução das taxas de mortalidade associadas à doença.

O uso de ML no diagnóstico precoce do câncer de mama por meio de exames de imagem representa uma verdadeira revolução na oncologia. As aplicações práticas dessa tecnologia não só aumentam a precisão diagnóstica, mas também oferecem a possibilidade de uma detecção mais precoce e, consequentemente, melhores taxas de sobrevivência. No entanto, os desafios associados à implementação e aceitação dessa tecnologia ainda precisam ser superados. A tendência é que, no futuro, os sistemas baseados em ML sejam uma ferramenta essencial para os profissionais de saúde, colaborando para um diagnóstico cada vez mais preciso e acessível.

2.5 Desafios e Considerações Éticas

Segundo Carter *et al., (*2020), embora frequentemente considerada neutra, a Inteligência Artificial (IA) inevitavelmente codifica valores, seja de forma explícita ou implícita. Podemos citar, algoritmos de DL e ML em programas de mamografia podem priorizar a redução de falsos negativos, dependendo das características do tecido mamário ou de fatores sociodemográficos. Isso levanta

questões éticas, já que esses valores nem sempre são transparentes. A problemática da "caixa-preta" (falta de explicabilidade) é central, visto que algoritmos mais precisos tendem a ser menos explicáveis, contrastando com a responsabilidade moral, legal e profissional exigida de médicos (Carter *et al.,* 2020).

Além disso, os pesquisadores discutem que o desenvolvimento de IA para melhorar os desfechos clínicos requer foco em reduzir morbidade e mortalidade, indo além da mera precisão diagnóstica. Porém, a falta de dados de treinamento sobre casos de sobrediagnóstico limita a eficácia dessa abordagem. Algoritmos futuros poderiam discriminar entre cânceres clinicamente significativos e aqueles que não progrediriam, mas os protocolos atuais priorizam sensibilidade e especificidade em critérios tradicionais (Carter *et al.,* 2020).

Outro ponto relevante é o viés nos dados de treinamento. Sistemas de IA refletem os dados nos quais são treinados, podendo perpetuar desigualdades já presentes na medicina. Por exemplo, subgrupos populacionais marginalizados frequentemente têm menor representação em bases de dados, o que resulta em resultados enviesados. Iniciativas como as do laboratório de IA do *Massachusetts Institute of Technology* (MIT) tentam mitigar isso, desenvolvendo algoritmos que funcionem de forma equitativa em diferentes grupos étnicos (Carter *et al.,* 2020).

Questões sobre propriedade de dados e consentimento também são críticas, especialmente com a entrada de grandes empresas tecnológicas na área da saúde. Casos como o da liberação de dados de saúde na ltália para o IBM *Watson* ilustram como o compartilhamento indiscriminado de informações pode comprometer a privacidade dos pacientes. Modelos alternativos, como o uso de contêineres de software que protegem os dados enquanto permitem treinamento de IA, são sugeridos como soluções potenciais (Carter *et al.,* 2020).

No aspecto legal, os autores destacam que a ausência de regulamentações específicas para IA cria um "vácuo regulatório". É necessário estabelecer diretrizes que atribuam responsabilidades claras, sejam elas para médicos, desenvolvedores ou instituições, antes da implementação de IA em grande escala. Isso inclui equilibrar a inovação com a proteção de pacientes vulneráveis (Carter *et al.,* 2020).

Embora a IA possa complementar o trabalho de médicos, sua falta de explicabilidade pode limitar a confiança dos pacientes e dificultar a responsabilização em casos de erros. Além disso, a automação pode levar à perda de habilidades clínicas humanas e ao viés de aceitação das decisões automáticas, exigindo estratégias para mitigar esses efeitos (Carter *et al.,* 2020).

2.6 Trabalhos relacionados

O artigo destaca a importância crescente das técnicas de aprendizado de máquina na análise de imagens de câncer de mama e a necessidade de um maior entendimento dessas tecnologias e dos desafios que elas envolvem. Este trabalho explora uma revisão da literatura existente sobre o uso de ML e DL no diagnóstico precoce do câncer de mama, destacando os avanços, as aplicações práticas e as implicações éticas dessas abordagens. Conforme discutido ao longo do artigo, as técnicas de aprendizado de máquina, especialmente as redes neurais profundas, têm mostrado um grande potencial para melhorar a acurácia diagnóstica, ajudando a superar limitações de métodos tradicionais, como a análise de mamografias em mulheres com mamas densas. Assim, o trabalho visa realizar uma pesquisa de revisão literária sobre os principais estudos e contribuições na área, além de identificar os benefícios, a eficácia e os desafios associados ao uso dessas tecnologias no contexto clínico.

