

UNIVERSIDADE FEDERAL DO MARANHÃO - UFMA  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET  
COORDENAÇÃO DO CURSO DE ENGENHARIA DA COMPUTAÇÃO

**ANTONIO FIALHO DA SILVA NETO**

**AVALIAÇÃO COMPARATIVA DE MÉTODOS DE EXTRAÇÃO DE FEATURES E  
CLASSIFICAÇÃO PARA DETECÇÃO DE MAMAS EM IMAGENS**

SÃO LUÍS – MA  
JANEIRO/2026

**ANTONIO FIALHO DA SILVA NETO**

**AVALIAÇÃO COMPARATIVA DE MÉTODOS DE EXTRAÇÃO DE FEATURES E  
CLASSIFICAÇÃO PARA DETECÇÃO DE MAMAS EM IMAGENS**

Trabalho de Conclusão de Curso II (TCC II) apresentado para obtenção do título de Bacharel em Engenharia da Computação, pela Universidade Federal do Maranhão - UFMA, Cidade Universitária: Campus Dom Delgado, São Luís, MA.

Orientador (a): Prof (a). Dr (a). Bruno Feres de Souza.

SÃO LUÍS – MA  
JANEIRO/2026

# **AVALIAÇÃO COMPARATIVA DE MÉTODOS DE EXTRAÇÃO DE FEATURES E CLASSIFICAÇÃO PARA DETECÇÃO DE MAMAS EM IMAGENS**

Trabalho de Conclusão de Curso II (TCC II) apresentado para obtenção do título de Bacharel em Engenharia da Computação, pela Universidade Federal do Maranhão - UFMA, Cidade Universitária: Campus Dom Delgado, São Luís, MA.

---

**ANTONIO FIALHO DA SILVA NETO**

Aprovado em: 28 / 01 / 2026

BANCA EXAMINADORA

---

*Orientador: Prof. Dr. Bruno Feres de Souza*

---

*Prof. Dr. Alex Oliveira Barradas Filho*

---

*Prof. Dr. Paulo Rogerio De Almeida Ribeiro*

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Fialho da Silva Neto, Antonio.

AVALIAÇÃO COMPARATIVA DE MÉTODOS DE EXTRAÇÃO DE  
FEATURES E CLASSIFICAÇÃO PARA DETECÇÃO DE MAMAS EM IMAGENS  
/ Antonio Fialho da Silva Neto. - 2026.

35 p.

Orientador(a): Bruno Feres de Souza.

Curso de Engenharia da Computação, Universidade Federal  
do Maranhão, São Luís - Ma, 2026.

1. Visão Computacional. 2. Detecção de Anomalias. 3.  
Classificação One-class. 4. Redes Neurais Convolucionais.  
5. Imagens Médicas. I. Feres de Souza, Bruno. II.  
Título.

*Dedico este trabalho a todos que me ajudaram  
ao longo desta caminhada, em especial, à  
minha família.*

## **AGRADECIMENTOS**

Agradeço a Deus por me conceder a oportunidade de realizar meu sonho e superar os desafios ao longo do caminho. Sou profundamente grato à minha família, que sempre esteve ao meu lado, oferecendo apoio incondicional, incentivo e compreensão nos momentos mais difíceis, sendo a base essencial para que eu pudesse seguir em frente e alcançar este objetivo.

Aos meus amigos do grupo de pesquisa do DARTiLab (Stellaris), em especial o Prof. Alex Barradas, deixo meu sincero agradecimento pelo companheirismo, pelas discussões enriquecedoras, pelas trocas de conhecimento e pelo apoio constante ao longo desta trajetória acadêmica, que tornaram o percurso mais leve e motivador.

Agradeço, de forma especial, ao Professor Dr. Bruno Feres, pela orientação dedicada, pela paciência, pelo rigor científico e pelos valiosos ensinamentos compartilhados ao longo dessa jornada, que foram fundamentais para meu crescimento acadêmico e profissional.

## RESUMO

A detecção automática da presença de mamas em imagens médicas constitui uma etapa fundamental para garantir a confiabilidade de sistemas computacionais voltados à estimativa automática do volume mamário. Este trabalho propõe e avalia uma abordagem baseada em classificação com uma única classe (one-class classification), empregando redes neurais convolucionais pré-treinadas como extratoras de características e algoritmos clássicos de detecção de anomalias como classificadores. Foram analisadas as arquiteturas VGG-16, ResNet50 e Xception para extração de embeddings, combinadas com os algoritmos One-Class Support Vector Machine (OC-SVM) e Isolation Forest (iForest). O treinamento foi realizado exclusivamente com imagens positivas, enquanto a avaliação considerou um conjunto de teste balanceado contendo imagens com e sem a presença de mamas. O desempenho das seis combinações avaliadas foi mensurado por meio das métricas acurácia, precisão, sensibilidade e F1-score, além da análise de falsos positivos e falsos negativos. Os resultados demonstraram que a combinação ResNet50 + Isolation Forest apresentou o melhor desempenho global, com maior acurácia e F1-score, elevada sensibilidade e ausência de falsos positivos, evidenciando maior robustez na detecção da presença de mamas. Conclui-se que a escolha adequada do extrator de características e do classificador é determinante para o sucesso da tarefa, e que a abordagem proposta é eficaz como etapa preliminar de verificação automática em sistemas de apoio à estimativa de volume mamário, contribuindo para maior segurança e confiabilidade em aplicações clínicas.

**Palavras-chave:** Visão Computacional; Detecção de Anomalias; Classificação One-Class; Redes Neurais Convolucionais; Imagens Médicas; Transfer Learning.

## ABSTRACT

The automatic detection of the presence of breasts in medical images is a fundamental step to ensure the reliability of computational systems aimed at automatic breast volume estimation. This work proposes and evaluates an approach based on one-class classification, employing pre-trained convolutional neural networks as feature extractors and classical anomaly detection algorithms as classifiers. The VGG-16, ResNet50, and Xception architectures were analyzed for embedding extraction, combined with the One-Class Support Vector Machine (OC-SVM) and Isolation Forest (iForest) algorithms. Training was performed exclusively using positive images, while evaluation was conducted on a balanced test set containing images with and without the presence of breasts. The performance of the six evaluated combinations was assessed using accuracy, precision, sensitivity (recall), and F1-score metrics, as well as the analysis of false positives and false negatives. The results demonstrated that the ResNet50 + Isolation Forest combination achieved the best overall performance, with higher accuracy and F1-score, high sensitivity, and no false positives, indicating greater robustness in detecting the presence of breasts. It is concluded that the appropriate choice of feature extractor and classifier is crucial for the success of the task, and that the proposed approach is effective as a preliminary automatic verification stage in breast volume estimation support systems, contributing to increased safety and reliability in clinical applications.

**Keywords:** Computer Vision; Anomaly Detection; One-Class Classification; Convolutional Neural Networks; Medical Images; Transfer Learning.

