



UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE TUCURUÍ
FACULDADE DE ENGENHARIA DE COMPUTAÇÃO

LUCAS VITOR LOCH PEREIRA

**AVALIAÇÃO DE MODELOS DE DETECÇÃO DE OBJETOS NA IDENTIFICAÇÃO
DE DOENÇAS PULMONARES E CARDÍACAS EM IMAGENS DE RAIO-X
TORÁCICOS**

TUCURUÍ
2024

LUCAS VITOR LOCH PEREIRA

**AVALIAÇÃO DE MODELOS DE DETECÇÃO DE OBJETOS NA IDENTIFICAÇÃO
DE DOENÇAS PULMONARES E CARDÍACAS EM IMAGENS DE RAIOS-X
TORÁCICOS**

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Orientador(a): Dr. Daniel da Conceição Pinheiro

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

P436a Pereira, Lucas Vitor Loch.
AVALIAÇÃO DE MODELOS DE DETECÇÃO DE
OBJETOS NA IDENTIFICAÇÃO DE DOENÇAS
PULMONARES E CARDÍACAS EM IMAGENS DE RAIOS-X
TORÁCICOS / Lucas Vitor Loch Pereira. — 2024.
58 f. : il. color.

Orientador(a): Prof. Dr. Daniel da Conceição Pinheiro
Trabalho de Conclusão (Graduação) - Universidade Federal do
Pará, Campus Universitário de Tucuruí, Faculdade de Engenharia
da Computação, Tucuruí, 2024.

1. Doenças pulmonares. 2. Doenças cardíacas. 3.
Radiografia torácica. 4. Detecção de objetos. 5. YOLO. I.
Título.

CDD 006.4

LUCAS VITOR LOCH PEREIRA

**AVALIAÇÃO DE MODELOS DE DETECÇÃO DE OBJETOS NA IDENTIFICAÇÃO
DE DOENÇAS PULMONARES E CARDÍACAS EM IMAGENS DE RAIO-X
TORÁCICOS**

Trabalho de Conclusão de Curso apresentado à Faculdade de Engenharia de Computação, do Campus Universitário de Tucuruí, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Bacharel em Engenharia de Computação.

Data da aprovação: ____/____/____

Conceito: _____

BANCA EXAMINADORA

Prof. Dr. Daniel da Conceição Pinheiro
Faculdade de Engenharia de Computação (UFPA)

Prof. Dr. Fabricio de Souza Farias
Sistema de Informação (UFPA)

Prof. Vigner Vieira dos Santos
Faculdade de Engenharia de Computação (UFPA)

Este trabalho é dedicado primeiramente a Deus, pois sem Ele nada eu seria, e a toda minha família pelo apoio incondicional durante essa caminhada.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por todas as bênçãos concebidas na minha vida e por ter me dado saúde e força para superar todos os desafios e obstáculos enfrentados durante esse percurso.

A toda minha família, por todo o amor e carinho dedicados, e pelo apoio incondicional para a concretização desse sonho.

Ao Professor Dr. Daniel da Conceição Pinheiro, orientador deste trabalho, pelos ensinamentos, pelo apoio e incentivo, os quais me ajudaram a conduzir este trabalho.

Aos meus amigos que sempre me apoiaram durante essa trajetória.

Aos colegas de curso, com quem formei grandes amizades, por todos os momentos compartilhados durante este percurso.

Aos demais professores pelo apoio e dedicação, e por todos os ensinamentos compartilhados ao longo destes anos de estudo.

E a todos que, de alguma forma, contribuíram ao longo da minha jornada acadêmica.

RESUMO

As doenças pulmonares e cardíacas representam um dos maiores desafios à saúde pública, sendo responsáveis por uma significativa taxa de mortalidade global, cenário que foi ainda mais agravado pela pandemia de COVID-19, que evidenciou a importância de diagnósticos precoces e precisos. Nesse contexto, a radiografia torácica destaca-se como um dos métodos mais eficazes para a detecção dessas patologias, pois permite uma análise detalhada da caixa torácica, pulmões e coração, fornecendo informações cruciais para o diagnóstico e acompanhamento clínico. Este trabalho propõe uma análise comparativa entre quatro modelos de detecção de objetos — YOLOv5, YOLOv8, Faster R-CNN e RetinaNet — com o objetivo de avaliar qual deles apresenta o melhor desempenho em precisão e sensibilidade na identificação de doenças pulmonares e cardíacas em imagens de raio-X torácico. A pesquisa examina as características específicas de cada modelo, considerando a eficácia na identificação de diversas patologias, como atelectasia, cardiomegalia, efusão, infiltração e pneumonia, e explora as métricas de avaliação, como precisão, sensibilidade e taxa de falsos positivos, para determinar qual modelo se destaca na prática clínica. Os resultados esperados visam contribuir para o avanço da detecção automatizada dessas doenças, oferecendo uma base sólida para a implementação de tecnologias de inteligência artificial em ambientes clínicos, com o intuito de melhorar a acurácia dos diagnósticos e, conseqüentemente, os desfechos dos pacientes.

Palavras-chave: doenças pulmonares; doenças cardíacas; radiografia torácica; detecção de objetos; YOLO.

ABSTRACT

Pulmonary and cardiac diseases represent one of the greatest challenges to public health, accounting for a significant global mortality rate, a scenario that has been further aggravated by the COVID-19 pandemic, which has highlighted the importance of early and accurate diagnoses. In this context, chest radiography stands out as one of the most effective methods for detecting these pathologies, as it allows a detailed analysis of the rib cage, lungs, and heart, providing crucial information for diagnosis and clinical follow-up. This work proposes a comparative analysis between four object detection models — YOLOv5, YOLOv8, Faster R-CNN, and RetinaNet — with the aim of evaluating which one presents the best performance in accuracy and sensitivity in identifying lung and heart diseases in chest X-ray images. The research examines the specific characteristics of each model, considering its effectiveness in identifying various pathologies, such as atelectasis, cardiomegaly, effusion, infiltration and pneumonia, and explores evaluation metrics, such as accuracy, sensitivity and false positive rate, to determine which model stands out in clinical practice. The expected results aim to contribute to the advancement of automated detection of these diseases, offering a solid basis for the implementation of artificial intelligence technologies in clinical settings, with the aim of improving diagnostic accuracy and, consequently, patient outcomes.

Keywords: pulmonary diseases; cardiac diseases; chest radiography; object detection; YOLO.

LISTA DE ILUSTRAÇÕES

Figura 1 - Raio-x de uma pessoa saudável e um caso de atelectasia.....	19
Figura 2 - Raio-x de uma pessoa saudável e um caso de cardiomegalia.....	19
Figura 3 - Raio-x de uma pessoa saudável e um caso de efusão	20
Figura 4 - Raio-x de uma pessoa saudável e um caso de infiltração	21
Figura 5 - Raio-x de uma pessoa saudável e um caso de pneumonia.....	22
Figura 6 - Modelo simplificado do método Faster R-CNN.....	24
Figura 7 - Modelo simplificado do método YOLO.....	24
Figura 8 - Exemplo de caso de IoU. A caixa vermelha indica a o objeto real da imagem e a caixa azul a tentativa de previsão do objeto.....	28
Figura 9 - Relação Entre as Doenças Cardíacas e Pulmonares	31
Figura 10 - Diagrama do Método de Classificação das Previsões	34
Figura 11 - Precisão Média para cada Versão do YOLOv5	35
Figura 12 - Precisão Média para cada Versão do YOLOv8	36
Figura 13 - Perda no Treino e Validação com Faster R-CNN.....	38
Figura 14 - Perda no Treino e Validação com RetinaNet.....	38
Figura 15 - Perda no Treino e Validação com YOLOv5N.....	39
Figura 16 - Perda no Treino e Validação com YOLOv8N.....	39
Figura 17 - Perda no Treino e Validação Após Ajuste nos Ciclos	40
Figura 18 - Matrizes de Confusão para Limar de Confiança $\geq 0\%$	42
Figura 19 - Frequência de Previsões por Classe para Faster R-CNN.....	43
Figura 20 - Frequência de Previsões por Classe para RetinaNet.....	43
Figura 21 - Frequência de Previsões por Classe para YOLOv5N	44
Figura 22 - Frequência de Previsões por Classe para YOLOv8N	44
Figura 23 - Matrizes de Confusão para Limar de Confiança $\geq 25\%$	45
Figura 24 - Histograma com Faixas de IoU para os Modelos	47
Figura 25 - IoU - Exemplo: Representações de caixas delimitadoras para três limiares diferentes de IoU, onde a caixa verde é a caixa verdadeira e a caixa vermelha é a prevista....	48
Figura 26 - Precisão Média para Dois Diferentes Limiares de IoU	49
Figura 27 - Comparação de Duas Previsões para um Raio-X – Em verde é possível ver a caixa delimitadora verdadeira e em vermelho as previsões, juntamente com suas confianças e valores de IoU.....	50
Figura 28 - Sensibilidade Média para dois Diferentes Limiares de IoU	51

Figura 29 - F1-score para Dois Diferentes Limiares de IoU.....52

LISTA DE TABELAS

Tabela 1 - Dados de Treinamento do Primeiro Treino	37
Tabela 2 - Dados de Treinamento do Treino Final	41

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Contextualização do Problema.....	14
1.2	Justificativa	16
1.3	Objetivos.....	16
1.4	Estrutura do Trabalho	17
2	REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS.....	18
2.1	Doenças Pulmonares e Cardíacas	18
2.1.1	Atelectasia	18
2.1.2	Cardiomegalia.....	19
2.1.3	Efusão	20
2.1.4	Infiltração.....	20
2.1.5	Pneumonia	21
2.2	Redes Neurais Convolucionais (CNN)	22
2.2.1	Faster R-CNN.....	23
2.2.2	You Only Look Once (YOLO).....	24
2.2.3	RetinaNet.....	25
2.2.4	Detectron2	26
2.3	Métricas de Avaliação em Modelos de Detecção de Objetos	26
2.3.1	Interseção sobre a União.....	27
2.3.2	Métricas de Análise Comparativa.....	28
3	METODOLOGIA.....	30
3.1	Método de Coleta de Dados	30
3.2	Ambiente de Desenvolvimento	32
3.3	Descrição dos Procedimentos	32
4	RESULTADOS E DISCUSSÕES	35
4.1	Escolha da Versão dos Modelos YOLO.....	35
4.1.1	YOLOv5	35
4.1.2	YOLOv8	36
4.2	Resultados dos Treinamentos	37
4.3	Escolha das Métricas de Avaliação	41
4.3.1	Definição do Limiar de Confiança	41
4.3.2	Definição do Limiar de Interseção	46

4.4	Avaliação e Análise Comparativa dos Modelos	48
5	CONCLUSÃO E TRABALHOS FUTUROS	53
	REFERÊNCIAS	55
	APÊNDICE A – REPOSITÓRIO COM CÓDIGOS	58

1 INTRODUÇÃO

1.1 Contextualização do Problema

As doenças pulmonares e cardíacas constituem um dos maiores desafios à saúde pública global e aumentam significativamente a morbidade e a mortalidade em vários grupos populacionais. Estas doenças não afetam apenas a qualidade de vida dos pacientes, mas também representam um pesado fardo para o sistema de saúde e para a economia. Além disso, conforme Angelim *et al.* (2024), a COVID-19 pode ter agravado esse quadro, visto o potencial desenvolvimento de comorbidades respiratórias e cardíacas durante a pandemia.

Conforme Carvalho *et al.* (2020), as doenças cardiovasculares estão entre as principais causas de morbidade e mortalidade no mundo, associadas principalmente ao sedentarismo, obesidade e tabagismo. Além disso, outros fatores como a idade e o sexo também podem influenciar quanto ao risco a esses tipos de doenças (MACARI, 2021).

Além de doenças cardiovasculares, as doenças pulmonares são tremendamente afetadas pelo tabagismo, principalmente entre os homens, apesar da diminuição do seu uso ao decorrer dos anos (LEAL, 2020). Entre as condições cardíacas mais prevalentes estão a insuficiência cardíaca e a cardiomegalia, enquanto doenças pulmonares comuns incluem a pneumonia e a Doença Pulmonar Obstrutiva Crônica (DPOC).

Estas doenças não apenas impactam profundamente a saúde dos indivíduos, mas também representam uma carga substancial para os sistemas de saúde em todo o mundo. Contudo, hoje em dia, com o advento das radiografias torácicas a identificação dessas doenças tornou-se cada vez mais recorrentes.

Atelectasia, pneumonia, efusão pleural, cardiomegalia e infiltração são condições específicas frequentemente detectáveis por meio de radiografias torácicas, que desempenham um papel crucial no diagnóstico e tratamento dessas doenças. Com base nisso, os modelos de detecção de objetos têm desempenhado um papel crucial na análise de imagens médicas, haja vista que oferecem soluções rápidas e precisas para a identificação de diversas patologias.

Anas *et al.*, em 2024, propuseram o uso do modelo YOLOv5 em conjunto com Mask R-CNN para melhorar a eficácia na identificação de tumores, onde o YOLOv5 foi ajustado para detectar massas e determinar sua natureza benigna ou maligna, enquanto o Mask R-CNN

identificou as bordas e tamanhos dos tumores. A combinação dos modelos melhorou a precisão em 0,06 pontos percentuais e pode auxiliar os médicos no diagnóstico precoce e no prognóstico mais preciso dos pacientes.

Ainda na linha do YOLOv5, o trabalho de Duraisamy, Arunkumar e Niranjani, publicado em 2024, apresenta o uso do modelo YOLOv5 para detecção da COVID-19 em imagens de raio-x de tórax, com foco na customização e aprimoramento do algoritmo. Nesse estudo os autores se propuseram a ajustar e treinar cuidadosamente o modelo com dados específicos, o que resulta em um modelo que se destaca na identificação de padrões intrincados relacionados à COVID-19 em imagens de raio-x, o que gerou uma precisão de 98% na detecção da patologia em imagens de radiografia.

O YOLOv8 também vem a ser utilizado cada vez mais na área médica, um exemplo disso é a aplicação proposta por Avanzato *et al.*, em 2024, que propõe uma estrutura chamada Lung-DT, que utiliza sensores de IoT e algoritmos de IA para estabelecer a representação digital da saúde respiratória de um paciente, e ao utilizar a rede neural YOLOv8, o sistema Lung-DT classifica com precisão as radiografias de tórax de até cinco categorias distintas de doenças pulmonares, com uma precisão média de 96,8%.

Bhojane *et al.*, em 2024, propôs uma outra abordagem para o uso do YOLOv8, que o utiliza para identificar lesões hepáticas, onde alcançou melhor uma precisão de 83,6% em um conjunto de dados de ressonância magnética, o que demonstra que o modelo é capaz de identificar anormalidades hepáticas com precisão e rapidez.

Ao olharmos para outra estrutura de detecção de objetos, há estudos que fizeram uso do Faster R-CNN com o mesmo propósito dos modelos anteriores, ser de utilidade na área médica. Ding *et al.*, em 2019, propôs o uso de um sistema de inteligência artificial em conjunto com o modelo Faster R-CNN para avaliar linfonodos metastáticos em pacientes com câncer retal. O estudo demonstrou que o uso da IA com Faster R-CNN conseguiu uma precisão superior à dos radiologistas, mesmo que não no nível de patologistas.

Por fim, o modelo RetinaNet também é estudado cada vez mais na detecção de patologias. Em 2023, Daoud *et al.* propuseram o uso de tal modelo para detectar áreas de interesse em imagens de ultrassom de mama, e as descobertas obtidas sugerem que o modelo RetinaNet fornece uma abordagem viável e eficaz para essa tarefa, além de chegar a alcançar 89% de precisão em suas detecções.

Com base no que foi apresentado, esse trabalho se propõe a apresentar uma avaliação comparativa de quatro diferentes modelos de detecção de objetos ao serem analisados no contexto clínico, em específico, na identificação de doenças pulmonares e cardíacas em imagens de raio-x torácico.

1.2 Justificativa

A radiografia de tórax é um dos exames radiológicos mais acessíveis e amplamente utilizados para a triagem e diagnóstico de diversas doenças pulmonares. Esse exame se tornou ainda mais presente na vida de muitos brasileiros durante e após a pandemia de COVID-19, que impôs um grande desafio à saúde pública global, destacando a necessidade de diagnósticos rápidos e precisos. Contudo, o problema da alta demanda por exames radiológicos vai além do cenário pandêmico, pois a triagem de doenças pulmonares é uma necessidade constante, especialmente em ambientes com limitações de recursos, onde pode não haver um técnico especializado disponível para realizar a análise imediata das imagens.

