



**UNIVERSIDADE FEDERAL DO MARANHÃO**  
**CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET**  
**BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO**

**JOSÉ ROBERTO DE ARAUJO ARANHA JUNIOR**

**UMA PROPOSTA DE SISTEMA PARA SEGMENTAÇÃO DE  
MAMAS EM MULHERES PORTADORAS DE HIPERTROFIA  
MAMÁRIA SINTOMÁTICA**

**São Luís**

**2023**



**UNIVERSIDADE FEDERAL DO MARANHÃO**  
**CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET**  
**BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO**

**JOSÉ ROBERTO DE ARAUJO ARANHA JUNIOR**

**UMA PROPOSTA DE SISTEMA PARA SEGMENTAÇÃO DE  
MAMAS EM MULHERES PORTADORAS DE HIPERTROFIA  
MAMÁRIA SINTOMÁTICA**

Monografia apresentada ao Bacharelado em Engenharia da Computação como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Prof. Dr. Bruno Feres de Souza

**São Luís**  
**2023**

Aranha Jr, José Roberto de Araújo

UMA PROPOSTA DE SISTEMA PARA SEGMENTAÇÃO DE MAMAS EM MULHERES PORTADORAS DE HIPERTROFIA MAMÁRIA SINTOMÁTICA/ JOSÉ ROBERTO DE ARAUJO ARANHA JUNIOR. – São Luís, 2023.

34 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Bruno Feres de Souza

Monografia – UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIAS - CCET  
BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO, 2023.

1. Hipertrofia mamária sintomática. 2. Mamoplastia. 3. Processamento digital de imagens. 4. Redes neurais convolucionais. 5. Segmentação

**JOSÉ ROBERTO DE ARAUJO ARANHA JUNIOR**

**UMA PROPOSTA DE SISTEMA PARA  
SEGMENTAÇÃO DE MAMAS EM MULHERES  
PORTADORAS DE HIPERTROFIA MAMÁRIA  
SINTOMÁTICA**

Monografia apresentada ao Bacharelado em Engenharia da Computação como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Data da Defesa:

Conceito:

**Banca Examinadora**

---

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

Universidade Federal do Maranhão - UFMA  
Orientador

---

**Profa. Dra. Vandecia Rejane Monteiro  
Fernandes**

Universidade Federal do Maranhão - UFMA  
Membro da Banca

---

**Prof. Dr. Wener Borges de Sampaio**

Universidade Federal do Piauí - UFPI  
Membro da Banca

São Luís  
2023

## **AGRADECIMENTOS**

Em primeiro lugar, gostaria de agradecer a minha mãe, por ser minha fortaleza em todos os momentos e por ter batalhado muito para que eu pudesse chegar até a realização desse sonho.

À minha avó e ao meu tio, por toda a sabedoria e imensurável ajuda durante toda a minha vida.

Ao meu orientador, Prof. Dr. Bruno Feres de Souza, por todo o incansável auxílio prestado para a consolidação desse trabalho.

Ao Prof. Dr. Wener Borges de Sampaio, por toda a mentoria e suporte em toda a minha trajetória acadêmica, durante os diversos anos que estivemos em projetos de pesquisa.

Aos meus amigos, especialmente Laryssa Adler e Roseana Mesquita, por todos os conselhos e incentivos, para que eu pudesse concluir mais essa etapa da minha trajetória acadêmica.

A todo o corpo docente do curso de Engenharia da Computação da Universidade Federal do Maranhão, pela excelente formação acadêmica e profissional fornecida, pelos ensinamentos e experiências de vida compartilhadas ao longo dos anos.

*“No que diz respeito ao empenho, ao compromisso,  
ao esforço, à dedicação, não existe meio-termo.  
Ou você faz uma coisa bem-feita ou não faz.”*  
*(Ayrton Senna)*

## RESUMO

A hipertrofia mamária sintomática (HMS) é uma alteração benigna caracterizada pelo aumento de uma ou ambas as mamas, que pode trazer uma série de consequências negativas à saúde física e psicossocial da mulher. Em casos mais graves, pode ser necessária a mamoplastia redutora, cirurgia que visa diminuir o volume excessivo. Na fase pré-operatória, é fundamental que o especialista possua informações quanto à forma, tamanho e volume da região. Entretanto, a região mamária é complexa de quantificar e fica sujeita à análise subjetiva do especialista ou a exames que nem sempre estão a disposição, como a ressonância magnética. O uso de ferramentas computacionais pode auxiliar a fornecer métricas mais precisas para embasar a tomada de decisão e o manejo adequado. Este trabalho propõe uma metodologia baseada em processamento digital de imagens e redes neurais convolucionais para a segmentação da região mamária, de modo a servir de etapa para cálculo futuro do volume. Os resultados obtidos são promissores com, entre outras métricas, precisão e *recall* superiores a 93%.

**Palavras-chave:** hipertrofia mamária sintomática; mamoplastia; processamento digital de imagens; redes neurais convolucionais; segmentação.

## ABSTRACT

Symptomatic breast hypertrophy (SBH) is a benign alteration characterized by the enlargement of one or both breasts, which can have a series of negative consequences on the physical and psychosocial health of women. In more severe cases, breast reduction surgery, known as mammoplasty, may be necessary to reduce excessive volume. In the pre-surgical phase, it is essential for the specialist to have information about the shape, size, and volume of the region. However, quantifying the breast region is complex and susceptible to the specialist's subjective analysis or examinations that may not always be available, such as magnetic resonance imaging. The use of computational tools can assist in providing more accurate metrics to support decision-making and proper management. This study proposes a methodology based on digital image processing and convolutional neural networks for segmentation of the breast region, as a step towards future volume calculation. The results obtained are promising, with precision and recall metrics exceeding 93%, among others.

**Keywords:** symptomatic breast hypertrophy; mammoplasty; digital image processing; convolutional neural networks; segmentation.



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>9</b>
<b>1.1</b>	<b>Objetivos</b>	<b>10</b>
1.1.1	Objetivo Geral	10
1.1.2	Objetivos Específicos	10
<b>1.2</b>	<b>Estrutura do Trabalho</b>	<b>10</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>11</b>
<b>2.1</b>	<b>Anatomia mamária humana e alterações</b>	<b>11</b>
<b>2.2</b>	<b>Hipertrofia mamária sintomática (HMS)</b>	<b>12</b>
<b>2.3</b>	<b>Processamento de imagens e segmentação das mamas</b>	<b>13</b>
<b>2.4</b>	<b><i>Deep Learning</i></b>	<b>14</b>
2.4.1	Camada de convolução	15
2.4.2	Camada de <i>pooling</i>	16
2.4.3	Camadas totalmente conectadas	16
<b>2.5</b>	<b>Métricas de avaliação</b>	<b>17</b>
<b>2.6</b>	<b>Arquitetura de segmentação do YOLOv8</b>	<b>19</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>22</b>
<b>3.1</b>	<b>Ambiente e configuração</b>	<b>22</b>
<b>3.2</b>	<b>Aquisição de imagens</b>	<b>23</b>
<b>3.3</b>	<b>Pré-processamento</b>	<b>23</b>
3.3.1	Preparação dos dados	23
3.3.2	Seleção dos dados e criação das máscaras	23
3.3.3	Adaptação das máscaras para o formato do YOLOv8	25
<b>3.4</b>	<b>Treinamento do modelo</b>	<b>25</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>27</b>
<b>4.1</b>	<b>Rótulos de validação</b>	<b>27</b>
<b>4.2</b>	<b>Rótulos de predição</b>	<b>27</b>
<b>4.3</b>	<b>Avaliação das métricas</b>	<b>28</b>
4.3.1	Precisão e <i>Recall</i>	29
4.3.2	mAP@50, mAP@50:95 e F1 <i>Score</i>	30
<b>5</b>	<b>CONCLUSÃO</b>	<b>31</b>
	<b>REFERÊNCIAS</b>	<b>32</b>

# 1 INTRODUÇÃO

Alterações benignas das mamas representam uma variedade considerável de alterações da glândula mamária, com frequência bem maior que o câncer de mama. Cerca de 90% das massas mamárias palpáveis são benignas (BODINE; HOLAHAN; MIXON, 2017; SILVA; OLIVEIRA, 2010).

A hipertrofia mamária sintomática (HMS) é uma condição benigna caracterizada por mamas aumentadas, de maneira desproporcional ao biotipo da mulher. Essa desproporcionalidade pode levar a uma série de disfunções físicas, problemas psicológicos e sensação de inadequação social (FONSECA et al., 2018).

Em alguns casos, quando esgotadas outras possibilidades mais conservadoras, o recomendado é o manejo cirúrgico, com a mamoplastia redutora. A cirurgia é indicada para aliviar dores e sinais físicos da HMS, além de promover maior qualidade de vida (DABBAH et al., 1995; SINGH; LOSKEN, 2012).

Na fase preliminar da mamoplastia redutora, é essencial que o cirurgião seja capaz de mensurar com precisão quanto de tecido mamário será removido. A massa tecidual removida pode variar de poucas centenas de gramas até mais que 2kg, a depender do grau de HMS (DESCAMPS et al., 2008).

Existem algumas técnicas na literatura para cálculo de volume mamário, como a utilização de imagens mamográficas e correlação com volume de espécimes cirúrgicos, ultrassonografias, tomografias, moldes termoplásticos e ressonância magnética (BULSTRODE; BELLAMY; SHROTRIA, 2001).