Com isso, nesta seção é apresentado o método utilizado para a realização desta pesquisa, foi utilizada a plataforma Google Scholar. A *string* de busca utilizada foi "*the use of Machine Learning in the early diagnosis of breast cancer through imaging exams"* (o uso de Machine Learning no diagnóstico precoce do câncer de mama por meio de exames de imagem), e foi empregada de forma completa, sem fragmentação. Esta plataforma foi escolhida devido à sua capacidade de filtrar publicações científicas de diversas bases de dados, proporcionando resultados abrangentes e relevantes para a pesquisa em questão, com isso optou-se por trabalhos publicados entre os anos de 2018 e 2024 que contenham relação com o assunto estudado, o escopo desta pesquisa é limitado a trabalhos científicos e acadêmicos e exclui interesses comerciais. Foram selecionados 80 artigos, no entanto, vale ressaltar que alguns desses artigos eram

repetidos ou continham informações que não estavam relacionados com a pesquisa. Esses artigos foram excluídos da seleção para garantir a qualidade e a precisão dos dados. Após essa etapa, continuou com a seleção dos artigos. A mesma foi feita com a análise através dos títulos e do resumo dos mesmos em relação ao foco da pesquisa. Os artigos menos relevantes para pesquisa foram eliminados, isso resultou em um total de 7 artigos selecionados que atendiam aos critérios estabelecidos para a pesquisa, apresentado conforme Tabela 1.

Tabela 1 – Identificação dos Estudos Selecionados.

Título	**Autor**	**Ano**	**Técnica Utilizada**	**Tecnologia Usada**
Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm	Heidari *et al.*	2018	*Locality Preserving Projection* (LPP), redução de dimensionalidade, classificador de ML	Algoritmos de ML, LPP
Prediction of Breast Cancer Using Supervised Machine Learning Techniques	Shravya *et al.*	2019	Regressão Logística, KNN, SVM, PCA para redução de dimensionalidade	Python, PCA, algoritmos de ML
Machine learning in breast MRI	Reig *et al.*	2019	Aprendizado supervisionado e não supervisionado, CNNs, radiômica	CNNs, radiômica, algoritmos de ML e DL
Artificial intelligence and machine learning in cancer imaging	Koh *et al.*	2022	CNNs, radiômica, análise de padrões sutis em imagens	Algoritmos de ML, radiômica
Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis	Jasti *et al.*	2022	SVM, KNN, *Random Forest*, *Naïve Bayes*, filtro de média geométrica, algoritmo Relief para seleção de características	Processamento de imagens, algoritmos de ML
Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis	Yue *et al.*	2018	*Particle Swarm Optimization* (PSO), ANN	Algoritmos de ML, banco de dados WBCD
Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis	Naji *et al.*	2021	SVM, *Random Forest*, Regressão Logística, *Decision Tree* (C4.5), KNN	Algoritmos de ML, conjunto de dados *Breast Cancer Wisconsin diagnostic*

Fonte: Autoria Própria

A eficácia dos algoritmos foi avaliada por meio de diversas métricas e métodos específicos em diferentes estudos. Um dos principais métodos utilizados foi a análise da precisão dos modelos, frequentemente comparando diferentes

algoritmos em um mesmo conjunto de dados. Além disso, métricas como sensibilidade, especificidade, AUC, e precisão foram amplamente utilizadas para medir o desempenho dos modelos na classificação e previsão de riscos. Cada trabalho utilizou abordagens distintas para medir o desempenho dos modelos de ML aplicados ao diagnóstico e prognóstico do câncer de mama. Esses trabalhos foram selecionados por apresentarem resultados consistentes, métricas robustas e aplicações práticas que contribuem para o avanço da detecção precoce e do tratamento do câncer de mama. Eles demonstram como a integração de ML com técnicas complementares podem transformar a prática clínica, promovendo soluções mais eficientes e precisas.