## LISTA DE FIGURAS

<b>Figura 1.</b> Arquitetura da VGG-16 .....	19
<b>Figura 2.</b> Arquitetura da ResNet-50 .....	20
<b>Figura 3.</b> Arquitetura da Xception .....	20
<b>Figura 4.</b> Pipeline do projeto .....	26
<b>Figura 5.</b> Algumas Imagens com resultados falsos positivos e negativos .....	28

## LISTA DE SIGLAS

<b>CNNs</b>	-	Redes Neurais Convolucionais
<b>VGG</b>	-	Visual Geometry Group
<b>Xception</b>	-	Extreme Inception
<b>One-Class SVM</b>	-	One-Class Support Vector Machine
<b>OC-SVM</b>	-	One-Class Support Vector Machine
<b>iForest</b>	-	Isolation Forest
<b>RGB</b>	-	Red, Green e Blue
<b>HSV</b>	-	Hue, Saturation, Value
<b>ML</b>	-	Machine Learning
<b>DL</b>	-	Deep Learning
<b>OCC</b>	-	One-Class Classification
<b>RBF</b>	-	kernel de base radial

## LISTA DE TABELAS

**Tabela 1.** Resultados comparativos das combinações CNN + Classificador ..... 26

**Tabela 2.** Resultados comparativos dos falsos positivos e negativos ..... 28

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>13</b>
<b>2. JUSTIFICATIVA.....</b>	<b>15</b>
<b>3. OBJETIVOS.....</b>	<b>16</b>
3.1 Objetivo Geral.....	16
3.2 Objetivos Específicos.....	16
<b>4. REFERENCIAL TEÓRICO.....</b>	<b>17</b>
4.1 Visão Computacional e Análise de Imagens.....	17
4.2 Aprendizado de Máquina Aplicado à Visão Computacional.....	17
4.3 Redes Neurais Convolucionais (CNNs).....	18
4.4 Transferência de Aprendizado (Transfer Learning).....	19
4.5 Arquiteturas de Redes Neurais Convolucionais Utilizadas.....	19
4.5.1 Arquitetura VGG.....	19
4.5.2 Arquitetura ResNet.....	20
4.5.3 Arquitetura Xception.....	21
4.6 Extração de Características e Embeddings.....	22
4.7 Classificação com Uma Única Classe.....	22
4.8 Algoritmos de Detecção de Anomalias Utilizados.....	23
4.8.1 One-Class Support Vector Machine (One-Class SVM).....	23
4.8.2 Isolation Forest.....	23
4.9 Métricas de Avaliação para Detecção de Anomalias.....	24
4.10 Aplicações em Sistemas de Pré-processamento de Imagens Médicas.....	24
<b>5. METODOLOGIA.....</b>	<b>25</b>
5.1 Conjunto de Dados e Pré-processamento.....	25
5.2 Extração de Embeddings.....	26
5.3 Modelos de Classificação One-Class.....	26
5.4 Protocolo Experimental e Avaliação.....	27
<b>6. RESULTADOS E DISCUSSÃO.....</b>	<b>28</b>
<b>7. CONCLUSÃO.....</b>	<b>31</b>
<b>REFERÊNCIAS.....</b>	<b>32</b>

## 1. INTRODUÇÃO

A estimativa automática do volume mamário por meio de imagens digitais é uma ferramenta valiosa para o planejamento cirúrgico em mamoplastias, promovendo padronização, redução da subjetividade e otimização do tempo. Recentemente, foi desenvolvido um sistema baseado em Aprendizado de Máquina que localiza, segmenta e calcula o volume das mamas a partir de fotografias bidimensionais, fornecendo suporte objetivo à decisão médica (SAUAIA, 2024).

No entanto, para garantir sua confiabilidade em cenários reais, é essencial incluir uma etapa preliminar de verificação automática da presença das mamas nas imagens de entrada. Esse pré-processamento evita que imagens inválidas, incompletas ou fora do escopo sejam enviadas às fases seguintes.

A detecção automática de objetos em imagens é uma área amplamente investigada no campo da Visão Computacional e do Aprendizado de Máquina, com aplicações relevantes em diversos domínios, tais como diagnóstico médico, inspeção industrial, segurança e monitoramento (GHELLERE, 2015).

De modo geral, métodos tradicionais de classificação supervisionada requerem conjuntos de dados extensos e balanceados, contendo amostras representativas tanto da presença quanto da ausência do objeto de interesse. Entretanto, em diversos cenários práticos, a obtenção de exemplos negativos é limitada, inviável ou economicamente custosa (RODRIGUES, 2021).

Nesse contexto, destaca-se o problema da classificação com uma única classe, também denominado detecção de anomalias (anomaly detection ou one-class classification). Nessa abordagem, o modelo é treinado exclusivamente com exemplos positivos, aprendendo as características intrínsecas do padrão considerado normal, de modo que amostras que apresentem desvios significativos em relação a esse comportamento sejam classificadas como anômalas, representando a ausência do objeto de interesse (LECUN ET AL., 2015 APUD GHELLERE, 2015).

A aplicação desse paradigma no domínio de imagens apresenta desafios adicionais em função da alta dimensionalidade dos dados e da presença de informações redundantes ou irrelevantes. Para mitigar essas dificuldades, abordagens recentes têm empregado redes neurais convolucionais (CNNs) pré-treinadas como extratoras de características, capazes de gerar representações

compactas e discriminativas, denominadas embeddings, que codificam atributos visuais relevantes, como textura, forma e padrões estruturais (SIMONYAN; ZISSERMAN, 2014).

Dentre as arquiteturas de CNNs amplamente utilizadas para esse fim, destacam-se a Visual Geometry Group (VGG), a ResNet e a Extreme Inception (Xception). A arquitetura VGG caracteriza-se por uma estrutura profunda e uniforme baseada em convoluções sequenciais, favorecendo a extração de padrões hierárquicos (SIMONYAN; ZISSERMAN, 2014).

A ResNet introduz conexões residuais, permitindo o treinamento de redes mais profundas e a extração de representações mais robustas (HE ET AL., 2016). Já a Xception explora convoluções separáveis em profundidade, buscando maior eficiência computacional e melhor captura de relações espaciais e de canal. Essas arquiteturas, quando pré-treinadas em bases de grande escala, como o ImageNet, são amplamente empregadas como extratoras de características fixas em cenários com dados limitados (CHOLLET, 2017).

Dessa forma, o presente trabalho propõe uma avaliação comparativa de métodos de extração de características baseados nas arquiteturas VGG, ResNet e Xception, combinadas com algoritmos clássicos de classificação com uma única classe, aplicados à detecção da presença de mamas em imagens. São investigados algoritmos baseados em distância e densidade, como One-Class Support Vector Machine (One-Class SVM) e Isolation Forest (iForest), considerando um cenário de aprendizado com apenas exemplos positivos.

## 2. JUSTIFICATIVA

O desenvolvimento de sistemas automáticos para análise de imagens no contexto médico tem se mostrado uma alternativa promissora para auxiliar profissionais da saúde, reduzindo a subjetividade das avaliações e otimizando o fluxo de processamento dos dados (LITJENS et al., 2017). Em aplicações reais, especialmente aquelas relacionadas à análise de imagens médicas, é comum a existência de restrições quanto à disponibilidade de dados rotulados, em particular no que se refere à obtenção de exemplos negativos, os quais podem demandar tempo, alto custo e a atuação de especialistas (RAJPURKAR et al., 2017).

