

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

**Aplicação de Visão Computacional na
Identificação de Tubos em Radiografias de Tórax**

Alexandre Rodrigues Bomfim Junior

**PROJETO FINAL DE GRADUAÇÃO
CENTRO TÉCNICO CIENTÍFICO - CTC
DEPARTAMENTO DE INFORMÁTICA
Curso de Graduação em Ciências da Computação**

Rio de Janeiro Julho 2024



Alexandre Rodrigues Bomfim Junior

**Aplicação de Visão Computacional na
Identificação de Tubos em Radiografias de
Tórax**

Projeto Final

Projeto Final apresentada ao Curso de Ciência da Computação da PUC-Rio como parte do cumprimento parcial dos requisitos para o grau de Bacharel em Ciência da Computação.

Orientador : Prof. Alberto Barbosa Raposo
Coorientador: Prof. César Augusto Sierra Franco

Rio de Janeiro
Junho de 2024

Agradecimentos

Primeiramente, agradeço a Deus pela força e sabedoria concedidas ao longo desta jornada acadêmica.

A minha querida mãe, Ivete Rosti Martins, por todo o amor, apoio e incentivos incondicionais. Você sempre acreditou no meu potencial e me inspirou a buscar o melhor de mim. Este trabalho é dedicado a você.

As minhas irmãs, Arielly Bomfim, Alessandra Bomfim e Amanda Bomfim, por estarem sempre ao meu lado, compartilhando momentos de alegria e superação. A força da nossa união me motiva a continuar em frente.

À minha namorada, Natália dos Santos, pelo carinho, paciência e compreensão durante todos os momentos desafiadores. Seu apoio constante foi fundamental para que eu pudesse alcançar este objetivo.

Ao meu querido amigo, João Pedro Maia, que tornou esta jornada mais leve e divertida. A amizade e os momentos compartilhados foram essenciais para manter o equilíbrio entre estudo e vida pessoal.

Ao meu orientador e co-orientador, Alberto Raposo e Cesar Sierra Franco, pela orientação, paciência e dedicação ao longo de todo o processo de desenvolvimento deste trabalho. Sua experiência e sabedoria foram indispensáveis para a realização deste projeto.

Agradeço também a todos os professores e colegas de curso que contribuíram, direta ou indiretamente, para minha formação acadêmica. Cada aprendizado e experiência compartilhada foram fundamentais para o meu crescimento pessoal e profissional.

Por fim, agradeço a todos que, de alguma forma, contribuíram para a concretização deste trabalho. Meu sincero agradecimento a cada um de vocês.

Resumo

Rodrigues Bomfim Junior, Alexandre; Barbosa Raposo, Alberto; Sierra Franco, César Augusto. **Aplicação de Visão Computacional na Identificação de Tubos em Radiografias de Tórax.** Rio de Janeiro, 2024. 47p. Projeto Final – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho investiga o uso de técnicas de *Deep Learning* para segmentação semântica de cateteres venosos centrais (CVC) em radiografias de tórax. O objetivo é auxiliar na identificação desses dispositivos para determinação de seu posicionamento, reduzindo complicações associadas a procedimentos invasivos. Inicialmente, abordamos desafios como o pequeno tamanho relativo dos dispositivos CVC nas imagens e o desbalanceamento dos dados. Para isso, utilizamos diferentes *backbones*, como *Resnets* e *EfficientNets*, além de realizar ajustes no redimensionamento das imagens. Para melhorar a robustez e a capacidade de generalização dos modelos, aplicamos técnicas de data augmentation. Implementamos um *ensemble* de modelos, combinando os resultados de várias arquiteturas individuais, o que se mostrou eficaz ao superar os modelos isolados em diversas métricas de desempenho. Um script adicional foi desenvolvido para identificar a presença do CVC nas predições do ensemble, analisando a contagem de pixels ativos e a detecção de contornos. Os resultados finais demonstraram que a abordagem de *ensemble* aprimora a precisão e a confiabilidade na detecção de CVCs. Futuras pesquisas devem focar na exploração da classificação do posicionamento do dispositivo, como uma etapa subsequente, visando melhorar mais a aplicabilidade clínica dessas técnicas.

Palavras-chave

Deep Learning; Segmentação Semântica; Radiografia de Tórax; Radiografia.

Abstract

Rodrigues Bomfim Junior, Alexandre; Barbosa Raposo, Alberto (Advisor); Sierra Franco, César Augusto (Co-Advisor). **Computer Vision Application in Identifying Tube Placement in Chest X-rays**. Rio de Janeiro, 2024. 47p. Final Project Proposal – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This paper investigates the use of Deep Learning techniques for semantic segmentation of central venous catheters (CVC) in chest X-rays. The goal is to assist in identifying these devices to determine their positioning, thereby reducing complications associated with invasive procedures. Initially, we address challenges such as the small relative size of the CVC devices in the images and data imbalance. For this, we utilize different backbones, such as Resnets and EfficientNets, in addition to making adjustments in image resizing. To improve the robustness and generalization capacity of the models, we apply data augmentation techniques. We implemented an ensemble of models, combining the results of various individual architectures, which proved effective by surpassing isolated models in various performance metrics. An additional script was developed to identify the presence of the CVC in the ensemble's predictions by analyzing the count of active pixels and contour detection. The final results demonstrated that the ensemble approach enhances the accuracy and reliability of CVC detection. Future research should focus on exploring the classification of device positioning as a subsequent step, aiming to further improve the clinical applicability of these techniques.

Keywords

Deep Learning; Semantic Segmentation; Chest X-ray; Radiograph.

Sumário

1	Introdução	8
2	Trabalhos Relacionados	10
3	Datasets	12
3.1	Origem dos Dados	12
3.2	Análise de Dados	13
4	Segmentação	18
4.1	Pré-Processamento e Geração de Datasets de Segmentação Semântica	18
4.2	Separação dos dados	21
4.3	Modelos de Segmentação	21
4.3.1	Arquitetura e Hiperparâmetros	21
5	Experimento e Etapas de Treinos	23
5.1	Primeira Etapa de Treinamento	23
5.2	Melhoria na Generalização	25
5.2.1	Data Augmentation	25
5.2.1.1	Transformação Geométrica	25
5.2.1.2	Alteração de Cor e Intensidade	26
5.2.1.3	Adição de Ruído e Desfoque	27
5.2.2	Resultados Treinamento	27
5.2.2.1	Amostras de Resultados	29
5.3	Terceira Etapa de Treinamento	30
5.3.1	Ensemble	30
5.4	Classificação de Presença do Dispositivo CVC na Previsão do Ensemble	32
6	Resultados	34
6.1	Métricas	34
6.1.1	Precisão (Precision)	34
6.1.2	Recall	34
6.1.3	F1-score	35
6.1.4	Intersection over Union (IoU) - Jaccard Index	35
6.1.4.1	Resultados das Métricas	35
6.1.5	Matriz de Confusão	39
7	Conclusão	42
	Referências bibliográficas	44

Lista de figuras

Figura 3.1	Porcentagem da presença ou não dos tubos CVC	13
Figura 3.2	Porcentagem das ocorrências das etiquetas	14
Figura 3.3	Ocorrências das etiquetas nas imagens, uma ou mais de uma etiqueta por imagem.	14
Figura 3.4	Ocorrências das etiquetas do CVC que aparecem juntas	15
Figura 3.5	Radiografias. À direita: com o três tubos CVC, sendo as etiquetas: 'Normal' azul; 'Borderline' Amarelo; 'Abnormal' vermelha; À esquerda: 'Normal' azul; 'Borderline' Amarelo;	15
Figura 3.6	Radiografias. À direita: 'CVC - Borderline'; À esquerda: 'CVC - Abnormal'	16
Figura 3.7	Radiografia 'CVC - Normal'	16
Figura 3.8	Ocorrências das etiquetas do CVC dos dados que possuem anotações da posição dos dispositivos médicos	17
Figura 4.1	Radiografias. À direita: Normal. À esquerda: Contrast Limited Adaptive Histogram Equalization (CLAHE)	19
Figura 4.2	Radiografias. À direita: Radiografia. À esquerda: Máscara gerada	20
Figura 4.3	Radiografia com a máscara sobreposta	20
Figura 5.1	Gráfico da função de perda dos modelos de segmentação	24
Figura 5.2	Gráfico da função de perda dos modelos de segmentação, imagens de 512x512	24
Figura 5.3	À direita: Radiografia com espelhamento. À esquerda: Máscara com espelhamento	26
Figura 5.4	À direita: Radiografia com rotação. À esquerda: Máscara com rotação	26
Figura 5.5	Radiografia com alteração de brilho e intensidade	27
Figura 5.6	Radiografia com ruído	27
Figura 5.7	Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet34, imagens de 512x512	28
Figura 5.8	Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone EfficientNetB3, imagens de 512x512	28
Figura 5.9	Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet50, imagens de 512x512	29
Figura 5.10	Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet101, imagens de 512x512	29
Figura 5.11	Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Resnet101; Quarta: Saída EfficientNetB3. O resultado esperado seria uma imagem completamente preta pois a radiografia não possui o dispositivo CVC	30

Figura 5.12 Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Resnet50; Quarta: Saída Esperada.	30
Figura 5.13 Primeira: Radiografia; Segunda: Saída EfficientNetB3; Terceira: Saída Resnet101; Quarta: Saída Ensemble.	31
Figura 5.14 Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Esperada; Quarta: Saída Ensemble.	32
Figura 5.15 Imagem considerada sem tubo por contagem de pixels	33
Figura 5.16 Imagem considerada sem tubo, pelo fato dos pixels brancos não estar próximo à alguma borda	33
Figura 6.1 Imagens: (1) Resnet34; (2) Resnet101; (3) Resnet50; (4) EfficientNetB3; (5) Ensemble; (6) Saída esperada	37
Figura 6.2 Imagens: (1) Resnet34; (2) Resnet101; (3) Resnet50; (4) EfficientNetB3; (5) Ensemble; (6) Saída esperada	37
Figura 6.3 À direita: Saída esperada. À esquerda: Saída do Ensemble	38
Figura 6.4 À direita: Saída esperada. À esquerda: Saída do Ensemble	38
Figura 6.5 À direita: Saída esperada. À esquerda: Saída do Ensemble	38
Figura 6.6 Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo Resnet34	39
Figura 6.7 Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo Resnet50	40
Figura 6.8 Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo EfficientNetB3	40
Figura 6.9 Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo Resnet101	40
Figura 6.10 Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do Ensemble	41

1 Introdução

Procedimentos invasivos que utilizam dispositivos como, sonda nasogástrica (NGT), tubo endotraqueal (ETT) e o cateter venoso central (CVC), quando mal posicionados podem desencadear complicações significativas, incluindo obstruções e agravamento do estado do paciente submetido a tais procedimentos [1, 2]. Além disso, observou-se que 25% dos casos envolvendo pacientes que são entubados fora da sala de cirurgia sofrem com o posicionamento inadequado de tubos de via aérea [3, 4].

A detecção precoce de tubos posicionados incorretamente desempenha um papel crucial na prevenção dessas complicações, sendo a confirmação mais confiável desse posicionamento por meio de radiografias de tórax [5, 6]. Por se tratar de imagens, devem-se realizar verificações visuais por profissionais de saúde, como médicos e radiologistas, para que se possa determinar a posição ideal dos dispositivos.

No entanto, esse processo pode ser afetado pela disponibilidade dos profissionais, podendo resultar em atrasos na identificação dessas situações. Além disso, de acordo com um relatório da National Patient Safety Agency, a principal causa de dano envolvendo dispositivos NGTs na Inglaterra, resultou da "interpretação inadequada de imagens de raio-X"[7].

Abordagens baseados em *Deep Learning* tem demonstrado ser excelentes recursos na área médica, devido à sua capacidade de aprender e identificar padrões em dados, incluindo aqueles que podem ser sutis e difíceis de detectar [8, 9, 10]. Estas técnicas vem sendo amplamente aplicadas na radiologia [11], principalmente pelo grande avanço na visão computacional [12, 13], possuindo uma precisão significativa em diferentes desafios como, na detecção de pneumonia [14, 15], detecção de tuberculose [16], detecção da COVID19 [17], utilizando radiografias de tórax como principal ferramenta diagnóstica. Além da detecção de doenças, modelos de *Deep Learning* possuem também o potencial de identificar o mal posicionamento de tubos e cateteres [18], contribuindo para a identificação precoce e eficaz desses dispositivos em radiografias de tórax.

Para realizar a identificação desses dispositivos em radiografias, uma possível etapa seria a utilização de segmentação semântica. Segmentação, do

ponto de vista da computação, refere-se ao processo de classificar cada pixel de uma imagem em uma categoria específica. A segmentação semântica é uma técnica que pode ser implementada com o uso da *Deep Learning*, na qual, redes neurais convolucionais (CNNs) e outras arquiteturas especializadas, como U-Net e SegNet, são desenvolvidas para realizar essa tarefa com alta precisão.

Dado este cenário, o objetivo principal deste projeto é propor uma solução que utilize segmentação semântica para identificar e definir na imagem a localização e extensão de dispositivos do tipo CVC em radiografias de tórax. Além de determinar a presença desses dispositivos, a solução também buscará localizá-los e marcá-los na imagem, fornecendo informações que podem ser posteriormente utilizadas para avaliar a qualidade do posicionamento dos tubos. Uma etapa essencial que precede a avaliação ou estimativa do quão correto está o posicionamento de dispositivos médicos. Essa avaliação pode ser realizada por meio de técnicas de classificação, garantindo a precisão e segurança dos procedimentos.

