



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**PEDRO HENRIQUE DE OLIVEIRA SILVA**

**LOCALIZAÇÃO DE FRATURAS NA COLUNA CERVICAL EM CT  
UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS**

**CAMPINA GRANDE - PB**

**2023**

**PEDRO HENRIQUE DE OLIVEIRA SILVA**

**LOCALIZAÇÃO DE FRATURAS NA COLUNA CERVICAL EM CT  
UTILIZANDO REDES NEURASIS CONVOLUCIONAIS**

**Trabalho de Conclusão Curso  
apresentado ao Curso Bacharelado em  
Ciência da Computação do Centro de  
Engenharia Elétrica e Informática da  
Universidade Federal de Campina  
Grande, como requisito parcial para  
obtenção do título de Bacharel em  
Ciência da Computação.**

**Orientador : Herman Martins Gomes**

**CAMPINA GRANDE - PB**

**2023**

**PEDRO HENRIQUE DE OLIVEIRA SILVA**

# **LOCALIZAÇÃO DE FRATURAS NA COLUNA CERVICAL EM CT UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS**

**Trabalho de Conclusão Curso  
apresentado ao Curso Bacharelado em  
Ciência da Computação do Centro de  
Engenharia Elétrica e Informática da  
Universidade Federal de Campina  
Grande, como requisito parcial para  
obtenção do título de Bacharel em  
Ciência da Computação.**

## **BANCA EXAMINADORA:**

**Herman Martins Gomes**

**Orientador – UASC/CEEI/UFCG**

**Marcus Salerno de Aquino**

**Examinador – UASC/CEEI/UFCG**

**Francisco Vilar Brasileiro**

**Professor da Disciplina TCC – UASC/CEEI/UFCG**

**Trabalho aprovado em: 28 de Junho de 2023.**

**CAMPINA GRANDE - PB**

## RESUMO

A cada 5 segundos uma vértebra da coluna é fraturada no mundo e com o envelhecimento natural da população esse tempo tende a diminuir. O principal método para analisar tais fraturas é a tomografia computadorizada (também conhecida como CT) e o tempo dispendido para a sua análise pode ser determinante para minimizar a ocorrência de sequelas nos pacientes. Sendo assim, o emprego de técnicas como redes neurais convolucionais pode auxiliar o profissional de saúde a tomar uma decisão com maior agilidade. Neste contexto, este artigo visa investigar uma abordagem fundamentada em uma rede neural convolucional para localizar fraturas em fatias de CT. A base de dados de detecção de fraturas da coluna cervical da RSNA (Sociedade Radiológica da América do Norte) 2022, disponibilizada em uma competição da plataforma Kaggle, foi utilizada para fins de treinamento e validação experimental de uma rede neural proposta. Espera-se que a investigação realizada neste trabalho e o modelo obtido possam ajudar profissionais de saúde na tarefa de detecção e localização de fraturas na coluna cervical.

# **LOCALIZATION OF CERVICAL COLUMN FRACTURES ON CT USING CONVOLUTIONAL NEURAL NETWORKS**

## **ABSTRACT**

Worldwide, every 5 seconds a spinal column vertebrae is fractured and with the natural population aging process this time interval tends to reduce. The main method for fracture analysis is the computed tomography (also known as CT) and the time required for its analysis is crucial to avoid complications to the patients. Therefore the usage of techniques such as convolutional neural networks may accelerate the healthcare professional's decision making process. This article aims at investigating a convolutional neural network approach to localize fractures in CT slices. The RSNA (Radiological Society of North America) 2022 fracture detection database, made available in the Kaggle platform, was explored with the purposes of training and validating a proposed neural network. It is expected that the investigation conducted in the work and the model produced may assist health professionals in the task of detection and localization of fractures in the cervical column.

# LOCALIZAÇÃO DE FRATURAS NA COLUNA CERVICAL EM CT UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS

Pedro Henrique de Oliveira Silva  
Unidade Acadêmica de Sistemas e Computação  
Universidade Federal de Campina Grande Campina  
Grande, Paraíba, Brasil

pedro.henrique.silva@ccc.ufcg.edu.br

Orientador: Herman Martins Gomes  
Unidade Acadêmica de Sistemas e Computação  
Universidade Federal de Campina Grande Campina  
Grande, Paraíba, Brasil

hmg@computacao.ufcg.edu.br

## RESUMO

A cada 5 segundos uma vértebra da coluna é fraturada no mundo e com o envelhecimento natural da população esse tempo tende a diminuir. O principal método para analisar tais fraturas é a tomografia computadorizada (também conhecida como CT) e o tempo dispendido para a sua análise pode ser determinante para minimizar a ocorrência de sequelas nos pacientes. Sendo assim, o emprego de técnicas como redes neurais convolucionais pode auxiliar o profissional de saúde a tomar uma decisão com maior agilidade. Neste contexto, este artigo visa investigar uma abordagem fundamentada em uma rede neural convolucional para localizar fraturas em fatias de CT. A base de dados de detecção de fraturas da coluna cervical da RSNA (Sociedade Radiológica da América do Norte) 2022, disponibilizada em uma competição da plataforma Kaggle, foi utilizada para fins de treinamento e validação experimental de uma rede neural proposta. Espera-se que a investigação realizada neste trabalho e o modelo obtido possam ajudar profissionais de saúde na tarefa de detecção e localização de fraturas na coluna cervical.

## PALAVRAS-CHAVE

CNN, localização de objetos, CT, coluna cervical, YOLO.

