



Detecção de câncer de mama por imagem com classificador híbrido

Detección de cáncer de mama mediante imágenes con clasificador híbrido

Breast cancer detection by imaging with hybrid classifier

Joaquim Osterwald Frota Moura Filho¹, Marcelo Estevão da Silva², Kamila Amélia Sousa Gomes², Sara Danielle de Souza³, Márcio André Baima Amora⁴

1 Me., Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI), Universidade Federal do Ceará, Fortaleza (CE), Brasil.

2 Me., Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC), Universidade Federal do Ceará, Sobral (CE), Brasil.

3 Esp., Hospital Regional Norte (HRN), Sobral (CE), Brasil.

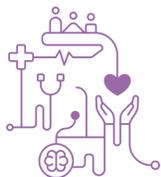
4 Prof. Dr., Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC), Universidade Federal do Ceará, Sobral (CE), Brasil.

Autor correspondente: Me. Joaquim Osterwald Frota Moura Filho
E-mail: joaquim1905@alu.ufc.br

Resumo

Objetivos: Desenvolver algoritmos de Machine Learning (ML) para classificação acurada de imagens de ultrassonografia para suporte ao diagnóstico de câncer de mama. **Método:** Implementação de um novo modelo de aprendizado híbrido que combina as técnicas de LightGBM, Rede Perceptron Multicamadas (MLP), Máquina de Vetores de Suporte (SVM) e otimização dos pesos por Enxame de Partículas Relativístico (RPSO). **Resultados:** O modelo classificador obtido resultou em acurácia de 98% nos dados de teste, oferecendo, portanto, alta acurácia. **Conclusão:** O modelo proposto obteve resultados superiores aos de trabalhos encontrados na literatura, sendo, então, uma promissora ferramenta de suporte ao diagnóstico.

Descritores: Aprendizado Híbrido; Câncer de Mama; Classificação por Imagem.



Abstract

Objectives: Develop Machine Learning (ML) algorithms for accurate classification of ultrasound images to support the diagnosis of breast cancer. **Method:** Implementation of a new hybrid learning model that combines the techniques of LightGBM, Multilayer Perceptron Network (MLP), Support Vector Machine (SVM) and Relativistic Particle Swarm weight optimization (RPSO). **Results:** The classifier model obtained resulted in an accuracy of 98% on the test data, therefore offering high accuracy. **Conclusion:** The proposed model obtained results superior to those of works found in the literature, making it a promising diagnostic support tool.

Keywords: Hybrid Learning; Breast Cancer; Image Classification.

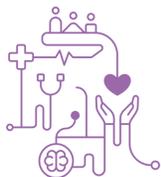
Resumen

Objetivos: Desarrollar algoritmos de aprendizaje automático (ML) para una clasificación precisa de imágenes de ultrasonido para respaldar el diagnóstico de cáncer de mama. **Método:** Implementación de un nuevo modelo de aprendizaje híbrido que combina las técnicas de LightGBM, Multilayer Perceptron Network (MLP), Support Vector Machine (SVM) y Relativistic Particle Swarm Weight Optimization (RPSO). **Resultados:** El modelo clasificador obtenido arrojó una precisión del 98% en los datos de prueba, por lo que ofrece una alta precisión. **Conclusión:** El modelo propuesto obtuvo resultados superiores a los trabajos encontrados en la literatura, lo que lo convierte en una herramienta de apoyo al diagnóstico prometedora. **Objetivos:** El cáncer de mama es una de las condiciones más prevalentes y letales entre las mujeres, con una alta tasa de mortalidad. Es esencial garantizar un diagnóstico preciso y precoz para iniciar el tratamiento en las etapas iniciales y aumentar las probabilidades de supervivencia.

Descriptores: Aprendizaje híbrido; Cáncer de mama; Clasificación de imágenes.

