



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS VII - PATOS
CENTRO DE CIÊNCIAS EXATAS E SOCIAIS APLICADAS
CURSO DE GRADUAÇÃO EM BACHARELADO EM COMPUTAÇÃO**

HALINE CARTAXO MILITÃO

**UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA NO DIAGNÓSTICO DE CÂNCER
DE MAMA: UMA REVISÃO INTEGRATIVA DA LITERATURA**

**PATOS - PB
2023**

HALINE CARTAXO MILITÃO

**UTILIZAÇÃO DE APRENDIZADO DE MÁQUINA NO DIAGNÓSTICO DE CÂNCER
DE MAMA: UMA REVISÃO INTEGRATIVA DA LITERATURA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Computação do Centro de Ciências Exatas e Sociais Aplicadas da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Computação.

Orientador: Me. Francisco Anderson Mariano da Silva

**PATOS - PB
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

M644u Militao, Haline Cartaxo.

Utilização de aprendizado de máquina no diagnóstico de câncer de mama [manuscrito] : uma revisão integrativa da literatura / Haline Cartaxo Militao. - 2023.

35 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas, 2023.

"Orientação : Prof. Me. Francisco Anderson Mariano da Silva, Coordenação do Curso de Computação - CCEA. "

1. Inteligência artificial. 2. Aprendizado de Máquina. 3. Diagnóstico. 4. Tratamento de saúde. I. Título

21. ed. CDD 005.3

HALINE CARTAXO MILITÃO

**Utilização de Aprendizado de Máquina no Diagnóstico de Câncer de Mama: Uma
Revisão Integrativa da Literatura**

Trabalho de Conclusão de Curso (TCC)
apresentado ao Curso de Graduação em
Bacharelado em Computação da Universidade
Estadual da Paraíba, como requisito parcial à
obtenção do título de Bacharel em Computação.

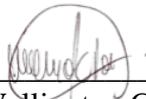
Área de concentração: Aprendizado de
Maquina

Aprovada em: 28/11/2023.

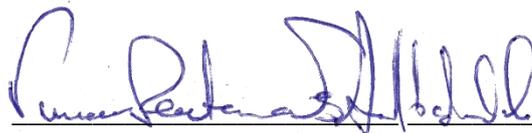
BANCA EXAMINADORA



Prof. Me. Francisco Anderson Mariano da Silva (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Wellington Candeia de Araujo
Universidade Estadual da Paraíba (UEPB)



Prof. Me. Vinicius Reuteman Feitoza Alves de Andrade
Universidade Estadual da Paraíba (UEPB)

Dedico este trabalho a minha família, amigos, professores e todos que me apoiaram nesta jornada. Este trabalho só foi possível graças a vocês.

AGRADECIMENTOS

À minha querida mãe, cujo amor incondicional e apoio constante foram a luz que me guiou durante os desafios acadêmicos. Seu exemplo de dedicação e perseverança sempre foi uma inspiração para mim, e eu não estaria aqui hoje sem o seu encorajamento.

Aos meus familiares, que também desempenharam um papel fundamental em minha jornada acadêmica, proporcionando incentivo e apoio que me ajudaram a superar os obstáculos que encontrei ao longo do caminho.

Além disso, gostaria de agradecer aos colegas de classe pelos momentos de apoio e amizade.

Por fim, gostaria de expressar minha sincera gratidão ao meu orientador, sua assistência e dedicação foram fundamentais para o desenvolvimento deste trabalho

*Nós só podemos ver um pouco do futuro,
mas o suficiente para perceber
que há muito a fazer.*

Alan Turing

RESUMO

A inteligência artificial (IA) representa um campo da ciência da computação que simula os processos cognitivos humanos, incluindo a capacidade de aprender e armazenar conhecimento. Atualmente, na área da medicina, a IA está sendo utilizada para aprimorar o diagnóstico, prognóstico e tratamento em várias especialidades, oferecendo inúmeros benefícios para a saúde humana. Essa aplicação da IA possibilita uma maior precisão na previsão da progressão das doenças, melhora o desempenho dos tratamentos e reduz os riscos para os pacientes. O propósito deste estudo consistiu em coletar artigos por meio de uma revisão integrativa, utilizando elementos da metodologia PRISMA, os quais foram obtidos em bases de dados como IEEE Xplore, PubMed e Scopus, com o intuito de evidenciar o uso da inteligência artificial na medicina voltado para o câncer de mama. Assim, este trabalho demonstra que o aprendizado de máquina (*machine learning*) apresenta um caminho para aprimorar a precisão e a confiabilidade das modalidades de diagnóstico, contribuindo de forma significativa para alcançar o objetivo da medicina de precisão.

Palavras-chave: Inteligência artificial; Aprendizado de Máquina; Diagnóstico; Tratamento de saúde.

ABSTRACT

Artificial intelligence (AI) represents a field of computer science that simulates human cognitive processes, including the ability to learn and store knowledge. Currently, in the field of medicine, AI is being employed to enhance the diagnosis, prognosis, and treatment across various specialties, offering numerous benefits for human health. This application of AI enables greater accuracy in predicting disease progression, improves treatment performance, and reduces risks for patients. The purpose of this study was to collect articles through an integrative review, utilizing elements of the PRISMA methodology, obtained from databases such as IEEE Xplore, PubMed, and Scopus. The aim was to highlight the use of artificial intelligence in medicine, specifically focused on breast cancer. Thus, this work demonstrates that machine learning provides a pathway to enhance the accuracy and reliability of diagnostic modalities, contributing significantly to achieving the goal of precision medicine.

Keywords: Artificial intelligence; Machine Learning; Diagnosis; Health treatment.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exame de mamografia	16
Figura 2 – Exemplo de Redes Neurais	17
Figura 3 – Treinamento de aprendizado de máquina	18
Figura 4 – Fluxo do processo de seleção dos artigos para a presente revisão integrativa da literatura, elaborado com base na metodologia prisma.	24

LISTA DE QUADROS

Quadro 1 – Critérios de inclusão e exclusão	22
Quadro 2 – Processo de pesquisa	23
Quadro 3 – Detalhes do Artigo	25

LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Network</i>
DT	<i>Decision Tree</i>
GBC	<i>Gradient Boosting</i>
IA	<i>Artificial Intelligence</i>
INCA	<i>Instituto Nacional de Câncer</i>
KNN	<i>K-nearest Neighbors</i>
ML	<i>Machine Learning</i>
OMS	<i>Organização Mundial da Saúde</i>
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</i>
RF	<i>Random Forest</i>
SVM	<i>Support-Vector Machine</i>
XGB	<i>XgBoost</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Questões Problema	13
1.2	Objetivos	13
1.2.1	<i>Objetivo Geral</i>	<i>13</i>
1.2.2	<i>Objetivos Específicos</i>	<i>13</i>
1.3	Justificativa	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Câncer de mama	15
2.2	Detecção e diagnóstico	15
2.3	Inteligencia Artificial Aplicada na saúde	16
2.4	Aprendizado profundo	16
2.5	Aprendizado de máquina	17
2.6	Algoritmos de aprendizado de máquina	18
2.6.1	<i>Máquina de vetores de suporte (SM)</i>	<i>18</i>
2.6.2	<i>Árvore de decisão (DT)</i>	<i>18</i>
2.6.3	<i>Floresta aleatória (RF)</i>	<i>18</i>
2.6.4	<i>Aumento de gradiente (GBC)</i>	<i>19</i>
2.6.5	<i>Rede neural convolucional (CNN)</i>	<i>19</i>
2.6.6	<i>XgBoost (XGB)</i>	<i>20</i>
2.6.7	<i>K-vizinho mais próximo (KNN)</i>	<i>20</i>
3	METODOLOGIA	21
3.1	Elaboração de pergunta norteadora	21
3.2	Crterios de Elegibilidade	22
3.3	Período de Busca	22
3.4	Fontes de Informação	22
3.5	Seleção dos Estudos	23
3.6	Fluxograma	23
4	RESULTADOS E DISCUSSÕES	25
5	CONCLUSÃO	30
	REFERÊNCIAS	31

1 INTRODUÇÃO

O câncer de mama é a segunda maior causa de óbitos por câncer entre mulheres globalmente, com um registro de cerca de 8,2 milhões de mortes (Mangukiya; Vaghani; Savani, 2022). Este tipo de câncer refere-se a uma forma de crescimento anormal de células malignas originadas no tecido mamário. Estimativas revelam que aproximadamente 9 milhões de pessoas, ou 1 em cada 6, morrem de câncer anualmente em nações com renda média e baixa. A Organização Mundial da Saúde (OMS) observa que o câncer de mama é atualmente o tipo mais comum em todo o mundo, principalmente em países desenvolvidos, mas está se expandindo rapidamente para regiões de renda média e baixa. Segundo o Instituto Nacional de Câncer (INCA) foram estimados 73.610 casos novos de câncer de mama em 2023, com um risco estimado de 66,54 casos a cada 100 mil mulheres. Existem duas formas distintas de câncer de mama: benigna e maligna. Os tumores benignos crescem lentamente e têm margens definidas que não invadem o tecido circundante. Por outro lado, os tumores malignos se disseminam rapidamente para outras partes do corpo. Normalmente, esse tipo de câncer se desenvolve nos ductos de leite e glândulas do tecido mamário.