## 1. INTRODUÇÃO

A coluna vertebral é formada por um conjunto de 33 ossos, e é uma das estruturas mais importantes do corpo humano. Além de garantir estrutura e mobilidade para o corpo, um de seus papéis mais importantes é a proteção da medula espinhal [1]. A medula serve como canal de comunicação entre o cérebro e as demais estruturas do corpo, sendo um componente fundamental do sistema nervoso. A medicina agrupa os ossos da coluna vertebral em 5 regiões, cervical, torácica, lombar, sacrum e coccyx. [1]

A cada 5 segundos uma vértebra da coluna é fraturada no mundo, e com o envelhecimento da população, esse tempo tende a diminuir. Apenas em 2019 a taxa de incidência de fraturas na coluna foi de pelo menos 8,58 milhões em todo o mundo [2]. Ao longo dos anos esse número vem crescendo. A avaliação por um profissional especializado de uma possível ocorrência de uma fratura na coluna, é geralmente realizada sobre um conjunto de imagens de tomografia computadorizada (CT). Entretanto, por muitas vezes essa análise é lenta. É crucial que o tempo de espera de um paciente aguardando a avaliação de um exame médico, dessa gravidade, seja o menor possível, a fim de evitar a ocorrência de sequelas. Diante do exposto, é proposto nesse documento o treinamento de uma rede neural convolucional, a partir do uso do modelo YOLO v8 [3], com o objetivo de localizar

fraturas na coluna cervical. O uso efetivo de redes neurais convolucionais para a resolução do problema de localização de fraturas, em imagens de radiografia e CT, já foi tratado anteriormente na comunidade científica [4, 5]. Entretanto, ainda não foram apresentados formalmente modelos treinados (apenas na comunidade Kaggle) que resolvam corretamente o problema de localização de fraturas da coluna cervical, com imagens de CT. Espera-se que o artefato produzido neste trabalho de conclusão de curso agilize o processo de análise do profissional de saúde, não com o intuito de substituí-lo, mas sim, de auxiliá-lo, reduzindo o tempo gasto na avaliação do exame, atingindo alto nível de sensibilidade. Deste modo, o diagnóstico mais rápido garantirá uma melhor prospeção ao estado do paciente.

## 2. OBJETIVOS

Esta Seção descreve os objetivos geral e específicos deste trabalho.

### 2.1 Objetivo Geral

Este trabalho tem como objetivo utilizar pré-processamentos de imagem e o treinamento da rede supervisionada de localização de objetos com o intuito de solucionar a tarefa de localização de fraturas em imagens de tomografia computadorizada da coluna cervical.

### 2.2 Objetivos Específicos

- Treinar e validar o modelo buscando atingir boa taxa de sensibilidade;
- Realizar inferências com baixo tempo de resposta;
- Facilitar acesso livre aos resultados e ao software necessário para sua reprodução

## 3. METODOLOGIA

O passo inicial deste trabalho foi a análise exploratória da base de dados. Em seguida foi realizado pré-processamento sobre os dados de interesse, tendo sido organizados e repartidos. Por fim, sendo realizado o treinamento de uma rede neural YOLO v8 com estes dados, a escolha do modelo é justificada adiante. Ao fim destas etapas foi desenvolvido um modelo capaz de localizar fraturas em imagens da coluna cervical, que está disponibilizado em um repositório aberto na plataforma *GitHub* [6].

A análise e treinamento foram realizados sobre a base de dados [7], disponibilizada pela Sociedade Norte Americana de Radiologia (RSNA), Sociedade Americana de Neurorradiologia (ASNR) e a Sociedade Americana de Radiologia Espinhal (ASSR), na comunidade aberta *Kaggle* [8]. Esta base é composta por mais de 700000 imagens de tomografia computadorizada capturadas de cerca de 2000 pacientes, e atende a diferentes

propósitos. Nela foram cedidos, além das imagens, informações em relação a qual vértebra ocorreu a fratura, segmentação das vértebras e retângulos envolventes da região em que houve a fratura, isto possibilita o uso desta base para solucionar tarefas como classificação, segmentação semântica e localização de objetos. No escopo deste trabalho, serão utilizadas as informações de retângulos envolventes para solucionar a tarefa de localização da fratura. As métricas que serão acompanhadas ao longo do trabalho serão mAP (mean Average Precision), IoU (intersecção sobre a união), precisão e revocação. A escolha desta base foi motivada pela escassez de acesso a dados de tomografia computadorizada da coluna cervical e pela confiança na qualidade dos dados, visto que foi realizada uma auditoria das anotações (*Ground Truth*) dos dados de treinamento por profissionais médicos.

Foi realizada uma busca por material bibliográfico na plataforma da *Google Scholar* [9]. Os critérios da busca e bibliografias citadas neste trabalho estão detalhadas na Seção 3.1.

A análise exploratória de dados (EDA) foi realizada sobre a base de dados disponibilizada pela RSNA, ASNR e ASSR na plataforma *Kaggle*, esta análise contempla a exploração das imagens de tomografia computadorizada e as anotações do local em que ocorreu a fratura. A EDA está presente na Seção 3.2.

A escolha da arquitetura para o treinamento foi a rede YOLO v8 [3], os critérios dessa escolha foram os bons resultados na tarefa de localização de objetos na base de dados COCO val2017 [10] e seu tempo de resposta relativamente baixo.

Os experimentos resguardados neste trabalho comportaram modificações no pré-processamento dos dados e particionamento das imagens de diferentes maneiras, como separação de treinamento e teste por caso positivo e negativo por paciente ou simplesmente por caso positivo e negativo. Mais detalhes dos experimentos realizados estão disponíveis na Seção 4.

Os resultados obtidos não atendem a todos os critérios estabelecidos como meta no planejamento do trabalho, contudo o modelo treinado ainda pode se mostrar útil em cenários específicos. A discussão destes resultados e cenários de aplicação do modelo estão descritos na Seção 5.

### 3.1 Revisão bibliográfica

A busca por material bibliográfico foi realizada na plataforma *Google Scholar*. Os critérios de busca foram artigos publicados nos últimos 3 anos com as seguintes palavras-chave: “CT”, “fracture”, “localization” e “CNN”.

Na busca por bibliografia não foi encontrado nenhum artigo que tenha o mesmo objetivo que este trabalho (localização de fraturas na coluna cervical em imagens de tomografia computadorizada, utilizando CNN). Foi encontrado um trabalho com objetivo similar [5], localização de fraturas da costela em imagens de tomografia computadorizada utilizando CNN. O trabalho citado apresenta resultados relevantes na tarefa de localização no contexto de fraturas de costela.