Introdução

Segundo o INCA (Instituto Nacional de Câncer), o câncer é uma das doenças mais graves que afetam os seres humanos. ⁽¹⁾ Caracterizado por uma replicação celular desorganizada, o câncer é uma doença originada por mutações nos genes responsáveis

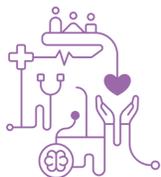


pela regulação do ciclo celular. Isso resulta em células cancerosas que demonstram capacidade de divisão independente de estímulos de crescimento, habilidade de metastatizar para outros tecidos e resistência à apoptose. ⁽²⁾ Entre as diversas formas de câncer, o de mama se destaca como um dos mais prevalentes e preocupantes, especialmente entre as mulheres, representando 30% dos novos casos de câncer em 2022 e uma taxa de mortalidade de 16,5% no mesmo ano no Brasil, tornando-o não apenas comum, mas também significativamente fatal. ⁽³⁾

Um tumor pode ser classificado em dois tipos principais: Tumor Benigno (TB) e Tumor Maligno (TM). Enquanto o TB cresce lentamente e é não cancerígeno, o TM é grave e pode se espalhar pelo corpo. No câncer de mama, a taxa de sobrevivência está associada ao tamanho do tumor, sendo quase 100% para tumores menores que 10mm. ⁽³⁾ Assim, um diagnóstico preciso e precoce desempenha um papel crucial na sobrevivência do paciente, permitindo que o tratamento seja iniciado ainda nos estágios iniciais da doença.

No diagnóstico do câncer de mama, as principais tecnologias podem ser divididas em dois tipos principais: baseadas em Imagem de Corpo e baseadas em Imagem de Micro-ondas. As técnicas de imagem de corpo incluem o Ultrassom, a Mamografia e a Imagem de Ressonância Magnética (IRM), entre outras. Além disso, no contexto do diagnóstico automático, as técnicas de *Deep Learning* (DL), especialmente as redes convolucionais (*Convolutional Neural Networks* - CNN), têm demonstrado eficácia. Apesar de demandarem grandes conjuntos de dados e alta capacidade computacional, as CNNs oferecem a vantagem de não exigir um pré-processamento extensivo dos dados. ⁽⁴⁾

Diversos trabalhos se dedicaram à tarefa de detecção de câncer de mama utilizando imagens e aprendizado de máquina. Em Raza *et al.* ⁽³⁾ um novo classificador baseado em DL foi proposto, chamado DeepBraestCancerNet, uma arquitetura de 6 camadas convolucionais, 9 módulos Inception e 1 camada totalmente conectada, que resultou uma acurácia superior a 99,3%. Já em Rehman *et al.* ⁽⁵⁾, as arquiteturas de DL utilizadas foram Mobile-NetV2 e DenseNet201, em uma abordagem de modelo *ensemble* profundo, que conferiram uma acurácia geral de 97,04%. Em Joshi *et al.* ⁽⁶⁾, uma rede de



aprendizado profundo foi proposto para classificação de câncer de mama a partir de imagens de ultrassom, utilizando uma rede CNN baseada na arquitetura da rede YOLO-v3, resultando em acurácia geral de 96,31%. Por outro lado, em Almutairi *et al.* ⁽⁷⁾, uma abordagem diferente do aprendizado profundo foi adotado, com o emprego do classificador (*EXtreme Gradient Boosting* - XGBoost) associado ao Filtro de Mediana Híbrido (FMH), resultando em acurácia geral de 97,87%. Em Manzari *et al.* ⁽⁸⁾ um novo classificador de DL chamado MedViT foi proposto, com uma proposta de tornar mais eficiente as operações de convolução, a robustez do modelo foi validado com uma acurácia geral de 89,6%. Além destes, existem trabalhos dedicados a adicionar interpretabilidade aos modelos de *Machine Learning* (ML), como o trabalho de Yahata *et al.* ⁽⁹⁾, que apresenta uma abordagem para explicabilidade de modelos preditivos de ML no câncer de mama com resultados de acurácia de 98%.

Dessa forma, visando não apenas melhorar a detecção precoce do câncer de mama, mas também tornar o processo de diagnóstico mais acessível e eficiente, contribuindo assim para melhores resultados e prognósticos para os pacientes, este trabalho propõe um novo sistema de detecção e classificação automática híbrido relativístico baseado em pesos, aprimorando a precisão e eficiência do diagnóstico. Ademais, busca-se desenvolver uma metodologia de classificação baseada em sistemas de baixo poder computacional, como uma alternativa acessível e eficaz para suporte à análise de imagens médicas.