Diante desse cenário, a implementação de técnicas automáticas de detecção de objetos em radiografias pode auxiliar os profissionais de saúde, de modo a proporcionar uma segunda opinião automatizada, o que agilizará o diagnóstico e, conseqüentemente, contribuirá para a tomada de decisões mais rápidas e precisas no tratamento de diversas condições médicas. Assim, o uso dessas novas tecnologias representa uma solução promissora para otimizar a eficiência e a precisão no diagnóstico por imagem, particularmente em situações onde o acesso a especialistas é limitado.

1.3 Objetivos

O objetivo desta pesquisa é utilizar o método comparativo para analisar o desempenho de diferentes modelos de detecção de objetos - YOLOv5, YOLOv8, Faster R-CNN e RetinaNet - e assim identificar o método mais eficaz em termos de precisão e eficiência na detecção de cinco classes específicas de doenças pulmonares, e assim ajudar a diagnosticar com maior precisão essas condições médicas.

Os objetivos específicos do trabalho são:

- Determinar qual a melhor versão dos modelos YOLOv5 e YOLOv8 para a análise proposta;
- Determinar boas métricas para o treinamento dos modelos de detecção de objetos;
- Definir um bom limiar de confiança para a análise das previsões;
- Definir um bom limiar de interseção sobre a união para a análise das previsões;
- Verificar o impacto de imagens em escala de cinza na tarefa de detecção de objetos.

1.4 Estrutura do Trabalho

Este trabalho está estruturado em cinco capítulos principais, onde o primeiro capítulo nos traz uma introdução e esboça uma visão geral sobre o trabalho desenvolvido, a justificativa e os objetivos do trabalho.

O segundo capítulo apresenta o embasamento teórico, que abrange pontos importantes sobre doenças pulmonares e cardíacas e os diferentes modelos de detecção de objetos, além de suas características.

O terceiro capítulo traz a metodologia utilizada, que consiste da apresentação dos materiais e métodos que foram utilizados para a escrita deste trabalho, a base de dados, métricas de avaliação, ambiente de desenvolvimento e os diferentes processos e procedimentos para a criação da sintaxe de comparação.

O capítulo quatro aborda as análises obtidas a partir da comparação do comportamento dos diferentes modelos para a base de dados utilizada.

Por fim, o capítulo cinco apresenta as conclusões alcançadas e delinea os objetivos pretendidos para trabalhos futuros.

2 REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

Neste capítulo, serão apresentados os principais conceitos sobre doenças pulmonares e cardíacas, a importância das radiografias torácicas, a aplicação da visão computacional na área médica, e os modelos de detecção de objetos utilizados no estudo. Esses tópicos fornecerão a base teórica necessária para compreender e contextualizar a solução proposta.

2.1 Doenças Pulmonares e Cardíacas

As doenças pulmonares e cardíacas são condições comuns que afetam milhões de pessoas em todo o mundo. As doenças pulmonares, como atelectasia, até mesmo infiltrações, comprometem a função respiratória, enquanto as doenças cardíacas, como cardiomegalia, impactam a circulação sanguínea. Ambas as condições representam desafios significativos para a saúde pública, tornando o diagnóstico precoce e preciso fundamental para melhorar os desfechos dos pacientes.

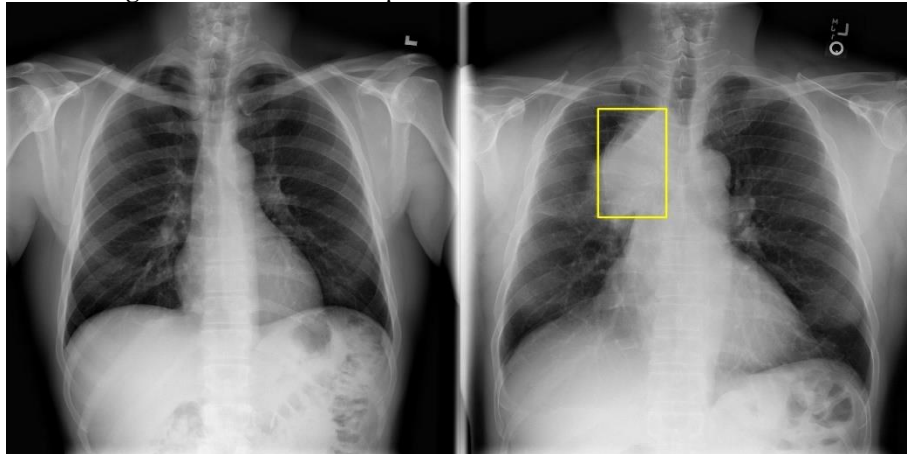
A seguir, serão apresentadas as doenças para as quais se propõe um método de identificação: atelectasia, cardiomegalia, efusão, infiltração e pneumonia.

2.1.1 Atelectasia

Segundo Dominguez e Alvares (2018), a atelectasia pulmonar consiste no colapso pulmonar, acompanhado de hipoventilação, que pode acometer um lobo, segmento ou todo o pulmão, o que resulta na diminuição da relação ventilação/perfusão. Além disso, conforme Serejo *et al.* (2007) tais complicações pulmonares são mais frequentes em pacientes submetidos a cirurgias abdominais, e sua incidência varia conforme a população estudada.

Shanthi *et al.* (2022) explica que entre os principais sintomas de tal doença estão a dificuldade em respirar, respiração rápida/superficial, chiado no peito e tosse. Ademais, elas podem ser identificadas através de um exame de raio-x como mostra a Figura 1, onde uma opacificação da zona superior, próxima ao centro, na parte esquerda, pode ser usada para identificá-la.

Figura 1 - Raio-x de uma pessoa saudável e um caso de atelectasia



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

2.1.2 Cardiomegalia

A cardiomegalia é uma condição médica caracterizada pelo aumento anormal do tamanho do coração. Para Amin e Siddiqui (2021) a cardiomegalia não é uma doença por si só, mas um sinal de outra condição subjacente que afeta o coração, como hipertensão, doença coronariana, ou insuficiência cardíaca. Além disso, estudos demonstraram que os homens são mais afetados pela cardiomegalia do que as mulheres (ALGHAMDI, 2020).

O diagnóstico de tal enfermidade pode ser feito através de exames de imagem, como raio-x torácicos, onde o aumento do coração é visível. Conforme Shanthi *et al.* (2022), a cardiomegalia está presente se a razão cardiororácica é maior que 0.5 (50%), ou seja, se a maior distância transversal entre as bordas esquerda e direita do coração da silhueta cardíaca é maior que a metade da largura do tórax. Tal aumento excessivo do coração pode ser observado na Figura 2.

Figura 2 - Raio-x de uma pessoa saudável e um caso de cardiomegalia



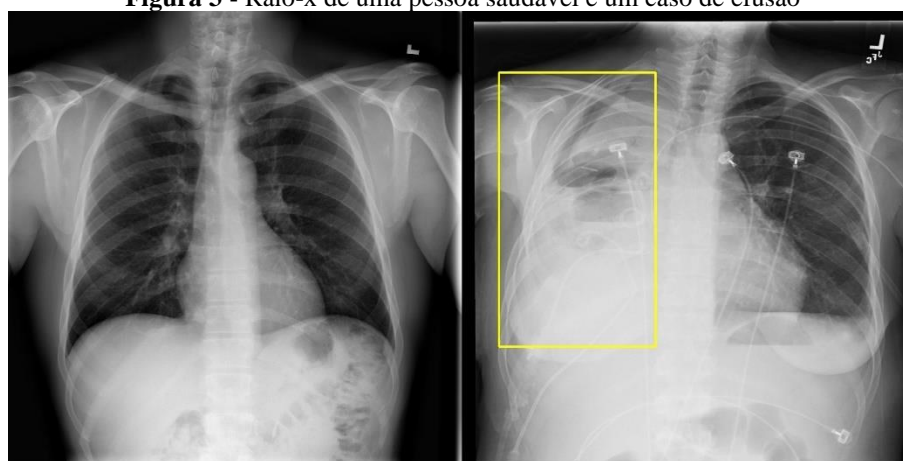
Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

2.1.3 Efusão

A efusão pleural, ou derrame pleural, é caracterizada pelo acúmulo de líquido dentro do tórax. Para Karkhanis e Joshi (2012) a apresentação clínica da efusão varia conforme a quantidade de fluido e a causa subjacente, e pode ser assintomática em muitos casos. Os sintomas possíveis incluem dor torácica pleurítica, dispneia e tosse seca. A dor torácica é causada pela inflamação da pleura parietal, exacerbada por movimentos como inspiração profunda, tosse e espirros, e pode aliviar com a imobilização do tórax ou aumento do acúmulo de fluido. Além disso, os autores afirmam que devido à inespecificidade dos sintomas, uma história clínica e exame físico cuidadosos são essenciais para o diagnóstico.

Shanthi *et al.* (2022) destaca que quando há muito líquido no corpo, o coração pode ficar tenso, assim, incapaz de bombear normalmente, o que demonstra a importância da identificação correta de tal problema, e a efusão aparece em imagens de raio-x como uma opacidade homogênea, que geralmente se aloja nas regiões basais e dorsais dos pulmões, áreas onde o fluido tende a se acumular devido à gravidade, especialmente quando o paciente está em posição ereta. A Figura 3 nos traz uma imagem de radiografia torácica, fundamental para o diagnóstico de derrame pleural, onde a efusão aparece como uma opacidade no pulmão à esquerda da imagem.

Figura 3 - Raio-x de uma pessoa saudável e um caso de efusão



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

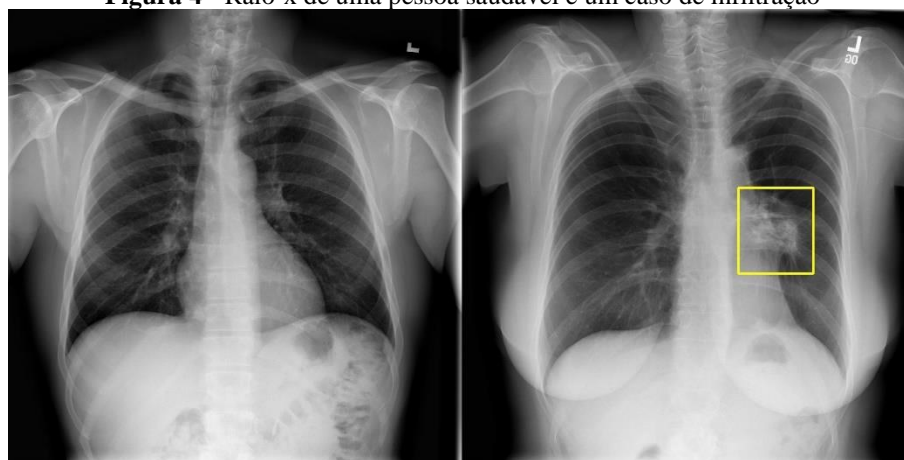
2.1.4 Infiltração

Considera-se como infiltração pulmonar quando é identificada a presença de substâncias anormais nos pulmões, como fluidos, células inflamatórias ou outros materiais, nos pulmões, que se acumulam nos espaços aéreos e nos tecidos pulmonares. Isso pode

ocorrer devido a várias condições, o que inclui infecções, doenças inflamatórias, neoplasias ou doenças pulmonares crônicas (MAIMON; ALMOG, 2003; HILDEBRAND *et al.*, 1990).

A infiltração pulmonar é frequentemente associada à atelectasia e efusão pleural devido ao impacto dessas condições na estrutura pulmonar e sua apresentação em exames de imagem. Isso acontece porque os três problemas afetam a radiografia torácica e outros exames pulmonares ao produzir áreas de opacidade e com densidade alterada. A Figura 4 nos traz um exemplo de infiltração pulmonar.

Figura 4 - Raio-x de uma pessoa saudável e um caso de infiltração



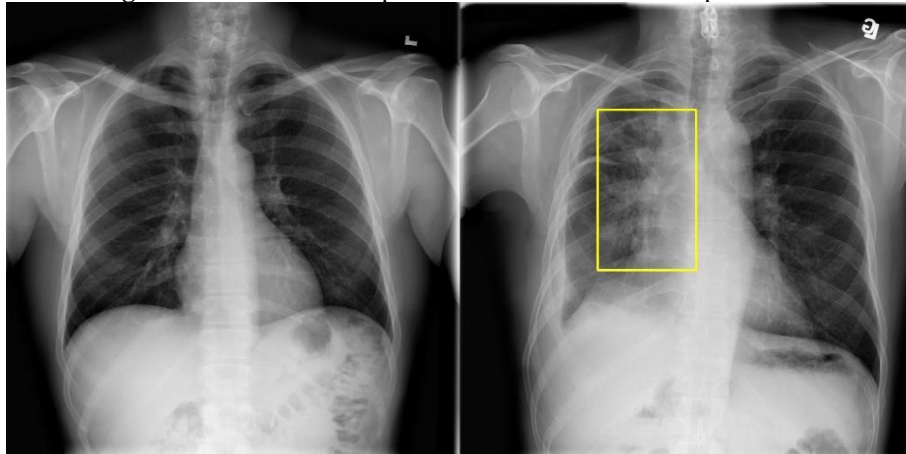
Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

2.1.5 Pneumonia

A pneumonia é uma infecção respiratória mortal e frequente que afeta predominantemente os pulmões e pode ser afetada por vários patógenos, como bactérias, vírus, fungos e outros microrganismos (DAS *et al.*, 2024). Além disso, Ibrahim *et al.* (2021) comenta que as pessoas em maior risco de pneumonia são idosas (acima de 65 anos), crianças (abaixo de 5 anos) e pessoas com outras complicações, como HIV/AIDS, diabetes, etc.

Para Jaiswal *et al.* (2019) a radiografia de tórax é um método crucial para o diagnóstico de pneumonia, mas exige conhecimento especializado para interpretar as imagens com precisão. O processo pode ser demorado e impreciso, pois outras condições médicas, como câncer de pulmão e excesso de fluidos, também podem apresentar opacidades semelhantes nas imagens. A Figura 5 nos traz um exemplo de caso de pneumonia.

Figura 5 - Raio-x de uma pessoa saudável e um caso de pneumonia



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

2.2 Redes Neurais Convolucionais (CNN)

Redes neurais convolucionais, ou *Convolutional Neural Networks* (CNN), são um tipo de rede neural especializado para tratar com processamento de dados, não à toa, vê-se seu uso com frequência para tarefas de classificação e visão computacional. O seu nome vem do fato de serem redes neurais que fazem uso de convolução, um tipo especializado de operação linear para o processamento dos dados, e como explica Zhang *et al.* (2021), redes convolucionais são simplesmente redes neurais que usam convolução no lugar de multiplicação de matrizes em pelo menos uma de suas camadas.

Além disso, segundo Lecun, Bengio e Hinton (2015), as CNNs derivam de princípios fundamentais como a invariância à translação, que garante que todas as partes de uma imagem sejam tratadas da mesma forma, e a localidade, que utiliza apenas pequenos bairros de pixels para calcular representações ocultas. Esses princípios tornam as CNNs especialmente adequadas para tarefas de visão computacional.

Compreender a evolução e as características dos principais modelos de detecção de objetos é fundamental para apreciar os avanços no campo da visão computacional. A seguir, exploraremos alguns modelos influentes na área, com uma breve introdução às suas arquiteturas, o que dará embasamento para o trabalho apresentado.

2.2.1 Faster R-CNN

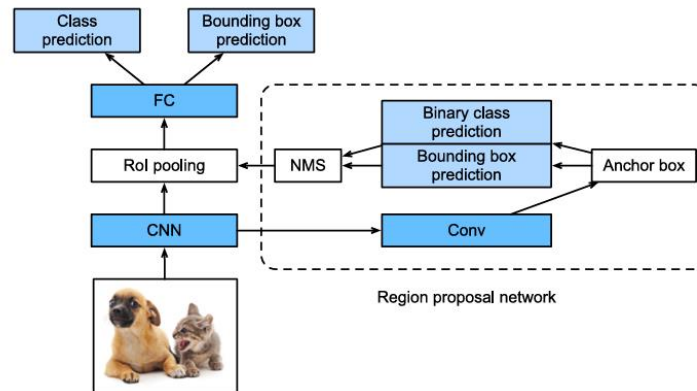
As Region-based CNNs (R-CNN), ou regiões com recursos de CNN, são pioneiros na no uso de aprendizagem profunda para detecção de objetos (ZHANG *et al.*, 2021). Este método foi proposto por Girshick *et al.* (2014), que visava a combinação de propostas de região com CNNs, onde são geradas propostas de regiões candidatas que podem conter objetos e, em seguida, é aplicada uma rede neural convolucional para classificar essas regiões e refinar as caixas delimitadoras.