Considerando os métodos tradicionais de detecção, como mamografia, apresentam limitações, resultando em baixa sensibilidade e especificidade, o que pode levar a diagnósticos equivocados e procedimentos invasivos desnecessários. O diagnóstico manual, embora exista um volume significativo de dados de imagens médicas, tende a ser demorado e suscetível a variações dependendo do observador, especialmente devido à escassez de médicos especialistas. Por conseguinte, identificar o câncer de mama pode ser um desafio, já que muitas vezes não apresenta sintomas. Todos esses exames médicos são dispendiosos.

Contudo, se o câncer for detectado em seus estágios iniciais, os pacientes podem evitar terapias invasivas desnecessárias. A interpretação precisa das imagens de mamografia e a identificação de características específicas do câncer de mama podem ser particularmente desafiadoras para os profissionais de saúde, resultando em taxas significativas de resultados falso-positivos e falso-negativos (Monticciolo, 2013).

Diante desse cenário, ressalta-se a necessidade crítica de métodos mais precisos e eficazes para o diagnóstico e o tratamento oportuno do câncer de mama. Para superar esses desafios, o campo do diagnóstico de câncer de mama tem explorado cada vez mais o potencial do aprendizado de máquina e da inteligência artificial. A implementação de algoritmos de aprendizado de máquina em sistemas de diagnóstico pode oferecer uma abordagem complementar e precisa para analisar imagens médicas e dados clínicos complexos (Ha, 2017).

Essas ferramentas podem auxiliar os médicos na detecção precoce, na diferenciação entre tumores malignos e benignos e no desenvolvimento de planos de tratamento mais personalizados e eficazes para os pacientes com câncer de mama.

Assim, esse trabalho visa reunir artigos demonstrando o uso da inteligência artificial no campo da saúde e demonstrar que o aprendizado de máquina fornece um caminho para

melhorar a precisão e a confiabilidade das modalidades diagnósticas, têm o potencial de contribuir significativamente para alcançar o objetivo da medicina de precisão. Para alcançar esse objetivo, este estudo se baseia em desenvolver conhecimentos teóricos sobre o tema por meio de uma revisão integrativa da literatura existente. Dessa forma, foi possível reunir, analisar e comparar informações para garantir a validade e relevância dos resultados obtidos.

1.1 Questões Problema

Q1 - Como o aprendizado de máquina tem sido aplicado no diagnóstico de câncer de mama?

Q2 - Quais são as perspectivas futuras para a aplicação do aprendizado de máquina no diagnóstico de câncer de mama e sua influência na medicina?

1.2 Objetivos

1.2.1 *Objetivo Geral*

O objetivo geral desta revisão integrativa da literatura é apresentar, por meio de publicações científicas, estudos já conduzidos sobre utilização de aprendizado de máquina no diagnóstico de câncer de mama. Dessa forma, o estudo visa contribuir para o desenvolvimento da área e abrir possibilidades para futuras pesquisas.

1.2.2 *Objetivos Específicos*

Para se alcançar o objetivo geral deste trabalho, foram necessários atingir os seguintes objetivos específicos:

- Identificar quais algoritmos de aprendizado de máquina são mais bem-sucedidos na previsão e diagnóstico do câncer de mama de acordo com as taxas de precisão.
- Avaliar os classificadores com base nos recursos e características extraídas.
- Identificar os benefícios e limitações enfrentados na área abordada.

1.3 Justificativa

A revisão integrativa da literatura sobre utilização de aprendizado de máquina no diagnóstico de câncer de mama, justifica-se pelos seguintes motivos. Em primeiro lugar, com os avanços tecnológicos em inteligência artificial e *machine learning*, existe uma oportunidade notável para aplicar essas técnicas de maneira inovadora no campo do diagnóstico do câncer de mama. Uma revisão abrangente da literatura sobre o uso dessas técnicas pode fornecer [insights] sobre o potencial dessas ferramentas no diagnóstico precoce e na melhoria da precisão diagnóstica.

A revisão integrativa sobre a aplicação de algoritmos de *machine learning* pode destacar as vantagens potenciais dessas técnicas no aumento da precisão do diagnóstico de câncer de

mama, levando a intervenções mais precisas e oportunas. Com a detecção precoce desse tipo de câncer, é possível melhorar significativamente as taxas de sobrevivência e a eficácia dos tratamentos. Sobre o uso de *machine learning* no diagnóstico de câncer de mama pode revelar como essas técnicas têm o potencial de reduzir o ônus emocional e financeiro para os pacientes, ao mesmo tempo que otimiza os recursos do sistema de saúde por meio de um diagnóstico mais rápido e preciso.

Além disso, pode ajudar a identificar lacunas atuais no conhecimento e na aplicação de técnicas de *machine learning* no diagnóstico de câncer de mama, fornecendo uma base para futuras pesquisas e desenvolvimento de novas abordagens que possam melhorar ainda mais a eficácia e a acessibilidade do diagnóstico.

Ao consolidar as descobertas da literatura existente, esta revisão integrativa pode fornecer uma base sólida para o desenvolvimento de políticas de saúde baseadas em evidências que visam implementar práticas de diagnóstico eficazes e acessíveis, garantindo assim melhores resultados de saúde para os pacientes.

Em última análise, esta revisão integrativa busca consolidar as informações atuais e destacar o potencial impacto positivo da aplicação de técnicas de *machine learning* no diagnóstico de câncer de mama, com a esperança de fornecer diretrizes valiosas para pesquisas futuras e práticas clínicas aprimoradas.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, são apresentadas, conceitos e informações relevantes já estudados na literatura acadêmica, nesta etapa faz um resgate do acervo bibliográfico sobre câncer de mama, inteligência artificial, aprendizado de máquina e revisão da literatura construindo uma base sólida para discutir sobre diferentes classificadores e seus resultados de precisão.

2.1 Câncer de mama

O câncer de mama é uma doença resultante da proliferação desordenada de células atípicas na região mamária, resultando na formação de um tumor com capacidade de se disseminar para outras partes do corpo (Brasil, 2021). Essa enfermidade afeta predominantemente o sexo feminino, representando cerca de 22% dos novos casos anualmente (LOPES, 2014).

Desse modo, os fatores de risco incluem idade avançada, exposição prolongada a hormônios, sobrepeso e antecedentes familiares ou mutações genéticas (UK, 2023). Além disso, é considerada uma das principais causas de óbito prematuro em muitos países, antes dos 70 anos de idade (Brasil, 2020). Portanto, a detecção precoce é crucial, pois resulta em tratamentos menos agressivos e diagnósticos mais precisos e eficazes.

As glândulas mamárias desempenham um papel crucial na produção de leite, sendo formado nos lóbulos e conduzido até os mamilos por ductos de pequeno porte. Quando ocorre uma proliferação desordenada das células mamárias, é comum a formação de um tumor maligno nos ductos, embora em ocasiões menos frequentes possa se desenvolver nos lóbulos (Souza, 2013).

2.2 Detecção e diagnóstico

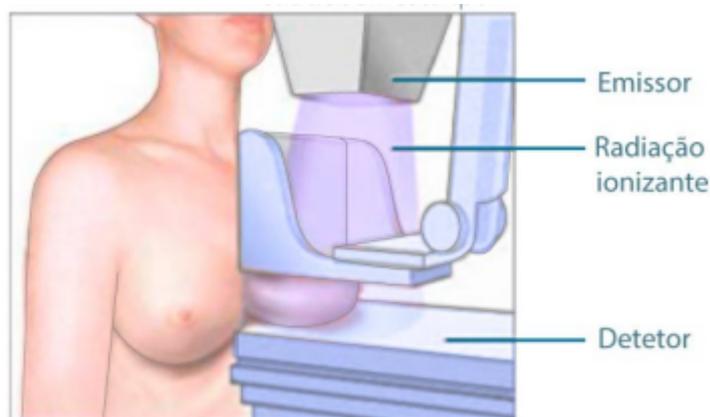
A detecção e diagnóstico precoces desempenham um papel crucial na melhoria dos desfechos clínicos e na redução da morbidade associada ao câncer de mama, uma das principais causas de mortalidade entre mulheres em todo o mundo (Brasil, 2022). A implementação de técnicas avançadas de imagem, como a mamografia digital e a ressonância magnética, desempenha um papel fundamental na detecção precoce de lesões suspeitas e na realização de avaliações diagnósticas precisas (Lehmann *et al.*, 2016).