Segundo Heidari *et al.* (2018), a aplicação do algoritmo *Locality Preserving Projection* (LPP) em um modelo de ML mostrou-se eficaz para melhorar a previsão do risco de câncer de mama a curto prazo. O estudo utilizou um conjunto de dados de 500 mamografias negativas, dividindo os casos em dois grupos pareados por idade: 250 mulheres de alto risco, nas quais o câncer foi detectado na triagem subsequente, e 250 de baixo risco, que permaneceram negativas. Inicialmente, foram extraídas 44 características relacionadas à assimetria na densidade do tecido mamário entre as mamas. Em seguida, um classificador de ML integrado ao algoritmo LPP foi desenvolvido, reduzindo a dimensionalidade para 4 características principais. Essa abordagem resultou em um aumento de 9,7% na precisão da previsão de risco. Além disso, os autores destacam que a aplicação do algoritmo LPP reduziu efetivamente a dimensionalidade dos recursos e resultou em um desempenho mais alto e potencialmente mais robusto na previsão do risco de câncer de mama a curto prazo. Em resumo esses resultados indicam o potencial dessa técnica para melhorar a estratificação de risco e apoiar a detecção precoce do câncer de mama.

A pesquisa de Shravya *et al,* (2019) exploraram o uso de técnicas supervisionadas de ML na predição de câncer de mama, focando na classificação de tumores benignos e malignos. Utilizando os algoritmos de *Logistic regression*, *K-Nearest Neighbors* (KNN) e SVM, o estudo analisou dados do repositório *University of California, Irvine* (UCI), implementando as soluções em Python com uso de *Principal Component Analysis* (PCA) para redução de dimensionalidade. Os

resultados apontaram que o SVM obteve a maior precisão de 92,7%, destacando-se como o modelo mais eficiente para análises preditivas. A pesquisa conclui que métodos baseados em ML podem auxiliar no diagnóstico precoce e na melhoria das taxas de sobrevivência ao câncer de mama.

Já Reig *et al.* (2019) explora o uso de ML em imagens de RM para melhorar o diagnóstico do câncer de mama. Ele destaca o uso de aprendizado supervisionado e não supervisionado, além das CNNs e DL que têm avançado a segmentação e a análise de imagens. A radiômica, que combina dados quantitativos de imagens com informações clínicas, também é mencionada como uma ferramenta promissora para prever o risco de câncer. O artigo conclui que o avanço em hardware, algoritmos e dados foi essencial para o crescimento exponencial dessas tecnologias, com potencial para transformar a prática clínica, melhorando a detecção precoce e o tratamento do câncer de mama.

No entanto a pesquisa de Koh *et al.* (2022) destaca o uso de ML na análise de imagens relacionadas ao câncer de mama, apontando como essas tecnologias podem transformar o diagnóstico e o tratamento. Técnicas de ML, como CNNs, são mencionadas por sua capacidade de identificar padrões sutis em imagens médicas, auxiliando na detecção precoce e na avaliação do prognóstico do câncer de mama. Um dos principais benefícios do ML é sua habilidade de processar grandes volumes de dados com rapidez e precisão, reduzindo o tempo e a carga de trabalho dos radiologistas. No entanto, os autores alertam para desafios, como a escassez de dados de alta qualidade e a necessidade de validação rigorosa para integrar essas ferramentas na prática clínica. O uso de radiômica, que combina dados de imagens com análises quantitativas, também é citado como uma abordagem promissora no câncer de mama, permitindo prever com maior precisão a resposta ao tratamento e melhorar o planejamento do tratamento.

No estudo Jasti *et al.* (2022) apresenta uma abordagem evolutiva para a detecção e classificação do câncer de mama, utilizando técnicas de ML e processamento de imagens. A proposta combina pré-processamento de imagens, extração e seleção de características com algoritmos de aprendizado de máquina, como SVM, KNN, *Random Forest* e *Naïve Bayes*, para identificar e classificar doenças, incluindo o câncer de mama. O modelo usa o filtro de média geométrica

para melhorar a qualidade das imagens e o algoritmo *Relief* para seleção de características. A abordagem foi testada usando o conjunto de dados *Mammographic Image Analysis Society* (MIAS), demonstrando a eficácia do método na detecção precisa do câncer de mama, destacando-se pela sua capacidade de classificar diferentes tipos de doenças mamárias. Os autores destacam que a detecção precoce, através de mamografia e análise computacional, pode melhorar significativamente as chances de recuperação total.