A ressonância magnética é considerada o padrão-ouro, pois pode calcular volume com precisão alta, por conta da possibilidade de geração de imagens 3D, além de ser um exame bem tolerado por todos os pacientes, com exceção dos portadores de claustrofobia. Entretanto, têm seu uso limitado pelo custo elevado (BULSTRODE; BELLAMY; SHROTRIA, 2001).

A medicina moderna, com problemas clínicos cada vez mais complexos, enfrenta desafios para adquirir, analisar e aplicar a grande quantidade de conhecimentos necessários para resolvê-los, além do acesso restrito a determinadas soluções devido a fatores externos, como o alto custo, citado anteriormente (RAMESH et al., 2004).

Nesse contexto, o uso de inteligência artificial (IA), com o desenvolvimento de algoritmos que visam executar tarefas associadas à inteligência humana, tem auxiliado médicos na tomada de decisões terapêuticas e na previsão de resultados, de forma a apoiar profissionais de saúde em suas tarefas diárias, especialmente as que envolvem manipulação de dados e conhecimento (HE et al., 2019; RAMESH et al., 2004).

*Deep learning*, subárea da IA inspirada na estrutura neural do cérebro humano, é atual-

mente o estado da arte das técnicas de aprendizado de máquina e o seu sucesso em tarefas de reconhecimento de padrões trouxe grandes mudanças na área da saúde, o que tem motivado novas pesquisas e esforços de desenvolvimento de sistemas de diagnóstico assistido por computador (CHAN et al., 2020; FERNANDES, 2021).

Ante o exposto, este trabalho propõe, por meio de técnicas de processamento de imagem e *deep learning*, realizar a segmentação frontal da estrutura mamária feminina, de modo a servir de etapa auxiliar para o cálculo futuro de volume mamário por meio de imagens obtidas por câmera de smartphone.

## 1.1 Objetivos

Para concretizar este trabalho, os objetivos a seguir devem ser alcançados:

### 1.1.1 Objetivo Geral

- Desenvolver um sistema computacional, baseado em Redes Neurais Convolucionais, para segmentação semântica da estrutura mamária feminina com a finalidade de auxiliar profissionais de cirurgia plástica a calcular o volume mamário de mulheres portadoras de hipertrofia mamária sintomática.

### 1.1.2 Objetivos Específicos

- Realizar a preparação do *dataset*, além do pré-processamento, para adequar as imagens e rótulos ao padrão aceito pelo *framework* utilizado no treinamento;
- Realizar o treinamento de um modelo de aprendizado de máquina baseado em Redes Neurais Convolucionais;
- Avaliar o modelo proposto por meio de métodos existentes na literatura.

## 1.2 Estrutura do Trabalho

O trabalho está dividido em cinco capítulos. No Capítulo 2 se encontra o referencial teórico utilizado, que fornece embasamento conceitual para a pesquisa. No Capítulo 3 está a metodologia, onde estão as etapas utilizadas para a construção do estudo. No Capítulo 4 são discutidos os resultados obtidos e, finalmente, no Capítulo 5 são apresentadas a conclusão e os trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

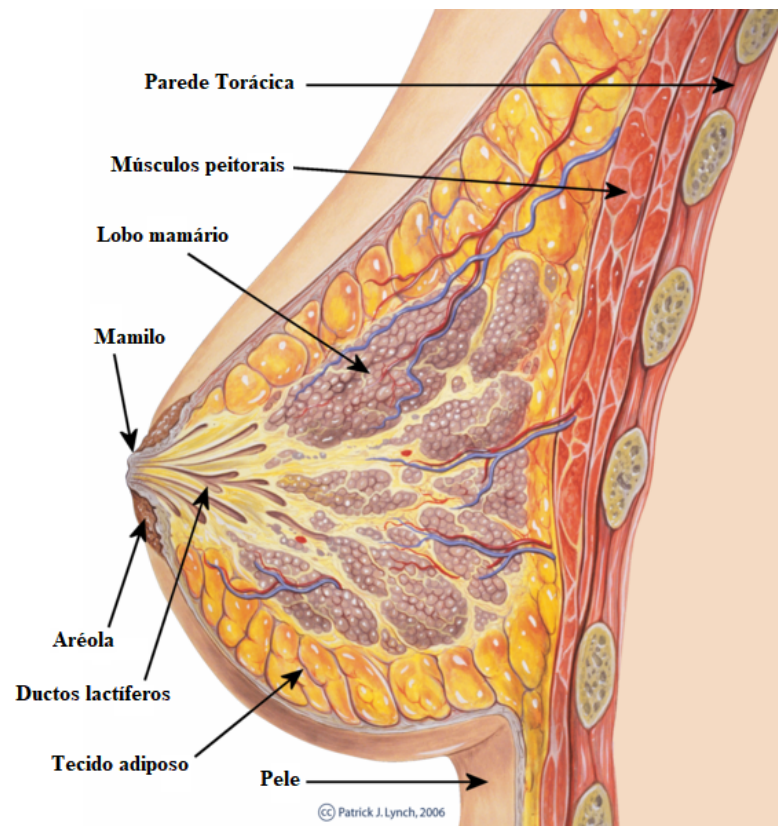
### 2.1 Anatomia mamária humana e alterações

O entendimento da anatomia mamária é de fundamental importância para garantir a eficiência e precisão dos exames físicos, dos diagnósticos diferenciais de neoplasias na região e de todos os seus possíveis usos na prática clínica e áreas correlatas, tais como: construção de próteses, tratamento de patologias, etc (BODINE; HOLAHAN; MIXON, 2017).

As mamas femininas se localizam sobre o músculo peitoral maior e cobrem grande parte da parede torácica. Frontalmente, podem se estender da clavícula até o meio do esterno; e lateralmente, podem se estender da axila até o músculo grande dorsal (GEFEN; DILMONEY, 2007).

A glândula mamária consiste, basicamente, de 15 a 20 lobos irregulares compostos por tecidos glandulares túbulo-alveolares. O estroma conjuntivo fibroso longitudinal forma uma estrutura de treliça que serve de sustentação para os lobos e o tecido adiposo preenche os espaços entre os lobos (BLAND et al., 2017). Essa estrutura está ilustrada na Figura 1.

**Figura 1 – Ilustração da anatomia mamária**



**Fonte: Khan e Sajjad (2019).**

Os tecidos mamários da mulher estão sujeitos a uma série de alterações, em sua maioria benignas, que ocorrem por razões diversas, tais como: fatores hereditários, idade da paciente, uso

de anticoncepcionais, entre outros. Estimativas apontam que cerca de 50% das mulheres terão alguma lesão benigna da mama em algum momento da vida (BODINE; HOLAHAN; MIXON, 2017).

## 2.2 Hipertrofia mamária sintomática (HMS)

Dentre as alterações benignas, a hipertrofia mamária é definida como o aumento de uma ou ambas as mamas além dos limites fisiológicos e não relacionado a traumas, tumores, hemorragias, processos inflamatórios ou de natureza gestacional (FOGED, 1953 apud GONZÁLEZ; ARTETA, 2019).

A etiologia da hipertrofia mamária não é bem determinada, mas existe correlação provável a fatores genéticos e hormonais. A excessiva sensibilidade hormonal tem como consequência a hiperplasia do componente estromal e do epitélio glandular da mama (BROWN et al., 1999 apud TAFURI; GOBBI, 2005).

O aumento do volume das mamas, quando demasiado, pode trazer uma série de consequências negativas à saúde física e psicossocial da mulher, entre as quais pode-se citar: cervicalgia, dores nas costas e nos ombros, eczemas crônicos, intertrigo, alterações posturais, ansiedade, depressão, etc (RUEDA, 2018).

O tratamento para HMS é multidisciplinar e a abordagem está diretamente relacionada ao volume da hiperplasia e seus sintomas. Incluem-se geralmente endocrinologistas, para descartar transtornos e síndromes hormonais, especialistas em medicina física e de reabilitação para avaliar a integridade do sistema músculo-esquelético e, em última instância, caso a resposta a tratamentos mais conservadores seja insatisfatória, opta-se por manejo cirúrgico, feito geralmente por cirurgiões plásticos (RUEDA, 2018).

No Brasil, é frequente a procura por tratamento cirúrgico para redução das mamas, de forma a melhorar a estética e auxiliar na correção de problemas posturais, ptose e dores ocasionadas pelo peso excessivo (GOBBI; LOUREIRO; PINTO, 1992 apud TAFURI; GOBBI, 2005).

A cirurgia indicada para redução do volume mamário é chamada de mamoplastia (ou mastoplastia) redutora e deve buscar também a proporção adequada entre mama e tórax, além de preservar sua função e sensibilidade e deixar cicatrizes discretas (PORTO et al., 2011).

Durante a fase de análise pré-operatória é imprescindível que o cirurgião obtenha informações quanto ao tamanho, formato e volume da região a ser reduzida. Entretanto, a estética da região das mamas é pouco compreendida e complexa de quantificar, o que gera dependência da análise subjetiva da aparência, além de dificultar a padronização das técnicas a serem utilizadas (KIM et al., 2008).

## 2.3 Processamento de imagens e segmentação das mamas

O processamento digital de imagens é uma área cada vez mais em evidência na sociedade moderna, devido principalmente ao avanço tecnológico. Uma imagem digital se configura basicamente como uma função discreta e finita  $f(x,y)$ , onde  $x$  e  $y$  referem-se às coordenadas espaciais em um ponto qualquer da imagem e a amplitude de  $f$  é a intensidade de luz nesse ponto. Cada elemento de uma imagem digital é comumente chamado de *pixel* (GONZALEZ; WOODS, 2000).