Este documento esta estrutura da seguinte forma: o Capítulo Trabalhos Relacionados oferece alguns trabalhos importantes para o problema em questão. O Capítulo Datasets exhibe e analisa o conjunto de dados utilizados no projeto. Já o Capítulo Segmentação aborda a estrutura base de solução utilizada. O Capítulo Experimento e Etapas de Treinos relata o desenvolvimento do trabalho e as técnicas utilizadas. Por fim os dois últimos Capítulos Resultados e Conclusão exibem os resultados alcançados e concluem o projeto.

2

Trabalhos Relacionados

O dispositivo médico cateter venoso central (CVC), é um dos diferentes tipos de procedimentos invasivos, sendo uma intervenção médica que envolve a introdução do dispositivo no corpo do paciente, com o acesso podendo ser feito através de incisões cirúrgicas, orifícios naturais ou técnicas percutâneas [19], com o objetivo de auxiliar nos cuidados dos pacientes.

O dispositivo CVC, é um dispositivo inserido em uma veia profunda calibrosa, proporcionando acesso ao sistema circulatório. Isso permite alguns usos, como a administração de medicamentos e fluídos intravenosos [20]. Desta forma, esses dispositivos devem ser inseridos e posicionados nos locais corretos, para que assim, funcionem corretamente e não causem complicações aos pacientes.

Para assegurar que esses dispositivos inseridos no corpo estão devidamente posicionados para o tratamento adequado, são utilizadas imagens radiográficas do tórax para realizar essa identificação. Esses dados são analisados por profissionais da saúde, podendo assim, tais análises estarem sujeitas a atrasos dada a disponibilidade do profissional, juntamente com possíveis erros devido há inexperiência [7].

Entretanto, a utilização de visão computacional pode ser um excelente recurso auxiliar para os profissionais da saúde. Técnicas de *Deep Learning* vêm sendo amplamente empregadas na radiologia e têm se mostrado uma ferramenta poderosa para abordar os desafios mencionados. [8, 9, 11, 21, 22].

Em um estudo para determinar a posição de ETTs em radiografia de tórax, Lakhani e outros [23], empregaram a transferência de aprendizado em uma DCNN Inception V3, para obter melhores resultados e agilizar o aprendizado. Durante o treinamento, técnicas de aumento de dados em tempo real, como rotação, cisalhamento, espelhamento horizontal e translação, foram aplicadas para diversificar o conjunto de dados de treinamento.

Em um problema semelhante, Lee et. al. [24], utilizam duas FCN para identificar o posicionamento de PICCs (Cateter Central de Inserção Periférica) em radiografias de tórax. O modelo foi treinado usando uma técnica de cascata de segmentação.

Contudo, Xin Yi, S. Adams e outros [25], concluem que a determinação

de cateteres em radiografias de tórax, exigirá uma integração de diferentes métodos, para conseguir determinar a presença do dispositivo, utilizando modelo para segmentação e posteriormente determinar a posição deste dispositivo utilizando modelo DCNN classificação. Além disso, eles destacam que a transferência de aprendizado pode ser um excelente recurso. No entanto, o *Deep Learning* ainda enfrenta alguns desafios na área da saúde, como dados insuficientes, interpretabilidade, privacidade de dados, modelos que possam ser integrados facilmente nos fluxos de trabalho clínicos existentes, entre outros [9, 25].

3 Datasets

Os dados consistem em um conjunto de radiografias de tórax, possuindo um total de 30.083 imagens. Além disso, há dois arquivos de anotações das imagens. O primeiro arquivo contém informações sobre o posicionamento dos tubos nas imagens para o problema de segmentação. O segundo arquivo possui as classes a que cada imagem pertence.

3.1 Origem dos Dados

O conjunto de dados utilizado neste projeto foi obtido através da competição "RANZCR CLiP - Catheter and Line Position Challenge", disponível no site Kaggle e organizado pela Royal Australian and NZ College of Radiologists [26]. Os dados estão disponíveis para participação na competição e/ou para pesquisa acadêmica e educação.

Foi utilizado um conjunto externo de imagens obtido através do conjunto "NIH Chest X-rays"[27], disponível no site Kaggle, de domínio público, podendo ser utilizado em pesquisas e estudos. Foram selecionadas imagens específicas para se adequarem ao primeiro conjunto de dados; ao todo, foram utilizadas 5.884 imagens de radiografia de tórax que não possuem o tubo CVC.

As imagens consistem em radiografias de tórax que podem conter as intervenções médicas NGT, ETT, CVC e o Swan Ganz, sendo este último um tipo de cateter distinto do CVC. Com exceção do Swan Ganz, o posicionamento desses dispositivos pode ser classificado em três categorias diferentes: 'Normal', 'Borderline' e 'Abnormal'. 'Normal' indica que o dispositivo está posicionado no local correto; 'Borderline' indica uma posição utilizável do dispositivo, porém não a ideal; e 'Abnormal' indica que o dispositivo está fora da posição. Nas radiografias de tórax, não há identificação do tipo de intervenção presente, além do próprio dispositivo médico inserido no corpo do paciente. As informações relacionadas a classe que o dispositivo pertence de acordo com a posição do mesmo, são obtidas através do arquivo de anotação.

3.2 Análise de Dados

As informações dos dados aqui analisados referem-se ao conjunto de radiografias que contêm a intervenção médica CVC. Ao todo, existem 32.979, como vemos em 3.1, ocorrências do tubo CVC em um total de 29.333 radiografias; aproximadamente 97% das imagens totais possuem esse dispositivo.

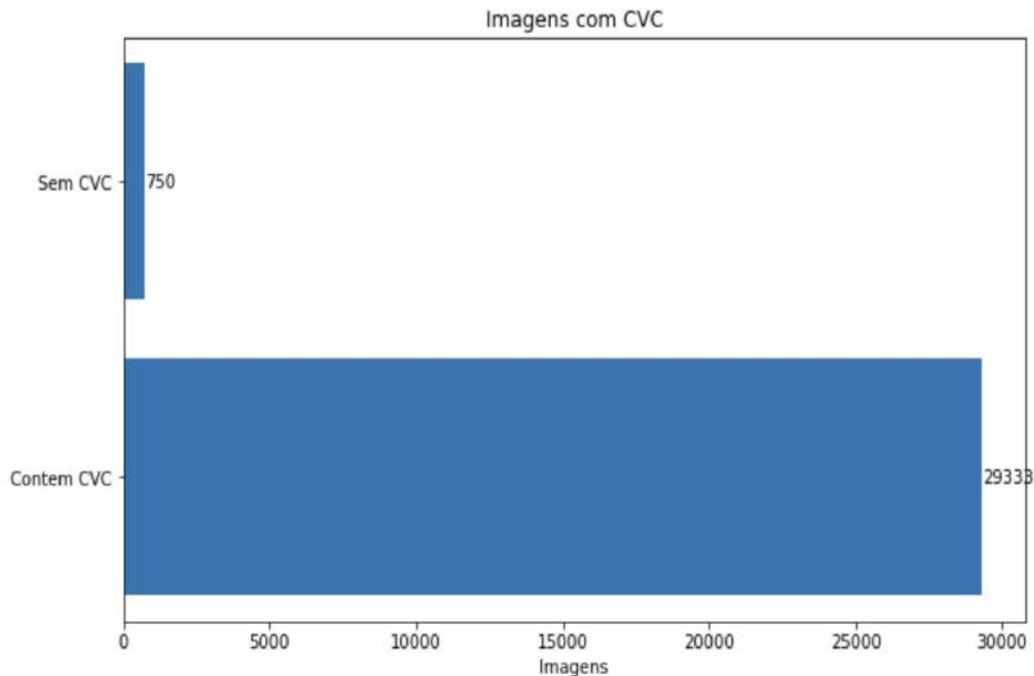


Figura 3.1: Porcentagem da presença ou não dos tubos CVC

Analisando as etiquetas, observa-se um desbalanceamento significativo nos dados. A categoria 'Normal' apresenta a maior disparidade em relação às outras categorias, possuindo 64,7% das ocorrências das etiquetas relacionadas ao CVC, com um total de 21.324 ocorrências de 'CVC - Normal', 3.195 de 'CVC - Abnormal' (9,7%) e 8.460 de 'CVC - Borderline' (25,6%), como exibido em 3.2.

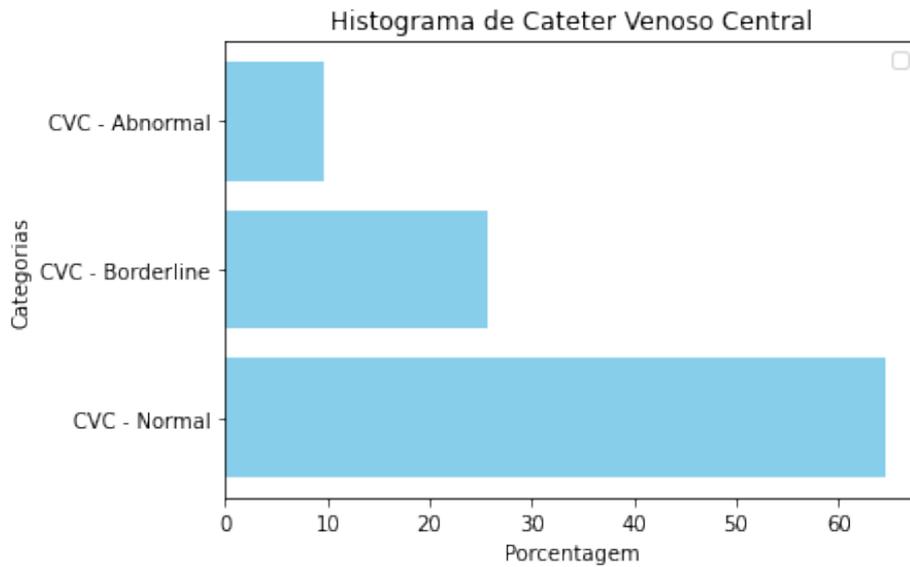


Figura 3.2: Porcentagem das ocorrências das etiquetas

Ao avaliar as ocorrências, pode-se compreender que uma radiografia pode possuir mais de uma etiqueta do tipo de intervenção médica CVC. Em 3.575 imagens, existe mais de uma possível etiqueta presente. Logo, 25.758 imagens possuem apenas uma etiqueta, como podemos ver em porcentagem em 3.3. Uma mesma radiografia pode ter uma etiqueta ou conter todas as três etiquetas existentes, ou seja, uma mesma imagem pode conter três ocorrências de CVC. Observa-se esses casos com as Figuras 3.5, 3.6, 3.7.

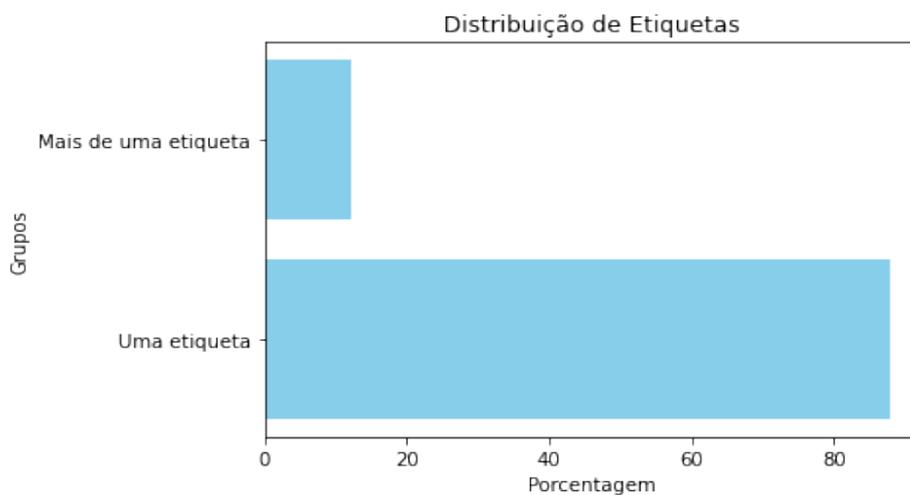


Figura 3.3: Ocorrências das etiquetas nas imagens, uma ou mais de uma etiqueta por imagem.

Existem 71 imagens que contêm os três casos (Normal, Borderline,

Abnormal), 2.536 envolvendo apenas Normal e Borderline, 658 que possuem Normal e Abnormal, e 310 imagens que possuem Borderline e Abnormal, Figura 3.4.