Um segundo trabalho [2] apresenta uma análise sistemática do estudo Global Burden of Disease Study 2019 (GBD 2019) [11], analisando principalmente a incidência e prevalência de fratura de ossos nos contextos nacional, regional e global. Parte das informações apresentadas na análise e no estudo são citadas no trabalho corrente.

A falta de bibliografia mais aprofundada no contexto deste trabalho representa um dos desafios nas etapas de experimentação.

### 3.2 Análise Exploratória de Dados

A base de dados fornecida através da plataforma *Kaggle* apresenta informações diversas que podem ser utilizadas para propósitos além da localização de fraturas, mas também para classificação de em qual vértebra ocorreu a fratura, segmentação semântica da vértebra na imagem ou uma análise de dados mais profunda. Visto que esta base atende a múltiplos propósitos, é esperado que nem todos os dados estejam rotulados para os objetivos deste trabalho, isto é descrito adiante.

Para a finalidade deste trabalho, foram utilizados os dados que dizem respeito às imagens (de tomografias computadorizadas) da coluna dos pacientes e os retângulos envolventes da localização em que ocorreu a fratura.

A base de dados fornece duas pastas contendo imagens monocromáticas de CT, separadas por paciente, com dimensões de 512x512 em sua grande maioria. Uma pasta com conjunto de treinamento e outra pasta com um conjunto de teste estão disponíveis. Contudo, o acesso aos rótulos (ou *Ground Truth*) deste conjunto de teste é restrito, dado que esta é uma base de competição e esses rótulos são utilizados internamente pela plataforma *Kaggle* para pontuar o nível de generalização destes modelos. Por esta razão, a pasta de teste não é utilizada e a pasta de treinamento (pasta principal) foi repartida para criar os conjuntos de treinamento, teste e validação.

A pasta principal comporta imagens de CT de 2019 pacientes diferentes, totalizando 711601 imagens. Destes pacientes, 1058 não sofreram nenhuma fratura, totalizando 320919 imagens. Os 961 pacientes restantes, que sofreram ao menos uma fratura, totalizam um volume de dados de 390682 imagens. A Figura 1 apresenta o histograma de presença, de pelo menos uma fratura por paciente.

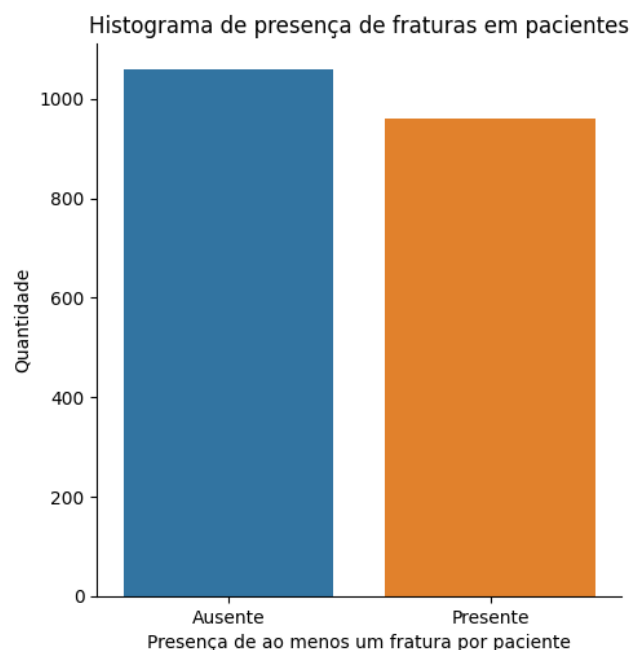


Figura 1: histograma da presença de fraturas por paciente.

A Figura 2 apresenta o boxplot do número de imagens por paciente, com a mediana de aproximadamente 352 e o

desvio-padrão de aproximadamente 132 imagens.  
Boxplot do número de imagens por paciente

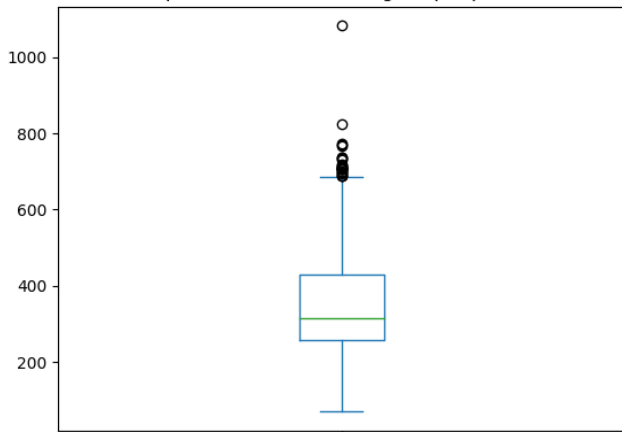


Figura 2: boxplot do volume de imagens por paciente.

Apesar deste grande volume de dados, existem apenas rótulos demarcando o local da fratura para 7217 imagens. Destas, 14 imagens não se enquadram nas dimensões de 512x512 e sofreram redimensionamento para 512x512. Os rótulos fornecidos pela base para localização de fraturas são as coordenadas na imagem que indicam o local em que ocorreu a fratura, conforme a Figura 3.

Exemplos de imagem com fratura delimitada por retângulo vermelho

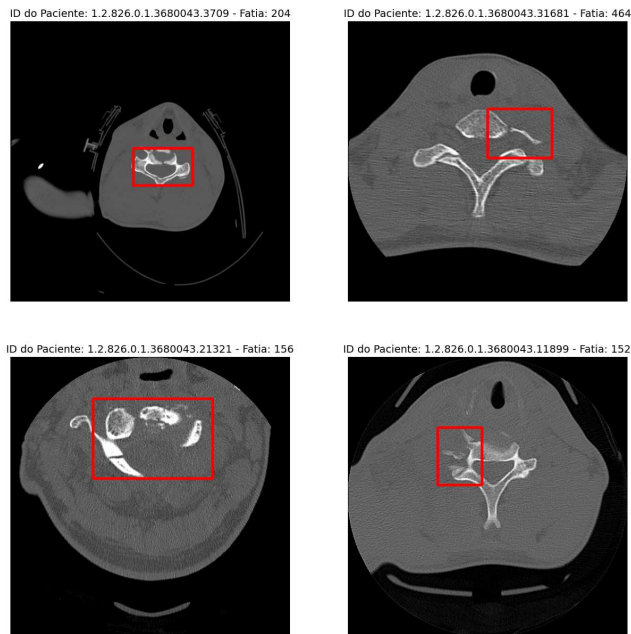


Figura 3: 4 exemplos de imagens pré-processadas com a fratura demarcada por um retângulo vermelho.