Processar uma imagem significa basicamente aplicar transformações sucessivas, lineares ou não, com o objetivo de extrair informações nela presentes com maior facilidade (ALBUQUERQUE; ALBUQUERQUE, 2000).

O processamento de imagens é normalmente composto por cinco etapas bem definidas: aquisição, pré-processamento, segmentação, representação (e/ou descrição) e interpretação (ALBUQUERQUE; ALBUQUERQUE, 2000).

A **aquisição** se refere à obtenção da imagem natural, isto é, a captura dela no mundo real e sua conversão para o digital. Essas imagens são geralmente obtidas por câmeras digitais, mas podem ser obtidas através de scanners, radares, radiografias, simulações, entre outras (SAMPAIO et al., 2009). Neste trabalho, foi utilizada uma base de imagens cosméticas frontais de mamas, obtida através de três testes clínicos, disponível na internet (GUO et al., 2022).

Na etapa de **pré-processamento** as imagens são tratadas de modo a garantir o sucesso nos passos seguintes. Nesta etapa, incluem-se tópicos como a diminuição de ruídos, realce, restauração e redimensionamento. Para o escopo da arquitetura do projeto a que este trabalho se limita foi utilizada apenas a técnica de redimensionamento, de modo a padronizar as imagens de entrada do modelo.

A **segmentação** é uma das etapas mais complexas de implementar, pois está diretamente relacionada às características intrínsecas ao domínio do problema, difíceis de serem traduzidos para a máquina (SAMPAIO et al., 2009).

É nessa etapa que há a separação da região de interesse (*Region of Interest* - ROI) do restante da imagem, o que deixa os processamentos subsequentes focados apenas no problema a ser solucionado. A saída da segmentação é geralmente um conjunto de dados primários em forma de *pixels* (ALBUQUERQUE; ALBUQUERQUE, 2000).

A **representação e descrição** servem para atribuir aos objetos de interesse um vetor de características, isto é, definir um conjunto descritivo de atributos mensuráveis, o que pode incluir perímetro, geometria, cor dos *pixels*, etc (SAMPAIO et al., 2009).

Por fim, a **interpretação** é a etapa onde é feito o reconhecimento dos objetos de acordo com suas particularidades e características, de modo a reuni-los em classes de acordo com os descritores obtidos no passo anterior (FREITAS, 2022).

Para as etapas de segmentação, representação e interpretação foi utilizado um modelo baseado em Redes Neurais Convolucionais (*Convolutional Neural Network* - CNN), de modo a agrupar os *pixels* correspondentes à região individual de cada uma das mamas.

## 2.4 *Deep Learning*

A tecnologia de aprendizado de máquina tem sido utilizada em larga escala, desde buscas na *web* até recomendações em sites de compras e filtros de conteúdo em redes sociais. Essas técnicas têm sido também aplicadas com êxito em identificação e segmentação de objetos em imagens, conversão de voz em texto, entre muitos outros exemplos (LECUN; BENGIO; HINTON, 2015).

No contexto específico de imagens médicas, nas últimas décadas, devido principalmente à grande variedade de patologias e potencial fadiga dos especialistas, a adoção de intervenções assistidas por computador, baseadas em técnicas de aprendizado de máquina, tem cada vez mais beneficiado pesquisadores e médicos (SHEN; WU; SUK, 2017).

Por muito tempo, entretanto, a construção de um sistema de aprendizado de máquina exigia considerável experiência no domínio do problema para tornar possível a conversão de uma série de dados brutos, como *pixels* de uma imagem de entrada, a priori sem significado para a máquina, em um vetor de características adequado para aprendizagem do modelo (LECUN; BENGIO; HINTON, 2015).

Diferentemente das técnicas convencionais, a arquitetura de *deep learning* utiliza uma série de métodos que permitem à máquina, a partir de dados brutos de entrada, estimar automaticamente as representações necessárias para a tarefa a que se destina, seja ela de detecção, classificação, segmentação, etc (LECUN; BENGIO; HINTON, 2015).

A menor dependência de intervenção humana na representação se deve ao fato do processo de aprendizagem ser feito de maneira hierárquica, onde o aprendizado de atributos de alto nível é uma composição do aprendizado de atributos de mais baixo nível. Além disso, geralmente encontra-se nesse tipo de arquitetura o que se define como aprendizagem por experiência, em que o conjunto de dados de treino percorre as camadas da rede várias vezes, em etapas iterativas chamadas de épocas (AGUIAR, 2017).

No contexto de *deep learning*, as Redes Neurais Convolucionais (do inglês *Convolutional Neural Networks* - CNN) são umas das redes mais representativas, especialmente nas áreas de visão computacional e processamento de linguagem natural (LI et al., 2021).

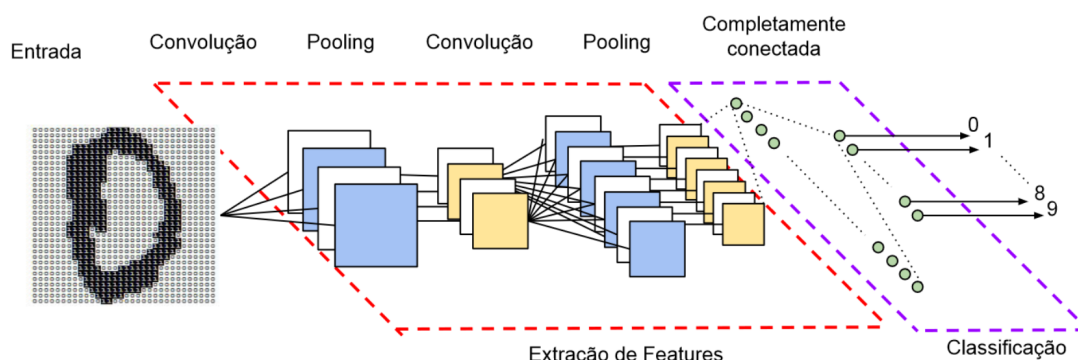
A popularização de sistemas baseados em CNNs se deve à conquista de feitos antes impensáveis, como reconhecimento facial, autonomia de veículos, tratamentos médicos inteligentes, além da evolução das *Graphical Processing Units* - GPUs, que dispõem atualmente de milhares de núcleos de processamento que podem otimizar as etapas de treinamento da rede ao

fazê-las em paralelo (AGUIAR, 2017; LI et al., 2021).

O funcionamento das CNNs é baseado no córtex visual dos animais. Uma CNN é capaz de aplicar filtros em imagens ao mesmo tempo que mantém a relação de vizinhança entre os seus *pixels* ao longo do processamento da rede (VARGAS; PAES; VASCONCELOS, 2016).

Redes neurais convolucionais são geralmente construídas com a composição de três tipos de camadas: camadas de convolução, camadas de *pooling* e camadas totalmente conectadas, conforme ilustrado na Figura 2 (SANTOS, 2021).

**Figura 2 – Ilustração das diferentes camadas de uma CNN**



**Fonte: Vargas, Paes e Vasconcelos (2016).**

### 2.4.1 Camada de convolução

Para o caso de aplicação em visão computacional, as camadas de convolução consistem basicamente de uma série de filtros aplicados à imagem de entrada, com o objetivo de extrair características. Cada uma dessas camadas é composta por diversos neurônios, que aplicam diferentes tipos de filtros em partes pequenas da entrada, onde a conexão de cada neurônio com o conjunto de *pixels* da camada anterior possui um peso atribuído, cujo conjunto representa um filtro de convolução da imagem no domínio espacial (VARGAS; PAES; VASCONCELOS, 2016).

Além disso, para garantir que sejam aplicados os mesmos filtros em diferentes posições da imagem é feito o compartilhamento dos pesos durante o treinamento, o que resulta em diminuição considerável do número de parâmetros a serem aprendidos e o tempo de treinamento da rede (VARGAS; PAES; VASCONCELOS, 2016).

As saídas produzidas pela aplicação das máscaras de filtragem, via operação linear de convolução, são matrizes intermediárias chamadas de mapa de recursos, onde cada elemento pode passar, opcionalmente, por uma função de ativação que introduz não-linearidade ao modelo antes da propagação para as outras camadas (LECUN; BENGIO; HINTON, 2015; SANTOS, 2021).

Em linhas gerais, especificamente em imagens, todos os componentes de um mapa de recursos utilizam o mesmo conjunto de pesos por duas razões principais: grupos de *pixels*



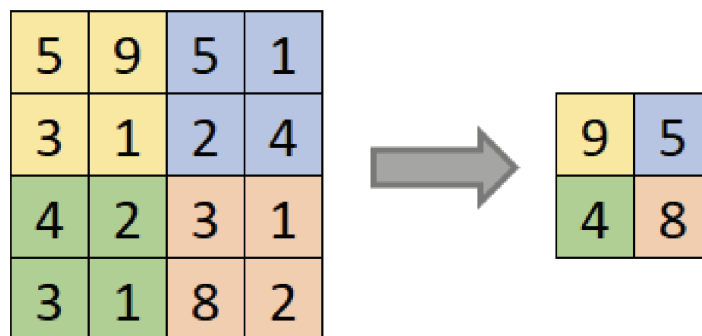
próximos na mesma vizinhança geralmente têm correlação alta entre si, de modo que diferentes elementos visuais significativos de partes específicas sejam facilmente detectados e pelo fato de estatísticas locais serem invariantes à localização (LECUN; BENGIO; HINTON, 2015).