Além disso, o desenvolvimento e a aplicação de biomarcadores específicos têm demonstrado um potencial significativo para aprimorar a precisão do diagnóstico e a estratificação de riscos em pacientes com câncer de mama. A abordagem integrada, que combina informações clínicas, radiológicas e genômicas, proporciona uma compreensão abrangente da complexidade do câncer de mama, permitindo a adoção de estratégias terapêuticas personalizadas e baseadas em evidências (Cardoso *et al.*, 2016).

Consoante com Houssami e Lord (2017) a interseção entre a inovação tecnológica e o conhecimento clínico está impulsionando avanços significativos na detecção precoce e no diagnóstico preciso do câncer de mama, com o potencial de melhorar substancialmente os

resultados e a qualidade de vida das pacientes afetadas. O exame mamografia é ilustrada por Sousa (2020) na **Figura 1**.

Figura 1 – Exame de mamografia



Fonte: Sousa (2020)

2.3 Inteligencia Artificial Aplicada na saúde

Neste campo, a inteligência artificial (IA) desempenha um papel crucial no auxílio ao diagnóstico, em conjunto com as avaliações dos médicos, na identificação e encaminhamento de pacientes com possíveis riscos de câncer. Além disso, os dados desses pacientes podem ser empregados em abordagens terapêuticas personalizadas, utilizando informações genéticas específicas do indivíduo (Dlamini *et al.*, 2020).

Na esfera oncológica, a aplicação da inteligência artificial envolve a análise de grandes volumes de dados e a identificação de padrões significativos (Santos; Baeßler, 2018), sendo empregada não apenas para a prevenção de doenças e o diagnóstico precoce, mas também para o planejamento de tratamentos e o acompanhamento contínuo dos pacientes.

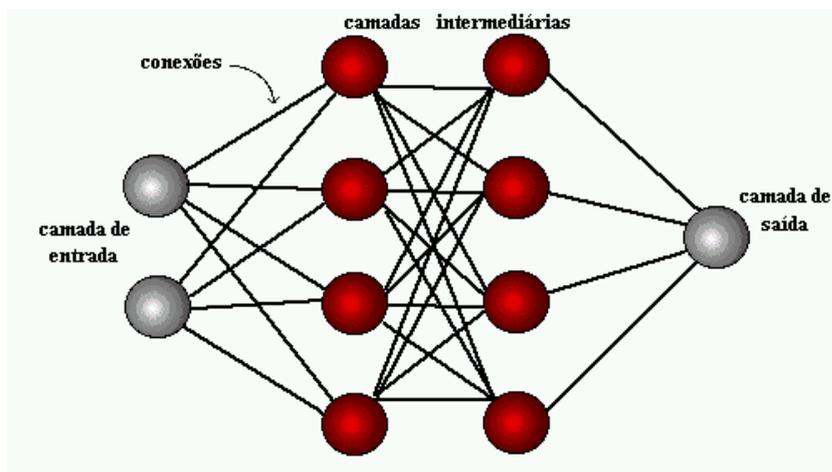
2.4 Aprendizado profundo

A aprendizagem profunda representa um procedimento que se baseia no uso iterativo da estrutura de árvore de decisão do processo de aprendizado de máquina, descartando certas características menos frequentes. Dessa forma, o sistema é capaz de re-processar os dados com base em seus resultados, possibilitando múltiplas iterações de processamento conforme necessário (Carvalho, 2009).

Devido à sua capacidade de identificar automaticamente diversos padrões presentes em um conjunto de dados, essa técnica se mostra adequada para capturar relações complexas entre imagens médicas e possíveis interpretações das mesmas. Essa característica tem impulsionado avanços significativos recentes, abrangendo áreas como a classificação de câncer de pele, a detecção de retinopatia diabética e a identificação de nódulos (Bien *et al.*, 2018).

Na **Figura 2** é mostrado como as funções escolhidas podem recriar outras funções subsequentes e, se a maioria das funções retorna a uma classificação, esta é a escolhida.

Figura 2 – Exemplo de Redes Neurais



Fonte: Carvalho (2022)

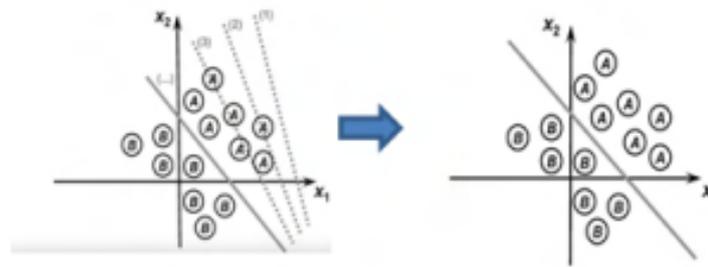
2.5 Aprendizado de máquina

Segundo Mitchell (1997), o aprendizado de máquina tem emergido como uma disciplina vital no campo da inteligência artificial, capacitando os sistemas computacionais a adquirir conhecimento e melhorar o desempenho de forma autônoma. Essa abordagem revolucionária permite que os algoritmos se adaptem e melhorem continuamente com base em experiências passadas, fornecendo *insights* valiosos e previsões precisas em uma variedade de domínios (Alpaydin, 2020).

Nesse sentido, a crescente disponibilidade de grandes conjuntos de dados e avanços significativos em tecnologias computacionais, o aprendizado de máquina desempenha um papel cada vez mais importante em áreas como reconhecimento de padrões, processamento de linguagem natural, análise de dados e tomada de decisões automatizadas. A aplicação eficaz do aprendizado de máquina requer uma compreensão profunda dos algoritmos subjacentes, bem como uma abordagem cuidadosa na preparação e análise dos conjuntos de dados relevantes (Hastie; Tibshirani; Friedman, 2009).

Separando os itens por similaridade, pode-se identificar uma função apropriada que efetivamente classifica os dados, como ilustrado na **Figura 3**. Diversas funções são testadas, e aquela que melhor distingue os dados é selecionada com base em seu desempenho.

Figura 3 – Treinamento de aprendizado de máquina



Fonte: Lama (2016)

2.6 Algoritmos de aprendizado de máquina

Os algoritmos de Aprendizado de Máquina, que abrangem uma variedade de técnicas estatísticas, probabilísticas e de otimização, possuem a capacidade de identificar prontamente padrões em conjuntos de dados extensos e complexos, adquirindo conhecimento por meio dos dados já presentes nos computadores.

2.6.1 Máquina de vetores de suporte (SM)

O método SVM estabelece um limite de decisão ótimo para a segregação de cada elemento dos dados em um plano onde os pontos são distribuídos no espaço n -dimensional (Tharwat, 2019). Esse plano é referido como hiperplano, com a intenção de maximizar a margem entre os pontos de ambas as classes (Li *et al.*, 2021). Esse algoritmo, fundamentado na teoria estatística de aprendizagem, constitui um tipo de aprendizado supervisionado em ML, útil para tarefas de classificação e regressão, aplicado para distinguir dados pertencentes a duas categorias distintas.

2.6.2 Árvore de decisão (DT)

A intenção do algoritmo de árvore de decisão é formular um modelo que consiste em uma ou várias árvores para categorizar o conjunto de dados inicial (Rajaguru; SR, 2019). Este método, baseado em árvores, simplifica um processo complexo em uma sequência de decisões elementares, começando a partir do topo e subdividindo-se em ramificações subsequentes (UK, 2020).

2.6.3 Floresta aleatória (RF)

O algoritmo Floresta Aleatória (RF) é uma das técnicas utilizadas na comunidade que envolve uma combinação de múltiplas árvores de decisão (Komura *et al.*, 2019). Essa abordagem foi inicialmente proposta por Leo Breiman em 2001. Constituindo um modelo que consiste em um conjunto de múltiplas árvores de decisão, o RF busca evitar o overfitting, um problema comum em árvores de decisão convencionais, fragmentando tanto o conjunto de dados quanto os recursos em várias partes e processando-os por meio de múltiplas árvores (Wang *et al.*, 2020).

O uso da *Random Forest* proporciona uma maior estabilidade em comparação com uma única árvore de decisão, uma vez que todos os nós da árvore são ramificados por meio da seleção aleatória dos melhores recursos em cada nó, e as árvores resultantes são desenvolvidas sem poda, utilizando seleção aleatória de características em cada conjunto de dados derivado do conjunto original.