A base de dados antes de realizar a partição é formada por 7217 imagens com rótulo de fratura (exemplos positivos) e outras 7200 escolhidas aleatoriamente do conjunto total de imagens sem fratura (exemplos negativos). Mais detalhes sobre o pré-processamento e seleção de imagens estão descritos na Seção 4.

## 4. Validação Experimental

A validação experimental deste trabalho inicia com o pré-processamento das imagens da base de dados, descrita na Seção 4.1. Seguida pela repartição dos dados em conjuntos de

treinamento, validação e teste, por fim realizado o treinamento, descrito na Seção 4.2.

A base de dados de dados é composta por imagens no padrão internacional *DICOM*. Este padrão regula o armazenamento e comunicação de imagens no contexto médico, como em radiografias, tomografias computadorizadas e ressonâncias magnéticas [12]. Arquivos que seguem este padrão armazenam não apenas os dados brutos da imagem, obtidos pela captura do equipamento, mas também um conjunto de metadados associados ao imageamento. Como por exemplo, dimensões da imagem, quantização de pixels na captura, parâmetros para operações de redimensionamento e janelamento [13].

### 4.1 Pré-processamento

O pré-processamento dos dados é fundamental para a generalização do modelo, visto que o armazenamento da informação extraída pelo equipamento de CT é dependente dos parâmetros e algoritmos específicos utilizados pelo equipamento. Sendo assim, é necessário transformar os dados para uma escala independente de equipamento, no contexto de tomografia computadorizada, geralmente a escala *Hounsfield* [14] é utilizada, cujas unidades são chamadas de *Hounsfield Unit* (HU). O equipamento é calibrado para que o coeficiente de atenuação (ou seja, o grau de penetrabilidade de um feixe) da água destilada tenha o valor de -1000 HU e o do ar tenha o valor de 0 HU. Esta calibração permite o uso consistente das informações em cenários independente do hardware específico [14]. Este pré-processamento é realizado através ou de uma substituição de valores a partir de uma tabela de busca, ou a partir de uma transformação linear com uma variável dependente e uma independente, um dos dois métodos [15] estará especificado nos metadados do arquivo DICOM.

Dado que a tomografia computadorizada mensura o coeficiente de atenuação de materiais (ou seja, o grau de penetrabilidade por um feixe eletromagnético), diferentes materiais possuem diferentes coeficientes. A partir destes coeficientes é possível diferenciar os tipos de materiais visualizados na tomografia. A etapa de janelamento é um processo de realce de materiais ou tecidos orgânicos de interesse, realizada em imagens pré-processadas. Por exemplo, o janelamento de valores de HU no intervalo de -120 a -90 ressalta o imageamento da gordura [16] e o tecido esponjoso do osso é ressaltado no intervalo de +300 a +400 [17]. A fim de facilitar a identificação de uma fratura óssea para fins de visualização humana e para a generalização da rede neural, é necessário realizar um enfoque nas informações de interesse. A etapa de janelamento pode ser realizada a partir de um recorte da visualização ou a partir de uma busca sobre a centralização da imagem e da uma transformação de valores a partir de uma tabela de busca. O janelamento por recorte é realizado a partir do centro da janela e da largura da janela, em que o centro é o valor que divide a janela em duas partes iguais [18]. Por exemplo, para realizar o janelamento da gordura para a escala de -120 HU a -90 HU, o centro da janela seria -105 HU e o tamanho da janela seria 30 HU. O tipo de janelamento a ser aplicado na imagem é especificado nos metadados do arquivo *DICOM*.

A última etapa do pré-processamento é a normalização dos dados, realizada após a sequência compreendida pela sequência de calibração da intensidade e do janelamento das imagens. As imagens desta base de dados possuem uma quantização de 16 bits de informação monocromática, ou seja, as imagens possuem 65536 possíveis valores de intensidade. Após o janelamento é garantido que a quantização das imagens permaneçam com 16 bits, porém o valor mínimo e máximo são diferentes para cada uma das imagens. Por esta razão é necessário normalizar os valores para o intervalo 0 a 65535. Esta



normalização é realizada a partir da subtração dos píxeis da imagem pelo mínimo e divisão pela diferença do máximo e mínimo.

A Figura 4 apresenta 4 imagens de sem nenhum pré-processamento, apenas normalização. A Figura 5 apresenta as mesmas 4 imagens sofrendo o pré-processamento completo.

Exemplos de imagem sem pré-processamento

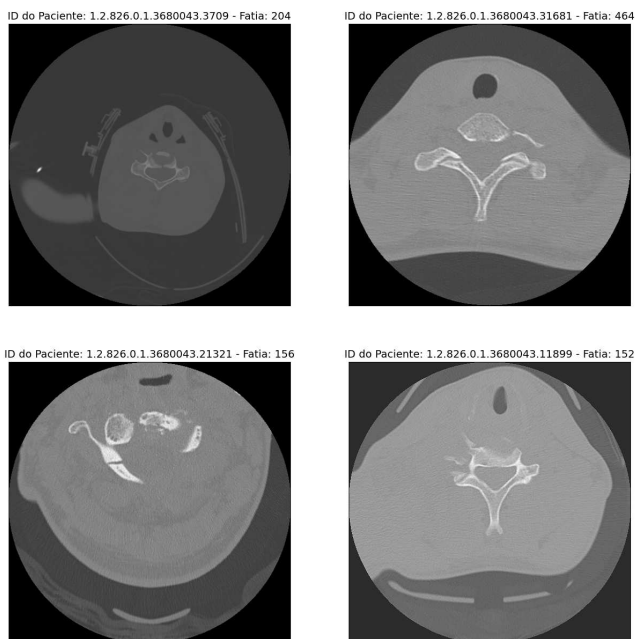


Figura 4: 4 exemplos de imagens sem pré-processamento.

