



Santiago Vallejo Silva

**Uma abordagem centrada em dados para o
aprimoramento de modelos de segmentação
com aprendizado profundo em imagens de
mamografia**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio.

Orientador : Prof. Alberto Barbosa Raposo
Co-orientador: Dr. Cesar Augusto Sierra Franco

Rio de Janeiro
Setembro de 2023



Santiago Vallejo Silva

**Uma abordagem centrada em dados para o
aprimoramento de modelos de segmentação
com aprendizado profundo em imagens de
mamografia**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Alberto Barbosa Raposo

Orientador

Departamento de Informática – PUC-Rio

Dr. Cesar Augusto Sierra Franco

Co-orientador

Departamento de Informática – PUC-Rio

Prof Paulo Ivson Netto Santos

Departamento de Informática – PUC-Rio

Prof Carlos Roberto Hall Barbosa

Programa de Pós-Graduação em Metrologia – PUC-Rio

Rio de Janeiro, 20 de Setembro de 2023

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Santiago Vallejo Silva

Possui graduação em Engenharia Mecatrônica pela Universidade Mariana (2021). Atualmente participa como bolsista no instituto Tecgraf e está no processo de finalização de um mestrado na area de computação gráfica no departamento de informática na Pontifícia Universidade Católica do Rio de Janeiro. Tem experiência na área de Automação e em programação em Python.

Ficha Catalográfica

Vallejo Silva, Santiago

Uma abordagem centrada em dados para o aprimoramento de modelos de segmentação com aprendizado profundo em imagens de mamografia / Santiago Vallejo Silva; orientador: Alberto Barbosa Raposo; co-orientador: Cesar Augusto Sierra Franco. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2023.

v., 55 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Machine Learning;. 2. Data-Centric;. 3. Segmentação semântica;. 4. Mamografia.. I. Barbosa Raposo, Alberto. II. Sierra Franco, Cesar Augusto. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Agradeço aos meus orientadores, sempre atentos e dispostos a solucionar as complicações que surgiram durante o desenvolvimento deste estudo. Da mesma forma, expresso minha gratidão aos participantes do projeto GE-Mammography pela colaboração essencial para a concretização dos objetivos deste trabalho.

Resumo

Vallejo Silva, Santiago; Barbosa Raposo, Alberto; Sierra Franco, Cesar Augusto. **Uma abordagem centrada em dados para o aprimoramento de modelos de segmentação com aprendizado profundo em imagens de mamografia**. Rio de Janeiro, 2023. 55p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A segmentação semântica das estruturas anatômicas em imagens de mamografia desempenha um papel significativo no apoio da análise médica. Esta tarefa pode ser abordada com o uso de um modelo de aprendizado de máquina, que deve ser capaz de identificar e delinear corretamente as estruturas de interesse tais como papila, tecido fibroglandular, músculo peitoral e tecido gorduroso. No entanto, a segmentação de estruturas pequenas como papila e peitoral é frequentemente um desafio. Sendo o maior desafio o reconhecimento ou detecção do músculo peitoral na vista craniocaudal (CC), devido ao seu tamanho variável, possíveis ausências e sobreposição de tecido fibroglandular. Para enfrentar esse desafio, este trabalho propõe uma abordagem centrada em dados para melhorar o desempenho do modelo de segmentação na papila mamária e no músculo peitoral. Especificamente, aprimorando os dados de treinamento e as anotações em duas etapas. A primeira etapa é baseada em modificações nas anotações. Foram desenvolvidos algoritmos para buscar automaticamente anotações fora do comum dependendo da sua forma. Com estas anotações encontradas, foi feita uma revisão e correção manual.

A segunda etapa envolve um downsampling do conjunto de dados, reduzindo as amostras de imagens do conjunto de treinamento. Foram analisados os casos de falsos positivos e falsos negativos, identificando as imagens que fornecem informações confusas, para posteriormente removê-las do conjunto.

Em seguida, foram treinados modelos usando os dados de cada etapa e foram obtidas as métricas de classificação para o músculo peitoral em vista CC e o IoU para cada estrutura nas vistas CC e MLO (Mediolateral Oblíqua).

Os resultados do treinamento mostram uma melhora progressiva na identificação e segmentação do músculo peitoral em vista CC e uma melhora na papila em vista MLO, mantendo as métricas para as demais estruturas.

Palavras-chave

Machine Learning; Data-Centric; Segmentação semântica; Mamografia.

Abstract

Vallejo Silva, Santiago; Barbosa Raposo, Alberto (Advisor); Sierra Franco, Cesar Augusto (Co-Advisor). **A data-centric approach to improving segmentation models with deep learning in mammography images**. Rio de Janeiro, 2023. 55p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The semantic segmentation of anatomical structures in mammography images plays a significant role in supporting medical analysis. This task can be approached using a machine learning model, which must be capable of identifying and accurately delineating the structures. However, segmentation of small structures such as nipple and pectoral is often challenging. Especially in the recognition or detection of the pectoral muscle in the craniocaudal (CC) view, due to its variable size, possible absences and overlapping of fibroglandular tissue.

To tackle this challenge, this work proposes a data-centric approach to improve the segmentation model's performance on the mammary papilla and pectoral muscle. Specifically, enhancing the training data and annotations in two stages. The first stage is based on modifications to the annotations. Algorithms were developed to automatically search for uncommon annotations depending on their shape. Once these annotations were found, a manual review and correction were performed.

The second stage involves downsampling the dataset, reducing the image samples in the training set. Cases of false positives and false negatives were analyzed, identifying images that provide confusing information, which were subsequently removed from the set.

Next, models were trained using the data from each stage, and classification metrics were obtained for the pectoral muscle in the CC view and IoU for each structure in CC and MLO (mediolateral oblique) views.

The training results show a progressive improvement in the identification and segmentation of the pectoral muscle in the CC view and an enhancement in the mammary papilla in the MLO view, while maintaining segmentation metrics for the other structures.

Keywords

Machine Learning; Data-centric; semantic segmentation; Mammography.

Sumário

1	Introdução	12
1.1	Motivação	12
1.2	Objetivo e questões de pesquisa	13
1.3	Caso de uso: segmentação de imagens de mamografia para análise de qualidade	14
1.4	Estrutura do trabalho	15
2	Trabalhos relacionados	16
2.1	Data-centric	16
2.2	Segmentação do músculo peitoral	17
3	Segmentação de Mamografia	20
3.1	Composição do dataset	20
3.2	Seleção da arquitetura da rede neural	22
4	Metodologia	24
4.1	Revisão do dataset	24
4.2	Criação de pipeline para correção de erros	25
4.3	Algoritmos para revisão de papila	25
4.4	Algoritmos para revisão de músculo peitoral	29
4.5	Correção das anotações	33
4.6	Redução de amostragem	36
5	Experimentos e resultados	39
5.1	Métricas de avaliação	39
5.2	Criação dos datasets	39
5.3	Resultados do Treinamento	40
5.4	Resultados no Conjunto de Validação	42
5.5	Resultados no Conjunto de Teste	45
5.6	Comparação de resultados Model-centric e Data-centric	46
6	Conclusões e trabalhos futuros	52
6.1	Principais contribuições	52
6.2	Limitações	52
6.3	Trabalhos futuros	53
7	Referências bibliográficas	54

Lista de figuras

Figura 1.1	Estruturas em vista CC e MLO	14
Figura 3.1	Anotação em imagens de mamografia em vista CC	21
(a)	Imagem Original CC	21
(b)	Polígono de Anotação CC	21
(c)	Máscara de anotação CC	21
Figura 3.2	Anotação em imagens de mamografia em vista MLO	22
(a)	Imagem Original MLO	22
(b)	Polígono de Anotação MLO	22
(c)	Máscara de anotação MLO	22
Figura 4.1	Fluxo do processo seguido	25
Figura 4.2	Exemplo de polígono criado a partir de pontos	26
Figura 4.3	Tipos de anotação de papila	27
Figura 4.4	Algoritmo para calcular distância entre papila e mama	28
Figura 4.5	Deteção de papila pequena	29
Figura 4.6	Anotação de papila repetida em uma mesma imagem	30
Figura 4.7	Peitoral pequeno	30
Figura 4.8	Peitoral repetido em uma mesma imagem	31
Figura 4.9	Exemplo de peitoral não convexo e peitoral convexo	33
Figura 4.10	Marca gerada nas anotações inconsistentes.	34
Figura 4.11	Modificações feitas no dataset	34
(a)	Total de modificações	34
(b)	Total de modificações por vista	34
Figura 4.12	Exemplo de cálculo de diferença de pixels	35
Figura 4.13	Caso de anotação esquecida em set de Treinamento	37
Figura 4.14	Imagem com alto contraste com anotação de músculo	38
Figura 4.15	Imagem com alto contraste sem anotação de músculo	38
Figura 5.1	Distribuição das imagens segundo o tamanho do músculo peitoral	40
(a)	Train	40
(b)	Validation	40
(c)	Test	40
Figura 5.2	Curvas de aprendizado vista CC	41
(a)	IoU Score	41
(b)	Loss	41
Figura 5.3	Curvas de aprendizado vista MLO	42
(a)	IoU Score	42
(b)	Loss	42
Figura 5.4	Matrizes de confusão vista CC no set de Validação	43
(a)	Baseline matriz	43
(b)	Etapa I matriz	43
(c)	Etapa II matriz	43
Figura 5.5	Resultados gráficos no set de Validação	44
(a)	Validação CC IoU	44

	(b) Peitoral CC Classificação	44
	(c) Validação MLO IoU	44
Figura 5.6	Matrizes de confusão vista CC no conjunto de teste	46
	(a) Baseline matriz	46
	(b) Etapa I matriz	46
	(c) Etapa II matriz	46
Figura 5.7	Resultados gráficos no conjunto de teste	47
	(a) Test CC IoU	47
	(b) Classificação peitoral CC	47
	(c) Test MLO IoU	47
Figura 5.8	Gráfico de Coordenadas Paralelas para Comparação de hiperparâmetros	49
Figura 5.9	Violin plots dos resultados de validação	50
	(a) Baseline	50
	(b) Etapa II	50

Lista de tabelas

Tabela 3.1	Resultados IoU vista CC	23
Tabela 3.2	Resultados IoU vista MLO	23
Tabela 4.1	Anotações modificadas em vista CC	35
Tabela 4.2	Anotações modificadas em vista MLO	36
Tabela 5.1	IoU vista CC no set de Validação	42
Tabela 5.2	Classificação do músculo peitoral no set de Validação	43
Tabela 5.3	IoU na vista MLO no set de Validação	44
Tabela 5.4	IoU vista CC no conjunto de teste	45
Tabela 5.5	Classificação do músculo peitoral no conjunto de teste	45
Tabela 5.6	IoU na vista MLO no conjunto de teste	46
Tabela 5.7	Ítems usados para combinações	47

Glossário

FPN – Feature Pyramid Network

IOU – Intersection Over Union

CC – Crânio-caudal

MLO – Oblíqua mediolateral

ACC – Accuracy

PREC – Precision

REC – Recall

TPR – True Positive Rate

TNR – True Negative Rate

FPR – False Positive Rate

FNR – False Negative Rate

1

Introdução

Em um modelo de aprendizado de máquina que é treinado para reconhecer padrões a partir de um conjunto de dados, um algoritmo obtém informações de dados fornecidos para posteriormente fazer previsões sobre novos dados. Os modelos de aprendizado possuem várias aplicações em diferentes campos, por exemplo na área médica pode ser usado como ferramenta de apoio para tomada de decisão, para agilizar procedimentos de análise ou diagnosticar doenças. Cada vez mais vemos diversas aplicações inovadoras que trazem contribuições significativas na indústria, educação e medicina. No entanto, o treinamento e a produção deste tipo de modelo dependem da quantidade e da qualidade dos dados coletados, bem como da existência de etiquetas adequadas para serem usadas pelos algoritmos de aprendizado de máquina.

No contexto da área médica, os dados para desenvolver um modelo de aprendizagem são geralmente obtidos de prontuários, se os prontuários contiverem dados visuais, como imagens, estes podem ser usados para criar modelos de classificação, detecção de objetos ou segmentação semântica. O estudo deste trabalho foca em segmentação semântica, que é uma técnica que permite atribuir uma determinada classe ou categoria aos pixels de uma imagem [6]. A segmentação semântica envolve a divisão de uma imagem em várias regiões ou objetos diferentes. Após atribuir uma classe a cada pixel, o modelo de segmentação é capaz de delinear as diferentes partes da imagem, criando contornos e realizando, assim, a segmentação completa da imagem.

1.1

Motivação

Este trabalho tem um foco em imagens de mamografias, nas quais um modelo de segmentação semântica pode ser uma ferramenta de grande relevância e ter mais de um uso, por exemplo, identificar diferentes tipos de doenças como câncer [7] ou, como neste caso, reconhecer as estruturas que compõem uma mama para realizar uma análise de qualidade.

A motivação para aprimorar o modelo de segmentação semântica utilizado está baseada no desenvolvimento de um modelo competente e preciso, pois a segmentação das estruturas da mama pode se tornar uma tarefa complexa em estruturas pequenas e difíceis de identificar, como a papila e o músculo peitoral na vista craniocaudal (CC). Por este motivo, esta dissertação

visa aprimorar os modelos de segmentação das estruturas mamárias a partir de uma abordagem data-centric que abrange a criação de algoritmos para a busca de anotações de baixa qualidade para posterior correção e a eliminação de informações ambíguas para a geração de um novo conjunto de dados de treinamento de alta qualidade.

1.2

Objetivo e questões de pesquisa

Para a criação do modelo é utilizada uma grande quantidade de dados para assegurar que o mesmo possa generalizar de forma eficaz a partir das informações fornecidas. No entanto, utilizar muitos dados no desenvolvimento de um modelo de segmentação de imagens pode levar à ocorrência de erros nas etiquetas feitas manualmente por um ou vários anotadores humanos. A tarefa de etiquetar imagens mamográficas pode ser complexa devido à presença de estruturas sutis e de tamanho reduzido, como a papila e o músculo peitoral, que necessitam de uma maior precisão. Devido a esta complexidade, podem se originar erros de diversas fontes, como diferenças de interpretação entre anotadores ou a geração de dados inconsistentes causados pelo esgotamento que o processo de etiquetagem manual traz, dado que é um trabalho repetitivo e prolongado. Erros como etiquetas pouco definidas, acidentais ou com ruídos podem piorar o desempenho do modelo, pois ele aprenderia com informações erradas causando resultados imprecisos. Por esse motivo, aprimorar os dados de treinamento é fundamental para garantir que o modelo seja treinado com informações confiáveis e de alta qualidade.