O trabalho desenvolvido por Yue *et al.* (2018) explora diferentes abordagens de ML aplicadas ao diagnóstico e prognóstico do câncer de mama, utilizando o banco de dados WBCD. Os autores destacam a eficácia das técnicas de ML na melhoria da precisão de classificação e previsão. Embora os algoritmos tenham alcançado altas taxas de precisão, como a *Particle Swarm Optimization* (PSO) e ANN, os autores sugerem que o desenvolvimento de vários algoritmos aprimorados ainda é necessário. Os autores enfatizam o potencial do ML para transformar o diagnóstico e o prognóstico do câncer de mama, fornecendo soluções mais precisas, eficientes e escaláveis do que os métodos tradicionais, sugerindo melhorias adicionais aos algoritmos de ML para abordar as complexidades dos dados médicos.

O estudo realizado por Naji *et al*. (2021) demonstra que os algoritmos de ML têm grande utilidade na previsão e diagnóstico do câncer de mama. A pesquisa avaliou cinco modelos: SVM, *Random Forest*, *Logistic Regression*, *Decision Tree (C4.5)* e KNN, utilizando o conjunto de dados *Breast Cancer Wisconsin Diagnostic*. Os resultados mostram que o SVM foi o mais eficiente, com acurácia de 97,2%, superando os outros algoritmos em métricas como precisão (*precision*), sensibilidade (*sensitivity*) e *Area Under the ROC Curve* (AUC). Isso demonstra que as técnicas de ML são ferramentas valiosas para diagnósticos precoces, ajudando a melhorar as opções de tratamento e as taxas de sobrevivência de pacientes. Apesar de promissor, o estudo reconhece a limitação dos resultados ao conjunto de dados utilizado, recomendando testes futuros em bases mais diversas para validação e refinamento das análises.

3. Considerações Finais

Conclui-se que, apesar dos benefícios significativos do ML na análise de imagens médicas, desafios éticos e técnicos precisam ser abordados, como a privacidade dos dados dos pacientes e a validação dos algoritmos. A pesquisa reforça a importância de um compromisso contínuo com a inovação tecnológica na área da saúde e a necessidade de estratégias robustas para a implementação de práticas de diagnóstico que integrem essas novas ferramentas.

O câncer de mama representa um grande desafio na saúde pública devido à sua elevada incidência e impacto na mortalidade feminina. A detecção precoce continua sendo um pilar essencial no combate à doença, aumentando significativamente as chances de sucesso no tratamento e a qualidade de vida das pacientes. Embora os métodos tradicionais, como mamografia, ultrassonografia e ressonância magnética, desempenhem papéis cruciais, suas limitações destacam a necessidade de avanços tecnológicos.

A integração de técnicas de ML, especialmente as CNN, surge como uma solução promissora para superar os desafios diagnósticos. Essas ferramentas tecnológicas oferecem maior precisão, eficiência e capacidade de análise, complementando os métodos convencionais e reduzindo erros, como falsos positivos e negativos. A automação proporcionada pelo aprendizado de máquina também contribui para uma melhor utilização dos recursos médicos, aliviando a carga de trabalho dos profissionais de saúde e permitindo intervenções mais rápidas.

Esses avanços tecnológicos não apenas aprimoram a acurácia diagnóstica, mas também criam oportunidades para abordagens mais personalizadas, acessíveis e eficazes no rastreamento do câncer de mama. No entanto, a implementação dessas inovações exige atenção a fatores como a qualidade dos dados, validação clínica rigorosa e integração harmônica com os fluxos de trabalho existentes. Esses estudos indicam que tanto o ML tradicional quanto o DL estão se tornando ferramentas essenciais para a análise de imagens médicas, com a capacidade de melhorar os diagnósticos de câncer de mama.

O uso de ML no diagnóstico precoce do câncer de mama por meio de exames de imagem apresenta diversos benefícios, como a melhoria na precisão da detecção de sinais iniciais e a agilidade na análise de grandes volumes de dados. A

tecnologia pode ajudar a identificar padrões sutis que os radiologistas podem não perceber, reduzindo a chance de erros e aumentando a confiabilidade do diagnóstico. Além disso, modelos de ML podem ser continuamente aprimorados com novos dados, tornando os diagnósticos mais precisos ao longo do tempo.