Nesse cenário, a utilização de técnicas de classificação com uma única classe surge como uma solução adequada, uma vez que permite o treinamento de modelos capazes de identificar a presença do objeto de interesse a partir exclusivamente de exemplos positivos (CHANDOLA; BANERJEE; KUMAR, 2009). Tal abordagem torna-se especialmente relevante quando se deseja implementar um estágio inicial de verificação automática, responsável por filtrar imagens inválidas ou inadequadas antes de submetê-las a etapas posteriores de análise mais complexas (PERERA; DOLGOV, 2019).

Dessa forma, a realização de uma análise comparativa entre arquiteturas consolidadas, como VGG (SIMONYAN; ZISSERMAN, 2014), ResNet (HE et al., 2016) e Xception (CHOLLET, 2017), associadas a algoritmos clássicos de classificação com uma única classe, como One-Class SVM (SCHÖLKOPF et al., 2001) e Isolation Forest (LIU; TING; ZHOU, 2008), justifica-se pela necessidade de identificar a configuração mais eficiente e robusta para a tarefa de detecção da presença de mamas em imagens. A identificação dessa combinação ideal é fundamental para garantir que apenas imagens válidas sejam encaminhadas para estágios subsequentes do sistema.

### **3. OBJETIVOS**

#### **3.1 Objetivo Geral**

Avaliar comparativamente o desempenho das arquiteturas VGG, ResNet e Xception como extratoras de características em conjunto com algoritmos de classificação com uma única classe para a detecção da presença de mamas em imagens.

#### **3.2 Objetivos Específicos**

1. Avaliar comparativamente o desempenho de arquiteturas de CNN (ResNet50, Xception e VGG-16) na extração de características para detecção de anomalias;
2. Avaliar comparativamente a influência dos classificadores One-Class SVM (OC-SVM) e Isolation Forest (iForest) no desempenho da detecção de anomalias;
3. Comparar as combinações propostas (CNN + classificador) por meio de métricas adequadas a problemas de detecção de anomalias;

## **4. REFERENCIAL TEÓRICO**

### **4.1 Visão Computacional e Análise de Imagens**

A visão computacional é uma área multidisciplinar da ciência da computação que busca possibilitar que máquinas interpretem e compreendam o conteúdo de imagens e vídeos de forma automática (SZYMKOWIAK et al., 2019). Seus fundamentos combinam conhecimentos de processamento de sinais, óptica, álgebra linear e geometria para extrair informação de dados visuais (GONZALEZ; WOODS, 2018). A representação digital de uma imagem, essencialmente uma matriz de pixels, é o ponto de partida para qualquer análise, onde cada elemento (pixel) carrega informações de intensidade luminosa (em níveis de cinza) ou de componentes de cor (como nos espaços RGB ou HSV). Esta transformação da realidade visual em dados numéricos estruturados é o que viabiliza a aplicação de algoritmos computacionais (SONKA; Hlavac; BOYLE, 2008).

No contexto específico das imagens médicas, a análise automática enfrenta desafios significativos, como a alta variabilidade anatômica entre pacientes, a presença de ruídos e artefatos de aquisição (ex: baixo contraste em ressonâncias), e a complexidade semântica das estruturas de interesse (LITJENS et al., 2017). Tais desafios evidenciam a importância crítica das etapas de pré-processamento e de extração de características robustas.

O pré-processamento, que pode envolver normalização, correção de contraste e filtragem, tem como objetivo isolar o sinal relevante, enquanto a extração de características visa capturar padrões discriminativos (como texturas, bordas ou formas) que serão posteriormente utilizados para classificação ou diagnóstico, servindo como base para a tomada de decisão assistida por computador (MAJUMDAR et al., 2020).

### **4.2 Aprendizado de Máquina Aplicado à Visão Computacional**

O aprendizado de máquina (Machine Learning – ML) consiste em um conjunto de técnicas que permitem a sistemas computacionais aprenderem padrões a partir de dados, melhorando seu desempenho em uma tarefa específica sem serem explicitamente programados para cada cenário (MITCHELL, 1997). No domínio da visão computacional, o ML tem sido fundamental para superar as

limitações das técnicas tradicionais baseadas em regras manuais, permitindo que modelos descubram representações hierárquicas e complexas diretamente dos pixels das imagens (GOODFELLOW; BENGIO; COURVILLE, 2016). Tradicionalmente, o aprendizado é categorizado em supervisionado (onde os dados de treinamento possuem rótulos conhecidos), não supervisionado (busca de padrões em dados não rotulados) e semi-supervisionado (que combina uma pequena quantidade de dados rotulados com muitos não rotulados) (BISHOP, 2006).

Em problemas de diagnóstico médico, é comum a escassez de grandes volumes de dados anotados por especialistas, o que impõe sérias limitações ao aprendizado supervisionado clássico, que tende a sofrer de overfitting (sobreajuste) em tais condições (RAJPURKAR et al., 2017). Essa limitação foi um dos principais motivadores para o surgimento e adoção em larga escala do aprendizado profundo, ou Deep Learning (DL). O DL, um subcampo do ML baseado em redes neurais artificiais com múltiplas camadas, demonstrou capacidade excepcional de aprender automaticamente características de alto nível e altamente discriminativas a partir de dados brutos, revolucionando tarefas como classificação e segmentação de imagens médicas ao abstrair a necessidade de uma extração de características manual e pré-definida (LITJENS et al., 2017).

### **4.3 Redes Neurais Convolucionais (CNNs)**

As CNNs constituem a principal arquitetura do Deep Learning para o processamento de dados visuais, sendo inspiradas no funcionamento do córtex visual biológico (LECUN et al., 2015). Elas são formadas por camadas convolucionais que aplicam filtros para extrair características locais da imagem, camadas de pooling que reduzem a dimensionalidade espacial e aumentam a robustez a variações, e camadas totalmente conectadas responsáveis pela classificação ou regressão (KIM, 2017).

O uso de funções de ativação não lineares, como a ReLU, permite a modelagem de relações complexas, enquanto a extração automática e hierárquica de características, desde padrões simples como bordas até representações abstratas e semânticas, torna as CNNs especialmente eficazes e adequadas para aplicações em análise de imagens médicas (LITJENS et al., 2017).

## **4.4 Transferência de Aprendizado (Transfer Learning)**

A transferência de aprendizado é uma técnica que permite aproveitar o conhecimento adquirido por modelos treinados em grandes bases de dados, como CNNs pré-treinadas na ImageNet, para melhorar o desempenho em tarefas relacionadas com conjuntos de dados limitados (YOSINSKI et al., 2014). No contexto visual, essas redes aprendem características genéricas como bordas, texturas e padrões que podem ser reutilizadas em novos domínios (RAHMAN; WANG, 2019).

Seu uso ocorre principalmente por meio da extração de embeddings, na qual a rede é mantida congelada e atua como extratora de características para um novo classificador, ou pelo ajuste fino, em que parte ou todas as camadas são retreinadas com dados da tarefa alvo (AZIZI et al., 2021). As principais vantagens dessa abordagem incluem a redução do tempo de treinamento, a necessidade de menos dados rotulados e a melhoria de desempenho, especialmente em áreas com escassez de dados como a medicina, embora possa haver limitações decorrentes do descompasso entre o domínio dos dados originais e o domínio específico da aplicação (RAGHU et al., 2019).

