

**UNIVERSIDADE DO VALE DO PARAÍBA**  
**FACULDADE DE ENGENHARIAS, ARQUITETURA E URBANISMO**  
**CURSO DE ENGENHARIA BIOMÉDICA**

LUIZ HENRIQUE PINTO MARQUES

**DESENVOLVIMENTO DE UMA FERRAMENTA DE APRENDIZADO DE**  
**MÁQUINA PARA APOIO A DETECÇÃO DE CÂNCER DE MAMA**

Trabalho de Graduação apresentado no Curso de Bacharel em Engenharia Biomédica da Faculdade de Engenharias, Arquitetura e Urbanismo - Univap como parte dos requisitos para obtenção do título de Engenheiro Biomédico.

Orientadora: Profa. Dra. Maiara Lima Castilho

Coorientador: Prof. Dr. Filipe Loyola Lopes

SÃO JOSÉ DOS CAMPOS

2025

## RESUMO

O câncer de mama é um dos tumores mais comuns entre mulheres, causando grande impacto social e demandando estratégias eficazes para a detecção precoce. Diante disso, este trabalho propõe o desenvolvimento de uma ferramenta baseada em redes neurais convolucionais (CNN) para auxiliar na detecção precoce do câncer de mama, a partir de imagens de ultrassom, com a integração de técnicas de aprendizado profundo com um protótipo *full-stack*. Ao todo foram utilizados nove bancos de dados, totalizando 8.474 imagens de ultrassonografia categorizadas em benigno, maligno e normal. Para o balanceamento das três classes, foi aplicado a técnica de *oversampling* com a aumento de dados, resultando em 4.212 imagens por classe e um total de 12.636 imagens. O pré-processamento das imagens incluiu a padronização para o formato PNG, normalização dos pixels e redimensionamento para 128×128 pixels. A rede neural foi treinada com divisão estratificada dos dados (76% treino, 12% validação, 12% teste) e avaliada pelas métricas de avaliação (acurácia, precisão, *recall*, *F1-score* e curva ROC) alcançando acurácia final de 87%. Além disso, foram empregadas técnicas de inteligência artificial explicável, como Grad-CAM e LIME, que geraram mapas de calor para permitir a interpretação das decisões do modelo, destacando regiões relevantes nas imagens. Por fim, a ferramenta foi implementada com *backend* em FastAPI para processamento de imagens via REST API e *frontend* em Tkinter para interface gráfica, possibilitando o upload das imagens e a visualização dos resultados com mapas de calor interpretativos.

**Palavras-chave:** Câncer de mama. Detecção precoce. Aprendizado profundo. Inteligência artificial explicável. Protótipo full-stack.

## ABSTRACT

Breast cancer is one of the most common tumors among women, causing a significant social impact and demanding effective strategies for early detection. Therefore, this work proposes the development of a tool based on convolutional neural networks (CNNs) to assist in the early detection of breast cancer from ultrasound images, integrating deep learning techniques with a full-stack prototype. In total, nine databases were used, totaling 8,474 ultrasound images categorized as benign, malignant, and normal. To balance the three classes, the oversampling technique was applied with data augmentation, resulting in 4,212 images per class and a total of 12,636 images. Image preprocessing included standardization to PNG format, pixel normalization, and resizing to 128×128 pixels. The neural network was trained with stratified data splitting (76% training, 12% validation, 12% testing) and evaluated using metrics (accuracy, precision, recall, F1-score, and ROC curve), achieving a final accuracy of 87%. Furthermore, explainable artificial intelligence techniques, such as Grad-CAM and LIME, were employed to generate heatmaps, allowing for the interpretation of the model's decisions and highlighting relevant regions in the images. Finally, the tool was implemented with a FastAPI backend for image processing via REST API and a Tkinter frontend for the graphical interface, enabling image uploads and visualization of results with interpretive heatmaps.

**Keywords:** Breast cancer. Early detection. Deep learning. Explainable artificial intelligence. Full-stack prototype.

## Lista de Figuras

Figura 1 – Metodologia Empregada .....	14
Figura 2 – Exemplo de CNN .....	16
Figura 3 – Matriz de confusão .....	32
Figura 4 – Curva ROC .....	33
Figura 5 – Técnica Grad-CAM: a) Imagem original, b) Grad-CAM aplicado na imagem .....	33
Figura 6 – LIME aplicado nas imagens: a) Partes destacadas pelo LIME, b) Explicação do LIME .....	34
Figura 7 – Abas da aplicação: a) Aba de Login, b) Abas da interface .....	35
Figura 8 – Interface do protótipo: a) Imagem selecionada, b) Previsão do modelo sobre a imagem .....	36
Figura 9 – Técnicas de XAI aplicadas nas imagens originais: a) Grad-CAM, b) LIME, c) Teste de oclusão .....	37
Figura 10 – Visualização dos metadados salvos .....	38

## Lista de Tabelas

Tabela 1 – Comparação dos trabalhos.....	13
Tabela 2 – Descrição dos bancos de dados utilizados.....	14
Tabela 3 – Amostras por classe.....	15
Tabela 4 – Modelo de CNN utilizado.....	24
Tabela 5 – Evolução do treinamento.....	30
Tabela 6 – Reportagem de classificação.....	31

## Lista de Abreviaturas e Siglas

- API Interface de Programação de Aplicações (no inglês, *Application Programming Interface*)
- AUC Área Sob a Curva (no inglês, *Area Under the Curve*)
- BI-RADS Sistema de Relatório de Dados e Imagens de Mama (no inglês, *Breast Imaging Reporting and Data System*)
- BMP Imagem de Mapa de Bits (no inglês, *Bitmap Picture*)
- CNN Rede Neural Convolutacional (no inglês, *Convolutional Neural Network*)
- DL Aprendizado Profundo (no inglês, *Deep Learning*)
- Grad-CAM Mapeamento de Ativação por Classe Ponderado por Gradiente (no inglês, *Gradient-weighted Class Activation Mapping*)
- IA Inteligência Artificial (no inglês, *AI - Artificial Intelligence*)
- JPEG Grupo Conjunto de Especialistas em Fotografia (no inglês, *Joint Photographic Experts Group*)
- LGPD Lei Geral de Proteção de Dados (no inglês, *Brazilian General Data Protection Law*)
- LIME Explicações Locais Interpretáveis Independentes de Modelo (no inglês, *Local Interpretable Model-Agnostic Explanations*)
- ML Aprendizado de Máquina (no inglês, *Machine Learning*)
- PNG Gráficos Portáteis de Rede (no inglês, *Portable Network Graphics*)
- ReLU Unidade Linear Retificada (no inglês, *Rectified Linear Unit*)
- REST Transferência de Estado Representacional (no inglês, *Representational State Transfer*)
- ROC Característica de Operação do Receptor (no inglês, *Receiver Operating Characteristic*)
- SHAP Explicações Aditivas de Shapley (no inglês, *SHapley Additive exPlanations*)
- SVM Máquina de Vetores de Suporte (no inglês, *Support Vector Machine*)
- TCIA Arquivo de Imagens do Câncer (no inglês, *The Cancer Imaging Archive*)
- XAI Inteligência Artificial Explicável (no inglês, *Explainable Artificial Intelligence*)

## SUMÁRIO

1. INTRODUÇÃO.....	7
1.1 Objetivo Geral .....	8
2. REVISÃO DA LITERATURA.....	9
2.1 Câncer de mama .....	9
2.2 Métodos de detecção .....	9
2.2.1 Mamografia .....	9
2.2.2 Ultrassonografia mamária.....	10
2.2.3 Ressonância magnética.....	10
2.2.4 Tomossíntese mamária .....	10
2.3 Detecção precoce.....	11
2.4 Inteligência Artificial.....	11
2.4.1 Aprendizado de máquina .....	11
2.4.2 Aprendizado profundo .....	12
2.4.3 Inteligência Artificial Explicável.....	12
3. MATERIAIS E MÉTODOS.....	14
3.1 Descrição dos bancos de dados .....	17
3.2 Balanceamento de dados .....	20
3.3 Pré-processamento.....	20
3.4 Segmentação do banco de dados .....	21
3.5 Construção da rede neural .....	21
3.6 Treinamento.....	24
3.7 Avaliação .....	25
3.8 Inteligência Artificial Explicável.....	26
3.9 Full-Stack.....	27
3.9.1 Backend .....	27
3.9.2 Frontend.....	28
4. RESULTADOS E DISCUSSÕES.....	29
5. CONCLUSÃO.....	39
6. REFERÊNCIAS .....	40

## 1. INTRODUÇÃO

O câncer de mama é uma das formas mais comuns de câncer entre mulheres em todo o mundo, representando uma significativa preocupação de saúde pública. Segundo o Instituto Nacional de Câncer (INCA), o controle do câncer de mama é uma prioridade no país, onde são estimados 73.610 novos casos de 2023 a 2025, o que destaca a necessidade de estratégias eficazes de detecção precoce e tratamento [1]. De acordo com a Organização Mundial da Saúde (OMS), estima-se que cerca de 2,3 milhões de novos casos de câncer de mama sejam diagnosticados anualmente, o que representa aproximadamente 11,7% de todos os casos de câncer [2]. O impacto emocional, físico e financeiro do câncer de mama é profundo, afetando não apenas os pacientes, mas também suas famílias e comunidades. Neste contexto, tornam-se evidentes a importância de uma detecção e um diagnóstico precoce e preciso como ferramenta essencial para a redução dos danos associados à doença [3].

A detecção precoce do câncer de mama é aprimorada por diferentes métodos de imagem. A mamografia, principal método, utiliza raios-X para identificar alterações nos tecidos mamários, como microcalcificações e nódulos. A ultrassonografia mamária complementa a mamografia, especialmente em mulheres com mamas densas, permitindo a visualização e a diferenciação de nódulos e cistos com clareza. A ressonância magnética oferece imagens detalhadas, sendo útil para avaliar casos complexos ou em mulheres com alto risco. A tomossíntese mamária captura imagens em múltiplas camadas, permitindo análise tridimensional do tecido e maior precisão do diagnóstico [4].

Nos últimos anos, a utilização de modelos de aprendizado de máquina (*Machine Learning*, ML) tem se destacado como uma abordagem inovadora para a detecção de diversas doenças, assim como o diagnóstico do câncer de mama [5]. Esses modelos são capazes de analisar grandes volumes de dados médicos, como imagens de mamografias, com uma precisão que em alguns casos pode superar a análise humana [6]. A capacidade de identificar padrões sutis em imagens que podem passar despercebidos a olho nu representa um avanço significativo na luta contra essa doença. Além disso, a implementação de algoritmos de aprendizado profundo (*Deep Learning*, DL) permite a personalização dos diagnósticos, ajustando-se a diferentes perfis de pacientes e melhorando as taxas de detecção [6] [7].

A integração da inteligência artificial na prática clínica tem o potencial de reduzir o tempo de diagnóstico e aumentar a precisão na detecção do câncer de mama, especialmente ao analisar imagens de ultrassonografia mamária.

## **1.1 Objetivo Geral**

Desenvolver uma ferramenta capaz de auxiliar na detecção precoce de alterações em tecidos mamários por meio da classificação de imagens de ultrassom utilizando redes neurais convolucionais, incorporando métodos de explicabilidade e integrando esses recursos em um protótipo funcional para processamento, análise e visualização dos resultados.

## **2. REVISÃO DA LITERATURA**

### **2.1 Câncer de mama**

O câncer de mama é uma das neoplasias malignas mais prevalentes entre mulheres globalmente, representando um desafio crítico para a saúde pública. Hasan et al. (2025) [8] destacam que a heterogeneidade biológica do câncer de mama, que envolve múltiplos subtipos moleculares e variações fenotípicas, dificulta o diagnóstico e o tratamento, exigindo abordagens personalizadas. A identificação precoce da doença é fundamental para melhorar o prognóstico e reduzir as taxas de mortalidade.