Com o intuito de melhorar o desempenho do modelo de segmentação utilizando um conjunto de dados de imagens de mamografia, este trabalho busca desenvolver algoritmos com abordagem centrada em dados, que consiste na otimização da qualidade dos dados utilizados para o treinamento. Segundo Miranda [5], o aprimoramento dos modelos de ML pode ser realizado com o aumento dos dados anotados. Porém, o foco deste trabalho é utilizar os dados disponíveis, realizando um processo de limpeza de dados (*Data Cleaning*) e correção de anotações, reconhecendo as imagens que compõem o conjunto de dados juntamente com os erros comuns e específicos nas etiquetas de cada uma das classes.

A pesquisa se concentrará em abordar duas questões de pesquisa: (a) Como a abordagem data-centric pode ser aplicada de maneira eficaz para aumentar a qualidade dos dados de treinamento a partir dos dados de mamografia disponíveis? e (b) Qual é o impacto da qualidade das anotações nas métricas de um modelo de segmentação de imagens mamográficas?

As questões de pesquisa propostas abrangem diretamente o impacto da qualidade dos dados de treinamento no desempenho do modelo.

1.3

Caso de uso: segmentação de imagens de mamografia para análise de qualidade

O intuito de realizar um modelo de segmentação semântica é reconhecer, em nível de pixel, as estruturas presentes em uma mamografia, tais como: músculo peitoral, papila e tecido fibroglandular. Posteriormente, a partir das segmentações das estruturas de interesse é possível realizar uma análise de qualidade de posicionamento seguindo os padrões de aquisição de imagem adequados para realização de avaliações médicas.

Nas duas vistas de mamografia utilizadas aparecem as mesmas estruturas, sendo mencionadas a seguir:

- Mama
- Tecido fibroglandular
- Peitoral
- Mamilo

A Figura 1.1 ilustra as estruturas de interesse para avaliação de qualidade de posicionamento.

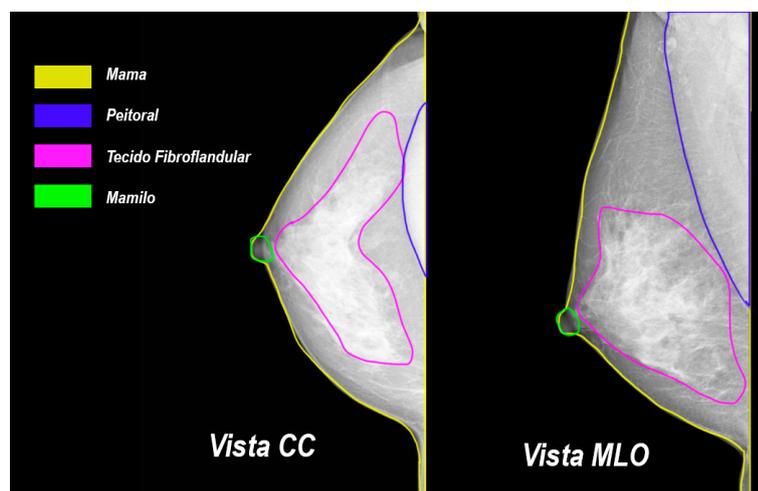


Figura 1.1: Estruturas em vista CC e MLO

O capítulo 3 fornecerá uma explicação mais aprofundada sobre a segmentação de imagens de mamografia. Também explicará com mais detalhe o tamanho e composição do conjunto de dados utilizado neste estudo e a seleção da rede neural para realizar os treinamentos.

1.4

Estrutura do trabalho

Esta dissertação está dividida em 5 capítulos: (1) Introdução, (2) Estado da arte, (4) Metodologia, (5) Resultados e (6) Conclusões. Na introdução, é apresentado o contexto da área do trabalho, a relevância do tema e os dados utilizados. No capítulo 2 é apresentado o estado da arte, é feita uma revisão da literatura para identificar os métodos desenvolvidos para diferentes tipos de dados na área de ML. Ao longo do capítulo 4 são detalhados os procedimentos para realizar o estudo da abordagem centrada em dados, descrevendo os algoritmos desenvolvidos. O capítulo 5 mostra todos os resultados obtidos, sobre os quais são analisadas as métricas dos modelos treinados. Para finalizar, no capítulo 6 são apresentadas as principais contribuições do trabalho, assim como as limitações do estudo e possíveis futuras pesquisas.

2

Trabalhos relacionados

Neste capítulo, é apresentada uma revisão de trabalhos relacionados que tratam de abordagens centradas em dados e técnicas de limpeza aplicadas a diferentes tipos de conjuntos de dados. Na seção 2.1, são expostos diversos estudos encontrados na literatura que fazem uso de processos centrados em dados. Na seção 2.2, são mostradas várias abordagens para a segmentação de mamografias com ênfase na segmentação do músculo peitoral.

2.1

Data-centric

Nesta seção, são apresentados trabalhos encontrados na literatura envolvendo abordagens centrada em dados aplicados a diferentes tipos de dados. Entre as técnicas utilizadas nas abordagens centradas nos dados estão: o aprimoramento de anotações geradas automaticamente, data augmentation e remoção de dados não representativos.

Chenaoua e seus colaboradores [1] apresentam um estudo explorando o uso de técnicas de limpeza de dados (*Data Cleaning*) em um conjunto de dados com anotações automáticas. Este conjunto de dados está composto por imagens de pessoas de diversas etnias e capturadas sob diferentes condições de iluminação. O objetivo principal deste estudo foi aprimorar o desempenho de um modelo de segmentação de pixels correspondentes à pele humana fazendo correções nos dados, uma vez que as anotações continham ruído devido a erros no processo de anotação. O processo de limpeza de dados realizado pelos autores envolveu a projeção de dois canais das imagens RGB e a criação de várias combinações, incluindo RG, RB e GB. Neste processo foram realizadas análises de histogramas 2D para extrair os picos dominantes de cada projeção. Como resultado dessas ações, os autores observaram uma redução significativa na quantidade de falsos positivos e falsos negativos nos resultados do modelo.

Ponzio e colaboradores [2] trabalharam em um estudo com imagens de tecido do cólon com o propósito de desenvolver um modelo capaz de classificar as imagens em três categorias (Adenocarcinoma, adenoma tubuloviloso e tecido saudável). A partir das imagens originais do tecido, eles realizaram a extração de regiões de interesse (ROIs). Estas ROIs representam pequenas porções de uma imagem que contêm informações relevantes [4]. O objetivo do projeto foi realizar um processo de limpeza de dados utilizando as ROIs, eliminando o *background* e outras partes das ROIs que continham informações irrelevantes.

Como resultado, eles geraram um novo conjunto de dados mais representativo.

O método apresentado no estudo de Liu [3] oferece uma abordagem interessante para realizar a limpeza de dados por meio do uso do *feedback*¹, que consiste no retorno de informações provenientes de um modelo de classificação de imagens. A principal meta desse estudo foi a detecção, dentro do conjunto de treinamento, das imagens que contribuíam menos para o processo de aprendizado, a fim de excluí-las do conjunto de treinamento. Esta abordagem resultou em um conjunto de dados mais eficaz, contendo somente as imagens que desempenham um papel relevante no aprimoramento do modelo de aprendizado de máquina, otimizando assim o processo de treinamento e melhorando a eficiência global do sistema.

No trabalho de Schmarje e colaboradores [16], foi desenvolvido um benchmark para classificação de imagens com dez conjuntos de dados, cada um contendo múltiplas anotações por imagem. Eles evidenciaram que uma única anotação não é suficiente e que o uso de múltiplas anotações melhora a aproximação da verdadeira distribuição de classes. Descobriram que as etiquetas rígidas não refletem a ambiguidade dos dados, resultando em modelos excessivamente confiantes. Demonstraram que o uso de etiquetas suaves é essencial para compreender a distribuição real e diminuir a sobreconfiança nos modelos. Identificaram a importância da qualidade dos dados nas métricas de classificação e distribuição, bem como a necessidade de métodos que operem em diferentes domínios de imagens. Seu trabalho propõe oportunidades para investigar métodos que abordem a estimação de ruído nas etiquetas e a aprendizagem semi-supervisionada para a classificação de imagens em problemas do mundo real.

Diferentemente dos trabalhos mencionados anteriormente, o presente estudo tem como objetivo a execução de um processo de *Data Cleaning* voltado para etiquetas de segmentação em imagens de mamografias. A abordagem adotada neste trabalho se fundamenta em análises detalhadas de características como forma, posicionamento e outras informações relevantes, especificamente relacionadas ao contexto do caso de uso selecionado

2.2

Segmentação do músculo peitoral

A segmentação semântica em mamografia é um tema que tem sido abordado por diversos autores utilizando técnicas de visão computacional e aprendizado de máquina. Dentre todos os trabalhos relacionados, há aqueles que focam na detecção e segmentação do músculo peitoral, tendo como

¹Retorno de informações

característica comum que na maioria deles o processo de segmentação é realizado em mamografias na vista MLO, pois esse tipo de vista proporciona melhor visualização das estruturas mamárias. Por outro lado, a mamografia CC mostra uma vista superior da mama, que pode fornecer informações importantes, mas não permite a análise das partes laterais. Devido às limitações desta vista, o número de estudos ou desenvolvimento de software para segmentação de mamografia para vista CC é significativamente menor em comparação com o número de artigos para a vista MLO.

Yongze Guo et ai. [9] aplicaram um processo em duas etapas para segmentar o músculo peitoral na vista MLO. Inicialmente, os autores utilizaram um modelo de aprendizado de máquina para localizar uma região da imagem contendo o músculo peitoral. Posteriormente, eles empregaram uma Rede Generativa Adversária (GAN) para criar uma imagem do segmento da forma do músculo peitoral. Essa abordagem possibilita uma segmentação precisa da estrutura, mesmo em casos de borrões ou manchas.

Dubrovina et ai. [10] implementou um modelo de aprendizado de máquina para segmentar o tecido fibroglandular na visualização MLO sem deixar de lado as outras estruturas principais. Em relação ao músculo peitoral, os autores enfatizam que essa estrutura está sempre localizada na mesma área, e o modelo pode detectá-la com precisão relativamente alta.

Usando outras técnicas, o estudo de Rubio e Montiel [11] mostra que uma abordagem model-centric pode contribuir para encontrar uma melhor estrutura de modelo para segmentar o músculo peitoral. Os autores compararam o comportamento e as métricas de diferentes combinações de arquiteturas de segmentação de dois estágios, uma para extração de recursos (VGG16 e ResNet50) e outra para segmentação (FCN-8, U-Net e PSPNet), usando o conjunto de dados mini-MIAS, que é composto de mamografias digitalizadas na visualização MLO. Depois de experimentar várias combinações usando o mesmo conjunto de dados, eles determinaram ResNet+PSPNet como a estrutura que produz os resultados com IoU mais altos e VGG+U-Net como aquela que produz os valores mais baixos em seu conjunto de dados.

Outros métodos de visão computacional incluem a classificação pixel a pixel usando uma modelagem Markov Random Field (MRF) [12]. Esta modelagem distribui as intensidades com base nas informações espaciais, auxiliando na busca da forma triangular curva do músculo peitoral na vista MLO e a forma semi-elíptica na vista CC. Este método é limitado a imagens sem ruído, pois pode falhar em imagens com sobreposição de tecido fibroglandular, dificultando a visualização do músculo.

Santle Camilus et al. [13] implementaram uma área de segmentação

peitoral em vista MLO usando a transformação watershed, que permite encontrar os contornos criando regiões de interesse e mesclando as regiões que correspondem ao músculo peitoral.

Recentemente, Sierra-Franco et al. [14] introduziram um grande conjunto de dados e experimentos baseados em aprendizado profundo para segmentação de imagens de mamografia, considerando visualizações MLO e CC. Os autores definem quatro estruturas de interesse, incluindo o músculo peitoral.

3

Segmentação de Mamografia

Neste capítulo, serão apresentados os detalhes da composição do dataset utilizado no treinamento do modelo, bem como as características da arquitetura utilizada para criar o modelo de segmentação, assim como os hiperparâmetros e configurações.

3.1

Composição do dataset

O dataset está composto por imagens de mamografias em vista Craniocaudal (CC) e Mediolateral Oblíqua (MLO) em formato DICOM, que é o formato mais comum utilizado para o armazenamento de exames médicos. Este tipo de arquivo permite armazenar informação do paciente, datas, equipamento utilizado e vários outros dados relevantes. A partir dos DICOMs são obtidas as informações de imagem para geração de imagens em formato PNG, as quais serão utilizadas como entrada para a rede neural. É importante mencionar que este conjunto de dados não considera imagens que apresentam implantes mamários, cirurgias complexas ou artefatos de imagem que complicam a visualização dos tecidos mamários e que está disponível apenas para membros do Instituto Tecgraf da Pontifícia Universidade Católica do Rio de Janeiro.

Como anotações de segmentação, foram utilizadas máscaras identificando a classe de cada pixel na imagem segundo a estrutura dentro da mamografia. Estas máscaras foram obtidas a partir de polígonos previamente desenhados por anotadores humanos utilizando uma ferramenta de anotação.

Foi selecionado um conjunto de 5137 imagens de mamografia com visualização CC e um conjunto de 5213 imagens com visualização MLO para construir o conjunto de dados, onde quatro estruturas foram consideradas para segmentação: mamilo ou papila mamária, músculo peitoral, tecido fibroglandular e tecido gorduroso. Todas as imagens coletadas foram anotadas usando uma ferramenta de desenho de contornos por uma equipe de anotadores treinados e supervisionados por especialistas clínicos, gerando um conjunto de anotações de polígonos densos para as estruturas mencionadas. Esta ferramenta permite desenhar contornos fechados para uma determinada estrutura. Os contornos são salvos como polígonos densos de alta resolução definidos no espaço da imagem. A ferramenta de anotação permite também uma inicialização automática do contorno da mama, acelerando o processo de

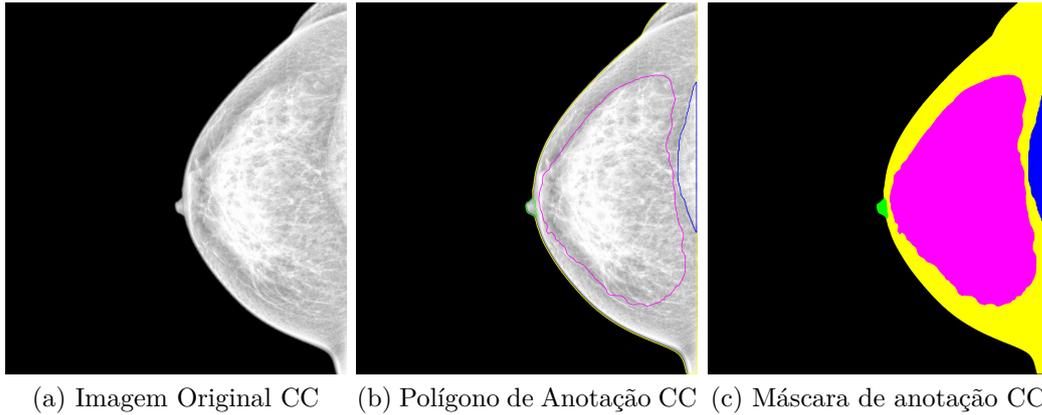


Figura 3.1: Anotação de segmentação em imagens de mamografia em vista CC. O músculo peitoral está colorido em azul, A papila em verde, o tecido fibroglandular em magenta, o tecido gorduroso em amarelo e o fundo (classe implícita) em preto.

anotação. Este contorno é feito através da segmentação da imagem usando a binarização Otsu [15] e computando o contorno externo do maior componente conexo. Como observação importante, a ferramenta permite criar polígonos baseados em aprendizado profundo realizados por dois modelos de segmentação provisionais treinados em versões parciais do conjunto de dados anotado.