### 2.4.2 Camada de *pooling*

As camadas de *pooling* são responsáveis por unir características semanticamente similares através da subamostragem dessas características por meio da substituição das saídas locais de pontos da rede por uma representação estatística que reduz as dimensões dos *inputs* (ALBAWI; MOHAMMED; AL-ZAWI, 2017; FERNANDES, 2021).

A função *max pooling* é uma das mais utilizadas em arquiteturas de CNN. Semelhantemente à aplicação de filtros espaciais, consiste em uma janela deslizante que percorre a matriz obtida na camada de convolução e aplica o operador *max*, que retorna o maior valor contido na vizinhança retangular de tamanho  $M \times N$  da janela e passo (ou *stride*) pré-definido (FERNANDES, 2021; SANTOS et al., 2020). A Figura 3 ilustra a aplicação da função de *max pooling* com janela de tamanho  $2 \times 2$ , em uma matriz de entrada  $4 \times 4$  e *stride* de 2.

**Figura 3 – Exemplo de aplicação de *max pooling***



Fonte: Santos et al. (2020).

### 2.4.3 Camadas totalmente conectadas

As camadas totalmente conectadas localizam-se geralmente na parte final da rede, onde utilizam-se as características extraídas previamente para obter as saídas de classificação do modelo (RODRIGUES, 2018).

As características obtidas nas camadas anteriores necessitam, normalmente, passar por um operador de *flattening*, que serve para vetorizar, isto é converter os dados, geralmente bidimensionais, em um arranjo unidimensional que será submetido à camadas densamente conectadas da rede (FERNANDES, 2021).

Nessa camada, espera-se que o modelo seja capaz de efetivamente classificar os atributos baseados em um conjunto de dados. Para que essa informação seja interpretada de maneira probabilística, utiliza-se a função *softmax*, definida conforme a Equação 2.1, que serve para

mapear a saída não-normalizada da rede para uma distribuição de probabilidades sobre as classes previstas (FERNANDES, 2021; SANTOS, 2021).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \text{ para } i = 1, \dots, k \text{ e } z = (z_1, \dots, z_k) \in \mathbb{R}^k \quad (2.1)$$

## 2.5 Métricas de avaliação

As métricas têm o papel de avaliar, sob diferentes aspectos, o desempenho de qualquer modelo desenvolvido (MALHOTRA et al., 2022). Para o caso específico de problemas de segmentação, algumas das mais populares são descritas a seguir.

Nas métricas discutidas abaixo, TP (*true positive*) e TN (*true negative*) significam os valores corretamente previstos, como positivos e negativos, respectivamente; enquanto FP (*false positive*) e FN (*false negative*) indicam resultados incorretamente previstos, onde FP indica um valor negativo erroneamente previsto como positivo e FN um valor real positivo classificado como negativo (MALHOTRA et al., 2022). Esses conceitos são obtidos através de uma matriz de confusão, ilustrada na Figura 4.

**Figura 4 – Matriz de confusão para duas classes**

		PREVISTO	
		POSITIVO	NEGATIVO
REAL	POSITIVO	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
	NEGATIVO	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

Fonte: O autor (2023).

**Precisão:** Representada na Equação 2.2, diz respeito a proporção de acerto no conjunto de dados reportados como positivos (VAKILI; GHAMSARI; REZAEI, 2020).

$$Precisao = \frac{TP}{TP + FP} \quad (2.2)$$

**Recall ou Sensibilidade:** Representada na Equação 2.3, se refere a porcentagem do total de resultados relevantes (positivos) que foram corretamente classificados pelo modelo (VAKILI; GHAMSARI; REZAEI, 2020).

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

**F1 Score:** Representada nas Equações 2.4 e 2.5, é definida como a média harmônica entre precisão e *recall* (VAKILI; GHAMSARI; REZAEI, 2020).

$$F1\ Score = \frac{2 * precisao * recall}{precisao + recall} \quad (2.4)$$

Em termos de TP, FN e FP, temos:

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \quad (2.5)$$

**Intersection over Union - IoU:** Também chamada de índice de Jaccard e representada na Equação 2.6, é obtida através da divisão do tamanho de interseção da máscara de base (A) e a máscara prevista na segmentação (B), pelo tamanho da união entre a máscara de segmentação e a máscara de base. Em linhas gerais, o IoU serve para medir a sobreposição entre a máscara de segmentação prevista pelo modelo e a máscara de base (LONG; SHELHAMER; DARRELL, 2015).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \text{ ou } \frac{TP}{TP + FP + FN} \quad (2.6)$$

**Coefficiente de Dice:** Representado na Equação 2.7, é definido como duas vezes o tamanho de interseção da máscara de base (A) e a máscara prevista na segmentação (B) dividido pela soma da quantidade de *pixels* na máscara prevista na segmentação e na máscara de base (BERTELS et al., 2019).

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (2.7)$$

Ainda na Equação 2.7, pode-se perceber, que  $|A \cap B| = TP$ , uma vez que a quantidade de *pixels* presentes na interseção entre máscara real com a máscara de saída do modelo representa os valores corretamente previstos.

Além disso, observa-se que  $|B| = TP + FP$ , uma vez que a quantidade de *pixels* presentes na máscara de saída corresponde ao conjunto de *pixels* corretamente e incorretamente previstos pelo modelo. E por fim, de maneira análoga, tem-se que  $|A| = FN + TP$ , que significa basicamente que a máscara de entrada é composta pelos valores corretamente previstos pelo modelo de saída, além dos *pixels* presentes na máscara real que o modelo deixou de prever.

Dessa forma, tem-se então que:

$$Dice = \frac{2TP}{FN + TP + TP + FP} = \frac{2TP}{2TP + FP + FN} \quad (2.8)$$

Logo, para o caso de segmentação de classes *booleanas*, isto é, representadas respectivamente, por uma classe de objeto de interesse e o restante é simplesmente *background* (não-objeto), o coeficiente de Dice corresponde ao mesmo valor de *F1 Score*, descrito na Equação 2.5.

**Mean Average Precision - mAP:** Métrica utilizada para avaliar a capacidade do modelo de fazer segmentação corretamente para diferentes IoUs. Em modelos de segmentação com apenas uma classe de objetos de interesse, o *mAP* é equivalente ao *AP* (*Average Precision*), que consiste em ordenar as predições através dos seus graus de confiança e calcular a área sob a curva de precisão-*recall*, conforme Equação 2.9 (DUMITRIU et al., 2023).

$$AP = \sum_n (Recall_n - Recall_{n-1}) \cdot Precisaon_n \quad (2.9)$$

As versões mais comuns de *mAP* são *mAP@50* (*AP* para *IoU* = 0.5) e *mAP@50 : 95* (média do *AP* para cada limiar de *IoU* de 0.5 à 0.95, com passo de 0.05, conforme Equação 2.10) (DUMITRIU et al., 2023).

$$mAP@50 : 95 = \frac{AP_{IoU=0.5} + AP_{IoU=0.55} + \dots + AP_{IoU=0.95}}{\text{quantidade de passos}} \quad (2.10)$$

## 2.6 Arquitetura de segmentação do YOLOv8

O *You Only Look Once* (YOLO) é um conjunto de algoritmos largamente utilizado no contexto de detecção e classificação de objetos, especialmente pelo fato de aliar velocidade e precisão. Foi lançado em 2015, por Redmon et al. (2016), e desde então tem sido bem-sucedido em executar tarefas diversas, que entre outras, incluem detecção de câncer e aplicações em robótica (TERVEN; CORDOVA-ESPARZA, 2023).

O YOLO é descrito como um detector *single-shot*, isto é, que percorre a matriz bidimensional que representa a imagem de entrada apenas uma vez. Inicialmente, divide-se a imagem em *grids* e, em seguida, utilizam-se de camadas convolucionais para prever as *bounding boxes*, retângulos utilizados para localização de objetos na imagem, e para estimar as probabilidades de classificação dos objetos da imagem entre as classes reais. Cada camada de convolução utiliza ainda *batch normalization* e função de ativação SiLU (ANG et al., 2023; REDMON et al., 2016).

Para cada uma das *bounding boxes* é calculado um grau de confiança e, em seguida, seleciona-se aquela que apresenta maior *score*. Pela forma como o YOLO é construído, geralmente acontece sobreposição entre as *bounding boxes* e, para manter apenas as que são mais promissoras, utiliza-se algoritmos de *non-maximum supression* (NMS) que, com base no IoU (Equação 2.6), servem para descartar sobreposições (ANG et al., 2023). A Figura 5 exemplifica a aplicação de NMS em uma típica imagem de detecção do YOLO.

O YOLOv8 é um modelo *state-of-the-art* – cuja arquitetura de detecção está ilustrada na Figura 6 – lançado pelo *Ultralytics* no presente ano (JOCHER; CHAURASIA; QIU, 2023). Ainda não há, até o momento, artigos oficiais do *Ultralytics* publicados sobre a arquitetura completa do YOLOv8. Em resumo, entretanto, sabe-se que a versão atual utiliza mecanismos de detecção *anchor-free*, onde a rede busca prever diretamente os objetos contidos na imagem, sem

**Figura 5 – Ilustração da aplicação de NMS**

Fonte: Terven e Cordova-Esparza (2023).

definir previamente candidatas a ROI, além de uma aprimorada rede de pirâmide de recursos e *loss function* modificada para obter melhor desempenho (DUMITRIU et al., 2023).

