Processamento de imagens de radiografia de mãos para a detecção automática de Maturidade Óssea utilizando Redes Neurais Convolucionais

Bianca Bertoldo de Oliveira

Universidade Federal de Uberlândia
Faculdade de Engenharia Elétrica
Programa de Graduação em Engenharia Elétrica

Uberlândia
2023

Bianca Bertoldo de Oliveira

Processamento de imagens de radiografia de mãos para a detecção automática de Maturidade Óssea utilizando Redes Neurais Convolucionais

Trabalho apresentado na Universidade Federal de Uberlândia como requisito para conclusão do curso de Engenharia Eletrônica e de Telecomunicações.

Área de concentração: Engenharia Eletrônica e de Telecomunicações

Orientador: Milena Bueno Pereira Carneiro

Uberlândia
2023

Dedico este trabalho a minha família e amigos.
Muito obrigada.

Agradecimentos

Ao meu pai **Durval**, minha madrasta **Cyntia** e meu irmão **Henrick** pelo apoio e incentivo.

À **Profª. Drª. Milena**, pela sua atenção dedicada ao longo de todo o projeto.

Ao meu companheiro **Gustavo** pelo amor e suporte.

Aos meus familiares e amigos que me apoiaram ao longo dessa jornada.

"A ignorância gera mais frequentemente confiança do que o conhecimento: são os que sabem pouco, e não aqueles que sabem muito, que afirmam de uma forma tão categórica que este ou aquele problema nunca será resolvido pela ciência."

(Charles Darwin)

Resumo

A avaliação da idade óssea de pacientes pediátricos é uma prática clínica padrão para determinar sua maturidade biológica, sendo um exame de grande importância na identificação das condições de crescimento e desenvolvimento em crianças e na previsão de sua altura futura. Normalmente, esse processo envolve a revisão manual de imagens de radiografia da mão usando a abordagem Greulich-Pyle (GP) ou Tanner-Whitehouse (TW). O método GP emprega um atlas de mão padronizado como ponto de referência para estimar a idade óssea, enquanto a abordagem TW usa um sistema de pontuação baseado em várias regiões de interesse.

No entanto, este método manual é demorado e propenso a erros no cálculo da idade esquelética. Portanto, um sistema inteligente de avaliação da idade óssea usando inteligência artificial poderia automatizar a tarefa repetitiva de análise de imagens. O algoritmo proposto, desenvolvido em Python com o auxílio de suas bibliotecas, tem duas etapas: primeiro, normalização da imagem e segmentação das imagens de raios-X e, em seguida, o uso de uma arquitetura de regressão de rede neural convolucional com múltiplas entradas.

Este método tem se mostrado promissor não apenas para demonstrações teóricas de teor acadêmico na área de Processamento Digital de Imagens, como também para uso clínico prático, com uma taxa de erro médio de 7,31 meses, possibilitando auxiliar profissionais médicos a avaliar a idade óssea de forma objetiva.

Palavras-chave: Idade Óssea, Inteligência Artificial, Redes Neurais Convolucionais, Python, Processamento Digital de Imagens.

Abstract

The evaluation of the bone age of pediatric patients is a standard clinical practice to determine their biological maturity, being an exam of great importance in identifying the conditions of growth and development in children and in predicting their future height. Typically, this process involves manual review of radiographic images of the hand using the Greulich-Pyle (GP) or Tanner-Whitehouse (TW) approach.The GP method employs a standardized hand atlas as the reference point for estimating bone age, while the TW approach uses a scoring system based on multiple regions of interest.

However, this manual method is time consuming and prone to errors in skeletal age calculation. Therefore, an intelligent bone age assessment system using artificial intelligence could automate the repetitive task of image analysis. The proposed algorithm, developed in Python with the aid of its libraries, has two stepsfirst, image normalization and segmentation of the X-ray images, and then the use of a multi-input convolutional neural network regression architecture. .

This method has shown to be promising not only for theoretical demonstrations of academic content in the area of Digital Image Processing, but also for practical clinical use, with an average error rate of 7,31 months, making it possible to help medical professionals to assess bone age in a more accurate way.

Keywords: Boneage, Artificial Intelligence, Convolutional Neural Network, Python, Digital Image Processing.

Lista de ilustrações

Lista de tabelas

Lista de siglas

BLPF *Butterworth Low-pass Filter* (Filtro Passa-Baixa Butterworth)

CNN *Convolutional Neural Network* (Rede Neural Convolucional)

CLAHE *Contrast Limited Adaptive Histogram Equalization* (Equalização de Histograma Adaptativa Limitada por Contraste)

ER Eklof & Ringertz

GP Greulich & Pyle

GLPF *Gaussian Low-pass Filter* (Filtro Passa-Baixa Gaussiano)

ILPF *Ideal Low-pass Filter* (Filtro Passa-Baixa Ideal)

LPF *Low-pass Filter* (Filtro Passa-Baixa)

MLP *Multilayer Perceptron* (Perceptron de Multicamada)

MAPE *Mean absolute percentage error* (Erro percentual absoluto médio)

MAE *Mean absolute error* (Erro médio absoluto)

RNA Rede Neural Artificial

RMSE *Root mean squared error* (Raiz quadrada do erro-médio)

RSNA *Radiological Society of North America* (Sociedade de Radiologia da América do Norte)

RGB *Red - Green - Blue* (Vermelho - Verde - Azul)

TW Tanner & Whitehouse

Sumário

Capítulo 1

Introdução

1.1 Considerações Iniciais

A estimativa da idade óssea é uma técnica amplamente utilizada para avaliar distúrbios do crescimento em pacientes pediátricos. Ele fornece informações valiosas sobre o desenvolvimento da maturidade óssea em relação à idade cronológica. Essa avaliação é particularmente crucial em regiões onde os registros de nascimento não são acessíveis, e uma estimativa precisa da idade é necessária para eventos como imigração e esportes (MUGHAL; HASSAN; AHMED, 2014). Além disso, os dentistas também estudam a maturidade óssea para determinar o tratamento ortodôntico específico para seus pacientes com base na idade (TAVANO, 2001).

No entanto, esse processo só é preciso para indivíduos entre 0 e 19 anos, onde o crescimento ósseo para completamente. Na prática clínica, uma idade óssea 20% abaixo ou acima da idade cronológica é considerada anormal (MELMED et al., 2015). Resultados tardios ou avançados na avaliação podem ser indicativos de distúrbios pediátricos mais graves, como problemas relacionados ao estado nutricional (ZEFERINO et al., 2003), precocidade puberal e hipertireoidismo (LONGUI, 2003).

A técnica padrão para análise da maturação óssea é baseada em uma varredura radiológica da mão não dominante e é comparada às referências dos métodos disponíveis de forma manual por um radiologista. Os métodos mais difundidos e utilizados são de Greulich & Pyle (GP), Tanner & Whitehouse (TW) e Eklof & Ringertz (ER), que possuem etapas próprias. No método GP, são examinadas certas características da imagem em regiões de interesse e comparadas com seu atlas (GREULICH; PYLE, 1959).Já para a avaliação TW, é realizada uma avaliação com base em sub pontuações dadas para partes específicas da imagem (TANNER et al., 1975), e por último, no método ER, dadas as medições dos centros de ossificação, os resultados são comparados com tabelas de padrões mínimos e máximos para cada sexo (EKLÖF; RINGERTZ, 1967).

Fazer estimativas precisas da idade óssea é uma tarefa complexa que requer uma compreensão completa de vários fatores relacionados ao desenvolvimento ósseo (GILSANZ;

RATIB, 2005). Portanto, esse método tem uma carga de trabalho alta e consome recursos significativos, resultando em um longo tempo para realizar a análise. Além disso, a precisão da estimativa da idade óssea é vulnerável ao julgamento humano, o que pode levar a incertezas na avaliação.

1.2 Objetivo

A integração de técnicas computacionais na análise de imagens médicas revolucionou o campo do diagnóstico médico. A utilização dessas técnicas tem possibilitado a identificação de estruturas de difícil visualização, bem como a extração de regiões de interesse, facilitando a análise do radiologista e melhorando a precisão do diagnóstico (ERICKSON; BARTHOLMAI, 2002).

Este trabalho deu passos significativos na pesquisa, uso e implementação de técnicas de pré-processamento para obter automaticamente regiões de interesse.Além disso, tem foco no desenvolvimento, parametrização e treinamento de uma Rede Neural Artificial (RNA) de regressão que pode ser aplicada para diagnosticar a idade óssea.Essa abordagem evita o uso de redes pré-treinadas para ambas as etapas, o que exigiria conhecimento prévio sobre suas arquiteturas e funcionamento, e também dispensa a aplicação de métodos básicos de avaliação descritos anteriormente.

Além disso, neste trabalho também foi proposto para teste da RNA o uso de três algoritmos diferentes que atendem à diferentes regiões.O primeiro algoritmo contém imagens de radiografias de um dos dedos das mãos que foram separados do fundo, eliminando elementos desnecessários como identificações. O segundo algoritmo inclui apenas o punho da mão segmentado, e o terceiro algoritmo engloba o dedo juntamente com os ossos do punho em mosaico, avaliando o desempenho entre eles.

O objetivo final deste trabalho é realizar testes que garantam a eficácia das técnicas propostas. Com base nos resultados desses testes, pontos de melhoria podem ser identificados e trabalhos futuros podem ser sugeridos. A intenção é desenvolver uma ferramenta que possa ajudar radiologistas e outros profissionais médicos a fazerem diagnósticos mais precisos, aumentando assim a eficiência no tratamento de distúrbios relacionadas ao crescimento.

1.3 Estrutura do trabalho

O presente trabalho foi dividido nos seguintes capítulos:

- Capítulo 1 - Introdução - Introdução geral do trabalho, contendo considerações sobre os métodos manuais existentes, a motivação e objetivos definidos e a estrutura do trabalho.

- Capítulo 2 - Estado da Arte - Apresenta uma revisão de trabalhos relacionados à análise automatizada da maturação óssea, métodos utilizados e resultados obtidos.
- Capítulo 3 - Fundamentação Teórica - Descreve as técnicas de pré-processamento de imagens e resumo sobre RNAs.
- Capítulo 4 - Metodologia - Descreve as técnicas e ferramentas computacionais aplicadas, desde a aquisição de imagens, processamento, arquitetura da RNA utilizada e método para a avaliação dos resultados.
- Capítulo 5 - Resultados - São apresentados os resultados quantitativos e qualitativos obtidos pelo método aqui desenvolvido.
- Capítulo 6 - Conclusão - Principais conclusões do trabalho e sugestões para trabalhos posteriores.

Capítulo 2

Estado da arte

2.1 Introdução

Muitas técnicas clássicas de visão computacional e processamento de imagem foram aplicadas ao problema de classificação de imagens médicas com taxas variáveis de sucesso. Esses métodos geralmente incluem como pré-processamento a segmentação das imagens em regiões de interesse, e cálculos dos centros de ossificação usando os métodos padrão de TW, GP e ER para a predição da idade óssea, ou treinamento de algoritmos RNA, geralmente usando redes pré-treinadas, com a execução de um classificador ou regressão nos resultados.