2.6.4 Aumento de gradiente (GBC)

O Aumento de Gradiente (GBC) representa uma técnica essencial de aprendizado de máquina que encontra aplicação em uma variedade de problemas, abrangendo desde tarefas de classificação até regressão. Essa abordagem combina modelos de previsão menos precisos para formar um modelo mais robusto, frequentemente baseado em árvores de decisão (Friedman, 2001).

O objetivo central de qualquer algoritmo de aprendizado supervisionado é identificar e minimizar uma função de perda, e o GBC, como um tipo de algoritmo de impulsionamento (boosting), se baseia na ideia de que ao incorporar um próximo modelo mais ajustado aos modelos anteriores, é possível reduzir o erro global de previsão (Chen; Guestrin, 2016). A lógica fundamental por trás do GBC é estabelecer metas para os resultados do modelo subsequente, com o intuito de minimizar os erros, resultando em um modelo mais preciso e confiável. A versatilidade do Aumento de Gradiente é evidente em sua capacidade de lidar tanto com tarefas de classificação, onde a atribuição de rótulos a dados é o foco principal, quanto com tarefas de regressão, onde a previsão de valores numéricos é o objetivo central (Friedman, 2002).

2.6.5 Rede neural convolucional (CNN)

A aplicação da técnica conhecida como Rede Neural Convolucional (CNN) destaca-se como uma abordagem fundamental no campo do processamento de imagens e reconhecimento de padrões (Goodfellow; Bengio; Courville, 2016). Essa arquitetura específica de rede neural é uma das diversas estruturas de aprendizado que se originam do conceito original da rede neural Perceptron, sendo adaptada para lidar com tarefas mais complexas de reconhecimento de padrões em imagens (LeCun; Bengio; Hinton, 2015).

A principal característica da CNN reside na sua capacidade de processar dados de entrada por meio de camadas de convolução, que identificam padrões visuais e características específicas em imagens, facilitando a compreensão de relações espaciais e hierárquicas entre diferentes elementos da imagem (Zhang; Wallace, 2015).

Considerando a aplicação de filtros e camadas de pooling, a CNN é capaz de extrair características importantes das imagens e aprender representações hierárquicas, tornando-se uma ferramenta poderosa em tarefas como reconhecimento de objetos, classificação e segmentação de imagens (Russakovsky *et al.*, 2015). Logo, a evolução contínua das redes neurais convolucionais tem promovido avanços significativos em áreas como visão computacional e processamento de

imagens, demonstrando seu potencial para resolver problemas complexos de reconhecimento e interpretação de informações visuais (Krizhevsky; Sutskever; Hinton, 2012).

2.6.6 *XgBoost (XGB)*

XGBoost é um algoritmo de aprendizado de máquina utilizado para classificações em tarefas de aprendizado supervisionado e é fundamentado em Árvores de Decisão (Yu *et al.*, 2020). O XGBoost se destaca por sua execução rápida e simplificação de cálculos em comparação com outros algoritmos, o que o estabelece como uma abordagem bem-sucedida de aprendizado de máquina. Devido a esses atributos, ele demonstra resultados eficazes em conjuntos de dados multidimensionais (Qiu *et al.*, 2021).

O XGB é uma versão otimizada e de alto desempenho do algoritmo *Gradient Boosting*, com ajustes variados. Suas características essenciais incluem a capacidade de alcançar alta precisão preditiva, evitar o *overfitting*, lidar com dados ausentes de forma eficiente e processá-los rapidamente. A otimização tanto do software quanto do hardware tem sido implementada para garantir resultados superiores com a utilização mínima de recursos. Sua lógica de funcionamento é similar ao *Gradient Boosting*.

O primeiro passo no XGBoost é a obtenção da pontuação base, que pode assumir qualquer valor. Em conjunto com as ações tomadas nas etapas seguintes, a previsão correta é alcançada. A qualidade dessa previsão é avaliada por meio dos erros de previsão do modelo (resíduos), os quais são determinados pela diferença entre o valor previsto e o valor observado.

2.6.7 *K-vizinho mais próximo (KNN)*

Após considerar a aplicação do algoritmo KNN em problemas de classificação, é crucial compreender que este método se baseia na identificação de pontos de dados semelhantes dentro do conjunto de dados de treinamento, e, a partir dessa descoberta, realiza estimativas para a classificação de novos pontos de dados (McKinney, 2018).

A escolha adequada do parâmetro K, que representa o número de vizinhos mais próximos a serem considerados, desempenha um papel crucial na precisão das previsões realizadas pelo KNN (Hastie; Tibshirani; Friedman, 2009). Ao analisar a operação do algoritmo, torna-se evidente que o KNN toma decisões com base na composição dos resultados dos vizinhos mais próximos, avaliando os padrões presentes na vizinhança de cada ponto (Alpaydin, 2020).

A seleção criteriosa do valor ideal de K, que minimiza os erros de classificação, é essencial para garantir a eficácia e a precisão do algoritmo KNN em diversas tarefas de classificação (Duda; Hart; Stork, 2015). Considerando seus princípios fundamentais, o KNN se destaca como uma ferramenta versátil e poderosa para abordar problemas de classificação em diversos domínios, proporcionando *insights* valiosos com base nas relações entre os pontos de dados em um determinado espaço de características (Bishop, 2006).

3 METODOLOGIA

De acordo com Whitemore e Knafl (2005), a revisão integrativa da literatura desempenha um papel fundamental na síntese e análise abrangente de um corpo diversificado de conhecimento, permitindo a identificação de lacunas e tendências significativas em uma determinada área de pesquisa. Ao integrar de forma sistemática e meticulosa as descobertas de estudos diversos, esse método de revisão proporciona uma compreensão mais profunda das complexidades e nuances de um tópico específico, oferecendo *insights* valiosos para a formulação de políticas, prática clínica e direcionamento de futuras investigações (Torraco, 2005). Nesse sentido, a revisão integrativa, serve como um instrumento crucial para a síntese crítica e a construção de uma base sólida para o avanço do conhecimento em uma ampla gama de disciplinas e campos de estudo.

Este estudo utilizou os métodos exploratório e descritivo. De acordo com Prodanov e Freitas (2013), o método exploratório busca ampliar a compreensão do tema investigado, enquanto o método descritivo tem como objetivo detalhar e analisar as características ou variáveis associadas ao fenômeno ou processo, utilizando técnicas padronizadas de coleta de dados (Tozoni; Campos, 2009).

Ademais, na literatura existem alguns métodos para o desenvolvimento de uma revisão integrativa da literatura nesta pesquisa optou-se pela metodologia PRISMA. Uma abordagem amplamente reconhecida na condução de revisões sistemáticas e meta-análises, desempenha um papel crucial na garantia da transparência e rigor metodológico durante o processo de síntese de evidências (Moher *et al.*, 2009).

Ao oferecer diretrizes claras e abrangentes para o planejamento, execução e relato de revisões sistemáticas, o PRISMA proporciona uma estrutura sólida para a identificação de estudos relevantes, a avaliação crítica da qualidade dos dados e a síntese objetiva dos resultados (Page *et al.*, 2021). A implementação eficaz da metodologia PRISMA não apenas promove a integridade e a replicabilidade dos estudos, mas também contribui para a produção de evidências confiáveis e a tomada de decisões informadas em diversas áreas de pesquisa e prática clínica.

3.1 Elaboração de pergunta norteadora

Para organizar a pesquisa e incorporar os elementos fundamentais, é fundamental estabelecer perguntas norteadoras claras e objetivas que estejam alinhadas com o escopo da pesquisa em questão. Nesse contexto, foram formuladas duas perguntas neste trabalho:

- **Q1** Como o aprendizado de máquina tem sido aplicado no diagnóstico de câncer de mama?
- **Q2** Quais são as perspectivas futuras para a aplicação do aprendizado de máquina no diagnóstico de câncer de mama e sua influência na medicina?

3.2 Critérios de Elegibilidade

A revisão integrativa requer critérios de elegibilidade, tanto para a exclusão quanto para a inclusão de estudos, ao aplicar as *strings* nas bases de dados. Essa abordagem robusta é essencial para garantir a seleção adequada dos artigos que serão objeto de estudo. Tais critérios de inclusão e de exclusão serão listados no **Quadro 1**.

Quadro 1 – Critérios de inclusão e exclusão

Critérios de inclusão	Critérios de exclusão
Os artigos devem conter no título as palavras-chave ("breast cancer", "machine learning", "diagnosis");	Não serão selecionados estudos não acadêmicos, livros ou revisões;
Artigos escritos em língua inglesa;	Serão excluídos artigos duplicados;
Artigos publicados no período de 2020 - 2023.	Não serão selecionados artigos sem acesso livre.