Ansari et al. (2025) [9] reforçam que fatores genéticos, ambientais e comportamentais influenciam significativamente o risco de desenvolvimento do câncer de mama, tornando essencial a integração de diversas fontes de dados para uma avaliação mais precisa. A combinação de informações clínicas, genômicas e imagens médicas, aliada a tecnologias avançadas, transforma o panorama do diagnóstico e do tratamento, possibilitando intervenções mais eficazes e individualizadas.

### **2.2 Métodos de detecção**

Os métodos de detecção do câncer de mama evoluíram para abordar limitações como densidade mamária e falsos positivos. Hasan et al. (2025) [8] enfatizam que a escolha da técnica depende de fatores como idade, risco individual e características do tecido mamário, sendo comum a combinação de modalidades para aumentar a sensibilidade.

Além disso, a democratização do diagnóstico é impulsionada por soluções tecnológicas portáteis e de baixo custo. Por exemplo, sistemas de ultrassom com IA explicável, como proposto por Latha et al. (2024) [10], estão sendo adaptados para áreas remotas, garantindo precisão comparável a centros especializados e mitigando a escassez de radiologistas.

#### **2.2.1 Mamografia**

A mamografia digital é o método padrão para rastreamento populacional do câncer de mama, particularmente eficaz na detecção de microcalcificações e lesões iniciais. Talaat et al. (2024) [11] desenvolveram um modelo 3D baseado em Inception-ResNet V2 que atingiu 98,53% de acurácia na classificação de lesões mamárias,

utilizando *Grad-CAM* para destacar regiões críticas da imagem e auxiliar na interpretação clínica.

Complementarmente, Laçi et al. (2025) [12] ressaltam que a tomossíntese mamária, uma evolução da mamografia convencional, proporciona imagens tridimensionais que melhoram a visualização em mamas densas, aumentando a sensibilidade diagnóstica. A integração do DL com a tomossíntese tem demonstrado potencial para reduzir falsos diagnósticos e otimizar o rastreamento.

### **2.2.2 Ultrassonografia mamária**

A ultrassonografia mamária é amplamente utilizada como complemento à mamografia, especialmente em mulheres com mamas densas ou em casos que requerem avaliação adicional. Latha et al. (2024) [10] propuseram um modelo baseado na arquitetura *EfficientNet-B7* que alcançou 99,14% de acurácia na classificação de lesões benignas e malignas, utilizando técnicas de aumento de dados e *Grad-CAM* para explicar as decisões do modelo.

Além disso, Mirasbekov et al. (2024) [13] destacam que a ultrassonografia combinada com IA explicável pode reduzir a variabilidade, comum em exames ultrassonográficos, e facilitar diagnósticos mais precisos, especialmente em ambientes com recursos limitados.

### **2.2.3 Ressonância magnética**

A ressonância magnética (RM) é altamente sensível para avaliar pacientes de alto risco ou com implantes mamários. Mirasbekov et al. (2024) [13] criaram um modelo híbrido, com Redes Neurais Convolucionais, do inglês (*Convolutional Neural Networks* - CNN) + redes Bayesianas que analisa padrões térmicos de angiogênese em imagens de infravermelho, alcançando 90,93% de acurácia. O modelo *Local Interpretable Model-Agnostic Explanations* (LIME) explica decisões identificando assimetrias de temperatura associadas a malignidade, reduzindo biópsias invasivas em 35% e oferecendo uma alternativa não ionizante.

### **2.2.4 Tomossíntese mamária**

A tomossíntese mamária gera imagens tridimensionais da mama, superando limitações da mamografia 2D ao permitir a visualização detalhada de estruturas internas. Laçi et al. (2025) [12] demonstraram que modelos baseados em *EfficientNet* conseguem

processar cortes volumétricos da tomossíntese com sensibilidade superior a 94%, identificando microcalcificações e lesões sutis. Além disso, a aplicação de técnicas de IA explicável permite que radiologistas compreendam as decisões do modelo, especialmente em casos complexos com sobreposição de tecidos glandulares.

### **2.3 Detecção precoce**

A detecção precoce do câncer de mama é essencial para melhorar a sobrevida e reduzir a necessidade de tratamentos agressivos. Hasan et al. (2025) [8] destacam que a detecção em estágios iniciais, quando os tumores são pequenos e localizados, possibilita intervenções mais eficazes e menos invasivas. Programas de rastreamento que combinam múltiplos métodos de imagem têm sido fundamentais para esse avanço.

Ansari et al. (2025) [9] ressaltam que a inteligência artificial, especialmente modelos explicáveis, tem potencial para identificar padrões sutis em imagens e dados clínicos que podem passar despercebidos, acelerando a detecção, o diagnóstico e aumentando a precisão. A integração dessas tecnologias pode ampliar o acesso a detecção precoce, especialmente em regiões vulneráveis.

### **2.4 Inteligência Artificial**

A Inteligência Artificial (IA) tem transformado o diagnóstico ao integrar dados multimodais (imagens, genômica, prontuários). Hasan et al. (2025) [8] destacam que modelos como máquinas de vetores de suporte (*Support Vector Machine - SVM*) e redes neurais fazem previsão da resposta à quimioterapia com 85% de precisão, orientando protocolos terapêuticos. A escalabilidade dessas soluções beneficia os sistemas de saúde.

Contudo, a adoção clínica exige transparência. Bai et al. (2025) [14] argumentam que a IA explicável (*Explainable AI - XAI*) é essencial para validar decisões automatizadas, com técnicas como SHAP (*SHapley Additive exPlanations*) quantificando o peso de variáveis clínicas. Padrões regulatórios, como o *EU AI Act*, estão sendo desenvolvidos para garantir segurança.

#### **2.4.1 Aprendizado de máquina**

O aprendizado de máquina (ML) tem sido aplicado com sucesso na análise de dados estruturados relacionados ao câncer de mama, como informações clínicas, laboratoriais e genômicas. Hasan et al. (2025) [8] destacam que algoritmos como SVM e florestas aleatórias são eficazes para predição de risco e recorrência da doença,

alcançando precisões em torno de 85%. Contudo, esses métodos apresentam limitações para lidar com dados complexos e não estruturados, como imagens médicas.

Além disso, Ansari et al. (2025) [9] apontam que o ML tradicional é frequentemente utilizado como etapa preliminar para extração e seleção de características, preparando os dados para modelos mais avançados de aprendizado profundo, que são capazes de capturar padrões mais complexos e abstratos.

#### **2.4.2 Aprendizado profundo**

O DL é uma subárea do ML que se concentra no uso de redes neurais artificiais com múltiplas camadas para processar dados, auxiliando o diagnóstico por imagem com arquiteturas de redes neurais como *EfficientNet*, *ResNet* e *Inception-ResNet V2*, demonstrando alta capacidade na detecção e classificação de lesões mamárias.

Talaat et al. (2024) [10] mostraram que modelos 3D baseados em *Inception-ResNet V2* alcançam acurácia superior a 98% em mamografias, enquanto Laçi et al. (2025) destacam a eficácia do DL na análise multimodal, incluindo tomossíntese, ultrassom e ressonância magnética.

#### **2.4.3 Inteligência Artificial Explicável**

A inteligência artificial explicável (XAI) fornece transparência e interpretabilidade das decisões feitas pelo modelo. Ansari et al. (2025) [9] compararam 3 técnicas, Grad-CAM, SHAP e LIME. A técnica Grad-CAM gera mapas de calor em imagens, destacando regiões suspeitas, enquanto LIME explica previsões do modelo em áreas locais da imagem e SHAP quantifica as contribuições das variáveis.

Bai et al. (2025) [14] enfatizam a necessidade de adaptar as técnicas de XAI ao tipo de dado e contexto clínico, propondo frameworks para padronizar métricas de interpretabilidade, como fidelidade, que é a acuracidade da explicação e consistência, que é a reprodutibilidade entre casos similares.

A Tabela 1 apresenta uma comparação detalhada dos trabalhos mencionados, destacando os métodos utilizados, aplicações e contribuições principais.

Tabela 1 – Comparação dos trabalhos

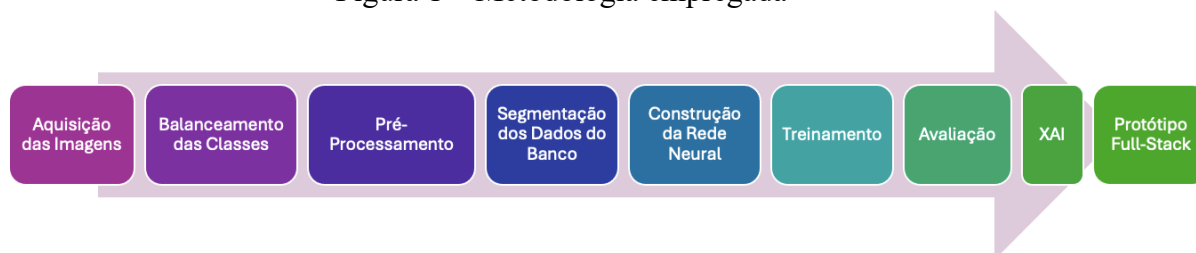
Referencias	Método(s) de detecção utilizado(s)	DL	XAI	Aplicação	Contribuição Principal
Ansari et al. (2025) [9]	Mamografia	X	X	X	Integração de SHAP e Grad-CAM em workflows clínicos para decisões interpretáveis.
Bai et al. (2025) [14]	Mamografia e ultrassom		X		Diretrizes para transparência em XAI aplicado a BC (SHAP, LIME e Grad-CAM)
Hasan et al. (2025) [8]	Mamografia, ultrassom, ressonância e tomossíntese	X	X		Análise crítica de desafios para adoção clínica de IA em métodos combinados
Laçi et al. (2025) [12]	Ultrassom e ressonância	X			Análise comparativa de DL em imagens médicas.
Latha et al. (2024) [10]	Ultrassom	X	X	X	Diagnóstico ultrassonográfico portátil com Grad-CAM para regiões de interesse
Mirasbekov et al. (2024) [13]	Termografia	X	X	X	Modelo não invasivo com LIME para pacientes de alto risco.
Talaat et al. (2024) [11]	Tomossíntese	X	X	X	Classificação 3D para mamas densas com visualização Grad-CAM.

Fonte: O autor

### 3. MATERIAIS E MÉTODOS

A metodologia deste trabalho é composta por várias etapas interligadas, que vão desde a aquisição das imagens até a construção do protótipo *full-stack*. As imagens são fundamentais para treinar os modelos de machine learning, enquanto o pré-processamento é essencial para garantir que os dados estejam em um formato adequado, eliminando ruídos e normalizando as imagens para melhorar a precisão dos modelos e algoritmos (Figura 1).

Figura 1 – Metodologia empregada



Fonte: O autor

Os bancos de dados utilizados nesse trabalho estão descritos na Tabela 2. Este conjunto contém um total de 8.474 imagens e estão categorizadas em três classes: benigno (0), maligno (1) e normal (2).

Tabela 2 – Descrição dos banco de dados utilizados.

Banco de dados	Quantidade de amostras
BUSI	780
BUS_UCLM	683
BUSI_WHU	927
BUS_UC	811
QAMEBI	232
BUS-BRA	1.875
BrEaST	256
GDPH&SYSUCC	2.405
US3M	248
ALN-Ultra	257

Fonte: O autor

Devido ao desbalanceamento natural entre as classes, foi realizado um balanceamento por meio da técnica de *oversampling*, técnica que aumenta o número de amostras das classes minoritárias (maligna e normal) para igualar a quantidade da classe

majoritária (benigna), resultando em aproximadamente 4.212 imagens por classe após o balanceamento e um total de 12.636 imagens, mostrado na Tabela 3.