Este capítulo fornece uma revisão bibliográfica do trabalho de alguns outros pesquisadores nas áreas de pré-processamento e classificação da maturação óssea em radiografias de mão. Os trabalhos escolhidos foram desenvolvidos de 2005 a 2022, e avaliados com relação a metodologia e resultados, através da elaboração de tabelas com as principais informações e considerações sobre o estado da arte. Por fim, são feitas as últimas considerações do capítulo.

2.2 Algoritmos de Determinação da Idade Óssea

2.2.1 Determinação por dimensões de regiões

Na avaliação dos métodos utilizando medidas de regiões para realizar a determinação da idade óssea, foram observados nos trabalhos o tipo de segmentação implementado, o número de amostras, o método teórico (TW, GP ou ER) e os resultados obtidos.

Observando-se a Tabela 1 é possível reparar que todos os trabalhos utilizaram centros de ossificação como imagem de análise, e possuem poucas imagens, o que pode indicar um método que não seja robusto a imagens que podem se desviar das analisadas em características como contraste e rotação.

Como o método proposto por esse trabalho utiliza regressão e métrica de erro percentual médio sem ponderar tolerâncias, nos trabalhos revisados foi considerado apenas a acurácia com menor desvio apresentada. O trabalho de (JÚNIOR, 2005) empregou o método ER, e obteve acerto de 30% e 50% considerando desvio padrão de 1,2 e 2,4 meses.

O método de (CASTRO et al., 2009) utilizou o método teórico de TW, e obteve acerto de 81,25%. Uma observação importante, é que no trabalho são apresentados os acertos considerando apenas os estágios catalogados por TW, onde as idades entre 0 e 18 anos são separadas em grupos de A até H, o que daria uma margem entre 2 e 3 anos para cada grupo. No método de TW, são considerados os estágios em conjunto com scores que vão de 0 até 1000, para estimar a idade óssea, aplicando-se o somatório dos scores obtidos, e para tanto, são apresentadas tabelas com os valores representativos para cada idade óssea e para cada sexo. Logo, essa taxa de acerto pode representar um desvio entre 1 e 2 anos, o que não é especificado no trabalho.

O método de (SILVA et al., 2016) pode ser considerado uma reavaliação do método de (CASTRO et al., 2009), já que utiliza as mesmas metodologias de segmentação e classificação, sendo que este obteve acerto de 90%, onde as mesmas observações anteriores são válidas, já que também são apresentados resultados apenas em estágios.

Tabela 1 – Comparação de resultados para avaliação por dimensões de regiões

Método	Segmentação	Número de amostras	Método Teórico	Resultados
(JÚNIOR, 2005)	10 centros de ossificação	200	ER	Acerto de 30% e 50% para desvio padrão de 1,2 e 2,4 meses
(CASTRO et al., 2009)	8 centros de ossificação	257	TW	Acerto de 81,25% para estágios
(SILVA et al., 2016)	8 centros de ossificação	265	TW	Acerto de 90% para estágios

2.2.2 Determinação por RNAs de Classificação

A segunda parte da análise de metodologia foi feita avaliando-se trabalhos elaborados utilizando RNAs com métricas de classificação. Para cada um dos métodos foram extraídas informações sobre segmentação, número de amostras, utilização ou não de *Data Augmentation* – geração de novas imagens utilizando cópias alteradas por exemplo em rotação, contraste, redimensionamento e orientação - e de *Transfer Learning* - reutilização de um modelo pré-treinado em um novo problema - e resultados da classificação.

Ao se avaliar a etapa de segmentação, foram utilizadas técnicas de segmentação de centros de ossificação, ossos inteiros ou mão inteira, com exceções de (CHEN, 2016), (POOSARLA, 2018) e (CHEN et al., 2020) que não realizaram nenhum tipo de segmentação em regiões de interesse.

Segundo (SHAHINFAR; MEEK; FALZON, 2020), em caso de recursos limitados, 150-500 imagens por classe são suficientes para alcançar uma precisão razoável. Isto posto, pode-se considerar que os métodos de (QUEIROZ, 2006), (SÉRGIO et al., 2011), (CHEN, 2016), (AN, 2017) e (SON et al., 2019) possuem poucos dados de treinamento.

Em relação ao uso de *Data Augmentation*, metade dos trabalhos analisados fizeram uso desse recurso, valendo ressaltar que é um método eficaz e de custo computacional relativamente baixo para melhorar o desempenho e a precisão dos modelos. Já para *Transfer Learning*, apenas (SÉRGIO et al., 2011) não empregou nenhum modelo.

(QUEIROZ, 2006) obteve um resultado de 95% de acerto para um desvio padrão de 3 meses. No trabalho, uma desvantagem é o método semi-automático, onde o usuário deve selecionar os centros de ossificação a serem analisados com o auxílio do mouse, o que pode ser oneroso em tempo. Além disso, as classificações foram realizadas apenas para a faixa entre 5 e 15 anos.

(SÉRGIO et al., 2011) alcançou um acerto de 81,25%, e no método de (LEE et al., 2017), o acerto médio alcançado foi de 59,36% porém, em ambos trabalhos, utilizou-se apenas os dados para crianças entre 8 e 15 anos, com a justificativa de ser a faixa etária que apresenta diferenças entre a idade óssea e a idade cronológica com maior frequência. No entanto, o sub/hiper desenvolvimento ósseo pode ser constatado desde o nascimento, onde já se pode observar complicações advindas de herança familiar, deficiência hormonal, falta de nutrientes na alimentação, doenças genéticas, uso de substâncias ilícitas durante a gestação e parto prematuro (CAVALLO et al., 2021). Além disso, das 205 imagens iniciais, cerca de 20% das radiografias não podiam ser utilizadas para análise devido a imperfeições.

Em (CHEN, 2016), o resultado observado foi um RMSE - raiz quadrada do erro-médio de 13 meses. (AN, 2017) obteve MAE - erro médio absoluto de 19 meses, onde houve uma diminuição de cerca de 30% da base inicial de 1380 imagens por motivos de más segmentações devido à dificuldade em padronizar um método para o pré-processamento das imagens com as variações de intensidades apresentadas, restando apenas 945 exemplares.

(POOSARLA, 2018) alcançou MAE de 6,28 meses, o que pode ser considerado o melhor resultado considerando a ausência de desvantagens e desenvolvimento do algoritmo para todas as faixas etárias. (SON et al., 2019) apesar de ter atingido MAE de 5,52 meses, também baseou os resultados apenas em acerto de estágios de TW, além de excluir os estágios A, B e C devido a poucas imagens, demonstrando uma avaliação com grande desvio.

(CHEN et al., 2020) e (HU et al., 2020) obtiveram respectivamente 62,6% e 41% na

acurácia top-1, onde apenas a classificação com maior certeza encontrada pelo modelo é considerada. Vale ressaltar também, que para treinamento da rede pré treinada YOLOV3 para detecção das zonas de interesse, em (HU et al., 2020) os dados de treinamento foram 190 imagens de ossos da mão marcadas manualmente, em contrapartida, o método desenvolvido nesse trabalho obteve bons resultados de segmentação realizando a marcação manual de apenas 80 imagens. A Tabela 2 apresenta um resumo do que foi discutido.

Tabela 2 – Comparação de resultados para avaliação por RNAs de classificação

Método	Segmentação	Número de amostras	Data Augmentation	Transfer Learning	Resultados
(QUEIROZ, 2006)	4 centros de ossificação	444	Não	Algoritmo de Levenberg-Marquardt	Acerto de 95% para desvio padrão de 3 meses
(SÉRGIO et al., 2011)	Centros de ossificação	205	Não	-	Acerto de 81,25%
(CHEN, 2016)	-	1400	Sim	VGGNet e GoogLeNet	RMSE de 13 meses
(LEE et al., 2017)	Mão inteira	8325	Sim	GoogLeNet e ImageNet	Acerto médio de 59,36%
(AN, 2017)	4 combinações diferentes usando 14 centros de ossificação, dedos e punho	945	Não	GoogLeNet, Alexnet e VGG-19	MAE de 19 meses
(POOSARLA, 2018)	-	12611	Sim	VGG-16	MAE de 6,28 meses
(SON et al., 2019)	13 ossos	3300	Sim	VGGNet	MAE de 5,52 meses
(CHEN et al., 2020)	-	12000	Não	ResNet	Acurácia top-1 de 62,6%
(HU et al., 2020)	13 ossos	7042	Sim	YOLOV3 e ResNet	Acurácia média top-1 de 41%

2.2.3 Determinação por RNAs de Regressão

Para a terceira e última parte de análise de metodologias, os trabalhos avaliados nessa seção utilizaram RNAs com métricas de regressão. Do mesmo modo da seção anterior, foram consideradas as escolhas de segmentação, número de amostras, utilização *Data Augmentation* e *Transfer Learning* e os resultados atingidos.

Na etapa de segmentação, foi obtida separação dos centros de ossificação, mão inteira, ou nenhum tipo de segmentação. Com relação ao número de amostras, todos os métodos propostos utilizaram a base de dados da *Radiological Society of North America* (Sociedade de Radiologia da América do Norte) (RSNA), que possui 12611 imagens de treino e validação. Em 4 dos 7 trabalhos citados foi empregado *Data Augmentation*, e todos utilizaram *Transfer Learning*.

Para avaliação dos resultados de predição, a métrica mais comum é MAE, e foi apresentada em todos os trabalhos. Pode-se notar que em média, os algoritmos de regressão

performaram melhor que os algoritmos de classificação.Em (SOUZA; OLIVEIRA, 2018), o MAE atingido foi de 6,44 meses. (IGLOVIKOV et al., 2018) encontrou MAE de 6,10 meses, porém, este possui algumas desvantagens. Ao todo, foram gerados 15 modelos diferentes, diferenciados em região segmentada, sexo, classificação e regressão. Primeiro, foi necessário a rotulação manual de 100 imagens para realizar a segmentação da mão inteira, e a rotulação de mais 800 imagens para segmentação de regiões de interesse menores, tais como centros de ossificação. Ainda, as idades entre 0 e 3 anos foram excluídas do treinamento. O melhor resultado obtido de MAE de 6,10 meses foi encontrado apenas realizando a técnica de *Ensemble Learning*, que consiste em combinar diversos modelos de predição mais simples e produzir a partir desses um modelo agrupado mais complexo, e para esse método, foram agrupados os 15 modelos treinados.

(MARROCOS et al., 2019) obteve MAE de 9 meses(PAN et al., 2020) alcançou MAE de 8,59, onde foi preciso rotular manualmente 300 imagens para a etapa de segmentação, e também foi utilizado *Ensemble Learning* para agrupar 2 modelos treinados.No método proposto por (WESTERBERG, 2020), o MAE atingido foi de 12 meses. Apesar de ter sido feito testes com centros de ossificação segmentados, onde teve-se o esforço de rotular 292 imagens, os modelos utilizando essas regiões segmentadas performaram pior do que os modelos sem segmentação alguma.