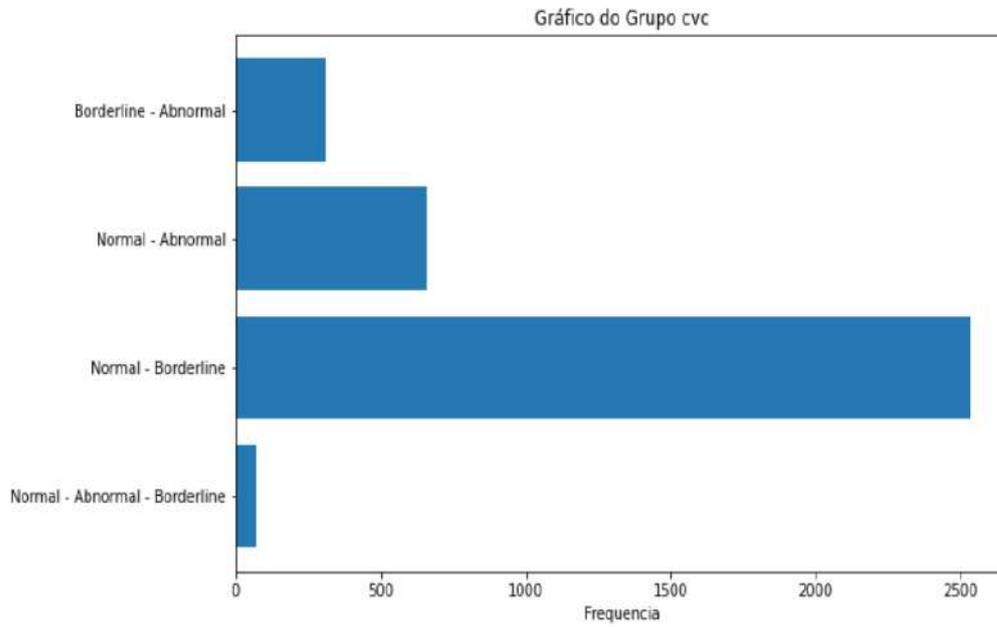


Figura 3.4: Ocorrências das etiquetas do CVC que aparecem juntas

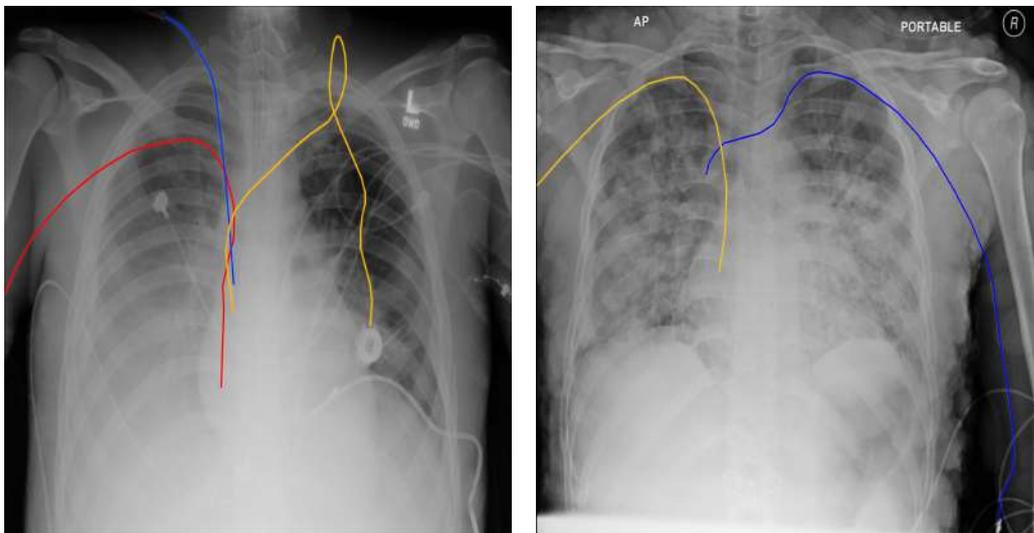


Figura 3.5: Radiografias. À direita: com o três tubos CVC, sendo as etiquetas: 'Normal' azul; 'Borderline' Amarelo; 'Abnormal' vermelha; À esquerda: 'Normal' azul; 'Borderline' Amarelo;

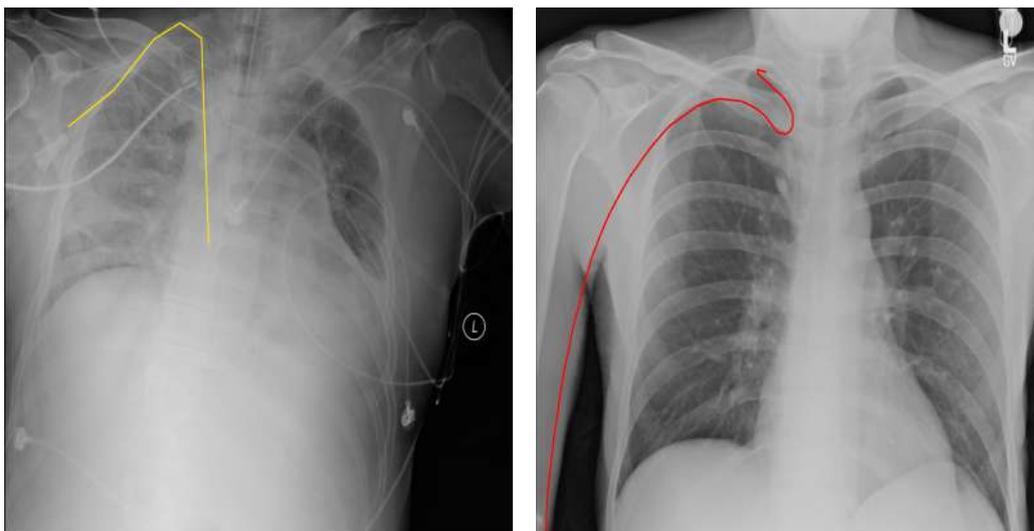


Figura 3.6: Radiografias. À direita: 'CVC - Borderline'; À esquerda: 'CVC - Abnormal'

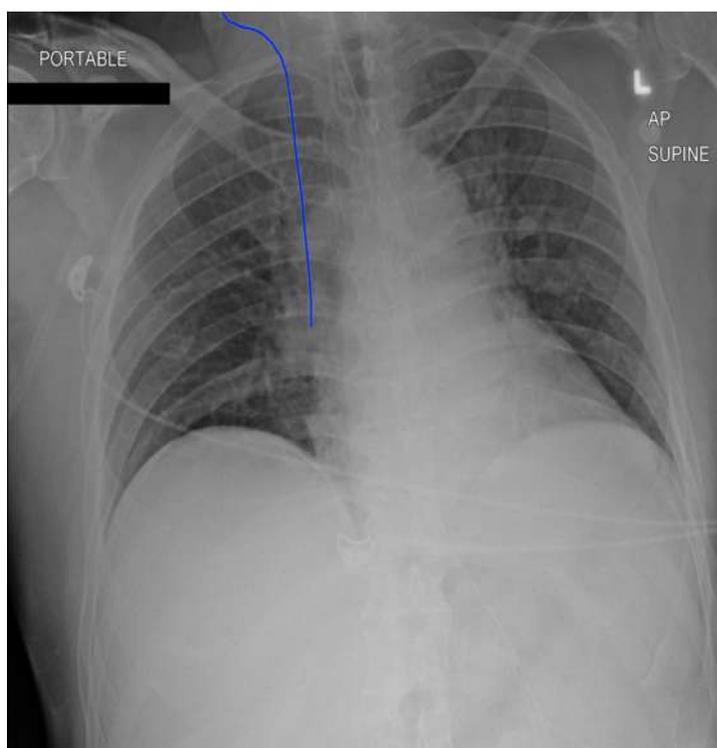


Figura 3.7: Radiografia 'CVC - Normal'

No que diz respeito aos dados utilizados para a geração de máscaras, que contêm anotações referentes às coordenadas dos dispositivos nas radiografias, é importante notar que nem todas as radiografias possuem essas informações. Das 9.095 radiografias que contêm essas anotações, 8.853 possuem ocorrências do dispositivo CVC.

Em relação às ocorrências das etiquetas, observa-se um total de 11.629, sendo 7.437 de 'CVC - Normal', 1.206 de 'CVC - Abnormal' e 2.986 de 'CVC - Borderline', vemos isso na Figura 3.8. Destacando mais uma vez que uma mesma radiografia pode conter mais de uma etiqueta.

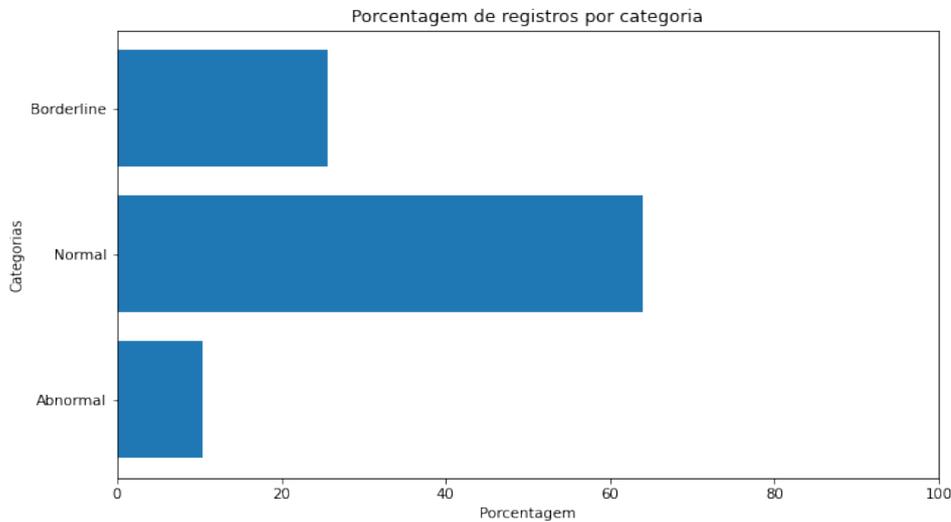


Figura 3.8: Ocorrências das etiquetas do CVC dos dados que possuem anotações da posição dos dispositivos médicos

Neste projeto, foi utilizado apenas o tipo de intervenção médica CVC. Essa escolha se deu porque após uma análise dos dados, constatou-se que os dispositivos CVC apresentavam o melhor conjunto de dados para a aplicação de técnicas baseadas em aprendizado supervisionado. Isso tornou o CVC o melhor ponto de partida devido à praticidade, disponibilidade e qualidade dos dados.

4 Segmentação

A segmentação semântica neste projeto tem como objetivo delinear a correta localização e extensão dos dispositivos presentes nas radiografias de tórax. Para atingir esse propósito, diversos modelos foram treinados e testados com o intuito de obter o melhor resultado na segmentação do cateter venoso central (CVC).

Os diferentes modelos desenvolvidos variaram tanto nos hiperparâmetros quanto na arquitetura utilizada, buscando identificar quais estruturas alcançariam os melhores resultados de aprendizado. O desenvolvimento deste projeto pode ser dividido em algumas etapas, uma etapa inicial para a análise de dados, pré-processamento e geração de conjuntos de dados de segmentação semântica, e outras três etapas de treinamento e análise de resultados, nas quais foram realizadas alterações na arquitetura, pré-processamento e hiperparâmetros, com o objetivo de aprimorar o aprendizado dos modelos de segmentação. Os dados foram selecionados e separados em três conjuntos: Treino, Teste e Validação. Após essa separação, foi aplicada uma etapa de pré-processamento nas radiografias e geração das máscaras.

A proposta inicial de solução foi utilizar dois modelos com diferentes arquiteturas e observar quais resultados os diferentes modelos obtiveram e quais as principais dificuldades enfrentadas no aprendizado. Com base nos resultados e nas observações dessa primeira fase, foi necessário realizar mudanças significativas, principalmente no pré-processamento dos dados. Essas modificações proporcionaram resultados superiores na segunda etapa de treinamento, além da utilização de arquiteturas mais robustas para um melhor aprendizado do problema. A terceira etapa consistiu em um *ensemble* dos modelos criados e treinados na segunda fase, obtendo assim resultados mais precisos ao utilizar as probabilidades de predição de mais de um modelo.

4.1 Pré-Processamento e Geração de Datasets de Segmentação Semântica

A etapa inicial que precede o modelo de segmentação, compreende o pré-processamento dos dados para a geração de máscaras. Nesta fase, as imagens passaram por etapas específicas para melhorar sua qualidade e facilitar

a segmentação. Técnicas de equalização de histograma foram aplicadas nas imagens aumentando o contraste para facilitar a identificação dos tubos nas radiografias, como exibo na Figura 4.1, visto que a equalização é um excelente recurso para melhorar os dados para o treinamento de modelos de segmentação [28].

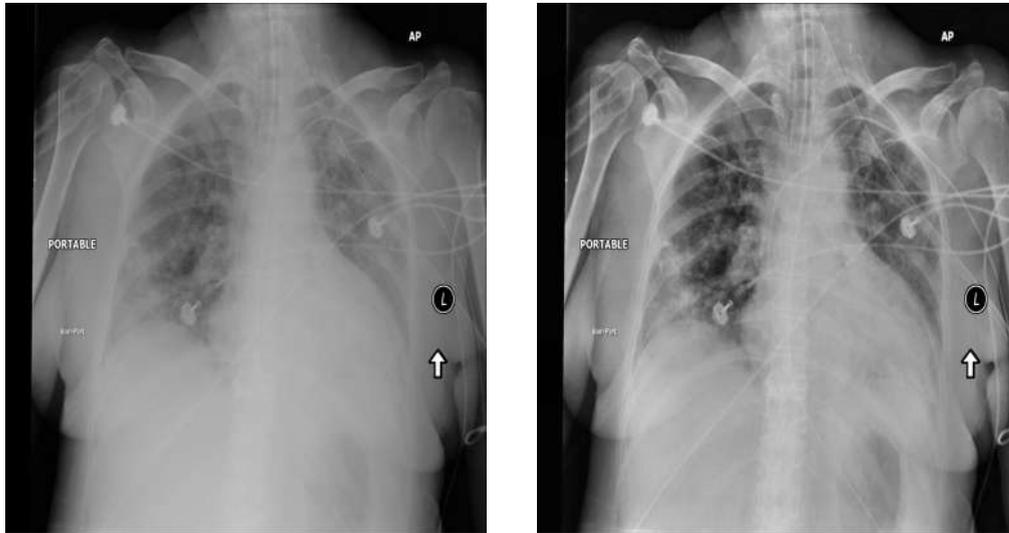


Figura 4.1: Radiografias. À direita: Normal. À esquerda: Contrast Limited Adaptive Histogram Equalization (CLAHE)

A etapa subsequente envolve o processo de mascaramento das imagens. Inicialmente, as anotações dos dados consistiam em uma série de coordenadas de pontos que indicavam as áreas dos dispositivos médicos de interesse nas radiografias. Utilizando um processo de rasterização, essas coordenadas foram transformadas em máscaras precisas. Essas máscaras, criadas a partir das anotações, representam a saída desejada do modelo de segmentação para cada entrada específica, permitindo a aplicação eficaz de técnicas de aprendizado supervisionado, Figuras 4.2, 4.3.

