MATHEUS GABRIEL PEREIRA NOGUEIRA

LUCAS GABRIEL NUNES GEREMIAS

JOEL SEPULVEDA MARTINS

LUCCA LUCCHIN DE CAMPOS COSTA

DRAYAN DA SILVA MAGALHÃES

Utilização de Técnicas de Aprendizagem de Máquina para Diagnóstico de Câncer

Projeto apresentado como requisito parcial da Disciplina de Pesquisa Aplicada do 6º Período, na Pontifícia Universidade Católica do Paraná.

Área de concentração: Ciência da Computação

Curitiba

2024

**RESUMO**

**Palavras-chaves:** Aprendizagem de Máquina Diagnóstico Médico; Tecnologia da Informação; Câncer.

**ABSTRACT**

**Keywords:** Machine Learning; Medical Diagnosis; Information Technology; Cancer

Sumário

[1. INTRODUÇÃO 5](#_Toc175495773)

[2. METODOLOGIA 6](#_Toc175495774)

[REFERÊNCIAS BIBLIOGRÁFICA 7](#_Toc175495775)

# 1. INTRODUÇÃO

O câncer é uma doença caracterizada pelo crescimento descontrolado de células anormais no corpo. Essas células podem invadir e destruir tecidos saudáveis, e, em casos avançados, espalhar-se para outras partes do corpo através do sangue e do sistema linfático, em um processo conhecido como metástase. Existem mais de 100 tipos diferentes de câncer, que podem afetar praticamente qualquer parte do corpo. A complexidade biológica e a diversidade dos tipos de câncer tornam o diagnóstico e o tratamento dessa doença um grande desafio para a medicina moderna.

O câncer é uma das principais causas de mortalidade em todo o mundo, representando um desafio significativo para os sistemas de saúde global. De acordo com a Organização Mundial da Saúde (OMS), o número de novos casos de câncer em 2020 foi estimado em 19,3 milhões, com aproximadamente 10 milhões de mortes associadas. Esses números alarmantes evidenciam a importância de diagnósticos precoces e precisos, que são cruciais para melhorar as taxas de sobrevivência dos pacientes.

O diagnóstico precoce do câncer não só aumenta as chances de sucesso do tratamento, como também pode reduzir os custos associados, aliviando a carga sobre os sistemas de saúde. No entanto, o diagnóstico de câncer é frequentemente desafiador, dado que envolve a análise de dados complexos, como imagens médicas, sequências genômicas e dados clínicos. Esse cenário destaca a necessidade de técnicas avançadas que possam auxiliar médicos e especialistas a identificar a presença de neoplasias malignas de forma mais eficaz e precisa.

Nesse contexto, as técnicas de Aprendizagem de Máquina (AM) emergem como ferramentas promissoras para o diagnóstico de câncer. A AM é um subcampo da inteligência artificial (IA) que envolve o desenvolvimento de algoritmos capazes de aprender a partir de dados, identificar padrões e tomar decisões baseadas nesses dados. Diferente de abordagens tradicionais de programação, onde as regras são explicitamente codificadas, na AM os sistemas são treinados em grandes conjuntos de dados e melhoram sua performance à medida que são expostos a novos exemplos.

Nos últimos anos, a aplicação de AM no diagnóstico médico tem demonstrado potencial para transformar a prática clínica, oferecendo suporte na interpretação de grandes volumes de dados e na detecção de anomalias sutis que poderiam passar despercebidas por métodos tradicionais.

Este artigo explora a aplicação de técnicas de AM no diagnóstico de câncer, focando em um estudo de caso que busca avaliar o desempenho de diferentes abordagens. Através da análise comparativa das principais métricas de performance, o estudo visa identificar quais técnicas se destacam na resolução dos desafios associados ao diagnóstico médico.

# 2. Revisão de Literatura

### 2.1 Dados

O câncer de mama é um tipo de tumor maligno que se desenvolve nas células do tecido mamário. Ele pode ser classificado em diferentes subtipos, dependendo da origem celular e das características moleculares do tumor. A detecção precoce é essencial para aumentar as chances de tratamento bem-sucedido, e uma das formas de detecção inclui o exame de imagens de tecidos mamários.

Algoritmos de aprendizado de máquina de imagem vem demonstrando grande potencial como aliado na medicina em tarefas de detecção, segmentação, diagnóstico e predição de chances de sobrevivência (B. MARLEEN, 2016), sendo amplamente utilizados em diagnósticos de câncer de mama (P. DEEPAK, B. SUYASH, K. UTKU; 2022).

Esses modelos comumente exigem imagens histopatológicas, que são obtidas a partir da análise microscópica de tecidos que foram removidos por biópsia ou cirurgia. Para tornar as células e suas estruturas mais visíveis, as amostras são coloridas usando técnicas como a coloração de Hematoxilina e Eosina (H&E). A hematoxilina tinge os núcleos das células em azul ou roxo, enquanto a eosina tinge o citoplasma e outros componentes em rosa.

Neve em cima

Descrição gerada automaticamente com confiança média Tecido com flores

Descrição gerada automaticamente com confiança média

Fig.1. Duas imagens histopatológicas; à esquerda, uma abnormalidade benigna. À direita, um carcinoma in-situ. (Z. XIAOMIN, L. CHEN, R. MAMUNUR et. Al; 2020).

Nós nos utilizaremos majoritariamente dos datasets, amplamente conhecidos e utilizados: Wisconsin breast cancer database (WBCD) para análise, e Cancer genome atlas breast invasive carcinoma (TCGA-BRCA) para classificação com imagem.

### 2.2 Pré-Processamento

Essas imagens precisam, antes de servir como entrada para os modelos, serem transformadas em conjuntos de dados mais informados. São necessários métodos de pré-processamento para a extração desses recursos, e para essa, é possível utilizar-se de soluções aproximadas matematicamente, como com o uso de funções histo-sigmoidianas (V. ANJI, S. BADAL, K. SUDHEER; 2020) para segmentação, e métodos probabilísticos como o uso de CNNs para extração de recursos (B. MARLEEN, 2016).

### 2.3 Estado-da-Arte

Dentre os métodos mais utilizados para tarefas de classificação de imagens de tumores, estão, desde classificadores mais antigos como Naïve Bayes, SVM, Bi-Clustering (K. IGOR, 2000), que seguem até os dias de hoje como benchmark para novas soluções, e mais atuais como HA-BiRNN e DNNs (V. ANJI, S. BADAL, K. SUDHEER; 2020).

Um dos mais bem sucedidos e amplamente usados atualmente, são os algoritmos de classificação com uso de DNNs. Esses modelos de aprendizagem profunda podem ser tanto supervisionados, necessitando dados bem anotados, para predição de resultados; quanto também não-supervisionados – não precisam de intervenção humana, descobrindo padrões escondidos nos dados e extraindo esses como recursos a serem utilizados por modelos classificadores para predição (S. MOHAMMAD, A. LAITH, S. QUSAI; 2022).

# REFERÊNCIAS BIBLIOGRÁFICA

1. World Health Organization. (2021). Cancer. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/cancer>.
2. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
3. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.