Em seguida, os polígonos são rasterizados e exportados como anotações, onde cada pixel é atribuído a uma única classe, ou seja, à estrutura correspondente. A imagem de mamografia é normalizada usando os percentis 2 e 98 como valores mínimo e máximo, em seguida, equalizada usando a Equalização Adaptativa de Histograma Limitado por Contraste (CLAHE) [8], com o tamanho do kernel sendo $1/8$ da altura e largura da imagem completa, e finalmente reescalada para a faixa $[0, 255]$. As Figuras 3.1 e 3.2 mostram exemplos de imagem pré-processada e suas anotações correspondentes.

As amostras anotadas são divididas nos três subconjuntos padrão considerados em um fluxo de aprendizado supervisionado convencional: treinamento, validação e teste. A formação dos subconjuntos apresenta certo balanceamento em relação à densidade do tecido fibroglandular e no tamanho do músculo peitoral, assim, evitando *data leakage*,¹ que ocorre quando os dados de avaliação incluem informações que não estiveram disponíveis no set de treinamento. A distribuição das amostras para a vista CC resultou em 3737 amostras para treinamento ($\sim 70\%$), 943 amostras para validação ($\sim 20\%$) e 457 amostras para teste ($\sim 10\%$) e para a vista MLO em 3450 amostras para treinamento ($\sim 70\%$), 1206 amostras para validação ($\sim 20\%$) e 557 amostras para teste ($\sim 10\%$)

¹Vazamento de dados

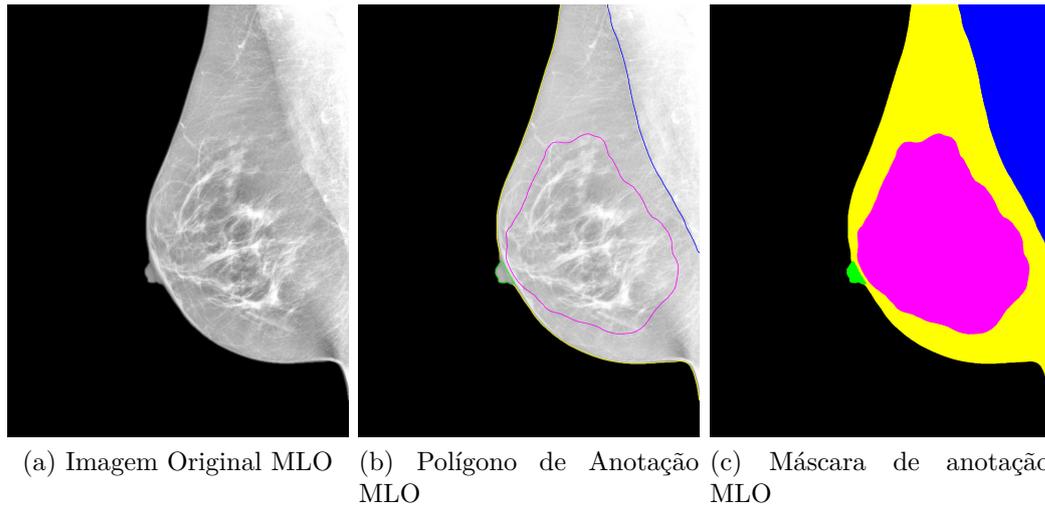


Figura 3.2: Anotação de segmentação em imagens de mamografia em vista MLO. Assim como na vista CC, a distribuição de cores na vista MLO é mantida para manter a uniformidade.

3.2

Seleção da arquitetura da rede neural

Devido às características distintas das vistas MLO e CC, foram treinados modelos separados para cada uma delas. Nos experimentos de treinamento, várias arquiteturas de segmentação semântica foram testadas, incluindo Unet, FPN, Linknet e PSPNet. Embora todas essas arquiteturas tenham o mesmo propósito, elas empregam estratégias diferentes. Foi selecionada a seguinte configuração de treinamento: a entrada do modelo é uma imagem única de 384×384 pixels, com valores de intensidade no intervalo $[0, 1]$. Assim, a imagem pré-processada deve ser redimensionada e reescalada para o intervalo e tamanho correspondentes. A saída do modelo é um tensor que representa os mapas de probabilidade das estruturas de interesse, incluindo o fundo. É utilizada a função de ativação softmax na camada final para ajudar o modelo a prever uma única classe para cada pixel, ou seja, um problema de múltiplas classes. Além disso, devido às capacidades das redes de segmentação que permitem a integração de um extrator de características convolucionais foi escolhido o EfficientnetB3 pela sua capacidade de extração de características juntamente com sua eficiência em relação ao uso de recursos computacionais.

Com base em experimentos de otimização de hiperparâmetros feitos em [14], foi utilizada a função de perda Jaccard com uma taxa de aprendizado de 0,001 e um batch size de 4. Foi implementado um esquema de parada antecipada com um máximo de 65 épocas e uma paciência de 20, selecionando os melhores pesos com base na função de perda no conjunto de validação.

As Tabelas 3.1 e 3.2 mostram os resultados do IoU por estrutura para

as vistas CC e MLO respectivamente. É possível observar valores de IoU adequados, acima de 0,7 para todas as estruturas segmentadas obtidas a partir do conjunto de teste. A rede PSPNet apresentou as métricas mais baixas, enquanto a FPN obteve os melhores resultados, com melhor detecção do músculo peitoral e papila.

Tabela 3.1: Resultados IoU vista CC

Arquitetura	Papila	Peitoral	Tecido	Mama	Mean
Unet	0,77	0,78	0,88	0,76	0,84
PSPNet	0,75	0,73	0,88	0,76	0,83
Linknet	0,76	0,75	0,90	0,81	0,84
FPN	0,77	0,79	0,90	0,81	0,85

Tabela 3.2: Resultados IoU vista MLO

Arquitetura	Papila	Peitoral	Tecido	Mama	Mean
Unet	0,70	0,95	0,90	0,80	0,87
PSPNet	0,71	0,96	0,89	0,80	0,87
Linknet	0,72	0,95	0,89	0,79	0,87
FPN	0,73	0,96	0,91	0,82	0,88

4 Metodologia

A abordagem centrada em dados proposta está dividida em duas etapas. Na primeira etapa, descrita da seção 4.1 até a seção 4.5, é apresentado todo o processo sobre busca, revisão e correção de anotações de músculo peitoral e papila. Na segunda etapa, que está focada no músculo peitoral, foi feita uma redução de tamanho no conjunto de imagens usadas para treinamento, deletando imagens que podem causar ruído a partir de uma análise visual.

4.1 Revisão do dataset

A ideia de modificar as etiquetas deste dataset surgiu a partir de uma verificação de erros. No início foi visto que algumas das etiquetas que foram feitas automaticamente usando uma primeira versão do modelo estavam erradas. Com uma revisão mais profunda identificou-se que a quantidade de etiquetas com erro era significativa; o modelo estava sendo treinado com etiquetas que na realidade são predicções, e muitas das etiquetas desse tipo apresentaram inconsistências.

Para realizar uma limpeza de dados, o primeiro passo é ter conhecimento sobre os principais tipos de erros de anotação que o dataset apresenta. Para reconhecer estes problemas foi feita uma revisão das anotações de cada estrutura, a qual foi feita em um subconjunto de dados. O critério para a seleção do subconjunto de dados levou em consideração os anotadores humanos envolvidos no processo de anotação. Em vez de corrigir os erros específicos de um único anotador, o objetivo foi identificar e abordar os erros gerais presentes nas anotações. Ao selecionar aleatoriamente partes das anotações de cada anotador, tornou-se possível detectar e abordar não apenas erros específicos de um indivíduo, mas também erros comuns a múltiplos anotadores.

Após analisar as anotações, foram reconhecidos os problemas mais comuns presentes nas anotações. Os principais erros foram encontrados nas anotações de papila e de músculo peitoral. Dentre esses erros, podem se mencionar os seguintes casos: segmentações pequenas, criação de polígonos acidentalmente, anotações repetidas e segmentações geradas automaticamente sem correção manual.

4.2

Criação de pipeline para correção de erros

Foi projetado um pipeline que leva em consideração o processo de início a fim para o aprimoramento de anotações. Este processo começa com a identificação de possíveis erros de anotação, correção de anotações e geração de estatísticas.

Como primeiro passo, é feita a preparação dos dados, na qual são reunidas todas as informações necessárias para executar os algoritmos de busca de erros. Com a execução dos algoritmos, estes geram anotações visíveis para os anotadores realizarem a revisão e, se for o caso, a respectiva correção da anotação. A imagem 4.1 mostra os passos seguidos no processo, desde a revisão do dataset até os treinamentos e análise de resultados.

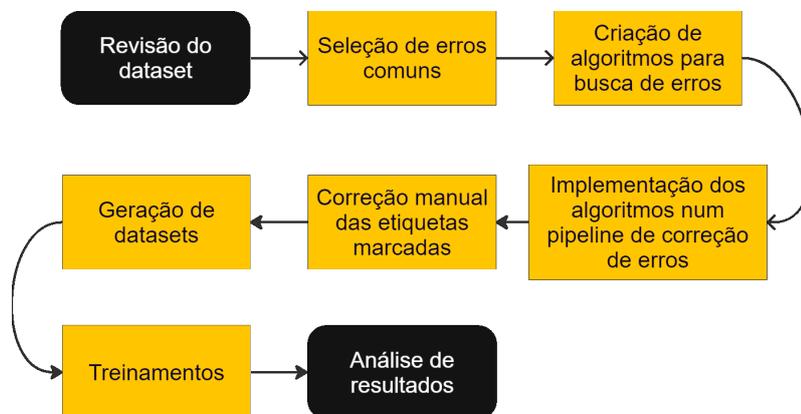


Figura 4.1: Fluxo do processo seguido

É importante mencionar que os algoritmos estão baseados na forma das anotações. As anotações dos polígonos são armazenadas em uma tabela que relaciona os dados dos polígonos com as imagens de mamografia. Assim, para cada algoritmo são utilizados os pontos nos eixos X e Y que formam o polígono. A imagem 4.2 ilustra como uma anotação é feita a partir de pontos que formam o polígono que é rasterizado para criar a máscara de anotação.

As próximas seções (4.3 e 4.4) descrevem os algoritmos criados para busca de erros na papila e no músculo peitoral, respectivamente.

4.3

Algoritmos para revisão de papila

– Segmentação automática:

O dataset contém casos de anotação automática que foram criados a partir de uma primeira versão de um modelo de segmentação. Esta primeira versão do modelo permite segmentar as mamografias automaticamente, reduzindo o tempo de anotação, mas deixando a

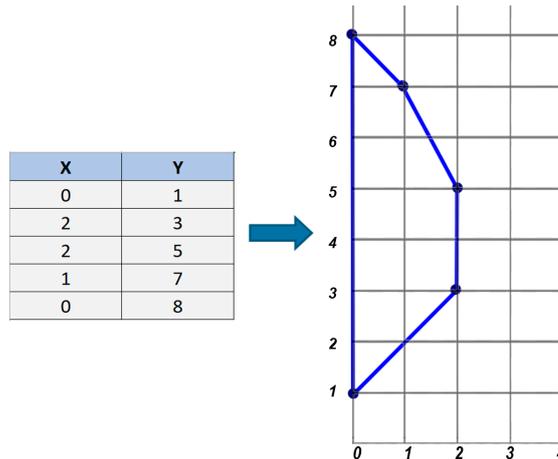


Figura 4.2: Exemplo de polígono criado a partir de pontos

critério do anotador o aperfeiçoamento das ditas estruturas. No entanto, o uso do modelo sem verificação humana pode resultar em novas anotações inconsistentes e imperfeitas.

As anotações das estruturas são compostas por listas de pontos em x e y . Esta lista de pontos foi utilizada para identificar se a anotação foi feita manual ou automaticamente. Percorrendo os pontos de uma anotação é possível identificar se é formada uma cadeia de pixels, o que significa que cada ponto está a uma distância de 1 pixel com relação aos outros. No caso da análise de pontos em direção diagonal, eles apresentam uma distância euclidiana de $\sqrt{2}$ pixels.

O algoritmo desenvolvido divide a lista de pontos ou perímetro P em 10 partes.

$$P = \text{Perímetro total do polígono.} \quad (4-1)$$

Estas partes são utilizadas para obter uma porcentagem de segmentação automática presente em uma anotação. Posteriormente é medida a distância entre pontos consecutivos em cada uma das partes sendo estas categorizadas em um dos três possíveis casos apresentados na figura 4.3

$$L_i = \frac{P}{10} \quad (\text{comprimento de cada parte } i). \quad (4-2)$$

Agora, para cada parte L_i , é calculada a distância euclidiana d_i entre pontos consecutivos nesta parte usando a fórmula:

$$d_i = \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2}, \quad \text{onde } j = 1, 2, \dots, N_j - 1. \quad (4-3)$$

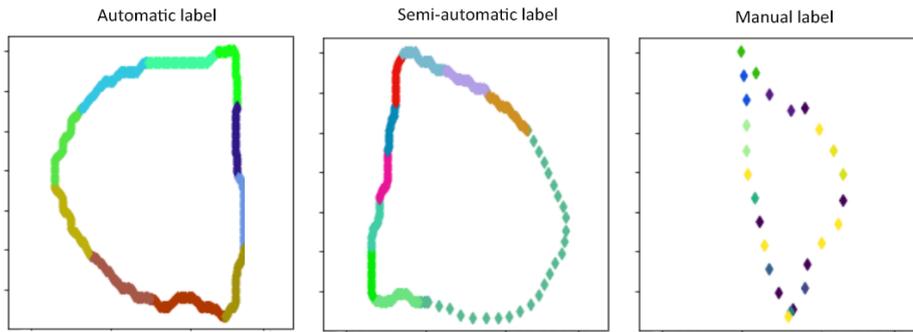


Figura 4.3: Tipos de anotação de papila

Onde N é a quantidade total de pontos da parte L_i e (x_j, y_j) são as coordenadas do ponto j na parte L_i do polígono, e (x_{j+1}, y_{j+1}) são as coordenadas do ponto seguinte ao ponto j na mesma parte L_i .

Em seguida, cada parte L_i é classificada como segmentada manualmente ($S_i = 0$) se a distância d_i nesta parte for maior que a raiz quadrada de 2 (a maior distância entre dois pontos separados por 1 unidade em uma direção diagonal), e como segmentada automaticamente ($S_i = 1$) se a distância for menor ou igual à raiz quadrada de 2.