Exemplos de imagem com pré-processamento

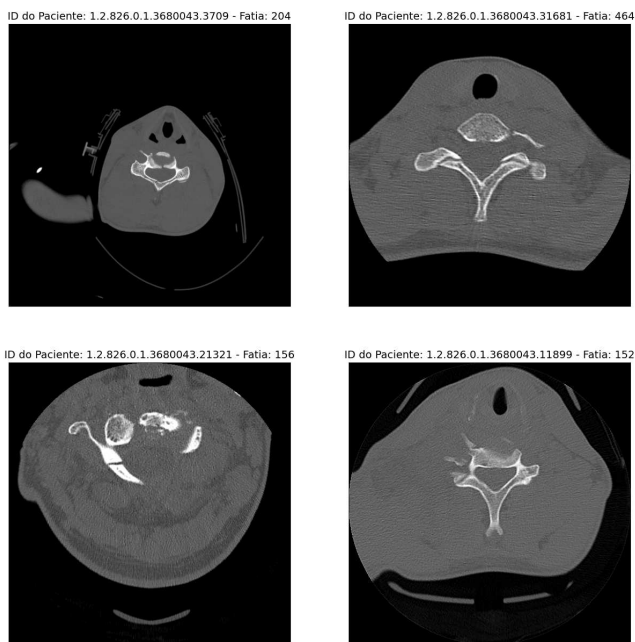


Figura 5: aplicação de pré-processamento sobre as imagens da Figura 4.

É notável o realce da região da coluna pelo pré-processamento, isto facilita sua visualização e o aprendizado

da rede. Uma vez com a imagem pré-processada, as etapas seguintes são o particionamento dos dados, o aumento dos dados e o treinamento, estas etapas estão descritas na Seção 4.2.

## 4.2 Treinamento

O procedimento para realizar o treinamento foi o particionamento dos dados, seguido do aumento de dados para garantir generalização e por fim o treinamento.

O particionamento dos dados foi feito na proporção de 70% para treinamento, 15% para validação e 15% para teste. Este particionamento foi realizado por paciente, ou seja, nunca duas imagens de um mesmo individuo estarão presentes em mais de um conjunto. Isto impede o enviesamento do modelo por paciente. Além disso, as imagens que não possuem fratura de pacientes que possuem ao menos uma fratura não foram utilizadas, dado que o conjunto de informações a respeito de quais imagens são fraturas é incompleto. Esta prática evita a rotulação incorreta de imagens, a rotulação de negativa (ausência de fratura) para imagem positiva (presença de fratura). Este particionamento resultou em uma separação de 10177 imagens, 2023 imagens e 2217 imagens para os conjuntos de treinamento, validação e teste, respectivamente. Para cada um destes conjuntos, metade das imagens é composta de exemplos positivos (possuem ao menos uma fratura) e a outra metade é composta de exemplos negativos (imagens sem fratura).

Uma vez com os dados particionados, é realizado o aumento de dados online pela própria ferramenta de interface de linha de comando (CLI) da rede YOLO v8 [19]. Vale ressaltar que as transformações realizadas na imagem também são aplicadas aos rótulos dela, ou seja, imagens aumentadas possuem rótulos consistentes. Os parâmetros customizados para o aumento de dados utilizados que garantem o melhor resultado no conjunto de validação foram o seguintes:

- Rotação em graus: 90.0 (rotação +/-);
- Aplicação de mosaico: 1.0 (probabilidade);
- Espelhamento horizontal: 0.5 (probabilidade);
- Espelhamento vertical: 0.5 (probabilidade);
- Redimensionar imagem: 512 (tamanho final);
- Redimensionamento: 0.5 (ganho +/-);
- Translação: 0.28 (probabilidade);
- Aumento de HSV:
  - Fração de H: 0.015;
  - Fração de S: 0.7;
  - Fração de V: 0.4.

Após o aumento de dados, é realizado o treinamento utilizando a ferramenta de CLI da rede. O modelo base selecionado foi o “yolov8m”, possuindo cerca de 26 milhões de parâmetros e atingindo o resultado de 52,9% de mAP com IoU de 50-95% no conjunto de validação COCO val2017 [10]. Este é o modelo médio, não possui tantos parâmetros quanto os modelos grandes e extra-grandes (“yolov8l” e “yolov8x” respectivamente) e ainda atinge um resultado relevante no conjunto COCO. O treinamento com melhor resultado foi realizado com pesos iniciais aleatórios (sem fine-tuning), utilizando o otimizador SGD, com 400 épocas, tamanho de *batch* de 16, paciência de 20 épocas e com *Dropout* de 22%.

Os demais parâmetros utilizados para o treinamento, tais como decaimento de pesos, *warmup* da taxa de aprendizagem, parâmetro de momentum, entre outros, estão disponibilizados no repositório do *GitHub* deste trabalho.

As técnicas e parâmetros apresentados anteriormente foram os que garantiram os melhores resultados no conjunto de validação, evitando ao máximo enviesamento. Vale mencionar que foram utilizados outros parâmetros e técnicas, porém causaram enviesamento do modelo ou piora da performance nos resultados do conjunto de validação. Foram experimentados:

- Variações de todos os parâmetros de aumento de dados descritos acima (com exceção do redimensionamento da imagem que foi mantido sempre em 512 pixels);
- Diferentes valores de *Droupout* para regularização do modelo em tempo de treinamento;
- Não aplicação do redimensionamento (em HU) e não aplicação do janelamento (ambos descritos na Seção 4.1);
- Uso de variações de arquitetura mais simples, “yolov8n” (nano) e “yolov8s” (small);
- Uso de variação de arquitetura mais complexa “yolov8l” (large);
- Uso de diferentes arquiteturas da YOLO v6 [20, 21] e suas variantes “yolov6m”, “yolov6l”;
- Uso de 1000 e 10000 imagens de background (exemplos negativos) ao invés dos 7200;
- Não separar imagens por paciente, apenas entre casos positivos e negativos (causa enviesamento);
- Normalização pela média e 3 desvios padrão, calculados sobre o conjunto de treinamento, ao invés do mínimo e máximo;
- Normalização pela média e 1 desvio padrão, calculados sobre o conjunto de treinamento, ao invés do mínimo e máximo;
- Uso dos dados de segmentação da coluna como máscara para as imagens da coluna.