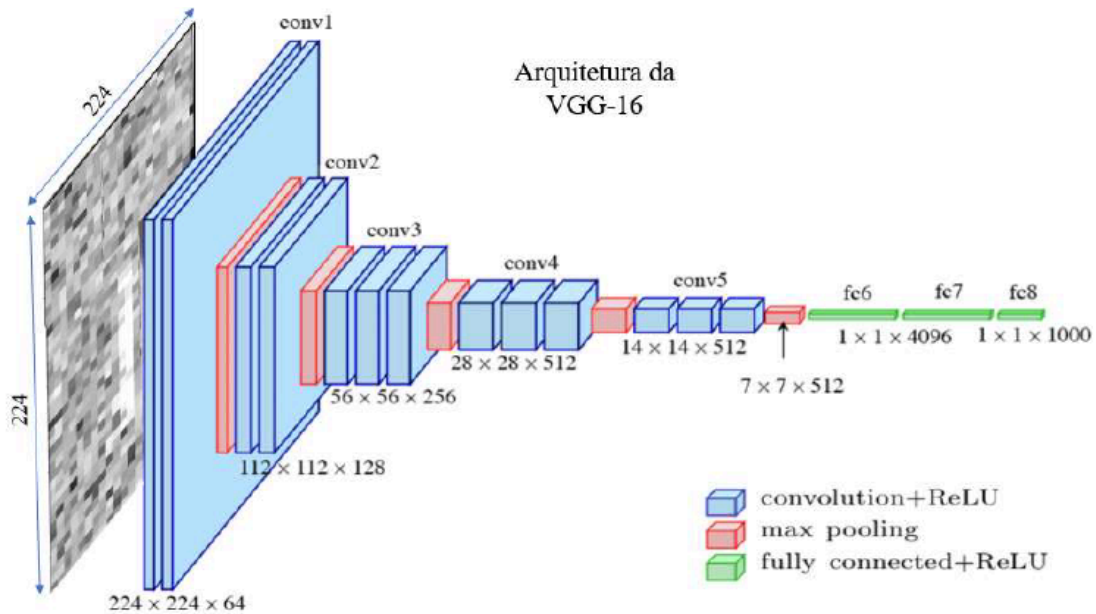
## **4.5 Arquiteturas de Redes Neurais Convolucionais Utilizadas**

### **4.5.1 Arquitetura VGG**

A arquitetura VGG mostrada na Figura 1, desenvolvida pelo Visual Geometry Group da Universidade de Oxford, marcou um avanço ao demonstrar que a profundidade das redes neurais é um fator crucial para seu desempenho (SIMONYAN; ZISSERMAN, 2015). Sua versão principal é a VGG16, caracterizada por uma estrutura profunda e uniforme, composta por uma sequência de camadas convolucionais 3x3 seguidas por camadas de max-pooling, culminando em camadas totalmente conectadas.

A simplicidade e a regularidade de seu design foram fatores-chave para sua ampla adoção e influência (SIMONYAN; ZISSERMAN, 2015). As principais vantagens da VGG incluem sua simplicidade conceitual e a capacidade de aprender representações robustas devido à sua profundidade, o que a tornou um excelente extrator de características para diversas tarefas (CHATFIELD et al., 2014).

Figura 1. Arquitetura da VGG-16



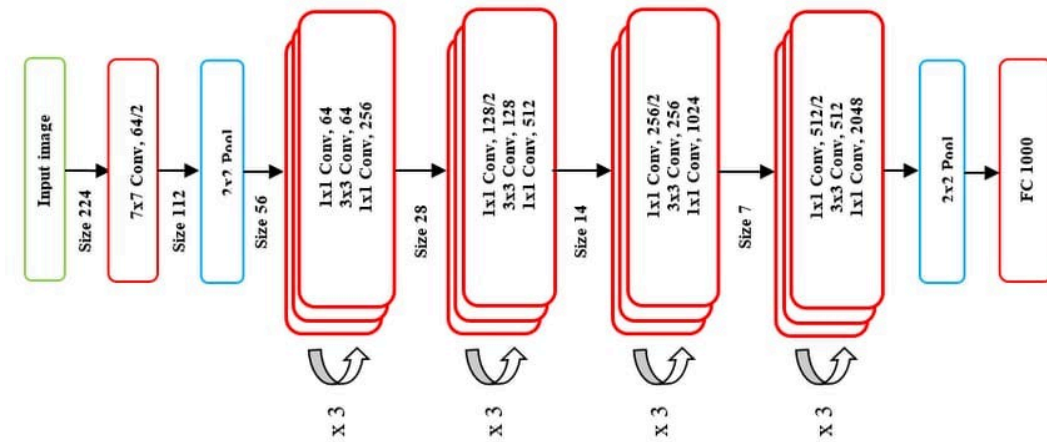
Fonte: FALQUETO et al., 2023.

#### 4.5.2 Arquitetura ResNet

A arquitetura Residual Network, exemplificada pela ResNet50, introduziu as conexões residuais como uma solução eficaz para o problema Vanishing Gradient (desaparecimento de gradiente) no treinamento de redes neurais profundas. Por meio de atalhos, chamados de skip connections, que permitem a soma direta da saída de camadas anteriores à entrada de camadas posteriores, os blocos residuais possibilitam que a rede aprenda funções residuais em vez de transformações complexas diretas, facilitando a otimização e o ajuste dos pesos (HE et al., 2016).

Esse mecanismo permite preservar informações relevantes ao longo de toda a profundidade da rede, resultando em uma extração de embeddings mais robusta, estável e rica em significado (HE et al., 2016). A ResNet50 torna-se especialmente adequada para tarefas de classificação e análise de imagens complexas devido à sua capacidade de lidar com arquiteturas muito profundas sem degradação de desempenho, a seguir é mostrada a arquitetura ResNet50 na Figura 2 (MUKHERJEE., 2022).

Figura 2. Arquitetura da ResNet-50

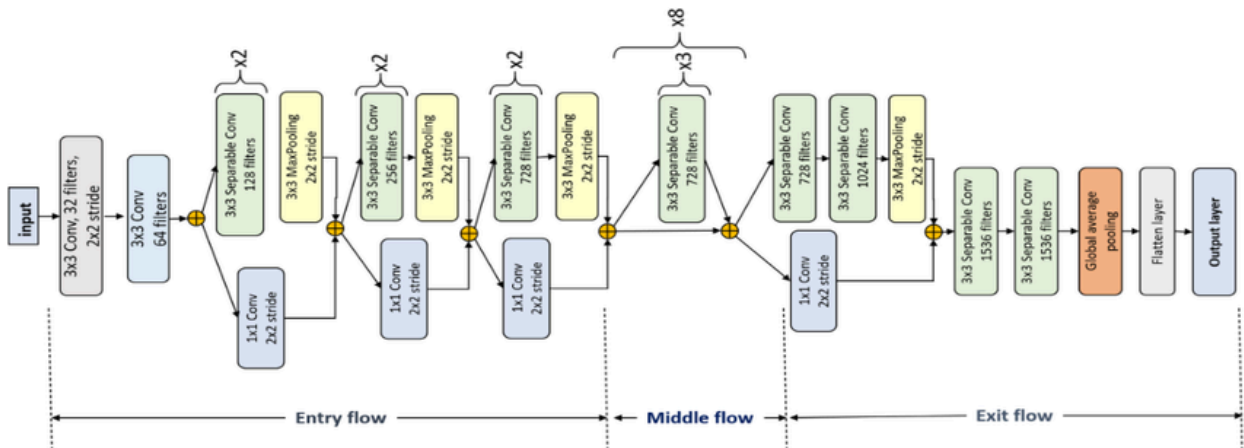


Fonte: MUKHERJEE., 2022.