Por fim, é calculada a porcentagem de segmentação automática (PSA) como:

$$PSA = \frac{s_1 + s_2 + \dots + s_{10}}{10} \quad (4-4)$$

Para classificar o polígono P em uma classe, as seguintes regras são levadas em consideração:

- Se em cada L_i existe uma distância maior que $\sqrt{2}$, o polígono será categorizado como anotação manual.
 - Se em apenas algumas das partes L_i existe uma distância maior que $\sqrt{2}$, a anotação será categorizada como anotação semi-automática
 - Se nenhuma parte L_i excede a distância máxima $\sqrt{2}$, é categorizada como segmentação automática.
- **Distância papila-mama:**

Na análise dos dados foi possível observar que várias das anotações automáticas da papila apresentam um erro recorrente. Este erro acontece quando a anotação não cobre a área completa da estrutura. Caso a área faltante sem segmentar seja grande, a anotação pode ser considerada como errada. Estas anotações podem ser encontradas a partir da medida entre a distância entre as bordas de papila e da mama.

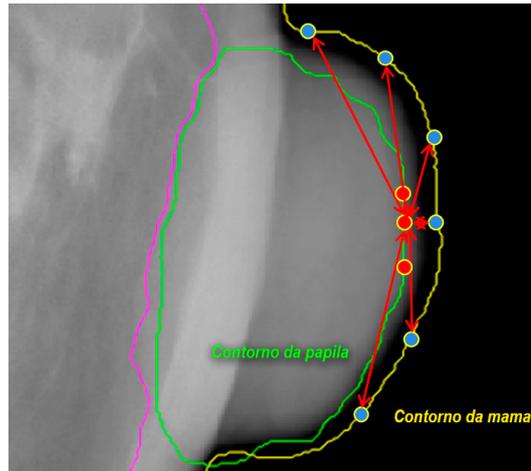


Figura 4.4: Algoritmo para calcular distância entre papila e mama

Na Figura 4.4, pode-se observar uma representação das informações utilizadas pelo algoritmo. O processo inicia selecionando o ponto mais distante no eixo X da papila, juntamente com outros dois pontos próximos, identificados como pontos vermelhos no exemplo da Figura 4.4. Inicialmente, foi explorada a possibilidade de utilizar o ponto mais próximo na direção à qual a papila está apontando para que o algoritmo não falhe em casos de papilas com formatos incomuns. No entanto, devido à forma padronizada da papila na maioria dos casos de mamografias, a diferença entre a seleção do ponto inicial da papila não se mostrou substancialmente representativa em nossos experimentos.

Na figura 4.4 pode-se observar uma representação das informações utilizadas pelo algoritmo. O algoritmo começa tomando o último ponto do eixo X da papila junto com outros dois pontos próximos, representados como pontos vermelhos no exemplo da figura 4.4. Para cada um destes pontos vermelhos se mede a distância entre eles e cada ponto da mama, representados como pontos azuis. Para cada ponto vermelho é obtida uma lista de distâncias das quais é utilizada unicamente a distância menor de cada uma delas, obtendo assim apenas 3 distâncias. Com as 3 distâncias menores é calculada a média. O resultado da média é uma estimativa em pixels da separação entre a borda da papila e da mama.

– **Papila pequena:**

Este algoritmo foi criado para identificar anotações de menor tamanho com suspeita de erro. Estas anotações podem ter sido feitas acidentalmente ou pelo uso do modelo para segmentar automaticamente e sem ter feito uma correção manual.

O algoritmo faz um cálculo da área sobre todas as anotações da papila, permitindo reconhecer desde as estruturas menores até as maiores. Para definir o que é considerado uma estrutura pequena, estabelecemos um critério: qualquer anotação com menos de 10000 pixels quadrados. Esta medida, ou threshold, foi determinada através de uma análise visual minuciosa.

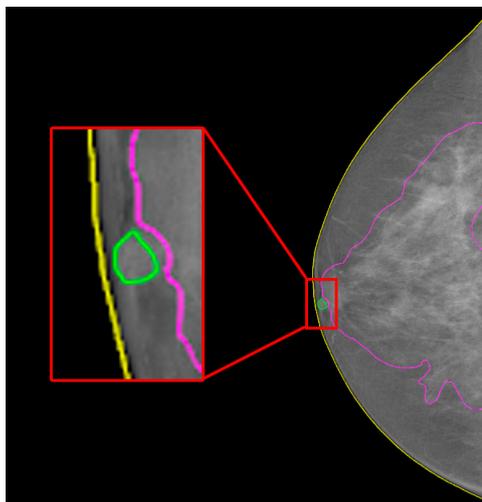


Figura 4.5: Detecção de papila pequena

– **Anotações repetidas:**

Uma imagem com mais de uma anotação por estrutura pode ser considerada como inconsistente e sua remoção pode evitar ambiguidades, por este motivo foi criado o algoritmo de busca de anotações repetidas. A busca das anotações é feita a partir de um ID único para cada estrutura. Se em uma imagem existem dois elementos com o mesmo ID, é porque existem duas anotações para uma mesma estrutura.

Na figura 4.6 é possível observar um exemplo de anotação repetida, tendo normalmente uma anotação correta e outra errada.

4.4

Algoritmos para revisão de músculo peitoral

– **Músculo peitoral pequeno:**

Foi adaptado para o músculo peitoral o algoritmo para encontrar estruturas pequenas. O algoritmo só foi utilizado em peitorais de imagens em vista CC, pois na vista MLO os peitorais têm um maior tamanho e não é comum que apresentem este tipo de erro. Porém, na vista CC foi visto que existem algumas estruturas ou áreas da imagem que foram confundidas com um músculo peitoral, por um anotador humano ou pelo

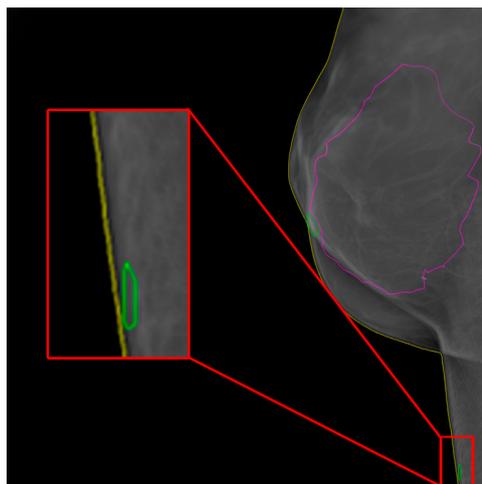


Figura 4.6: Anotação de papila repetida em uma mesma imagem

modelo de anotação automática. Um exemplo é mostrado na Figura 4.7, na qual a anotação automática segmentou um músculo peitoral em uma área muito distante de onde o músculo peitoral realmente está localizado. O critério para definir um músculo peitoral pequeno é o mesmo utilizado na papila: qualquer anotação com menos de 10000 pixels quadrados é considerada pequena.

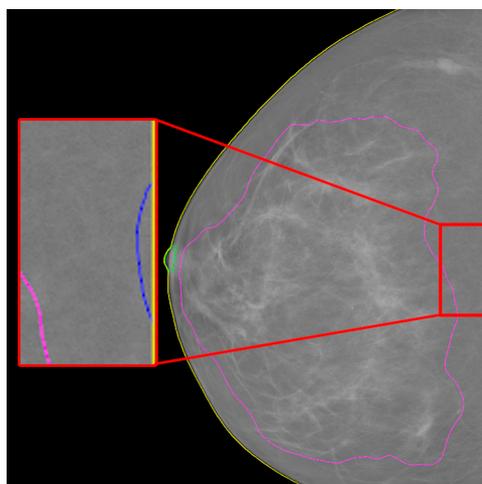


Figura 4.7: Peitoral pequeno

– **Anotações de peitoral repetidas:**

Também foi considerada a busca das imagens que apresentam mais de uma anotação de peitoral na mesma imagem. Utilizando este algoritmo foram descobertos alguns casos de imagens com anotação de peitoral repetida. Pode-se ver um exemplo na Figura 4.8.

– **Convex hull:**

Este algoritmo permite encontrar anotações de músculos peitorais que possuem um formato indesejável a partir de comparações com seu

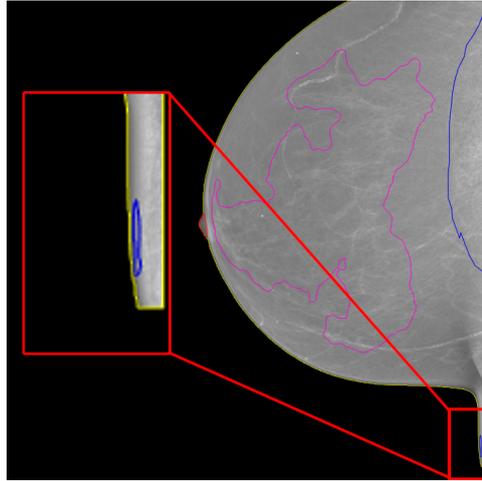


Figura 4.8: Peitoral repetido em uma mesma imagem

formato Convex Hull correspondente, que é o contorno convexo mais simples que engloba ou envolve todos os pontos que compõem um polígono. Portanto é capaz de identificar formas alongadas, além de criar uma representação convexa a partir dos pontos presentes na anotação e comparando a representação criada com a anotação atual, com o objetivo de identificar diferenças de área e diferenças de forma, permitindo encontrar anotações com formas côncavas. Este algoritmo é unicamente utilizado para o músculo em vista CC, pois é uma estrutura que usualmente apresenta uma forma convexa. As equações a seguir explicam o comportamento do algoritmo em detalhe.

Para este algoritmo, foi implementada uma função que retorna um valor entre 0 e 1. Os valores próximos a 0 indicam que a forma apresenta um comportamento inesperado, enquanto os valores próximos a 1 indicam que a forma é a esperada.

Para uma imagem, é denotada a rasterização de todas as suas instâncias de polígonos correspondentes como uma imagem binária \mathbf{P} e a função de análise de forma como s , que é uma função composta definida da seguinte forma:

$$s(\mathbf{P}) = \min\{s_{\text{ecc}}(\mathbf{P}), s_{\text{sol}}(\mathbf{P}), s_{\text{top}}(\mathbf{P}), s_{\text{hau}}(\mathbf{P})\}, \quad (4-5)$$

nesta equação, s_{ecc} , s_{sol} , s_{top} , e s_{hau} são funções para penalizar diferentes comportamentos indesejados da forma analisada.

s_{ecc} é uma função de pontuação que utiliza a excentricidade de \mathbf{P} para penalizar formas arredondadas. Pois é esperado que a projeção do músculo peitoral seja representada por formas alongadas. A pontuação é definida pelo seguinte kernel gaussiano:

$$s_{\text{ecc}}(\mathbf{P}) = 1 - \exp\left(-\frac{f_{\text{ecc}}(\mathbf{P})^2}{2\sigma_{\text{ecc}}^2}\right), \quad (4-6)$$

onde $f_{\text{ecc}}(\mathbf{P})$ calcula a excentricidade de uma elipse ajustada a \mathbf{P} e σ_{ecc} fixado em 0.3 é o desvio padrão do kernel.

De forma semelhante, s_{sol} utiliza a solidez de \mathbf{P} para penalizar formas cuja área difere da forma do correspondente Convexhull (CH). Essa função é definida da seguinte maneira:

$$s_{\text{sol}}(\mathbf{P}) = \exp\left(-\frac{|f_{\text{sol}}(\mathbf{P}) - 1|^2}{2\sigma_{\text{sol}}^2}\right), \quad (4-7)$$

Nesta equação, $s_{\text{sol}}(\mathbf{P})$ é a função que determina a solidez da forma \mathbf{P} em relação ao seu Convex Hull (CH). Na mesma equação, $|f_{\text{sol}}(\mathbf{P}) - 1|^2$ representa a diferença entre a solidez calculada e um valor de referência (neste caso, 1). E é elevado ao quadrado para garantir que o resultado seja sempre positivo. σ_{sol} é o desvio padrão do kernel e foi fixado em 0.1. Este valor influencia a forma como a penalização é aplicada com base na diferença de áreas entre a forma e seu Convex Hull (CH). O cálculo de s_{sol} é semelhante ao cálculo do IoU, porém, o IoU é uma métrica de avaliação comum para sobreposição de regiões em tarefas de segmentação de objetos, enquanto $s_{\text{sol}}(\mathbf{P})$ é uma função destinada a penalizar formas com áreas muito diferentes do seu Convex Hull (CH)

Continuando com o terceiro elemento da equação 4-5, s_{top} é um valor booleano que penaliza topologias inesperadas de \mathbf{P} . Assim, s_{top} retorna um se \mathbf{P} apresentar um único componente conectado sem furos; caso contrário, retorna zero.

Por fim, em lugar de penalizar a diferença de área entre \mathbf{P} e sua forma Convex Hull (CH), s_{hau} penaliza a distância entre os contornos correspondentes, permitindo detectar regiões altamente côncavas. A pontuação também é modelada como um kernel gaussiano:

$$s_{\text{hau}}(\mathbf{P}) = \exp\left(-\frac{d_{\text{hau}}(\mathcal{C}(\mathbf{P}), \mathcal{C}(\mathbf{P}_{\text{CH}}))^2}{2\sigma_{\text{hau}}^2}\right), \quad (4-8)$$

d_{hau} calcula a distância de Hausdorff direcionada para um par de contornos, \mathcal{C} calcula os contornos de uma imagem binária, \mathbf{P}_{CH} é a imagem Convex Hull (CH) de \mathbf{P} , e σ_{hau} é o desvio padrão do kernel fixado em 0.012 vezes a largura da imagem.

Na parte esquerda da Figura 4.9, pode se observar um exemplo de quais anotações são consideradas como não convexas.

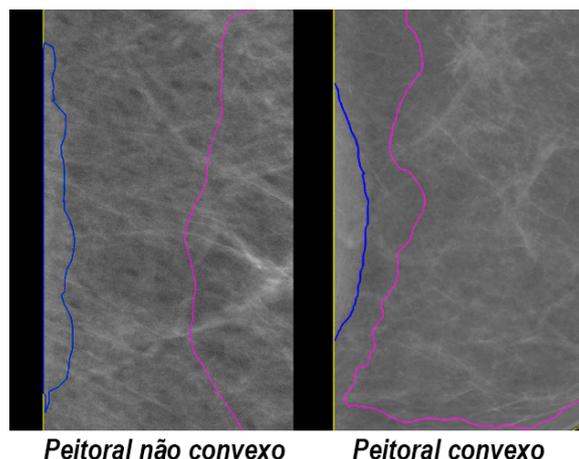


Figura 4.9: Exemplo de peitoral não convexo e peitoral convexo

4.5

Correção das anotações

Para a seleção das imagens para revisão foram utilizados thresholds para os algoritmos de estruturas pequenas, distância papila mama e convexhull, uma vez que os algoritmos fazem o correspondente cálculo para todas as anotações. A partir de uma análise visual sobre um subconjunto do dataset foram determinados os thresholds, que são os valores dentro dos quais as anotações são consideradas como casos de revisão, pois podem conter algum tipo de erro. Para as estruturas pequenas foi utilizado o valor de 10^4 pixels, sendo igual para papila e músculo peitoral, assim, se uma anotação está formada por uma quantidade de pixels menor que o threshold, é marcada para revisão. Para o algoritmo de distância papila mama foi utilizado um threshold de 3 pixels. Se a distância calculada for maior que este valor, a anotação da papila é marcada para revisão na ferramenta de anotação mencionada na seção 3.1 (Figura 4.10). Para o algoritmo de convexhull foi definido um threshold de 0.22. Se as anotações de músculo peitoral têm um nível de convexidade menor que o valor definido, são marcadas para revisão. No total, foram marcadas 3794 anotações para revisão.