Fonte: Elaborada pela autora (2023)

3.3 Período de Busca

Para a delimitação do período de busca, necessitou-se levar em consideração um tempo significativo para ser possível extrair informações relevantes e atualizadas para esta pesquisa. Portanto, o período escolhido foi de 2020 até 2023.

3.4 Fontes de Informação

O Quadro 2 apresenta as diferentes fases do processo de pesquisa conduzido para selecionar os materiais mais relevantes para os objetivos propostos no contexto do estudo.

A pesquisa começou com a inserção dos termos "breast cancer", "machine learning" e "diagnosis" nas bases de dados selecionadas, resultando em 4.552 registros. A busca foi refinada ao adicionar o operador [Title] entre os termos, reduzindo os resultados para 230 materiais. Posteriormente, um limite de período de publicação foi estabelecido, resultando em 105 publicações. Após a aplicação de critérios de exclusão, restaram 53 resultados relevantes.

A leitura dos artigos da base de dados IEEE Xplore foi realizada, seguida pela análise e seleção dos materiais obtidos na base de dados PubMed. No processo subsequente, os artigos da base de dados Scopus foram lidos, e foram selecionados aqueles que respondiam à pergunta norteadora.

Dessa maneira, os materiais escolhidos para atender aos objetivos do trabalho e responder às perguntas norteadoras serão provenientes das três bases de dados (IEEE Xplore, PubMed e Scopus), que demonstraram conter periódicos relevantes para o tema em questão.

Quadro 2 – Processo de pesquisa

Data	Ação
16/08/2023	Início da pesquisa geral Termos de pesquisa: - "breast cancer" - "machine learning" - "diagnosis" Resultados iniciais: 4.552
15/09/2023	Refinamento da pesquisa com o operador [Title] Resultados após refinamento: 230 materiais Delimitação do período de publicação reduz resultados para 105 publicações Aplicação de critérios de exclusão resulta em 53 resultados
25/09/2023	Leitura de artigos na IEEE Xplore Leituras, análises e seleção na base de dados PubMed
26/10/2023	Leitura de artigos na base de dados Scopus Seleção de artigos relevantes Materiais finais selecionados a partir das três bases de dados (IEEE Xplore, Pubmed e Scopus) que apresentaram periódicos relevantes

Fonte: Elaborada pela autora (2023)

3.5 Seleção dos Estudos

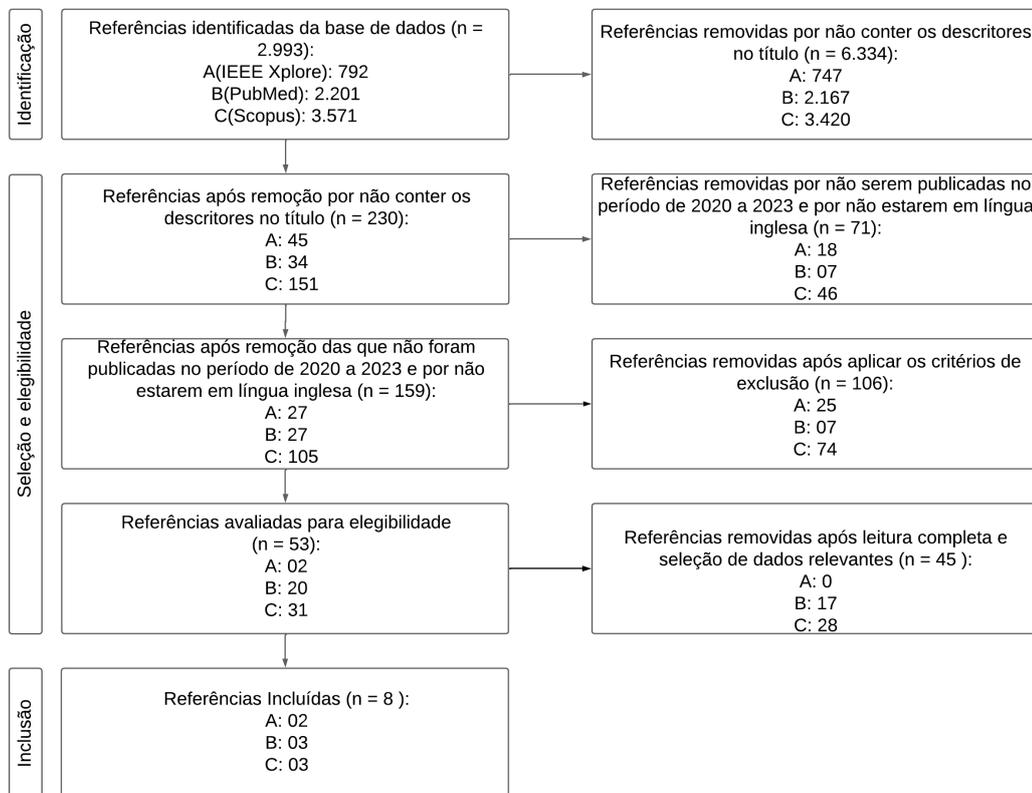
A seleção dos estudos foi realizada com base nos Critérios de Inclusão e Exclusão estabelecidos e a partir disso foi possível coletar um número significativo de artigos para a realização da pesquisa e aquisição de resultados.

3.6 Fluxograma

O fluxograma apresentado na **Figura 4** foi criado com o objetivo de proporcionar uma representação visual e concisa do processo de desenvolvimento da pesquisa. Através dessa representação gráfica, torna-se possível compreender a sequência de etapas seguidas nos estudos selecionados. A lista a seguir descreve as ações realizadas:

1. Foram utilizados os indicadores A, para referenciar as buscas no banco de dados IEEE Xplore, B, para referenciar as buscas no banco de dados PubMed e C para referenciar as buscas no banco de dados Scopus.
2. Eliminar materiais que não continham os descritores no título.
3. Excluir artigos que estivessem fora do período de tempo estabelecido.
4. Aplicar todos os critérios de exclusão definidos.
5. Ler e avaliar se o conteúdo correspondia à pergunta norteadora da pesquisa.

Figura 4 – Fluxo do processo de seleção dos artigos para a presente revisão integrativa da literatura, elaborado com base na metodologia prisma.



Fonte: Elaborada pela autora (2023)

4 RESULTADOS E DISCUSSÕES

Esta seção introduz o que foi estudado até o momento, abrangendo a exposição dos resultados da revisão integrativa e do conteúdo abordado na seção anterior (procedimentos metodológicos), após a aplicação dos critérios, a leitura e análise dos artigos escolhidos. O **Quadro 3** ilustra um total de oito artigos que foram selecionados para a pesquisa. Assim, a partir dos artigos selecionados, deu-se início ao processo de avaliação das informações.

A seleção dos estudos foi baseada na sua relevância e adequação quanto aos objetivos da revisão integrativa, sendo excluídos aqueles que não tivessem associação com o tema ou fossem redundantes. No **Quadro 3**, apresentam-se os artigos selecionados em todas as bases de dados utilizando as *strings* de busca "Breast cancer", "Machine learning" e "Diagnosis" no período de "2020 até 2023". Para organização das informações o quadro foi dividido em cinco colunas, na qual cada uma trata-se respectivamente sobre título, autores, ano de publicação, base de dados e algoritmo destacado nos resultados de cada artigo.

Quadro 3 – Detalhes do Artigo

Artigo	Autores	Ano	Base de Dados	Algoritmo Destacado
An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers	Naseem et al.	2022	IEEE Xplore	SVM
Auxiliary Diagnosis of Breast Cancer Based on Machine Learning and Hybrid Strategy	Chen et al.	2023	IEEE Xplore	RF XGB
Comparison of Classification Success Rates of Different Machine Learning Algorithms in the Diagnosis of Breast Cancer	Ozcan, Aydin, Cektinkaya	2022	Pub Med	GBC
Comparison of machine learning models for breast cancer diagnosis	Rania, Mohammed	2023	Scopus	GBC

CT-ML: Diagnosis of Breast Cancer Based on Ultrasound Images and Time- Dependent Feature Extraction Methods Using Contourlet Transformation and Machine Learning	Hajipour et al.	2022	Pub Med	DT
Presentation of Novel Architecture for Diagnosis and Identifying Breast Cancer Location Based on Ultrasound Images Using Machine Learning	Pourasad et al.	2021	Pub Med	CNN
Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis	Durga et al.	2022	Scopus	SVM
Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis	Abdur et al.	2022	Scopus	SVM

Fonte: Elaborada pela autora (2023)

Naseem *et al.* (2022) apresenta neste artigo, a utilização dois conjuntos de dados, o Wisconsin Breast Cancer (Diagnóstico) e o Wisconsin Breast Cancer (Prognóstico), consistindo em 569 e 198 instâncias, respectivamente, para avaliar a eficácia de um modelo proposto. Estratégias de abordagem de algoritmo e de dados foram empregadas para lidar com o problema de classificação desequilibrada. A padronização dos dados foi aplicada para garantir consistência. A precisão foi adotada como a métrica de avaliação, com o SVM mostrando o desempenho mais forte em ambos os conjuntos de dados, atingindo 98,10% e 78,35% de precisão no diagnóstico e no prognóstico, respectivamente. Além disso, a RNA também apresentou um desempenho notável, alcançando 98,24% de precisão no diagnóstico e 90,22% de precisão no prognóstico.