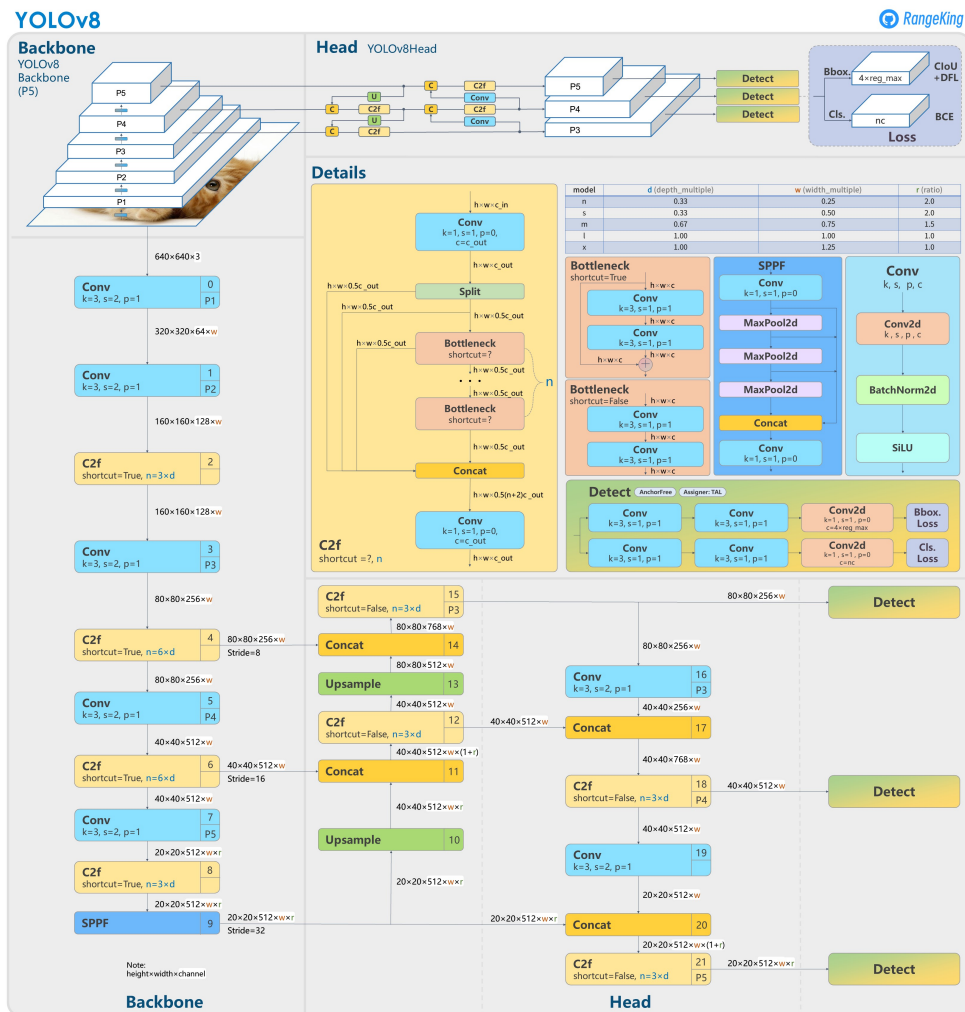
O YOLOv8 utiliza como *backbone* – parte da estrutura que serve como extrator de características da imagem – um modelo modificado do *CSPDarknet53*. O *CSPDarknet53* trata-se de uma evolução do modelo de CNN *Darknet53*, com a adição de Conexões Parciais entre Estágios (do inglês *Cross Stage Partial Connections* - CSPC), que servem para particionar o mapa de recursos da base, e por mesclá-los entre as camadas de maneira parcial por meio de uma hierarquia de estágio cruzado, que integra mapas de recursos do estágio inicial ao estágio final da rede (MAHASIN; DEWI, 2022; WANG et al., 2020).

Em seguida, são utilizados módulos C2f (2 modelos de convolução e *n bottlenecks*) introduzidos para otimizar a rede ao paralelizar mais fluxos de gradientes para obter informações mais ricas (YU; ZHOU, 2023).

Como é possível observar, ainda no diagrama da Figura 6, ao final do *backbone* há um processo chamado de SPPF, que se refere basicamente a uma implementação customizada, com menos operações de ponto flutuante por segundo (FLOPs), do *Spatial Pyramid Pooling* - SPP. Esse algoritmo faz o papel da camada de *pooling* de modo que não seja necessário limitar o tamanho da imagem (largura x altura) de *input* do modelo. Em linhas gerais, o SPP agrupa informações em estágios mais profundos da rede e gera saídas de comprimentos fixos, que serão passadas para as últimas camadas (HE et al., 2015).

Neste trabalho, utiliza-se o módulo YOLOv8-Seg, voltado para segmentação semântica, que é uma extensão do módulo de detecção. Ambos são iguais nas etapas previamente descritas e a diferença entre eles se dá pelo uso de dois *segmentation heads*, que aprendem a prever as máscaras de segmentação da imagem de entrada. Por fim, os *detection heads* são similares aos modelo da Figura 6, e consiste em cinco módulos de detecção e uma camada de predição (JOCHER; CHAURASIA; QIU, 2023).

Figura 6 – Arquitetura do YOLOv8

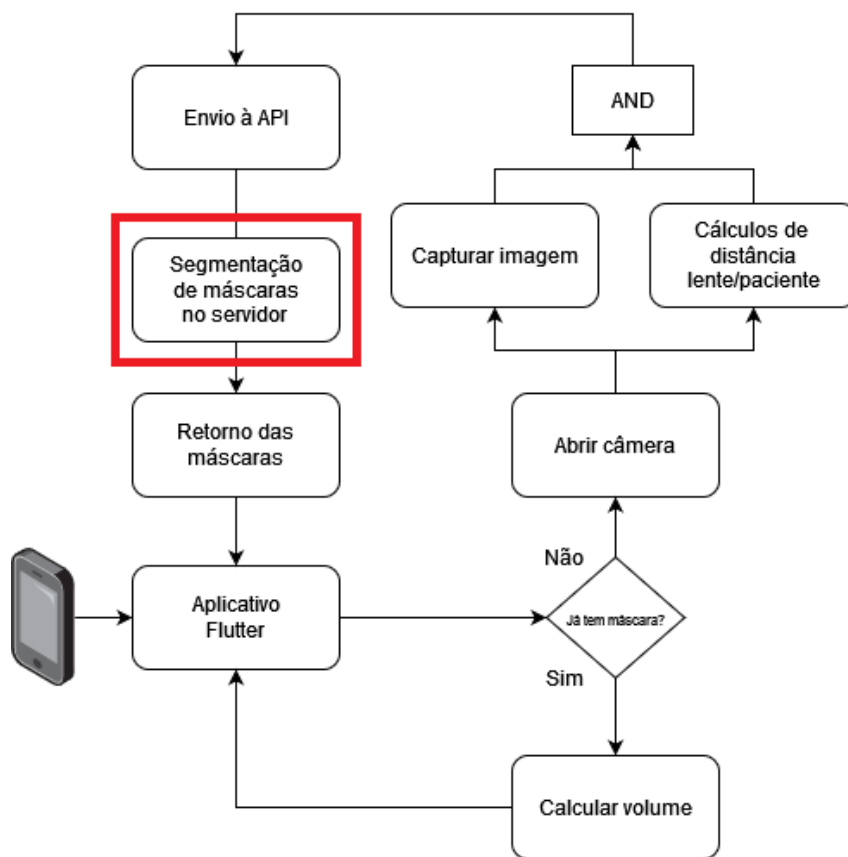


Fonte: RangeKing (2023).

### 3 METODOLOGIA

No contexto deste trabalho, é importante inicialmente ressaltar que o presente estudo está incluso em um projeto mais abrangente, cujo objetivo final é utilizar uma aplicação móvel integrada a uma API para calcular volume mamário, com base em uma imagem frontal que contenha as duas mamas e duas imagens laterais, cada uma contendo a região lateral das mamas esquerda e direita, individualmente. A aplicação deverá ser utilizada por cirurgiões plásticos capacitados, que terão a possibilidade de ajustar as máscaras, caso a predição não seja bem-sucedida. A arquitetura simplificada prevista para a aplicação, assim como em qual região este trabalho se situa, estão ilustradas na Figura 7. Portanto, todas as decisões aqui tomadas visam a viabilidade de tais funcionalidades no futuro.

Figura 7 – Diagrama simplificado do fluxo da aplicação



Fonte: O autor (2023).

#### 3.1 Ambiente e configuração

Implementou-se as etapas de pré-processamento, criação das máscaras necessárias para o treinamento, assim como o modelo, em Python 3.10. Utilizou-se o *framework* de visão computacional Ultralytics, nomeadamente o módulo YOLOv8-Seg, voltado para segmentação. A máquina utilizada para os treinamentos e testes possui as seguintes especificações: processador

Intel i5 9600K, 16GB de memória RAM, placa de vídeo Nvidia RTX 3060 e sistema operacional Windows 11. Todas as configurações foram feitas com a utilização de ambientes virtuais Python, criados com o gerenciador de distribuição Anaconda.

## 3.2 Aquisição de imagens

A base utilizada para treinamento e teste do modelo foi obtida de repositório público, na plataforma *CodeOcean*, e é oriunda de três estudos clínicos: NRG NSABP B39/RTOG 0413, NRG RTOG 1005 e NRG RTOG 1014, sob números de registro: NCT00103181, NCT01349322 e NCT01082211, respectivamente (GUO et al., 2022).