O treinamento foi realizado em cerca de 5h, assim que a paciência do modelo foi exaurida e o loss convergiu no conjunto de validação.

O treinamento foi realizado no seguinte hardware:

- Placa de Vídeo Asus NVIDIA Geforce RTX3070, 8GB, GDDR6
- 32 GB RAM 3600Mhz
- Processador Ryzen 5 5500 3.6GHz-4.2GHz 6/12 cores.

Os resultados nos conjuntos de treinamento, validação e teste estão descritos na Seção 5.

## 5. Resultados

As métricas utilizadas para avaliar os resultados foram mAP (ou AP), IoU, precisão e revocação. Usualmente essas métricas são utilizadas em conjunto na tarefa de localização de objetos para avaliar o quão preciso é o modelo com determinado limiar. A métrica IoU significa intersecção sobre união (*Intersection over Union*), e representa a proporção entre a intersecção do rótulo (*Ground Truth* ou GT) e da predição sobre a união entre o GT e a predição. O intervalo do IoU é entre 0.0 e 1.0. Quanto mais próximo do valor 1.0, melhor o resultado, quanto mais próximo de zero, pior o resultado. A métrica mAP representa a média da precisão média ao longo de diferentes classes, no contexto deste trabalho ela se comporta exatamente como a métrica AP, visto que não há múltiplas classes. A métrica de interesse neste trabalho é a mAP 50% IoU, isto é, a precisão média, considerando um acerto como as predições cujo IoU tem ao menos 50% de similaridade com o GT.

As funções de perda avaliadas foram três: *box\_loss*, *cls\_loss* e *dfl\_loss*. O *box\_loss* (ou *IoU loss*) representa a função de perda que otimiza o IoU, quanto menor o valor da função, mais parecidos são as dimensões, razão de aspecto e localização entre o GT e a predição. O *cls\_loss* (ou *classification loss*) representa a função de perda que otimiza a classificação dos objetos, quanto menor o seu valor, melhor é a classificação dos objetos nas predições. Por fim o *dfl\_loss* (ou *distribution focal loss*) representa a função de perda que otimiza o nível de ajuste de diferentes amostras no treinamento, amostras mais difíceis de serem aprendidas se destacam de amostras mais simples a partir

de um mecanismo chamado fator focal, os critérios para classificar uma amostra como difícil são valores baixos de confiança e tamanho pequenos de predição.

A Figura 6 apresenta os resultados obtidos das funções de perda para os conjuntos de treinamento e validação, e apresenta as métricas de precisão, revocação e mAP (IoU 50%) para o conjunto de validação.

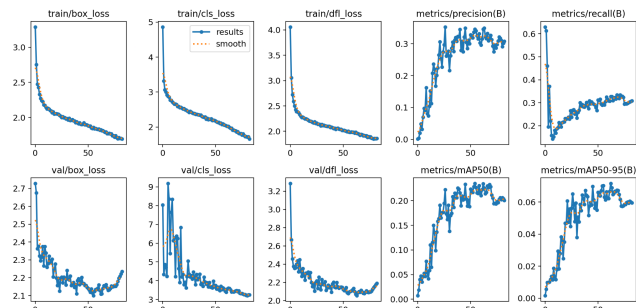


Figura 6: resultados do treinamento da rede YOLO v8, apresentando as métricas e funções de perda ao longo das épocas no eixo x.

Pelos resultados obtidos, é perceptível um comportamento adequado para o treinamento da rede, tendo as funções de custo *val/box\_loss* e *val/dfl\_loss* atingido visível convergência, assim como a métrica de mAP (IoU 50%). A rede encerrou o treinamento na época 82 e o melhor resultado foi obtido na época 62. A Tabela 1 apresenta o melhor resultado obtido.

Conjunto	Precisão	Revocação	mAP (50% IoU)
Validação	<b>34,1%</b>	<b>32,8%</b>	<b>23,3%</b>
Teste	33,1%	28,9%	22,5%

Tabela 1: melhores resultados obtidos nos conjuntos de validação e teste para as métricas de precisão, revocação e mAP (IoU 50%).

O tempo de inferência por imagem é de aproximadamente 7,8 ms em GPU e 268,8 ms em CPU.

A Figura 7 apresenta a curva de precisão e revocação do conjunto de validação. E a Figura 8 apresenta a curva de precisão e revocação no conjunto de teste.

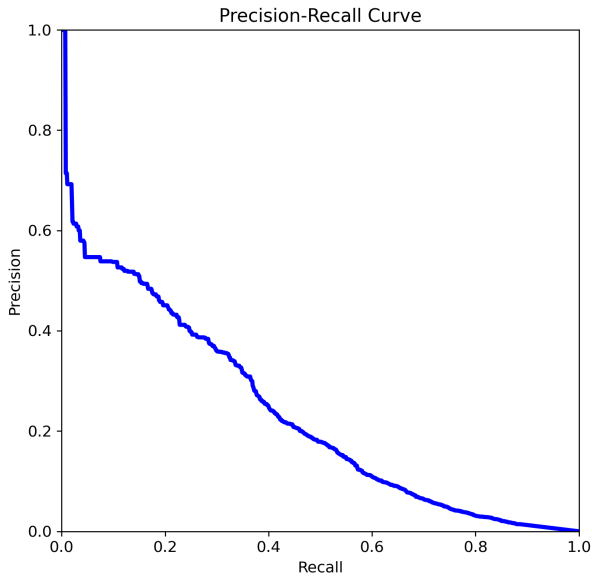


Figura 7: curva de precisão e revocação no conjunto de validação.