#### 4.5.3 Arquitetura Xception

A arquitetura Xception é uma evolução da Inception que propõe o desacoplamento entre a modelagem de correlações espaciais e entre canais por meio do uso extensivo de convoluções separáveis em profundidade, nas quais a convolução espacial é aplicada de forma independente a cada canal e posteriormente combinada por uma convolução pontual 1×1 (CHOLLET, 2017). Essa abordagem resulta em elevada eficiência computacional, com significativa redução no número de parâmetros e operações, sem comprometer a capacidade de representação do modelo mostrado na Figura 3 (MUKHLIF et al., 2023).

Figura 3. Arquitetura da Xception



Fonte: MUKHLIF et al., 2023.

Como consequência, a Xception frequentemente apresenta desempenho competitivo ou superior a arquiteturas mais complexas, tornando-se especialmente adequada para a extração de embeddings em tarefas de visão computacional e análise de imagens médicas, inclusive em cenários com restrições de recursos computacionais (CHOLLET, 2017).

#### **4.6 Extração de Características e Embeddings**

No aprendizado de máquina, características são atributos mensuráveis que descrevem e diferenciam os dados, enquanto embeddings são representações vetoriais compactas obtidas por transformações aprendidas que mapeiam os dados para um espaço de características de menor dimensão e maior densidade semântica (BENGIO; COURVILLE; VINCENT, 2013). As Redes Neurais Convolucionais desempenham um papel central nesse processo ao extrair embeddings de forma hierárquica, convertendo imagens de alta dimensionalidade em vetores que concentram as informações visuais mais relevantes (ZEILER; FERGUS, 2014).

A qualidade dessas representações é essencial para o desempenho de tarefas posteriores, como classificação e detecção de anomalias, pois embeddings bem estruturados facilitam a identificação de padrões e a tomada de decisões precisas (KINGMA; WELING, 2014).

#### **4.7 Classificação com Uma Única Classe**

A classificação com uma única classe (One-Class Classification – OCC) é um paradigma de aprendizado supervisionado no qual o modelo é treinado apenas com exemplos da classe normal, aprendendo suas características intrínsecas para posteriormente identificar se novas amostras pertencem a essa classe ou representam anomalias (CHANDOLA; BANERJEE; KUMAR, 2009). Diferente da classificação tradicional, essa abordagem é amplamente associada à detecção de anomalias e é especialmente útil em cenários onde a obtenção de exemplos da classe anormal é limitada, como em aplicações médicas (PERERA; DOLGOV, 2019).

Apesar de sua relevância, a OCC apresenta desafios significativos, pois o modelo deve definir os limites da classe alvo sem exemplos contrastivos, o que

pode resultar em dificuldades de calibração, sensibilidade a parâmetros e aumento das taxas de falsos positivos ou negativos (SCHÖLKOPF et al., 2001).

## **4.8 Algoritmos de Detecção de Anomalias Utilizados**

### **4.8.1 One-Class Support Vector Machine (One-Class SVM)**

O One-Class Support Vector Machine (OC-SVM) é uma variação do Support Vector Machine tradicional desenvolvida especificamente para cenários em que apenas amostras de uma única classe estão disponíveis durante o treinamento. Nesse contexto, o objetivo do algoritmo não é realizar uma classificação multiclasse ou binária convencional, mas aprender a representação estatística da normalidade dos dados, identificando como anômalas as amostras que se desviam desse padrão (SCHÖLKOPF et al., 2001).

O OC-SVM constrói uma fronteira de decisão no espaço de características que engloba a maior parte das amostras de treinamento, maximizando a separação em relação à origem ou a regiões consideradas fora do padrão. A utilização de funções kernel, como o kernel de base radial (RBF), permite projetar os dados em espaços de alta dimensionalidade, possibilitando a modelagem de fronteiras não lineares complexas, o que é particularmente relevante em problemas envolvendo descritores extraídos de imagens e dados de alta dimensionalidade. Neste trabalho, o OC-SVM é adotado como classificador único, sendo responsável pela etapa final de decisão, a partir de vetores de características previamente extraídos (SCHÖLKOPF et al., 2001).

### **4.8.2 Isolation Forest**

O Isolation Forest é um algoritmo de detecção de anomalias baseado em árvores de decisão que se diferencia por buscar isolar explicitamente as anomalias em vez de modelar apenas o comportamento dos dados normais. Ele constrói uma floresta de árvores de isolamento por meio de divisões aleatórias dos atributos, explorando o fato de que amostras anômalas tendem a ser isoladas com caminhos mais curtos da raiz até as folhas (LIU et al., 2008).

O comprimento médio desses caminhos define o score de anomalia, sendo valores menores indicativos de maior anormalidade (LIU et al., 2008). Entre suas principais vantagens destacam-se a alta eficiência computacional, a boa

escalabilidade e a adequação a dados de alta dimensionalidade, o que o torna especialmente apropriado para a identificação de padrões atípicos em embeddings extraídos por CNNs (LIU; TING; ZHOU, 2012).

#### **4.9 Métricas de Avaliação para Detecção de Anomalias**

A avaliação de modelos de detecção de anomalias em cenários one-class apresenta desafios específicos, uma vez que, em geral, o treinamento é realizado apenas com exemplos da classe normal, tornando necessária a utilização de um conjunto de teste contendo amostras normais e anômalas para simular a operação real do sistema (CHANDOLA; BANERJEE; KUMAR, 2009). Nessa avaliação, as métricas derivadas da matriz de confusão assumem papel central, permitindo a análise de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos.

Considerando a criticidade de aplicações médicas, nas quais erros podem acarretar consequências significativas, torna-se essencial empregar múltiplas métricas, como as taxas de falsos positivos e falsos negativos, bem como precisão, recall e F1-score, que fornecem uma visão equilibrada do desempenho do modelo, especialmente em cenários com classes desbalanceadas, característicos da detecção de anomalias (DAVIS; GOADRIC, 2006).

#### **4.10 Aplicações em Sistemas de Pré-processamento de Imagens Médicas**

Em sistemas de auxílio diagnóstico baseados em aprendizado profundo, a qualidade das imagens de entrada é determinante para o desempenho do modelo final, tornando o pré-processamento automático uma etapa fundamental não apenas para ajustes técnicos, mas também para a filtragem semântica do pipeline de análise (LITJENS et al., 2017). Nesse contexto, classificadores one-class podem ser integrados como uma etapa inicial de triagem inteligente, validando se a imagem pertence ao domínio de interesse antes de ser encaminhada a etapas posteriores mais complexas e computacionalmente custosas, como predição de volume (PERERA; DOLGOV, 2019).