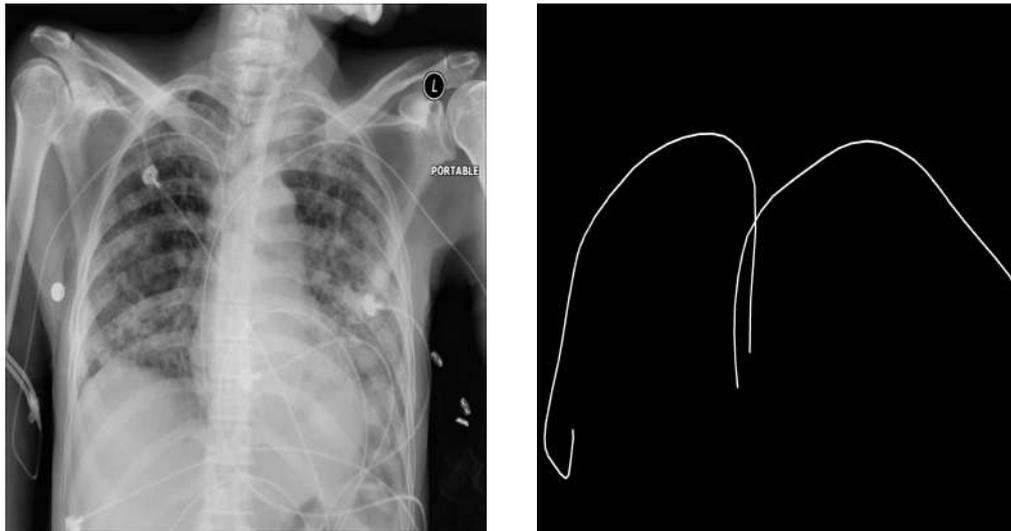


Figura 4.2: Radiografias. À direita: Radiografia. À esquerda: Máscara gerada

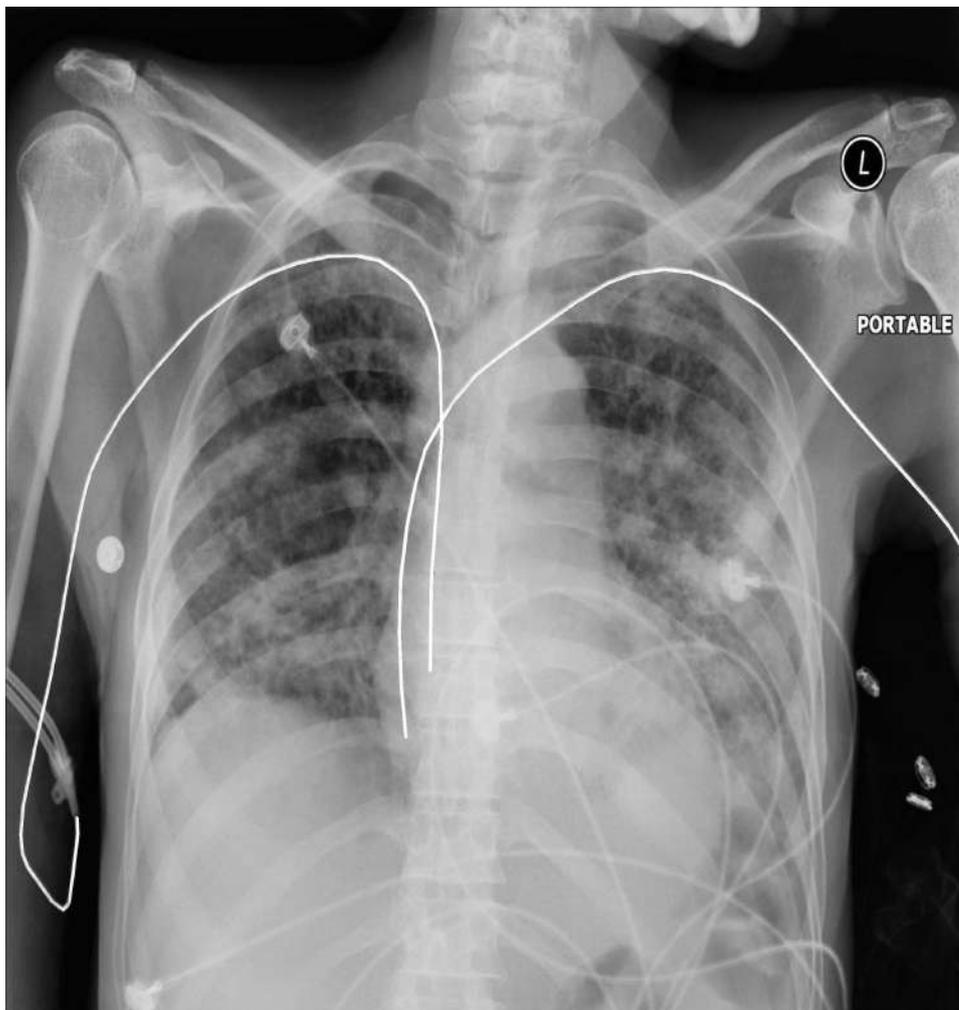


Figura 4.3: Radiografia com a máscara sobreposta

4.2

Separação dos dados

Os dados utilizados para o treino dos modelos de segmentação correspondem a 8.853 radiografias, nas quais é possível gerar as máscaras através das anotações. Conforme discutido na Seção Análise de Dados, a quantidade de imagens com o dispositivo CVC presente são consideravelmente maiores do que as correspondentes a radiografias sem a presença do CVC, que totalizam 750 imagens.

A separação dos dados para os conjuntos de Treino, Teste e Validação ocorreu da seguinte maneira: 60% dos dados foram alocados para o conjunto de treino, enquanto os 40% restantes foram divididos igualmente entre os outros dois conjuntos, ficando 20% para cada.

Os dados foram agrupados em cinco rótulos diferentes para manter um balanceamento ao separar aleatoriamente as radiografias em cada conjunto. Os rótulos consistem em 'CVC - Normal', 'CVC - Abnormal' e 'CVC - Borderline', que correspondem a radiografias com apenas uma etiqueta do rótulo ao qual pertencem. Os demais rótulos correspondem a radiografias que possuem mais de uma etiqueta, sendo separadas em '2 - tubos' e '3 - tubos'. Posteriormente foi agregado as imagens sem correspondência dos dispositivos CVC.

4.3

Modelos de Segmentação

4.3.1

Arquitetura e Hiperparâmetros

Após a preparação das imagens e das máscaras e separação dos conjunto de dados, inicia-se a etapa de estruturação e treinamento do modelo de segmentação. Optou-se pelo uso de uma arquitetura Unet, com a utilização de diferentes backbones, sendo as principais a Resnet34 e a EfficientNet B2 e B3. Para a função de ativação, foi escolhida a sigmoid, além disso, os modelos foram treinados com um batch de 4, no máximo 100 épocas para treino com um early stopping com paciência de 10 épocas, com o otimizador Adam inicializado com valores default.

Durante a fase de treino, optou-se por utilizar Dice Loss, como função de perda. Ela é eficaz nesse contexto, pois foca na similaridade entre as áreas segmentadas, ela traz uma vantagem significativa para o modelo, na qual, a área a ser analisada em comparação com a imagem total é pequena, já que os dispositivos representam uma pequena parte em comparação com toda a imagem. Além disso, no momento em que o modelo carrega as imagens, elas

são normalizadas para conterem valores entre 0 e 1 facilitando o aprendizado do modelo. Além disso, é fundamental considerar métricas para avaliar a performance do modelo treinado. Neste projeto, as métricas utilizadas são IoU (Intersection over Union), F1-score, Precision (Precisão) e Recall. A métrica IoU mede a sobreposição entre a área predita e a área real dos dispositivos segmentados, indicando a precisão da segmentação. O F1-score é a média harmônica entre precisão e recall, fornecendo uma única medida de desempenho do modelo. A Precision calcula a proporção de verdadeiros positivos em relação ao total de positivos preditos, enquanto o recall mede a proporção de verdadeiros positivos em relação ao total de positivos reais. Para obter mais informações sobre essas métricas, veja a Seção 6.1.

5 Experimento e Etapas de Treinos

Neste Capítulo, será apresentado detalhadamente os modelos de segmentação criados, que tinham como objetivo aprimorar os resultados obtidos durante cada etapa de treinos, destacando as mudanças realizadas entre as etapas de treinamento, bem como análises de desempenho obtidos em cada fase. Todos os modelos criados e treinados que foram detalhados neste Capítulo utilizam as mesmas informações dadas no Capítulo Segmentação na Seção Modelos de Segmentação.

5.1 Primeira Etapa de Treinamento

A Primeira proposta de solução para a segmentação semântica dos dispositivos médicos CVC, consiste na utilização de dois modelos diferentes treinados com o mesmo objetivo, para a entrada dos modelos as imagens foram padronizadas através do redimensionamento para o formato 384x384 com três canais. A saída esperada do modelo possui os mesmos valores de 384x384, porém, com apenas um canal. Foram treinados dois modelos com backbones diferentes: um Resnet34 e um EfficientNetB2. A Figura 5.1 exibe a evolução do treinamento com a função de perda Dice Loss. Foi observado que os modelos enfrentaram dificuldades de aprendizado, principalmente por dois motivos: os pontos de interesse (o CVC) são significativamente menores em comparação com as demais áreas da radiografia, e quanto menor a imagem, menor a proporção dessa área de interesse, o que pode dificultar o aprendizado do modelo. Os modelos possuem uma dificuldade em aprender quando não há a ocorrência dos dispositivos CVC, isso principalmente pelo desbalanceamento na quantidade de radiografias com e sem, abordado na Seção Separação dos dados.

Ao observar os resultados obtidos na Figura 5.1, algumas mudanças foram realizadas para melhorar o treinamento dos modelos. A primeira foi o redimensionamento do formato das imagens de 384x384 para 512x512, tanto para a entrada quanto a saída. Além disso, foi realizado um acréscimo nos dados com imagens sem a presença do dispositivo CVC, visando um melhor balanceamento nos conjuntos de dados. Foram adicionadas 5.884 imagens de

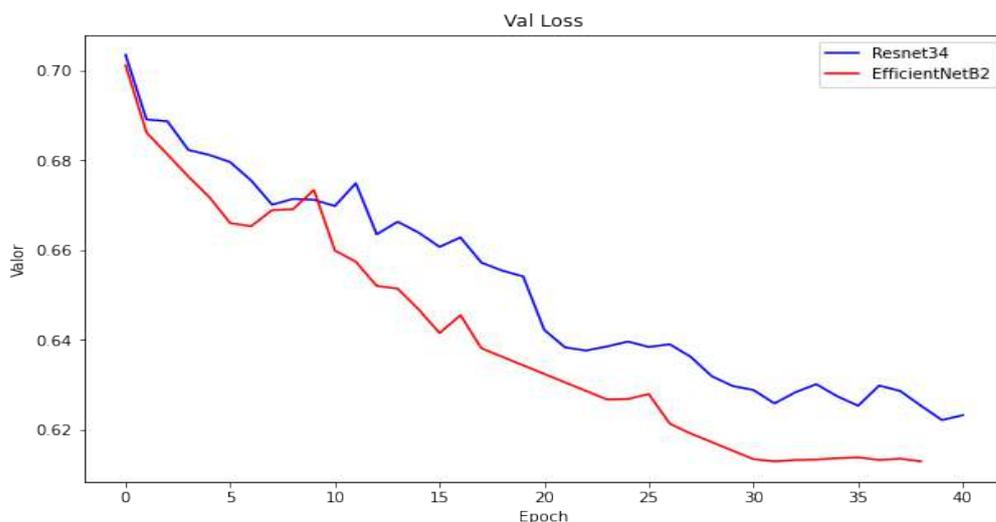


Figura 5.1: Gráfico da função de perda dos modelos de segmentação

um conjunto externo, conforme destacado na Seção Origem dos Dados, nos três conjuntos de Treino, Teste e Validação.

Após o treino dos modelos com as mudanças realizadas, porém, mantendo a mesma arquitetura dos modelos, os valores obtidos na função de perda no conjunto de validação melhorou em relação aos treinos anteriores. Contudo, desafios ainda foram encontrados durante a aprendizagem do modelo, um overfitting nos dados de treino e uma instabilidade na convergência nos dados de Validação, como destacado na Figura 5.2.

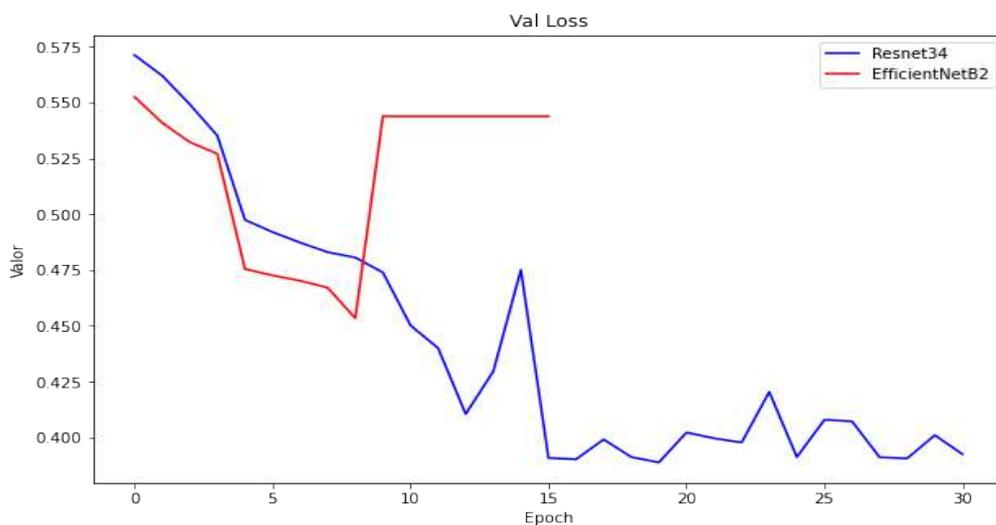


Figura 5.2: Gráfico da função de perda dos modelos de segmentação, imagens de 512x512

5.2

Melhoria na Generalização

Para evitar o overfitting durante os treinos e obter uma melhor generalização, problemas enfrentados na primeira etapa de treinos dos modelos, utilizou-se um método de data augmentation. Ao todo foram treinados 4 modelos, com entradas de 512x512 com três canais e saídas esperadas de 512x512 com um canal. Nesta segunda etapa, manteve-se a arquitetura UNet, mas utilizou-se arquiteturas mais robustas para os backbones, como: Resnet34, Resnet50, Resnet101, EfficientNetB3.