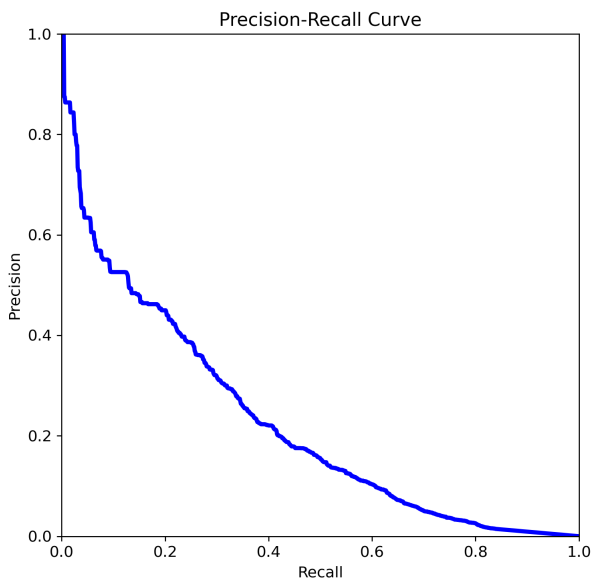


Figura 8: curva de precisão e revocação no conjunto de teste.

A Figura 9 apresenta 16 imagens e seus rótulos destacados em vermelho quando há algum tipo de fratura e a Figura 10 apresenta estas mesmas 16 imagens com suas respectivas predições demarcadas em vermelho com o nível de confiança.

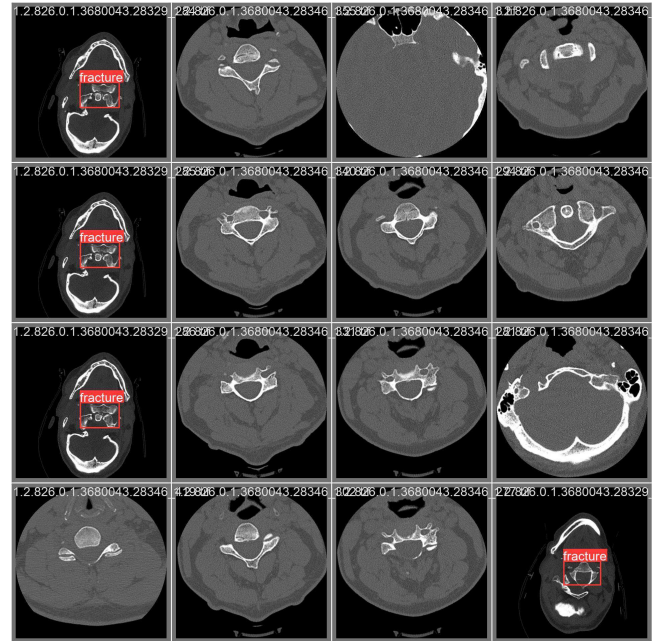


Figura 9: 4 exemplos de imagens com fraturas demarcadas por retângulos e 12 exemplos de imagens sem fratura.

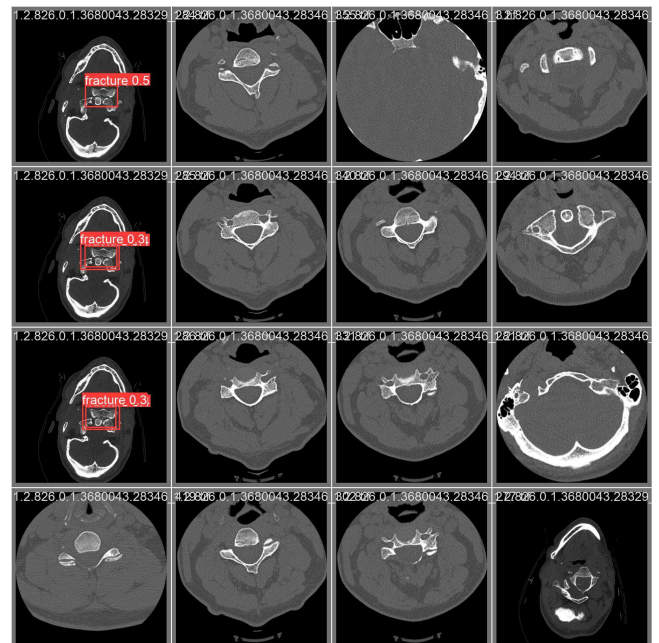


Figura 10: predição da rede sobre as 16 imagens da Figura 9, as fraturas localizadas estão localizadas com retângulos vermelhos.

Na Seção 6 são discutidos os resultados obtidos neste trabalho, suas possíveis aplicações e as considerações finais, incluindo propostas de trabalhos futuros.

## 6. Discussão

A tarefa de localização de fraturas em imagens de tomografia computadorizada é uma tarefa difícil, ainda não é bem resolvida em todos os cenários. Isto se deve principalmente pela escassez de dados rotulados com boa qualidade, variedade de equipamentos diferentes que realizam a captura (com hardware e software específicos). No contexto de fraturas da coluna, há presença de outros materiais na imagem de natureza semelhante (como por exemplo presença dos ossos da mandíbula e ossos da coluna na mesma imagem). Estes obstáculos dificultam o acesso a bons

resultados com alta revocação. Outros trabalhos atingiram bons resultados com revocação acima de 94,8% [5], no entanto isto ocorreu no contexto de fratura de costelas e a quantidade de dados rotulados é cerca 50x maior que o deste trabalho corrente (340501 anotações de fraturas de costela contra 7217 anotações de fraturas da coluna). A revocação atingida neste trabalho foi de 28,9% no conjunto de teste, que é um resultado relevante para um tarefa difícil, entretanto não é adequado para um contexto de aplicação hospitalar. A Figura 8 apresenta a curva de precisão e revocação no conjunto de teste, nela podemos notar que calibrando o limiar para atingir uma revocação de pelo menos 60%, a precisão fica abaixo dos 10%, levantando a questão de: “é possível aproveitar este modelo para algum cenário específico?”. Este modelo pode ser aplicado no contexto de revisão médica, entretanto isso deve ocorrer de modo moderado e somente como uma segunda opinião a respeito do problema, não como analisador principal. A opinião do autor é que o valor deste trabalho não se configura apenas em atingir um resultado digno de aplicação final, mas também em apresentar que esta tarefa difícil é resolvível com a aplicação de mais recursos, e demonstrar as etapas que funcionaram e as que não.