Chen *et al.* (2023) investiga nesse estudo experimentos com o conjunto de dados WDBC, (*Wisconsin Diagnostic Breast Cancer*) foi empregado nos experimentos os modelos XGBoost, RF, SVM, KNN e LR, foram aplicados para a classificação de massas mamárias como benignas

ou malignas. A configuração dos parâmetros dos experimentos envolveu o uso de 70% dos dados para treinamento e 30% para teste, com busca em grade e validação cruzada de 10 vezes para encontrar a melhor combinação de parâmetros para cada modelo. Uma análise comparativa de modelos revelou que o XGBoost teve os melhores resultados para o grupo de controle, enquanto o RF apresentou a melhor performance para o grupo experimental.

Além disso, as métricas de avaliação foram melhoradas significativamente com o uso da estratégia híbrida proposta no estudo. O tempo de treinamento dos modelos sob diferentes estratégias foi analisado, revelando que a estratégia híbrida exigiu mais tempo, mas apresentou o melhor desempenho de classificação, especialmente com os modelos XGBoost e RF. A comparação dos resultados com estudos anteriores demonstrou que a pesquisa atual alcançou uma precisão de 99,52%, superando estudos anteriores devido à utilização da estratégia híbrida proposta para processamento de dados.

Ozcan, Aydin e Cetinkaya (2022) analisam a previsão do câncer de mama usando algoritmos de aprendizado de máquina, reconhecendo sua importância crítica no diagnóstico e tratamento eficaz dessa doença. Com a aplicação de algoritmos de SVM, NB, RF, DT, KNN, LR, MLP, LDA, XGB, ABC e GBC, foram classificados os dados de tumores benignos e malignos, obtendo taxas de sucesso na previsão do câncer de mama. A análise comparativa revelou que o algoritmo GBC obteve a maior taxa de precisão, alcançando 99,12%. Os resultados ressaltam a eficácia dos algoritmos de impulsionamento (GB), ensacamento (RF) e redes neurais artificiais (MLP) na classificação precisa de pacientes com câncer de mama, reforçando a importância do uso de métodos de aprendizado de máquina para diagnóstico médico preciso e eficaz.

Kadhim e Kamil (2023) discutem o estudo envolvendo o conjunto de dados WDBC, visou classificar tumores em benignos e malignos utilizando dez métodos de classificação diferentes: DT, LDA, FRT, GB, LR, PA, Perceptron, NB, NC e SVC. A análise de desempenho de cada modelo foi avaliada em cinco aspectos: especificidade, sensibilidade, precisão, acurácia e escore F1. Os resultados indicaram que os modelos LDA, LR, PA, NB e SVC alcançaram uma especificidade máxima de 100%. Em relação à sensibilidade, o modelo Perceptron obteve o valor mais alto, atingindo 97,87%. A análise de precisão revelou que os modelos LDA, LR, PA, NB e SVC apresentaram valores de 100%, enquanto o Perceptron registrou o menor valor, 82,14%. A análise de acurácia mostrou uma alta acurácia entre quase todos os modelos, com o modelo GB alcançando a maior precisão de 97,36%.

Além disso, o escore F1 mostrou que o modelo GB também obteve o valor mais alto, com 96,77%. Em suma, os modelos de aprendizado de máquina mostraram uma taxa mais alta de precisão de classificação, reduziram os falsos positivos e obtiveram uma taxa de desempenho mais elevada, indicando que esses modelos podem ser valiosos para auxiliar os radiologistas no diagnóstico preciso do câncer de mama. Este estudo fornece uma base sólida para pesquisas futuras que possam continuar aprimorando a eficácia e o uso desses modelos na prática clínica.

Masjidi *et al.* (2022) introduz uma abordagem inovadora de extração de recursos baseada em transformação de contorno e modelos dependentes do tempo, revelando parâmetros visuais

ocultos de imagens de ultrassom de câncer de mama. A análise das características extraídas demonstrou a presença de associações significativas entre as subbandas da transformação de contorno, contribuindo para uma classificação precisa das classes de câncer. Diversos métodos de aprendizado de máquina foram empregados, resultando em diferentes níveis de sensibilidade e precisão. O método de árvore de decisão (DT) foi identificado como o mais adequado para o diagnóstico do câncer de mama, com sensibilidades de 87,8%, 92,0% e 87,0% para as classes normais, benignas e malignas, respectivamente. Além disso, as taxas de precisão para SVM, LDA, KNN, DT e RF foram avaliadas em 65%, 62,3%, 71,5%, 88,9% e 74,9%, respectivamente. O uso da metodologia híbrida proposta confirmou a eficácia da arquitetura DT para o diagnóstico preciso de câncer de mama.

Pourasad *et al.* (2021) analisa um conjunto de dados de 400 mulheres com imagens de ultrassonografia mamária, incluindo 780 imagens classificadas em três classes: normais, benignas e malignas. Seis métodos de ML foram aplicados, incluindo o método fractal para extração de características e os classificadores KNN, SVM, DT e NB. A arquitetura de rede neural de convolução (CNN) também foi utilizada para a classificação dos pacientes com base nas imagens de ultrassom. O método da árvore de decisão obteve uma precisão de 81%, seguido por KNN, SVM e NB com 67,7%, 40,1% e 44,9% de precisão, respectivamente. O modelo CNN proposto alcançou uma precisão de 99,8% no conjunto de treinamento e 76,1% no conjunto de teste, demonstrando sua eficácia no diagnóstico de câncer de mama.

Além disso, o modelo CNN foi capaz de identificar com precisão a localização e o volume dos tumores, utilizando operações morfológicas como algoritmo de pós-processamento. Essas descobertas podem ser aplicadas para monitorar e prevenir o crescimento de áreas infectadas, representando uma contribuição significativa para a medicina.

Reddy *et al.* (2022) apresenta resultados da análise, demonstra que a precisão do LS-SVM superou a de todos os outros classificadores, enquanto a sensibilidade e especificidade do KNN foram mais favoráveis em comparação com os outros classificadores. A metodologia adotada mostrou-se eficaz na detecção de doenças mamárias, especialmente ao destacar a importância da seleção de recursos relevantes para melhorar o desempenho da classificação. A análise foi realizada utilizando o banco de dados MIAS, que consiste em várias imagens de mamografias com diferentes tipos de ruído de fundo. A remoção adequada de ruídos das imagens foi essencial para a interpretação precisa dos resultados. As métricas de precisão, sensibilidade e especificidade foram adotadas para avaliar o desempenho dos diferentes algoritmos utilizados no estudo.

Os resultados apresentados destacaram a viabilidade e eficácia da abordagem proposta na detecção de doenças de câncer de mama. A análise comparativa dos diferentes classificadores permitiu a identificação dos pontos fortes de cada método, com o LS-SVM se destacando em termos de precisão, enquanto o KNN se mostrou mais sensível e específico para a detecção da doença. A aplicação bem-sucedida dessa metodologia sugere seu potencial para contribuir significativamente para o campo da detecção e diagnóstico de câncer de mama.

Rasool *et al.* (2022) mostra a avaliação dos modelos preditivos, que foram aplicados

vários algoritmos, incluindo SVM, LR, KNN e EC, em diferentes iterações e configurações. Os resultados indicaram que o *kernel* polinomial SVM e o LR com *Recursive Feature Elimination* (RFE) apresentaram desempenhos notáveis, atingindo pontuações de precisão e F1 superiores a 98%. O desempenho do modelo KNN com hiperparâmetro também foi notável, com uma pontuação F1 de 97,35%. A análise comparativa entre os modelos destacou a superioridade do LR com RFE em comparação com outros métodos, demonstrando a capacidade desse modelo de previsão na detecção precisa do câncer de mama. Além disso, a comparação com estudos anteriores que utilizaram os mesmos conjuntos de dados WDBC confirmou a eficácia dos modelos propostos neste estudo. Os resultados alcançados pelos modelos propostos superaram significativamente as técnicas anteriores, com o *kernel* polinomial SVM atingindo uma precisão de 99,03% e o LR com RFE alcançando uma precisão de 98,06%.