Foram modificadas 2290 imagens com um total de 2950 anotações, correspondentes a 77 % de todas as anotações revisadas. Com isto, observa-se que os algoritmos de busca e sugestão de erros tiveram uma aceitável taxa de acerto sem incluir grande quantidade de falsos positivos, como pode se observar nos gráficos da Figura 4.11

Para uma melhor compreensão das alterações realizadas, para cada anotação modificada foi feita uma comparação com sua respectiva anotação antiga para obter uma porcentagem de mudança em relação aos pixels. A figura 4.12 mostra um exemplo de como a comparação é feita. Caso a anotação

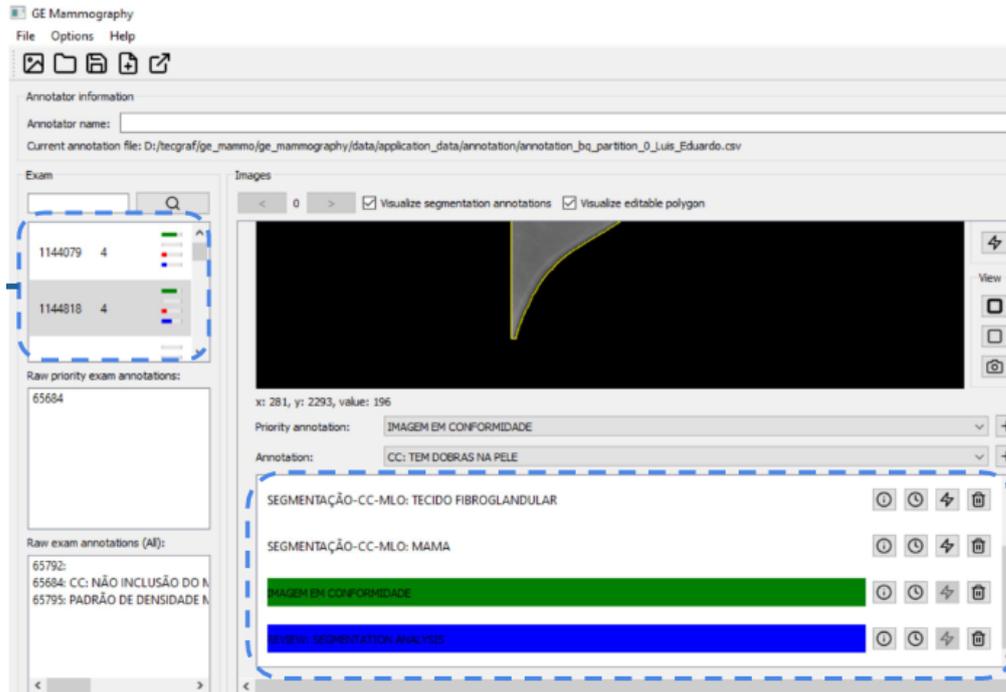


Figura 4.10: Marca gerada nas anotações inconsistentes.

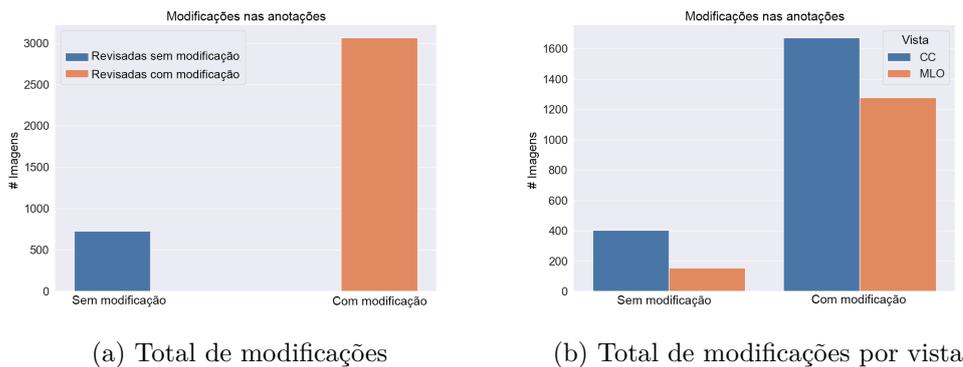


Figura 4.11: Modificações feitas no dataset

modificada tenha um tamanho menor que a anotação original, é retornado um percentual negativo, nesse caso é tomado seu valor absoluto para reconhecer quais anotações tiveram um percentual de modificação superior a 25 % ou superior a 50 %.

4.5.1

Correções na vista CC

A tabela 4.1 apresenta o número de anotações com modificação nos conjuntos de treinamento, validação e teste para a vista CC. Pode-se observar que existem casos nos quais o músculo peitoral teve modificações de pixels maiores que 50 %. No caso da vista CC, algumas das anotações

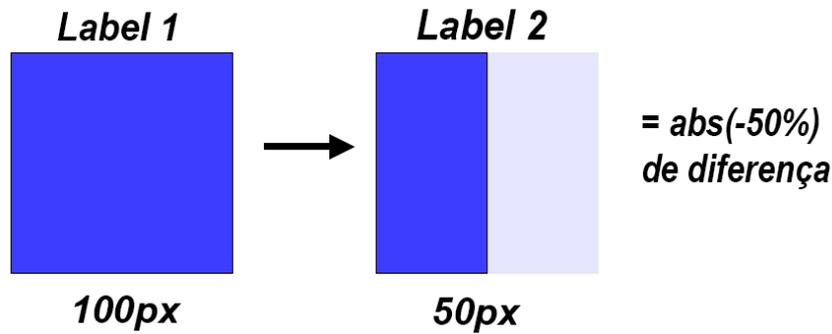


Figura 4.12: Exemplo de cálculo de diferença de pixels

do músculo peitoral apresentam modificações grandes, que posteriormente poderão influenciar positivamente nas métricas do modelo. Com um total de 5137 imagens, foram feitas 1036 modificações na papila, o que equivale a 20 % do total de imagens, e 642 modificações no músculo peitoral, o equivalente a 12,5 % do total de imagens.

Tabela 4.1: Anotações modificadas em vista CC

Set	Imagens	Mudanças	Mudanças na papila			Mudanças no peitoral		
			>0 %	>25 %	>50 %	Total	>25 %	>50 %
Training	3737	1161	793	252	83	476	125	11
Validation	943	248	163	37	10	105	31	4
Test	457	131	80	25	9	61	13	2
Total	5137	1540	1036	314	102	642	169	17

Observar que a maioria das modificações na papila e no músculo peitoral representa diferenças menores que 25 % em relação às anotações originais sugere a possibilidade de ajustes desnecessários. Esta tendência pode indicar uma pressão percebida pelos responsáveis pelas correções para modificar as anotações, mesmo quando a versão original provavelmente era precisa. Porém, as alterações acima de 25 %, especialmente aquelas acima de 50 %, poderiam influenciar significativamente na eficácia do modelo. Apesar da quantidade total de imagens com modificações importantes ser limitada, é crucial compreender como essas alterações podem afetar positiva ou negativamente a precisão e confiabilidade do modelo final.

4.5.2

Correções na vista MLO

Da mesma forma, na tabela 4.2 é mostrada a quantidade de anotações modificadas em vista MLO. Nesta vista, o músculo pectoral é, geralmente, uma estrutura bem definida e de tamanho considerável, não exigindo correções. Em contrapartida, foram observados casos nos quais a papila sofreu alterações na

sua posição, sem necessariamente afetar a quantidade de pixels. Apesar de haver uma menor quantidade de modificações na papila em comparação com a vista CC, essas mudanças podem ter um impacto mais significativo, uma vez que alteraram a localização da papila. Dessa forma, mesmo que as alterações na papila sejam menos frequentes, seu impacto pode ser substancial devido às modificações na sua posição, afetando potencialmente a interpretação dos dados.

Tabela 4.2: Anotações modificadas em vista MLO

Set	Imagens	Mudanças	Mudanças na papila			Mudanças no peitoral		
			>0 %	>25 %	>50 %	Total	>25 %	>50 %
Training	3450	888	888	293	77	10	0	0
Validation	1206	256	255	70	22	3	0	0
Test	557	112	111	25	10	5	0	0
Total	5213	1256	1254	388	109	18	0	0

4.6

Redução de amostragem

Depois de treinar um modelo com todas as mudanças nas anotações feitas na Etapa 1, foi feita uma análise numérica para avaliar se o modelo está classificando corretamente a presença e ausência de músculo peitoral na vista CC. Esta análise foi feita com base em métricas obtidas no conjunto de treinamento, nas quais foi possível observar que existiam vários casos de falsos positivos e falsos negativos. No entanto, durante uma avaliação visual dos casos em que ocorreram detecções incorretas do músculo peitoral no conjunto de treinamento, foi observado que nem todos os casos considerados falsos positivos eram realmente erros de segmentação do modelo.

Este tipo de caso pode causar ruído no processo de treinamento, por causa disso, a análise numérica foi complementada com uma avaliação visual dos casos em que ocorreu uma detecção incorreta do músculo peitoral no set de treinamento. O objetivo de revisar o set de treinamento é garantir que os dados dos que o modelo está aprendendo sejam de alta qualidade, consistentes e confiáveis. Além disso, este tipo de casos não puderam ser analisado na etapa 1 deste trabalho, pois este tipo de informação não confiável não está relacionada com o formato das anotações, mas sim com a própria imagem. Assim, durante a avaliação visual dos casos em que ocorreram detecções incorretas do músculo peitoral, foi observado que nem todos os casos considerados falsos positivos foram erros de segmentação do modelo. Como pode se observar na Figura 4.13, fica evidente que algumas anotações do músculo foram esquecidas. Estas anotações correspondem principalmente ao pequeno músculo peitoral, indicado

com uma seta vermelha na imagem original. Nota-se que a estrutura não pode ser facilmente identificada, o que pode ter ocasionado a não realização da anotação.

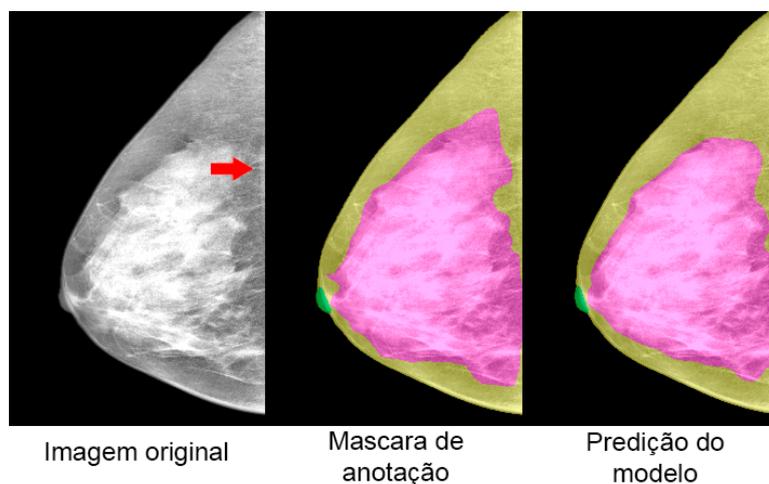


Figura 4.13: Caso de anotação esquecida em set de Treinamento

Como realizar um novo processo de anotação não é uma opção viável neste ponto do projeto, as imagens com anotação esquecida de músculo peitoral foram consideradas como possível fonte de informação errada e não foram consideradas para um próximo treinamento.

Alem disso, a análise visual do modelo da Etapa 1 permitiu ver também vários casos com um problema em comum em imagens complexas, as quais podem causar dificuldade de visualização devido ao pré-processamento da imagem. O pré-processamento aplicado pode gerar um alto contraste na zona do músculo peitoral. Nestas imagens se apresentam casos com anotação de músculo peitoral e outros casos sem anotação, o que pode ser considerado informação ambígua, pois imagens muito semelhantes são apresentadas com duas possíveis respostas. Por este motivo, várias imagens apresentam informação confusa, como os exemplos mostrados nas Figuras 4.14 e 4.15, as quais possuem uma área de alto contraste

As imagens do centro correspondem à máscara de anotação e as da direita à predição do modelo. Na Figura 4.14 a anotação indica a presença de músculo peitoral, e na Figura 4.15 a anotação indica ausência de músculo. Ambos os casos apresentam grande dificuldade de visualização do músculo, portanto não é possível indicar se há presença ou ausência desta estrutura, assim, este tipo de exame não fornece o tipo de informação esperada para treinar o modelo, por isso optou-se por eliminar as imagens com alto contraste na região do músculo peitoral, nas quais não foi possível determinar se a anotação estava correta ou errada. Seguindo este critério, no total foram removidas 469 imagens do conjunto de treinamento da Etapa 1, obtendo assim um novo set

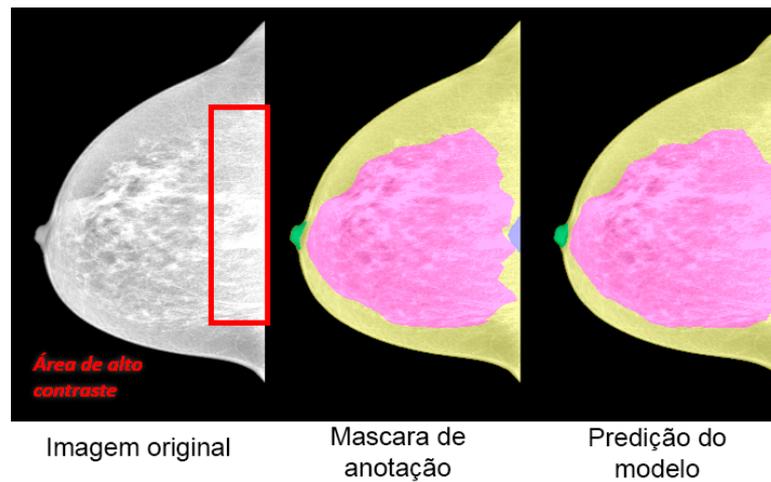


Figura 4.14: Imagem com alto contraste com anotação de músculo

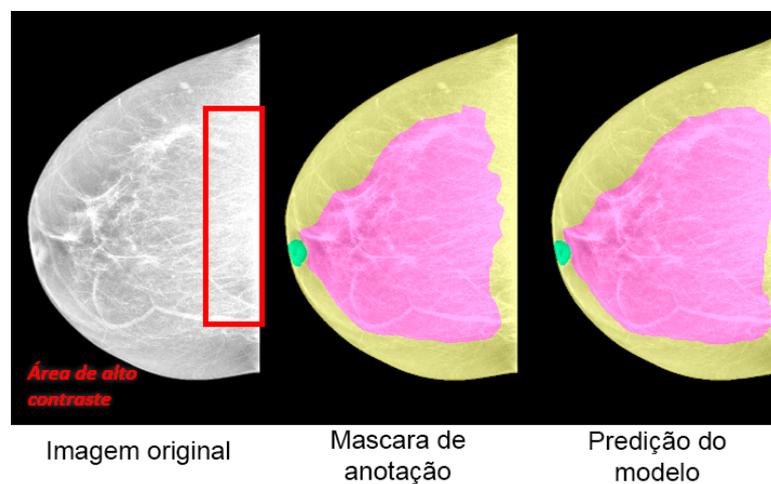


Figura 4.15: Imagem com alto contraste sem anotação de músculo

de treinamento para a vista CC com 3268 imagens em lugar das 3737 iniciais, mantendo as mesmas informações para os conjuntos de validação e teste.