Com o decorrer do tempo tal método veio a sofrer melhorias, até chegar ao Faster R-CNN, um método proposto por Ren *et al.*, em 2015. Conforme explicam Zhang *et al.*, em 2021, dado que a versão mais recente até aquele momento, Fast R-CNN, tinha que gerar muitas propostas de região em busca seletiva, o que ocasionava em um elevado custo computacional, foi proposto um método para reduzir a quantidade de propostas regionais sem perda de precisão, onde foi substituída a busca seletiva por uma rede de proposta de região (RPN), assim dando início ao Faster R-CNN (REN *et al.*, 2015).

As RPNs geram propostas de regiões diretamente das características convolucionais, o que elimina a necessidade de métodos lentos e computacionalmente caros de extração de regiões. Isso resulta em um modelo mais rápido e eficiente, ao mesmo tempo que mantém uma alta precisão na detecção de objetos em diversas escalas e contextos.

A Figura 6 traz uma breve explicação do modelo Faster R-CNN, e como explica Zhang *et al.* (2021), primeiro, uma camada de convolução com preenchimento de 1 é usada para transformar a saída CNN em uma nova saída com “C” canais. Em seguida, são geradas diversas caixas âncoras de diferentes escalas e tamanhos, centradas em cada pixel dos mapas de funcionalidade. Com o uso do vetor de características de comprimento “C” no centro de cada caixa de âncora, a classe binária (plano de fundo ou objeto) e a caixa delimitadora são previstas. Por fim, as caixas delimitadoras esperadas como objetos são refinadas com uma compressão não máxima, o que resulta nas propostas de região necessárias para a camada de pooling de interesse regional.

Figura 6 - Modelo simplificado do método Faster R-CNN

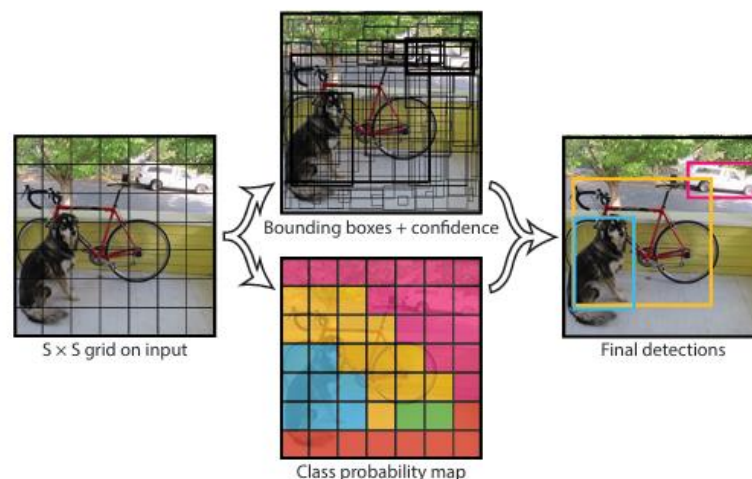


Fonte: imagem retirada do livro Deep Into Deep Learning (ZHANG, 2021, p. 670).

2.2.2 You Only Look Once (YOLO)

A arquitetura You Only Look Once (YOLO - Você Só Olha Uma Vez) foi desenvolvida por Redmon *et al.* (2016), como uma nova abordagem à detecção de objetos, onde a ideia principal por trás do YOLO é dividir uma imagem de entrada em uma grade de células e prever caixas delimitadoras e probabilidades de classe para cada célula, dessa forma a detecção de objetos é tratada como um problema de regressão (JANI *et al.*, 2023), conforme pode ser observado pela Figura 7. Além disso, ao contrário de outros modelos, o YOLO faz uso de uma única rede neural que prevê caixas delimitadoras e probabilidades de classe diretamente das imagens inteiras em uma única análise. Visto que todo o processo de detecção é executado por uma única rede, é possível otimizar todo o sistema de detecção com base no desempenho diretamente.

Figura 7 - Modelo simplificado do método YOLO



Fonte: REDMON *et al.* (2016, p. 780).

Com o passar do tempo o modelo YOLO foi sofreu evoluções e novas versões foram lançadas, até que em 2020 o Ultralytics introduziu o YOLOv5, que conforme Jani *et al.* (2023) é dividida em três partes principais: (1) espinha dorsal, que utiliza uma estrutura *CSPDarknet53* que é baseada no Darknet53 original, mas com a adição da estratégia *Cross Stage Partial* (CSP), técnica essa divide a rede em duas partes para reduzir cálculos e melhorar a eficiência, o que permite uma extração de características mais precisa e rápida; (2) pescoço, que utiliza CSP-PAN, que melhora a fusão de características, e uma versão otimizada do bloco SPP chamada SPPF, que aumenta a velocidade de processamento, e por fim; (3) cabeça, que faz uso de caixas de referência para gerar os resultados finais da detecção, como classes, pontuações e caixas delimitadoras.

Em 2023 o Ultralytics, mesmo desenvolvedor do YOLOv5, lança mais uma versão do modelo, o YOLOv8, uma versão ainda mais aprimorada, que incorpora novas funcionalidades e uma melhor otimização, o que corroborou para melhorias em performance e eficiência. Essa nova versão possui uma arquitetura onde utiliza uma versão que adota algumas modificações em relação ao YOLOv5, primeiramente a coluna dorsal *CSPDarknet53* foi modificada, ao substituir o *CSPLayer* do YOLOv5 pelo módulo *C2f*. A camada SPPF acelera o processamento ao agrupar características em um mapa fixo. As camadas de convolução possuem normalização por lote e ativação SiLU, enquanto a cabeça da rede lida separadamente com a detecção de objetos, a classificação e a regressão (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

2.2.3 RetinaNet

O método RetinaNet foi desenvolvido em 2017 por Lin *et al.*, um modelo de detecção de objetos cujo foco é a resolução de problemas associados ao desequilíbrio de classes e à detecção densa. Segundo o autor, o modelo RetinaNet, em sua estrutura, faz uso de uma CNN, geralmente baseada em ResNet, para processar imagens e criar uma série de mapas de características através de várias camadas de convolução. Esses mapas são então enviados para uma Feature Pyramid Network (FPN), que usa essas características em diferentes escalas. A partir da FPN, duas sub-redes são usadas para classificar objetos e ajustar as caixas delimitadoras.

Além disso, a principal inovação nesse modelo, que ajuda a resolver o problema de desequilíbrio de informações enfrentado por detectores gerais de primeira ordem durante o

treinamento, foi a aplicação de uma função chamada Focal Loss (perda focal), conforme Wu *et al.* (2019), que também ressalta que essa função é crucial, pois as informações de fundo geralmente são muito mais abundantes do que as informações de primeiro plano. Com o uso da Focal Loss, o modelo consegue alcançar alta precisão de detecção, ao mesmo tempo em que mantém um baixo atraso no tempo de processamento, assim é possível manter o desempenho equilibrado em situações de desequilíbrio de dados.

2.2.4 Detectron2

O Detectron2 foi desenvolvido por cientistas do *Facebook AI Research* (FAIR), como uma compilação de bibliotecas de modelos de detecção e segmentação de objetos de última geração (WU *et al.* 2019, apud MERZ *et al.* 2023). A plataforma é implementada em PyTorch e possui um grande conjunto de modelos de referência e modelos pré-treinados no conjunto de dados COCO disponíveis para download, que conforme Merz *et al.* (2022), ao aproveitar a aprendizagem por transferência, é possível reduzir o tempo e os custos de treinamento com esses modelos pré-treinados.

Segundo Ackermann *et al.* (2022), o Detectron2 é uma estrutura que inclui implementações de alta qualidade de algoritmos de detecção de objetos de última geração, como Faster R-CNN, Mask R-CNN, RetinaNet e DensePose. Com base nisso, Amerikanos e Maglogiannis (2022) descrevem o Detectron2 como uma estrutura de aprendizagem profunda de última geração, que aborda uma variedade de tarefas importantes de visão computacional, por exemplo, detecção de objetos, segmentação de instâncias, além de iterar rapidamente por meio de diferentes arquiteturas e parâmetros.

Tal plataforma foi utilizada neste trabalho, exclusivamente, em conjunto com os modelos Faster R-CNN e RetinaNet, de modo a otimizar o treinamento dos modelos e, ao mesmo tempo, garantir um ambiente de avaliação mais justo entre as duas técnicas.

2.3 Métricas de Avaliação em Modelos de Detecção de Objetos

No desenvolvimento de modelos de detecção de objetos, é fundamental utilizar métricas adequadas para medir a eficácia e a precisão dos algoritmos. Essas métricas permitem avaliar o desempenho do modelo em diferentes aspectos, como a capacidade de

localizar corretamente os objetos em uma imagem e a precisão com que as classes dos objetos são identificadas. Neste tópico serão apresentados os principais conceitos e fórmulas por trás dessas métricas, essenciais para garantir que os resultados obtidos sejam consistentes e confiáveis na aplicação clínica.

2.3.1 Interseção sobre a União

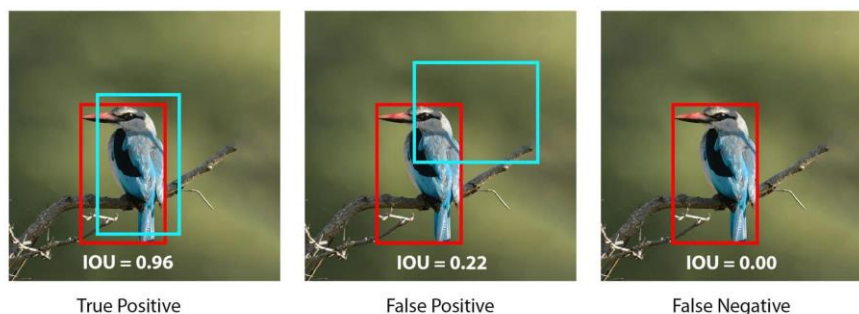
A Interseção sobre a União (IoU - *Intersection Over Union*) é uma das métricas mais utilizadas em tarefas de detecção de objetos e segmentação semântica, visto que é simples e, ao mesmo tempo, bastante poderosa. O IoU é um método utilizado para quantificar a sobreposição percentual entre a caixa delimitadora da imagem real e a previsão (MOBSITE, 2023).

A fórmula do IoU (1) é dada pela divisão da área de interseção das duas caixas pela área de união entre elas, sendo que o valor do IoU pode variar de 0 a 1 (0% a 100%), onde 0 representa a ausência total de sobreposição entre a previsão e a verdade de terreno, enquanto 1 indica que há uma correspondência perfeita, ou seja, as duas áreas estão completamente alinhadas.

$$IoU = \frac{CAIXA\ REAL \cap PREVISÃO}{CAIXA\ REAL \cup PREVISÃO} \quad (1)$$

A Figura 8 mostra um exemplo do uso do cálculo do IoU para classificar uma previsão, ao considerar um limiar de 0.5 (50%), no primeiro caso o IoU foi de 0.96 (96%), um valor muito alto e que está acima do limiar definido, sendo assim a previsão é considerada verdadeiro positivo (VP), no segundo caso vemos que a pontuação foi de apenas 0.22 (22%), o que está abaixo do limiar, mas como houve uma previsão ela é marcada como falso positivo (FP). No último caso não houve uma previsão, nesse caso marcamos como falso negativo (FN). Para essa pesquisa, como há muitas classes foi levada em consideração também o acerto da classe, desse modo a previsão deve fornecer a classe correta e uma pontuação de IoU acima do limiar que será definido.

Figura 8 - Exemplo de caso de IoU. A caixa vermelha indica a o objeto real da imagem e a caixa azul a tentativa de previsão do objeto



Fonte: site LearnOpenCV ¹ (2022).

2.3.2 Métricas de Análise Comparativa

A avaliação do desempenho de sistemas de detecção de objetos é um processo essencial para garantir a eficácia de modelos em contextos práticos. Em áreas críticas, como a saúde, onde esses sistemas podem ser aplicados, a precisão no diagnóstico ou na identificação de padrões em imagens é de extrema importância. Através de métricas de análise comparativa, é possível não apenas verificar a precisão de um modelo, mas também entender seus pontos fortes e limitações. Essas métricas fornecem informações valiosas que permitem otimizar algoritmos, melhorar a tomada de decisão automatizada e, sobretudo, aumentar a confiança na aplicação dessas tecnologias no cenário clínico.

Antes de tudo é necessário entender algumas técnicas de detecção de objetos e como é feita sua classificação, assim identificando a classe corretamente. O método mais utilizado nesses casos é a classificação binária (GRAY *et al.*, 2011), onde para cada previsão feita pelo modelo só pode uma dentre as quatro opções a seguir:

Verdadeiro Positivo (TP): Quando um dado positivo é corretamente classificado como positivo.

Verdadeiro Negativo (TN): Quando um dado negativo é corretamente classificado como negativo.

Falso Positivo (FP): Quando um dado negativo é incorretamente classificado como positivo.

¹ Disponível em: <<https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/>>. Acesso em 31 jul. 2024.

Falso Negativo (FN): Quando um dado positivo é incorretamente classificado como negativo.

Essas classificações são essenciais para poder avaliar da melhor forma os modelos, visto que são utilizadas nos principais métodos de mensuração dos modelos, como a precisão, a sensibilidade (também chamada de recall) e o F1-score. As fórmulas (2), (3) e (4), fornecidas por Gray *et al.* (2011) apresentam como foi feito o cálculo dessas métricas:

- Precisão (*Precision*): Mede a proporção de verdadeiros positivos em relação ao total de predições positivas feitas pelo modelo.

$$P = \frac{VP}{VP + FP} \quad (2)$$

- Sensibilidade (*Recall*): Mede a proporção de verdadeiros positivos em relação ao total de reais positivos presentes na base de dados.

$$R = \frac{VP}{VP + FN} \quad (3)$$

- F1-score: É a média harmônica da precisão e da sensibilidade, fornece uma única métrica que considera tanto a precisão quanto a capacidade de recuperação do modelo.

$$F1 = \frac{2 \times (P \times R)}{(P + R)} \quad (4)$$

Para o cálculo da precisão e sensibilidade foi calculado separadamente para cada classe, de modo a obter a precisão por classe, ao final foi feito um novo cálculo para obter a precisão média do modelo conforme a fórmula (5), onde é somada as precisões de cada classe e em seguida dividido pelo número total de classes.

$$\text{Precisão Média (PM)} = \frac{1}{N} \sum_{i=1}^N P_i \quad (5)$$

Outras métricas foram utilizadas na análise de treinamento dos modelos, como a Precisão Média (AP - *Average Precision*) e a Média das Precisões Médias (mAP - *Mean Average Precision*), calculadas internamente por alguns modelos. Sua principal diferença se dá no fato de serem calculadas em diferentes limiares de IoU, contudo na avaliação final dos modelos foi optado por não a utilizar para que seja possível seguir um mesmo padrão na avaliação dos modelos.

3 METODOLOGIA

Prodanov e Freitas (2013) definem a metodologia como “a aplicação de procedimentos e técnicas que devem ser observados para construção do conhecimento, com o propósito de comprovar sua validade e utilidade nos diversos âmbitos da sociedade”. Com base nisso, surge a necessidade de definir os procedimentos adotados para a pesquisa. Este tópico abrange a escolha dos métodos de coleta e análise de dados, os instrumentos e procedimentos adotados.

Ao seguir as classificações de pesquisa definidas por Moresi *et al.* (2003), a seguinte pesquisa pode ser definida como de natureza aplicada, que utiliza abordagens tanto quantitativas quanto qualitativas, visto que busca medir o desempenho das técnicas de detecção de objetos aplicadas ao problema proposto. Além disso, pode-se dizer que o trabalho utiliza o método comparativo, visto que compara os resultados obtidos de cada modelo para verificar suas semelhanças e diferenças em relação às métricas de avaliação propostas.