Dessa maneira, com base nos estudos abordados, após análise comparativa desses estudos, foi possível constatar que o algoritmo de aprendizado de máquina que mais se destacou em termos de precisão e bons resultados foi o *Support Vector Machine* (SVM). O SVM demonstrou consistentemente um desempenho notável em vários estudos, alcançando altas taxas de precisão em diferentes conjuntos de dados relacionados ao diagnóstico e prognóstico do câncer de mama. Em alguns estudos específicos, o SVM atingiu precisões tão altas quanto 99,03%, 98,10% e 78,35%, demonstrando sua capacidade de lidar efetivamente com problemas de classificação desequilibrada e fornecer resultados robustos e confiáveis para o diagnóstico de câncer de mama.

Os estudos sobre o uso de algoritmos de aprendizado de máquina no diagnóstico e prognóstico do câncer de mama apresentaram avanços promissores, destacando o SVM, XGBoost e GBC como os algoritmos mais eficazes em termos de precisão. No entanto, essas pesquisas enfrentaram desafios, incluindo o desequilíbrio de dados, complexidade dos dados médicos e interpretabilidade dos modelos. Além disso, pontos fracos como limitações de generalização, viés nos conjuntos de dados e complexidade dos algoritmos utilizados foram identificados. A abordagem dessas limitações é crucial para garantir a confiabilidade e a aplicabilidade prática dos resultados, visando contribuir de forma significativa para o avanço no diagnóstico e tratamento do câncer de mama.

5 CONCLUSÃO

Ao longo deste estudo, foi possível observar os avanços notáveis na aplicação do aprendizado de máquina no diagnóstico de câncer de mama. Diversos estudos revisados revelaram a eficácia e a confiabilidade do algoritmo *Support Vector Machine* (SVM), juntamente com outros modelos como XGBoost e *Gradient Boosting Classifier* (GBC), na identificação precisa de padrões complexos relacionados a essa doença. A precisão alcançada por esses modelos destaca sua capacidade de lidar com conjuntos de dados desafiadores e oferecer *insights* valiosos para os profissionais de saúde, potencialmente melhorando os resultados para os pacientes afetados.

No entanto, é crucial reconhecer os desafios enfrentados, incluindo o desequilíbrio de dados, a complexidade dos dados médicos e a interpretabilidade dos modelos. Para garantir a confiabilidade e a aplicabilidade prática dessas abordagens, é necessário abordar essas limitações por meio da implementação de estratégias de balanceamento de dados, métodos de interpretabilidade de modelo aprimorados e uma abordagem mais abrangente para o treinamento e validação dos modelos.

Quanto às perspectivas futuras, é evidente que o aprendizado de máquina tem o potencial de transformar significativamente o diagnóstico de câncer de mama e a prática médica em geral. A continuidade do desenvolvimento e aplicação dessas técnicas pode levar a avanços significativos no campo da medicina de precisão, oferecendo diagnósticos personalizados e terapias mais eficazes, e proporcionando uma base sólida para estratégias proativas de prevenção e intervenção precoce.

Além disso, a integração de dados e sistemas de saúde, o desenvolvimento de novas tecnologias de imagem e detecção, a automação de processos clínicos e o desenvolvimento de ferramentas de suporte à decisão clínica são áreas que podem se beneficiar significativamente do aprendizado de máquina, melhorando a eficiência e a qualidade dos cuidados de saúde e promovendo uma abordagem mais holística e centrada no paciente.

Para efetivar essas perspectivas futuras, é fundamental continuar investindo em pesquisa e desenvolvimento, promover colaborações interdisciplinares e interinstitucionais, e incentivar a implementação responsável e ética das tecnologias de aprendizado de máquina no campo da saúde. Somente assim poderemos maximizar o potencial transformador dessas tecnologias, garantindo resultados de saúde mais eficazes, acessíveis e centrados no paciente para aqueles afetados pelo câncer de mama e outras doenças complexas.

REFERÊNCIAS

- ALPAYDIN, E. **Introduction to machine learning**. Londres, Inglaterra: MIT press, 2020. Citado 2 vezes nas páginas 17 e 20.
- BIEN, N. *et al.* Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of mrnet. **PLOS Medicine**, Public Library of Science, v. 15, p. 1–19, 11 2018. Disponível em: <https://doi.org/10.1371/journal.pmed.1002699>. Acesso em: 18 out. 2023. Citado na página 16.
- BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006. 778 p. Disponível em: <https://link.springer.com/book/9780387310732>. Acesso em: 18 out. 2023. Citado na página 20.
- BRASIL. **Incidência de Câncer no Brasil**. Instituto Nacional de Câncer - INCA, 2020. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2020-incidencia-de-cancer-no-brasil.pdf>. Acesso em: 18 out. 2023. Citado na página 15.
- BRASIL. **Estatísticas de câncer**. Instituto Nacional de Câncer - INCA, 2021. Disponível em: <https://www.gov.br/inca/pt-br/assuntos/cancer/numeros>. Acesso em: 18 out. 2023. Citado na página 15.
- BRASIL. Dados e números sobre câncer de mama - relatório anual 2023. Instituto Nacional de Câncer, 2022. Disponível em: <https://www.inca.gov.br/publicacoes/relatorios/dados-e-numeros-sobre-cancer-de-mama-relatorio-anual-2023>. Acesso em: 18 out. 2023. Citado na página 15.
- CARDOSO, F. *et al.* 70-gene signature as an aid to treatment decisions in early-stage breast cancer. **The New England journal of medicine**, v. 375, n. 8, p. 717–729, 2016. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/27557300/>. Acesso em: 18 out. 2023. Citado na página 15.
- CARVALHO, A. **Redes Neurais Artificiais**. 2009. Disponível em: <https://sites.icmc.usp.br/andre/research/neural/>. Acesso em: 18 out. 2023. Citado na página 16.
- CARVALHO, A. **Redes Neurais Artificiais**. 2022. Disponível em: <https://sites.icmc.usp.br/andre/research/neural/>. Acesso em: 18 out. 2023. Citado na página 17.
- CHEN, H. *et al.* Auxiliary diagnosis of breast cancer based on machine learning and hybrid strategy. **IEEE Access**, PP, p. 1–1, 2023. Disponível em: https://www.researchgate.net/publication/373700744_Auxiliary_Diagnosis_of_Breast_Cancer_Based_on_Machine_Learning_and_Hybrid_Strategy. Acesso em: 18 out. 2023. Citado na página 26.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. **CoRR**, abs/1603.02754, 2016. Disponível em: <http://arxiv.org/abs/1603.02754>. Acesso em: 18 out. 2023. Citado na página 19.
- DLAMINI, Z. *et al.* Artificial intelligence (ai) and big data in cancer and precision oncology. **Computational and structural biotechnology journal**, v. 18, p. 2300–2311, 2020. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/32994889/>. Acesso em: 18 out. 2023. Citado na página 16.

DUDA, R.; HART, P.; STORK, D. irspot-gaense: identificando pontos de recombinação via classificador de conjunto e estendendo o conceito de pseaac de chou para formular amostras de dna. 2015. Citado na página 20.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 29, p. 1189–1232, 2001. Disponível em: <https://doi.org/10.1214/aos/1013203451>. Acesso em: 18 out. 2023. Citado na página 19.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics Data Analysis**, v. 38, p. 367–378, 2002. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167947301000652>. Acesso em: 18 out. 2023. Citado na página 19.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Acesso em: 18 out. 2023. Citado na página 19.

HA, R. Artificial intelligence in breast imaging. radiologic clinics of north america. 2017. Citado na página 12.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. Stanford, California: [s.n.], 2009. 767 p. Disponível em: <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>. Acesso em: 18 out. 2023. Citado 2 vezes nas páginas 17 e 20.

HOUSSAMI, N.; LORD, S. The changing landscape of breast cancer diagnosis and treatment in the era of personalised medicine. *medical journal of australia*. 2017. Citado na página 15.

KADHIM, R.; KAMIL, M. Comparison of machine learning models for breast cancer diagnosis. **IAES International Journal of Artificial Intelligence (IJ-AI)**, v. 12, p. 415–421, 2023. Disponível em: https://www.researchgate.net/publication/365615067_Comparison_of_machine_learning_models_for_breast_cancer_diagnosis. Acesso em: 18 out. 2023. Citado na página 27.

KOMURA, N. *et al.* Constrained sialic acid donors enable selective synthesis of -glycosides. **Science**, v. 364, n. 6441, p. 677–680, 2019. Disponível em: <https://www.science.org/doi/abs/10.1126/science.aaw4866>. Acesso em: 18 out. 2023. Citado na página 18.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. *et al.* **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2012. v. 25. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Acesso em: 18 out. 2023. Citado na página 20.