5.2.1

Data Augmentation

Para cada um dos diferentes modelos, foi utilizado os mesmos parâmetros para o *data augmentation*. Ao aplicar técnicas de data augmentation, é possível introduzir variações nas imagens de treinamento, como rotação, escala, translação, mudanças de brilho e contraste. Essas variações adicionam diversidade aos dados de treinamento, ajudando o modelo a aprender padrões mais robustos e generalizáveis [29][30]. Ao gerar novas amostras de dados de treinamento a partir das amostras existentes, o modelo recebe mais variedade, o que pode ajudar a reduzir o *overfitting* e estabilizar o treinamento. Isso pode ser observado no treinamento de cada um dos 4 modelos.

5.2.1.1

Transformação Geométrica

Durante os treinamentos, diversas técnicas de transformação geométrica foram empregadas nos dados, incluindo espelhamento horizontal, deslocamento, escalonamento e rotação, todas com uma probabilidade de 50%, (Figuras 5.3 e 5.4). Além disso, uma distorção elástica foi aplicada com uma probabilidade de 15%. Devido à complexidade dessa transformação, que aumenta significativamente o tempo de treinamento do modelo, sua probabilidade de aplicação foi mantida baixa. Apenas as técnicas geométricas são aplicadas nas saídas esperadas do modelo (as máscaras).



Figura 5.3: À direita: Radiografia com espelhamento. À esquerda: Máscara com espelhamento

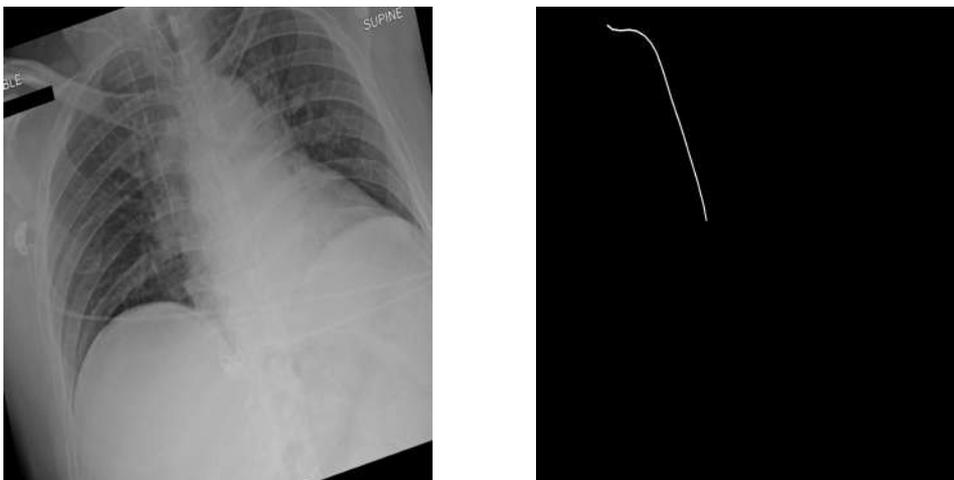


Figura 5.4: À direita: Radiografia com rotação. À esquerda: Máscara com rotação

5.2.1.2 Alteração de Cor e Intensidade

Foram utilizadas técnicas de ajuste aleatório de brilho e contraste nas imagens, aplicadas às radiografias com um limite de 20%, Figura 5.5. Cada técnica foi aplicada com uma probabilidade de 50%, sendo que apenas uma delas é aplicada.



Figura 5.5: Radiografia com alteração de brilho e intensidade

5.2.1.3

Adição de Ruído e Desfoque

Técnicas de adição de ruído e desfoque gaussiano foram aplicadas, com limites variando entre 5 e 10 para o ruído e entre 3 e 5 para o desfoque, ambas com uma probabilidade de 5% de serem aplicadas, Figura 5.6. Apenas uma dessas técnicas é aplicada.



Figura 5.6: Radiografia com ruído

5.2.2

Resultados Treinamento

Após a aplicação do *data augmentation*, os quatro modelos foram treinados, obtendo os seguintes resultados (Figuras 5.7 - 5.10). Os modelos com diferentes *backbones*, quando treinados com as mudanças realizadas e a utilização de técnicas de *data augmentation*, apresentaram melhores resultados

de perda tanto no conjunto de treino quanto no conjunto de validação. Eles conseguiram evitar o *overfitting* e generalizar melhor o aprendizado.

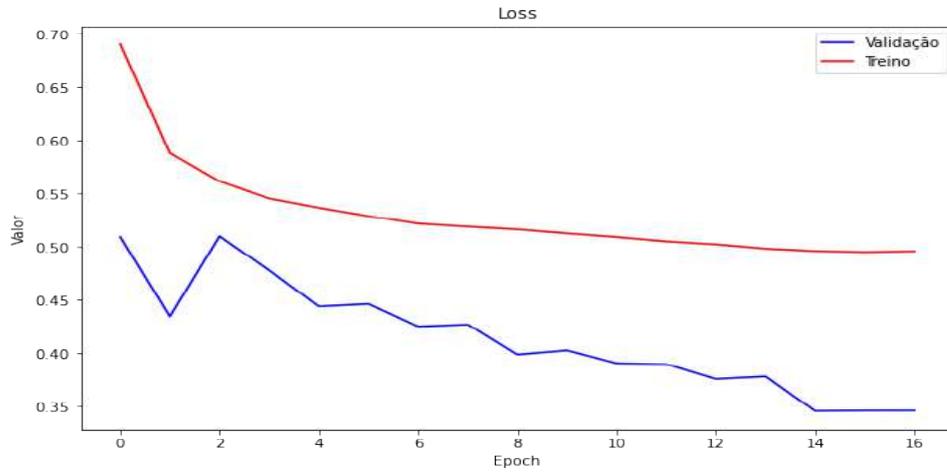


Figura 5.7: Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet34, imagens de 512x512

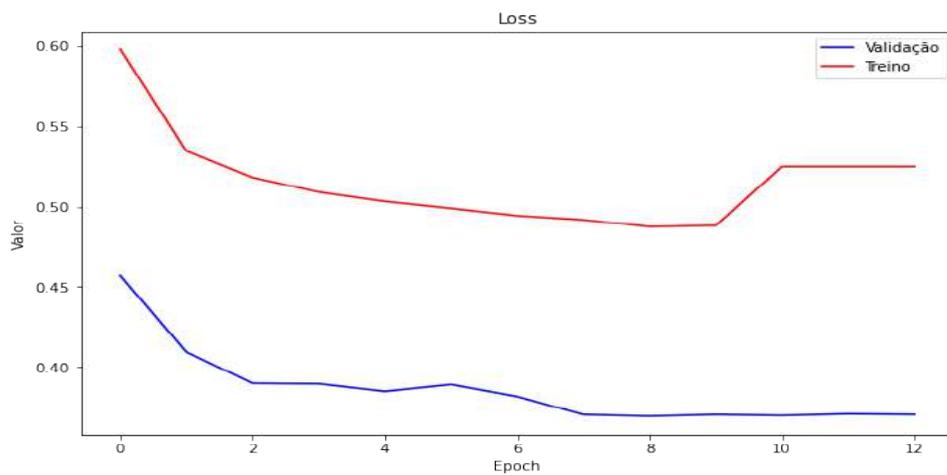


Figura 5.8: Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone EfficientNetB3, imagens de 512x512

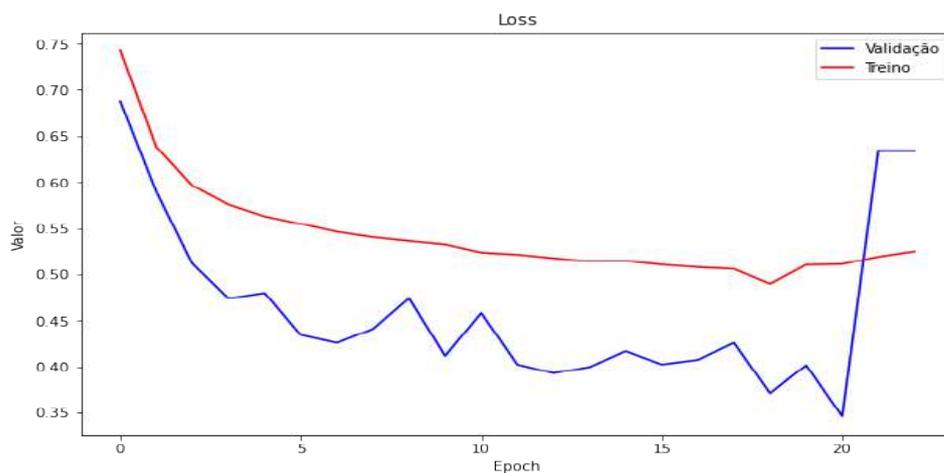


Figura 5.9: Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet50, imagens de 512x512

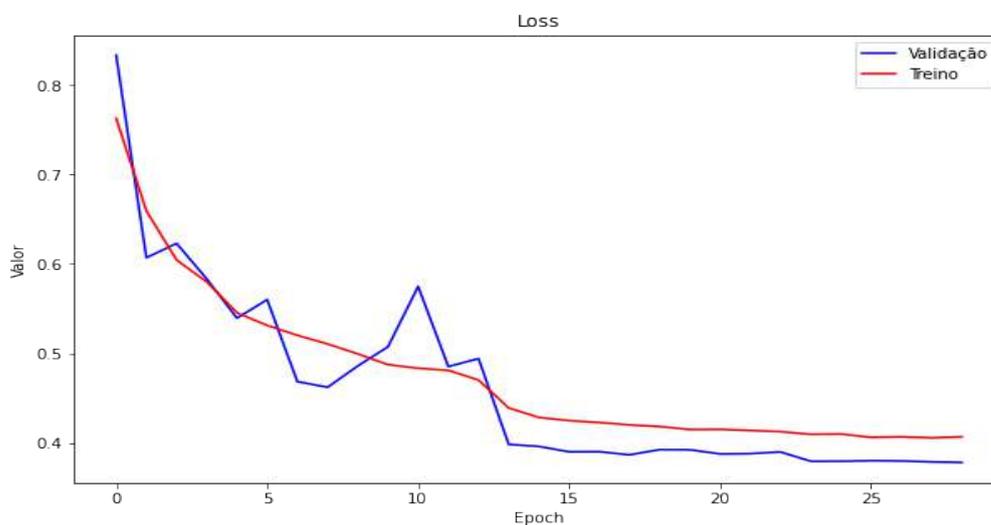


Figura 5.10: Gráfico da função de perda: Treino e Validação do modelo de segmentação com backbone Resnet101, imagens de 512x512

5.2.2.1

Amostras de Resultados

Através dos resultados visuais que os modelos obtiveram, fica mais fácil identificar os resultados, as qualidades e dificuldades dos diferentes modelos. Alguns têm mais facilidades para identificar quando há a ocorrência de apenas um dispositivo, enquanto outros possuem melhores resultados quando não há ocorrência dos dispositivos CVC (Figuras 6.1, 6.2).



Figura 5.11: Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Resnet101; Quarta: Saída EfficientNetB3. O resultado esperado seria uma imagem completamente preta pois a radiografia não possui o dispositivo CVC

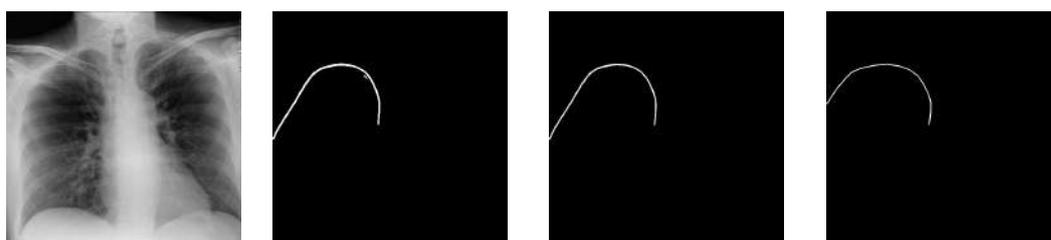


Figura 5.12: Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Resnet50; Quarta: Saída Esperada.

5.3

Terceira Etapa de Treinamento

Após os treinamentos e a análise dos resultados apresentadas na Seção 5.2, a terceira etapa do projeto visa aprimorar a precisão dos modelos através da implementação de um ensemble, uma técnica que combina os resultados de vários modelos individuais. A motivação por trás dessa abordagem é maximizar a qualidade das saídas, aproveitando as forças de cada modelo individualmente. O ensemble considera as probabilidades geradas por cada modelo para uma entrada específica. Combinar essas probabilidades de múltiplos modelos permite extrair o melhor aprendizado de cada um, resultando em saídas mais precisas e robustas.

É importante observar que os diferentes modelos apresentam resultados variados. As Figuras 5.13 e 5.14 ilustram essa variação. A combinação dos resultados desses modelos em um ensemble oferece a oportunidade de agregar essas variações de uma maneira que aumenta a qualidade das saídas, compensando as deficiências individuais de cada modelo.