Os pesquisadores envolvidos na consolidação da base de dados, entre eles médicos e especialistas em biologia computacional, padronizaram as imagens, principalmente com ajuste de cores, além de anotar manualmente os landmarks dos contornos das mamas. Incluíram ainda as coordenadas dos retângulos de detecção com o algoritmo proposto no artigo associado à base.

Além disso, a equipe responsável optou pela exclusão completa de imagens muito ruidosas ou de baixa resolução da base, cortaram a imagem de modo a excluir a anatomia da clavícula e da região abaixo do umbigo, além da remoção de artefatos ruidosos provocados por problemas de iluminação ou inerentes ao ambiente de captura. Por fim, removeram todos os rótulos e características que pudessem vir a identificar os pacientes, a fim de preservar o sigilo.

## 3.3 Pré-processamento

### 3.3.1 Preparação dos dados

A base original possui 3762 imagens no total, em resoluções diferentes, com ausência de padronização prévia nesse sentido. O *framework* utilizado para a tarefa de segmentação, no entanto, utiliza internamente fator de escala quadrado para treinamento. A resolução utilizada foi 512x512 *pixels*, com o objetivo de preservar as informações corporais das pacientes na imagem. Por conta disso, optou-se por remover da base as imagens cujas resoluções fossem inferiores a 512x512, uma vez que o *upscale* de imagens menores geraria muitas distorções. Por esse critério, foram removidas 432 imagens da base.

### 3.3.2 Seleção dos dados e criação das máscaras

Nessa etapa foi feita a separação da base em treino e teste. Dada a relevante quantidade de imagens na base e o ambiente restrito do *dataset*, assim como a utilização de apenas duas classes (objeto e *background*), para o foco desta pesquisa utilizou-se 70% da base para treinamento do modelo (2331 imagens) e 30% para teste (999 imagens).



Por se tratar de um problema de segmentação, o treinamento da rede necessita, além da imagem de entrada, de máscaras pré-determinadas, de modo a indicar onde estão os objetos de interesse.

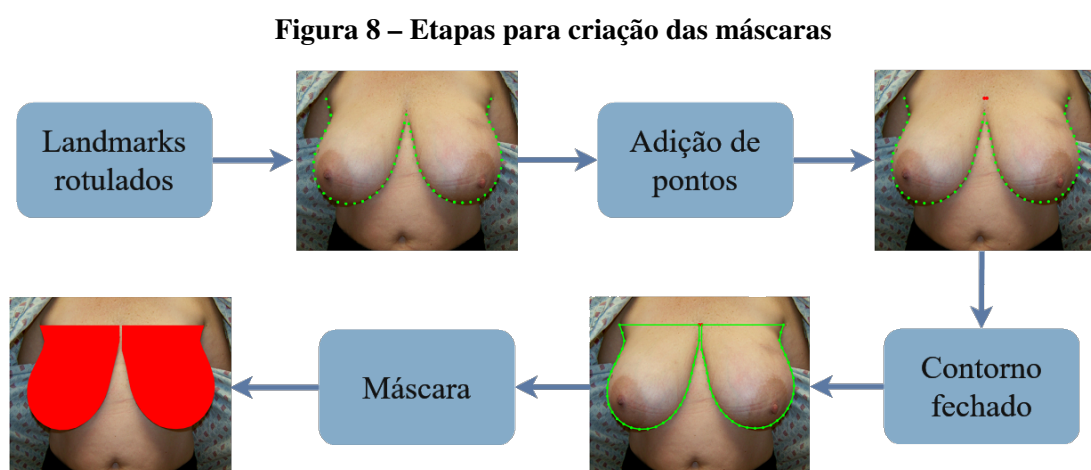
A base contém, juntamente às imagens, arquivos de texto com os *landmarks* dos contornos das mamas rotulados por profissionais. Cada arquivo possui 30 pontos igualmente espaçados, que representam o contorno externo de cada mama, ordenados de forma que o ponto inicial corresponda a região próxima às axilas e o ponto final a região próxima ao esterno.

Com o auxílio das bibliotecas OpenCV e NumPy, converteu-se cada arquivo de texto em um *array* de pontos, plotados sobrepostos à imagem da base a que se referem, conforme ilustrado na etapa “Landmarks rotulados” da Figura 8.

Observou-se, entretanto, que os *landmarks* referem-se apenas ao contorno aberto externo da mama e o modelo de segmentação aqui proposto, para o propósito de cálculo posterior de volume, precisa além de ser um contorno fechado, incluir toda a estrutura mamária.

De posse da lista com os trinta *landmarks* por mama, utilizou-se os pontos inicial e final, respectivamente, como coordenadas adicionais para delimitar a estrutura completa. Foi adicionado mais um ponto ao fim da lista, com a coordenada  $y$  do ponto inicial original e coordenada  $x$  do ponto final original. Esse processo está ilustrado na etapa “Adição de pontos” da Figura 8.

Com a inclusão da nova coordenada, foi possível delimitar a região de interesse completa. Apenas para fins ilustrativos, para a etapa de “Contorno fechado” da Figura 8, foram traçadas polilinhas, objetos compostos por segmentos de reta e arcos, para demonstrar o contorno completo da região desejada.



**Fonte: O autor (2023).**

Por fim, com base nos pontos de contorno, criou-se o polígono preenchido, também com o auxílio do OpenCV, que consiste na máscara necessária para treinamento do modelo, ilustrado na etapa “Máscara” da Figura 8.

### 3.3.3 Adaptação das máscaras para o formato do YOLOv8

Como dito anteriormente, a arquitetura utilizada foi baseada no YOLOv8 do *Ultralytics*, que padroniza a forma que as coordenadas das máscaras de *input*, isto é, as coordenadas das máscaras *ground truth*, devem estar descritas nos arquivos de texto. Cada imagem possui um arquivo de texto correspondente, e os rótulos têm que estar da seguinte ordem:

$$mascara = cls \ x_1 \ y_1 \ \dots \ x_n \ y_n \quad (3.1)$$

- *cls* é um número inteiro que representa a classe que a máscara pertence. No modelo aqui proposto existe apenas a classe "*breast*", definida como 0;
- $x_n$  representa a coordenada  $x$  normalizada do ponto  $n$ ;
- $y_n$  representa a coordenada  $y$  normalizada do ponto  $n$ .

Dessa forma, foi necessário normalizar os pontos dos contornos obtidos na subseção 3.3.2, de modo que cada coordenada tivesse seus valores limitados pelo intervalo  $[0.0, 1.0]$ . As normalizações são feitas com a estimativa de coordenadas relativas com base nas larguras e alturas das imagens originais. As Equações 3.2 e 3.3 foram utilizadas para essa etapa.

$$\text{coordenada normalizada } x = \frac{x}{\text{largura da imagem original}} \quad (3.2)$$

$$\text{coordenada normalizada } y = \frac{y}{\text{altura da imagem original}} \quad (3.3)$$

Via de regra, por cada imagem conter geralmente duas mamas, cada arquivo de texto possui duas linhas, onde cada linha representa uma máscara com as coordenadas normalizadas. Um exemplo desta etapa está ilustrado na Figura 9.

**Figura 9 – Exemplo de arquivo de texto com os *labels* para uma imagem**

```
0 0.18531228551818807 0.27413479052823314 0.18462594371997254 ... 0.2750455373406193
0 0.543582704186685 0.30510018214936246 0.5428963623884695 ... 0.30510018214936246
```

**Fonte: O autor (2023).**

## 3.4 Treinamento do modelo

Utilizou-se para treinamento o Ultralytics YOLOv8.0.128 com Python 3.10.11 e Torch-2.0.1+cu118, pacote compatível com os núcleos CUDA da Nvidia e que habilita treinamento em GPU.

Fez-se em seguida o carregamento do modelo pré-treinado YOLOv8-Seg *Medium*. Por padrão, o *framework* fornece uma série de modelos pré-treinados, e o *Medium* costuma ser recomendado pois tem o objetivo de fornecer equilíbrio entre velocidade e desempenho.

Para o processo de treinamento, utilizou-se os seguintes parâmetros:

- **data:** Inclui-se o *path* para o arquivo YAML que contém os caminhos para as pastas, com os rótulos de *train* e *test*. Cada pasta possui em seu interior outras duas subpastas, com as imagens e os rótulos (coordenadas das máscaras). Esse arquivo contém também o valor inteiro que representa a classe do objeto de interesse, aqui definido como 0, e para o nome da mesma, definido como “*breast*”;
- **epochs:** Representa o número total de iterações em que o conjunto de dados deve ser processado. Utilizou-se *epochs* = 50;
- **imgsize:** Tamanho para o qual a imagem será redimensionada após o SPPF. Utilizou-se *imgsize* = 512, isto é, serão utilizadas imagens 512x512 antes da entrada na camada totalmente conectada;
- **patience:** Quantidade máxima de épocas que o modelo deve percorrer a mais ao não perceber melhoria das métricas de avaliação, que ajuda a evitar *overfitting*. Utilizou-se *patience* = 10, isto é, 20% do total de épocas;
- **batch:** Indica o número de amostras processadas ao mesmo tempo antes de haver atualização dos pesos. Manteve-se o valor padrão: *batch* = 16;
- **save:** Valor *booleano* que indica se os *checkpoints* de treinamento devem ser salvos, além de realizar ou não predição após o treinamento. Manteve-se o valor padrão: *save* = *True*;
- **iou:** valor mínimo de IoU (Equação 2.6) para considerar a amostra como verdadeira. Foi utilizado *iou* = 0.7, que indica que a máscara predita e a máscara verdadeira devem ter pelo menos 70% de sobreposição.