3.1 Método de Coleta de Dados

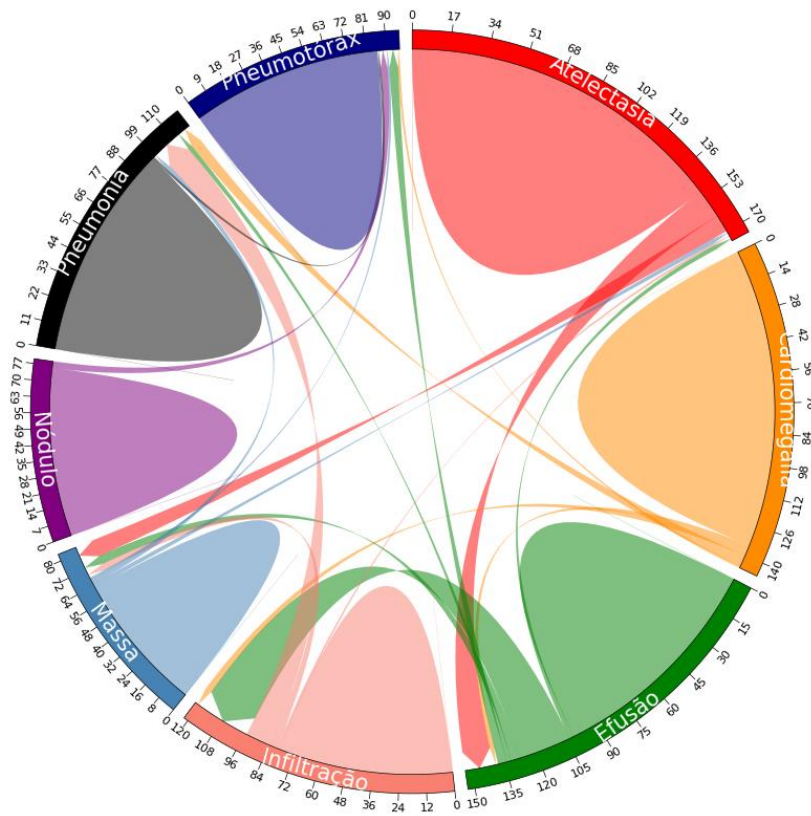
Após escolha do tema de estudo, a elaboração deste trabalho começou por definir a base de dados a ser utilizada, tendo em vista que a busca de uma base de dados de qualidade influencia diretamente nos resultados da pesquisa.

A base de dados utilizada como fonte de estudo foi a *NIH Chest X-rays* (WANG *et al.*, 2017), disponível na plataforma Kaggle, que conta com mais de 30.000 imagens, separadas em 14 classes de doenças pulmonares e respiratórias, além de um grupo de imagens sem nenhuma enfermidade. Para esta pesquisa foram utilizados apenas os exemplos que continham caixas delimitadoras, cerca de 984 imagens, categorizadas em 8 classes distintas. Além disso, deve-se lembrar que uma mesma imagem pode apresentar mais de uma patologia diferente.

Para esta pesquisa foi executado um teste preliminar que buscava verificar a quantidade de imagens por classe e a relação entre elas, que pode ser observado na Figura 9. Observa-se que Massa, Nódulo e Pneumotórax são as doenças com a menor quantidade de exemplos disponíveis, o que se mostra um problema, visto que para melhorar o desempenho dos modelos será feita uma redução na quantidade geral de imagens, de modo a obter, ao fim,

a mesma quantidade de imagens para cada classe. Essa técnica nos ajuda a evitar o desequilíbrio entre as classes, que poderia prejudicar o desempenho dos modelos.

Figura 9 - Relação Entre as Doenças Cardíacas e Pulmonares



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

Para garantir o equilíbrio entre as classes foram selecionadas as 5 categorias com o maior número de imagens:

- Cardiomegalia
- Atelectasia
- Efusão
- Infiltração
- Pneumonia

Dentre essas classes, foram selecionadas, aleatoriamente, 600 imagens (120 por classe) para treinamento e avaliação. A divisão da base de dados foi feita da seguinte forma:

- Treinamento: 80% (480 imagens)
- Validação: 10% (60 imagens)

- Teste: 10% (60 imagens)

3.2 Ambiente de Desenvolvimento

O ambiente utilizado para o treinamento dos modelos foi o próprio Kaggle, visto que a base de dados já se encontrava disponível no mesmo, além de proporcionar recursos computacionais de alto nível, como processadores potentes e acesso a GPUs, o que possibilitou a realização das atividades de processamento de dados e o treinamento de modelos de forma eficiente, ao aproveitar de uma infraestrutura em nuvem capaz de lidar com grandes volumes de dados e operações que exigem alta capacidade computacional.

O *notebook* utilizado para esse estudo contou com uma unidade da GPU Nvidia Tesla P100, acompanhada de 4 núcleos de CPU e 29 GB de memória RAM, sendo também disponibilizado um espaço em disco de 20 GB. Cada sessão de uso da GPU conta com o tempo máximo de 12 horas, sendo 30 horas de uso o máximo liberado para as GPUs.

Já para a análise de desempenho e desenvolvimento dos gráficos foi optado por utilizar o ambiente do Google Colaboratory, também chamado de Google Colab. O *notebook* utilizado não fez uso de recursos de GPU, pois não foi necessário, mas contava com uma memória RAM de 12 GB e 107 GB de espaço em disco.

3.3 Descrição dos Procedimentos

A treinamento e análise dos modelos foram realizados com o uso de um código desenvolvido no ambiente Kaggle e no Google Colab. Estes ambientes foram escolhidos pelas suas capacidades de fornecer os recursos computacionais necessários para lidar com grandes volumes de dados e pela facilidade de compartilhar e reproduzir os resultados. O código foi desenvolvido em Python, com o uso de bibliotecas como PyTorch e Detectron2 para implementar e treinar os modelos de detecção de objetos Faster R-CNN e RetinaNet, quanto ao YOLOv5 e YOLOv8, fora utilizado a biblioteca de cada modelo, já disponível para implementação em Python.

Antes de tudo é necessário registrar que, por os modelos YOLO fazerem uso de épocas para medir os ciclos durante o treinamento, enquanto os modelos Faster R-CNN e

RetinaNet fazem uso de iterações, foi optado por manter cada modelo com suas métricas padrões para melhor explicação dos mesmos na seção 4, RESULTADOS E DISCUSSÕES.

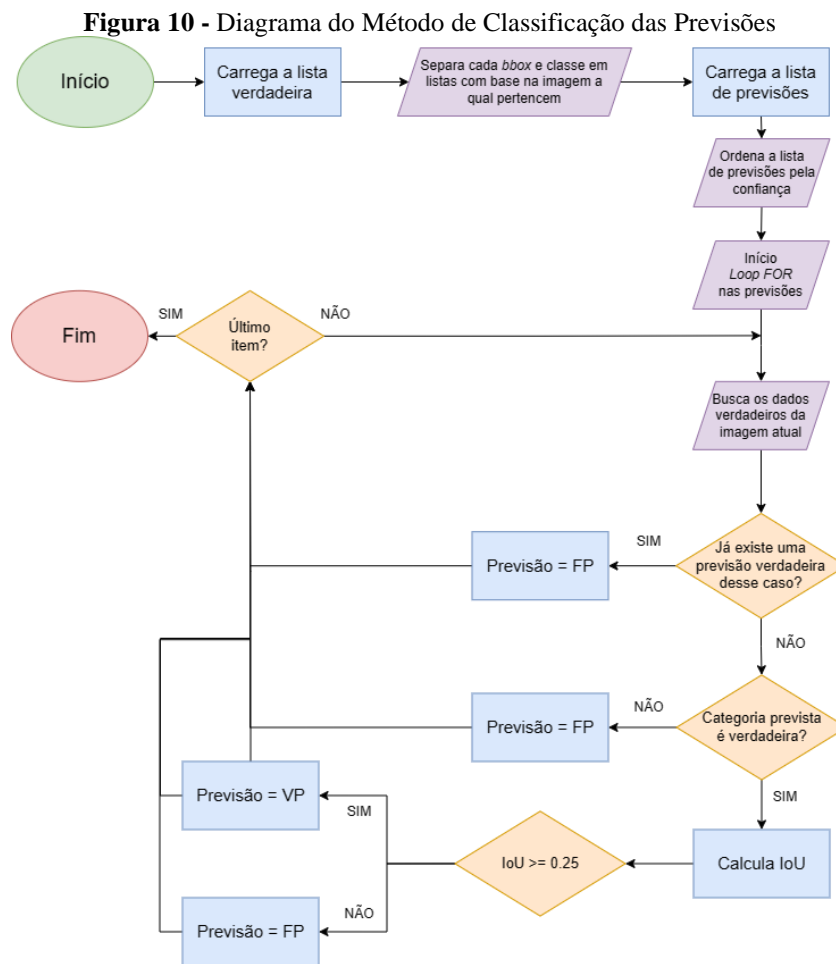
Após a escolha da base de dados, modelos e ambiente de desenvolvimento foi dado início ao treinamento e análise dos modelos. Primeiramente foi criado um código de modo a seguir o que fora citado anteriormente, ou seja, diminuir a quantidade de classes para poder obter um número maior de imagens, visto que será utilizado o mesmo número de imagens para cada classe.

Com a base de dados devidamente organizada foi em seguida necessário verificar qual a melhor versão a ser utilizada dos modelos YOLOv5 e YOLOv8, visto que ambos apresentam mais de uma versão, variando o nível de complexidade e conseqüentemente o tempo de treino. Para realizar essa avaliação foi realizado um treinamento por um intervalo de 500 épocas e com tamanho do lote de 16, ao final verificado qual obteve o melhor desempenho em termos de tempo de treino e mAP:50, métrica essa utilizada pelos modelos YOLO como mecanismo de avaliação. Para os modelos Faster R-CNN e RetinaNet não foi necessário visto que cada um apresenta apenas uma versão, estas já completas.

Após isso foi necessário delimitar as melhores métricas de treinamento para os modelos, onde mais uma vez foi utilizado o máximo de recursos do ambiente e colhido dados referentes ao tempo de treino e precisão. Em seguida foi analisado os gráficos de perda de cada modelo, de modo a identificar um possível sobreajuste (*overfitting*) durante o treinamento, que ocorre quando um modelo se ajusta tão bem aos dados de treinamento que não consegue generalizar para novos dados, o que faz com que possua um bom resultado em treinamento, mas um mal desempenho quando colocado contra os dados de teste. Tal análise ocasionou em uma melhoria no tempo de treino e precisão em geral. Assim, com as métricas definidas foi realizado o treino de cada modelo e obtido as previsões referentes aos dados de teste, dessa forma foi iniciada a análise de desempenho dos modelos.

Antes de tudo é necessário entender o funcionamento do algoritmo de classificação das previsões, ou seja, o método utilizado para identificar se as previsões foram corretas ou não. O diagrama exibido na Figura 10 resume a forma como foi ordenada a sequências de passos para classificação das previsões, onde o código desenvolvido segue uma regra em que apenas a previsão correta de maior confiança será marcada como VP, as demais, mesmo ao acertar a classe e com bom IoU serão marcados como FP, dessa forma evita-se que uma mesma imagem possua duas previsões marcadas como correta. Esse método pode diminuir

um pouco a precisão do modelo, mas servirá para uma análise justa dos modelos, o que evita que uma imagem de maior facilidade possua várias previsões corretas e distorça o resultado final.



Fonte: o autor (2024).

Com a definição desse modelo de classificação acertada, foi então necessário avaliar qual o melhor limiar de IoU e de confiança a ser utilizado, visto que um valor baixo pode gerar muitas previsões errôneas, enquanto um muito alto pode exigir demais do modelo e provocar um péssimo desempenho na tarefa de detecção. As análises para definição dos limiares foram feitas utilizando principalmente matrizes de confusão, frequência de previsão dos modelos e histogramas com as faixas de IoU que os modelos previram.

Por fim, com as métricas a serem utilizadas definidas, mais uma vez as previsões dos modelos foram avaliadas, de modo a identificar as previsões corretas e incorretas, seguindo o diagrama na Figura 10. Após isso foi realizado o cálculo das métricas de desempenho definidas, como precisão, sensibilidade e F1-score, onde foi possível fazer uma análise de cada modelo separadamente, assim obtendo respostas para os objetivos.

4 RESULTADOS E DISCUSSÕES

Nesta seção serão apresentados e avaliados os resultados obtidos ao longo do processo de treinamento e análise dos diferentes métodos de detecção de objetos propostos.

4.1 Escolha da Versão dos Modelos YOLO

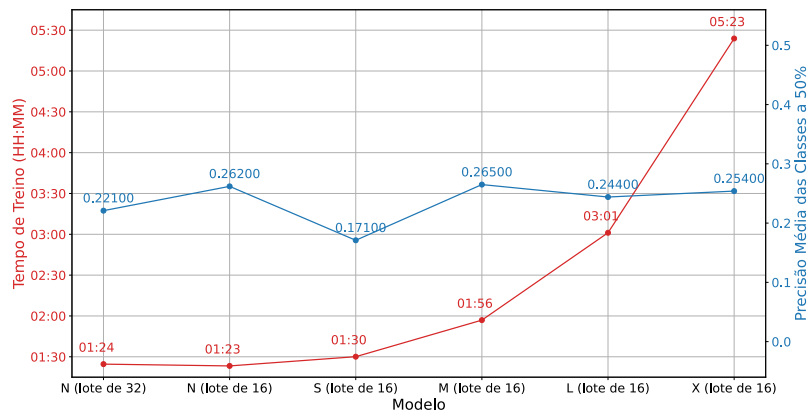
Este tópico busca apresentar as técnicas aplicadas para escolha da melhor variação de cada modelo YOLO, que serão usadas na comparação entre os modelos na seção 4.4.

4.1.1 YOLOv5

Para determinar o modelo de detecção de objetos mais eficiente, foi realizada uma série de testes comparativos com as diferentes variações do modelo YOLOv5 - N, S, M, L e X - para avaliar o desempenho em termos de precisão e tempo de execução. Cada versão foi treinada por um intervalo de 500 épocas e com tamanho do lote de 16, pelo fato desse valor ser o máximo permitido pela memória do ambiente de testes. Além disso, o modelo YOLOv5n foi treinado duas vezes, com o tamanho do lote de 16 e de 32, devido a ser menos custoso em termos de recursos operacionais.

Ao fim de cada treinamento foi anotado o tempo de execução e a média de precisão, denominado mAP, conforme mostrado na Figura 11. Este procedimento foi fundamental para avaliar o melhor custo-benefício de cada modelo

Figura 11 - Precisão Média para cada Versão do YOLOv5



Fonte: o autor (2024).

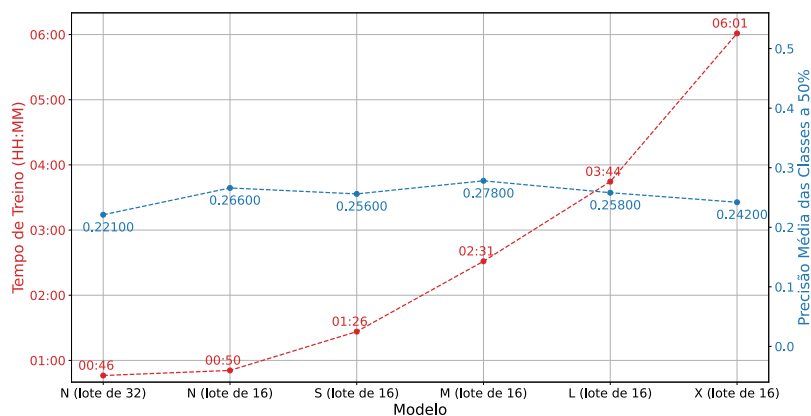
Os resultados desses testes preliminares revelaram que o modelo YOLOv5n com tamanho do lote de 16 apresentou um equilíbrio superior entre tempo de execução e mAP. Este modelo foi então selecionado para um treinamento mais intensivo e prolongado.

A estratégia de realizar um treinamento inicial com várias versões do YOLO e selecionar o modelo com melhor custo-benefício para um treinamento intensivo permitiu uma comparação aprofundada e justa com os modelos mais avançados, como YOLOv8 e os modelos Faster R-CNN e RetinaNet, implementados através do Detectron2. Esta abordagem garantiu uma avaliação completa e precisa das capacidades e limitações de cada modelo, o que forneceu *insights* valiosos para futuras implementações e pesquisas na área de detecção de objetos em imagens médicas.

4.1.2 YOLOv8

Para o modelo YOLOv8 foi utilizada a mesma aproximação aplicada aos modelos do YOLOv5, conforme descrito na seção anterior. A Figura 12 ilustra os resultados obtidos onde, assim como ocorreu com a versão 5 do YOLO, a versão que teve o melhor custo-benefício foi a versão YOLOv8n quando aplicado um tamanho de lote de 16, portanto ela foi escolhida para a realização das comparações com os demais modelos por apresentar menor tempo de execução ao mesmo tempo que mantém a média da precisão a par das demais variações.