5

Experimentos e resultados

Após todas as mudanças nos conjuntos de dados foi validado o desempenho dos modelos de segmentação treinados em variações do conjunto de dados de mamografia. Esta seção descreve a formação dos diferentes conjuntos de dados usados para treinamento e as métricas usadas para avaliar o desempenho de cada modelo.

5.1

Métricas de avaliação

Para avaliar a segmentação do músculo peitoral e das outras estruturas, foi utilizada a métrica de IoU, uma métrica amplamente utilizada em tarefas de segmentação. O IoU mede a sobreposição entre as estruturas previstas pelo modelo e as anotações de referência. Além do IoU, também foram calculadas métricas de classificação para avaliar o desempenho do modelo na detecção do músculo peitoral. Foram consideradas duas classes: presença e ausência do músculo peitoral. A classe de presença foi atribuída se pelo menos um pixel da máscara prevista pertencer à estrutura do músculo peitoral. Esta abordagem permitiu o cálculo de métricas como acurácia, precisão, recall, F1 e a matriz de confusão especificamente para a detecção do músculo peitoral na visão CC. Foram definidas 4 classes para o músculo peitoral de acordo com seu tamanho usando o K-Means, o algoritmo rotulou cada imagem com a presença do músculo peitoral em sua classe correspondente (NONE, SMALL, MEDIUM, LARGE). Os gráficos de barras da Figura 5.1 mostram a distribuição das imagens em cada um dos conjuntos de acordo com o tamanho do músculo peitoral. Pode-se notar que nos conjuntos de Validação e Test é apresentada a seguinte distribuição: LARGE (~ 8 %), MEDIUM (~ 12 %), NONE (~ 60 %) e SMALL (~ 20 %). E como mostra a figura 5.1a, para o conjunto de treinamento a distribuição apresenta maior percentual de imagens para o peitoral pequeno e médio. Assim, obtendo com as seguintes porcentagens: LARGE (~ 5 %), MEDIUM (~ 19 %), NONE (~ 38 %) e SMALL (~ 38 %). Esta distribuição foi feita para garantir um melhor aprendizado dos peitorais pequenos e médios.

5.2

Criação dos datasets

A partir dos dados base e dos dados obtidos nas etapas mencionadas na seção 4 foram estruturados três conjuntos de treinamento para a vista

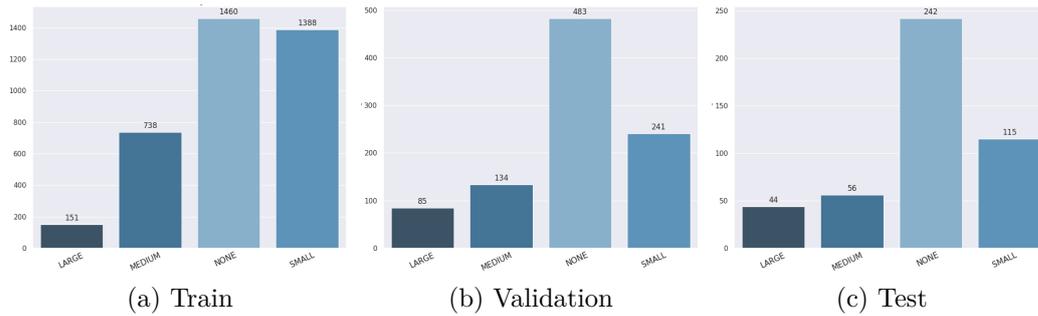


Figura 5.1: Distribuição das imagens segundo o tamanho do músculo peitoral

CC e dois para a vista MLO. Para as duas vistas, o primeiro conjunto de treinamento contém os dados originais, sem nenhum tipo de modificação. Para as duas vistas, o segundo conjunto contém as imagens com modificações nas etiquetas feitas na Etapa 1 mencionada na seção 4. Para a vista CC foi criado um terceiro set de treinamento com a quantidade de imagens reduzida, o que foi feito a partir do processo de redução de amostragem explicado na seção 4.6, correspondente à Etapa II.

Os cenários de treinamento mencionados foram avaliados usando os mesmos conjuntos de validação e teste, formado pelos dados corrigidos obtidos na Etapa I. Como o treinamento pode ser afetado pela inicialização dos pesos, existe a probabilidade de obter resultados diferentes usando os mesmos dados. Como o processo de treinamento pode ser influenciado pela inicialização dos pesos ou pela qualidade dos dados, cada conjunto foi treinado seis vezes e foi selecionado um modelo com métricas médias para cada conjunto. No caso da vista CC, os modelos foram selecionados usando as médias das métricas de classificação de músculo peitoral e para a vista MLO, as médias dos resultados de IoU.

5.3

Resultados do Treinamento

Na análise das curvas de aprendizado, de IoU e perda, foram observados padrões que apoiam a abordagem Data-centric e evidenciam a generalização dos modelos com dados corrigidos. Os gráficos das Figuras 5.2 e 5.3 mostram as curvas de IoU e perda dos conjuntos de treinamento e validação comparando apenas o modelo Baseline e o modelo da Etapa II para uma melhor visualização. As curvas foram denotadas da seguinte forma: Treinamento do Baseline com preto sólido, validação do Baseline com preto pontilhado, treinamento da Etapa II com verde sólido e validação de Etapa II com verde pontilhado.

Começando com a análise das curvas dos modelos da vista CC, ao

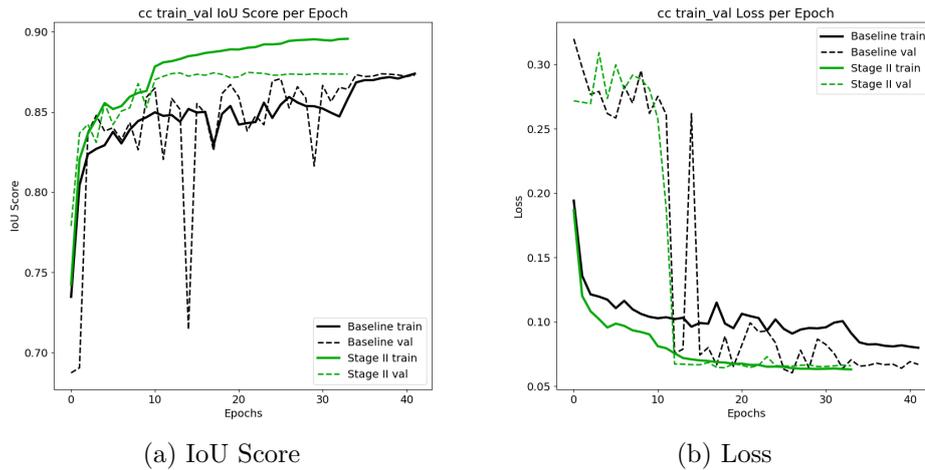


Figura 5.2: Curvas de aprendizado vista CC

comparar os valores de IoU entre o conjunto de treinamento e o conjunto de validação, notou-se que o IoU na validação na segunda etapa permaneceu acima da curva do IoU de treinamento no modelo de referência (Baseline). Isso sugere uma melhora na generalização do modelo da segunda etapa. Além disso, o IoU (treinamento e validação) do baseline apresentou flutuações mais marcantes ao longo das épocas, enquanto o IoU (treinamento e validação) do modelo da Etapa II mostrou uma rápida estabilização. Isso poderia indicar que o modelo da segunda etapa possui menor variabilidade nas previsões, o que pode ocorrer devido ao fato de estar sendo alimentado com dados mais representativos e coerentes.

A perda de treinamento no Baseline apresenta altas flutuações em comparação com a perda da Etapa II, sendo essas flutuações ainda mais evidentes na perda de validação. Isso significa que o modelo Baseline não está generalizando os dados com facilidade. Isso pode ser atribuído a todos os erros contidos nas anotações do conjunto de treinamento do modelo base. Ao contrário do baseline, o modelo da Etapa II está generalizando os dados de forma mais estável e rápida.

Os dados aprimorados no modelo da segunda etapa podem ter tido um impacto positivo na inicialização dos pesos do modelo durante as primeiras épocas do treinamento. Isso pode ter acelerado o processo de generalização do modelo, resultando em uma melhoria na capacidade do modelo de extrair características e compreender os padrões dos dados. Essa melhoria é refletida em curvas mais estáveis de IoU e perda ao longo das épocas no modelo da segunda etapa.

Para a vista MLO, a análise das curvas de aprendizagem mostra um desempenho ligeiramente melhor no modelo da Etapa I, porém o processo

de treinamento dos dois modelos apresenta comportamento semelhante, como pode se observar nos gráficos do IoU e Loss da Figura 5.3.

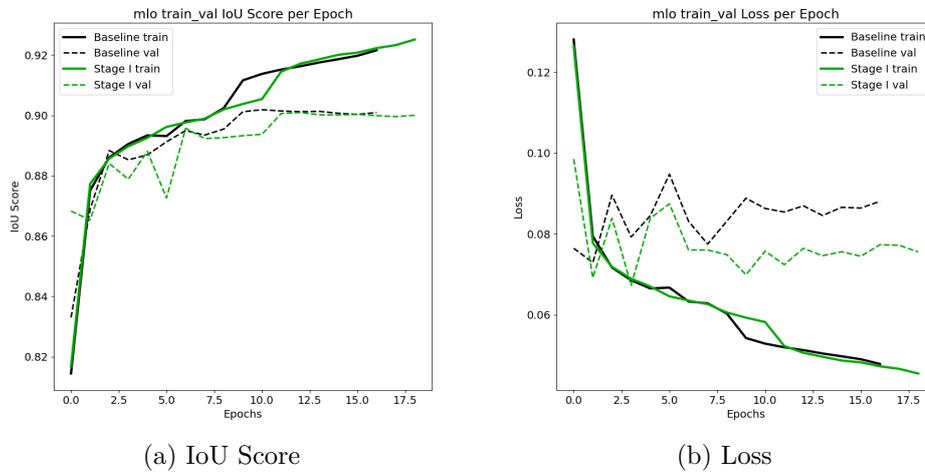


Figura 5.3: Curvas de aprendizado vista MLO

Para esta vista não houve modificações no músculo peitoral nem uma Etapa II focado nesta estrutura. Além disso, o aumento do IoU da papila é pequeno e não representa uma mudança significativa na curva do IoU do modelo da Etapa I em comparação ao modelo base.

5.4

Resultados no Conjunto de Validação

A tabela 5.1 mostra os resultados de IoU obtidos nos três modelos da vista CC enfatizando os diferentes tamanhos do músculo peitoral, pois é a estrutura objetivo a melhorar.

Tabela 5.1: IoU vista CC no set de Validação

Etapa	Mean	Papila	Tecido	Mama	Peitoral			
					Global	Grande	Médio	Pequeno
Baseline	0,85	0,75	0,90	0,81	0,79	0,9	0,85	0,66
Etapa I	0,86	0,76	0,90	0,80	0,82	0,89	0,83	0,60
Etapa II	0,87	0,76	0,92	0,85	0,84	0,90	0,87	0,70

Como parte do objetivo, se quer reduzir o número de casos nos que o modelo segmenta incorretamente um músculo peitoral, pois o modelo base apresenta uma alta taxa de Falsos Positivos (FPR). O modelo da Etapa II apresentou maior equilíbrio entre taxa de Falsos Positivos e Falsos Negativos, tendo as métricas mais altas devido à redução da amostragem no conjunto de treinamento. A Tabela 5.2 apresenta os resultados relacionados à classificação das imagens, como acurácia, precisão, recall e F1, além das taxas de verdadeiros

e falsos. Na tabela mencionada pode se observar que com uma quantidade de imagens menor o modelo da Etapa II manteve o valor do FNR em 0,04 e conseguiu diminuir o FPR de 0,18 para 0,10.

Tabela 5.2: Classificação do músculo peitoral no set de Validação

Etapa	Acc	Prec	Rec	F1	TPR	TNR	FPR	FNR	n
Baseline	0,89	0,84	0,96	0,90	0,96	0,82	0,18	0,04	3737
Etapa I	0,91	0,90	0,92	0,91	0,92	0,93	0,07	0,10	3737
Etapa II	0,93	0,91	0,96	0,93	0,96	0,90	0,10	0,04	3268

Para complementar as métricas de classificação, as matrizes de confusão da figura 5.4 fornecem uma melhor visão das classificações corretas e incorretas feitas pelos modelos em relação aos verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. A matriz de confusão para o modelo base tem a quantidade de falsos positivos consideravelmente alta em comparação aos falsos negativos, com um FPR de 0,18. Este valor indica que o modelo apresenta tendência a identificar erroneamente casos em que o músculo peitoral não está presente. O modelo da Etapa I teve uma melhoria na taxa de falsos positivos, tendo o valor de FPR mais baixo dos três modelos, porém com a taxa de falsos negativos mais alta. O modelo da Etapa II conseguiu manter a mesma taxa de falsos negativos do modelo base além de simultaneamente reduzir a taxa de falsos positivos, por este motivo considera-se que o modelo da Etapa II possui uma capacidade aprimorada de identificar corretamente a presença do músculo peitoral.