Na Seção 6.1 serão tratadas ameaças a validação e seus devidos méritos no contexto do trabalho. E a Seção 6.2 apresenta as considerações finais.

### 6.1 Ameaça a validação

O contexto de dados hospitalares é bastante sensível, não só no esforço da captura dos dados, mas também na responsabilidade da aplicação destes no contexto final. A repartição dos dados (separação entre conjuntos de treinamento, teste e validação) por paciente ao invés de uma repartição simples que só considera casos positivos e negativos, causou uma queda no mAP (IoU 50%) de 92,5% para 22,5% no conjunto de validação, porém este procedimento não introduz viés como o caso da repartição simples. Apesar desta queda, isto indica que imagens de diferentes pacientes são mais diferentes entre si do que entre as imagens dos próprios pacientes, naturalmente, mas também indica que a rede não generalizou suficientemente os ruídos de background entre os diferentes pacientes. Conclui-se que um treinamento com mais dados rotulados de diferentes pacientes traria uma maior grau de generalização à rede e consequentemente, um melhor resultado.

Os procedimentos aplicados na experimentação do trabalho evitaram ao máximo ameaçar a validade e garantir a reprodutibilidade do experimento.

### 6.2 Considerações finais

Conclui-se que os resultados obtidos têm relevância, porém são passíveis de melhoria, principalmente se o modelo for retreinamento com mais dados rotulados. O final da Seção 4.2 pontua experimentos realizados que não trouxeram benefícios nesta base para este problema. Possíveis sugestões para trabalhos futuros são:

- Retreinamento com mais dados rotulados;
- Uso de arquiteturas detectoras de duas fases, como R-CNN, FPN e Mask R-CNN;
- Revisão manual da base de dados;
- Treinamento de classificador binário de fratura e uso da rede YOLO após classificação.

### 7. Agradecimentos

Primeiramente agradeço aos meus pais, graças a eles eu estou onde estou hoje. Agradeço aos professores que me auxiliaram em minha formação ao longo de toda minha trajetória, desde o ensino fundamental até o ensino superior. E por fim agradeço aos meus amigos e familiares que me deram suporte ao longo de todos os anos.

### 8. Referências

- [1] Health Encyclopedia. Anatomia da coluna. (2023) <<https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=85&contentid=P07391>>.
- [2] Wu, Ai-min et al. “Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019” *The LANCET Healthy Longevity*. (2021).
- [3] Joker, Glenn, et al. “YOLO by Ultralytics”. Repositório no *GitHub*. (2023). <<https://github.com/ultralytics/ultralytics>>.
- [4] Thian, Yee Liang, et al. “Convolutional Neural Networks for Automated Fracture Detection and Localization on Wrist Radiographs”. *Radiology: Artificial Intelligence*. (2019). <<https://doi.org/10.1148/ryai.2019180001>>.
- [5] Zhou, Qing-Qing, et al. “Precise anatomical localization and classification of rib fractures on CT using a convolutional neural network”. (2022) <<https://www.sciencedirect.com/science/article/abs/pii/S089970712100382X>>.
- [6] Repositório deste trabalho no *GitHub*. <<https://github.com/pedrohos/cervical-spine-fracture-detection>>.
- [7] Flanders, Adam, et al. “RSNA 2022 Cervical Spine Fracture Detection”. *Kaggle*. (2022). <<https://kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection>>.
- [8] Site da comunidade *Kaggle*. <<https://www.kaggle.com/>>.
- [9] Site do motor de busca *Google Scholar*. <<https://scholar.google.com/>>.
- [10] Base de dados COCO. <<https://cocodataset.org/#detection-2017>>.
- [11] Dados do estudo *Global Burden of Disease Study 2019* (GBD 2019). <<https://ghdx.healthdata.org/gbd-2019>>.
- [12] Documentação do padrão DICOM, seção Visão Geral. <[https://dicom.nema.org/medical/dicom/current/output/chtml/part01/chapter\\_1.html#sect\\_1.1](https://dicom.nema.org/medical/dicom/current/output/chtml/part01/chapter_1.html#sect_1.1)>.
- [13] Documentação do padrão DICOM, seção Informação das Definições de Objeto. <[https://dicom.nema.org/medical/dicom/2016a/output/chtml/part03/sect\\_C.8.2.html](https://dicom.nema.org/medical/dicom/2016a/output/chtml/part03/sect_C.8.2.html)>.
- [14] Science Direct. Escala Hounsfield. (2023) <<https://www.sciencedirect.com/topics/medicine-and-dentistry/hounsfield-scale>>.
- [15] Medical Connections. Descrição de operações de pré-processamento de imagens de CT. (2023). <<https://www.medicalconnections.co.uk/kb/Lookup-Tables#modality-lut>>.
- [16] Página 83 em Herbert Lepor (2000). *Prostatic Diseases*. W.B. Saunders Company. ISBN 9780721674162.
- [17] Patrick S, Birur NP, Gurushanth K, Raghavan AS, Gurudath S. “Comparison of gray values of cone-beam computed tomography with hounsfield units of multislice computed tomography: an in vitro study”. *Indian J Dental Res: Off Public Indian Soc Dental Res*. 2017;28(1):66–70. Disponível em: <[https://doi.org/10.4103/ijdr.IJDR\\_415\\_16](https://doi.org/10.4103/ijdr.IJDR_415_16)>.
- [18] Innolitics. Atributo *Window With* no explorador de documentação do padrão DICOM. (2023). <<https://dicom.innolitics.com/ciods/vl-endoscopic-image/vl-image/00281051>>.

- [19] Joker, Glenn, et al. "YOLO by Ultralytics". Ferramenta de CLI. (2023). <<https://github.com/ultralytics/ultralytics/cli/>>.
- [20] Li, Chuyi, et al. "YOLOv6 v3. 0: A Full-Scale Reloading." arXiv preprint arXiv:2301.05586 (2023). <<https://arxiv.org/abs/2301.05586>>.
- [21] Meituan Vision AI Department. Repositório de implementação da YOLO v6 no GitHub. <<https://github.com/meituan/YOLOv6>>.