Figura 12 - Precisão Média para cada Versão do YOLOv8



Fonte: o autor (2024).

4.2 Resultados dos Treinamentos

Avaliar o desempenho dos modelos de detecção de objetos requer uma análise minuciosa dos resultados obtidos durante o processo de treinamento. Nesta seção, apresentamos uma síntese dos resultados alcançados após os treinamentos dos modelos investigados. Esses resultados são fundamentais para compreender não apenas o potencial de cada modelo, mas também suas limitações e áreas de melhoria.

Primeiramente foi realizado o treinamento com a capacidade máxima disponível no ambiente de estudo, ou seja, a maior quantidade de épocas e iterações disponível pela memória do ambiente, de modo a permitir uma boa análise para as 12 horas de uso contínuo disponibilizada pelo ambiente de estudo. Sendo assim, o resultado acabou por ser o apresentado na Tabela 1.

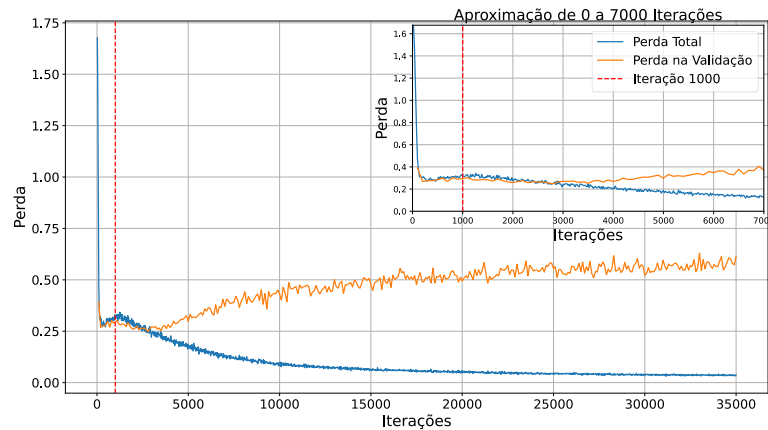
Tabela 1 - Dados de Treinamento do Primeiro Treino

Modelo	Tempo de Treino	Ciclos	Tamanho do Lote	mAP:0.50
Faster R-CNN	11:53:22	35.000 iterações	8	0,127 (12%)
RetinaNet	11:40:54	32.000 iterações	8	0,145 (14%)
Yolov5N	10:00:02	3.500 épocas	16	0,228 (22%)
Yolov8N	06:45:12	3.500 épocas	16	0,237 (23%)

Fonte: o autor (2024).

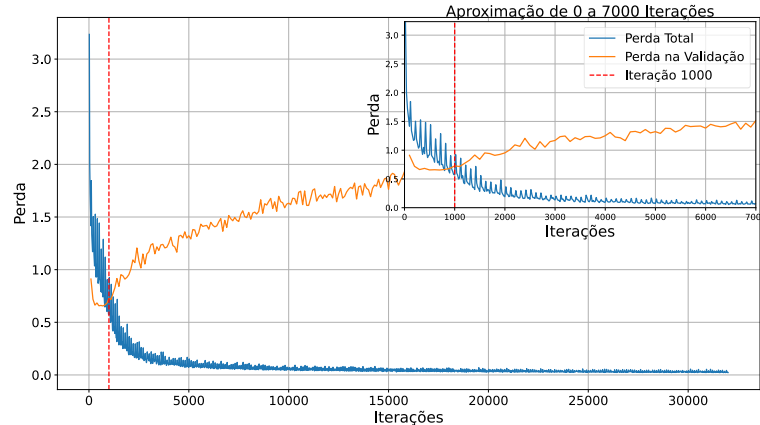
A partir da tabela, observa-se uma maior velocidade ao treinar com os modelos YOLO, em especial o YOLOv8, que por ser a versão mais recente, apresenta um melhor desempenho quando comparado aos demais. O modelo RetinaNet por sua vez não conseguiu treinar durante as 35000 iterações, devido a um alto consumo de memória e baixo desempenho. Mas para termos certeza dos resultados e evitar sobreajuste devemos visualizar também a perda durante o treinamento e a validação.

Para o modelo Faster R-CNN, o gráfico de perda na Figura 13 revela um padrão de sobreajuste significativo a partir da iteração 3000, onde se vê que por mais que a perda durante o treinamento continue a diminuir, a linha que representa a perda na validação começa a subir, o que expressa uma perda cada vez maior na validação. Este sobreajuste indica que, após este ponto, o modelo começa a se ajustar excessivamente aos dados de treinamento, e assim passa a perder a capacidade de generalizar para novos dados.

Figura 13 - Perda no Treino e Validação com Faster R-CNN

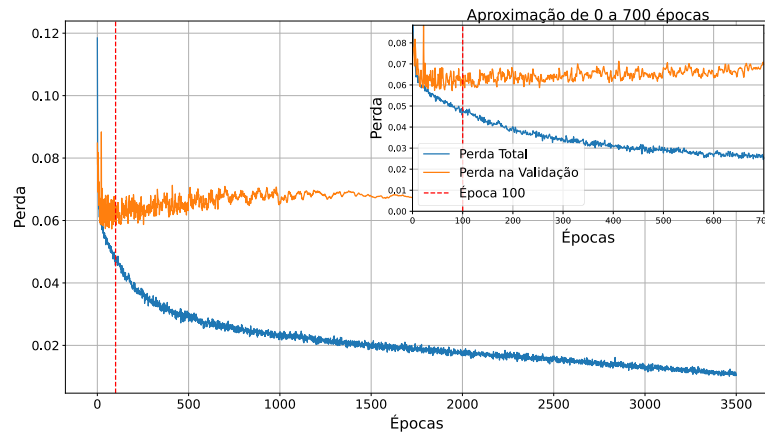
Fonte: o autor (2024).

No caso do modelo RetinaNet, a análise do gráfico de perda na Figura 14 também nos mostra que há um claro sobreajuste a partir da iteração 1000. Isso sugere que, após essa marca, o modelo começa a aprender os ruídos específicos dos dados de treinamento, o que compromete sua capacidade de generalização.

Figura 14 - Perda no Treino e Validação com RetinaNet

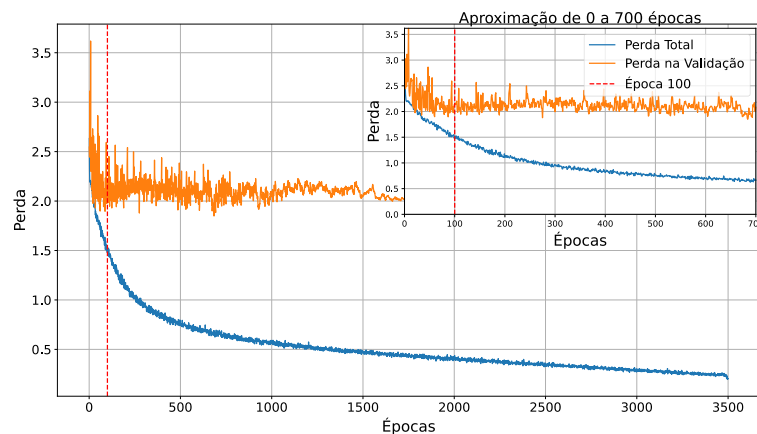
Fonte: o autor (2024).

Ao observar o gráfico de perda na Figura 15 para o modelo YOLOv5, percebe-se um ruído inicial o que indica que o modelo ainda está identificando os padrões, contudo ele sofre um pequeno aumento na perda a partir da época 100, o que pode indicar um possível sobreajuste. Contudo, ao notar que a perda se estabiliza após cerca de 1000 épocas, pode ser vantajoso reduzir o número total de épocas para otimizar o treinamento sem comprometer a performance, visto que por mais que o modelo tenha estabilizado, não houve grande alteração na perda a partir da época 100.

Figura 15 - Perda no Treino e Validação com YOLOv5N

Fonte: o autor (2024).

Para o modelo YOLOv8n, cujo gráfico de perda é exibido na Figura 16, pode-se realizar a mesma análise que a do modelo YOLOv5n, onde há um ruído inicial, mas da mesma forma que seu modelo antecessor há uma estabilização após algumas épocas de treinamento, mas ainda assim sem grande variação na perda. Haja vista a estabilização da perda durante a validação após cerca de 1000 épocas, diminuir o número total de épocas pode ser uma estratégia eficaz para evitar o sobreajuste e poupar recursos computacionais sem comprometer a performance.

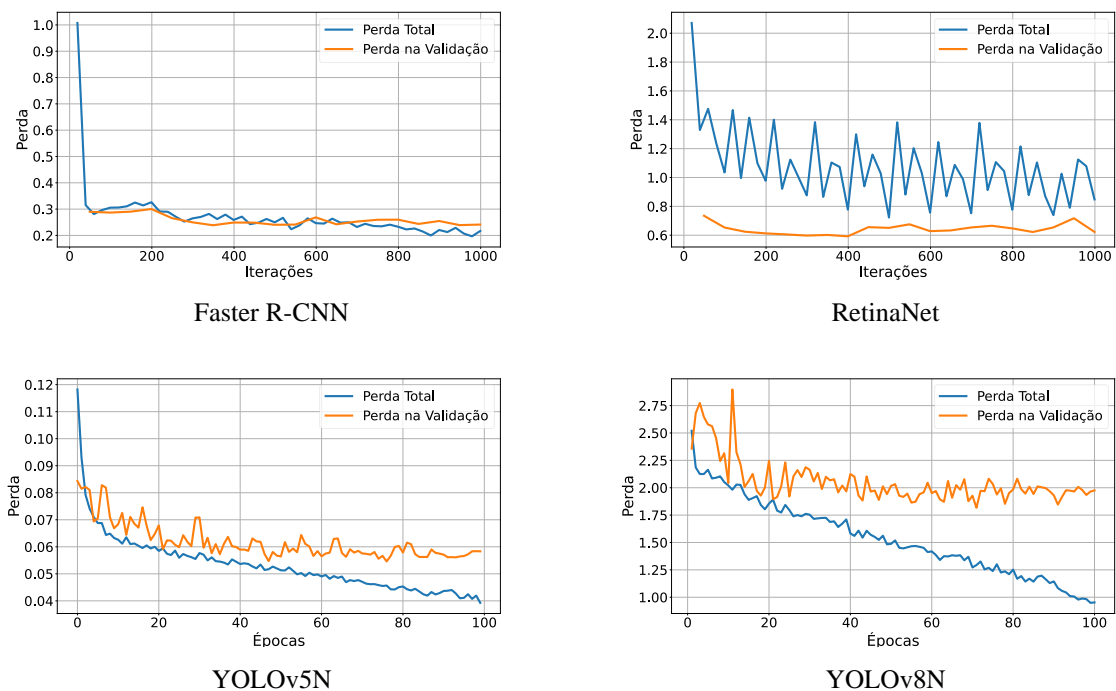
Figura 16 - Perda no Treino e Validação com YOLOv8N

Fonte: o autor (2024).

Em todos os modelos analisados - Faster R-CNN, RetinaNet, YOLOv5n e YOLOv8n - o sobreajuste a partir da iteração 1000 (ou época 100 para os modelos YOLO) indica a necessidade de considerar a redução do número de épocas. Isso pode ajudar a otimizar o treinamento e ainda manter uma boa capacidade de generalização do modelo para novos dados. Sendo assim, foi optado por limitar as épocas a 100 e as iterações a 1000, visto que a

perda não sofre grande alteração após esses ciclos. Após a mudança foi demonstrada uma melhoria para a capacidade do modelo, conforme apresentado na Figura 17, onde a partir das imagens apresentadas podemos notar, que por mais que ainda haja um pouco de ruído, a perda ainda manteve-se de certa forma estável. A opção por reduzir o Faster R-CNN para 1000 iterações, por mais que sofra sobreajuste apenas a partir de 3.000 iterações, se deu para manter a proximidade nos valores em comparação ao RetinaNet, assim como nas versões do YOLO se manteve os mesmos valores.

Figura 17 - Perda no Treino e Validação Após Ajuste nos Ciclos



Fonte: o autor (2024).

Agora, com a quantidade de ciclos definidos para cada modelo, que será utilizado durante o treinamento, foram-se obtidos os resultados apresentados na Tabela 2, onde é possível observar um leve aumento no mAP, o que se pode aumentar ainda mais com a escolha correta de métricas ao se realizar a avaliação com os dados de teste e com a definição adequada dos hiperparâmetros de cada modelo. Além disso, a diminuição no tempo de treinamento é um bom sinal no âmbito das análises rápidas, onde muitas vezes é necessário realizar um novo treinamento nos modelos. diferentes métodos de detecção de objetos propostos. A diminuição no tempo de treinamento e leve aumento no mAP também nos mostra que os modelos não precisam de muito tempo para determinar padrões nesse tipo de imagem, contudo, nem sempre apresentam uma precisão adequada devido à complexidade das mesmas.

Tabela 2 - Dados de Treinamento do Treino Final

Modelo	Tempo de Treino	Ciclos	Tamanho do Lote	mAP:0.50
Faster R-CNN	00:16:26	1000 iterações	8	0,201 (20%)
RetinaNet	00:27:59	1000 iterações	8	0,190 (19%)
YOLOv5	00:16:04	100 épocas	16	0,222 (22%)
YOLOv8	00:10:37	100 épocas	16	0,263 (26%)

Fonte: o autor (2024).

4.3 Escolha das Métricas de Avaliação

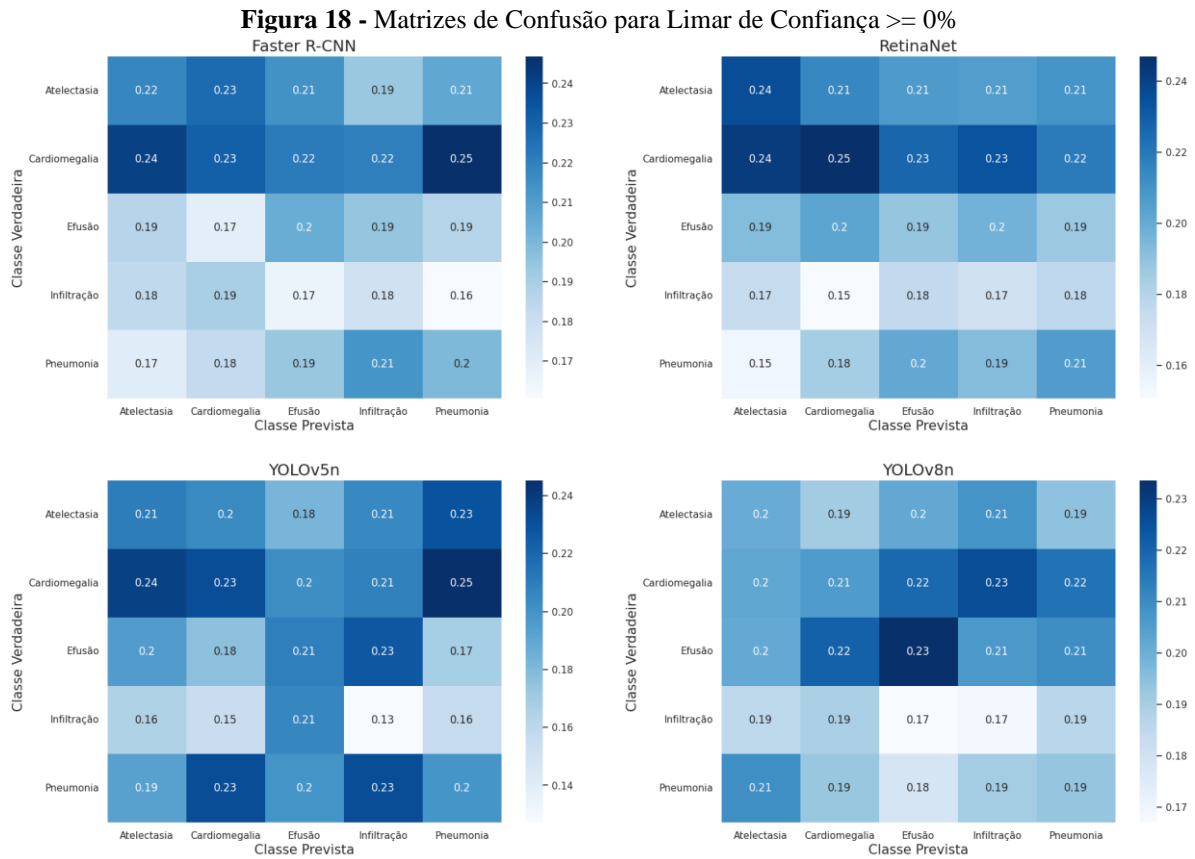
Nesta seção, apresentamos uma compilação abrangente dos resultados dos testes realizados para avaliar o desempenho e a eficácia dos modelos de detecção de objetos investigados neste estudo quando confrontados com as imagens de teste. Os resultados são discutidos em várias categorias para fornecer uma análise detalhada e abrangente das capacidades e limitações dos modelos avaliados. A seguir, apresentamos os principais tópicos abordados na compilação dos resultados do teste.