5.3.1

Ensemble

Para implementar o ensemble, foi adotada uma abordagem que considera as probabilidades geradas por cada modelo para uma entrada específica. Cada

modelo produz uma probabilidade para cada pixel da imagem, indicando a possibilidade de ser um ponto de interesse.

Uma parte fundamental do processo é a definição de um limiar de decisão (threshold). Esse limiar determina como as probabilidades combinadas serão convertidas em uma previsão binária. Neste caso, foi estabelecido um limiar de 0,5. Isso significa que qualquer valor de probabilidade acima de 0,5 será considerado como pertencente à classe positiva, e qualquer valor abaixo será considerado como pertencente à classe negativa. Os resultados de cada modelo são combinados através do cálculo da média das probabilidades para cada pixel da imagem. Essa média representa a probabilidade agregada de existência de um ponto de interesse em cada pixel, conforme avaliado por todos os modelos.

Após calcular a média das probabilidades, o resultado é binarizado utilizando o threshold definido anteriormente. Se a probabilidade média de um pixel for maior que 0,5, ele será considerado como pertencente à classe positiva. A previsão binária final gerada pelo ensemble é uma representação das áreas consideradas como pontos de interesse. Essa saída pode ser comparada com as saídas individuais dos modelos para avaliar a eficácia do ensemble em melhorar a precisão e a robustez das previsões.

Podemos comparar os resultados exibidos nas Figuras 5.13 e 5.14 com os resultados obtidos pelo ensemble entre os modelos. Na Figura 5.13, a radiografia não possui um dispositivo CVC, com isso, pode-se observar que o resultado do *ensemble* de forma correta, não atribui pixels de interesse a essa radiografia, diferente dos resultados menos precisos obtidos pelos modelos sozinhos, treinados com diferentes *backbones*, que atribuem pixels de interesse aleatórios na imagem. Na Figura 5.14, o *ensemble* possui melhor fluidez e menos ruídos na imagem gerada, em relação a saída esperada.

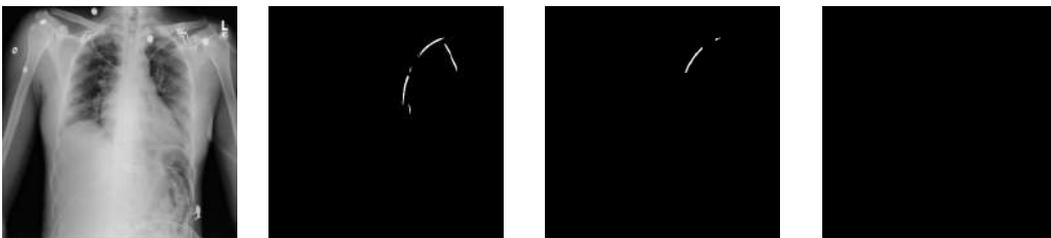


Figura 5.13: Primeira: Radiografia; Segunda: Saída EfficientNetB3; Terceira: Saída Resnet101; Quarta: Saída Ensemble.

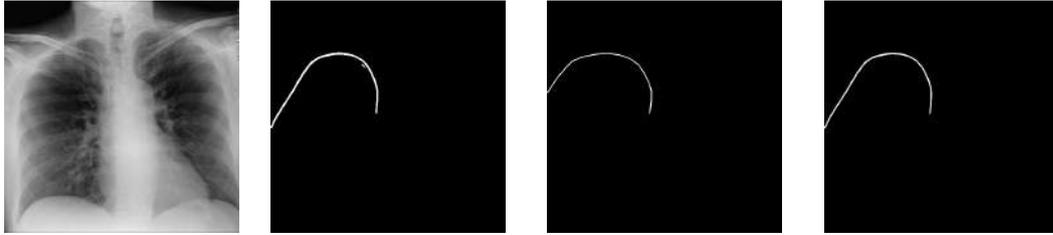


Figura 5.14: Primeira: Radiografia; Segunda: Saída Resnet34; Terceira: Saída Esperada; Quarta: Saída Ensemble.

5.4

Classificação de Presença do Dispositivo CVC na Previsão do Ensemble

Após a criação do ensemble e a obtenção de melhores precisões nas predições, foi desenvolvido um script capaz de identificar se as predições geradas pelo ensemble contêm a presença de um dispositivo CVC ou não. Isso permite determinar a presença do dispositivo e avaliar a capacidade dos modelos de não segmentar quando não há presença do tubo.

Esse script pode ser dividido em duas partes responsáveis por determinar se as predições possuem ou não o dispositivo. A etapa inicial consiste em contar quantos pixels estão ativos na imagem, ou seja, são diferentes de zero. Por meio dessa contagem, é possível identificar através de um limiar quando há ou não a presença de um tubo. Se essa contagem for menor que o limiar estabelecido, então não há tubo na imagem.

A segunda verificação é iniciada caso a primeira etapa indique a presença de um possível dispositivo CVC. Essa segunda etapa tem como objetivo detectar anomalias que possam sugerir que os pixels presentes não formam um dispositivo CVC conforme observado durante o projeto. Primeiramente, é realizado um fechamento morfológico para suavizar as linhas presentes na imagem, e então são encontrados os contornos. O maior contorno é selecionado com base na área, e o retângulo delimitador do contorno é calculado para determinar o comprimento do contorno. Dois limiares são usados para determinar se é um possível dispositivo médico CVC ou não. Se o comprimento do contorno encontrado for maior que o maior limiar, o dispositivo é confirmado. Se for menor que o menor limiar, é confirmado que não há CVC. Caso o comprimento esteja entre os dois limiares, verifica-se se os pixels ativos estão próximos à borda da imagem, pois foi observado que dispositivos médicos CVC presentes no dataset abordado nesse projeto, presente na Seção 3.1, com comprimento detectável entre esses dois limiares geralmente ficam próximos à borda da radiografia.

Essa estrutura do script auxilia na identificação da presença do dispositivo CVC, melhorando a robustez e a confiabilidade das previsões feitas pelo

ensemble. A Figura 5.15 mesmo contendo pixels de interesse, é definida como sem tubo pelo script, por falta de pixels suficientes para ser considerado uma imagem, a Figura 5.16, é definida como sem tubo pelo script, pelo comprimento do tubo identificado não ser o suficiente para ser considerado um dispositivo CVC.

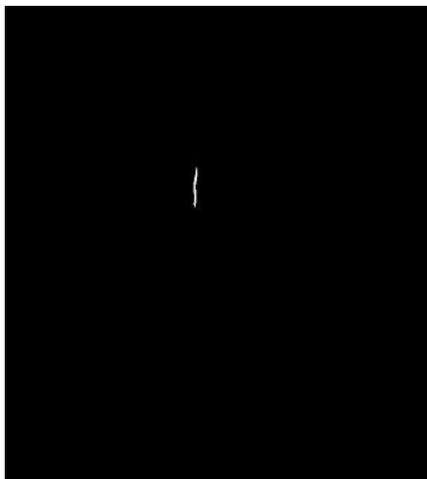


Figura 5.15: Imagem considerada sem tubo por contagem de pixels

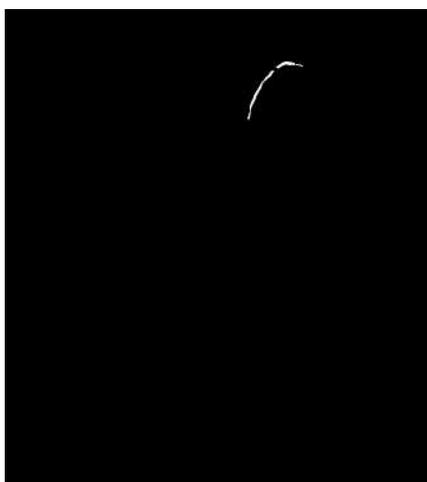


Figura 5.16: Imagem considerada sem tubo, pelo fato dos pixels brancos não estar próximo à alguma borda

6

Resultados

Para avaliar o desempenho dos modelos e treinos realizados no projeto é fundamental considerar métricas para avaliar a performance do modelo treinado. Muitas métricas de avaliação têm sido aplicadas no campo da segmentação de imagens médicas para avaliar os resultados da previsão da rede sob diferentes perspectivas. Durante a fase de avaliação, foram empregadas diversas métricas, tais como o IoU (Intersection over Union) que mensura a sobreposição entre as áreas segmentadas pelo modelo e as áreas verdadeiras nos dados, isso permite compreender a precisão do modelo ao gerar uma saída comparada com a saída esperada. Outras métricas importantes incluem o F1-score, Precision (Precisão) e Recall, que oferecem uma visão abrangente do desempenho do modelo na identificação e delimitação dos dispositivos médicos nas radiografias de tórax.

6.1

Métricas

6.1.1

Precisão (Precision)

A precisão mede a exatidão das predições positivas. Ela é calculada como a proporção de verdadeiros positivos (TP) entre o total de amostras preditas como positivas (TP + FP) [31]. Isso significa que a precisão avalia quantas das predições positivas foram corretas. A fórmula é:

$$\text{Precision} = \frac{TP}{TP + FP}$$

onde:

- TP (True Positives) é o número de verdadeiros positivos,
- FP (False Positives) é o número de falsos positivos.

6.1.2

Recall

O recall, também conhecido como sensibilidade, mede a capacidade do modelo de identificar todas as amostras positivas. É calculado como a

proporção de verdadeiros positivos (TP) entre o total de amostras que são realmente positivas (TP + FN) [31]. Isso significa que o recall avalia quantas das amostras positivas reais foram corretamente identificadas pelo modelo. A fórmula é:

$$\text{Recall} = \frac{TP}{TP + FN}$$

onde:

- TP (True Positives) é o número de verdadeiros positivos,
- FN (False Negatives) é o número de falsos negativos.

6.1.3

F1-score

O F1-score é a média harmônica da precisão e do recall. Ele oferece uma medida equilibrada entre precisão e recall, especialmente útil quando há um desequilíbrio entre as classes positivas e negativas. A fórmula é:

$$F1 = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}}$$

ou substituindo as fórmulas de precisão e recall:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

6.1.4

Intersection over Union (IoU) - Jaccard Index

A métrica IoU, também conhecida como Jaccard Index, mede a sobreposição entre a área predita e a área real. É calculada como a razão entre a área de interseção e a área de união dos conjuntos preditos e reais. A fórmula é:

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

onde:

- TP (True Positives) representa os verdadeiros positivos,
- FP (False Positives) representa os falsos positivos,
- FN (False Negatives) representa os falsos negativos.

6.1.4.1

Resultados das Métricas

A função IoU (Intersection over Union) permite uma compreensão mais precisa da performance dos modelos de acordo com suas saídas. Observando os

resultados de cada modelo para os conjuntos de validação pode-se compreender a precisão das saídas geradas por eles. Observando a Tabela 6.1 com resultados para o conjunto de validação e a Tabela 6.2 com resultados para o conjunto de Testes, pode-se identificar diferentes resultados obtidos pelos diferentes modelos, mostrando uma melhora no desempenho quando realizado o ensemble. Além de compreender que os modelos tem a capacidade de generalizar, obtendo resultados semelhantes nos dois conjuntos.

Métrica	Modelo	Valores
IoU	Resnet34	0.5352
	Resnet50	0.5200
	Resnet101	0.5277
	EfficientNetB3	0.5417
	Ensemble	0.5828
Precision	Resnet34	0.4680
	Resnet50	0.4646
	Resnet101	0.4353
	EfficientNetB3	0.4146
	Ensemble	0.4714
Recall	Resnet34	0.7600
	Resnet50	0.7511
	Resnet101	0.7925
	Ensemble	0.7989
F1-Score	Resnet34	0.5698
	Resnet50	0.5629
	Resnet101	0.5534
	Ensemble	0.5717

Tabela 6.1: Resultados no Conjunto de Validação

Métrica	Modelo	Valores
IoU	Resnet34	0.5017
	Resnet50	0.5084
	Resnet101	0.5229
	EfficientNetB3	0.5374
	Ensemble	0.5731
Precision	Resnet34	0.4421
	Resnet50	0.4512
	Resnet101	0.4224
	EfficientNetB3	0.3986
	Ensemble	0.4465
Recall	Resnet34	0.7734
	Resnet50	0.7599
	Resnet101	0.8026
	Ensemble	0.8255
F1-Score	Resnet34	0.5771
	Resnet50	0.5538
	Resnet101	0.5596
	Ensemble	0.5702

Tabela 6.2: Resultados no Conjunto de Testes

Os resultados visuais das predições das máscaras indicam que o modelo desenvolvido está funcionando corretamente com seu propósito, gerando máscaras que destacam as áreas dos dispositivos médicos. Essa observação visual confirma a capacidade do modelo de realizar segmentações precisas (Figuras 6.1 6.2 6.3 6.4 6.5).