Em (ZULKIFLEY et al., 2021), o MAE foi de 7,69 meses.Para (GUO et al., 2022), o MAE foi de 6,07 meses, onde utilizando *Ensemble Learning* foram agrupados 6 modelos diferentes obtidos treinando cada centro de ossificação separado. A Tabela 3 apresenta todas as informações observadas.

Tabela 3 – Comparação de resultados para avaliação por RNAs de regressão

Método	Segmentação	Número de amostras	Data Augmentation	Transfer Learning	Resultados
(SOUZA; OLIVEIRA, 2018)	-	12611	Sim	Inception-V3	MAE de 6,44 meses
(IGLOVIKOV et al., 2018)	Mão inteira e regiões de interesse	12611	Sim	U-Net e VGG	MAE de 6,10 meses
(MARROCOS et al., 2019)	-	12611	Sim	Inception V3 e MobileNET	MAE de 9 meses
(PAN et al., 2020)	Mão inteira	12611	Não	U-Net, ImageNet e Inception-ResNet-V2	MAE de 8,59 meses
(WESTERBERG, 2020)	-	12611	Não	Mask R-CNN, Yolo, RetinaNet, Xception, InceptionV3, VGG19, e ResNet152	MAE de 12 meses
(ZULKIFLEY et al., 2021)	Mão inteira	12611	Sim	DeepLab V3+ e MobileNet V1	MAE de 7,69 meses
(GUO et al., 2022)	6 centros de ossificação	12611	Não	Faster R-CNN	MAE de 6,07 meses

2.3 Considerações Finais

Tendo em vista os diferentes métodos avaliados em estado da arte, é possível concluir que os algoritmos de regressão performam em geral melhor do que o resto, como esperado, em razão de que idade é uma variável contínua (ANDRADE, 2021). Os métodos de segmentação foram dos mais variados, indicando que para a tarefa de estimativa da idade óssea, outros parâmetros como robustez da rede utilizada e quantidade de amostras pode ser mais importante do que a região em si fornecida à RNA.

Neste trabalho, buscamos uma abordagem mais geral em que segmentamos apenas 1 dos dedos da mão e o pulso, realizando testes com essas regiões separadas e por fim as duas em mosaico. Evitamos a utilização de recursos pesados com limitações de tamanho de imagem de entrada, canais de cores e tipo de imagem, como a maioria de redes para *Transfer Learning* disponíveis, se limitando a utilizar apenas YOLO v3, que é uma rede facilmente treinável para qualquer problema de segmentação proposto, tendo vantagens em cima de outras redes pela sua rapidez na detecção, sem limitações na entrada (TAN et al., 2021). Além disso, será implementado *Data Augmentation* a fim de obter exemplares suficientes para todas as idades, e a RNA de regressão será elaborada do zero, experimentando com hiper parâmetros diferentes buscando o melhor resultado em comparação com o estado da arte.

Capítulo 3

Fundamentação Teórica

3.1 Processamento Digital de Imagens

O processamento digital de imagens é baseado na transformação de um campo de imagem contínua em uma estrutura digital comparável, por meio de um conjunto de componentes discretos e tamanhos limitados, chamados pixels, definidos em um arranjo bidimensional. Cada pixel está associado a um valor, em imagens em tons de cinza, ou um conjunto de três valores RGB (vermelho, verde e azul) para representar uma cor. A área de processamento de imagens alude à manipulação desses pixels visando melhora das informações visuais para a interpretação humana, o realce ou eliminação de certas características e a extração de informações (GONZALES; WINTZ, 1987).

Ao contrário da visão humana, que é restrita à faixa visual do alcance eletromagnético (EM), os dispositivos de manipulação de imagem cobrem quase todo o alcance EM, variando de ondas gama a ondas de rádio.Eles podem atuar em imagens criadas por fontes que as pessoas não estão acostumadas a associar com imagens, tais como ultrassom, microscopia eletrônica e imagens criadas por computadores. Nessa linha, o tratamento computadorizado de imagens incorpora um campo de usos amplo e variado.

Para (GONZALES; WINTZ, 1987), as etapas de processamento pode ser fragmentada conforme ilustrado na Fig. 1. O diagrama não significa que todo processo se aplique a uma imagem, e sim, que todas as metodologias podem ser aplicadas a imagens para diferentes propósitos e objetivos.

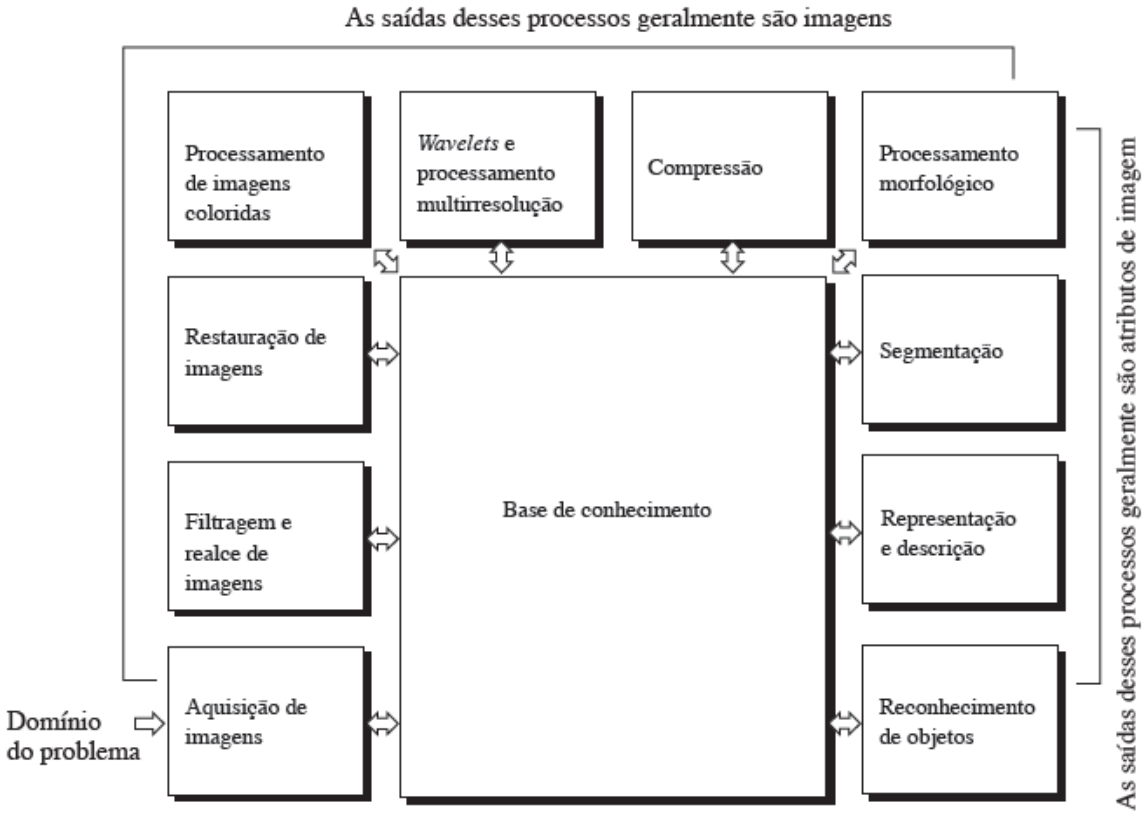

Figura 1 – Passos fundamentais em processamento digital de imagens. Fonte: (GONZALES; WINTZ, 1987). Adaptado pelo autor.

3.1.1 Equalização de Imagem

As imagens digitais adquiridas geralmente apresentam várias distorções e distúrbios. Por exemplo, a luminância de uma imagem natural que foi digitalizada pode ter sido altamente distorcida em direção aos níveis mais escuros, onde a maioria dos pixels possui uma luminância menor que a média. Nessas imagens, os detalhes nas regiões mais escuras geralmente não são perceptíveis. Um meio de aprimorar esses tipos de imagens é uma técnica chamada modificação do histograma, na qual a imagem original é alterada para que o histograma da imagem aprimorada siga alguma forma desejada (PRATT, 2007). Além de realce, a manipulação de histogramas pode ainda ser utilizada para compressão e segmentação de imagens.

O histograma de uma imagem digital com intervalo $[0, L-1]$ é uma função discreta $h(r_k) = n_k$, onde L é o numero de níveis de intensidade, r_k é o k-ésimo valor de intensidade e n_k é o número de pixels da imagem com intensidade r_k. Uma das técnicas de aprimoramento de contraste mais populares é a equalização de histograma, que tenta explorar totalmente a faixa dinâmica criando um histograma de saída uniformemente distribuído. Na equalização de histograma, a função de transformação $T(r)$ é dada pela equação (1):

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} \frac{n_j}{N}, \quad 0 \leq r_k \leq 1 \quad \text{e} \quad k = 0, 1, \ldots, L-1 \tag{1}$$

onde:

- $T(r_k)$ - Transformação que mapeia um valor de pixel r em um valor de pixel s.
- $p_r(r_j)$ - Função densidade de probabilidade do nível r_j da imagem de entrada.
- n_j - Número de pixels na imagem de entrada que têm nível r_j de tom de cinza.
- N - Número total de pixels na imagem de entrada.

A equalização possui duas técnicas diferentes: global e adaptativa. O método global é simples e eficaz, mas seu efeito é muito severo para muitas aplicações, pois não pode se adaptar aos recursos de brilho local da imagem de entrada (TOET; WU, 2014). Isso geralmente resulta em uma deterioração do contraste do plano de fundo e de objetos pequenos, sendo mais indicado para imagens que possuem seus níveis de cinza minimamente distribuídos. Na equalização adaptativa, histogramas separados são calculados para diferentes regiões da imagem, que são então usados para redistribuir os valores locais de nível de cinza (KIM; KIM; HWANG, 2001).

Nas Figuras 2 e 3 é mostrado o efeito das equalizações globais e adaptativas em duas imagens com brilho e contraste diferentes.