## 5. METODOLOGIA

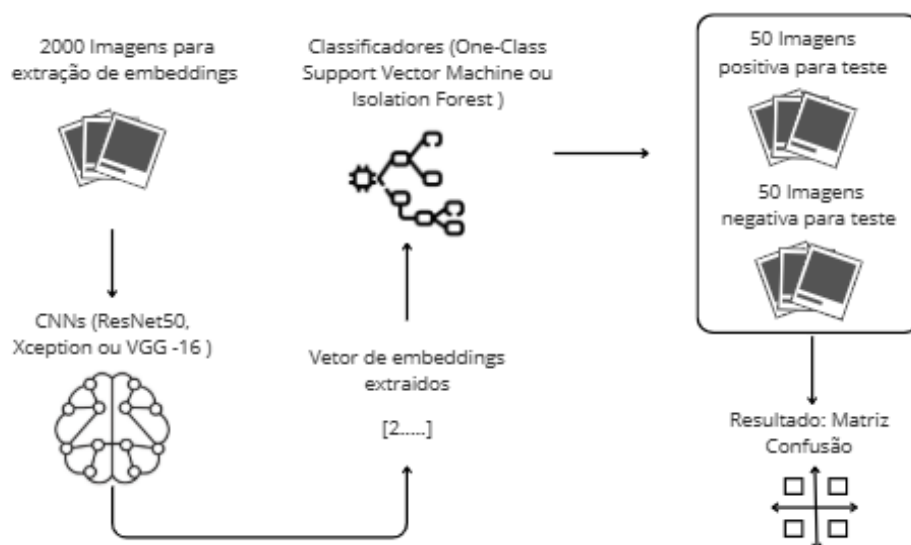
### 5.1 Conjunto de Dados e Pré-processamento

O presente estudo utilizou uma base pública de imagens médicas disponibilizada por Guo et al. (2022), composta por imagens frontais de pacientes participantes de três ensaios clínicos relacionados à terapia de radiação após cirurgia conservadora da mama. As imagens foram adquiridas em alta resolução, com as pacientes posicionadas em postura akimbo, sobre fundo neutro, e submetidas a critérios rigorosos de padronização quanto à iluminação, ao contraste e à qualidade visual, conforme descrito por Wang et al. (2021). Do total disponível, 2.000 imagens foram destinadas à extração de embeddings, enquanto 50 imagens foram reservadas para a etapa de teste positivo.

Adicionalmente, foi empregada a técnica de web scraping para a obtenção de 50 imagens de casos negativos a partir do mecanismo de busca Google Imagens. Esse procedimento foi realizado com o objetivo de selecionar imagens com dimensões padronizadas de 500 × 500 pixels de casos negativos.

No contexto deste trabalho, a base de dados foi utilizada em um cenário de classificação de classe única, no qual o conjunto de treinamento foi composto exclusivamente por imagens positivas, contendo mamas. Para a etapa de avaliação, foi construído um conjunto de teste balanceado, formado por imagens positivas e negativas (não mamárias), permitindo uma análise adequada do desempenho dos modelos. As imagens foram previamente redimensionadas de acordo com os requisitos das arquiteturas empregadas, 224×224 pixels para VGG-16 e ResNet50, e 299×299 pixels para Xception, normalizadas conforme os parâmetros das redes pré-treinadas no conjunto ImageNet. A Figura 4 apresenta o pipeline do projeto proposto.

Figura 4. Pipeline do projeto



Fonte: Autoria própria, 2026.

## 5.2 Extração de Embeddings

As três arquiteturas de CNN (VGG-16, ResNet50 e Xception), pré-treinadas no conjunto ImageNet, foram utilizadas como extratores de características fixos (congelados). Para cada imagem, os embeddings foram extraídos da penúltima camada antes da classificação (camada totalmente conectada de 4096 dimensões para VGG-16, global average pooling de 2048 dimensões para ResNet50, e global average pooling de 2048 dimensões para Xception) (ZEILER; FERGUS, 2014). Este processo resultou em três conjuntos de vetores de características de alta dimensionalidade, um para cada arquitetura.

## 5.3 Modelos de Classificação One-Class

Dois algoritmos clássicos de detecção de anomalias foram utilizados neste estudo: o One-Class Support Vector Machine e o Isolation Forest. O OC-SVM foi implementado com kernel RBF, sendo o parâmetro  $\nu$  ( $\nu$ ), que define o limite superior da fração de outliers permitidos durante o treinamento, ajustado por meio de validação cruzada utilizando apenas os dados positivos (SCHÖLKOPF et al., 2001). Já o Isolation Forest foi configurado com 100 estimadores (árvores), e a proporção de contaminação foi estimada a partir da análise da distribuição dos escores de anomalia gerados pelo modelo (LIU et al., 2008). Ambos os métodos

foram treinados exclusivamente sobre os embeddings extraídos das imagens positivas (classe normal), obtidos a partir de cada uma das redes neurais convolucionais (CNNs) avaliadas.

#### **5.4 Protocolo Experimental e Avaliação**

O conjunto de dados foi dividido em um conjunto de treinamento composto exclusivamente por embeddings de imagens positivas e um conjunto de teste contendo embeddings de imagens positivas e negativas em proporções iguais (50% cada), todas não vistas durante o treinamento. Para cada uma das seis combinações avaliadas (três CNNs e dois classificadores), os modelos foram treinados apenas com os embeddings positivos do conjunto de treinamento e posteriormente avaliados no conjunto de teste. O desempenho foi analisado por meio das métricas Acurácia, Precisão, Sensibilidade (Recall da classe positiva) e F1-Score, calculadas a partir da matriz de confusão, sendo a Sensibilidade considerada a métrica mais relevante devido ao contexto médico, no qual falsos negativos podem resultar na exclusão indevida de exames importantes (ANSARI et al., 2025).

## 6. RESULTADOS E DISCUSSÃO

Os experimentos conduzidos avaliaram a eficácia de diferentes combinações entre arquiteturas de redes neurais convolucionais (CNN) para extração de embeddings e algoritmos de detecção de anomalias no contexto de classificação one-class. Foram testadas três arquiteturas de CNN - ResNet50, Xception e VGG-16 - em conjunto com dois algoritmos classificadores - One-Class Support Vector Machine e Isolation Forest. A Tabela 1 resume os resultados obtidos para cada combinação.

Tabela 1. Resultados comparativos das combinações CNN + Classificador

<b>Combinação</b>	<b>Acurácia</b>	<b>Precisão</b>	<b>Sensibilidade</b>	<b>F1-Score</b>
ResNet50 + OC-SVM	0,9400	1,0000	0,8800	0,9362
ResNet50 + iForest	0,9600	1,0000	0,9200	0,9583
Xception + iForest	0,9400	0,9400	0,9400	0,9400
Xception + OC-SVM	0,9400	1,0000	0,8800	0,9362
VGG -16 + iForest	0,9500	0,9778	0,9200	0,9485
VGG-16 + OC-SVM	0,9200	0,9773	0,8600	0,9149

A análise comparativa das arquiteturas evidencia a ResNet50 como a mais robusta para a classificação one-class, alcançando os melhores resultados de F1-Score (0,9583) e acurácia (0,9600) quando combinada ao Isolation Forest, além de desempenho consistente com o OC-SVM. Esse comportamento indica que suas conexões residuais favorecem a extração de embeddings altamente discriminativos, mesmo em um cenário restritivo de aprendizado com uma única classe.