4.3.1 Definição do Limiar de Confiança

Primeiramente, antes de fazer as análises, deve-se verificar as previsões e analisar a frequência das previsões dos modelos. Dessa forma, conseguimos verificar qual seria um bom limiar de confiança para separar as previsões que analisaremos e as que não. Afinal, analisar uma previsão que possui baixo nível de confiança não faz sentido, pois essas previsões têm uma alta probabilidade de serem incorretas e podem distorcer os resultados das análises subsequentes. Ao estabelecer um limite de confiança adequado, podemos focar nas previsões mais robustas e confiáveis, o que garante que nossas conclusões sejam baseadas em dados de alta qualidade e, assim, reduz a influência de previsões incertas. É importante ressaltar que para a criação das matrizes de confusão foi levado em consideração apenas a primeira previsão correta como VP (Verdadeiro Positivo), enquanto as demais são classificadas como FP (Falso Positivo).

A Figura 18 exibe a matriz de confiança para quando o limite de confiança é de 0 (0%), ou seja, a matriz retorna a frequência considerando todas as previsões dos modelos, e como pode ser observado, as frequências tendem a ficar muito espalhadas, por isso deve-se

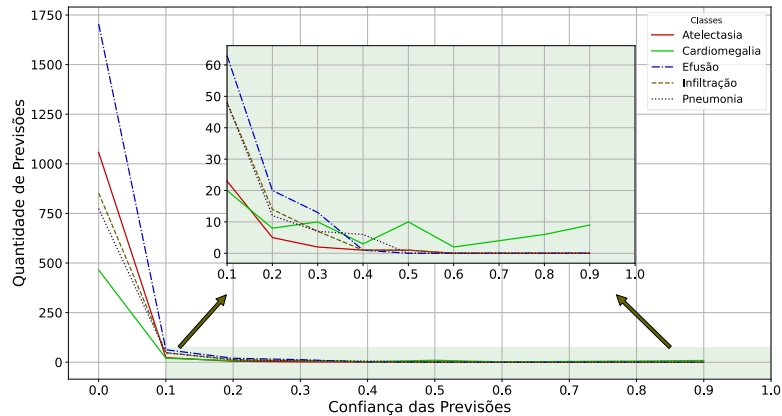
escolher um bom limite de confiança, para que assim a matriz de confusão seja mais afinada, de modo que se observe uma melhor taxa de acerto.



Fonte: o autor (2024).

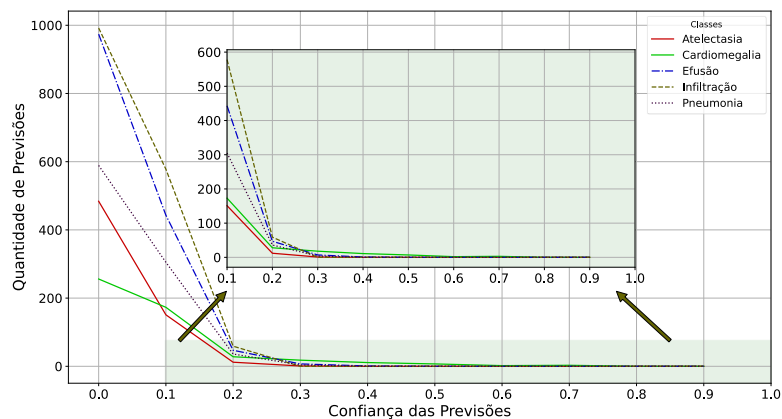
Com as figuras a seguir podemos verificar a distribuição de frequências das previsões dos modelos para cada classe e dessa forma tentar escolher um limiar de confiança que garanta uma maior “afinação” das previsões e ainda assim mantenha um bom número de previsões.

Como é apresentado na Figura 19, o modelo Faster R-CNN gerou um grande número de previsões com baixa confiança (menor ou igual a 0.1 ou 10%), o que indica uma alta sensibilidade, mas também potencial para muitos falsos positivos. A seguir a confiança estabiliza, próximo a 0.4 (40%) de confiança e, ao fim, o número maior de previsões para uma classe específica, nesse caso a cardiomegalia, reflete que o modelo passou a aprender características discriminativas fortes, e assim tende a prever essa classe com mais frequência.

Figura 19 - Frequência de Previsões por Classe para Faster R-CNN

Fonte: o autor (2024).

Já o RetinaNet, conforme a Figura 20, apresenta ainda mais previsões com baixa confiança, o que pode ser devido à sua arquitetura, o que leva a uma maior sensibilidade a objetos mais difíceis de se identificar, mas assim como no modelo anterior ele estabiliza, dessa vez mais cedo, próximo a 0.3 (30%) de confiança, o que gerou previsões mais certas e com alta confiança. O fato de o RetinaNet não variar tanto nas previsões com maior confiança pode indicar que ele tenha se adaptado melhor às características de cada classe, quando comparado ao modelo anterior.

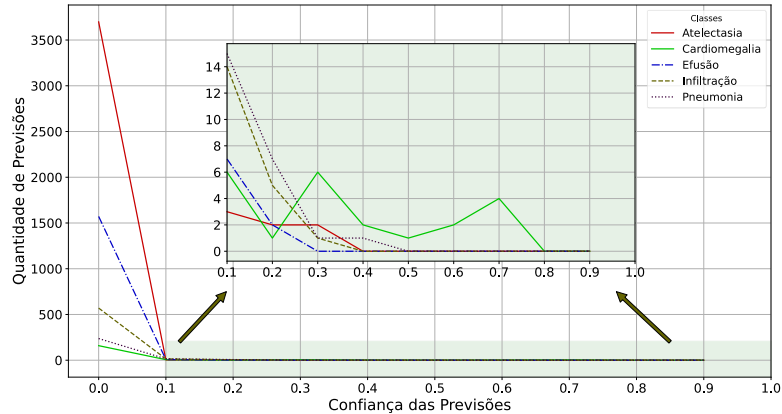
Figura 20 - Frequência de Previsões por Classe para RetinaNet

Fonte: o autor (2024).

Para o modelo YOLOv5n, conforme exibido na Figura 21, pode-se observar um comportamento similar ao Faster R-CNN, onde tende a prever para a classe cardiomegalia com mais frequência do que as demais mesmo para os níveis de confiança acima de 0.1 (10%), mas com um número geral de previsões menor quando avaliado para níveis médios de confiança (maior ou igual a 0.5 ou 50%), o que sugere ser um modelo mais conservador e

menos certo de suas previsões, então, embora o modelo consiga capturar uma ampla gama de previsões, sua confiança quanto às previsões ainda é baixa, o que pode ser uma estratégia para evitar falsos positivos em níveis mais altos de confiança.

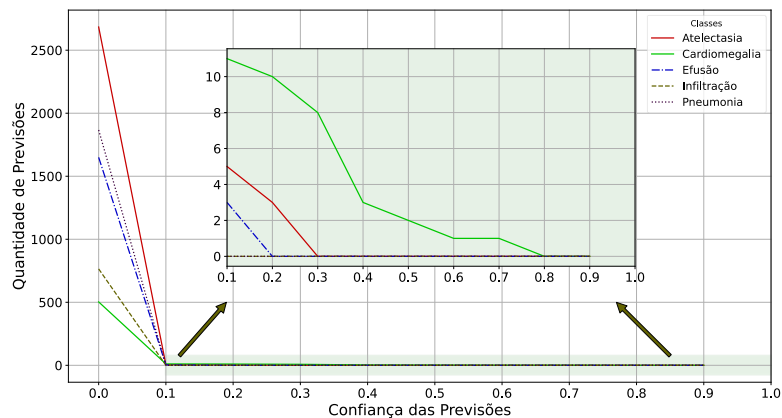
Figura 21 - Frequência de Previsões por Classe para YOLOv5N



Fonte: o autor (2024).

A respeito da Figura 22, referente ao YOLOv8n, percebe-se que ele se concentra ainda mais na classe cardiomegalia em comparação aos modelos anteriores, esse comportamento pode indicar a dificuldade do modelo em identificar objetos mais difíceis e/ou pequenos, o que causa uma pobre habilidade de generalização. Nesse sentido percebe-se que o modelo YOLOv5n está mais apto a trabalhar com imagens de raio-x quando comparado à sua versão posterior, isso pode ser devido à mudança na arquitetura do YOLOv8n, o que pode torná-lo mais apto para detecção em outros tipos de imagem e tarefas.

Figura 22 - Frequência de Previsões por Classe para YOLOv8N

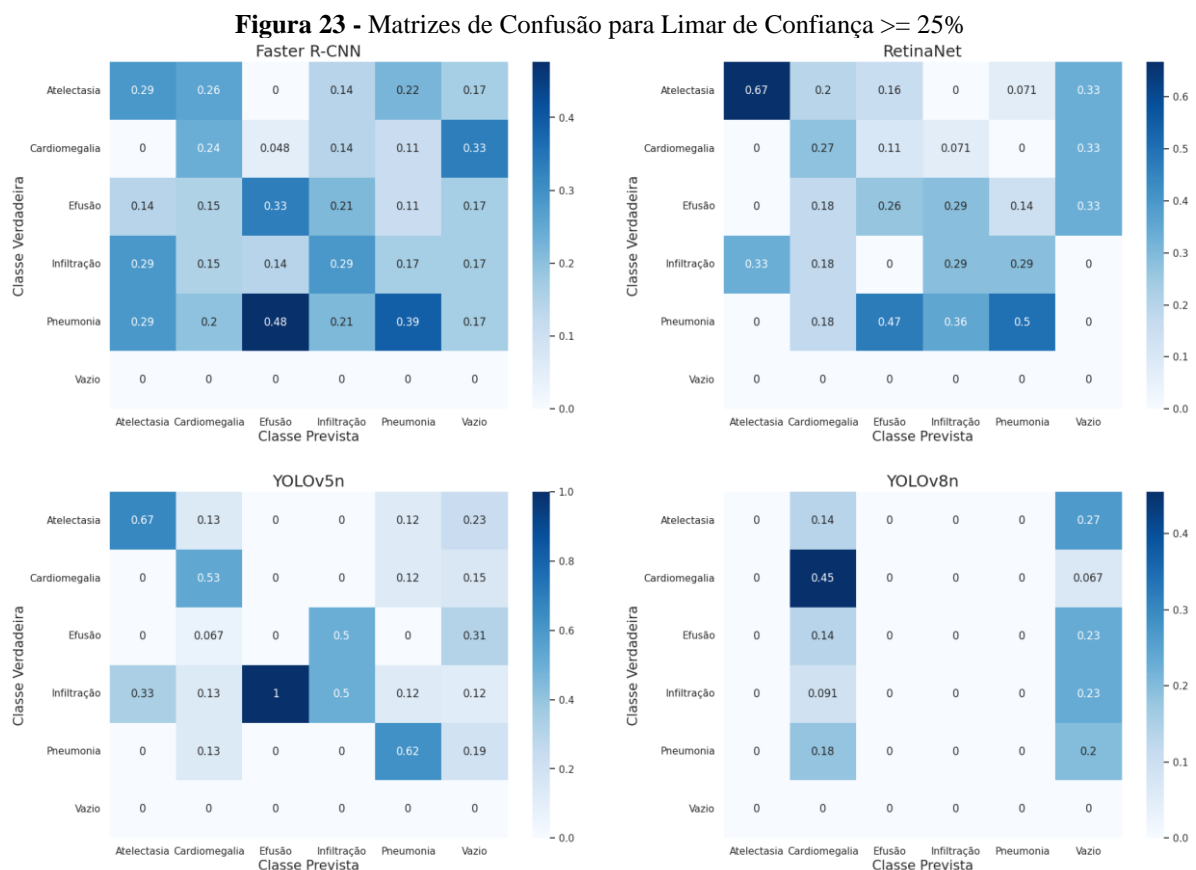


Fonte: o autor (2024).

A partir das figuras com as frequências de previsões por classe exibidas anteriormente vemos que há um padrão seguido pelas previsões, onde os níveis de confiança mais baixo (0.1

ou 10%) possuem uma maior quantidade de previsões, que posteriormente começa a estabilizar, próximo à faixa de 0.2 a 0.3 (20% a 30%), isso pode ser visto principalmente ao analisar as Figura 19 e 20.

Ao levar em conta o que fora citado, foi escolhido um valor de 0.25 (25%) para atuar como limiar de confiança das previsões, o que pode nos garantir previsões de maior certeza e ainda manter uma boa média de previsões, de modo a variar a até 10 previsões por faixa de confiança. Foi optado por utilizar esse valor de limiar pois ofereceu uma boa afinação da matriz de confusão para os modelos, além de ser um limiar já comumente utilizado em benchmarks reconhecidos, como o COCO (*Common Objects in Context*), em tarefas de detecção de objetos (Lin *et al.*, 2014). Por conseguinte, todas as previsões com confiança abaixo desse limiar foram desconsideradas durante a avaliação, o que nos deixa as matrizes de confusão salientadas na Figura 23.



Fonte: o autor (2024).

Ao analisar as matrizes de confiança sem aplicação do limiar de confiança, como visto na Figura 18, nota-se uma quantidade semelhante de previsões para cada classe, no entanto, quando o limiar de confiança é ajustado para 0,25 ou 25%, conforme a Figura 23, as matrizes, principalmente para os modelos YOLOv5n e YOLOv8n, sofreram grandes alterações.

Para o modelo YOLOv5n, a matriz sofreu uma afinação, o que gerou um resultado mais próximo do esperado, mas ainda apresentou valores zerados para algumas classes. Já para o modelo YOLOv8n, a matriz sofreu um grande revés, visto que gerou previsões apenas para uma classe, o que pode indicar uma tendência para aquela classe, causado por um grande sobreajuste. Este valor de limiar foi inadequado para o YOLOv8n, o que comprometeu a precisão e a capacidade de generalização do modelo, isso destaca a importância de definir um limiar de confiança adequado para obter resultados mais precisos e equilibrados.

Quanto aos modelos Faster R-CNN e RetinaNet, as matrizes de confusão continuaram por ter um bom número de previsões ainda assim dispersas, mas menos em comparação às matrizes mostradas na Figura 18. Ademais, essas matrizes nos mostram que os modelos podem ter começado a se especializar na identificação de alguns tipos específicos de objetos, ao invés de serem capazes de lidar igualmente bem com todas as classes, conforme descrito na análise das frequências de previsão dos modelos.