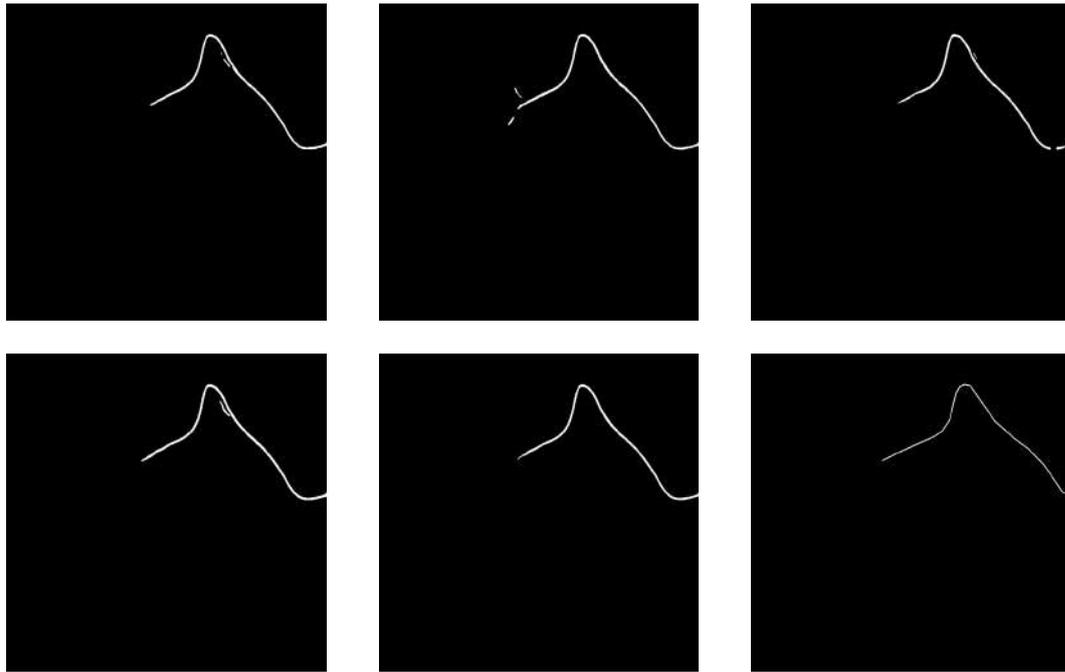


Figura 6.1: Imagens: (1) Resnet34; (2) Resnet101; (3) Resnet50; (4) Efficient-NetB3; (5) Ensemble; (6) Saída esperada

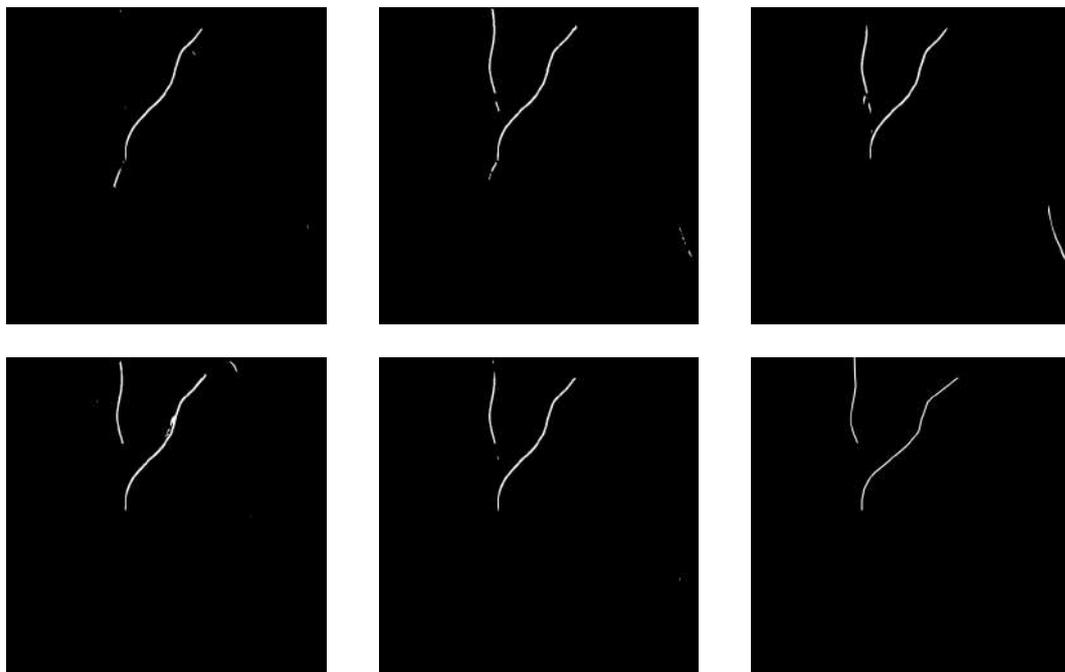


Figura 6.2: Imagens: (1) Resnet34; (2) Resnet101; (3) Resnet50; (4) Efficient-NetB3; (5) Ensemble; (6) Saída esperada

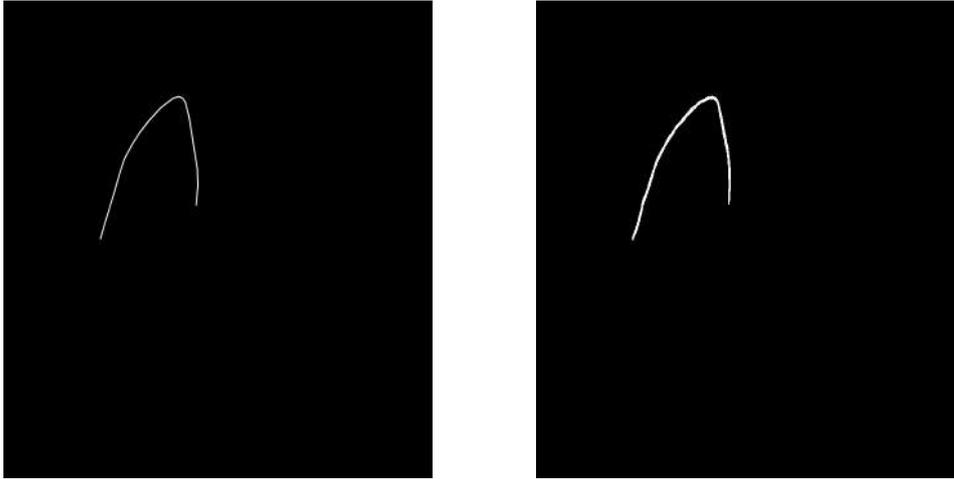


Figura 6.3: À direita: Saída esperada. À esquerda: Saída do Ensemble



Figura 6.4: À direita: Saída esperada. À esquerda: Saída do Ensemble

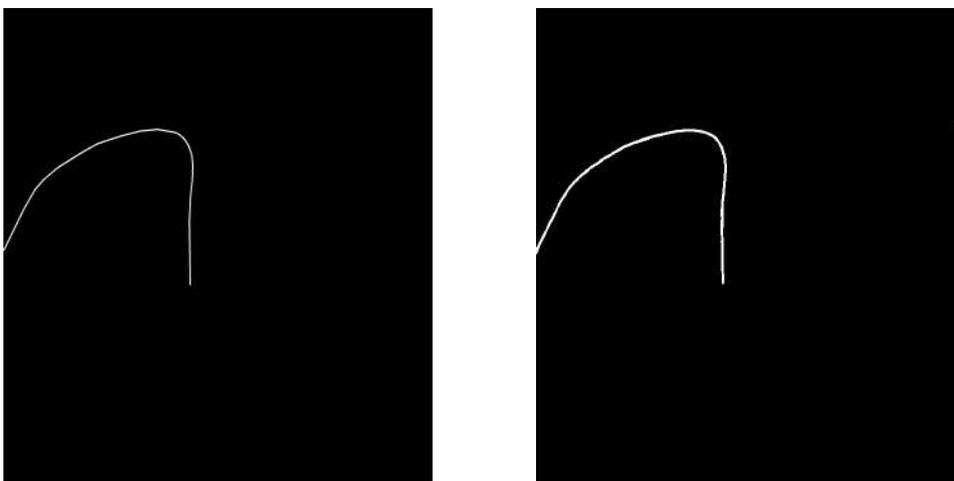


Figura 6.5: À direita: Saída esperada. À esquerda: Saída do Ensemble

6.1.5 Matriz de Confusão

A matriz de confusão fornece uma visão detalhada do desempenho de um modelo, permitindo a análise das previsões em termos de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Com base nesses valores, várias métricas de desempenho podem ser calculadas, conforme apresentado anteriormente na Seção 6.1. A matriz de confusão é uma excelente ferramenta para verificar os resultados do script de classificação de presença do dispositivo CVC utilizando previsões geradas pelo ensemble e comparar esses resultados com os dos modelos individuais, conforme apresentado nas Seções 5.4 e 5.3.1. As Figuras 6.6 6.7 6.8 6.9 6.10 que exibem as matrizes de confusão demonstram que, além do ensemble apresentar melhores resultados nas outras métricas, ele também gera previsões mais precisas para a identificação dos dispositivos.

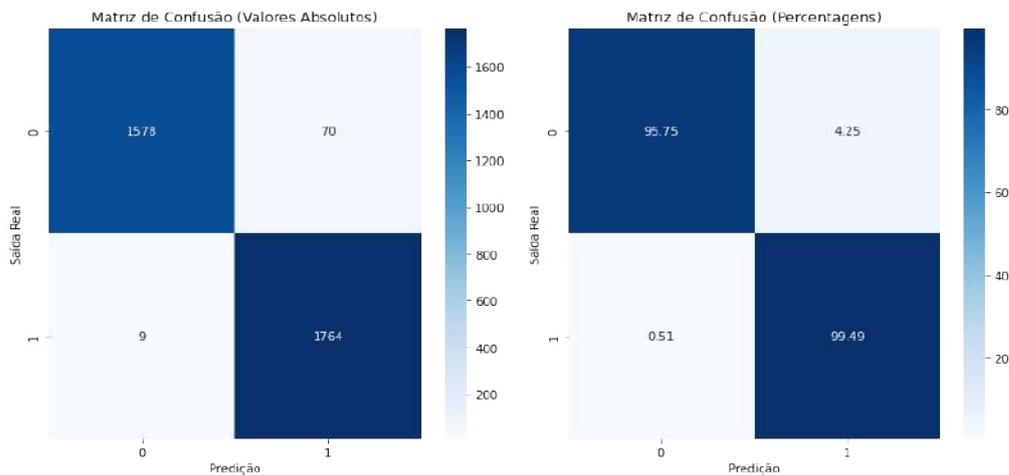


Figura 6.6: Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as previsões do modelo Resnet34

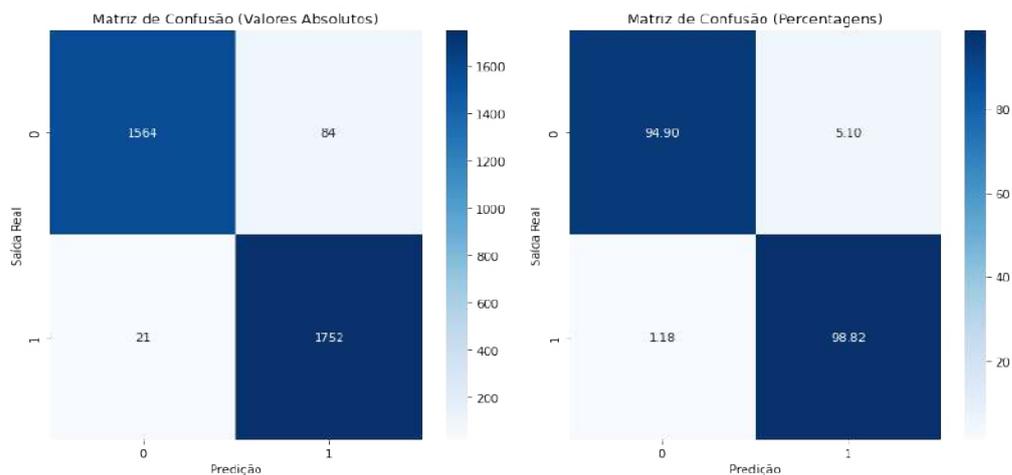


Figura 6.7: Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo Resnet50

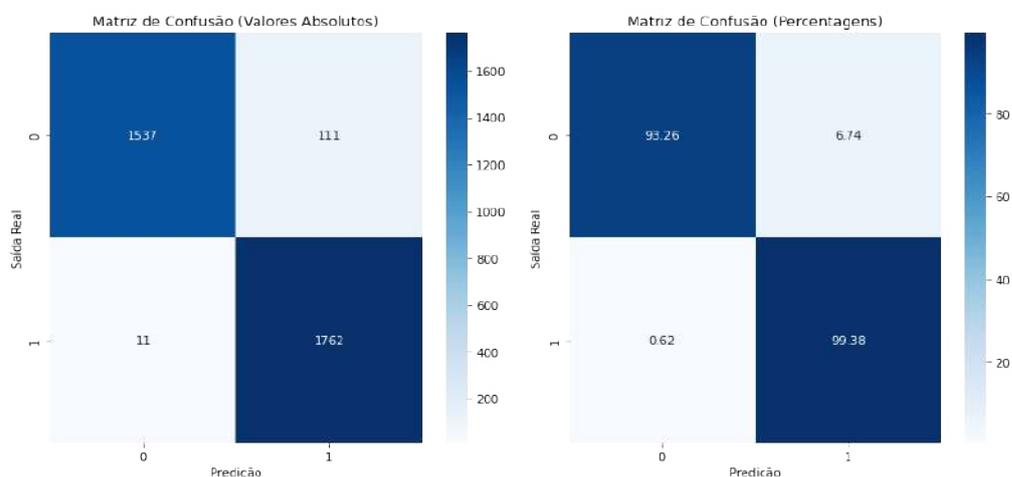


Figura 6.8: Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo EfficientNetB3

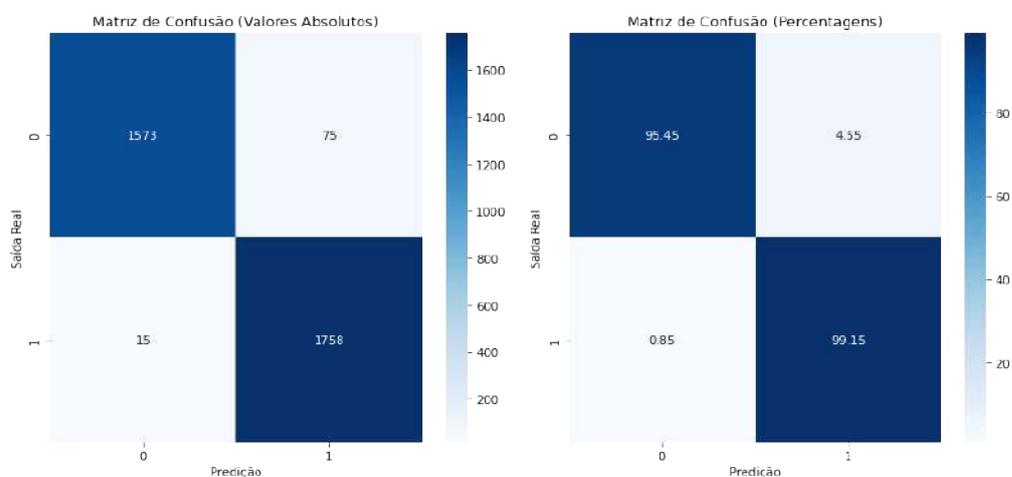


Figura 6.9: Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as predições do modelo Resnet101