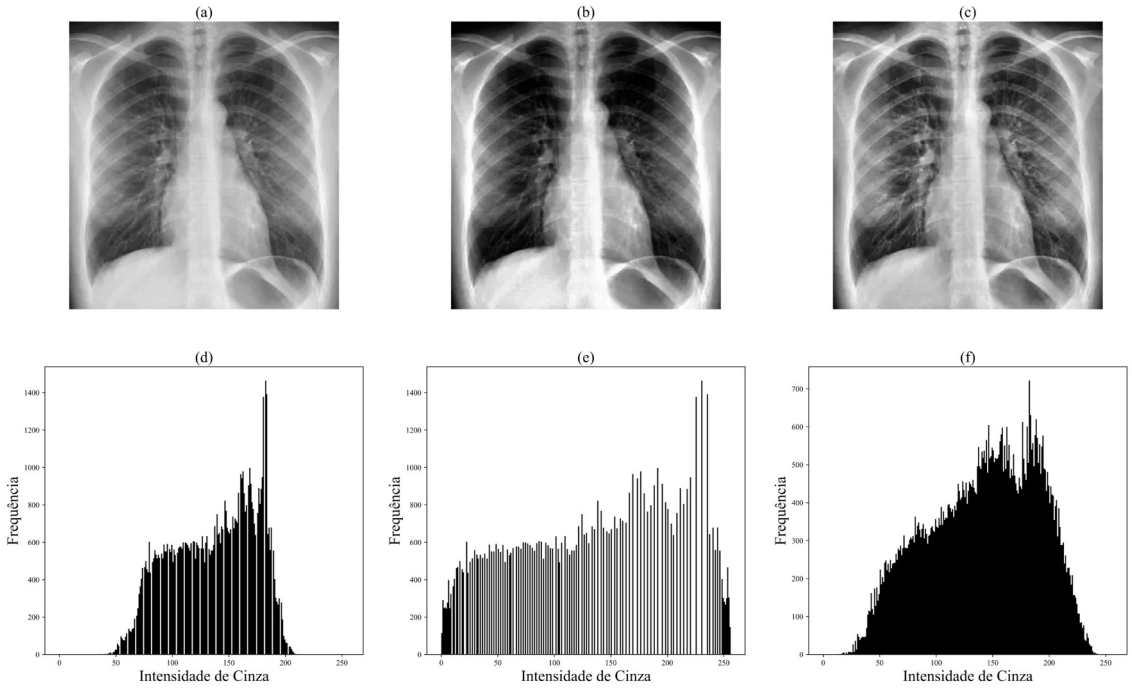

Figura 2 – Equalização em imagem de radiograma do tórax. (a) Imagem original. (b) Imagem equalizada globalmente. (c) Imagem equalizada adaptativamente. (d) a (f) Histograma das imagens (a), (b) e (c) respectivamente. Elaborado pelo autor.

Na Figura 2, apesar de ter um histograma mais distribuído ao invés de picos condensados, nota-se a prevalência de tons de cinza mais claros. Nesse caso, para os dois tipos de equalização, observamos uma melhora no contraste da imagem sem deterioração de aspectos importantes.

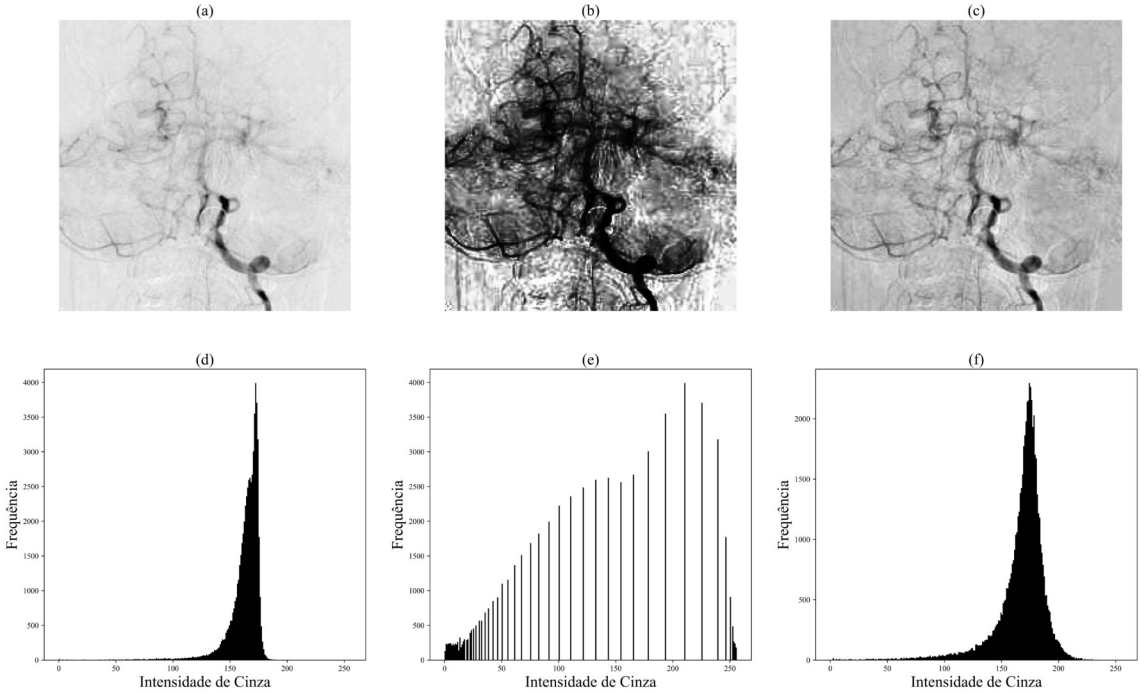

Figura 3 – Equalização em imagem de angiografia. (a) Imagem original. (b) Imagem equalizada globalmente. (c) Imagem equalizada adaptativamente. (d) a (f) Histograma das imagens (a), (b) e (c) respectivamente.Elaborado pelo autor.

Já para a Figura 3, que possui um histograma original bem concentrado, a equalização global causou degradação na imagem com um realce excessivo de muitos tons escuros espalhados. A equalização do tipo adaptativa, nesse caso, produz então um resultado superior para a análise, com os vasos sanguíneos e suas ramificações mais eminentes sem se misturar ao fundo da imagem.

Em conclusão, a equalização de histograma é uma técnica amplamente utilizada no processamento de imagens, que pode ser implementada em várias aplicações, como imagens médicas, visão computacional e reconhecimento de padrões. Embora a equalização do histograma tenha suas limitações, ela continua sendo uma escolha popular para transformar uma imagem em uma representação visualmente mais atraente e informativa.

3.1.2 Filtros de suavização

Uma imagem pode ser afetada por ruído e interferência de várias fontes, como ruído de sensores elétricos, ruído fotográfico e erros de canal, o que pode vir a afetar o diagnóstico de um especialista se, por exemplo, uma pequena estrutura como um tumor, desaparecer da imagem devido ao ruído excessivo. Esses efeitos podem ser reduzidos com o uso

de técnicas tradicionais de filtragem, enquanto proporcionam uma pequena perda de informação (GEDRAITE; HADAD, 2011). Por isso, é fundamental entender os efeitos da filtragem e as situações em que devem ser usadas.

O ruído em imagens geralmente se manifesta como variações discretas de pixels isolados que não são correlacionados espacialmente. Os pixels de erro frequentemente parecem ser visualmente distintos de seus vizinhos (PRATT, 2007).A suavização de ruído é realizada no domínio da frequência, atenuando elementos de faixas altas, denominada filtragem passa-baixa. Os *Low-pass Filters* (Filtros Passa-Baixa) (LPFs) mais comuns são: Ideal, Butterworth e Gaussiano. Essas três categorias cobrem todo o espectro de filtragem, desde muito brusca (Ideal) até muito amenizado (Gaussiano).

Um LPF que deixa passar, sem atenuação, todas as frequências em um círculo de raio D_0 a partir da origem e remove todas as frequências contidas fora é chamado de *Ideal Low-pass Filter* (Filtro Passa-Baixa Ideal) (ILPF). Diferentemente do ILPF, a função de transferência do *Butterworth Low-pass Filter* (Filtro Passa-Baixa Butterworth) (BLPF) com ordem de filtro n, não tem uma descontinuidade abrupta, onde é definido um local de frequência de corte para os quais pixels são reduzidos a uma determinada fração de seu valor máximo. Valores altos de ordem de filtro o aproximam do ILPF, valores baixos o tornam mais parecido com o Gaussiano. O *Gaussian Low-pass Filter* (Filtro Passa-Baixa Gaussiano) (GLPF), possui também corte suave, maior suavização quando comparado aos outros filtros, e não apresenta ruído oscilatório.

A função transferência de cada filtro são definidas pelas equações (2), (3) e (4):

$$H(u, v)_{ILPF} = \begin{cases} 1 & D(u, v) \leq D_0 \\ 0 & D(u, v) > D_0 \end{cases} \tag{2}$$

$$H(u, v)_{BLPF} = \frac{1}{1 + \left[\frac{D(u,v)}{D_0}\right]^{2n}} \tag{3}$$

$$H(u, v)_{GLPF} = e^{-D^2(u,v)/2D_0^2} \tag{4}$$

onde:

D_0	-	Frequência de corte.
$D(u, v)$	-	Distância entre um ponto (u, v) no domínio da frequência e o centro do retângulo de frequência.
n	-	Ordem de filtro.
σ	-	Dispersão ao redor do centro.

As Figuras 4, 5 e 6 demonstram gráficos em perspectiva de H(u, v), e o filtro exibido como uma imagem.

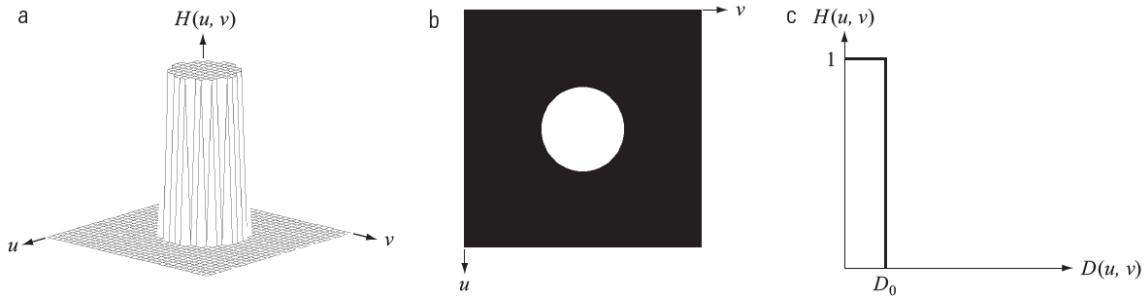

Figura 4 – (a) Gráfico em perspectiva de uma função de transferência de ILPF. (b) Filtro exibido como uma imagem. (c) Corte transversal radial do filtro. Fonte: (GONZALES; WINTZ, 1987).

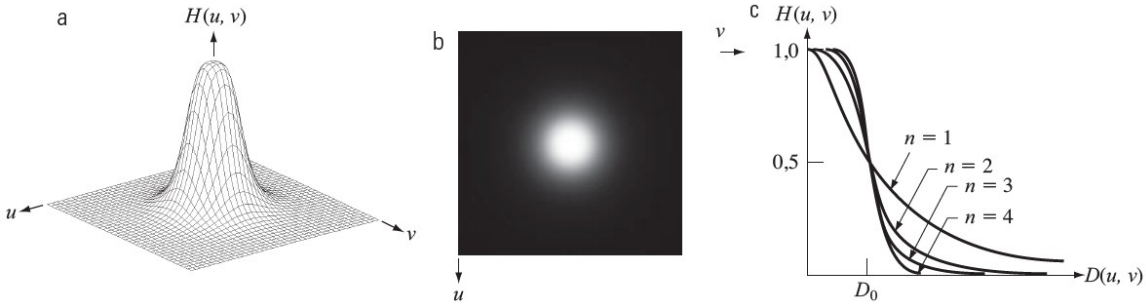

Figura 5 – (a) Gráfico em perspectiva de uma função de transferência de BLPF. (b) Filtro exibido como uma imagem. (c) Cortes transversais radiais do filtro de ordens 1 a 4. Fonte: (GONZALES; WINTZ, 1987).