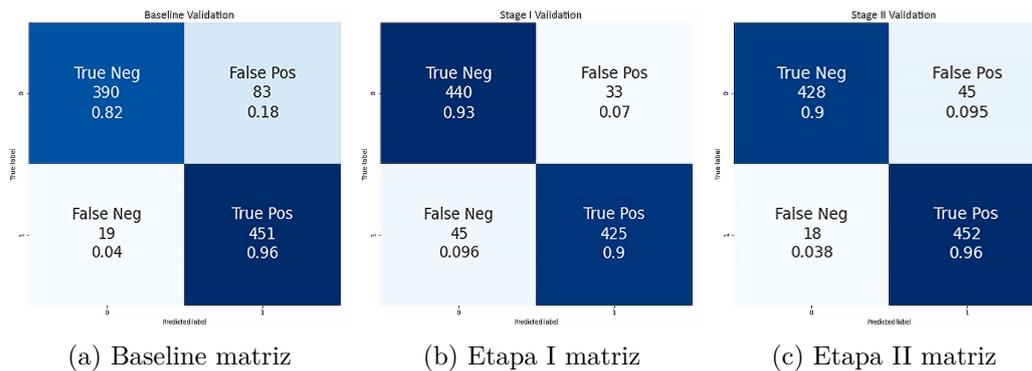


Figura 5.4: Matrizes de confusão vista CC no set de Validação

Para a vista MLO, as mudanças feitas no dataset tiveram um maior impacto no delineado da papila, mantendo o valor do IoU para as outras estruturas. Nesta vista o algoritmo do cálculo da distância entre papila e mama encontrou varias anotações de papila não perfilada na localização incorreta dentro da imagem. Portanto, elas não sofreram grande alteração em seu número

de pixels, mas sim em sua localização na imagem, o que pode gerar um impacto maior no treinamento. A tabela 5.3 mostra um aumento no IoU da papila, com um valor no modelo base de 0,71 e chegando até 0,75 na Etapa I, o que pode se considerar como um aumento significativo.

Tabela 5.3: IoU na vista MLO no set de Validação

Etapa	Mean	Papila	Tecido	Mama	Peitoral
Baseline	0,85	0,71	0,88	0,78	0,94
Etapa I	0,88	0,75	0,88	0,80	0,95

A figura 5.5 mostra de forma visual os resultados no set de validação. Os resultados são exibidos em gráficos de radar e com linhas coloridas. Para as duas vistas (CC e MLO), a linha azul representa o modelo base, a linha verde a Etapa I e a linha vermelha a Etapa II. Ressaltando que apenas os modelos da vista CC apresentam a linha vermelha, pois não existe nenhum experimento na Etapa II para a vista MLO.

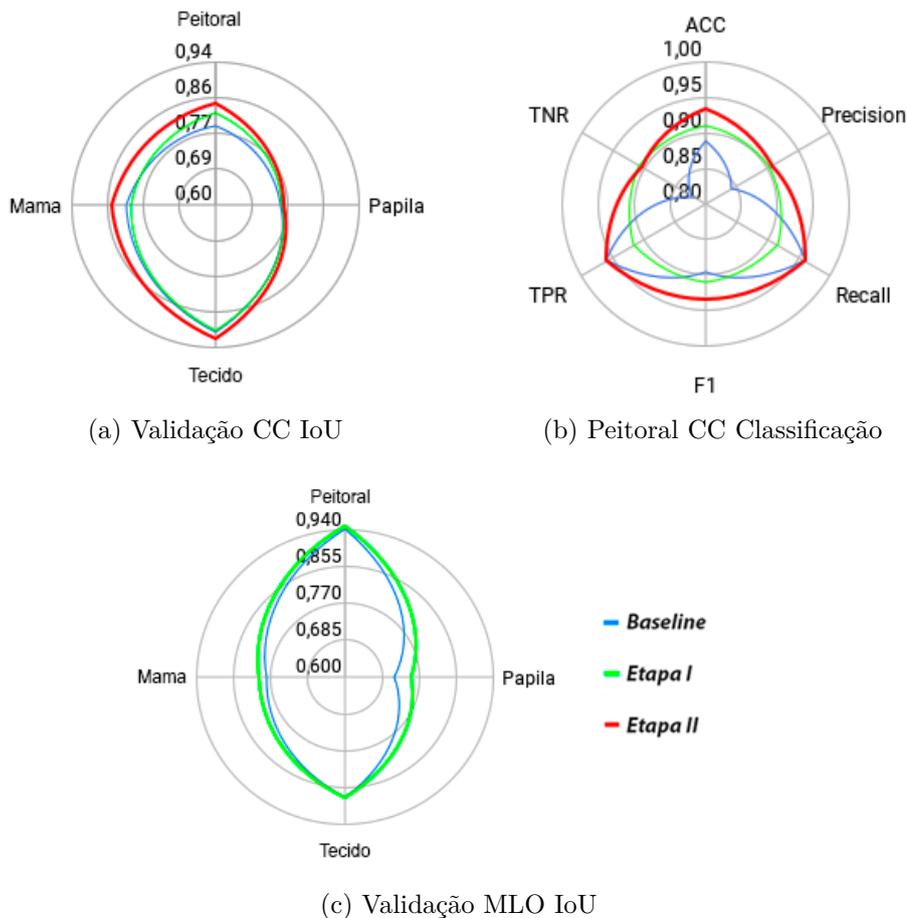


Figura 5.5: Resultados gráficos no set de Validação

5.5

Resultados no Conjunto de Teste

Inicialmente, foram feitas as modificações com o objetivo de avaliar seu impacto no conjunto de validação. No entanto, para demonstrar que o impacto positivo não se limita apenas a este conjunto, também foram calculadas as métricas no conjunto de teste. Os resultados obtidos neste conjunto reafirmam as melhorias observadas no conjunto de validação, o que indica que o modelo está mais generalizado e é capaz de segmentar imagens de maneira eficaz em diferentes conjuntos de dados.

Na tabela 5.4 pode se observar um aumento no IoU global do músculo peitoral e também em cada uma das classes atribuídas (SMALL, MEDIUM, LARGE). Estes resultados comprovam que a abordagem tem uma boa efetividade no aprimoramento do desempenho do modelo da vista CC, especialmente no cenário da Etapa II. Pode-se observar melhorias na delimitação precisa em todos os tamanhos de músculo peitoral, abordando efetivamente casos que demandam maior precisão e detalhes mais finos.

Tabela 5.4: IoU vista CC no conjunto de teste

Etapa	Mean	Papila	Tecido	Mama	Peitoral			
					Global	Grande	Médio	Pequeno
Baseline	0,85	0,77	0,90	0,81	0,79	0,84	0,83	0,68
Etapa I	0,86	0,76	0,92	0,84	0,80	0,83	0,82	0,70
Etapa II	0,87	0,76	0,92	0,85	0,84	0,86	0,84	0,73

Os resultados sobre este novo conjunto de dados mostra uma queda importante na taxa de falsos positivos na classificação do músculo peitoral, começando em 0,22 no modelo base até 0,12 no modelo da Etapa II, indicando que a tendência do modelo em detectar falsos músculos peitorais diminuiu. Portanto, houve melhorias na capacidade do modelo em detectar e delimitar o músculo peitoral em diferentes tamanhos, como evidenciado pelo aumento das métricas de classificação, do IoU global e pelos IoUs mais altos em cada classe específica.

Tabela 5.5: Classificação do músculo peitoral no conjunto de teste

Etapa	Acc	Prec	Rec	F1	TPR	TNR	FPR	FNR
Baseline	0,86	0,80	0,96	0,87	0,96	0,78	0,22	0,04
Etapa I	0,90	0,83	0,98	0,90	0,98	0,83	0,17	0,02
Etapa II	0,93	0,88	0,97	0,93	0,97	0,88	0,12	0,03

Para complementar as métricas de classificação do conjunto de teste, as matrizes de confusão da figura 5.6 mostram as quantidades de

verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Diferentemente do conjunto de validação, o número de falsos negativos nos 3 modelos permaneceu baixo e estável, porém pode ser notada uma diminuição progressiva no número de falsos positivos, reduzindo a porcentagem de FP em 50 %, começando com 53 casos no modelo base, passando por 42 casos no modelo da Etapa I e conseguindo 28 casos de FP no modelo da Etapa II.

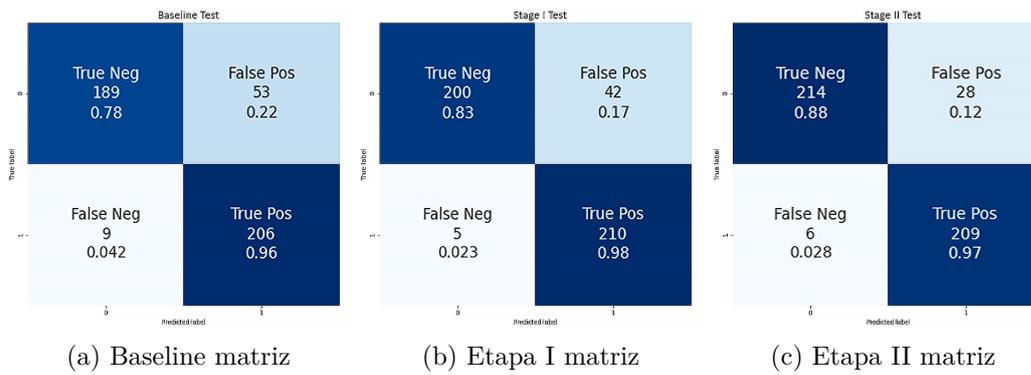


Figura 5.6: Matrizes de confusão vista CC no conjunto de teste

Para a vista MLO, a análise dos IoUs no conjunto de teste mostra uma ligeira melhoria no IoU da papila de 0,73 para 0,75, mantendo IoUs semelhantes nas outras estruturas. A tabela 5.6 indica que para a vista MLO, o modelo da Etapa I apresenta uma vantagem leve na precisão do delineamento da papila.

Tabela 5.6: IoU na vista MLO no conjunto de teste

Etapa	Mean	Papila	Tecido	Mama	Peitoral
Baseline	0,88	0,73	0,91	0,82	0,96
Etapa I	0,89	0,75	0,90	0,81	0,96

A figura 5.5 mostra de forma visual os resultados no conjunto de teste. Como foi mostrado nos gráficos do conjunto de validação, para o conjunto de teste os resultados são exibidos em gráficos de radar, usando as mesmas cores para as linhas. Para as duas vistas (CC e MLO), a linha azul representa o modelo base, a linha verde a Etapa I e a linha vermelha a Etapa II. Ressaltando que apenas os modelos da vista CC contém a linha vermelha, pois não existe nenhum experimento na Etapa II para a vista MLO.

5.6

Comparação de resultados Model-centric e Data-centric

Foi realizado um estudo model-centric através da implementação de vários modelos que compartilham o mesmo conjunto de dados. Esses modelos se distinguem uns dos outros por suas diversas estruturas de modelo, backbones,

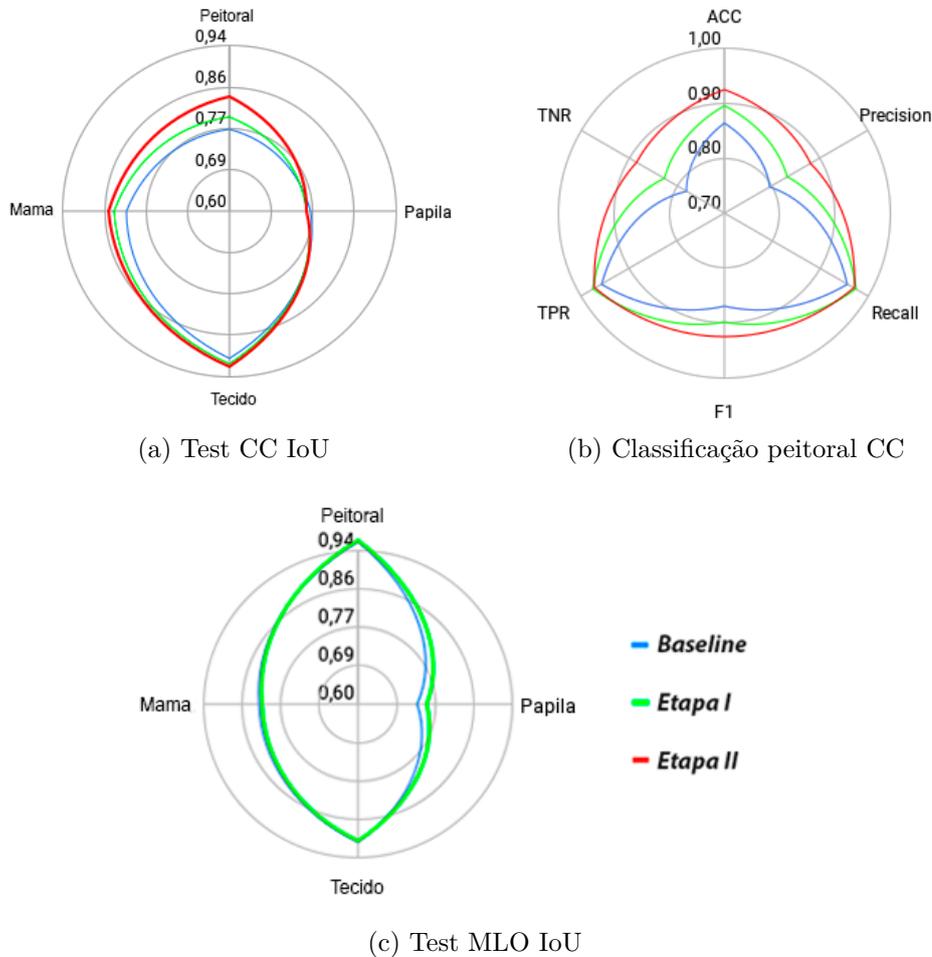


Figura 5.7: Resultados gráficos no conjunto de teste

otimizadores e taxas de aprendizado. A tabela 5.7 mostra os itens utilizados para criar as combinações.

Model	Backbone	Optimizer	Learning rate
Unet	efficientnetb3	Adam	0.001
FPN	resnet50	SGD	0.01
PSPNet	efficientnetb5	rmsprop	0.1

Tabela 5.7: Ítems usados para combinações

O objetivo por trás desta abordagem foi identificar a combinação ideal desses componentes para alcançar o máximo desempenho na segmentação semântica de mamografias. Foi implementada uma estratégia grid-search para realização dos experimentos, gerando no total 81 combinações dos hiperparâmetros. Cada variante do modelo possui uma configuração única que pode influenciar na sua capacidade de extrair características das imagens. Com os treinamentos realizados e adicionando os resultados do modelo das etapas I e II, foi feita uma comparação através de um Parallel Plot, mostrando

como cada ajuste afetou o desempenho da segmentação em termos de IoU de papila e músculo peitoral, além da quantidade de Verdadeiros Positivos e Verdadeiros Negativos na classificação do músculo peitoral. A Figura 5.8 mostra os resultados de interesse para este trabalho dos modelos feitos com todas as possíveis combinações. A figura mostra nas quatro primeiras colunas os itens utilizados para fazer as combinações de estrutura do modelo (model, backbone, optimizer, learning rate), e no restante das colunas os resultados de IoU do músculo e papila além de algumas métricas de classificação do músculo peitoral, como F1, TNR e PRF.