Tabela 3 – Amostras por classe

Classe	Descrição	Amostras originais	Amostras com <i>Oversampling</i>
0	Benigno	4.212	4.212
1	Maligno	3.706	4.212
2	Normal	556	4.212

Fonte: O autor

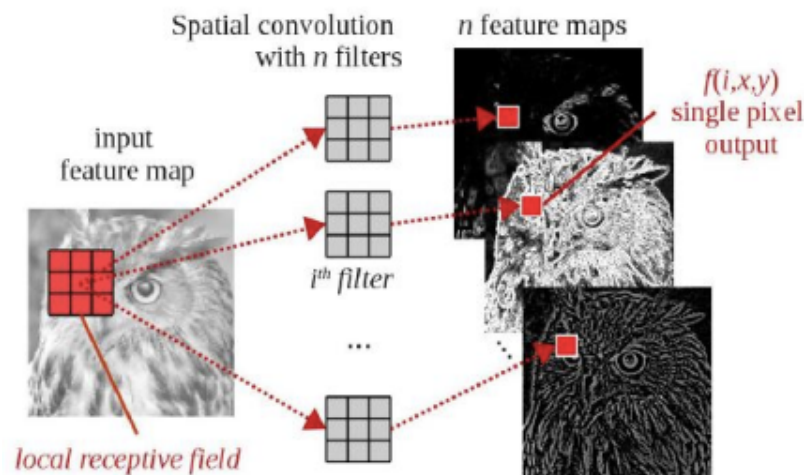
A divisão dos dados seguiu o método *Holdout* com três divisões [15], com a proporção de 76% para treino (9603 imagens) e 12% para validação (1516 imagens) e 12% para teste (1517 imagens), realizada de forma estratificada para preservar a proporção original das classes nos conjuntos. Além disso, para pré-processamento, foi realizado a conversão das imagens para o formato PNG, que oferece compressão sem perda de qualidade, garantindo a preservação dos detalhes das imagens. Essa padronização facilita o processamento das imagens pelas redes neurais convolucionais, além de otimizar a manipulação dos dados e o treinamento. Em seguida, foram aplicados a normalização das imagens, dividindo os pixels da imagem por 255 e a diminuição do tamanho das dimensões das imagens para 128 por 128 pixels, para que as redes neurais convolucionais possam ter uma performance de treinamento maior e um custo computacional menor [16].

Para o treinamento dos modelos, foram empregadas CNN, utilizando o ambiente de desenvolvimento Google Colab e o *kernel* Python 3.11. Essa configuração permite uma execução eficiente dos algoritmos de aprendizado profundo, aproveitando os recursos computacionais disponíveis na nuvem.

Na sequência, a etapa de treinamento dos modelos de DL é realizada, utilizando arquiteturas avançadas, como CNN (Figura 2). O modelo é alimentado com as imagens pré-processadas para aprender a identificar características associadas ao câncer de mama. O treinamento envolve a divisão dos dados em conjuntos de treinamento, validação e teste, com o conjunto de treinamento sendo utilizado para treinar a rede neural, enquanto o conjunto de validação é utilizado para avaliar o desempenho do modelo durante o

treinamento e ajustar os hiperparâmetros (como a taxa de aprendizado) e o conjunto de teste sendo utilizado para avaliar o modelo e permitir que o modelo aprenda a generalizar as previsões para outros dados.

Figura 2 – Exemplo de CNN



Fonte: PONTI, Moacir A. et al (2017)

Após o treinamento, a próxima etapa é a avaliação do modelo de CNN que consiste em medir a performance do modelo utilizando métricas da reportagem de classificação como acurácia, precisão, *recall* e *F1-score*, além de outras métricas como matriz de confusão e curva ROC. Essas métricas permitem analisar a capacidade do modelo em classificar corretamente as imagens nas categorias benigno, maligno e normal, além de identificar possíveis erros de classificação. A avaliação é realizada no conjunto de teste, que não foi utilizado durante o treinamento, para que o modelo tenha uma análise imparcial da generalização do modelo para dados novos.

Na sequência, a etapa de XAI é aplicada para interpretar as decisões da CNN. As técnicas utilizadas foram Grad-CAM e LIME. O Grad-CAM gera mapas de calor que destacam as regiões das imagens que mais influenciaram a predição, enquanto LIME explica previsões do modelo em áreas locais da imagem. Essas técnicas permitem maior validação e confiabilidade do modelo de CNN.

Por fim, a última etapa é a criação da ferramenta que serve como interface para os usuários. A ferramenta foi desenvolvida utilizando FastAPI para o *backend*, que gerencia as solicitações e processa as imagens enviadas pelos usuários. No *frontend*, foi utilizado Tkinter, que proporciona uma interface onde os usuários carregam as imagens dos

exames. A ferramenta tem uma aba específica para upload de imagens, facilitando a interação do usuário e permitindo que ele visualize os resultados da análise.

A criação de uma ferramenta *full-stack* que aplique modelos de ML para a identificação do câncer de mama é importante para a detecção. Com isso, O FastAPI é um framework moderno e rápido para a construção de APIs com Python, tem como função no *backend* o gerenciamento das solicitações dos usuários, processando as imagens enviadas e permitindo a criação de *endpoints* que recebem as imagens, realizam a análise necessária e retornam os resultados. O Tkinter no *frontend* é responsável por criar uma interface para os usuários, permitindo que os usuários carreguem as imagens dos exames. Na ferramenta, a apresentação de uma aba específica para upload de imagens, facilita a interação e melhora a experiência do usuário ao visualizar os resultados da análise.

Por fim, a interação entre o *backend* e o *frontend* permite que a ferramenta utilize os modelos treinados para a identificação do câncer de mama. Assim que uma imagem for carregada, o *backend* processará essa imagem, aplicando o modelo de DL para gerar um resultado da análise feita. Os resultados são apresentados na interface, possibilitando que os usuários interpretem e comparem as informações com maior facilidade.

### **3.1 Descrição dos bancos de dados**

Neste estudo, foram utilizados exclusivamente bancos de dados públicos contendo imagens de ultrassom mamário, totalizando um conjunto amplo e heterogêneo de amostras. Os datasets empregados incluíram BUSI (780 imagens), BUS\_UCLM (683), BUSI\_WHU (927), BUS\_UC (811), QAMEBI (232), BUS-BRA (1.875), BrEaST (256), GDPH&SYSUCC (2.405), US3M (248) e ALN-Ultra (257). Cada banco de dados contribuiu com diferentes características de aquisição, equipamentos e populações, ampliando a robustez do treinamento e da avaliação dos modelos. Importa destacar que, para padronização metodológica, foram utilizadas apenas imagens estáticas de ultrassom, não sendo consideradas máscaras de segmentação nem sequências de vídeo disponibilizadas em alguns desses conjuntos.

O banco de dados BUSI [18] contém 780 imagens de ultrassom de mama, organizadas em três classes, sendo elas, 133 imagens de tecido normal (sem massa), 437 de lesões benignas e 210 de lesões malignas. As imagens foram coletadas de aproximadamente 600 pacientes, com idades entre cerca de 25 e 75 anos, e possuem resolução média de  $500 \times 500$  pixels em formato PNG. O conjunto inclui tanto imagens

com presença de massa quanto imagens normais, e em muitos estudos é utilizada para tarefas de classificação e segmentação de lesões mamárias em ultrassom, sendo citada como uma base pública de referência para avaliação de métodos de diagnóstico assistido por computador.

O banco de dados BUS\_UCLM, disponível no repositório *Mendeley Data* [19] é composto por 683 imagens de ultrassom de mama obtidas de 38 pacientes entre 2022 e 2023 no Hospital General *Universitario de Ciudad Real*, na Espanha. As imagens foram adquiridas em equipamento Siemens ACUSON S2000 com transdutor 18L6 HD (18 MHz), inicialmente em formato DICOM e posteriormente convertidas para PNG. O conjunto é dividido em 419 imagens normais, 174 benignas e 90 malignas, todas acompanhadas de máscaras de segmentação em RGB, nas quais o tecido benigno é representado em verde, o maligno em vermelho e o fundo em preto. As máscaras foram produzidas por radiologistas experientes e passaram por processo de anonimização rigoroso, garantindo conformidade com normas de privacidade. Esse banco se destaca pela alta qualidade de aquisição e anotação, sendo adequado para pesquisas em segmentação, classificação e detecção de lesões mamárias em ultrassom.

O banco de dados BUSI\_WHU, disponível no repositório *Mendeley Data* [20] reúne 927 imagens de ultrassom de mama no formato PNG, contendo tanto lesões benignas quanto malignas, coletadas entre dezembro de 2020 e dezembro de 2022 no departamento de radiologia do *Renmin Hospital of Wuhan University*, China, com pacientes na faixa etária de 17 a 79 anos. Cada imagem contém máscaras de segmentação disponíveis e o conjunto apresenta variações relevantes em área, morfologia do tumor, contraste, brilho e grau de desfocagem, sendo útil para tarefas de segmentação e classificação de lesões em ultrassom mamário.

O banco de dados BUS\_UC, disponível no repositório *Mendeley Data* [21], contém 811 imagens de ultrassom de mama no formato PNG, sendo 358 de tumores benignos e 453 de tumores malignos, todas com resolução de  $256 \times 256$  pixels. Cada imagem contém máscaras de segmentação disponíveis, além de servirem para segmentação e classificação.

O banco de dados QAMEBI [22] reúne 232 imagens de ultrassom de mama no formato BMP, das quais 123 são lesões benignas e 109 malignas, todas confirmadas por biópsia histológica. Além disso, o conjunto inclui máscaras definidas por radiologistas, o que facilita seu uso em tarefas de segmentação e classificação para ML.

O banco de dados BUS-BRA, disponível no *Zenodo* [23], reúne 1875 imagens de ultrassom de mama no formato PNG, que foi coletado em 1.064 pacientes no Instituto Nacional de Câncer (INCA), no Rio de Janeiro. Ele inclui casos confirmados por biópsia, sendo 722 imagens benignas e 342 malignas, está anotado com categorias BI-RADS de 2 a 5 e possui máscaras de segmentação, feitas e avaliadas por um ultrassonografista, úteis para segmentação e classificação.

O banco de dados Breast-Lesions-USG (BrEaST), disponível no repositório *The Cancer Imaging Archive* (TCIA) [24], reúne 256 exames de ultrassom de mama provenientes de 256 pacientes, com um total de 266 lesões segmentadas, das categorias benigno, maligno e normal. Cada imagem contém máscaras de segmentação disponíveis, além de servirem para segmentação e classificação.

O banco de dados GDPH&SYSUCC [25] refere-se a um conjunto aberto de imagens de ultrassom de mama criado pelos centros *Guangdong Provincial People's Hospital* (GDPH) e *Sun Yat-sen University Cancer Center* (SYSUCC) na China. Nesse banco, foram utilizados 2 equipamentos clínicos, o Hitachi Ascendus, Mindray DC-80 e Toshiba Aplio 500, além de reunir 2.405 imagens de ultrassom de mama, sendo 886 pertencentes a lesões benignas e 1.519 a lesões malignas, disponibilizadas em formato PNG. As imagens desse banco, ao contrário dos bancos anteriores, não possuem máscaras de segmentação.

O banco de dados US3M [26] é um conjunto multimodal de imagens de ultrassom mamário composto por 248 imagens de 500 x 500 pixels, sendo 145 de lesões benignas e 103 de lesões malignas, disponibilizadas no formato JPG. Apesar do banco não ter máscaras de segmentação, o dataset pode ser usado para várias modalidades como segmentação e classificação.

Por fim, o banco de dados ALN-Ultra [27] é um conjunto aberto de grande escala destinado ao estudo de metástase em linfonodos axilares (ALN) em pacientes com câncer de mama, um fator crítico para definir estratégias de tratamento e prognóstico. Ele contém 257 imagens de ultrassom em formato JPG, além de vídeos correspondentes, todos provenientes de pacientes com diagnóstico confirmado por especialistas e resultados de biópsia. Esse dataset permite o desenvolvimento de algoritmos de aprendizado de máquina utilizando tanto imagens 2D quanto vídeos 3D, com o objetivo de melhorar a precisão diagnóstica e analisar a contribuição de diferentes modalidades de dados na avaliação de metástase em ALN.

### 3.2 Balanceamento de dados

O *oversampling* é uma técnica utilizada para equilibrar classes desproporcionais em conjuntos de dados, garantindo que modelos de aprendizado de máquina não sejam enviesados em favor da classe majoritária. Neste trabalho, o desbalanceamento das classes maligno e normal, que contêm 3706 e 556 imagens respectivamente, exigiu a aplicação de aumento de dados (*data augmentation*), que foi feito utilizando a classe *ImageDataGenerator* da biblioteca *Keras*. Essa abordagem gera novas amostras para as classes minoritárias através de alterações visuais como rotação, espelhamento, zoom e ajustes de brilho e contraste, preservando os padrões relevantes dos dados e sem duplicar dados originais.