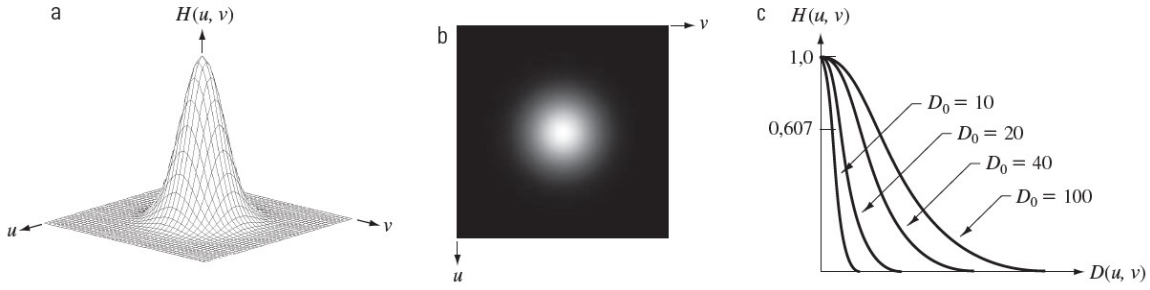

Figura 6 – (a) Um gráfico em perspectiva de uma função de transferência do GLPF. (b) Filtro exibido como uma imagem. (c) Cortes transversais radiais do filtro para vários valores de D_0. Fonte: (GONZALES; WINTZ, 1987).

O principal propósito de apresentar esses filtros foi mostrar como é simples desenvolver e montar filtros no domínio da frequência.Embora os filtros espaciais sejam normalmente utilizados na implementação final de uma solução, como empregado neste trabalho, os conceitos obtidos pela pesquisa no domínio da frequência como guia na seleção de filtros espaciais são extremamente importantes.

3.1.3 Binarização

A binarização é crucial no processamento de imagens digitais, particularmente em aplicações de visão computacional. Devido à sua simplicidade e eficácia, as imagens binárias são úteis em muitos aplicativos de processamento de imagem. A binarização pode ser usada para identificar texto e símbolos em aplicativos como processamento de documentos, identificação de objetos com formas distintas e determinação da orientação de objetos (SEZGIN; SANKUR, 2004). Uma imagem binária é criada reduzindo os níveis de cinza da imagem para dois valores, normalmente 0 e 1.

Limiarização é uma técnica de binarização eficaz, e a escolha da técnica de limiarização usada para binarizar uma imagem é crucial. Para determinar o melhor valor de limiar, vários métodos foram explorados e recomendados. Em muitas aplicações de processamento de imagem, os níveis de cinza dos pixels correspondentes ao objeto diferem significativamente dos níveis de cinza dos pixels pertencentes ao plano de fundo. A limiarização tornou-se uma ferramenta simples, porém eficaz, para distinguir objetos do plano de fundo. Ruído não estacionário e correlacionado, iluminação do ambiente, presença de níveis de cinza aleatórios dentro do objeto e seu fundo, contraste insuficiente e tamanho do objeto fora de proporção com a cena, são fatores que complicam a operação de limiarização.

Em geral, existem duas técnicas para limiarização: limiarização global e limiarização local (LEE; CHUNG; PARK, 1990). O método global sugere que a imagem tem um pico claramente definido ou histograma bimodal e, portanto, o objeto pode ser extraído do plano de fundo usando uma operação simples que avalia os valores da imagem de acordo com um valor limite T. Suponha que temos uma imagem com o histograma representado na Figura 7. Os níveis de cinza nos pixels do objeto são categorizados em um modo dominante. Selecionar um limite T que separa esses modos do resto é um método óbvio para separar o objeto do fundo. A limiarização global é computacionalmente direta e rápida. Ele funciona bem em imagens com objetos com valores de intensidade uniformes em um fundo contrastante.

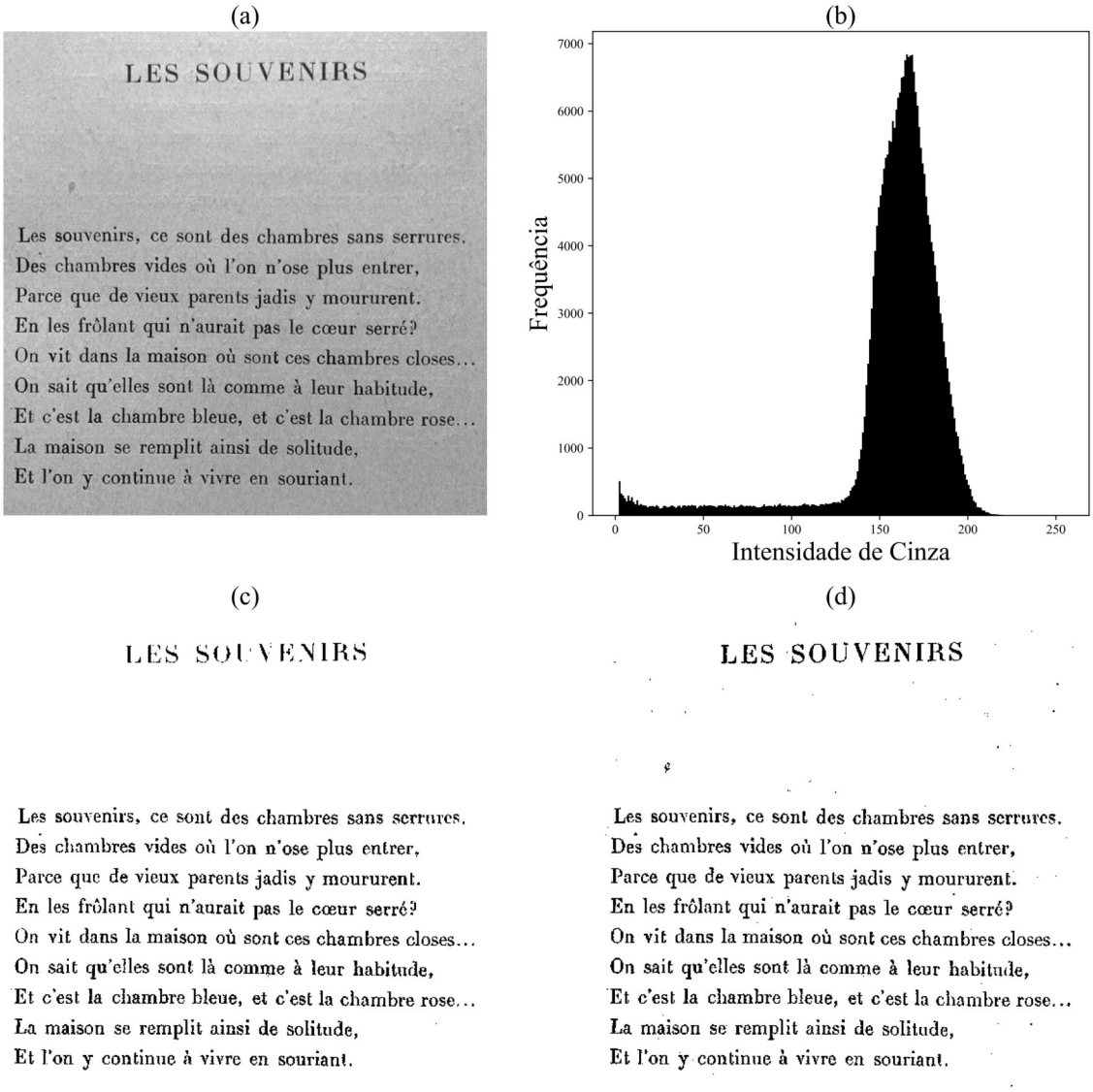

Figura 7 – (a) Imagem original. (b) Histograma da imagem original. (c) Binarização usando limiarização global. (d) Binarização usando limiarização local. Elaborado pelo autor.

Uma das técnicas de binarização global mais referenciadas é o método de Otsu, essencial em problemas de decisão não supervisionados de reconhecimento de padrão. É proposto na perspectiva da análise discriminante, selecionando um limiar ideal automaticamente. Para determinar esse limite, Otsu propôs minimizar a soma ponderada das variações dentro da classe dos pixels do primeiro plano e do plano de fundo. Quando os números de pixels em cada classe estão próximos uns dos outros, esse método produz resultados satisfatórios. No entanto, se o contraste do objeto e do fundo for baixo, a imagem for ruidosa ou o brilho do fundo variar muito na imagem, o processo pode ser dificultado (BANKMAN, 2008).

A limiarização local divide uma imagem em sub-imagens menores, e o limiar para cada sub-imagem é determinado pelas propriedades ou posição local do ponto. Antes do cálculo do limiar, o método de binarização apropriado é escolhido de acordo com uma avaliação das propriedades da imagem local (SAUVOLA et al., 1997). A Figura 8 mostra um exemplo de uma imagem que requer limiarização adaptativa. O tamanho das subimagens usadas deve ser grande o suficiente para incorporar o objeto e os pixels de fundo. Se uma subimagem tiver um histograma bimodal, a distância mínima entre os picos do histograma deve ser usada para determinar um limiar local. Se um histograma for unimodal, o limiar pode ser determinado pela interpolação dos limiares de subimagem próximos. Uma segunda interpolação é necessária na etapa final para identificar os limiares corretos em cada pixel.