O treinamento teve seu *early stopping* na época 48, uma vez que não houve melhoria nas métricas com base no *patience* = 10 e foi finalizado em cerca de 1h2min. O arquivo com melhor avaliação foi o de época 38, automaticamente salvo como “best.pt”.

Por fim, por conta do atributo *save* = *True*, o modelo, após o treinamento, automaticamente executou as validações e plotou os gráficos com as curvas de métricas, assim como as ilustrações de alguns resultados.

## 4 RESULTADOS E DISCUSSÕES

Para iniciar a discussão é necessário ressaltar que todo objeto detectado/segmentado pelo YOLO possui um grau de confiança atrelado. Por a aplicação proposta ser de caráter médico, necessita-se de grau de confiança tão alto quanto possível.

### 4.1 Rótulos de validação

Os rótulos de validação dizem respeito às máscaras *ground truth* das imagens de teste, que espera-se que estejam perfeitamente alinhadas, uma vez que dizem respeito às coordenadas rotuladas nas subseções 3.3.2 e 3.3.3. A Figura 10 ilustra um lote particularmente importante para a discussão.

Figura 10 – Exemplo de lote de validação



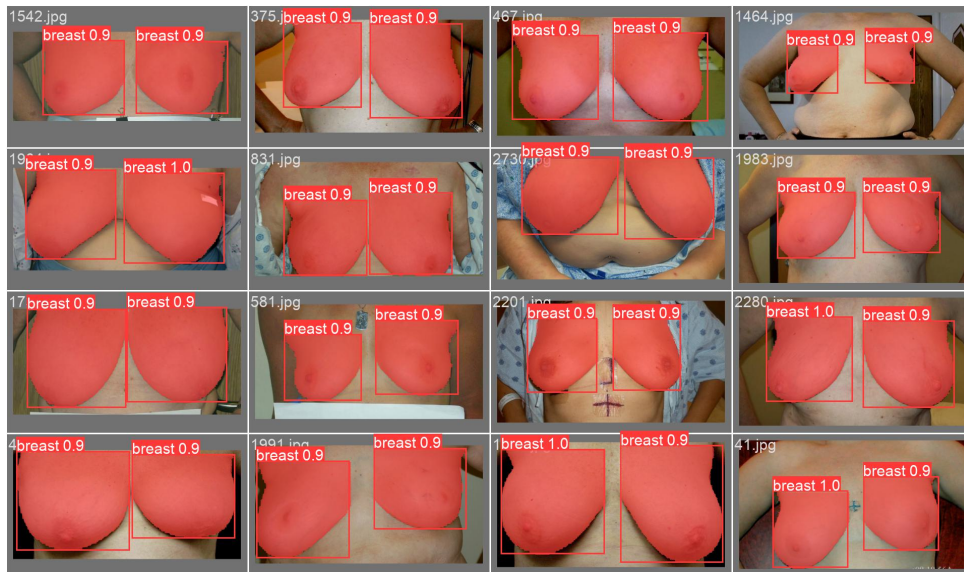
Fonte: O autor (2023).

Pode-se observar que duas imagens apresentam ausência de máscara fechada. Entretanto, os pontos de *landmarks* estão presentes no contorno da mama, o que indica falha provável ou no script de geração da curva fechada com adição de pontos extras, apenas em algumas amostras, ou na conversão das máscaras para o modelo normalizado do YOLO.

### 4.2 Rótulos de predição

Os rótulos de predição, por outro lado, se referem às máscaras previstas pelo modelo treinado, com as mesmas imagens da etapa anterior, o que dá uma ideia de como é a *performance* do modelo. A Figura 11 ilustra as máscaras para esse lote.

**Figura 11 – Exemplo de lote de predição**



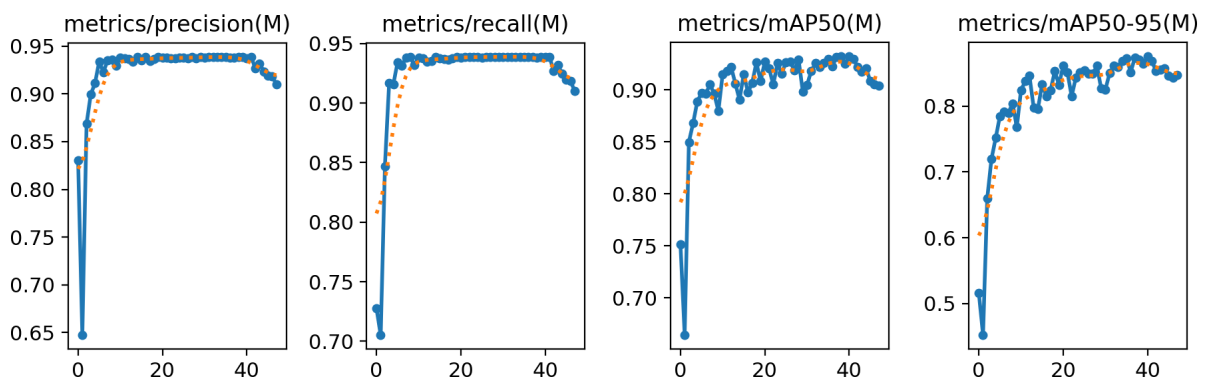
**Fonte: O autor (2023).**

Observa-se, que mesmo com o problema anterior da criação das máscaras em algumas imagens, o modelo treinado foi capaz de prever corretamente as regiões em questão, com bom grau de confiança.

### 4.3 Avaliação das métricas

A Figura 12 ilustra a evolução de algumas métricas durante a etapa de treinamento, onde observa-se que os melhores resultados ocorreram por volta da época 38, conforme anteriormente descrito.

**Figura 12 – Evolução das métricas por época durante o treinamento**

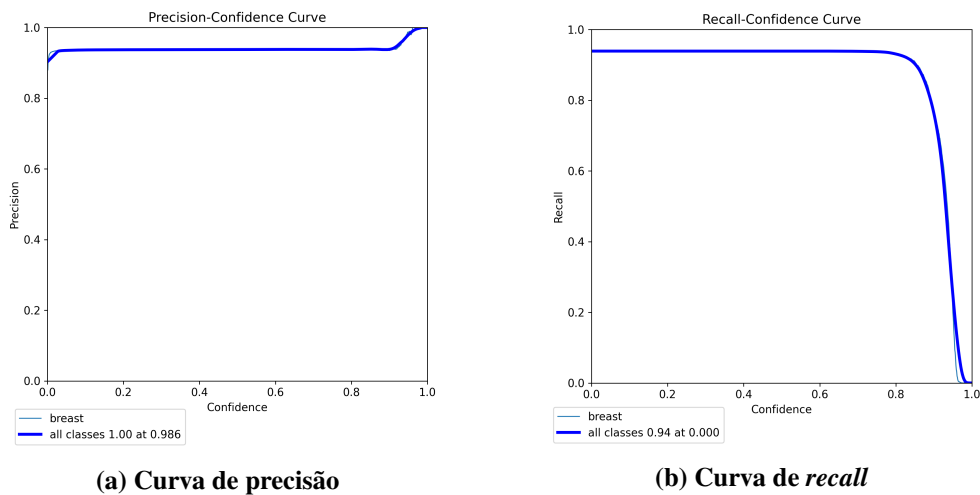


**Fonte: O autor (2023).**

### 4.3.1 Precisão e Recall

O modelo treinado obteve precisão de 93.8% e *recall* de 93.9% para as máscaras, mas é importante compreender as curvas das respectivas métricas, que revelam informações sobre como o modelo se comporta a depender do grau de confiança desejado.

**Figura 13 – Curvas de precisão e *recall* em função da confiança**



(a) Curva de precisão

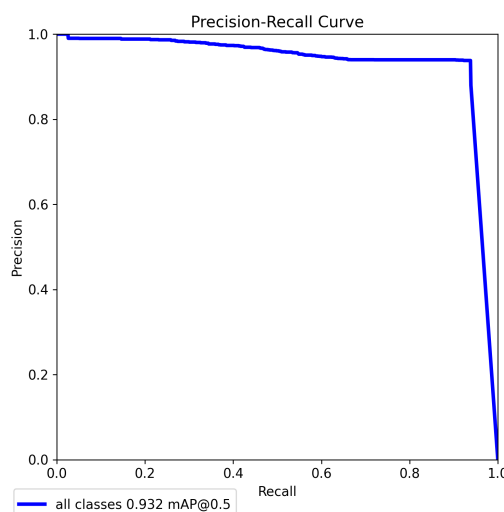
(b) Curva de *recall*

Fonte: O autor (2023).

A Subfigura 13a indica que, para um grau de confiança igual ou superior a 0.986, o modelo atinge precisão de 100%, mas isso impacta negativamente o *recall*, ilustrado na subfigura 13b, por conta do *trade-off* entre ambos.

A curva de precisão-*recall* está descrita na Figura 14, para o  $mAP@50$  e indica que o valor máximo para precisão e *recall* é em cerca de 0.932, para que o *trade-off* seja mínimo.