O artigo está organizado nas seções de introdução, onde o problema de classificação é apresentado, seguido da seção de fundamentação teórica, onde estão as técnicas de classificação empregadas, na sequência há a seção de métodos, onde a metodologia é detalhada, por conseguinte, na seção de resultados estão os resultados obtidos e comparativo com a literatura, e, finalmente, na seção de conclusão estão os resultados obtidos e sugestões de trabalhos futuros.



Fundamentação teórica

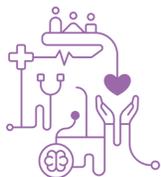
Para compreender o artigo, é necessário ter conhecimento sobre algumas técnicas desenvolvidas durante o seu desenvolvimento. Nesta seção, serão abordados os algoritmos LightGBM ⁽¹⁰⁾, Rede Neural Perceptron de Multicamadas (*Multilayer Perceptron* - MLP) ⁽¹¹⁾, Máquina de Vetores de Suporte (*Support Vectors Machine* - SVM) ⁽¹²⁾, descritos na Tabela 1, Enxame de Partículas Relativístico (*Relativistic Particle Swarm Optimization* - RPSO) ⁽¹³⁾.

Tabela 1 – Técnicas para construção dos classificadores

Algoritmo	Descrição
LightGBM	Algoritmo que utiliza conjuntos de AD para estimar ou classificar um banco de dados. Implementação mais eficiente do que o <i>Gradient Boosting</i> e o <i>XGBoost</i> , especialmente em cenários com grandes volumes de dados e alta dimensionalidade. ⁽¹⁰⁾
MLP	Algoritmo composto por neurônios capazes de processar valores e detectar padrões nos dados. Sua estrutura inclui camadas de entrada, oculta e de saída, onde a camada de entrada transfere os dados, a camada oculta processa as informações e a camada de saída apresenta os resultados. ⁽¹¹⁾
SVM	.Algoritmo que busca encontrar o melhor hiperplano para separar as regiões de diferentes classes em um conjunto de dados. Utilizado para classificação binária, o SVM pode ser estendido para sistemas multiclases por meio da estratégia um contra um, que envolve a criação de múltiplas SVMs, cada uma responsável por uma classificação binária específica. ⁽¹²⁾

Enxame de Partículas Relativístico (RPSO)

RPSO ⁽¹³⁾ é uma meta-heurística inspirada no comportamento de enxames de partículas. Essa abordagem busca encontrar soluções ótimas para problemas complexos por meio do movimento iterativo de "partículas" em um espaço de busca multidimensional. O RPSO se destaca por incorporar conceitos da teoria da relatividade de Einstein, como a distorção espaço-temporal, nas equações que governam o movimento das partículas. Essa abordagem permite que as partículas atualizem sua posição e velocidade conforme o momento relativístico, parâmetros cognitivos e sociais, além de considerar a melhor posição encontrada até o momento. Durante o processo de otimização, as partículas colaboram e competem entre si, explorando o espaço de busca de forma eficiente em



busca da solução ótima para o problema em questão. O RPSO tem sido amplamente aplicado em uma variedade de áreas, incluindo otimização de funções matemáticas, problemas de engenharia, aprendizado de máquina e design de sistemas complexos, demonstrando sua eficácia e versatilidade como uma técnica de otimização robusta e poderosa. ^(13, 14)

Métodos

O modelo proposto é baseado na combinação de técnicas avançadas de aprendizado de máquina, como LightGBM, MLP e SVM, com a otimização sendo feita pelo RPSO. Essa abordagem visa aumentar a eficiência e a precisão na detecção e classificação do câncer de mama.