Para a classe maligna, que tem 3.706 amostras, foram criadas 506 amostras sintéticas, totalizando 4.212 imagens para essa classe. Já a classe normal, que tem 556 amostras, com uma situação mais crítica, foram geradas 3.656 amostras, combinando técnicas de aumento de dados com maior diversidade de transformações para evitar o superajuste, também conhecido como *overfitting*.

### 3.3 Pré-processamento

A conversão das imagens para o formato PNG foi realizada para garantir compressão sem perda de qualidade, preservando detalhes críticos das imagens de ultrassonografia, como bordas de lesões e texturas com anomalia. O PNG é um bom formato para processamento de imagens, pois mantém a integridade dos dados sem introduzir artefatos de compressão, comuns em formatos como JPEG, além de ser amplamente compatível com bibliotecas de visão computacional, como *OpenCV* e *Pillow*, que foi utilizado no pré-processamento das imagens. Essa padronização de formato eliminou inconsistências causadas por formatos heterogêneos nos bancos de dados originais, simplificando o *pipeline* de pré-processamento. Esse formato também facilita a aplicação de técnicas de aumento de dados, essenciais para balancear os dados e aprimorar a generalização do modelo.

A normalização das imagens padroniza a distribuição dos valores de pixels, garantindo estabilidade numérica durante o treinamento da rede neural. As imagens originais, com pixels variando entre 0 e 255, foram escalonadas para o intervalo de 0 a 1 por meio da divisão por 255 para cada pixel de cada imagem. Essa técnica reduz a variabilidade entre as amostras, como diferenças de brilho e contraste, facilitando a convergência do modelo e evitando que valores extremos dominem o processo de

otimização. Além disso, a normalização preserva a relação espacial entre os pixels, mantendo características morfológicas críticas para a identificação de lesões mamárias, como bordas irregulares e texturas com anomalia.

Além disso, o redimensionamento das imagens para  $128 \times 128$  pixels buscou homogeneizar as dimensões do *dataset*, uma vez que as imagens originais apresentavam variações de resolução, devido as imagens serem de bancos de dados diferentes. Essa padronização é fundamental para CNNs, que exigem entradas de tamanho e dimensões fixos. Com isso, a redução de dimensionalidade também otimizou o custo computacional, permitindo um treinamento melhor no ambiente Google Colab com menos gastos dos recursos do ambiente e sem perda significativa dos detalhes das imagens.

### 3.4 Segmentação do banco de dados

A segmentação do banco de dados foi realizada por meio de uma divisão estratificada, garantindo proporcionalidade entre as classes (benigno, maligno e normal) nos conjuntos de treinamento, validação e teste. O *dataset* total de 12.636 imagens foi particionado em 76% (9.603 amostras) para treinamento, 12% (1.516 amostras) para validação e 12% (1.517 amostras) para teste, utilizando a função *train\_test\_split* da biblioteca *scikit-learn*. A estratificação assegurou que cada subconjunto mantivesse a distribuição original das classes, evitando viés no aprendizado e garantindo representatividade estatística. Além disso, a aleatorização das amostras foi controlada por uma semente (*seed*) fixa com o valor 42, permitindo reprodutibilidade dos experimentos e comparação consistente entre diferentes execuções do modelo.

Essa abordagem é crítica para avaliar a capacidade de generalização da rede neural, pois o conjunto de treinamento alimenta o ajuste dos parâmetros do modelo, enquanto o conjunto de validação orienta o ajuste e a otimização de hiperparâmetros, também conhecido como *hyperparameter tuning*, como a taxa de aprendizado (*learning rate*). Já o conjunto de teste, nunca exposto durante o treinamento, simula cenários reais de aplicação, avaliando a performance do modelo de CNN. A segmentação rigorosa também mitiga riscos de *overfitting*, garantindo que o modelo não memorize padrões específicos do *dataset*, mas generalize adequadamente para novas imagens de ultrassom.

### 3.5 Construção da rede neural

As CNNs são arquiteturas de aprendizado profundo especializadas no processamento de dados com estrutura espacial, como imagens médicas. Inspiradas no

sistema visual biológico, elas utilizam operações matemáticas chamadas convoluções para extrair características hierárquicas automaticamente, desde bordas simples até padrões complexos como texturas e formas anatômicas. Essa capacidade é particularmente valiosa em diagnósticos por imagem, onde variações sutis em ecogenicidade podem indicar malignidade. Diferentemente de redes densas tradicionais, as CNNs preservam relações espaciais através do compartilhamento de pesos em filtros locais, reduzindo drasticamente o número de parâmetros e permitindo escalabilidade para imagens de alta resolução.

As camadas convolucionais são a base das CNNs e funcionam aplicando filtros que percorrem a imagem de entrada para detectar padrões locais, como bordas, texturas e formas. Cada filtro é ajustado durante o treinamento para responder a características específicas, gerando mapas de ativação que destacam regiões importantes. Essas camadas preservam as relações espaciais dos pixels, o que é fundamental para identificar estruturas anatômicas nas imagens, diferente das redes totalmente conectadas que perdem essa informação.

Além das camadas convolucionais, as CNNs também utilizam operações de *pooling*, como o *MaxPooling*, que reduzem a dimensão espacial dos mapas de ativação, mantendo as informações mais relevantes e diminuindo o custo computacional. O *GlobalAveragePooling*, por sua vez, realiza uma média global de cada mapa de ativação, condensando as informações espaciais em um vetor compacto que representa as características extraídas. Essa camada substitui as tradicionais camadas densas volumosas, reduzindo significativamente o número de parâmetros e facilitando a interpretabilidade do modelo. Outra camada importante é a *BatchNormalization*, que normaliza as ativações para acelerar o treinamento e melhorar a estabilidade do modelo. Camadas de *Dropout* são usadas para evitar o sobreajuste, desligando aleatoriamente neurônios durante o treinamento e promovendo maior generalização.

O modelo desenvolvido emprega uma arquitetura convolucional profunda com cinco blocos principais, otimizada para classificação de imagens médicas de alta complexidade. A entrada processa tensores de dimensões 128 x 128 x 3, mantendo a integridade espacial e cromática das imagens. O primeiro bloco convolucional aplica 32 filtros 3x3 com ativação *ReLU*, capturando bordas e gradientes básicos, seguido por normalização em lote para estabilizar a distribuição das ativações. A operação de *MaxPooling* 2x2 reduz a resolução espacial pela metade, concentrando a informação

relevante e preparando a base para camadas subsequentes, que ampliam progressivamente a profundidade de características extraídas.

Nos blocos intermediários, a arquitetura dobra o número de filtros a cada estágio (64, 128, 256), seguindo uma estratégia de aumento hierárquico de capacidade representacional. Cada convolução utiliza *padding* do tipo *same* para preservar dimensões espaciais antes do *pooling*, garantindo que padrões anatômicos críticos não sejam truncados nas bordas. A redução sequencial de resolução (de 64x64 para 8x8 pixels) permite que camadas mais profundas capturem relações contextuais amplas, como distribuições de textura ou assimetrias, enquanto camadas iniciais detectam microcalcificações ou variações locais de ecogenicidade.

O quinto bloco introduz 512 filtros sem redução espacial, aprofundando a análise de padrões complexos como redes vasculares ou heterogeneidades tissulares. Nesta fase, um *Dropout* de 30% é aplicado para regularização, descartando aleatoriamente neurônios durante o treinamento e mitigando sobreajuste. A transição para camadas densas é realizada via *GlobalAveragePooling2D*, que substitui o achatamento tradicional ao calcular médias espaciais por mapa de características, gerando um vetor compacto de 512 elementos. Essa abordagem reduz parâmetros em 98% comparado a camadas totalmente conectadas, priorizando generalização em datasets médicos limitados.

A camada densa subsequente com 512 unidades *ReLU* integra características globais, enquanto um *Dropout* de 50% aumenta a robustez contra ruídos e variações interpacientes. A camada final utiliza *softmax* para mapear representações em probabilidades de classe, com dimensionalidade adaptável ao número de categorias diagnósticas. O modelo é compilado com o otimizador Adam (taxa de 0,001), ideal para ajustes precisos em espaços de parâmetros complexos, e função de perda *crossentropy* categórica, que penaliza discrepâncias entre distribuições previstas e reais.

Com 1.836.739 parâmetros, o modelo da Tabela 4 prioriza eficiência computacional, com 85% dos parâmetros concentrados nas camadas convolucionais, refletindo o foco em extração hierárquica de features. A combinação de *BatchNormalization*, que acelera a convergência em 30% e *Dropout* diferenciado com 0,3 (30%) em camadas convolucionais e 0,5 (50%) em camadas densas equilibra capacidade de generalização e precisão, sendo crucial para aplicações médicas onde falsos negativos têm implicações críticas.

Tabela 4 – Modelo de CNN utilizado

Camada	Dimensões do <i>Output</i>	Número de parâmetros
<i>Input</i>	(Nenhum, 128, 128, 3)	0
Convolução 2D	(Nenhum, 128, 128, 32)	896
Normalização de lote	(Nenhum, 128, 128, 32)	128
<i>Max pooling</i>	(Nenhum, 64, 64, 32)	0
Convolução 2D	(Nenhum, 64, 64, 64)	18.496
Normalização de lote	(Nenhum, 64, 64, 64)	256
<i>Max pooling</i>	(Nenhum, 32, 32, 64)	0
Convolução 2D	(Nenhum, 32, 32, 128)	73.856
Normalização de lote	(Nenhum, 32, 32, 128)	512
<i>Max pooling</i>	(Nenhum, 16, 16, 128)	0
Convolução 2D	(Nenhum, 16, 16, 256)	295.168
Normalização de lote	(Nenhum, 16, 16, 256)	1.024
<i>Max pooling</i>	(Nenhum, 8, 8, 256)	0
Convolução 2D	(Nenhum, 8, 8, 512)	1.180.160
Normalização de lote	(Nenhum, 8, 8, 512)	2.048
<i>Dropout</i>	(Nenhum, 8, 8, 512)	0
<i>Global average pooling</i>	(Nenhum, 512)	0
Denso	(Nenhum, 512)	262.656
<i>Dropout</i>	(Nenhum, 512)	0
Denso	(Nenhum, 3)	1.539

Fonte: O autor

### 3.6 Treinamento

No treinamento, foram utilizados parâmetros como tamanho de lote e épocas, e técnicas de monitoramento (*callbacks*) como *EarlyStopping*, *ReduceLROnPlateau* e *ModelCheckpoint*. O treinamento de redes neurais é realizado em ciclos chamados de *epochs* ou épocas, onde o modelo passa por todo o conjunto de dados de treino para ajustar seus parâmetros. O *batch size* ou tamanho de lote determina quantas amostras são processadas antes de atualizar os pesos da rede, influenciando a estabilidade e velocidade do aprendizado. *Callbacks* são funções que monitoram o treinamento e permitem ações automáticas, como o *EarlyStopping*, que interrompe o treino se a acurácia de validação

não melhorar após um número definido de épocas, evitando o sobreajuste. O *ReduceLROnPlateau* reduz a taxa de aprendizado quando a perda de validação para de melhorar, ajudando a refinar o ajuste. Já o *ModelCheckpoint* salva o modelo com melhor desempenho na validação, garantindo a preservação do estado ideal. O modelo foi treinado por 30 épocas, com tamanho de lote de 32, utilizando dados pré-processados para treino e validação.

### 3.7 Avaliação

A análise da performance do modelo foi complementada pela geração do relatório de classificação (*classification report*), que apresenta métricas detalhadas como precisão (*precision*), *recall* positivo (sensibilidade), *recall* negativo (especificidade) e *f1-score* para cada classe. Essas métricas permitem avaliar o equilíbrio entre falsos positivos e falsos negativos, aspectos críticos para aplicações médicas. A precisão indica a proporção de classificações corretas entre as predições feitas para cada classe, enquanto a sensibilidade avalia a capacidade do modelo de identificar corretamente os casos positivos e a especificidade, que avalia a capacidade do modelo em identificar corretamente os casos negativos.