A Xception apresentou desempenho equilibrado, destacando-se pela boa calibração entre precisão e sensibilidade, enquanto a VGG-16 obteve resultados inferiores, especialmente com o OC-SVM. Assim, para a tarefa de detecção da presença de mamas, a ResNet50 mostrou-se a arquitetura mais eficaz, seguida pela Xception e, por último, pela VGG-16, conforme indicado pelas métricas de F1-Score e acurácia.

A análise dos classificadores evidenciou desempenho consistentemente superior do iForest em relação ao OC-SVM em todas as arquiteturas avaliadas. As combinações com iForest resultaram em F1-Scores mais elevados e maior sensibilidade, métrica essencial para detecção de anomalias, indicando melhor capacidade de identificar corretamente a presença de mama e reduzir falsos negativos em comparação ao OC-SVM.

Embora o OC-SVM tenha alcançado precisão perfeita em algumas combinações, esse resultado ocorreu à custa de menor sensibilidade, sugerindo um modelo excessivamente conservador. Assim, os resultados confirmam que o iForest apresenta melhor equilíbrio entre precisão e recall, refletindo desempenho global superior e maior adequação para a tarefa de detecção proposta.

A avaliação comparativa das combinações confirmou que ResNet50 + Isolation Forest é a abordagem mais eficaz para a detecção da presença de mamas, alcançando os melhores resultados em acurácia (0,9600) e F1-Score (0,9583), além de alta sensibilidade e precisão perfeita. Essa combinação apresentou o melhor equilíbrio entre detecção correta da classe de interesse e rejeição de anomalias, com baixa incidência de falsos negativos e ausência de falsos positivos.

Embora VGG-16 + iForest tenha mostrado desempenho competitivo, as combinações com OC-SVM ficaram em posições inferiores, especialmente com a VGG-16. Assim, os resultados demonstram que a escolha conjunta do extrator de características e do classificador é determinante, destacando a ResNet50 associada ao iForest como a solução mais robusta e recomendada para a tarefa proposta.

A Tabela 2 evidencia diferenças claras entre as combinações quanto à incidência de erros. A ResNet50 + iForest destacou-se por não gerar falsos positivos e por apresentar o menor número de falsos negativos, reforçando sua robustez. As combinações com OC-SVM concentraram maior quantidade de falsos negativos, indicando um comportamento mais conservador do classificador. Por outro lado, a Xception + iForest e as combinações com VGG-16 introduziram falsos positivos, sugerindo que, em alguns casos, as representações extraídas confundiram anomalias com a classe normal.

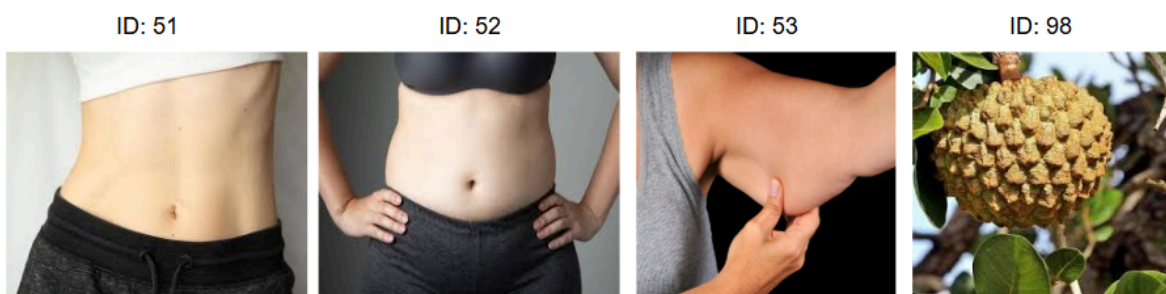
Tabela 2. Resultados comparativos dos falsos positivos e negativos

<b>Combinação</b>	ID Imagem Falso Positivas	ID Imagem Falso Negativas
ResNet50 + OC-SVM	-	5, 22, 23, 24, 28, 34
ResNet50 + iForest	-	5, 22, 23, 24
Xception + iForest	51, 52, 98	22, 23, 24
Xception + OC-SVM	-	22, 23, 24, 34, 38, 48
VGG-16 + iForest	53	13, 22, 24, 34
VGG-16 + OC-SVM	53	13, 19, 22, 23, 24, 34, 44

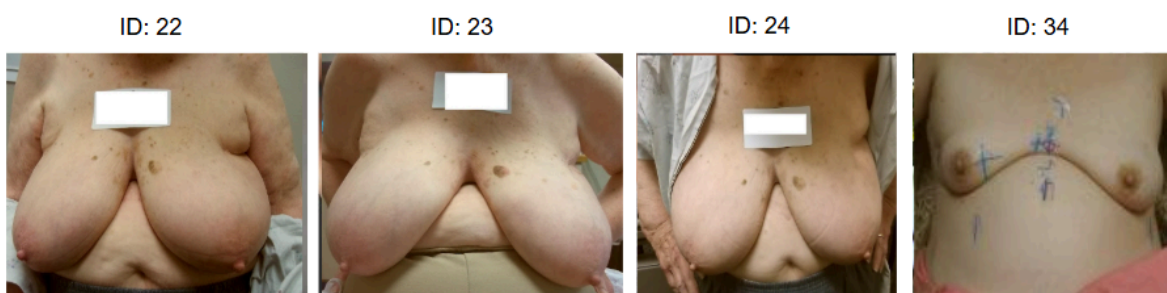
A seguir na Figura 5, é possível verificar alguns exemplos de falsos positivos e falsos negativos, o que permite uma análise qualitativa que colabora com os resultados das métricas, reforçando a superioridade da ResNet50, a qual demonstrou maior capacidade de distinguir regiões do corpo que não correspondem à mama como anômalas.

Figura 5. Algumas Imagens com resultados falsos positivos e negativos

### Falsos Positivos



### Falsos Negativos



Fonte: Autoria própria, 2026.

## 7. CONCLUSÃO

Este trabalho apresentou uma avaliação comparativa de arquiteturas de redes neurais convolucionais pré-treinadas (ResNet50, Xception e VGG-16) combinadas com algoritmos de classificação one-class (One-Class SVM e Isolation Forest) para a detecção automática da presença de mamas em imagens. Os resultados demonstraram que a escolha conjunta do extrator de características e do classificador exerce influência direta no desempenho do sistema, especialmente em um cenário restritivo de aprendizado baseado exclusivamente em exemplos positivos. Entre as arquiteturas avaliadas, a ResNet50 destacou-se de forma consistente, evidenciando maior capacidade de gerar embeddings discriminativos e robustos para a tarefa proposta.

A análise quantitativa, baseada em métricas como acurácia, precisão, sensibilidade e F1-Score, indicou que a combinação ResNet50 + Isolation Forest apresentou o melhor desempenho global, alcançando os maiores valores de acurácia e F1-Score, além de manter precisão perfeita e elevada sensibilidade. Complementarmente, a análise dos falsos positivos e falsos negativos corroborou esses resultados, mostrando que essa combinação foi capaz de minimizar erros críticos e distinguir de forma mais eficaz regiões do corpo que não correspondem à mama como anômalas.