Ambiente e Ferramentas de Desenvolvimento

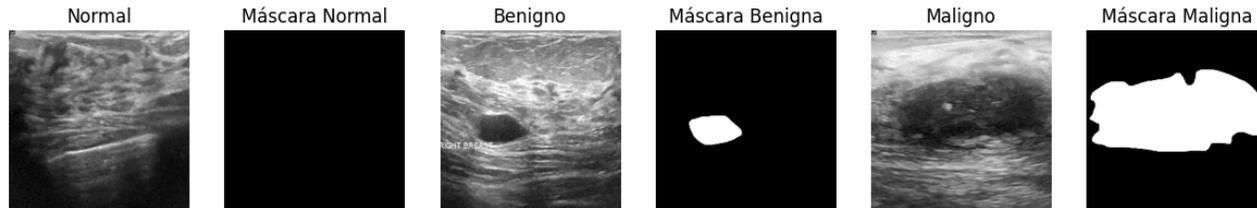
O projeto foi realizado empregando o uso da linguagem de programação Python dentro da plataforma *Google Colab*. As bibliotecas utilizadas na criação do código foram: Scikit-Learn ⁽¹⁵⁾, Skimage ⁽¹⁶⁾, Scipy ⁽¹⁷⁾, Pandas ⁽¹⁸⁾, Numpy ⁽¹⁹⁾, LightGBM ⁽¹⁰⁾, Matplotlib ⁽²⁰⁾, Seaborn ⁽²⁰⁾ e Opytimizer ⁽²¹⁾.

Base de Dados

O banco de dados utilizado neste estudo sobre câncer de mama está disponível em Al-Dhabyani *et al.* ⁽²²⁾, essas imagens de ultrassom mamário foram obtidas de mulheres com idades variadas, entre 25 e 75 anos.

As imagens são divididas em três categorias distintas - imagens normais, benignas e malignas - esses dados oferecem uma representação abrangente das diferentes condições que podem ser observadas durante o exame de ultrassom. Na Figura 1 são mostrados exemplos das imagens disponibilizadas pelo banco de dados.

Figura 1 – Exemplo das imagens do banco de dados ⁽²³⁾



Essa coleta foi realizada no ano de 2018 e envolveu um grupo significativo de 600 pacientes do sexo feminino. Com um total de 780 imagens disponíveis, cada uma com uma resolução média de 500x500 pixels e armazenadas no formato PNG. Esse conjunto de dados serve como base crucial para a pesquisa em detecção precoce e diagnóstico preciso do câncer de mama. Cada imagem original é acompanhada de sua contraparte (máscara segmentada da imagem original), permitindo uma análise detalhada e comparativa das características identificadas nos exames de ultrassom mamário.

Pré-Processamento e Extração de Características

As imagens extraídas deste banco de dados foram previamente segmentadas pelo autor, cujo processo facilitou significativamente a etapa de extração de características das mesmas.

Para a etapa de extração, utilizou-se a biblioteca Skimage, a fim de capturar propriedades essenciais como área, perímetro, excentricidade, extensão e diâmetro de Feret das regiões segmentadas. Na Figura 3 são apresentados exemplos de como a etapa de extração foi realizada na máscara.

Figura 3 – Exemplos do processo de extração de características nas máscaras





Após a obtenção do banco de dados, contendo as características mencionadas e 780 amostras, procedeu-se com uma verificação da importância de cada atributo. Para isso, foi empregada a capacidade de técnicas baseadas em árvores de decisão, como o LightGBM, que determina o grau de relevância de cada atributo utilizando a fórmula do ganho de informação. Os resultados dessas análises estão apresentados na Tabela 2, fornecendo informações sobre quais atributos têm maior impacto na classificação das imagens.

Tabela 2 – Importância dos atributos do banco de dados

Atributo	Importância
Excentricidade	2053
Extensão	1962
Área	1393
Perímetro	1161
Diâmetro de Feret	797

Como mostrado na Tabela 2, a excentricidade foi identificada como o atributo mais importante, com uma importância de 2053, enquanto o diâmetro de Feret foi considerado o menos importante, com uma pontuação de apenas 797.

A excentricidade é uma medida que descreve o quão longe um objeto é de ser perfeitamente circular. Por outro lado, o diâmetro de Feret é uma medida que representa a maior distância entre dois pontos ao longo de um objeto.