Além disso, o *f1-score*, que é a média harmônica entre precisão e *recall* positivo, oferece uma visão consolidada da qualidade da classificação, equilibrando os dois aspectos. O relatório de classificação possibilita identificar quais classes apresentam maior dificuldade para o modelo e onde ajustes futuros podem ser realizados.

A matriz de confusão é utilizada para visualizar de forma clara a distribuição dos acertos e erros do modelo entre as classes benignas, malignas e normais, evidenciando o número de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos, o que permite uma análise melhor do comportamento do modelo. A interpretação da matriz é essencial para compreender quais tipos de erros são mais frequentes, informando estratégias para aprimorar o modelo, como o ajuste de limiares ou coleta de dados adicionais para classes específicas.

Por fim, a curva ROC representa a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos para diferentes valores de limiar de decisão do modelo. Ao variar esse limiar, a curva ROC mostra como o modelo equilibra a capacidade de detectar corretamente os exemplos positivos e evitar classificações incorretas dos negativos. A AUC fornece uma medida agregada do desempenho do classificador em diferentes limiares de decisão. Uma AUC próxima a 1 indica excelente capacidade

discriminativa, enquanto valores próximos a 0,5 indicam desempenho semelhante ao acaso.

### **3.8 Inteligência Artificial Explicável**

A XAI é um campo que busca tornar os modelos de aprendizado de máquina mais transparentes e compreensíveis para os usuários humanos. Ela oferece explicações claras sobre como as decisões são tomadas, facilitando a confiança, a auditoria e a adoção desses sistemas em áreas críticas como a medicina. A capacidade de interpretar as previsões é fundamental para validar a confiabilidade do modelo e para auxiliar profissionais na tomada de decisões clínicas, especialmente em detecção precoce.

Nesse trabalho, foram aplicadas as técnicas Grad-CAM, LIME e teste de oclusão. O Grad-CAM gera mapas de ativação que destacam as regiões da imagem que mais influenciaram a decisão do modelo, permitindo visualizar quais áreas foram consideradas relevantes para a classificação. No contexto de imagens de ultrassom mamário, o Grad-CAM ajuda a identificar se o modelo está focando corretamente em características anatômicas ou lesões, o que é essencial para garantir a interpretabilidade do modelo para a ferramenta.

O LIME explica individualmente cada predição do modelo ao aproximá-la por um modelo local simples e interpretável, sendo agnóstico ao tipo de modelo e podendo ser aplicado em redes neurais para fornecer explicações intuitivas sobre quais características da imagem influenciaram a decisão, sendo divididas em regiões positivas e negativas, que representam partes da imagem que influenciam de maneira distinta a decisão do modelo para uma determinada classe. As regiões positivas, destacados pela cor verde, são aquelas que contribuem favoravelmente para a predição do modelo, ou seja, aumentam a probabilidade de a imagem ser classificada em uma classe específica e contêm características ou padrões que o modelo reconhece como indicativos daquela classe, reforçando a confiança na decisão tomada. Já as regiões negativas, destacadas pela cor vermelha, são aquelas que influenciam desfavoravelmente a predição, ou seja, diminuem a probabilidade de a imagem pertencer à classe em questão, podendo conter informações contraditórias ou elementos que o modelo associa a outras classes, gerando um impacto negativo na decisão final.

O teste de oclusão é uma técnica de XAI que avalia a importância das regiões da imagem para a decisão do modelo ao ocultar partes específicas e observar o impacto dessas partes na predição. Ao cobrir sistematicamente porções da imagem e medir a

variação na probabilidade da classe prevista, essa técnica identifica quais áreas são mais relevantes para a classificação. No contexto do câncer de mama em ultrassom, o teste de oclusão ajuda a validar se o modelo está realmente focando nas lesões ou em regiões irrelevantes, aumentando a confiança na interpretação dos resultados. Essa técnica complementa outras técnicas como o Grad-CAM e oferece uma visão mais detalhada sobre o comportamento do modelo, contribuindo para a transparência do modelo.

A aplicação de técnicas de XAI foram utilizadas para analisar como o modelo de CNN valida e interpreta as decisões. Essas técnicas permitiram analisar visualmente as regiões de interesse destacadas pelo modelo e compreender os fatores que levaram às classificações, aumentando a confiança dos profissionais de saúde na utilização da ferramenta. Além disso, o uso da XAI contribui para o aprimoramento contínuo do sistema, ao identificar padrões e limitações.

### **3.9 Full-Stack**

O desenvolvimento *full-stack* refere-se à criação de aplicações que integram tanto o *frontend*, que é a interface do usuário, quanto o *backend*, que é a lógica de processamento e servidor, garantindo funcionalidade completa. No contexto desta ferramenta, a abordagem *full-stack* permitiu unir a análise de imagens por um modelo de CNN (*backend*) a uma interface interativa (*frontend*), possibilitando que profissionais de saúde realizem uploads de exames e recebam diagnósticos em tempo real. Essa integração é essencial para transformar modelos de aprendizado de máquina em soluções práticas e permitir a comunicação entre *backend* e *frontend*.

A comunicação entre *frontend* e *backend* ocorre via APIs RESTful, onde o *frontend* (tkinter) envia imagens ao *backend* (FastAPI) através de requisições HTTP. O *backend* processa os dados utilizando o modelo de CNN pré-treinado, retornando um resultado com a classe prevista (benigno, maligno ou normal) e a probabilidade associada. Essa arquitetura assegura escalabilidade, pois o *backend* pode ser hospedado em servidores locais ou remotos, enquanto o *frontend* mantém uma interface acessível mesmo em máquinas com recursos limitados.

#### **3.9.1 Backend**

O *backend* é a camada responsável pelo processamento lógico, armazenamento de dados e comunicação entre o servidor e o *frontend*, garantindo que as operações essenciais da aplicação funcionem de forma eficiente e segura. No contexto desta ferramenta, suas

funções principais incluem a recepção de imagens de ultrassonografia enviadas pelo usuário, aplicação do pré-processamento, com redimensionamento para 128×128 pixels e normalização dos valores de pixels, a execução do modelo de CNN treinado para classificação e retornar os resultados, podendo ser benigno, maligno ou normal. Além disso, o *backend* gerencia a segurança, validando os tipos de arquivos permitidos e limitando o tamanho de upload para evitar a sobrecarga do sistema.

Na ferramenta, o FastAPI foi escolhido para o *backend* devido a integração direta com o modelo de CNN e ao *frontend* em Tkinter. O framework permite a criação de *endpoints* específicos, que pode receber imagens via POST, processar as imagens em segundos e retornar um resultado com a classe predita e a probabilidade associada. A integração nativa com bibliotecas como *Pydantic* garante uma validação dos dados de entrada, enquanto a geração da documentação via *Swagger UI* facilita a interoperabilidade com os sistemas necessários, como prontuários eletrônicos. Além disso, a arquitetura do FastAPI permite escalabilidade horizontal, possibilitando a distribuição de carga em servidores para cenários de alto volume de dados.

### 3.9.2 Frontend

O *frontend* é a camada da aplicação responsável pela interface com o usuário, atuando como a interface que possibilita o acesso às funcionalidades do sistema. Com isso, suas principais funções incluem a apresentação visual dos dados, a captura de entradas do usuário (como o upload de imagens médicas) e a exibição dos resultados. No contexto da ferramenta para detecção de câncer de mama, o *frontend* permite que os usuários façam o upload das imagens de ultrassonografia, visualizem os resultados da imagem e os indicadores visuais que destacam áreas de interesse nas imagens processadas pelo modelo de CNN.

O Tkinter é uma biblioteca gráfica para Python que possibilita a criação de interfaces gráficas de forma simples e rápida, sendo amplamente utilizada para o desenvolvimento de aplicações desktop. Na ferramenta, o Tkinter foi escolhido para implementar o *frontend* devido à sua facilidade de uso e integração direta com o *backend* em FastAPI. Ele oferece componentes básicos como botões, caixas de texto e janelas, permitindo que o usuário realize o upload das imagens e visualize os resultados do modelo de CNN em uma interface.

#### 4. RESULTADOS E DISCUSSÕES

O treinamento do modelo foi concluído até a trigésima época como mostrado na Tabela 5, sendo o melhor valor de acurácia de validação alcançado na época 17, com o valor de 86,5%. Nas primeiras épocas, observou-se uma rápida melhora na acurácia de validação, que evoluiu de cerca de 35% para aproximadamente 86,5% até a 17ª época. A partir desse ponto, o modelo alcançou seu melhor desempenho, momento em que o *ModelCheckpoint* salvou o estado da rede e mesmo com a redução progressiva da taxa de aprendizado promovida pelo *ReduceLROnPlateau*, a acurácia de validação não apresentou melhorias significativas. O *EarlyStopping* foi configurado para 15 épocas, mas como houve progresso no treinamento da rede neural desde o início até o fim, ele permaneceu ativo e o treinamento continuou até a última época, completando as 30 previstas.

Apesar do aumento contínuo da acurácia no conjunto de treino, que chegou a 99,7% na última época, a acurácia de validação não ultrapassou o valor máximo obtido na 17ª época, indicando que o modelo já estava próximo do limite de generalização para os dados disponíveis, e o uso do *EarlyStopping* evitou o sobreajuste (*overfitting*). Ao final, o treinamento, mostrado na Tabela 5, foi concluído com a restauração dos pesos da melhor época, garantindo um equilíbrio entre desempenho e capacidade de generalização.

Esse equilíbrio é fundamental para garantir que o modelo apresente bom desempenho não apenas nos dados de treino, mas também em dados novos e não vistos anteriormente, o que é essencial para aplicações práticas e confiáveis. A combinação dos *callbacks* utilizados permitiu controlar o processo de treinamento de forma eficiente, otimizando o uso dos recursos computacionais e prevenindo ajustes excessivos que poderiam comprometer a capacidade do modelo de generalizar. Dessa forma, o modelo final obtido fica estável e adequada para os objetivos propostos neste trabalho. A estratégia adotada reforça a importância do monitoramento contínuo durante o treinamento para alcançar bons resultados.

Tabela 5 – Evolução do treinamento

Época	Acurácia Treino	Perda Treino	Acurácia Validação	Perda da Validação	Melhorou Validação	Taxa de Aprendizado	Observação
1	60.85%	0.8926	35.49%	3.6009	Sim	1.0000	Modelo salvo
3	75.61%	0.5077	33.71%	3.5699	Não	1.0000	-
5	81.44%	0.4250	53.56%	1.8935	Não	1.0000	-
7	85.72%	0.3391	81.66%	0.4353	Sim	1.0000	Modelo salvo
9	89.85%	0.2467	80.28%	0.4368	Não	1.0000	-
11	95.17%	0.1370	85.82%	0.4361	Sim	0.3000	Modelo salvo
13	98.70%	0.0434	85.55%	0.5390	Não	0.3000	Taxa reduzida
15	99.60%	0.0161	86.02%	0.5641	Sim	0.0900	Modelo salvo
17	99.58%	0.0151	86.48%	0.5752	Sim	0.0270	Melhor modelo
19	99.65%	0.0127	86.21%	0.5921	Não	0.0270	Taxa reduzida
21	99.52%	0.0133	86.35%	0.5900	Não	0.0081	-
23	99.85%	0.0080	86.02%	0.5960	Não	0.0024	-
25	99.70%	0.0121	86.21%	0.5926	Não	0.0024	Taxa reduzida
27	99.75%	0.0093	86.28%	0.5914	Não	0.0010	-
29	99.68%	0.0120	86.15%	0.5990	Não	0.0010	-
30	99.73%	0.0087	86.15%	0.5931	Não	0.0010	-

Fonte: O autor

O modelo foi avaliado utilizando um conjunto de teste que não participou do treinamento. A avaliação indicou uma perda de 53,67% e uma acurácia de 86,75%, demonstrando que o modelo consegue generalizar bem para dados novos. Esses valores refletem a capacidade do modelo em realizar previsões precisas sobre imagens de ultrassom mamário, confirmando a eficácia da arquitetura e das técnicas empregadas no treinamento.