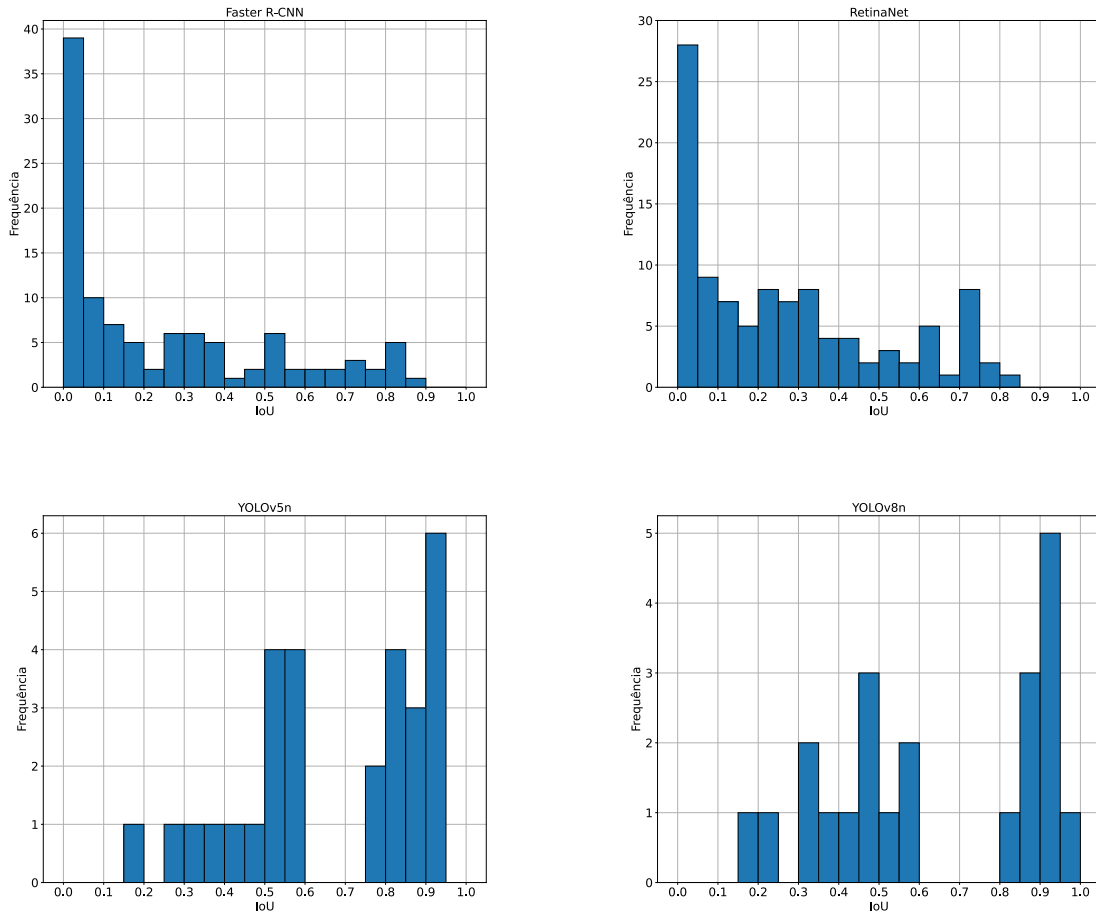
Por fim, a abordagem qualitativa nos permitiu entender melhor como cada modelo performa em relação às diferentes classes, observando que, embora um dos modelos tenha gerado previsões apenas para uma das cinco classes, os outros três modelos se adaptaram bem ao limiar proposto, ao abranger todas as classes. Essa discrepância pode ocorrer porque as previsões para as demais classes ficaram abaixo do limiar de 0.25 (25%), mesmo que tenham havido muitas previsões. Dessa forma, foi escolhido fazer uso do valor de 0.25 (25%) como atuante do limiar de confiança que será aplicado na fase de análise dos modelos, e assim gerados os resultados para cada um dos modelos.

4.3.2 Definição do Limiar de Interseção

Imagens de raio-x torácico podem ser complexas, com estruturas anatômicas sobrepostas e variações sutis de contraste. Para compensar isso devemos definir um limiar de IoU, que como apresentado na seção 3.3, nada mais é do que o cálculo da interseção sobre a união, que usaremos para validar se uma caixa delimitadora prevista é um caso VP ou FP. Apesar de serem utilizadas métricas de avaliação comparativa já comumente vistas, de 0.5 ou 50%, é necessário também verificar o melhor valor para os modelos estudados, afinal o objetivo não é apenas identificar os casos VP, mas os FP também podem ser bem-vindos no contexto de avaliação. Com isso em mente, é necessário definir um bom valor para essa medida.

Para uma boa análise da distribuição de IoU nos dados podemos utilizar Figura 24, que traz um histograma para cada faixa de IoU, com variação de 0 a 1 (0% a 100%), já com o aplique do limite de confiança definido anteriormente (0.25 ou 25%). Analisar um histograma de IoU é uma etapa crucial para definir o limiar ideal de IoU para os nossos modelos e através dele podemos averiguar a variabilidade nas predições.

Figura 24 - Histograma com Faixas de IoU para os Modelos



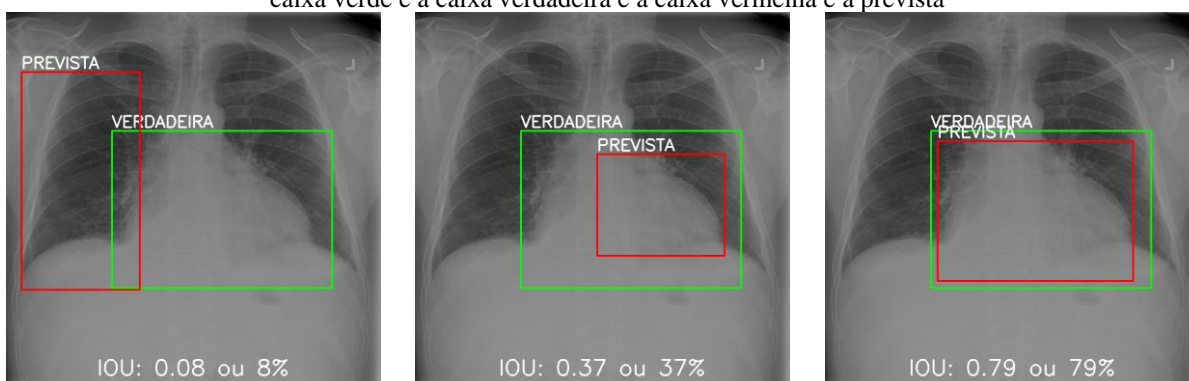
Fonte: o autor (2024).

Esse limiar mais baixo permite ao modelo ser mais permissivo, assim sendo mais tolerante a pequenas variações na posição e tamanho dos objetos detectados, e aceitar predições que podem não estar perfeitamente alinhadas com as anotações verdadeiras, mas ainda assim reconhecendo a presença dos objetos.

A partir da Figura 25, onde é exibida a mesma imagem, mas com três caixas em diferentes limiares para o IoU, nota-se que mesmo que a imagem com IoU de 0.37 (37%) possua menos precisão em comparação à de 0.79 (79%), ela ainda assim consegue identificar onde está a área de interesse, o que permite uma avaliação primária do que foi identificado, já

a de menor valor para IoU indica um caso não desejado, onde a caixa delimitadora prevista não está de acordo com a caixa real. Isso pode ser particularmente importante em contextos onde a ausência de detecção tem consequências mais graves do que a detecção imprecisa, como é no contexto clínico estudado.

Figura 25 - IoU - Exemplo: Representações de caixas delimitadoras para três limiares diferentes de IoU, onde a caixa verde é a caixa verdadeira e a caixa vermelha é a prevista



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

A partir das métricas definidas para o limiar de confiança e de IoU, podemos começar o processo de avaliação do modelo.

4.4 Avaliação e Análise Comparativa dos Modelos

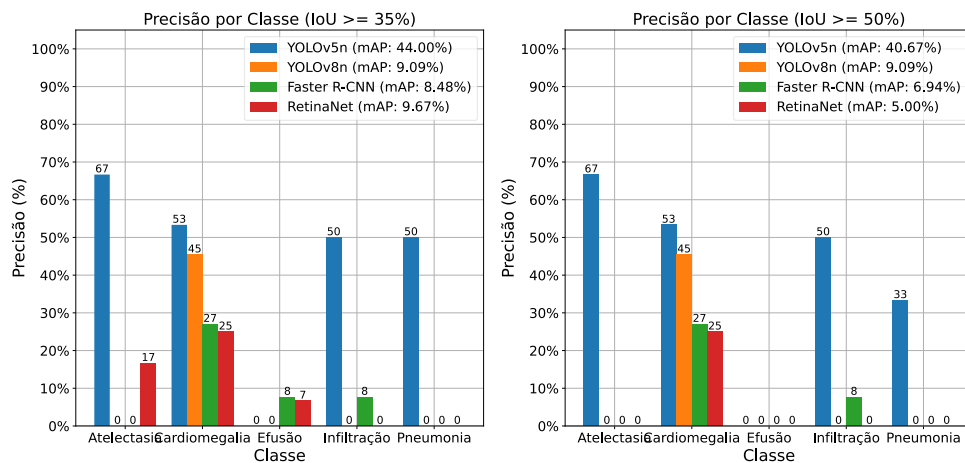
Na verificação dos casos de Falso Positivo (FP) e Verdadeiro Positivo (VP) seguem como explicado na Figura 10 da seção 3.3, mas aplicados os valores de IoU e limiar de confiança definidos nas seções anteriores. Com essas informações em mãos e as previsões geradas para as imagens de teste, a avaliação do modelo pode ser feita.

Inicialmente duas análises foram realizadas, uma para o limiar de 0.5 (50%) para verificação do IoU de uma caixa delimitadora, e em seguida para o limiar de 0.35 (35%) conforme decidido na seção 4.3.2. Para ambos os casos foram levadas em consideração tanto a precisão média quanto a sensibilidade média para avaliação dos modelos, visto que são de extrema valia quando estamos no contexto clínico.

A partir da Figura 26 nos traz uma análise da distribuição da precisão média para cada classe analisada durante a pesquisa, além do valor final de precisão média para cada modelo com base nos diferentes limiares de IoU utilizados para estudo. Para o limite de IoU maior ou igual a 35% (0.35), o YOLOv5n se destaca com uma precisão geral superior, especialmente para Cardiomegalia (53% ou 0.53) e Pneumonia (50% ou 0.5). Isto sugere que o YOLOv5n

tem um bom controle sobre as detecções falsas positivas para estas classes. No entanto, o modelo luta com classes como Efusão e Atelectasia, onde a precisão é muito baixa ou nula, o que pode indicar que o modelo frequentemente detecta esses casos incorretamente. O YOLOv8n apresenta precisão nula em muitas classes, o que reflete dificuldades em identificar corretamente as detecções. Similarmente, o Faster R-CNN mostra precisão nula para algumas classes e baixa para outras, isso indica problemas consistentes de desempenho em várias categorias. O RetinaNet também exibe uma precisão baixa para a maioria das classes, exceto Cardiomegalia, o que sugere uma dificuldade em manter uma boa precisão em um contexto geral.

Figura 26 - Precisão Média para Dois Diferentes Limiares de IoU



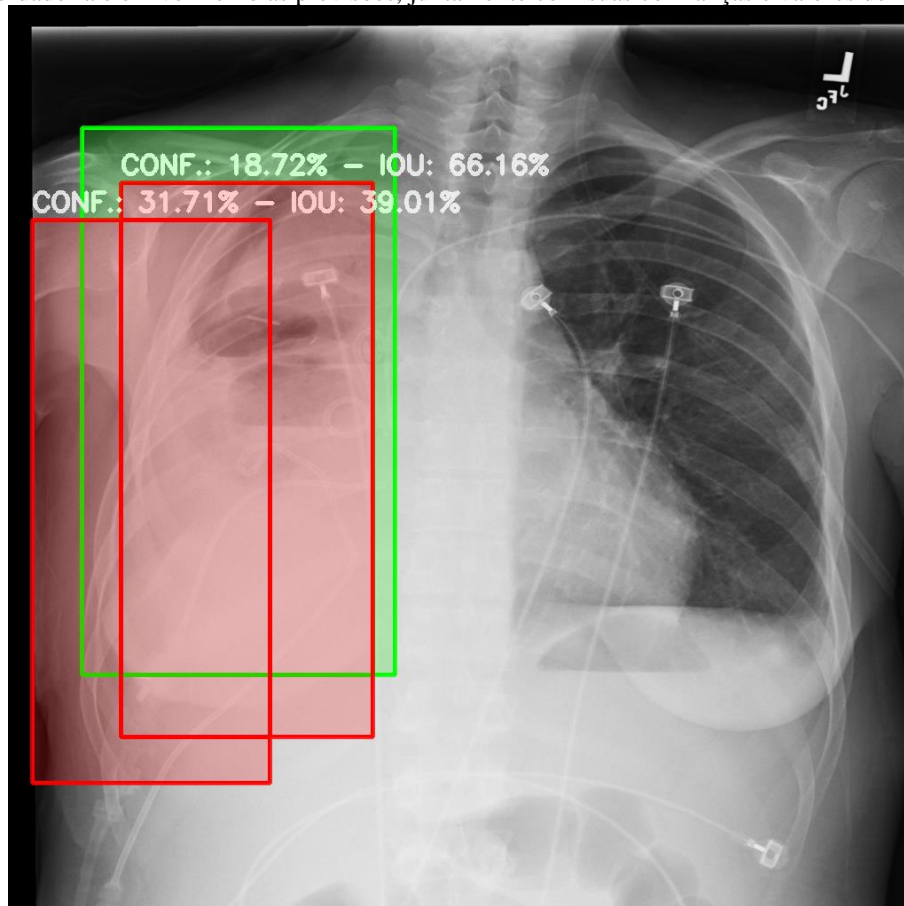
Fonte: o autor (2024).

Ao avaliar as precisões médias para IoU maior ou igual a 50% (0.5) há uma diminuição notável na precisão geral para todos os modelos. O YOLOv5n ainda apresenta a maior precisão para Cardiomegalia (53% ou 0.53), mas as precisões em outras classes são drasticamente reduzidas. O YOLOv8n continua com precisão nula em quase todas as classes, e o desempenho do Faster R-CNN e RetinaNet também é bastante reduzido. Esse comportamento pode indicar que, ao aumentar o limiar de IoU, os modelos têm dificuldades adicionais em manter uma alta precisão, possivelmente devido a uma maior rigidez na definição do que constitui uma detecção positiva.

A diminuição na precisão ao aumentar o IoU sugere que o modelo consegue prever as classes dos objetos, porém as caixas delimitadoras previstas são um pouco menos precisas nesse cenário, já que o modelo será mais permissivo ao avaliar a sobreposição entre a caixa prevista e a caixa real.

A Figura 27 nos mostra que, por vezes, mesmo uma previsão que possua uma grande compatibilidade com o objeto analisado pode possuir uma baixa confiança por parte do modelo, como é o caso da primeira previsão, com uma sobreposição de 66% (0.66) e uma confiança de apenas 18% (0.18), o que pode acontecer em casos onde a região detectada é semelhante ao objeto, mas existem muitos outros elementos na imagem que podem causar confusão, o que nesse caso poderia ser a tonalidade das imagens, visto que apresentam apenas tons de preto, branco e cinza, o que torna a identificação difícil. Por outro lado, a segunda previsão apresenta um valor de IoU menor, 39% (0.39), ainda assim sua confiança naquela previsão ainda é acima da margem de corte, 31% (0.31). Isso pode ocorrer quando a caixa delimitadora está parcialmente sobreposta ao objeto verdadeiro, mas não está perfeitamente alinhada com ele.

Figura 27 - Comparação de Duas Previsões para um Raio-X – Em verde é possível ver a caixa delimitadora verdadeira e em vermelho as previsões, juntamente com suas confianças e valores de IoU.



Fonte: elaborado pelo autor a partir da base de dados *NIH Chest X-rays* (2024).

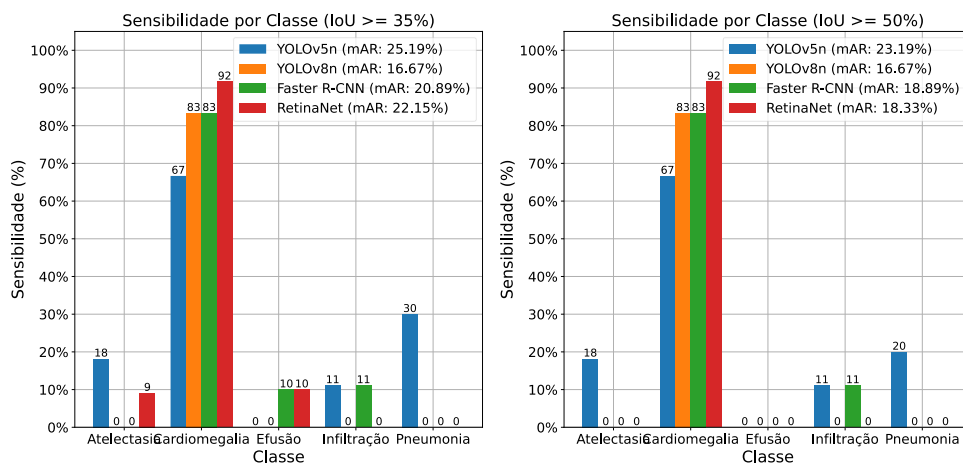
No contexto deste estudo, consideramos como verdadeiros positivos (VP) apenas uma previsão correta para cada caso, a de maior confiança. Essa abordagem pode levar à exclusão de casos que poderiam ser considerados verdadeiros, o que potencialmente resulta em uma

diminuição da precisão do modelo. É importante reconhecer que, ao adotar esse critério, podemos inadvertidamente ignorar situações em que a classe é corretamente identificada, mas a caixa delimitadora associada não é tão precisa.