Treinamento

Os modelos LightGBM, SVM e MLP são combinados utilizando um processo de votação por média ponderada, onde os pesos são determinados utilizando a meta-heurística RPSO. O conjunto de dados foi dividido em 80% para treinamento com validação cruzada de 5 *folds* e 20% para teste, com otimização de hiperparâmetros e com ajuste de pesos via RPSO. Esses pesos variam entre 0 e 1, representando a contribuição relativa de cada modelo na previsão final. Os pesos finais são determinados calculando a média dos pesos encontrados em cada *fold*. Essa abordagem permite uma combinação adaptativa dos diferentes modelos, ajustando seus pesos conforme o desempenho.



Métricas de Desempenho

Neste trabalho, as métricas empregadas são: acurácia (Acc), precisão (P), F-score e sensibilidade (S) que analisam relações entre Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (NP) e Falsos Negativos (FP) e podem ser descritos segundo as equações (1)-(4).

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

$$P = \frac{VP}{VP + FP} \quad (2)$$

$$S = \frac{VP}{VP + FN} \quad (3)$$

$$F - Score = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (4)$$

Resultados e Discussão

Após a otimização dos hiperparâmetros foram construídos três algoritmos de aprendizado de máquina: LightGBM, SVM e MLP. Para a MLP, foram definidas duas camadas, cada uma com 100 neurônios, utilizando a função de ativação logística, e o número total de iterações foi de 9795. No algoritmo de otimização RPSO, os parâmetros foram ajustados com um $c1 = c2 = 1,7$. Para a máquina de vetor de suporte, o kernel escolhido foi a função de base radial, onde o valor do parâmetro gamma foi gerado de forma escalar, e o parâmetro C foi definido como 5000. Já para o LightGBM, o modelo foi configurado com 559 estimadores, 217 folhas, uma profundidade de 233 e uma taxa de aprendizagem de 0,01.

Os resultados obtidos para os classificadores individuais e o classificador combinados desenvolvidos a partir da validação cruzada são apresentados na Tabela 2.

Tabela 2 – Resultados para a validação cruzada de 5 *folds*.

Modelo	Acurácia Média	Desvio Padrão
LightGBM	0,931058	0,016640
MLP	0,959948	0,008717
SVM	0,969574	0,010578
Votação por média - RPSO	0,975974	0,005018



Os resultados apresentados na Tabela 3 destacam que o algoritmo de votação ponderada em conjunto com RPSO obteve a melhor acurácia média. Além disso, esse modelo também demonstrou o menor desvio padrão entre as técnicas analisadas, indicando uma consistência mais robusta em suas previsões.

Para avaliação dos classificadores no conjunto de teste, foram escolhidos os melhores modelos para cada classificador individual durante o processo de validação cruzada. Já o modelo combinado foi construído utilizando a média dos pesos determinados em cada *fold* da validação cruzada, resultando nos pesos de 0,34 para o LightGBM, 0,30 para a MLP e 0,68 para o SVM. Essa combinação de pesos reflete a contribuição relativa de cada modelo para a previsão final, conforme determinado pelo processo de otimização. Para uma comparação direta entre os modelos e sua capacidade de generalização para o conjunto de teste, a Tabela 3 apresenta os resultados de acurácia obtidos para cada modelo testado individualmente.

Tabela 3 – Resultados de acurácia

Modelo	Acurácia
LightGBM	0,942307
MLP	0,961538
SVM	0,974358
Votação por média - RPSO	0,980769

Como evidenciado na Tabela 3, a técnica empregada alcançou o melhor resultado de acurácia, obtendo uma taxa de classificação correta de 98,07% das amostras para o diagnóstico de câncer de mama. Esse desempenho destaca a eficácia do modelo na identificação precisa de casos positivos e negativos. Se o objetivo for minimizar o custo computacional, uma abordagem viável seria empregar apenas o SVM, que alcançou uma acurácia de 97,43%. O SVM ainda mantém uma taxa de classificação alta, tornando-se uma alternativa viável para situações onde a eficiência computacional é uma consideração importante. Para uma avaliação mais abrangente do desempenho dos algoritmos, é importante considerar outras métricas além da acurácia. Na Tabela 4 são apresentadas as métricas como precisão, sensibilidade e F-score considerando o conjunto de teste.