Dessa forma, a abordagem proposta atende aos objetivos estabelecidos e configura-se como uma solução robusta e adequada para atuar como uma etapa preliminar de verificação automática em sistemas de estimativa do volume mamário, contribuindo para o aumento da confiabilidade e da segurança nas aplicações clínicas subsequentes. Como perspectiva para trabalhos futuros, sugere-se a investigação de modelos baseados em redes neurais convolucionais treinados com conjuntos de dados contendo casos positivos e negativos, a fim de avaliar se tais abordagens apresentam desempenho superior ao método proposto.

## REFERÊNCIAS

AZIZI, S. et al. Big self-supervised models advance medical image classification. *Nature*, v. 617, p. 762–768, 2021.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, n. 8, p. 1798–1828, 2013.

BISHOP, C. M. *Pattern recognition and machine learning*. New York: Springer, 2006.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: a survey. *ACM Computing Surveys*, v. 41, n. 3, p. 1–58, 2009.

CHATFIELD, K. et al. Return of the devil in the details: delving deep into convolutional nets. In: *Proceedings of the British Machine Vision Conference*, 2014.

CHOLLET, F. Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, Honolulu. p. 1251–1258. Disponível em: <https://arxiv.org/abs/1610.02357>. Acesso em: 17 jan. 2026.

DAVIS, J.; GOADRICH, M. The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*, 2006, Pittsburgh. p. 233–240. Disponível em: <http://www.cs.utexas.edu/~olga/Courses/fall08/papers/DavisG06.pdf>. Acesso em: 17 jan. 2026.

FALQUETO, Leonan; PAES, Rafael; PASSARO, Angelo. KNN e rede neural convolucional para o reconhecimento de plataformas de petróleo em imagens SAR do Sentinel-1. *Aplicações Operacionais em Áreas de Defesa*, v. 24, p. 29–33, 2023. DOI: 10.55972/spectrum.v24i1.395.

GHELLERE, J. S. Detecção de objetos em imagens por meio da combinação de extração de características e reconhecimento de padrões. 2015. 90 f. Trabalho de Conclusão de Curso (Graduação em Tecnologia em Sistemas para Internet) – Universidade Tecnológica Federal do Paraná, Medianeira, 2015. Disponível em: [https://repositorio.utfpr.edu.br/jspui/bitstream/1/12518/4/MD\\_COCIC\\_2015\\_1\\_02.pdf](https://repositorio.utfpr.edu.br/jspui/bitstream/1/12518/4/MD_COCIC_2015_1_02.pdf). Acesso em: 17 jan. 2026.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. 4. ed. Pearson, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

GUO, C.; SMITH, T. L.; FENG, Q.; BENITEZ-QUIROZ, F.; VICINI, F.; ARTHUR, D.; WHITE, J.; MARTINEZ, A. A fully automatic framework for evaluating cosmetic results of breast conserving therapy. *Machine Learning with Applications*, v. 10, n. 100430, dez. 2022.

HE, K. et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, Las Vegas. p. 770–778. Disponível em: <https://arxiv.org/abs/1512.03385> Acesso em: 17 jan. 2026.

KIM, J. Understanding convolutional neural networks: from theory to practice. Journal of Machine Learning Research, 2017.

KINGMA, D. P.; WELLING, M. Auto-encoding variational Bayes. In: Proceedings of the International Conference on Learning Representations, 2014. Disponível em: <https://arxiv.org/abs/1312.6114> . Acesso em: 17 jan. 2026.

LECUN, Y. et al. Deep learning. Nature, v. 521, n. 7553, p. 436–444, 2015.

LITJENS, G. et al. A survey on deep learning in medical image analysis. Medical Image Analysis, v. 42, p. 60–88, 2017.

LIU, F. T.; TING, K. M.; ZHOU, Z.-H. Isolation forest. In: Proceedings of the Eighth IEEE International Conference on Data Mining, 2008, Pisa. p. 413–422. Disponível em: <https://arxiv.org/abs/0811.02172> . Acesso em: 17 jan. 2026.

LIU, F. T.; TING, K. M.; ZHOU, Z.-H. Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, v. 6, n. 1, p. 1–39, 2012.

MAJUMDAR, A. et al. A comprehensive survey on deep learning based medical image processing. Multimedia Tools and Applications, 2020. DOI: 10.1007/s11042-020-09534-5.

MITCHELL, T. M. Machine learning. New York: McGraw-Hill, 1997.

MUKHERJEE, Suvaditya. The Annotated ResNet-50: Explaining how ResNet-50 works and why it is so popular. Towards Data Science, 18 ago. 2022. Disponível em: <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758/>. Acesso em: 20.01.2026

MUKHLIF, Abdulrahman; AL-KHATEEB, Belal; MOHAMMED, Mazin. Breast cancer images classification using a new transfer learning technique. Iraqi Journal for Computer Science and Mathematics, 2023. DOI: 10.52866/ijcsm.2023.01.01.0014

PERERA, P.; DOLGOV, V. Deep one class classification. In: Proceedings of the IEEE International Conference on Computer Vision, 2019, Long Beach. p. 2542–2550.

RAGHU, M. et al. Direct perception: learning invariant representations from data. arXiv preprint arXiv:1905.11120, 2019.

RAHMAN, M. M.; WANG, Y. Optimizing deep feature representations for fine-grained medical image classification. Medical Image Analysis, v. 56, p. 1–14, 2019.

RAJPURKAR, P. et al. ChexNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225, 2017.

RODRIGUES, P. H. B. Técnicas de visão computacional para classificação de peças. 2021. Disponível em: <https://dspace.mackenzie.br/items/1cc2806d-38df-42ac-88d2-a2f16d0c4012>. Acesso em: 17 jan. 2026.

SAUAIA, Rayssa Yasmin Pereira. *Estimação de volume mamário em cirurgias plásticas do Sistema Único de Saúde, utilizando aprendizado de máquina*. 2024. 101 f. Dissertação (Mestrado em Saúde Coletiva) – Universidade Federal do Maranhão, São Luís, 2024.

SCHÖLKOPF, B. et al. Estimating the support of a high-dimensional distribution. *Neural Computation*, v. 13, n. 7, p. 1443–1471, 2001. Disponível em: <https://arxiv.org/abs/cond-mat/0106116>. Acesso em: 17 jan. 2026.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations (ICLR)*, 2015. Disponível em: <https://arxiv.org/abs/1409.1556>. Acesso em: 17 jan. 2026.

SONKA, M.; HLAVAC, V.; BOYLE, R. *Image processing, analysis and machine vision*. 3. ed. Thomson, 2008.

SZYMKOWIAK, A. et al. *Computer vision: algorithms and applications*. Springer, 2019.

WANG, C.; LIU, C.; CHENG, H.; CHEN, L.; ENDO, Y.; LUAN, J. The effect of arm position on breast volume measurement using three-dimensional imaging. *Aesthetic Plastic Surgery*, v. 45, n. 5, p. 2009–2014, 2021.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: *Advances in Neural Information Processing Systems*, 2014, Montreal. p. 3320–3328.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: *Proceedings of the European Conference on Computer Vision*, 2014, Zurique. p. 818–833. Disponível em: <https://arxiv.org/abs/1311.2901>. Acesso em: 17 jan. 2026.