**Figura 14 – Curva de precisão-*recall***

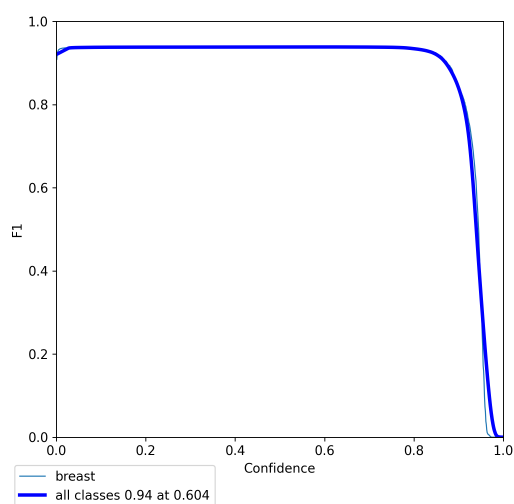


Fonte: O autor (2023).

### 4.3.2 mAP@50, mAP@50:95 e F1 Score

O modelo obteve mAP@50 de 0.932, mAP@50:95 de 0.874 e F1 Score de 0.938  $\approx$  0.94 para um grau de confiança de aproximadamente 0.604, conforme ilustrado na Figura 15.

**Figura 15 – F1 Score em função da confiança**



**Fonte: O autor (2023).**

## 5 CONCLUSÃO

Nesse estudo, apresentou-se uma proposta para segmentação frontal completa da estrutura mamária feminina, validada com as principais métricas encontradas na literatura, o que demonstra o potencial do YOLOv8 para segmentação de mamas. Vale ressaltar que não foi feito *data augmentation* pois a base foi diversa e bem rotulada o suficiente para o caso de uso proposto, uma vez que a aplicação será utilizada por médicos, em ambiente controlado.

Para trabalhos futuros deseja-se empregar esse mesmo modelo para imagens laterais de mamas, que devem ser coletadas por profissionais, para avaliar a capacidade de generalização da rede, uma vez que estão presentes elementos similares de cor e textura, que estão dispostos em ângulos completamente diferentes, com diferença de vizinhança de *pixels*, entre outros fatores. Além disso, deseja-se integrar o modelo a uma API, que será consumida por uma aplicação móvel para calcular o volume mamário.



## REFERÊNCIAS

- AGUIAR, R. d. L. **Classificação de gêneros musicais utilizando convolutional neural network e data augmentation**. Dissertação (Mestrado) — Universidade Estadual de Maringá, 2017.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: IEEE. **2017 international conference on engineering and technology (ICET)**. [S.l.], 2017. p. 1–6.
- ALBUQUERQUE, M. P. de; ALBUQUERQUE, M. P. de. **Processamento de imagens: métodos e análises**. Centro Brasileiro de Pesquisas Físicas MCT, 2000.
- ANG, G. J. N. et al. A novel application for real-time arrhythmia detection using yolov8. **arXiv preprint arXiv:2305.16727**, 2023.
- BERTELS, J. et al. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In: SPRINGER. **Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22**. [S.l.], 2019. p. 92–100.
- BLAND, K. I. et al. **The Breast-E-Book: Comprehensive Management of Benign and Malignant Diseases**. [S.l.]: Elsevier Health Sciences, 2017.
- BODINE, A. M.; HOLAHAN, B.; MIXON, A. Benign breast conditions. **Journal of Osteopathic Medicine**, American Osteopathic Association, v. 117, n. 12, p. 755–760, 2017.
- BULSTRODE, N.; BELLAMY, E.; SHROTRIA, S. Breast volume assessment: comparing five different techniques. **The Breast**, Elsevier, v. 10, n. 2, p. 117–123, 2001.
- CHAN, H.-P. et al. Deep learning in medical image analysis. **Deep Learning in Medical Image Analysis: Challenges and Applications**, Springer, p. 3–21, 2020.
- DABBAH, A. et al. Reduction mammoplasty: an outcome analysis. **Annals of plastic surgery**, v. 35, n. 4, p. 337–341, 1995.
- DESCAMPS, M. J. et al. A formula determining resection weights for reduction mammoplasty. **Plastic and reconstructive surgery**, LWW, v. 121, n. 2, p. 397–400, 2008.
- DUMITRIU, A. et al. Rip current segmentation: A novel benchmark and yolov8 baseline results. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2023. p. 1261–1271.
- FERNANDES, V. R. M. **Uma metodologia para a criação de arquiteturas de redes neurais convolucionais otimizadas para a detecção e o diagnóstico entre pneumonias**. Tese (Doutorado) — Universidade Federal do Maranhão, 2021.
- FONSECA, C. C. et al. Breast hypertrophy, reduction mammoplasty, and body image. **Aesthetic surgery journal**, Oxford University Press US, v. 38, n. 9, p. 972–979, 2018.
- FREITAS, G. P. C. **Diagnóstico de glaucoma via deep learning**. UFMA, 2022.
- GEFEN, A.; DILMONEY, B. Mechanics of the normal woman’s breast. **Technology and Health Care**, IOS Press, v. 15, n. 4, p. 259–271, 2007.

- GONZÁLEZ, D.; ARTETA, B. M. M. Dolor crónico asociado con hipertrofia mamaria. **Revista Dolor**, v. 8, n. 2, p. 94–1, 2019.
- GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. [S.l.]: Editora Blucher, 2000.
- GUO, C. et al. A fully automatic framework for evaluating cosmetic results of breast conserving therapy. **Machine learning with applications**, Elsevier, v. 10, p. 100430, 2022.
- HE, J. et al. The practical implementation of artificial intelligence technologies in medicine. **Nature medicine**, Nature Publishing Group US New York, v. 25, n. 1, p. 30–36, 2019.
- HE, K. et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 37, n. 9, p. 1904–1916, 2015.
- JOCHER, G.; CHAURASIA, A.; QIU, J. **YOLO by Ultralytics**. 2023. <<https://github.com/ultralytics/ultralytics>>. Acesso em 14 de julho de 2023.
- KHAN, Y. S.; SAJJAD, H. Anatomy, thorax, mammary gland. 2019.
- KIM, M. S. et al. Assessment of breast aesthetics. **Plastic and reconstructive surgery**, NIH Public Access, v. 121, n. 4, p. 186e, 2008.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.
- LI, Z. et al. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, IEEE, 2021.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 3431–3440.
- MAHASIN, M.; DEWI, I. A. Comparison of cspdarknet53, cspresnext-50, and efficientnet-b0 backbones on yolo v4 as object detector. **International Journal of Engineering, Science and Information Technology**, v. 2, n. 3, p. 64–72, 2022.
- MALHOTRA, P. et al. Deep neural networks for medical image segmentation. **Journal of Healthcare Engineering**, Hindawi, v. 2022, 2022.
- PORTO, R. da R. et al. Impacto da mastoplastia redutora na qualidade de vida física e emocional. **Boletim Academia Paulista de Psicologia**, Academia Paulista de Psicologia, v. 80, n. 1, p. 112–120, 2011.
- RAMESH, A. et al. Artificial intelligence in medicine. **Annals of the Royal College of Surgeons of England**, Royal College of Surgeons of England, v. 86, n. 5, p. 334, 2004.
- RANGEKING. **Arquitetura do YOLO**. 2023. <<https://user-images.githubusercontent.com/27466624/239739723-57391d0f-1848-4388-9f30-88c2fb79233f.jpg>>. Acesso em 14 de julho de 2023.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 779–788.

- RODRIGUES, D. A. Deep learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento automotivo. Universidade Federal da Paraíba, 2018.
- RUEDA, J. D. A. Patología mamaria benigna: diagnóstico y tratamiento. **Revista Médicas UIS**, Universidad Industrial de Santander, v. 31, n. 3, p. 65–71, 2018.
- SAMPAIO, W. B. d. et al. Detecção de massas em imagens mamográficas usando redes neurais celulares, funções geoestatísticas e máquinas de vetores de suporte. Universidade Federal do Maranhão, 2009.
- SANTOS, D. C. et al. Real-time traffic sign detection and recognition using cnn. **IEEE Latin America Transactions**, IEEE, v. 18, n. 03, p. 522–529, 2020.
- SANTOS, E. **Uma aplicação móvel para classificação de imagens utilizando Deep Learning: Um Estudo de Caso sobre Doenças Pulmonares**. Dissertação (Mestrado) — Universidade Estadual do Maranhão, 2021.
- SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. **Annual review of biomedical engineering**, Annual Reviews, v. 19, p. 221–248, 2017.
- SILVA, T. S. da; OLIVEIRA, C. F. de. Doença benigna da mama. 2010.
- SINGH, K. A.; LOSKEN, A. Additional benefits of reduction mammoplasty: a systematic review of the literature. **Plastic and reconstructive surgery**, LWW, v. 129, n. 3, p. 562–570, 2012.
- TAFURI, L. S. d. A.; GOBBI, H. Hiperplasias epiteliais em espécimes de mamoplastia redutora estética bilateral e mamoplastia redutora contralateral a câncer de mama. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, SciELO Brasil, v. 41, p. 133–138, 2005.
- TERVEN, J.; CORDOVA-ESPARZA, D. A comprehensive review of yolo: From yolov1 and beyond. **arXiv preprint arXiv:2304.00501**, 2023.
- VAKILI, M.; GHAMSARI, M.; REZAEI, M. Performance analysis and comparison of machine and deep learning algorithms for iot data classification. **arXiv preprint arXiv:2001.09636**, 2020.
- VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the xxix conference on graphics, patterns and images**. [S.l.], 2016. v. 1, n. 4.
- WANG, C.-Y. et al. Cspnet: A new backbone that can enhance learning capability of cnn. In: **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops**. [S.l.: s.n.], 2020. p. 390–391.
- YU, G.; ZHOU, X. An improved yolov5 crack detection method combined with a bottleneck transformer. **Mathematics**, MDPI, v. 11, n. 10, p. 2377, 2023.