Porém, há desafios, como a dependência da qualidade dos dados. Se as imagens forem de baixa qualidade, o modelo pode cometer erros. A implementação de ML também exige investimentos em infraestrutura e dados rotulados, o que pode ser um obstáculo. Outro problema é a dificuldade de interpretar as decisões de alguns modelos, o que pode gerar desconfiança entre os profissionais de saúde. Além disso, a integração da tecnologia aos sistemas médicos existentes pode ser complexa e enfrentar resistência. No entanto, todos os autores concordam que, embora essas tecnologias mostrem grande promessa, elas devem ser usadas como um complemento ao julgamento clínico humano, nunca como um substituto completo.

Referências

Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). **Deep convolutional neural networks for mammography: advances, challenges and applications**. *BMC Bioinformatics*, v. 20, n. 1, p. 210, 2019. Disponível em: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2823-4. Acesso em: 03 de nov. de 2024.

Arefan, D., Mohamed, A. A., Berg, W. A., Zuley, M. L., Sumkin, J. H., & Wu, S. (2019). **Deep learning in medical imaging: A review of algorithms and applications in breast cancer diagnosis**. *Medical Physics*, 46(10), 6172-6186. https://doi.org/10.1002/mp.13886. Acesso em: 12 de nov. de 2024.

Berg, W. A., Blume, J. D., Cormack, J. B., Mendelson, E. B., Lehrer, D., Böhm-Vélez, M., ... & Pisano, E. D. (2008). **Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer**. *JAMA*, 299(18), 2151-2163. Disponível em: https://jamanetwork.com/journals/jama/fullarticle/181896. 12 de nov. de 2024.

Boyd, N. F., Guo, H., Martin, L. J., Sun, L., Stone, J., Fishell, E., ... & Yaffe, M. J. (2007). **Mammographic density and the risk and detection of breast cancer**. *New England Journal of Medicine*, 356(3), 227-236. Disponível em: https://www.nejm.org/doi/full/10.1056/NEJMoa062790. 12 de nov. de 2024.

CARTER, Stacy M.; ROGERS, Wendy; WIN, Khin Than; FRAZER, Helen; RICHARDS, Bernadette; HOUSSAMI, Nehmat. **The ethical, legal and social**

implications of using artificial intelligence systems in breast cancer care. *The Breast*, v. 49, p. 25–32, Feb. 2020. Disponível em: https://doi.org/10.1016/j.breast.2019.10.001. Acesso em: 15 de nov. de 2024.

Chaurasia, V., & Pal, S. (2020). **Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer**. *SN Computer Science, 1*(270). Disponível em: https://doi.org/10.1007/s42979-020-00296-8. Acesso em: 15 de nov. de 2024.

Egwom, O. J., Hassan, M., Tanimu, J. J., Hamada, M., & Ogar, O. M. (2022). **An LDA–SVM Machine Learning Model for Breast Cancer Classification**. *BioMedInformatics*, 2(3), 345-358. Disponível em: https://doi.org/10.3390/biomedinformatics2030022. Acesso em: 21 de nov. de 2024

Heidari, M., Zargari Khuzani, A., Hollingsworth, A. B., Danala, G., Mirniaharikandehei, S., Qiu, Y., Liu, H., & Zheng, B. (2018). **Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm**. *Physics in Medicine & Biology, 63*(3), 035020. Disponível em: https://doi.org/10.1088/1361-6560/aaa1ca. Acesso em: 23 de nov. de 2024

Houssein, E. H., Emam, M. M., Ali, A. A., & Suganthan, P. N. (2021). Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*, *167*, 114161. Disponível em: https://doi.org/10.1016/j.eswa.2020.114161. Acesso em: 23 de nov. de 2024

International Agency for Research on Cancer (IARC). **Breast Cancer Fact Sheet**. (2020) Disponível em: https://gco.iarc.who.int/media/globocan/factsheets/cancers/20-breast-fact-sheet.pdf. Acesso em: 11 de set. de 2024.

Instituto Nacional de Câncer. (2023). **Estimativa 2023: Incidência de Câncer no Brasil**. Rio de Janeiro: INCA. Disponível em: https://www.inca.gov.br/publicacoes/livros/estimativa-2023-incidencia-de-cancer-no-brasil. Acesso em: 11 de set. de 2024.