Tabela 4 – Resultados do algoritmo votação por média (RPSO)

Classe	Precisão (%)	Sensibilidade (%)	F-Score (%)
Maligno	92%	100%	96%
Benigno	100%	97%	98%
Normal	100%	100%	100%

Como observado na Tabela 4, o classificador por votação ponderada beneficiou-se das vantagens dos outros classificadores, obtendo resultados significativos. A combinação dos diferentes modelos logrou desempenho mais robusto, superando as limitações de cada algoritmo. Além disso, é possível notar o alto valor de acerto na classe maligno, alcançando 100% de precisão. A alta precisão na detecção de casos malignos é crucial, considerando que essa é a classe mais importante a ser diagnosticada no contexto do câncer de mama. É importante destacar que um dos problemas identificados é a possibilidade de erro do classificador em classificar elementos da classe benigno como maligno. Esses falsos positivos podem representar um desafio clínico, pois podem levar a tratamentos desnecessários ou preocupações adicionais para os pacientes.

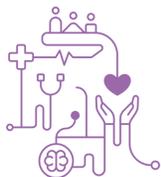
Comparativo dos resultados

Na Tabela 5 estão apresentados os principais trabalhos da literatura que se dedicaram a detecção de câncer de mama a partir das mesmas imagens do presente trabalho. Nota-se que há predominância do uso de técnicas de DL, que necessitam de maiores volumes de dados e poder computacional.

Tabela 5 – Resultados dos trabalhos analisados que usaram o mesmo banco de dados

Autor	Algoritmo	Acurácia
Raza <i>et al.</i> ⁽³⁾ (2023)	DeepBraestCancerNet DL	99,35%
Rehman <i>et al.</i> ⁽⁵⁾ (2023)	Mobile-NetV2 + DenseNet201	97,04%
Joshi <i>et al.</i> ⁽⁶⁾ (2022)	YOLO-V3	96,31%
Almutairi <i>et al.</i> ⁽⁷⁾ (2023)	XGBoost + FAM	97,87 %
Manzari <i>et al.</i> ⁽⁸⁾ (2023)	MedViT T	89,60 %
Obtido pelos autores	LightGBM+MLP+SVM+RPSO	98,07 %

Embora os modelos propostos neste trabalho tenham obtido acurácia 1,28% menor do que o maior resultado obtido da literatura, de Raza *et al.* ⁽³⁾, este trabalho possui a



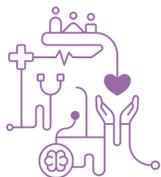
vantagem de utilizar técnicas de baixo poder computacional em comparação com as de aprendizado profundo. Rehman *et al.* ⁽⁵⁾ construiu um modelo ensemble profundo, mas obteve resultado inferior às técnicas tradicionais propostas neste trabalho. Por outro lado, Joshi *et al.* ⁽⁶⁾ com sua arquitetura profunda também não superou os resultados aqui propostos, além de possuir a desvantagem de usar considerável poder computacional. Almutairi *et al.* ⁽⁷⁾ obteve um resultado satisfatório na relação custo computacional e acurácia com sua segmentação associada ao XGBoost. Finalmente, Manzari *et al.* ⁽⁸⁾ embora tenha uma inovadora arquitetura de CNN, não conseguiu obter 90% de acurácia geral, dependendo, para isso, grande poder computacional.

Portanto, o método aqui proposto, por ser de fácil implementação e rápido treinamento em comparação com os classificadores de aprendizagem profunda, permite maior liberdade na construção de dispositivos de diagnóstico para câncer de mama, permitindo que esses algoritmos possam ser executados de forma local, embarcados nos dispositivos de suporte ao diagnóstico.

Conclusão

O presente estudo apresentou um modelo híbrido para o diagnóstico do câncer de mama, e os resultados obtidos demonstraram ser satisfatórios, evidenciando a eficiência da metodologia proposta. Em 98% dos casos, o classificador acertou na classificação dos dados, e mesmo em situações de erro, a classe normal (elemento sem nenhum tipo de tumor) não foi afetada. Além disso, em comparação com outros estudos na área, os resultados alcançados foram equivalentes, validando a eficácia do modelo proposto. Para cenários onde se busca alta eficiência com baixo custo computacional, a utilização apenas do SVM para classificar os dados se mostrou uma opção viável, oferecendo resultados satisfatórios.