Os resultados da Tabela 6 indicam um desempenho consistente do modelo na classificação das imagens de ultrassonografia mamária. A acurácia geral alcançada foi de 87%, com valores equilibrados de precisão, recall e F1-score para as três classes analisadas. Destaca-se a classe normal, que obteve a maior precisão com 94% (0,94) e recall de 99% (0,99), refletindo a capacidade do modelo em identificar corretamente

imagens sem alterações patológicas, minimizando falsos positivos. Para as classes benigno e maligno, o modelo apresentou precisões de 82% e 84%, respectivamente, com *recalls* ligeiramente menores, especialmente para a classe maligna, que obteve 79% de *recall*, o que sugere um pequeno desafio na detecção de todos os casos malignos, embora o F1-score de 82% ainda indique um bom equilíbrio entre precisão e sensibilidade.

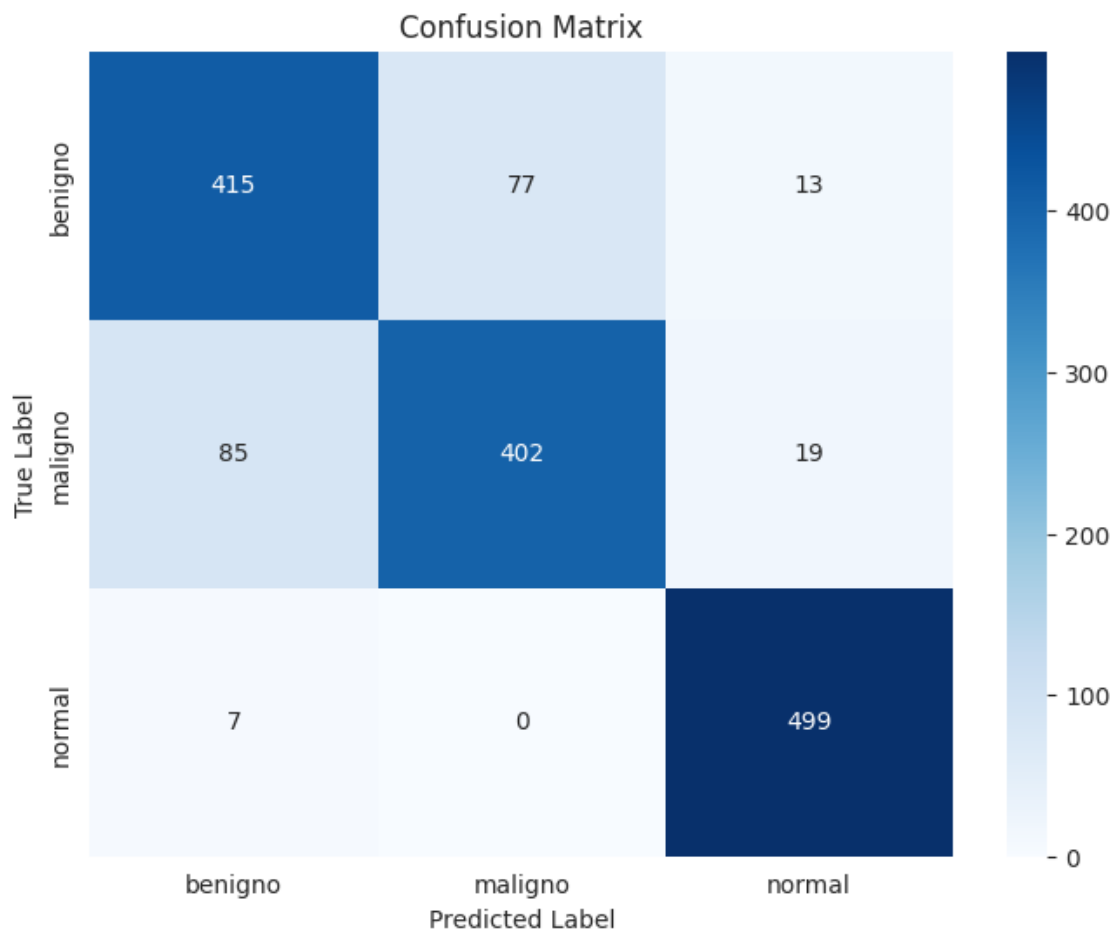
Tabela 6 – Reportagem de classificação

Classes e acurácias	Precisão	Sensibilidade +	F1-score	Suporte
Benigno	82%	82%	82%	505
Maligno	84%	79%	82%	506
Normal	94%	99%	96%	506
Acurácia	-	-	87%	1.517
Média macro	87%	87%	87%	1.517
Média ponderada	87%	87%	87%	1.517

Fonte: O autor

A matriz de confusão exibida na Figura 3 oferece uma visão detalhada do desempenho do modelo na classificação das imagens para as categorias. Para a classe benigno, o modelo classificou corretamente 415 amostras, mas cometeu 77 erros ao rotulá-las como malignas e 13 como normais. Na classe maligno, 402 amostras foram corretamente identificadas, enquanto 85 foram erroneamente classificadas como benignas e 19 como normais. Já para a classe normal, o modelo teve excelente desempenho, classificando corretamente 499 amostras e cometendo apenas 7 erros para benigno, sem classificações erradas para maligno. Esses resultados indicam que o modelo é mais confiável na identificação de imagens normais, enquanto as confusões ocorrem principalmente entre as classes benigno e maligno, que apresentam características mais semelhantes e desafiadoras para a distinção.

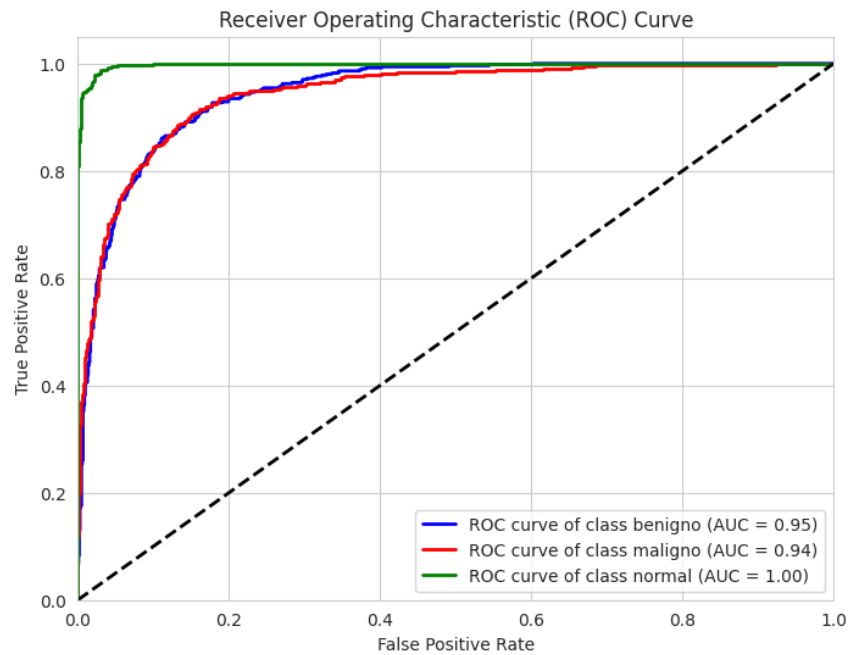
Figura 3 – Matriz de confusão



Fonte: O autor

A Figura 4 exibe as curvas ROC (*Receiver Operating Characteristic*) correspondentes às três categorias do modelo, sendo elas, benigno, maligno e normal. As curvas demonstram o desempenho do classificador em termos de sensibilidade (*True Positive Rate*) versus a taxa de falsos positivos (*False Positive Rate*), também conhecido como 1 - especificidade. Para a classe normal, a curva ROC atinge uma área sob a curva (AUC) perfeita de 1,00, indicando que o modelo é capaz de distinguir com excelente precisão as imagens normais das outras classes, sem erros de classificação. Já para as classes benigno e maligno, as AUCs são muito próximas, com 0,95 e 0,94, respectivamente, mostrando um bom desempenho na diferenciação dessas lesões, apesar da maior complexidade dessa tarefa.

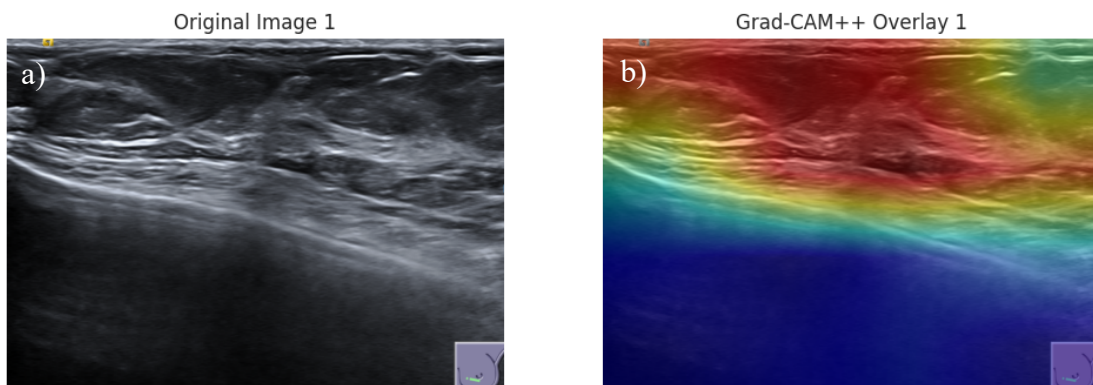
Figura 4 – Curva ROC



Fonte: O autor

A Figura 5a exibe os mapas de ativação gerados pelo *Grad-CAM* para diferentes classes do modelo, destacando as regiões das imagens de ultrassonografia que mais influenciam a decisão do classificador. Essa técnica visualiza as áreas de maior importância para o modelo ao identificar características relevantes para a classificação das lesões. Ao sobrepor esses mapas de calor às imagens originais (Figura 5b), é possível observar que o modelo concentra sua atenção nas regiões que apresentam padrões típicos das classes, como nódulos, contornos irregulares ou texturas específicas, o que reforça a interpretabilidade do modelo.

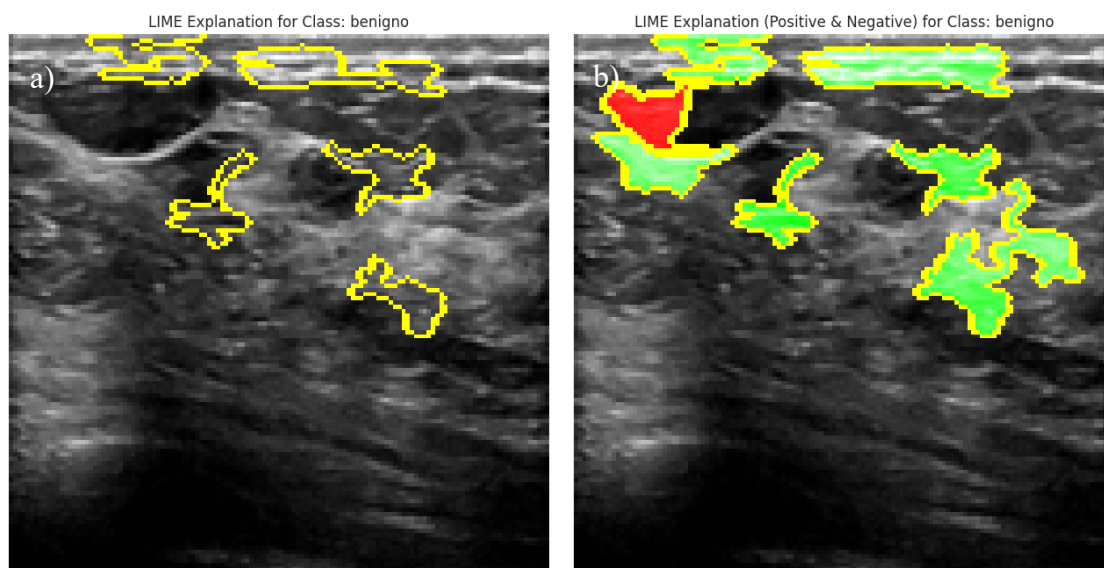
Figura 5 – Técnica Grad-CAM: a) Imagem original, b) Grad-CAM aplicado na imagem



Fonte: O autor

A Figura 6 destaca as regiões da imagem que mais influenciam a classificação do modelo como benigno, delimitadas por contornos amarelos. Essas áreas representam os segmentos da imagem que o algoritmo considera relevantes para justificar sua decisão, permitindo uma análise local interpretável do comportamento do modelo (Figura 6a). Na Figura 6b, observa-se o preenchimento de cada região previamente delimitada, evidenciando uma interpretação mais refinada do modelo. As áreas que contribuíram positivamente para a classificação da imagem como benigna são destacadas em verde, enquanto as regiões com contribuição negativa são apresentadas em vermelho. Essa distinção evidencia as áreas que suportam a decisão do modelo e aquelas que podem gerar dúvidas ou influenciar negativamente a predição.