Jasti, V. D. P., Zamani, A. S., Arumugam, K., Naved, M., Pallathadka, H., Sammy, F., Raghuvanshi, A., & Kaliyaperumal, K. (2022). **Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis**. *Security and Communication Networks*, 2022, Article ID 1918379. Disponível em: https://doi.org/10.1155/2022/1918379. Acesso em: 12 de nov. de 2024.

Koh, D.-M., Papanikolaou, N., Bick, U., Illing, R., Kahn, C. E. Jr., Kalpathi-Cramer, J., Matos, C., Martí-Bonmatí, L., Miles, A., Mun, S. K., Napel, S., Rockall, A., Sala, E., Strickland, N., & Prior, F. (2022). **Artificial intelligence and machine learning in cancer imaging**. *Communications Medicine, 2*, 133. Disponível em: https://doi.org/10.1038/s43856-022-00199-0. Acesso em: 23 de nov. de 2024

Lehman, C. D., Wellman, R. D., & Buist, D. S. (2015). **Diagnostic Accuracy of Digital Screening Mammography with and without Computer-Aided Detection**. *JAMA Internal Medicine*, 175(3), 394-401. Disponível em: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2443369. Acesso em: 12 de nov. de 2024.

Lehman, C. D., Mercaldo, S., Lamb, L. R., King, T. A., Ellisen, L. W., Specht, M., & Tamimi, R. M. (2022). **Artificial intelligence for breast cancer diagnosis: Current applications and future directions**. *JNCI: Journal of the National Cancer Institute*, 114(10), 1355-1363. Disponível em: https://doi.org/10.1093/jnci/djac142. Acesso em: 12 de nov. de 2024.

McKinney, S. M., Sieniek, M., & Godbole, V. (2020). **International Evaluation of an AI System for Breast Cancer Screening**. *Nature*, 577(7788), 89-94. Disponível em: https://www.nature.com/articles/s41586-019-1799-6. Acesso em: 12 de nov. de 2024.

Mann, R. M., Cho, N., & Moy, L. (2019). **Breast MRI: State of the Art.** *Radiology*. Disponível em: https://doi.org/10.1148/radiol.2019182947. Acesso em: 12 de nov. de 2024.

NAJI, Mohammed Amine; EL FILALI, Sanaa; AARIKA, Kawtar; BENLAHMAR, EL Habib; AIT ABDELOUHAHID, Rachida; DEBAUCHÉ, Olivier. **Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis**. Procedia Computer Science, v. 191, p. 487–492, 2021. Disponível em: https://doi.org/10.1016/j.procs.2021.07.062. Acesso em: 24 de nov. de 2024.

Reig, B., Heacock, L., Geras, K. J., & Moy, L. (2019). **Machine learning in breast MRI**. *Journal of Magnetic Resonance Imaging*, *50*(1), 1-18. Disponível em: https://doi.org/10.1002/jmri.26852. Acesso em: 23 de nov. de 2024.

SHRAVYA, Ch.; PRAVALIKA, K.; SUBHANI, Shaik. **Prediction of Breast Cancer Using Supervised Machine Learning Techniques**. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, v. 8, n. 6, p. 1106, abr. 2019. Disponível em: https://www.academia.edu/download/79315323/F3384048619.pdf. Acesso em: 23 de nov. de 2024.

Sociedade Brasileira de Mastologia. (2023). **A importância da mamografia no rastreamento do câncer de mama**. São Paulo: SBM. Disponível em: https://www.sbmastologia.com.br/sociedades-medicas-brasileiras-recomendam-mamografia-anual-a-partir-dos-40-anos/. Acesso em: 03 de nov. de 2024.

World Health Organization. (2024). **Breast cancer. Geneva: WHO**. Disponível em: https://www.who.int/news-room/fact-sheets/detail/breast-cancer. Acesso em: 03 de nov. de 2024.

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). **A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction**. *Radiology*, 291(2), 532-540. Disponível em: https://doi.org/10.1148/radiol.2019182716. Acesso em: 12 de nov. de 2024.

Yala, A., Schuster, T., Miles, R., Barzilay, R., & Lehman, C. (2019). **Artificial intelligence in breast cancer screening: A collaborative study**. *Radiology*, 291(3), 682-690. Disponível em: https://doi.org/10.1148/radiol.2019182908. Acesso em: 12 de nov. de 2024.

Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). **Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis**. *Designs*, 2(2), 13. Disponível em: https://doi.org/10.3390/designs2020013. Acesso em: 23 de nov. de 2024.