Como perspectivas para trabalhos futuros, espera-se explorar e aplicar novas técnicas de aprendizado de máquina no processo de diagnóstico do câncer de mama. Isso pode incluir novas combinações de classificadores, métodos de otimização e a incorporação de dados adicionais, como informações genéticas e histopatológicas, para



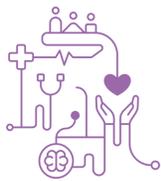
aprimorar ainda mais a precisão e eficácia do modelo de diagnóstico. Essas abordagens permitem aprimorar significativamente o diagnóstico precoce e o tratamento do câncer de mama, contribuindo para melhores resultados clínicos e uma qualidade de vida aprimorada para os pacientes.

Agradecimentos

À Agência Financiadora, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro por bolsas de pesquisa.

Referências

1. INCA. (2022). Instituto Nacional de Câncer. Recuperado de: <<https://www.gov.br/inca/pt-br/assuntos/cancer/numeros/>>. Acessado em 18 mar 2023.
2. Chamorro HM, Colturato PL, Fattori NCM. "Câncer de mama: fatores de risco e a importância da detecção precoce." *Revista Científica Eletrônica de Ciências Aplicadas*. 2021; 1 (1).
3. Raza A, et al. "DeepBreastCancerNet: A novel deep learning model for breast cancer detection using ultrasound images." *Applied Sciences*. 2023; 13(4): 2082.
4. Abdul Halim AA, Andrew AM, Mohd Yasin MN, Abd Rahman MA, Jusoh M, Veeraperumal V, Rahim HA, Illahi U, Abdul Karim MK, Scavino E. Existing and Emerging Breast Cancer Detection Technologies and Its Challenges: A Review. *Applied Sciences*. 2021; 11(22):10753.
5. Rehman MZU, et al. "An Efficient Automated Technique for Classification of Breast Cancer Using Deep Ensemble Model." *Computer Systems Science and Engineering*. 2023; 46(1).
6. Joshi RC et al. "An efficient deep neural network based abnormality detection and multi-class breast tumor classification." *Multimedia Tools and Applications*. 2022; 81(10).
7. Almutairi SM et al. "An efficient USE-Net deep learning model for cancer detection." *International Journal of Intelligent Systems* 2023..
8. Manzari ON et al. "MedViT: a robust vision transformer for generalized medical image classification." *Computers in Biology and Medicine*. 2023; 157: 106791.
9. Yahata E, Winnikow EP, Suyama R, Simões PW. Explicabilidade em Modelos Preditivos de Machine Learning no Câncer de Mama. *Journal Health Informatics*. 2023 jul; 15(1): 1-14.
10. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Anais do 31st Conference on Neural Information Processing Systems*. 2017; 3146-3154.
11. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986; 323(1): 533–536.
12. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20(3): 273-297.



13. Roder M et al. "Harnessing particle swarm optimization through relativistic velocity." 2020 IEEE Congress on Evolutionary Computation (CEC). IEEE. 2020.
14. Procópio, Fábio. "Otimização por enxames de partículas: usando uma adaptação de serendipidade". 2018.
15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Duchesnay MPE. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12: 2825-2830.
16. Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. Scikit-image: image processing in Python. PeerJ. 2014; 2: 2-18.
17. Virtanen P, Gommers R, Oliphant T.E et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020; 17: 261-272.
18. McKinney W. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference (SciPy 2010). 2010; 56-61.
19. Harris CR, Millman KJ, Walt SJ, Gommers R, Virtanen P, Cournapeau D et al. Array programming with NumPy. Nature. 2020;585: 357-362.
20. Lemenkova P. Python Libraries Matplotlib, Seaborn and Pandas for Visualization Geo-spatial Datasets Generated by QGIS. Anais do Universitatii "Alexandru Ioan Cuza" din Iasi - seria Geografie. 2020; 64(1): 13-32.
21. Rosa GH, Rodrigues D, Papa JP. Opytimizer: A Nature-Inspired Python Optimizer. Universidade de São Paulo. 2020.
22. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief. 2020; 28: 104863.