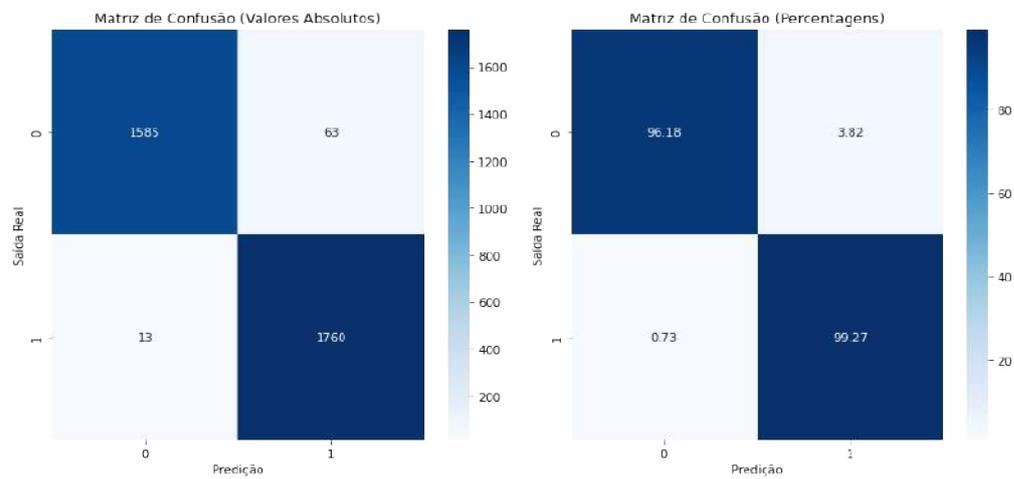


Figura 6.10: Matriz de Confusão gerada pelo script de classificação de presença do dispositivo CVC com as previsões do Ensemble

7

Conclusão

Este trabalho explorou a aplicação de técnicas de Deep Learning, especificamente segmentação semântica, para a detecção e classificação de dispositivos médicos, como o cateter venoso central (CVC), em radiografias de tórax. A motivação por trás deste estudo reside na possibilidade de técnicas de Deep Learning futuramente auxiliar profissionais da saúde na identificação de dispositivos mal posicionados, sendo a segmentação semântica uma etapa necessária anterior a de classificação para esse problema, essa identificação reduz possíveis complicações associadas aos procedimentos invasivos.

Inicialmente, a pesquisa abordou os desafios enfrentados durante o treinamento dos modelos de segmentação, como o pequeno tamanho relativo dos dispositivos CVC nas imagens e o desbalanceamento dos dados. A primeira etapa de treinamento utilizou modelos com diferentes backbones, ajustando o redimensionamento das imagens e adicionando mais dados para balancear os conjuntos de treino, teste e validação. Apesar de melhorias observadas na função de perda, problemas de overfitting e instabilidade na convergência ainda persistiram.

Para mitigar esses desafios, a segunda etapa de treinamento incorporou técnicas de data augmentation, introduzindo variações nas imagens para aumentar a robustez e a capacidade de generalização dos modelos. A utilização de backbones mais robustos, como Resnet101 e EfficientNetB3, juntamente com data augmentation, resultou em um desempenho aprimorado dos modelos.

A etapa final envolveu a implementação de um ensemble, combinando os resultados de vários modelos individuais para maximizar a precisão das predições. A técnica de ensemble demonstrou ser eficaz ao agregar as variações individuais dos modelos, resultando em saídas mais precisas e robustas. A abordagem de ensemble considerou as probabilidades geradas por cada modelo. Além disso, foi desenvolvido um script para identificar a presença do CVC nas predições do ensemble. O script analisa a contagem de pixels ativos e seguido por uma detecção de contornos para verificar a presença do dispositivo. Esta abordagem melhorou a precisão e a confiabilidade na detecção do CVC.

Os resultados finais, apresentados através de matrizes de confusão, demonstraram que o ensemble superou os modelos individuais em várias métricas

de desempenho, mostrando-se promissores. A capacidade de detectar corretamente a presença ou ausência do dispositivo foi aprimorada, porém ainda há a possibilidade de melhorias significativas nos resultados obtidos.

O aumento na quantidade de dados e a utilização de arquiteturas mais complexas contribuíram para a melhoria dos resultados. Com a continuidade desses aumentos, os resultados podem ser ainda mais robustos e precisos. No entanto, essa abordagem exige recursos computacionais significativos, o que representa um desafio prático para sua implementação em larga escala.

Futuras pesquisas podem focar na criação de modelos para classificar ou avaliar a qualidade do posicionamento do tubo, uma área que não foi o foco principal deste projeto. A otimização dos parâmetros de ensemble e a integração de novas técnicas de aprendizado também são áreas promissoras para melhorar ainda mais a precisão e a aplicabilidade clínica dessas abordagens.

Referências bibliográficas

- [1] ZADEH, M. K.; SHIRVANI, A.. **The role of routine chest radiography for detecting complications after central venous catheter insertion.** Saudi Journal of Kidney Diseases and Transplantation, 25(5):1011–1016, September 2014.
- [2] HALLORAN, O.; GRECU, B. ; SINHA, A.. **Methods and complications of nasoenteral intubation.** Journal of Parenteral and Enteral Nutrition, 35:61–66, 2011.
- [3] JEMMETT, M.. **Unrecognized misplacement of endotracheal tubes in a mixed urban to rural emergency medical services setting.** Academic Emergency Medicine, 10:961–965, 2003.
- [4] LOTANO, R.; GERBER, D.; ASERON, C.; SANTARELLI, R. ; PRATTER, M.. **Utility of postintubation chest radiographs in the intensive care unit.** Crit Care, 4(1):50–53, Jan 2000.
- [5] ABOOD, G.; DAVIS, K.; ESPOSITO, T.; LUCHETTE, F. ; GAMELLI, R.. **Comparison of routine chest radiograph versus clinician judgment to determine adequate central line placement in critically ill patients.** J Trauma, 63:50–56, 2007.
- [6] GLADWIN, M.; SLONIM, A.; LANDUCCI, D.; GUTIERREZ, D. ; CUNNION, R.. **Cannulation of the internal jugular vein: Is postprocedural chest radiography always necessary?** Crit Care Med, 27:1819–1823, 1999.
- [7] LAMONT, T.; BEAUMONT, C.; FAYAZ, A.; HEALEY, F.; HUEHNS, T.; LAW, R.; LECKO, C.; PANESAR, S.; SURKITT-PARR, M.; STROUD, M. ; OTHERS. **Checking placement of nasogastric feeding tubes in adults (interpretation of x-ray images): Summary of a safety report from the national patient safety agency.** BMJ, 342:d2586, 2011.
- [8] MIOTTO, R.; WANG, F.; WANG, S.; JIANG, X. ; DUDLEY, J. T.. **Deep learning for healthcare: review, opportunities and challenges.** Briefings in Bioinformatics, 19(6):1236–1246, 2018.

- [9] YANG, S.; ZHU, F.; LING, X.; LIU, Q. ; ZHAO, P.. **Intelligent health care: Applications of deep learning in computational medicine.** *Frontiers in Genetics*, 12:607471, 2021.
- [10] BUCHLAK, Q. D.; ESMAILI, N.; LEVEQUE, J.-C. ; OUTROS. **Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review.** *Neurosurgical Review*, 43:1235–1253, 2020.
- [11] MCBEE, M. P.; AWAN, O. A.; COLUCCI, A. T.; GHOBADI, C. W.; KADOM, N.; KANSAGRA, A. P.; TRIDANDAPANI, S. ; AUFFERMANN, W. F.. **Deep learning in radiology.** *Academic Radiology*, 25(11):1472–1480, 2018.
- [12] KRIZHEVSKY, A.; SUTSKEVER, I. ; HINTON, G. E.. **Imagenet classification with deep convolutional neural networks.** In: *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, p. 1097–1105, 2012.
- [13] HE, K.; ZHANG, X.; REN, S. ; SUN, J.. **Deep residual learning for image recognition.** In: *2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)*, p. 770–778, 2016.
- [14] KIM, J. H.; KIM, J. Y.; KIM, G. H.; KANG, D.; KIM, I. J.; SEO, J.; ANDREWS, J. R. ; PARK, C. M.. **Clinical validation of a deep learning algorithm for detection of pneumonia on chest radiographs in emergency department patients with acute febrile respiratory illness.** *Journal of Clinical Medicine*, 9(6):1981, 2020.
- [15] RAJPURKAR, P.; IRVIN, J.; ZHU, K.; YANG, B.; MEHTA, H.; DUAN, T.; DING, D.; BAGUL, A.; LANGLOTZ, C.; SHPANSKAYA, K. ; OTHERS. **Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning.** *arXiv preprint arXiv:1711.05225*, 2017.
- [16] QIN, Z.; SANDER, M.; RAI, B. ; OUTROS. **Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems.** *Scientific Reports*, 9:15000, 2019.
- [17] KHASAWNEH, N.; FRAIWAN, M.; FRAIWAN, L.; KHASSAWNEH, B. ; IBNIAN, A.. **Detection of covid-19 from chest x-ray images using deep convolutional neural networks.** *Sensors*, 21(17):5940, 2021.
- [18] SINGH, V.; DANDA, V.; GORNIK, R. ; OUTROS. **Assessment of critical feeding tube malpositions on radiographs using deep learning.** *Journal of Digital Imaging*, 32:651–655, 2019.

- [19] COUSINS, S.; BLENCOWE, N. S. ; BLAZEY, J. M.. **What is an invasive procedure? a definition to inform study design, evidence synthesis and research tracking.** *BMJ Open*, 9:e028576, 2019.
- [20] BERTOLIN, D. C.. **Cateterismo venoso central: Revisão atualizada das técnicas no procedimento.** *Revista Corpus Hippocraticum*, 2020. Acesso em: 22/10/2023.
- [21] AHMAD, H. K.; MILNE, M. R.; BUCHLAK, Q. D.; EKTAS, N.; SANDERSON, G.; CHAMTIE, H.; KARUNASENA, S.; CHIANG, J.; HOLT, X.; TANG, C. H. M. ; OUTROS. **Machine learning augmented interpretation of chest x-rays: A systematic review.** *Diagnostics*, 13(4):743, 2023.
- [22] TANG, C. H. M.; SEAH, J. C. Y.; AHMAD, H. K.; MILNE, M. R.; WARDMAN, J. B.; BUCHLAK, Q. D.; ESMAILI, N.; LAMBERT, J. F. ; JONES, C. M.. **Analysis of line and tube detection performance of a chest x-ray deep learning model to evaluate hidden stratification.** *Diagnostics*, 13(14):2317, 2023.
- [23] LAKHANI, P.; FLANDERS, A. ; GORNIK, R.. **Endotracheal tube position assessment on chest radiographs using deep learning.** *Radiology: Artificial Intelligence*, 3(1):e200026, 2021.
- [24] LEE, H.; MANSOURI, M.; TAJMIR, S. ; OUTROS. **A deep-learning system for fully-automated peripherally inserted central catheter (picc) tip detection.** *Journal of Digital Imaging*, 31:393–402, 2018.
- [25] YI, X.; ADAMS, S. J.; HENDERSON, R. D. E. ; BABYN, P.. **Computer-aided assessment of catheters and tubes on radiographs: How good is artificial intelligence for assessment?** *Radiology: Artificial Intelligence*, 2(1):e190082, 2020.
- [26] SEAH, J.; JEN; MAGGIE; LAW, M.; CULLITON, P. ; DOWD, S.. **Ranzcr clip - catheter and line position challenge.** <https://www.kaggle.com/competitions/ranzcr-clip-catheter-line-classification>, 2020.
- [27] CRAWFORD, C.; OF HEALTH CHEST X-RAY DATASET, N. I.. **Nih chest x-rays.** <https://www.kaggle.com/datasets/nih-chest-xrays/data>, 2018.
- [28] MAHESHAN, M.; HARISH, B. ; NAGADARSHAN, N.. **On the use of image enhancement technique towards robust sclera segmentation.** *Procedia Computer Science*, 143:466–473, 2018.

- [29] WANG, Y.; PAN, X.; SONG, S.; ZHANG, H.; WU, C. ; HUANG, G.. **Implicit semantic data augmentation for deep networks**. arXiv preprint arXiv:1909.12559, 2019. Submitted on 26 Sep 2019 (v1), last revised 25 Apr 2020 (this version, v5).
- [30] WU, X.; LIANG, J.; ZHANG, Y. ; TIAN, X.. **A segmentation method based on pdnet for chest x-rays with targets in different positions and directions**. Applied Sciences, 13(8):5000, 2023.
- [31] POWERS, D. M. W.. **Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation**. Journal of Machine Learning Technologies, 2(1):37–63, 2011.