Ao aplicar essas observações ao cenário clínico, surge a discussão sobre a importância de identificar a presença de uma condição médica em comparação à precisão detalhada na primeira análise. A identificação correta da condição é crucial para decisões clínicas imediatas, o que torna essencial avaliar a capacidade do modelo de detectar todos os objetos.

Como pode ser observado na Figura 28, para IoU maior ou igual a 35% (0.35), os valores de recall mostram que o YOLOv5n tem um desempenho relativamente bom em detectar a maioria dos objetos de Cardiomegalia (66% ou 0.66) e Pneumonia (30% ou 0.3). Isso sugere que o modelo é eficaz em identificar objetos dessas classes. No entanto, há uma performance fraca para Atelectasia, onde a sensibilidade é baixa, o que indica que muitos objetos dessa classe não são detectados. O YOLOv8n exibe uma sensibilidade muito alta para Cardiomegalia (83% ou 0.83) mas falha em identificar corretamente outras classes. O Faster R-CNN e o RetinaNet também têm uma alta sensibilidade para Cardiomegalia, mas apresentam resultados variados para outras classes, então embora possam detectar alguns objetos, eles não são consistentes em detectar todas as instâncias relevantes.

Figura 28 - Sensibilidade Média para dois Diferentes Limiares de IoU



Fonte: o autor (2024).

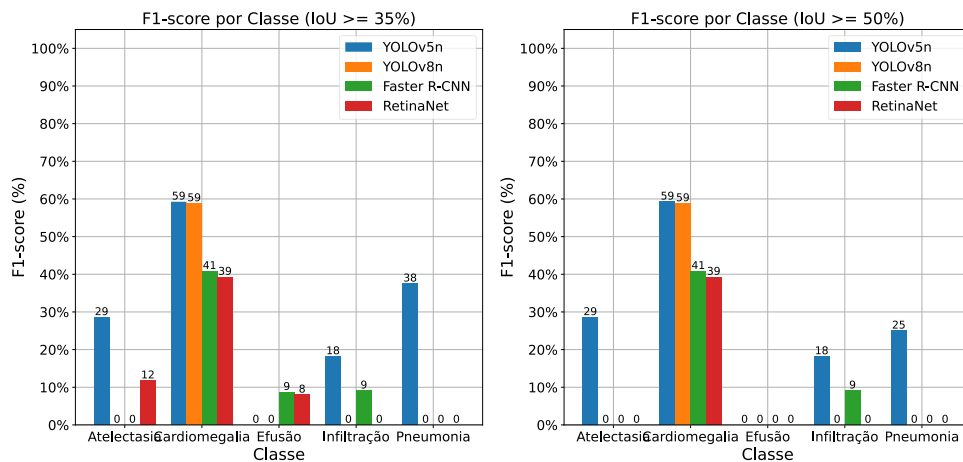
Para IoU maior ou igual a 50% (0.5), a sensibilidade para todos os modelos mostra uma tendência de queda. O YOLOv5n ainda apresenta uma sensibilidade relativamente boa para Cardiomegalia (66% ou 0.66), mas diminui para outras classes. O YOLOv8n e o RetinaNet continuam a ter dificuldades em detectar a maioria das classes, exceto

Cardiomegalia. Isso indica que, ao aumentar o limiar de IoU, todos os modelos lutam para manter a sensibilidade, o que pode ser devido à necessidade de uma correspondência mais precisa para que as detecções sejam consideradas válidas.

O F1-score combina precisão e sensibilidade em uma única métrica, o que oferece uma visão equilibrada do desempenho do modelo, especialmente útil para balancear precisão e detecção de verdadeiros positivos.

Através da Figura 29, que apresenta o F1-score dos modelos, podemos analisar que em ambos os limiares de IoU o YOLOv5n se destaca, pois possui pontuação para todas as classes, exceto Efusão. Além disso, conforme analisado para precisão e sensibilidade, os modelos apresentam uma maior estabilidade quando olhamos para a classe Cardiomegalia, principalmente para os modelos YOLO, ambos com 59%, contudo, ao avaliarmos para o Faster R-CNN e RetinaNet notamos um grande vácuo na pontuação desses modelos, isso se deve principalmente devido ao desempenho dos modelos em identificar as demais classes.

Figura 29 - F1-score para Dois Diferentes Limiares de IoU



Fonte: o autor (2024).

Algo a se analisar também é que os modelos tiveram uma maior facilidade em prever objetos para a classe Cardiomegalia, sendo assim, mais sensíveis nesses casos, além de também, em alguns casos, conseguirem boas previsões para pneumonia, enquanto para atelectasia as métricas retornaram valores muito baixos, o que indica sua dificuldade em ser observada a partir de modelos de detecção de objetos.

5 CONCLUSÃO E TRABALHOS FUTUROS

No trabalho apresentado foi realizada uma comparação entre os modelos YOLOv5 e YOLOv8, além de Faster R-CNN e RetinaNet com o uso do Detectron2, aplicados a uma base de dados de imagens de raio-x torácicos, com o objetivo de analisar qual seria a melhor abordagem em um contexto clínico. As apurações demonstraram que o modelo YOLOv5n apresentou os melhores resultados, com um desempenho relativamente equilibrado, além de conseguir realizar previsões para todas as classes, com exceção de Efusão. O YOLOv8, embora tenha se destacado em Cardiomegalia, teve um desempenho ruim em outras classes. O Faster R-CNN, apesar de ter uma sensibilidade elevada, apresentou uma precisão geralmente baixa, o que indica uma maior taxa de previsões incorretas. O RetinaNet conseguiu manter um desempenho relativamente equilibrado, mas ainda enfrentou dificuldades em precisar corretamente algumas classes específicas. De maneira geral, todos os modelos demonstraram uma tendência maior para identificar a classe Cardiomegalia, enquanto a classe Atelectasia foi a mais difícil de detectar com precisão.

Além disso, demais perguntas feitas na definição dos objetivos gerais foram respondidas, onde foi identificado que dentre os modelos YOLOv5 e YOLOv8 a melhor versão em ambos os casos foi a N quando utilizada um tamanho de lote de 16, o que é uma surpresa, pois era esperado que a versão X, por ser mais complexa, apresentaria o melhor resultado.

Outro ponto de destaque foi a possibilidade de avaliar, nesse contexto, quais as melhores métricas de treino para os modelos, que mostraram que um grande período de treino pode deixá-los suscetíveis a desenvolverem tendências para classes mais fáceis. Da mesma forma, a definição de bons limiares de IoU (35% ou 0.35) e de confiança (25% ou (0.25) pode muitas vezes ajudar um modelo a desempenhar melhor sem comprometer sua precisão, enquanto limiares mais altos podem garantir previsões mais corretas, mas tendem a diminuir a precisão média.

Tal variação no desempenho dos modelos pode estar ligado, principalmente, ao fato de imagens apresentarem baixa variação de contraste e natureza monocromática, o que se mostrou um desafio para os modelos, já que imagens em tons de cinza tendem a limitar informações de textura e cor. Essa limitação quanto às cores das imagens fez com que os modelos não fossem capazes de identificar padrões e terem dificuldade em generalização, como consequência houve um aumento na ocorrência de falsos negativos e falsos positivos.

Este estudo destaca a importância de ajustar os parâmetros dos modelos e considerar diferentes métricas de avaliação ao comparar modelos de detecção de objetos. A análise de desempenho por classe também é crucial para identificar quais classes são mais desafiadoras para cada modelo e ajustar as estratégias de treinamento e detecção conforme necessários.

Além disso, em um contexto clínico como o apresentado nesse trabalho, é necessário avaliar sua aplicação em um ambiente real, afinal a depender do tipo de doença um falso positivo pode causar pânico desnecessário em pessoa mais frágeis, enquanto falsos negativos geram uma sensação de falsa segurança, sendo assim o uso de tais modelos é indicado apenas como análise prévia, sem ser considerada como verdade absoluta, então deve-se considerar friamente a implementação ou não de diferentes modelos em diferente áreas da saúde e do cotidiano público.

Para trabalhos futuros pode-se realizar um incremento nos dados de treino para uma melhor variação no treinamento e teste ou alterações nos hiperparâmetros dos modelos. A aplicação de diferentes técnicas também pode ser feita, como mudança da escala de cores, para facilitar a identificação de objetos mais difíceis. Além disso, durante a pesquisa para esse trabalho foi optado por utilizar o YOLOv8 por ser a versão mais recente, mas no decorrer de sua escrita novas versões foram lançadas, o que pode se mostrar interessante, ao buscar avaliar as capacidades das novas versões nesse tipo de teste.

REFERÊNCIAS

- ACKERMANN, Marc *et al.* Automated segmentation of martensite-austenite islands in bainitic steel. **Materials Characterization**, v. 191, p. 112091, 2022.
- ALGHAMDI, Salem Saeed *et al.* Study of cardiomegaly using chest x-ray. **Journal of Radiation Research and Applied Sciences**, v. 13, n. 1, p. 460-467, 2020.
- AMERIKANOS, Paris; MAGLOGIANNIS, Ilias. Image analysis in digital pathology utilizing machine learning and deep neural networks. **Journal of Personalized Medicine**, v. 12, n. 9, p. 1444, 2022.
- AMIN, H.; SIDDIQUI, W.J. *Cardiomegaly*. In: **StatPearls [Internet]**. Treasure Island, FL: StatPearls Publishing, 2021. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK542296/>. Acesso em: 11 ago. 2024.
- ANAS, Muhammad *et al.* Advancing Breast Cancer Detection: Enhancing YOLOv5 Network for Accurate Classification in Mammogram Images. **IEEE Access**, 2024.
- ANGELIM, Fabio Pereira *et al.* Manifestações clínicas e complicações da Síndrome Pós-COVID-19 em indivíduos adultos. **Revista Eletrônica Acervo Saúde**, v. 24, n. 4, p. e15247-e15247, 2024.
- AVANZATO, Roberta *et al.* Lung-DT: an AI-powered digital twin framework for thoracic health monitoring and diagnosis. **Sensors**, v. 24, n. 3, p. 958, 2024.
- BHOJANE, Rhugved *et al.* Liver Lesion Detection from MR T1 In-Phase and Out-Phase Fused Images and CT Images Using YOLOv8. In: **International Conference on Data Science and Applications**. Singapore: Springer Nature Singapore, 2023. p. 121-135.
- CARVALHO, Tales de *et al.* Diretriz Brasileira de Reabilitação Cardiovascular 2020. **Arquivos Brasileiros de Cardiologia**, v. 114, p. 943-987, 2020.
- DAOUD, Mohammad I. *et al.* Detecting the Regions-of-Interest that Enclose the Tumors in Breast Ultrasound Images Using the RetinaNet Model. In: **2023 10th International Conference on Electrical and Electronics Engineering (ICEEE)**. IEEE, 2023. p. 36-40.
- DAS, Ruchika *et al.* Deep Learning Techniques for Identification of Pneumonia: A CNN Approach. In: **2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)**. IEEE, 2024. p. 1-5.
- DING, Lei *et al.* Artificial intelligence system of faster region-based convolutional neural network surpassing senior radiologists in evaluation of metastatic lymph nodes of rectal cancer. **Chinese Medical Journal**, v. 132, n. 04, p. 379-387, 2019.
- DOMINGUEZ, Mariana Chiaradia; ALVARES, Beatriz Regina. Pulmonary atelectasis in newborns with clinically treatable diseases who are on mechanical ventilation: clinical and radiological aspects. **Radiologia Brasileira**, v. 51, n. 1, p. 20-25, 2018.

- DURASAMY, Premkumar; ARUNKUMAR, D.; NIRANJANI, V. Enhancing COVID-19 Detection with YOLOv5 Object Detection Algorithm in Chest X-ray Images. *In: 2024 3rd International Conference for Innovation in Technology (INOCON)*. IEEE, 2024. p. 1-6.
- GIRSHICK, Ross *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. p. 580-587.
- GRAY, David *et al.* Further thoughts on precision. *In: 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*. IET, 2011. p. 129-133.
- HILDEBRAND, Fredric L. *et al.* Pulmonary complications of leukemia. *Chest*, v. 98, n. 5, p. 1233-1239, 1990.
- IBRAHIM, Abdullahi Umar *et al.* Pneumonia classification using deep learning from chest X-ray images during COVID-19. *Cognitive computation*, p. 1-13, 2021.
- JANI, Mohammad *et al.* Model compression methods for YOLOv5: A review. *arXiv preprint arXiv:2307.11904*, 2023.
- JAISWAL, Amit Kumar *et al.* Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*, v. 145, p. 511-518, 2019.
- KARKHANIS, Vinaya S.; JOSHI, Jyotsna M. Pleural effusion: diagnosis, treatment, and management. *Open access emergency medicine*, p. 31-52, 2012.
- LEAL, Lisiane Freitas *et al.* Epidemiology and burden of chronic respiratory diseases in Brazil from 1990 to 2017: analysis for the Global Burden of Disease 2017 Study. *Revista Brasileira de Epidemiologia*, v. 23, p. e200031, 2020.
- LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *nature*, v. 521, n. 7553, p. 436-444, 2015.
- LIN, Tsung-Yi *et al.* Focal loss for dense object detection. *In: Proceedings of the IEEE international conference on computer vision*. 2017. p. 2980-2988.
- LIN, Tsung-Yi *et al.* Microsoft coco: Common objects in context. *In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer International Publishing, 2014. p. 740-755.
- MACARI, Douglas Binotto. Prevalência de fatores de risco cardiovasculares em ingressantes de um programa de reabilitação cardíaca. *In: Repositório Institucional da UCS*. Universidade de Caxias do Sul, 2021. Disponível em: <https://repositorio.ucs.br/11338/9913>. Acesso em: 11 ago. 2024.
- MAIMON, Nimrod; ALMOG, Yaniv. Pulmonary infiltrates in the immunocompromised host. *IMAJ-RAMAT GAN-*, v. 5, n. 2, p. 112-115, 2003.
- MERZ, Grant *et al.* Detection, instance segmentation, and classification for astronomical surveys with deep learning (DEEPPDISC): DETECTRON2 implementation and demonstration

with Hyper Suprime-Cam data. **Monthly Notices of the Royal Astronomical Society**, v. 526, n. 1, p. 1122-1137, 2023.

MOBSITE, Sara *et al.* Semantic segmentation-based system for fall detection and post-fall posture classification. **Engineering Applications of Artificial Intelligence**, v. 117, p. 105616, 2023.

MORESI, Eduardo *et al.* Metodologia da pesquisa. **Brasília: Universidade Católica de Brasília**, v. 108, n. 24, 2003.

PRADO, Ana Cecilia de Sousa *et al.* ANÁLISE DAS ETIOLOGIAS E COMPLICAÇÕES DA ATELECTASIA. **Revista Contemporânea**, v. 4, n. 3, p. e3665-e3665, 2024.

PRODANOV, Cleber Cristiano; DE FREITAS, Ernani Cesar. Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição. **Editores Feevale**, 2013. *E-book*. Disponível em: <https://www.feevale.br/institucional/editora-feevale/metodologia-do-trabalho-cientifico---2-edicao>. Acesso em: 27 jul. 2024.

REDMON, Joseph *et al.* You only look once: Unified, real-time object detection. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 779-788.

REN, Shaoqing *et al.* Faster r-cnn: Towards real-time object detection with region proposal networks. **Advances in neural information processing systems**, v. 28, 2015.

SEREJO, Livia Goreth Galvão *et al.* Risk factors for pulmonary complications after emergency abdominal surgery. **Respiratory medicine**, v. 101, n. 4, p. 808-813, 2007.

SHANTHI, K. G. *et al.* Atelectasis, Cardiomegaly and Effusion Disease Prediction using Image Processing (Convolutional Neural Networks). *In: 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2022. p. 1470-1473.

TERVEN, Juan; CORDOVA-ESPARZA, Diana-Margarita; ROMERO-GONZÁLEZ, Julio-Alejandro. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. **Machine Learning and Knowledge Extraction**, v. 5, n. 4, p. 1680-1716, 2023.

WANG, Xiaosong *et al.* Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. p. 2097-2106.

WU, Wenhao *et al.* RetinaNet-based visual inspection of flexible materials. *In: 2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*. IEEE, 2019. p. 432-435.

ZHANG, Aston *et al.* Dive into deep learning. **arXiv preprint arXiv:2106.11342**, 2021.

APÊNDICE A – REPOSITÓRIO COM CÓDIGOS

Abaixo é informado o *link* para o repositório Github, onde é possível acessar e visualizar os códigos para treino e análise dos modelos.

<https://github.com/lucasloch/TCC>