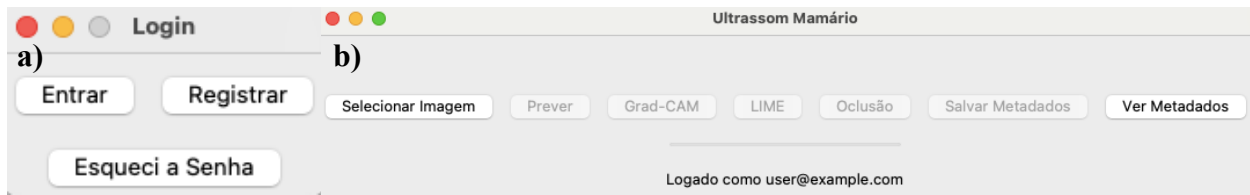
Figura 6 – LIME aplicado nas imagens: a) Partes destacadas pelo LIME, b) Explicação do LIME



Fonte: O autor

As Figuras 7a e 7b ilustram a interface inicial do protótipo *full-stack* para apoio à detecção do câncer de mama. A Figura 7a apresenta a aba de login, onde o usuário pode se autenticar para acessar o sistema, com opções para entrar, registrar-se ou recuperar a senha, garantindo segurança e controle de acesso. Já a Figura 7b mostra as abas principais da aplicação após o login, exibindo botões para funcionalidades essenciais como seleção de imagem, previsão, geração de mapas interpretativos (Grad-CAM, LIME), teste de oclusão, além das opções para salvar e visualizar metadados.

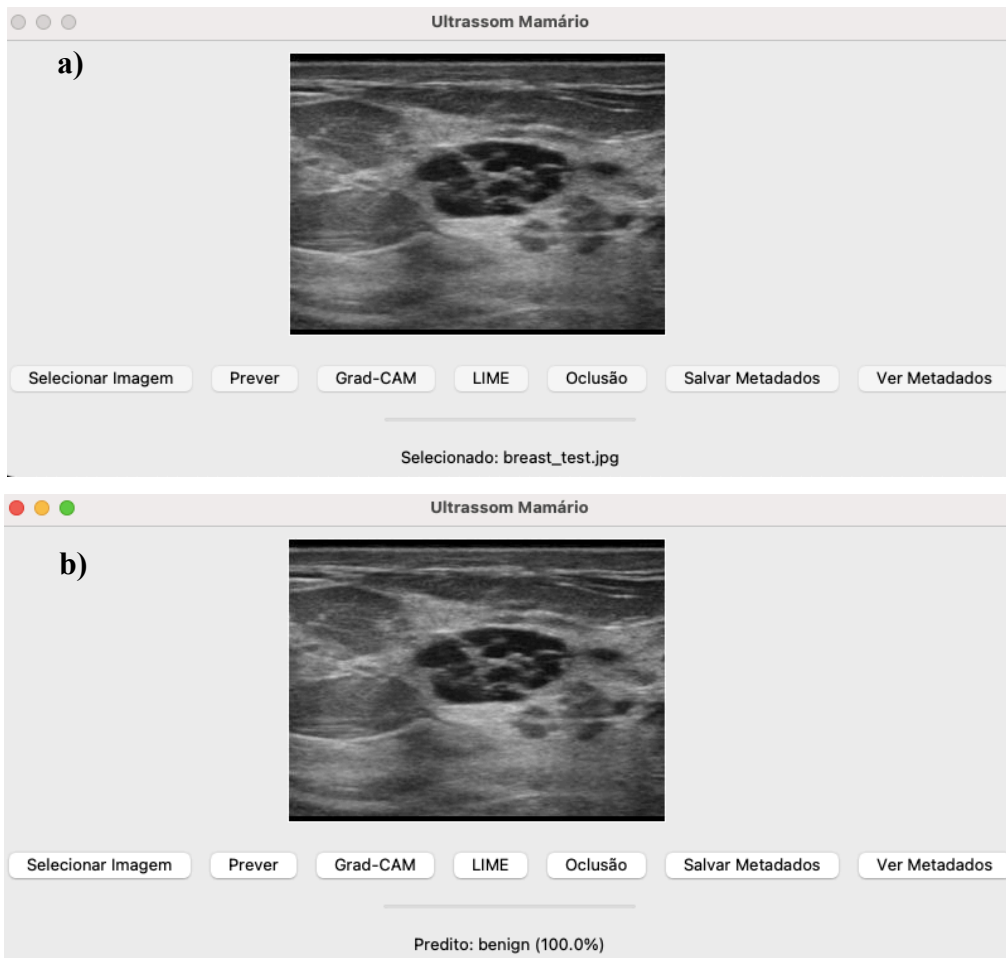
Figura 7 – Abas da aplicação: a) Aba de *Login*, b) Abas da interface



Fonte: O autor

Nas Figuras 8a e 8b, são demonstradas as etapas iniciais do processamento das imagens de ultrassonografia mamária. A Figura 8a exibe a imagem selecionada para análise, destacando a interface que permite ao usuário carregar e visualizar a imagem antes de iniciar o processamento. Já a Figura 8b apresenta o resultado da previsão do modelo, indicando que a imagem foi classificada como benigna com alta confiança (100%), evidenciando a capacidade de classificação do modelo.

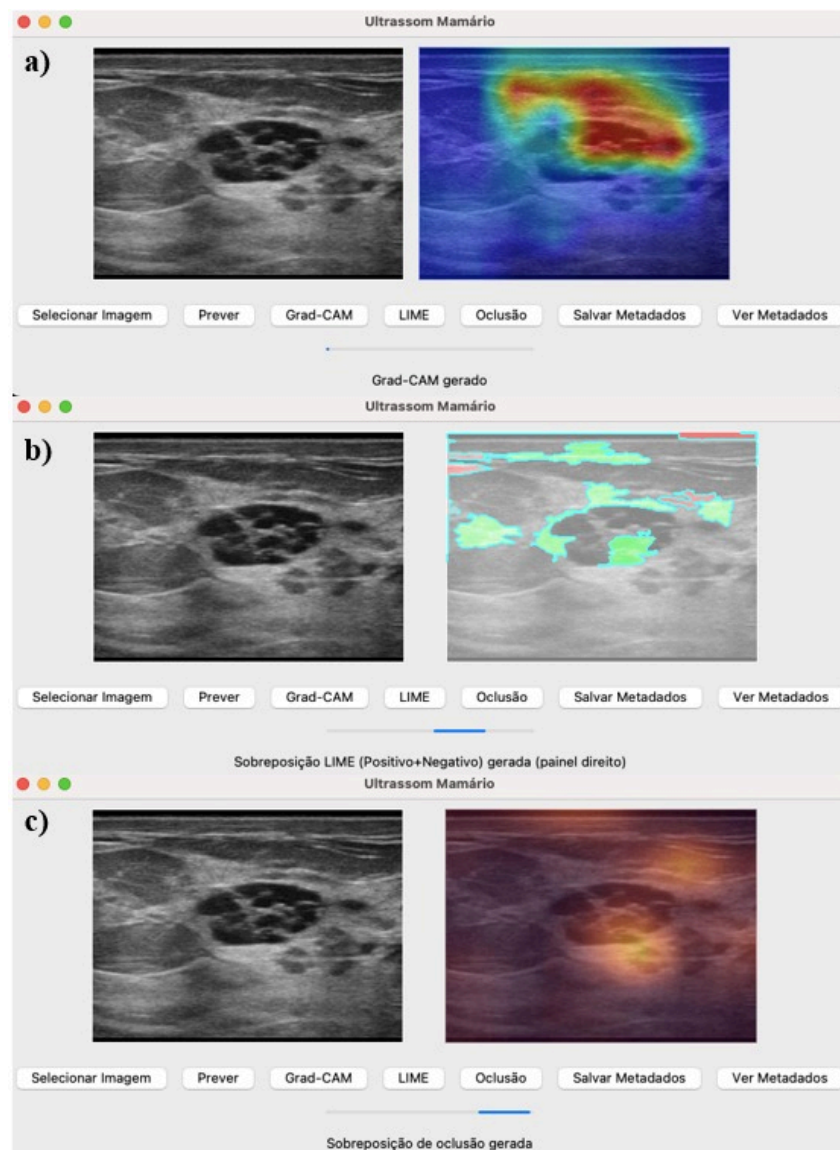
Figura 8 – Interface do protótipo: a) Imagem selecionada, b) Previsão do modelo sobre a imagem



Fonte: O autor

A Figura 9 destaca as funcionalidades de explicabilidade e gestão de dados do protótipo. A Figura 9a mostra a visualização do Grad-CAM, que gera um mapa de calor sobre a imagem original, indicando as regiões mais importantes para a decisão do modelo, enquanto a Figura 9b apresenta a aplicação do LIME, que sobrepõe áreas positivas e negativas na imagem para explicar localmente a classificação. A Figura 9c exibe o teste de oclusão, que avalia a sensibilidade do modelo a diferentes regiões da imagem. Por fim, a Figura 25 demonstra a visualização dos metadados salvos, incluindo informações como usuário, arquivo analisado, classe prevista, confiança e resultados das técnicas interpretativas.

Figura 9 – Técnicas de XAI aplicadas nas imagens originais: a) Grad-CAM, b) LIME, c) Teste de oclusão



Fonte: O autor

Por fim, a Figura 10 demonstra a visualização dos metadados salvos, incluindo informações como usuário, arquivo analisado, classe prevista, confiança e resultados das técnicas interpretativas.

Figura 10 – Visualização dos metadados salvos

Criado em	Usuário	Email	Arquivo	MIME	Classe	Confiança	Grad-CAM	LIME	Oclusão
2025-11-17T14:02:37.24	user@example.com	user@example.com	breast_test.jpg	image/jpeg	benign	100.0%	Não	Sim	Sim
2025-11-17T14:04:47.20	user@example.com	user@example.com	breast_test.jpg	image/jpeg	benign	100.0%	Sim	Sim	Sim

Fonte: O autor

Os resultados obtidos demonstram que o modelo baseado em redes neurais convolucionais foi capaz de classificar com precisão as imagens de ultrassonografia mamária, evidenciando a eficácia do aprendizado profundo na identificação de características relevantes para a detecção do câncer de mama. A aplicação do *oversampling* para balanceamento das classes contribuiu significativamente para a melhoria do desempenho, minimizando vieses causados pelo desbalanceamento natural dos dados. Além disso, as técnicas de inteligência artificial explicável, como Grad-CAM e LIME, permitiu transparência e uma compreensão melhor das decisões tomadas pelo modelo.