LAMA, R. D. **Perceptron multicamadas**. 2016. Disponível em: <https://commons.wikimedia.org/w/index.php?curid=53027287>. Acesso em: 18 out. 2023. Citado na página 18.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <https://doi.org/10.1038/nature14539>. Acesso em: 18 out. 2023. Citado na página 19.

LEHMANN, B. D. *et al.* Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. **PLOS ONE**, Public Library of Science, v. 11, n. 6, p. 1–22, 2016. Disponível em: <https://doi.org/10.1371/journal.pone.0157368>. Acesso em: 18 out. 2023. Citado na página 15.

LI, J. *et al.* Prevendo a sobrevivência em 5 anos do câncer de mama usando aprendizado de máquina: uma revisão sistemática. **PLOS ONE**, 2021. Disponível em: <https://www.x-mol.com/paper/1390993493914701824/t?recommendPaper=1298678527628120064>. Acesso em: 18 out. 2023. Citado na página 18.

LOPES, L. A. S. Câncer de mama: diagnóstico e tratamento. 2014. Citado na página 15.

MANGUKIYA, M.; VAGHANI, A.; SAVANI, M. Breast cancer detection with machine learning. **International Journal for Research in Applied Science and Engineering Technology**, v. 10, p. 141–145, 2022. Disponível em: <https://www.ijraset.com/best-journal/breast-cancer-detection-with-machine-learning>. Acesso em: 18 out. 2023. Citado na página 12.

MASJIDI, B. H. K. *et al.* Ct-ml: Diagnosis of breast cancer based on ultrasound images and time-dependent feature extraction methods using contourlet transformation and machine learning. **Computational Intelligence and Neuroscience**, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37800039/>. Acesso em: 18 out. 2023. Citado na página 27.

MCKINNEY, W. **Python Data Science Handbook**. 2018. Disponível em: <https://jakevdp.github.io/PythonDataScienceHandbook/05.02-introducing-scikit-learn.html>. Acesso em: 18 out. 2023. Citado na página 20.

MITCHELL, T. **Machine Learning**. McGraw-Hill Science, 1997. 432 p. Disponível em: <https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf>. Acesso em: 18 out. 2023. Citado na página 17.

MOHER, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. **PLOS Medicine**, Public Library of Science, v. 6, p. 1–6, 2009. Disponível em: <https://doi.org/10.1371/journal.pmed.1000097>. Acesso em: 18 out. 2023. Citado na página 21.

MONTICCIOLO, D. L. Breast imaging: the art and science of mammography. 2013. Citado na página 12.

NASEEM, U. *et al.* An automatic detection of breast cancer diagnosis and prognosis based on machine learning using ensemble of classifiers. **IEEE Access**, v. 10, p. 78242–78252, 2022. Disponível em: <https://ieeexplore.ieee.org/document/9773160>. Acesso em: 18 out. 2023. Citado na página 26.

OZCAN, I.; AYDIN, H.; CETINKAYA, A. Comparison of classification success rates of different machine learning algorithms in the diagnosis of breast cancer. **Asian Pacific Journal of Cancer Prevention : APJCP**, v. 23, n. 10, p. 3287–3297, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36308351/>. Acesso em: 18 out. 2023. Citado na página 27.

PAGE, M. J. *et al.* The prisma 2020 statement: An updated guideline for reporting systematic reviews. **PLOS Medicine**, Public Library of Science, v. 18, n. 3, p. 1–15, 2021. Disponível em: <https://doi.org/10.1371/journal.pmed.1003583>. Acesso em: 18 out. 2023. Citado na página 21.

POURASAD, Y. *et al.* Presentation of novel architecture for diagnosis and identifying breast cancer location based on ultrasound images using machine learning. **Diagnostics**, 2021. Disponível em: https://www.researchgate.net/publication/355210168_Presentation_of_Novel_Architecture_for_Diagnosis_and_Identifying_Breast_Cancer_Location_Based_on_Ultrasound_Images_Using_Machine_Learning. Acesso em: 18 out. 2023. Citado na página 28.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo, Rio Grande do Sul: [s.n.], 2013. 276 p. Disponível em: <https://www.feevale.br/Comum/midias/0163c988-1f5d-496f-b118-a6e009a7a2f9/E-book%20Metodologia%20do%20Trabalho%20Cientifico.pdf>. Acesso em: 18 out. 2023. Citado na página 21.

QIU, Y. *et al.* Performance evaluation of hybrid woa-xgboost, gwo-xgboost and bo-xgboost models to predict blast-induced ground vibration. **Engineering with Computers**, Springer, p. 4145–4162, 2021. Disponível em: <https://link.springer.com/article/10.1007/s00366-021-01393-9>. Acesso em: 18 out. 2023. Citado na página 20.

RAJAGURU, H.; SR, S. C. Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. **Asian Pacific journal of cancer prevention: APJCP**, Shahid Beheshti University of Medical Sciences, v. 20, n. 12, p. 3777, 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7173366/>. Acesso em: 18 out. 2023. Citado na página 18.

RASOOL, A. *et al.* Improved machine learning-based predictive models for breast cancer diagnosis. **International Journal of Environmental Research and Public Health**, v. 19, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35328897/>. Acesso em: 18 out. 2023. Citado na página 28.

REDDY, G. T. *et al.* Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis. **Security and Communication Networks**, 2022. Disponível em: <https://doi.org/10.1155/2022/1918379>. Acesso em: 18 out. 2023. Citado na página 28.

RUSSAKOVSKY, O. *et al.* Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, v. 115, p. 211–252, 2015. Disponível em: <https://doi.org/10.1007/s11263-015-0816-y>. Acesso em: 18 out. 2023. Citado na página 19.

SANTOS, D. Pinto dos; BAESSLER, B. Big data, artificial intelligence, and structured reporting. **European Radiology Experimental**, v. 2, n. 1, p. 42, 2018. Disponível em: <https://doi.org/10.1186/s41747-018-0071-4>. Acesso em: 18 out. 2023. Citado na página 16.

SOUSA, D. P. **Mamografia - o que é, indicações exame, como é realizada, preço**. 2020. Disponível em: <https://www.saudebemestar.pt/pt/exame/imagiologia/mamografia/>. Acesso em: 18 out. 2023. Citado na página 16.

SOUZA, V. S. **Altas habilidades e superdotação: uma reflexão sobre o tema**. 51 p. Monografia (MONOGRAFIA DE ESPECIALIZAÇÃO) — Universidade Tecnológica Federal do Paraná, Medianeira, 2013. Disponível em: <https://doi.org/10.1007/s11263-015-0816-y>. Acesso em: 18 out. 2023. Citado na página 15.

THARWAT, A. Parameter investigation of support vector machine classifier with kernel functions. **Knowledge and Information Systems**, Springer, v. 61, p. 1269–1302, 2019. Disponível em: <https://link.springer.com/article/10.1007/s10115-019-01335-4>. Acesso em: 18 out. 2023. Citado na página 18.

TORRACO, R. J. Writing integrative literature reviews: Guidelines and examples. **Human Resource Development Review**, v. 4, p. 356–367, 2005. Disponível em: <https://doi.org/10.1177/1534484305278283>. Acesso em: 18 out. 2023. Citado na página 21.

TOZONI, R.; CAMPOS, M. F. **Metodologia da Pesquisa**. 2. ed. Curitiba: [s.n.], 2009. 136 p. Citado na página 21.

UK, C. R. Um estudo comparativo da classificação de tumores de câncer de mama por métodos clássicos de aprendizado de máquina e método de aprendizado profundo. 2020. Citado na página 18.

UK, C. R. **Breast Cancer Risk Factors**. 2023. Disponível em: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/risk-factors#heading-Zero.%20Acesso%20em:%2017%20out.%202023>. Acesso em: 18 out. 2023. Citado na página 15.

WANG, S. *et al.* An improved random forest-based rule extraction method for breast cancer diagnosis. **Applied Soft Computing**, v. 86, p. 105941, 2020. ISSN 1568-4946. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1568494619307227>. Acesso em: 18 out. 2023. Citado na página 18.

WHITTEMORE, R.; KNAFL, K. The integrative review: Updated methodology. **Journal of Advanced Nursing**, v. 52, p. 546–553, 2005. Acesso em: 18 out. 2023. Citado na página 21.

YU, D. *et al.* Copy number variation in plasma as a tool for lung cancer prediction using extreme gradient boosting (xgboost) classifier. **Thoracic cancer**, Wiley Online Library, v. 11, n. 1, p. 95–102, 2020. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31694073/>. Acesso em: 18 out. 2023. Citado na página 20.

ZHANG, Y.; WALLACE, B. C. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. **CoRR**, 2015. Disponível em: <http://arxiv.org/abs/1510.03820>. Acesso em: 18 out. 2023. Citado na página 19.