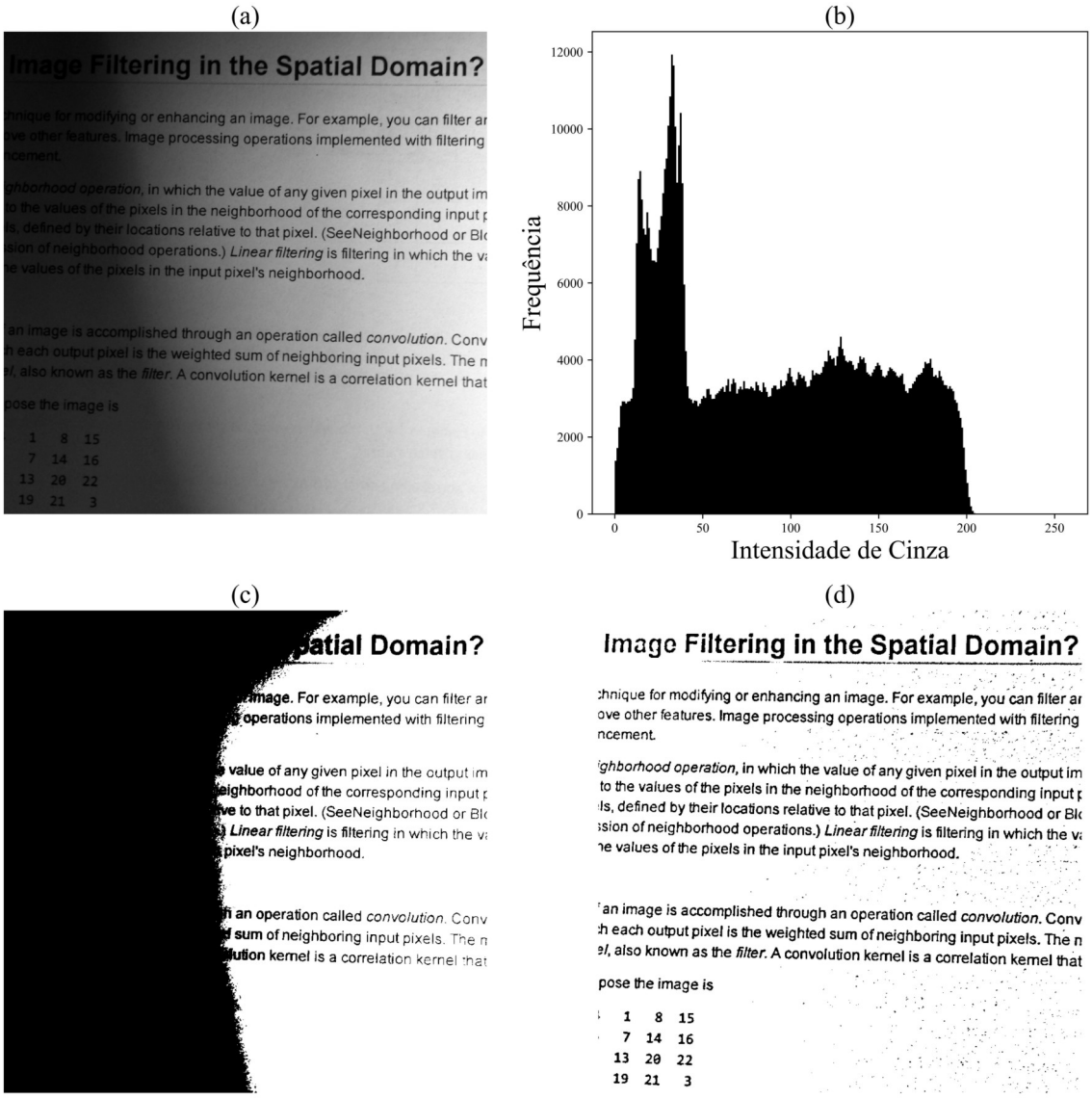

Figura 8 – (a) Imagem original. (b) Histograma da imagem original. (c) Binarização usando limiarização global. (d) Binarização usando limiarização local. Elaborado pelo autor.

3.1.4 Processamento Morfológico

O processamento morfológico é uma abordagem baseada na forma e formato dos objetos. Os métodos morfológicos usam um elemento estrutural para criar uma imagem de saída do mesmo tamanho da imagem de entrada, onde uma comparação do pixel correspondente na imagem de entrada com seus vizinhos determina o valor de cada pixel na imagem de saída. Uma operação morfológica sensível a formas específicas na imagem de entrada é construída variando o tamanho e a forma do elemento estruturante utilizado (SREEDHAR; PANLAL, 2012).

A morfologia, neste contexto, é uma ferramenta para selecionar componentes de imagens que podem ser usados para retratar e descrever a forma de uma região, como limites, esqueletos e fechos convexos. Um dos principais usos da morfologia no caso de imagens binárias é a extração de componentes de imagem úteis na representação de formas, além de preenchimento de buracos, afinamento e espessamento de regiões, que são frequentemente empregados como procedimentos de pré ou pós-processamento em conjunto com esses algoritmos. Deve-se notar que procedimentos morfológicos por si só são insuficientes, por exemplo, erodir sem restrições diminuirá as informações até que elas desapareçam. Para regular o tamanho e a forma final da região, testes devem ser aplicados ao processo de morfologia.

Operações lógicas bit a bit são necessárias para procedimentos morfológicos em imagens binárias. As operações fundamentais da morfologia incluem os operadores booleanos AND e OR. Os procedimentos morfológicos essenciais são dilatação e erosão.A dilatação é um operador booleano OR na morfologia binária, e visa inflar características esparsas para torná-las mais espessas. Já a erosão é um operador booleano AND, e tenta eliminar recursos esparsos até que restem apenas traços fortes (KRIG, 2014).

A dilatação pode ser definida matematicamente e implementada de várias maneiras. A dilatação generalizada é expressa simbolicamente de acordo com a equação (5).

$$G(j,\ k) = F(j,\ k) \oplus H(j,\ k) \tag{5}$$

onde:

- $G(j,\ k)$ - Imagem binária resultante.
- $F(j,\ k)$ - Imagem original com valor binário.
- $H(j,\ k)$ - Elemento estruturante, matriz de tamanho L x L.

Para simplicidade de notação $F(j,\ k)$ e $H(j,\ k)$ são consideradas matrizes quadradas. A erosão generalizada é expressa simbolicamente de acordo com a equação (6).

$$G(j,\ k) = F(j,\ k) \ominus H(j,\ k) \tag{6}$$

Outras duas operações morfológicas cruciais adicionais são abertura e fechamento, e ambas são geradas a partir das ações básicas de erosão e dilatação (PRIYA; NAWAZ, 2017). A operação de abertura morfológica é uma erosão seguida de uma dilatação, com ambas as operações utilizando o mesmo elemento estruturante. A abertura geralmente suaviza o contorno de um objeto, removendo saliências finas. A operação de fechamento morfológico é uma dilatação seguida de uma erosão, e como a abertura, suaviza os contornos ao mesclar descontinuidades estreitas e ao longo de campos finos, eliminando pequenos buracos e preenchendo lacunas em um contorno.

Figura 9 – (a) Imagem original representada por níveis de cinza. (b) Imagem binarizada. (c) Resultado do processo de erosão. (d) Resultado do processo de dilatação. Elaborado pelo autor.

3.1.5 Segmentação

A segmentação de imagens é uma etapa fundamental no aprendizado sobre processamento de imagens, reconhecimento de imagens e visão computacional. Segmentação de imagem é a divisão de uma imagem em várias seções que possuem características semelhantes e diferentes, como cor, intensidade ou textura. Este processo, que normalmente é direto e rápido para o sistema visual humano, pode representar uma barreira significativa no desenvolvimento de algoritmos, e vários métodos para segmentação de imagens foram criados. Observou-se que não existe uma abordagem perfeita para a segmentação de imagens, pois cada imagem possui um tipo único. A escolha de uma técnica de segmentação em detrimento de outra depende principalmente das características peculiares do problema a ser considerado (KHAN, 2013).

A maioria dos algoritmos de segmentação de imagens são classificados em três tipos: segmentação baseada em binarização, segmentação baseada em regiões e segmentação baseada em bordas.

As técnicas de segmentação por binarização dividem as imagens em sub-regiões com base nas mudanças nos valores da escala de cinza. Também pode ser empregado para

separar objetos de primeiro plano e de fundo com base em um valor limite.Uma imagem em tons de cinza pode ser transformada em uma imagem binária. A imagem binária deve conter todas as informações necessárias sobre a localização e forma dos objetos de interesse. A obtenção de uma imagem binária tem a vantagem de reduzir a complexidade dos dados e simplificar a operação de reconhecimento (ABDULATEEF; SALMAN, 2021).

A segmentação com base na região combina crescimento de regiões e a divisão e fusão da região. O crescimento de regiões é uma técnica onde sua abordagem básica é começar com um conjunto de pontos "semente" e, a partir deles, fazer as regiões crescerem anexando a cada semente aqueles pixels vizinhos que têm propriedades predefinidas semelhantes às das sementes (como os intervalos específicos de intensidade ou cor). Essa abordagem, no entanto, tem as seguintes desvantagens: sensibilidade a ruído, lacunas ou descontinuidades na região recuperada e complexidade computacional significativa. O método de divisão e fusão subdivide uma imagem inicialmente em um conjunto de regiões distintas e arbitrárias e, em seguida, funde e/ou sub-divide as regiões em uma tentativa de satisfazer as condições de segmentação.

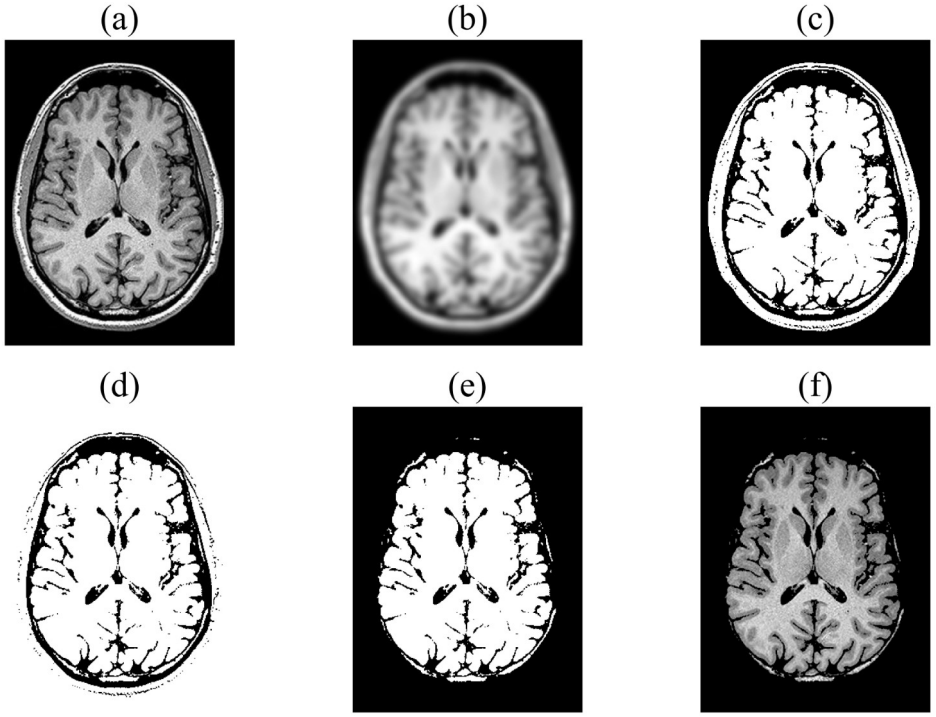

Figura 10 – (a) Imagem original representada por níveis de cinza. (b) Imagem suavizada utilizando GLPF. (c) Imagem binarizada pelo método de Otsu. (d) Primeiro crescimento de regiões adicionando pixels brancos a partir do canto superior esquerdo da imagem até atingir os limites.(e) Segundo crescimento de regiões com pixels pretos. (f) Resultado da operação lógica AND entre (a) e (e). Elaborado pelo autor.