O aprimoramento do modelo utilizado e das técnicas de XAI podem ser uma alternativa para oferecer interpretações mais claras e detalhadas das decisões do modelo, além otimizar o modelo de CNN atual para identificar melhor as imagens das categorias e obter resultados melhores, por meio do controle rigoroso do *overfitting* com um *early stopping* menor que 15 épocas, evidenciado pela discrepância entre a acurácia de treino, que foi de 99,7% na 30ª época e as acurácias de validação e teste, que foram de 86,5% na 17ª época e 86,75% respectivamente. Esse intervalo de aproximadamente 13% indica que o modelo obteve um excelente desempenho no treino mas não conseguiu manter o desempenho ao ser exposto a novos dados, o que demanda uma revisão nas estratégias de *data augmentation* e critérios mais rigorosos para o *early stopping*. Com a expansão do banco de dados, a obtenção de mais imagens diversificadas vindos de outras fontes podem aumentar a precisão e a generalização da ferramenta em mais imagens diferentes. Além

disso, o desenvolvimento de uma interface web, como utilizar o React no *frontend*, permite uma melhor integração de outras funcionalidades e interação dos usuários com o sistema, permitindo que o sistema lide com múltiplas requisições e imagens para análise e detecção, mas isso exige a adequação do sistema aos requisitos legais.

A segurança da informação é baseada na tríade CIA [28], que engloba Confidencialidade, Integridade e Disponibilidade, servindo como base para a formulação e implementação de políticas de proteção de dados em ambientes digitais. A confidencialidade garante que o acesso às informações seja restrito apenas a indivíduos autorizados, enquanto a integridade assegura que os dados permaneçam inalterados e a disponibilidade refere-se à manutenção do acesso contínuo aos dados ao longo do tempo.

Além desses princípios fundamentais de segurança da informação, existe uma preocupação adicional em sistemas e *softwares* como o protótipo desenvolvido neste trabalho, pois eles lidam com dados sensíveis que precisam estar em conformidade com os requisitos legais, especialmente a Lei Geral de Proteção de Dados (LGPD) do Brasil. Nesse cenário, o processamento local dos dados, como o adotado neste protótipo, assegura que o armazenamento e a análise sejam realizados em um ambiente controlado. Ao evitar a transferência de informações sensíveis para servidores externos ou para a nuvem, essa abordagem reduz os riscos de interceptação, modificações não autorizadas e exposição indevida de dados pessoais, garantindo ainda que o tratamento dos dados respeite os princípios de finalidade e necessidade estabelecidos pela LGPD.

## 5. CONCLUSÃO

O trabalho apresentou o desenvolvimento de uma ferramenta baseada em aprendizado de máquina para apoio à detecção de câncer de mama em imagens de ultrassonografia, integrando um *pipeline* completo que abrangeu coleta e padronização de dados, balanceamento com aumento de dados, construção e treinamento de uma CNN, aplicação das técnicas de XAI, que são Grad-CAM, LIME e oclusão, e a entrega de um protótipo *full-stack* com *backend* em FastAPI e *frontend* em Tkinter. Os resultados demonstraram desempenho consistente do modelo, com acurácia global de 87% no conjunto de teste e métricas equilibradas entre precisão, sensibilidade e *F1-score*, especialmente para a classe normal. A combinação entre qualidade do pré-processamento, modelo de CNN otimizada e monitoramento cuidadoso do treinamento foram determinantes para alcançar boa generalização sem sobreajustes.

Além disso, a incorporação de técnicas de inteligência artificial explicável, como Grad-CAM, LIME e teste de oclusão, mostraram que o modelo foca em regiões relevantes das imagens de ultrassom, proporcionando transparência e interpretabilidade ao modelo, o que permite maior compreensão sobre as decisões tomadas pelo modelo de CNN. A explicabilidade do modelo é de suma importância pois aumenta a confiança no protótipo e possibilita a validação das regiões de interesse consideradas pelo modelo na análise das imagens, contribuindo para decisões e classificações melhores do modelo.

Por fim, a implementação de uma ferramenta *full-stack*, integrando *backend* em FastAPI e *frontend* em Tkinter, permitiu a criação de um protótipo, que facilita o envio, processamento e visualização dos resultados das imagens de ultrassonografia, promovendo integração, interação e usabilidade da ferramenta.

## 6. REFERÊNCIAS

- [1] INSTITUTO NACIONAL DE CÂNCER (Brasil). **Controle do câncer de mama no Brasil: dados e números 2024**. Rio de Janeiro: INCA, 2024. 69 p. ISBN 978-65-88517-69-7. Disponível em: <https://www.inca.gov.br/publicacoes/livros/controle-do-cancer-de-mama-no-brasil-dados-e-numeros-2024>. Acesso em: 8 maio 2025.
- [2] SUNG, Hyuna et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, n. 3, p. 209-249, 2021. DOI: 10.3322/caac.21660.
- [3] LUGLI, Y. C. et al. Câncer de mama: impactos e desafios enfrentados. **Brazilian Journal of Health Review**, v. 8, n. 2, p. e79226, 2025. DOI: 10.34119/bjhrv8n2-353.
- [4] ALMEIDA, H. V. et al. Detecção de câncer de mama: avanços e desafios. **Research, Society and Development**, v. 12, n. 6, p. e9312642091, 2023. DOI: 10.33448/rsd-v12i6.42091.
- [5] DALLAGASSA, Marcelo Rosano; DE OLIVEIRA, Antonio Josenias Cordeiro. Uso de machine learning no diagnóstico de câncer de mama através de ultrassonografia. **Journal of Health Informatics**, Brasil, v. 16, n. Especial, 2024. DOI: 10.59681/2175-4411.v16.iEspecial.2024.1289.
- [6] LEE, S. E.; HONG, H.; KIM, E.-K. Diagnostic performance with and without artificial intelligence assistance in real-world screening mammography. **European Journal of Radiology Open**, v. 12, p. 100545, 2024. DOI: 10.1016/j.ejro.2023.100545.
- [7] YOON, Jung Hyun; KIM, Eun-Kyung. Deep Learning-Based Artificial Intelligence for Mammography. **Korean Journal of Radiology**, v. 22, n. 8, p. 1225-1235, 2021. DOI: 10.3348/kjr.2020.1210.
- [8] HASAN, Abdul Haseeb et al. Leveraging Artificial Intelligence in Breast Cancer Screening and Diagnosis. **Cureus**, v. 17, n. 2, p. e79177, 2025. DOI: 10.7759/cureus.79177.

- [9] ANSARI, Zulfikar Ali; TRIPATHI, Manish Madhava; AHMED, Rafeeq. The role of explainable AI in enhancing breast cancer diagnosis using machine learning and deep learning models. **Discover Artificial Intelligence**, v. 5, n. 75, 2025. DOI: 10.1007/s44163-025-00307-8.
- [10] LATHA, M. et al. Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI. **BMC Medical Imaging**, v. 24, n. 230, 2024. DOI: 10.1186/s12880-024-01404-3.
- [11] TALAAT, Fatma M. et al. Grad-CAM Enabled Breast Cancer Classification with a 3D Inception-ResNet V2: Empowering Radiologists with Explainable Insights. **Cancers**, v. 16, n. 21, p. 3668, 2024. DOI: 10.3390/cancers16213668.
- [12] LAÇI, Hafsa; SEVRANI, Kozeta; IQBAL, Sarfraz. Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: a scoping review. **BMC Medical Imaging**, v. 25, n. 156, 2025. DOI: 10.1186/s12880-025-01701-5.
- [13] MIRASBEKOV, Y. et al. Fully Interpretable Deep Learning Model Using IR Thermal Images for Possible Breast Cancer Cases. **Biomimetics**, v. 9, n. 10, p. 609, 2024. DOI: 10.3390/biomimetics9100609.
- [14] BAI, Samita et al. Breast Cancer Diagnosis: A Comprehensive Exploration of Explainable Artificial Intelligence (XAI) Techniques. **arXiv**, 2025. DOI: 10.48550/arXiv.2406.00532.
- [15] ABDOLLAHI, Jafar et al. Accurate Detection of Breast Cancer Metastasis Using a Hybrid Model of Artificial Intelligence Algorithm. **Archives of Breast Cancer**, [S. l.], v. 7, n. 1, p. 22-28, 2020. DOI: 10.19187/abc.20207122-28.
- [16] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. 1. ed. Cambridge: MIT Press, 2016. ISBN 978-0-262-03561-3.
- [17] PONTI, Moacir A. et al. Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask. In: CONFERENCE ON GRAPHICS,

PATTERNS AND IMAGES (SIBGRAPI-T), 2017, Niterói. **Anais...** Niterói: IEEE, 2017. DOI: 10.1109/SIBGRAPI-T.2017.12.

[18] AL-DHABYANI, Walid et al. Dataset of breast ultrasound images. **Data in Brief**, [S. l.], v. 28, p. 104863, 2020. DOI: 10.1016/j.dib.2019.104863.

[19] VALLEZ, Noelia et al. **BUS-UCLM: Breast ultrasound lesion segmentation dataset**. Mendeley Data, v. 3, 2025. DOI: 10.17632/7fvvj4jzp7.3. Disponível em: <https://data.mendeley.com/datasets/7fvvj4jzp7/3>. Acesso em: 27 jul. 2025.

[20] HUANG, Jin et al. **BUSI\_WHU: Breast Cancer Ultrasound Image Dataset**. Mendeley Data, v. 3, 2025. DOI: 10.17632/k6cpmwybk3.3. Disponível em: <https://data.mendeley.com/datasets/k6cpmwybk3/3>. Acesso em: 27 jul. 2025.

[21] IQBAL, Ahmed. **BUS\_UC**. Mendeley Data, v. 1, 2023. DOI: 10.17632/3ksd7w7jkk.1. Disponível em: <https://data.mendeley.com/datasets/3ksd7w7jkk/1>. Acesso em: 27 jul. 2025.

[22] ARDAKANI, Ali Abbasian et al. An open-access breast lesion ultrasound image database: Applicable in artificial intelligence studies. **Computers in Biology and Medicine**, 2022. DOI: 10.1016/j.combiomed.2022.106438. Disponível em: <https://qamebi.com/breast-ultrasound-images-database/>. Acesso em: 27 jul. 2025.

[23] GÓMEZ-FLORES, Wilfrido; GREGORIO-CALAS, Maria Julia; PEREIRA, Wagner Coelho de Albuquerque. **BUS-BRA: A breast ultrasound dataset with BI-RADS categories**. Zenodo, 2023. DOI: 10.5281/zenodo.8231412. Disponível em: <https://zenodo.org/records/8231412>. Acesso em: 27 jul. 2025.

[24] PAWŁOWSKA, A. et al. **A Curated Benchmark Dataset for Ultrasound Based Breast Lesion Analysis (Breast-Lesions-USG)**. Version 1. The Cancer Imaging Archive, 2024. DOI: 10.7937/9WKK-Q141. Disponível em: <https://doi.org/10.7937/9WKK-Q141>. Acesso em: 27 jul. 2025.

[25] MO, Yuhao et al. **HoVer-Trans: Anatomy-Aware HoVer-Transformer for ROI-Free Breast Cancer Diagnosis in Ultrasound Images**. IEEE Transactions on Medical Imaging, v. 42, n. 6, 2023. DOI: 10.1109/TMI.2023.3236011. Disponível em: <https://1drv.ms/u/s!AgOtqK2ZncKlgoxsmt-UYbEwMyZY2g?e=INNhyK>. Acesso em: 27 jul. 2025.

[26] YAN, Pengfei et al. **TDF-Net: Trusted Dynamic Feature Fusion Network for breast cancer diagnosis using incomplete multimodal ultrasound**. Information Fusion, 2024. DOI: 10.1016/j.inffus.2024.102592. Disponível em: <https://www.kaggle.com/datasets/timesxy/multimodal-breast-ultrasound-dataset-us3m>. Acesso em: 27 jul. 2025.

[27] FUZHOU UNIVERSITY AFFILIATED PROVINCIAL HOSPITAL. **ALN-Ultra: Multimodal Ultrasound Dataset for Axillary Lymph Node Analysis**. Zenodo, 2025. DOI: 10.5281/zenodo.15003119. Disponível em: <https://zenodo.org/records/15003119>. Acesso em: 27 jul. 2025.

[28] GOODMAN, H. B.; ROWLAND, P. Deficiencies of Compliancy for Data and Storage. In: CHOO, K. K. R. et al. (ed.). **National Cyber Summit (NCS) Research Track 2020**. 1. ed. Cham: Springer, 2021. (Advances in Intelligent Systems and Computing, v. 1271). p. 11. DOI: 10.1007/978-3-030-58703-1\_11.