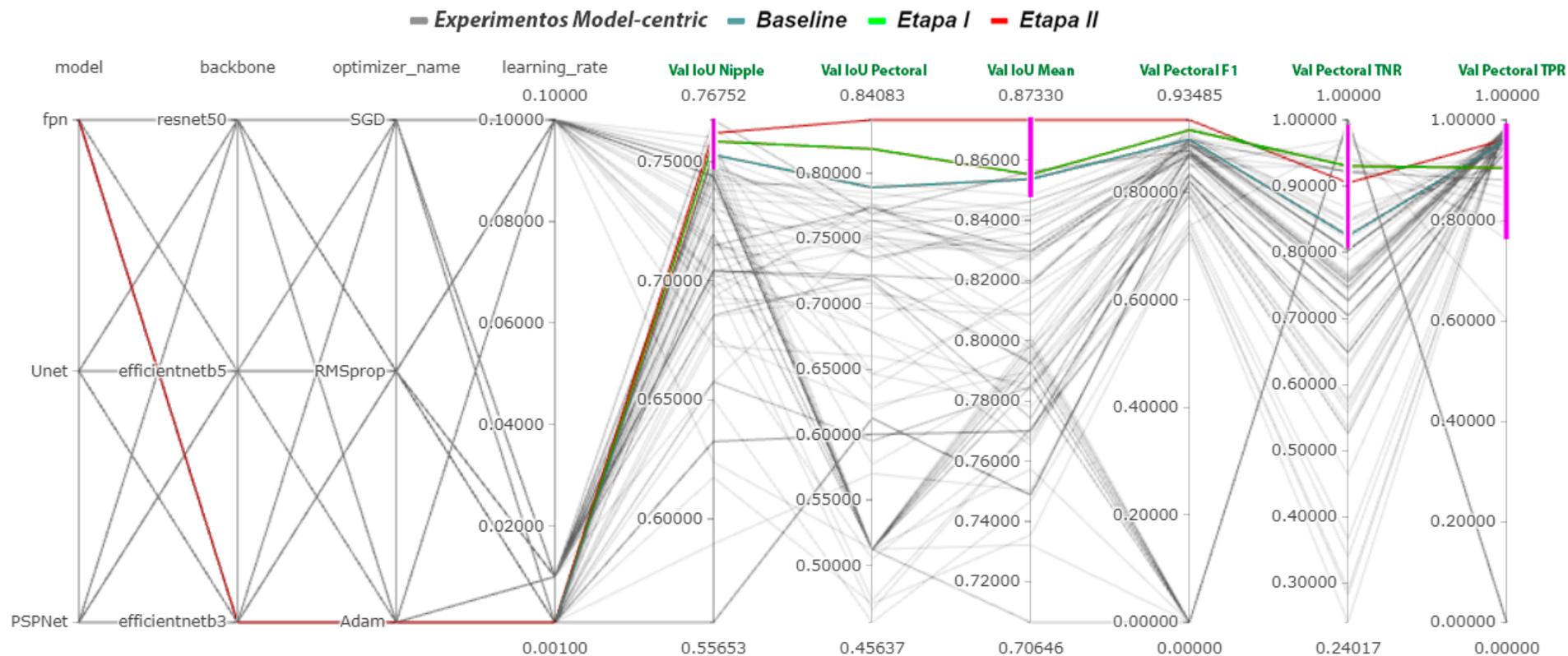


Figura 5.8: O gráfico mostra os resultados da IoU da papila e do músculo peitoral, além das métricas de classificação do músculo. As linhas cinza mostram todos os resultados obtidos na abordagem centrada no modelo, a linha azul (Baseline) mostra o melhor resultado desta abordagem, que foi a utilizada para realizar os experimentos focados nos dados. As linhas verdes e vermelhas mostram os resultados da abordagem centrada em dados e demonstram como o foco nos dados pode melhorar as métricas da melhor abordagem centrada em modelo.

Para mostrar os modelos com os maiores resultados, foram utilizados thresholds representados como linhas roxas. Assim, no gráfico resultante, todas as linhas cinzas representam os resultados dos modelos da abordagem model-centric. Depois de comparar os resultados de todos os modelos avaliados sobre os mesmos dados, a configuração destacada foi aquela utilizada para os treinamentos feitos neste trabalho, a qual está formada por um modelo FPN suportado pela arquitetura eficiente netb3. Este modelo foi otimizado usando o algoritmo Adam com taxa de aprendizado de 0,001. Os resultados dos modelos que utilizam esta estrutura estão marcados na Figura 5.8 e correspondem ao modelo base, ao modelo da Etapa I e ao modelo da Etapa II, em azul, verde e vermelho respectivamente. As linhas cinzas correspondem aos resultados dos modelos feitos com as demais combinações. Se pode observar que em la abordage model-centric los resultados. Percebe-se que os resultados máximos obtidos na abordagem model-centric (linhas cinza e azuis) foram melhorados pela abordagem centrada em dados (linhas verde e vermelha) . Estes últimos obtiveram maiores IoUs nas estruturas de interesse e um alto F1 na detecção do músculo peitoral, o que significa que apresentam uma taxa menor de falsos positivos e falsos negativos. Estes resultados ressaltam a eficácia da abordagem data-centric, na qual aprimorar a qualidade dos dados de treinamento se mostrou fundamental para obter melhorias no desempenho do modelo.

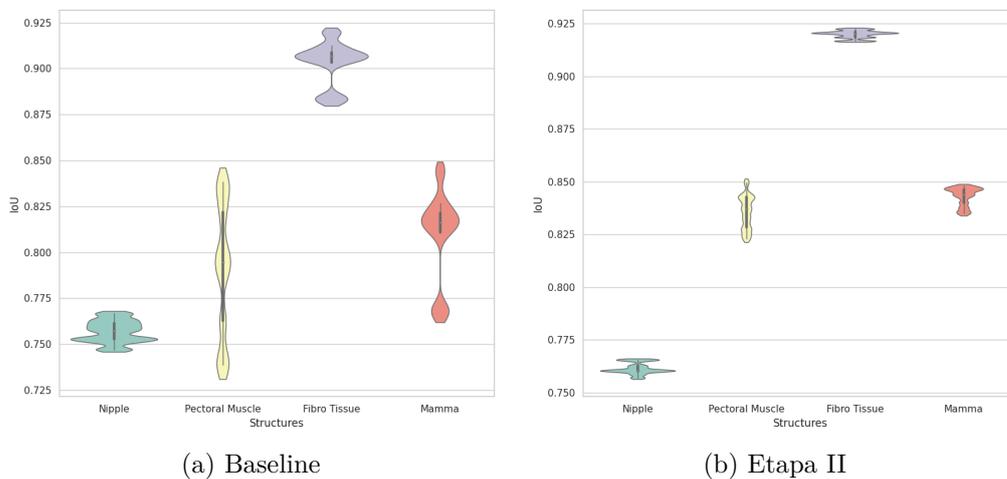


Figura 5.9: Violin plots dos resultados de validação

Adicionalmente, para comparar o comportamento do modelo base e do modelo da Etapa II para a vista CC, foram plotados os IoUs das 4 classes principais em violin plots, os quais representam os diferentes valores de IoU em função da quantidade de treinamentos realizados no eixo X, com o IoU no eixo Y. As áreas mais largas dos violinos no eixo X indicam uma maior densidade de experimentos para uma estrutura específica. Os gráficos da Figura 5.9 mostram

faixas mais amplas de valores no modelo base (Figura 5.9a), enquanto o modelo 2 apresenta faixas de valores mais estreitas (Figura 5.9b). Na papila, o modelo base tende a obter um IoU entre 0,747 e 0,765 enquanto na Etapa II são obtidos valores entre 0,754 e 0,766, mostrando que para a vista CC, ao contrário da vista MLO, a papila não obteve uma melhora considerável. Porém, para o músculo peitoral a melhora é evidente, o modelo base obteve valores entre 0,739 e 0,839, obtendo geralmente um valor de aproximadamente 0,78 na maioria dos casos, enquanto o modelo da Etapa II obteve valores de IoU entre 0,823 e 0,850, obtendo na maioria dos casos um valor de aproximadamente 0,84.

No próximo capítulo, todos os resultados obtidos serão levados em consideração para concluir pontos importantes, além de mencionar algumas limitações encontradas neste trabalho e possíveis trabalhos futuros.

6

Conclusões e trabalhos futuros

O objetivo deste trabalho consistiu em aprimorar as métricas de um modelo de segmentação semântica de mamografias por meio de uma abordagem data-centric. O foco principal foi o delineamento da papila nas vistas CC e MLO e a detecção do músculo peitoral na vista CC, porque esta estrutura frequentemente apresenta desafios significativos em termos de identificação.

O estudo realizado neste trabalho prova que a abordagem data-centric é uma estratégia eficiente para aumentar as métricas de um modelo de segmentação semântica. O modelo de segmentação treinado em conjuntos de dados com modificações apropriadas e eliminação de informações ambíguas conseguiu fornecer uma melhoria na detecção e na segmentação do músculo peitoral na vista CC, bem como uma melhoria na segmentação da papila na vista MLO. O aumento nas métricas de avaliação, como o IoU e as métricas de classificação, destacam a capacidade do modelo para reconhecer e delinear com mais precisão as estruturas de interesse, tornando-o mais confiável em diferentes cenários.

6.1

Principais contribuições

A principal contribuição deste trabalho é a geração de um processo semiautomático para facilitar a detecção de erros para a melhoria de anotações em um conjunto de dados mamográficos, que incluiu a criação de algoritmos que visam analisar e compreender as formas das anotações feitas para as estruturas anatômicas mamárias, especificamente para as anotações de papila e músculo peitoral. Estes algoritmos permitiram encontrar anotações inconsistentes com precisão e conseqüentemente aumentar as métricas do modelo de segmentação semântica.

Como resultado de este trabalho será publicado um artigo intitulado: *'Data-centric Approach for Pectoral Muscle Deep Learning Segmentation Enhancements in Mammography Images'* no 18º Simpósio Internacional de Computação Visual (ISVC'23).

6.2

Limitações

Esta pesquisa mostrou a implementação um pipeline centrado em dados para a segmentação de estruturas anatômicas da mama, obtendo bons

resultados. No entanto, é importante reconhecer que ainda existem limitações nesta abordagem. Uma das limitações mais notáveis é a dependência de correção manual das anotações sinalizadas pelos algoritmos de busca de erros. O objetivo dos algoritmos criados é apenas a busca de anotações com algum tipo de inconsistência, reduzindo significativamente o tempo que este processo levaria em uma abordagem manual, porém a intervenção humana ainda é necessária para validar e ajustar as correções propostas pelos algoritmos.

6.3

Trabalhos futuros

Embora os resultados obtidos nesta pesquisa demonstrem um aprimoramento na segmentação e detecção do músculo peitoral, uma nova abordagem para investigações futuras é a implementação de uma seleção automática dos thresholds utilizados nos algoritmos de busca de erros ou de um processo de correção automática de anotações através de técnicas de processamento de imagem. A correção automática não substituiria a segmentação, mas poderia melhorar a qualidade das anotações existentes, reduzindo as imprecisões introduzidas por anotadores humanos diferentes. Assim, este tipo de anotação corrigida poderia gerar um impacto significativo nas métricas de avaliação.

7

Referências bibliográficas

- [1] CHENAOUA, K.; KURUGOLLU, F. ; BOURIDANE, A.. **Data cleaning and outlier removal: Application in human skin detection**. In: 2014 5TH EUROPEAN WORKSHOP ON VISUAL INFORMATION PROCESSING (EUVIP), p. 1–6. IEEE, 2014. 2.1
- [2] PONZIO, F.; DEODATO, G.; MACII, E.; DI CATALDO, S. ; FICARRA, E.. **Exploiting “uncertain” deep networks for data cleaning in digital pathology**. In: 2020 IEEE 17TH INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI), p. 1139–1143. IEEE, 2020. 2.1
- [3] LIU, D.; MENG, Y. ; WANG, L.. **Data cleaning of irrelevant images based on transfer learning**. In: 2020 INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING, AUTOMATION AND SYSTEMS (ICICAS), p. 450–456. IEEE, 2020. 2.1
- [4] KAUR, M.; WASSON, V.. **Roi based medical image compression for telemedicine application**. *Procedia Computer Science*, 70:579–585, 2015. 2.1
- [5] MIRANDA, L. J.. **Towards data-centric machine learning: a short review**. *ljvmiranda921.github.io*, 2021. 1.2
- [6] ORELLANA RUEDA, P. I.. **Segmentación semántica y reconocimiento de lugares usando características cnn pre-entrenadas**. 2019. 1
- [7] MARTINS, R. A. P.; OTHERS. **Aplicação de redes convolucionais profundas para detecção de massas em mamografias**. 2019. 1.1
- [8] ZUIDERVELD, K.. **Contrast Limited Adaptive Histogram Equalization**, p. 474–485. Academic Press Professional, Inc., USA, 1994. 3.1
- [9] GUO, Y.; ZHAO, W.; LI, S.; ZHANG, Y. ; LU, Y.. **Automatic segmentation of the pectoral muscle based on boundary identification and shape prediction**. *Physics in Medicine & Biology*, 65(4):045016, 2020. 2.2
- [10] DUBROVINA, A.; KISILEV, P.; GINSBURG, B.; HASHOUL, S. ; KIMMEL, R.. **Computational mammography using deep neural networks**.

- Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 6(3):243–247, 2018. 2.2
- [11] RUBIO, Y.; MONTIEL, O.. **Multicriteria evaluation of deep neural networks for semantic segmentation of mammographies**. Axioms, 10(3):180, 2021. 2.2
- [12] GE, M.; MAINPRIZE, J. G.; MAWDSLEY, G. E. ; YAFFE, M. J.. **Segmenting pectoralis muscle on digital mammograms by a markov random field-maximum a posteriori model**. Journal of Medical Imaging, 1(3):034503–034503, 2014. 2.2
- [13] CAMILUS, K. S.; GOVINDAN, V. ; SATHIDEVI, P.. **Pectoral muscle identification in mammograms**. Journal of applied clinical medical physics, 12(3):215–230, 2011. 2.2
- [14] SIERRA-FRANCO, C. A.; HURTADO, J.; DE A. THOMAZ, V.; DA CRUZ, L. C.; SILVA, S. V. ; RAPOSO, A. B.. **Towards automated semantic segmentation in mammography images**, 2023. 2.2, 3.2
- [15] OTSU, N.. **A threshold selection method from gray-level histograms**. IEEE transactions on systems, man, and cybernetics, 9(1):62–66, 1979. 3.1
- [16] SCHMARJE, L.; GROSSMANN, V.; ZELENKA, C.; DIPPEL, S.; KIKO, R.; OSZUST, M.; PASTELL, M.; STRACKE, J.; VALROS, A.; VOLKMANN, N. ; OTHERS. **Is one annotation enough?-a data-centric image classification benchmark for noisy and ambiguous label estimation**. Advances in Neural Information Processing Systems, 35:33215–33232, 2022. 2.1