Por fim, a segmentação usando detecção de borda é realizada reconhecendo as bordas da imagem no gradiente da imagem para produzir os limites do objeto. Objetos são

reconhecidos usando este método usando arestas como critério. Os três tipos de características da imagem em que estamos interessados são os pontos isolados, as linhas e as bordas. É comumente usado para identificar itens como o operador de derivada/gradiente de primeira ordem, o operador de segunda derivada e o detector de borda idealOs pixels de borda são pixels com mudanças repentinas na intensidade de uma função de imagem, enquanto as bordas são grupos de pixels de borda conectados.Os detectores de borda são técnicas de processamento de imagens locais distintas para detectar pixels de bordaUma linha pode ser considerada como um segmento de borda no qual o brilho de fundo em ambos os lados da linha é significativamente maior ou muito menor do que a intensidade de pixel da linha. Da mesma forma, cada ponto pode ser representado por uma linha com comprimento e largura iguais a um pixel.A Figura 11 demonstra em (a) e (b) a aplicação da detecção de bordas em uma imagem sem tratamento, e em (c) e (d) a aplicação da detecção em conjunto com a limiarização.

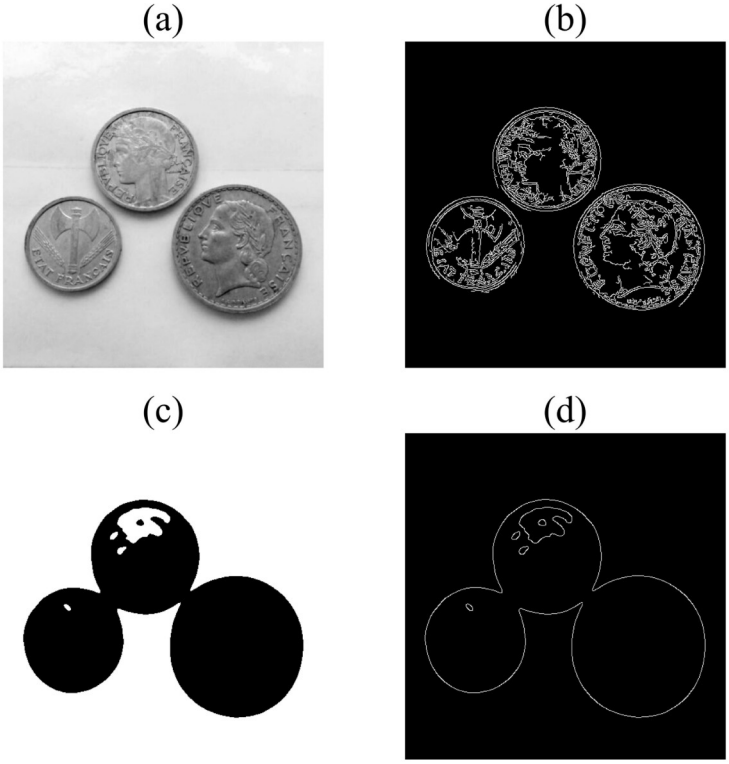

Figura 11 – (a) Imagem original representada por níveis de cinza. (b) Resultado da detecção de bordas na imagem original. (c) Imagem binarizada. (d) Resultado da detecção de bordas na imagem binarizada. Elaborado pelo autor.

3.1.6 Resolução espacial e redimensionamento

A resolução espacial é uma medida do menor detalhe observável em uma imagem e pode ser expressa de várias maneiras, sendo a mais popular a de pontos por unidade de distância (DPI - pontos por polegada) ou pixels por unidade de distância (PPI - pixels por polegada). O número de pontos impressos contidos dentro de uma polegada de uma

imagem produzida por impressora é referido como DPI. O número de pixels contidos dentro de uma polegada de uma imagem digital exibida em uma tela, por outro lado, é referido como PPI. Métricas de resolução espacial devem ser representadas em termos de unidades espaciais para serem significativas. Apenas o tamanho da imagem em si não transmite tudo, pois afirmar que uma imagem tem uma certa resolução faz pouco sentido a menos que as dimensões espaciais da imagem sejam declaradas. O tamanho sozinho pode ser usado para comparar as capacidades dos sistemas de captura de imagem. Por exemplo, uma câmera digital com placa de aquisição de imagens de 20 megapixels captura mais informações do que uma câmera de 8 megapixels, desde que as duas câmeras tenham lentes equivalentes e as fotografias comparativas sejam tiradas à mesma distância.

Qual é a melhor resolução espacial para uma imagem? De acordo com o teorema da amostragem, a frequência espacial mais alta nos dados da imagem precisa ser menor que a metade da frequência de amostragem para que a imagem amostrada replique com precisão a original. No entanto, o teorema da amostragem não é um bom indicador da facilidade com que os programas de computador podem reconhecer objetosMuitas vezes, os objetos podem ser identificados mais facilmente em imagens com uma taxa de amostragem muito baixa porque os cálculos são menores devido à redução na dimensionalidade, e as informações confusas presentes nas versões de alta resolução das imagens podem não existir na resolução reduzida. No entanto, enquanto alguns objetos são mais facilmente encontrados em baixas resoluções, uma descrição de objeto geralmente requer detalhes que só são revelados em resoluções mais altas (BALLARD; BROWN, 1982).

O redimensionamento é definido pelo mapeamento uniforme entre os pixels na imagem de origem e os pixels na imagem final define a escala. O método de redimensionamento mais popular é a interpolação dos pixels da imagem original. Essencialmente, a interpolação é um processo que utiliza dados conhecidos para estimar valores em pontos desconhecidos. Os três métodos de interpolação mais prevalentes são: interpolação do vizinho mais próximo, interpolação bilinear e interpolação bicúbica. Quando métodos de interpolação são usados, o dimensionamento da imagem pode ser feito em tempo real, preservando os efeitos visuais globais. Esses métodos de dimensionamento de interpolação, no entanto, podem introduzir artefatos indesejáveis. Se a proporção da imagem de entrada for claramente diferente da imagem de saída, o dimensionamento gera distorção visível (DIGHE; GURU, 2014).

A interpolação do vizinho mais próximo é a mais simples e requer menos tempo de processamento de todos os algoritmos de interpolação. Sua abordagem seleciona o valor de pixel mais próximo arredondando as coordenadas do ponto de interpolação desejado. Usando este método, encontra-se o pixel correspondente mais próximo na imagem de origem para cada pixel na imagem final (PARSANIA; VIRPARIA et al., 2014). Esta forma de interpolação sofre de efeitos normalmente inaceitáveis tanto para ampliação quanto para redução de imagens, já que tem a tendência de produzir artefatos indesejáveis,

o que pode distorcer linhas retas. Desse modo, raramente é utilizado na prática.

A interpolação bilinear, que usa os quatro vizinhos mais próximos para estimar a intensidade de uma determinada posição, é uma técnica melhor.Com um ligeiro aumento no esforço de computação, a interpolação bilinear produz resultados muito melhores do que a interpolação do vizinho mais próximo. A interpolação bicúbica, que incorpora os 16 vizinhos mais próximos de um ponto, é o próximo grau de complexidade. Em geral, a interpolação bicúbica preserva características mais finas melhor do que a interpolação bilinear. Em softwares comerciais de edição de imagens, a interpolação bicúbica é o padrão.

A Figura 12 mostra o exemplo de uma imagem de 600 x 600 pixels, redimensionada utilizando os três métodos para 60% de seu tamanho original.

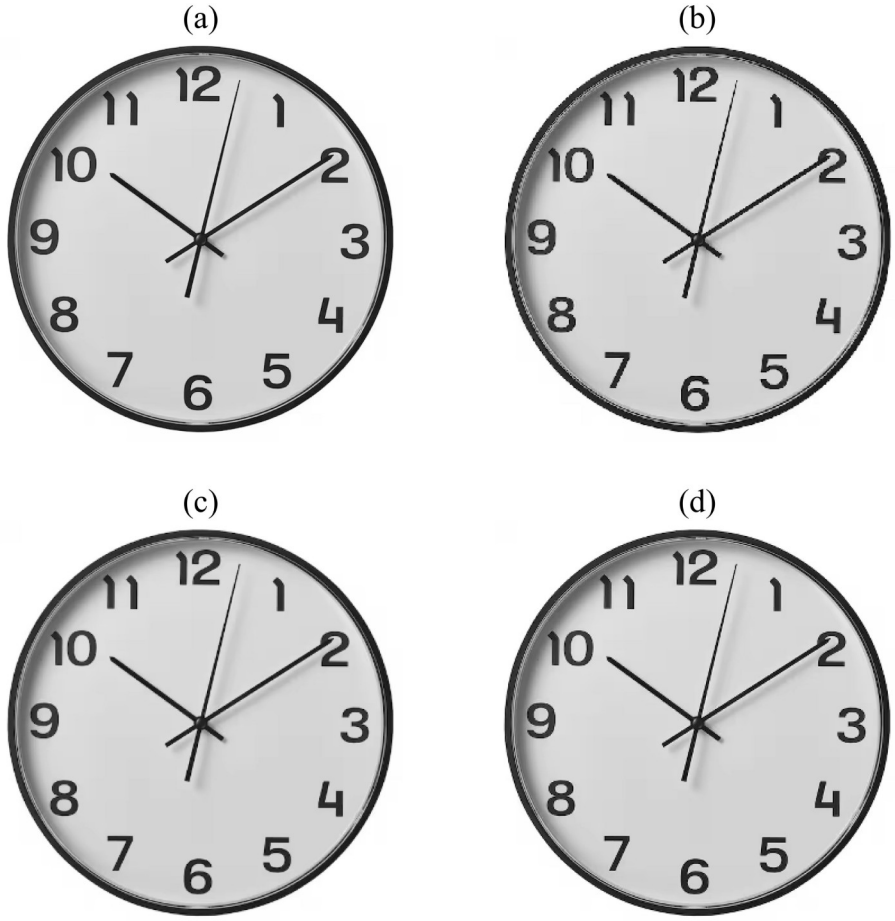

Figura 12 – (a) Imagem original representada por níveis de cinza. (b) Resultado do redimensionamento por interpolação do vizinho mais próximo. (c) Resultado do redimensionamento por interpolação bilinear. (d) Resultado do redimensionamento por interpolação bicúbica. Elaborado pelo autor.

3.2 RNA - Redes Neurais Artificiais

As estruturas de Redes Neurais Artificiais (RNAs) recentemente ganharam destaque como a principal metodologia para análise de imagens, devido às suas grandes capacidades de aprendizado e vantagens em lidar com padrões complicados.A maioria dos algoritmos modernos de aprendizado de máquina inclui um estágio de pré e/ou pós-processamento que é integrado a uma rede neural profunda. Essas fases, baseadas em abordagens clássicas de processamento de imagem, são usadas para ajudar a resolver desafios de classificação, detecção ou segmentação resultantes. Vários estudos mostraram que a incorporação de métodos de pré e pós-processamento em um pipeline de aprendizado profundo pode melhorar o desempenho do modelo quando comparado à rede sozinha.

RNA é uma estrutura de rede de predição e processamento de recursos não lineares de autoaprendizagem. Com base em pesquisas modernas de neurociência, a RNA processa informações imitando a rede neural do cérebro. É organizada em três camadas: uma camada de entrada, camadas ocultas e uma camada de saída. As camadas ocultas são compostas por um determinado número de unidades de nós conhecidas como neurônios. Cada neurônio está ligado a cada unidade de nó na camada anterior. O papel dos neurônios é fazer uma transformação linear e uma transformação não linear nos dados de entrada da camada anterior (HOU et al., 2021).

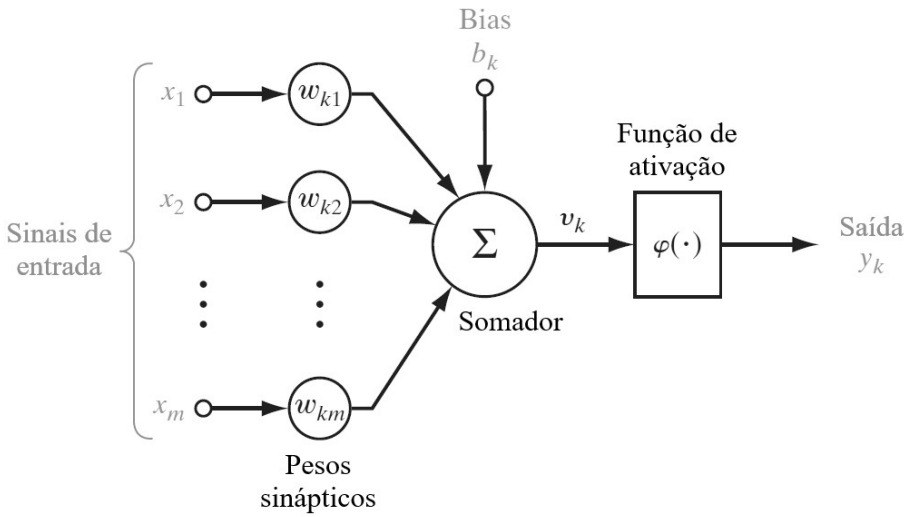

Figura 13 – Modelo não linear de um neurônio k. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

Um neurônio é uma unidade de processamento de informações crucial na operação de uma rede neural. O modelo de um neurônio k é representado no diagrama de blocos da Figura 13. Define-se três características fundamentais do modelo neural:

- 1. Um grupo de sinapses ou elos de conexão, cada um com seu próprio peso ou força. Em particular, o peso sináptico w_{kj} é multiplicado por um sinal x_j na entrada da sinapse j conectada ao neurônio k.

- 2. Um somador para adicionar os sinais de entrada, que são ponderados pelas forças sinápticas do neurônio; os processos descritos aqui criam um combinador linear.

- 3. Uma função de ativação para restringir a magnitude, ou faixa de amplitude, da saída de um neurônio (HAYKIN, 2009).

Em termos matemáticos, podemos descrever o neurônio k representado na Figura 13 escrevendo as equações (7), (8) e (9).

$$u_k = \sum_{j=1}^{n} w_{kj} x_j \tag{7}$$

$$y_k = \phi(u_k + b_k) \tag{8}$$

$$v_k = u_k + b_k \tag{9}$$

onde:

x_j	-	Sinais de entrada.
w_{kj}	-	Pesos sinápticos do neurônio.
u_k	-	Saída do combinador linear.
b_k	-	*Bias* aplicado a um neurônio.
$\phi(\cdot)$	-	Função de ativação.
v_k	-	Potencial de ativação do neurônio k.
y_k	-	Sinal de saída do neurônio.

O *bias* é um elemento que serve para aumentar o grau de liberdade dos ajustes dos pesos, de forma a transladar a função de ativação no eixo.

A topologia dos neurônios de uma rede neural está intimamente ligada ao método de aprendizado usado para treinar a rede. Em geral, existem três grupos essencialmente distintos de arquiteturas de rede:

Redes *feedforward* de camada única: Para neurônios em uma rede neural em camadas, temos uma camada de entrada de nós de origem que se projeta diretamente em uma camada de saída de neurônios, mas não vice-versa. Em certo sentido, esta rede é totalmente *feedforward*, ou direta. A Figura 14 descreve o cenário de quatro nós nas camadas de entrada e saída. Uma rede desse tipo é conhecida como rede de camada única, com o termo "camada única"referindo-se à camada de saída dos nós de computação. A camada de entrada dos nós de origem não é considerada porque não há nenhuma computação nesse estágio.

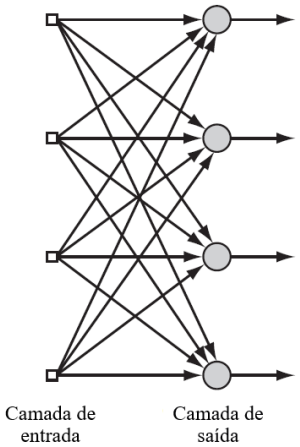

Figura 14 – Rede *feedforward* com uma única camada de neurônios. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

Redes *Feedforward* Multicamadas: Possui uma ou mais camadas ocultas, onde o termo "oculto" refere-se ao fato de que esta parte da rede neural não é visível diretamente da entrada ou da saída da rede. A função dos neurônios ocultos é atuar como uma ponte entre a entrada externa e a saída da rede. A rede pode extrair estatísticas de ordem superior de sua entrada adicionando uma ou mais camadas ocultas. Os nós de origem da rede fornecem elementos correspondentes do padrão de ativação, que compõem os sinais de entrada aplicados aos neurônios na segunda camada. Os sinais de saída da segunda camada são usados como entradas para a terceira camada e assim por diante em toda a rede. O conjunto de sinais de saída dos neurônios na camada de saída representa a resposta geral da rede ao padrão de ativação fornecido pelos nós de origem. A rede na Figura 15 é descrita como uma rede 10-4-2, pois compreende 10 nós de origem, 4 neurônios ocultos e 2 neurônios de saída.

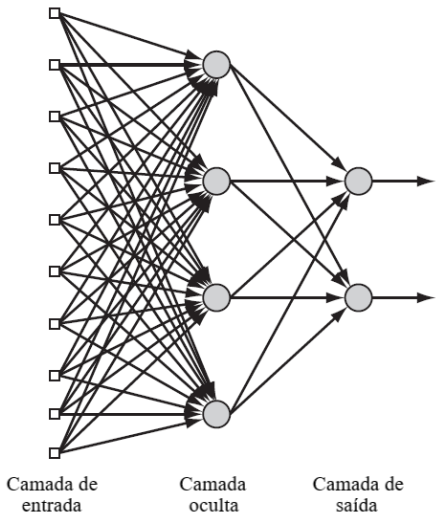

Figura 15 – Rede *feedforward* totalmente conectada com uma camada oculta e uma camada de saída. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

Redes Recorrentes: Uma rede neural recorrente se distingue de uma rede neural *feedforward* pela presença de pelo menos um *loop* de *feedback*. Essa rede pode ser composta de uma única camada de neurônios, com cada neurônio alimentando seu sinal de saída de volta para as entradas de todos os outros neurônios, conforme mostrado no gráfico arquitetônico da Figura 16, ou possuir camadas ocultas também. A inclusão de *loops* de realimentação tem um impacto significativo no potencial de aprendizado e no desempenho da rede. Além disso, os *loops* de realimentação fazem uso de ramificações específicas compostas de elementos de atraso de tempo unitário (denotados por z-1), resultando em comportamento dinâmico não linear, assumindo que a rede neural contém unidades não lineares.

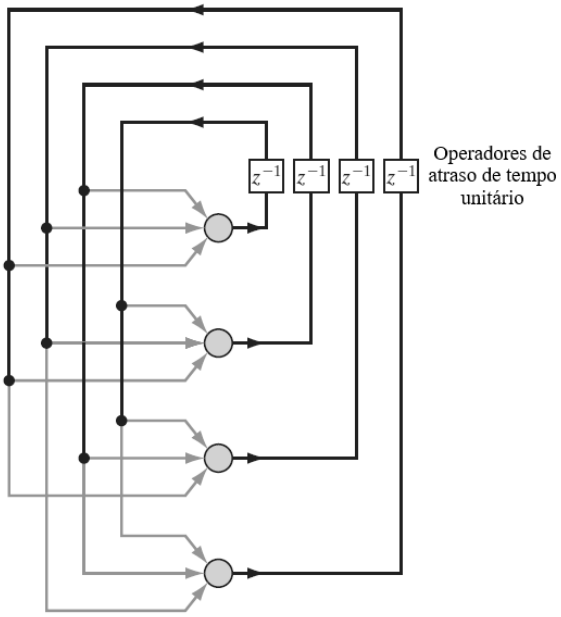

Figura 16 – Rede recorrente sem neurônios ocultos. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

3.2.1 MLP - Multilayer Perceptron

A primeira rede neural descrita algoritmicamente foi o *perceptron*. O *perceptron* é o tipo mais básico de rede neural utilizada para categorização de padrões linearmente separáveis. É basicamente composto por um único neurônio com pesos sinápticos e *bias* configuráveis. O *perceptron* que é baseado em um único neurônio, só pode fazer classificação de padrões com duas classes (hipóteses). Podemos realizar a classificação com mais de duas classes aumentando a camada de saída (computação) do *perceptron* para incluir mais de um neurônio, no entanto, para que o *perceptron* funcione corretamente, as classes devem ser linearmente separáveis.

O *perceptron* é uma rede que opera usando aprendizado de correlação de erro.O procedimento de aprendizado de classificação de padrão é concluído após um determinado número de iterações. Para resolver as restrições práticas do *perceptron*, recorremos a uma

estrutura de rede neural conhecida como *Multilayer Perceptron* (Perceptron de Multicamada) (MLP). Os três pontos a seguir enfatizam as características fundamentais dos MLPs:

- Cada neurônio da rede possui uma função de ativação não linear diferenciável em seu modelo.
- A rede tem uma ou mais camadas que são invisíveis para os nós de entrada e saída.
- A rede possui alto grau de conexão, cujo valor é definido pelos pesos sinápticos da rede.

Um método popular para o treinamento de MLPs é o algoritmo de retropropagação, e decorre em duas fases:

1. Os pesos sinápticos da rede são fixados na fase de avanço, e o sinal de entrada é propagado pela rede, camada por camada, até atingir a saída. Assim, durante esta fase, as mudanças estão limitadas aos potenciais de ativação e saídas dos neurônios da rede.

2. Na fase reversa, um sinal de erro é gerado comparando a saída da rede com uma resposta desejada. O sinal de erro resultante é enviado pela rede novamente, camada por camada, mas desta vez de maneira inversa. Durante a segunda fase, os pesos sinápticos da rede são gradualmente ajustados.

A Figura 17 mostra o gráfico arquitetônico de um MLP com duas camadas ocultas e uma camada de saída.

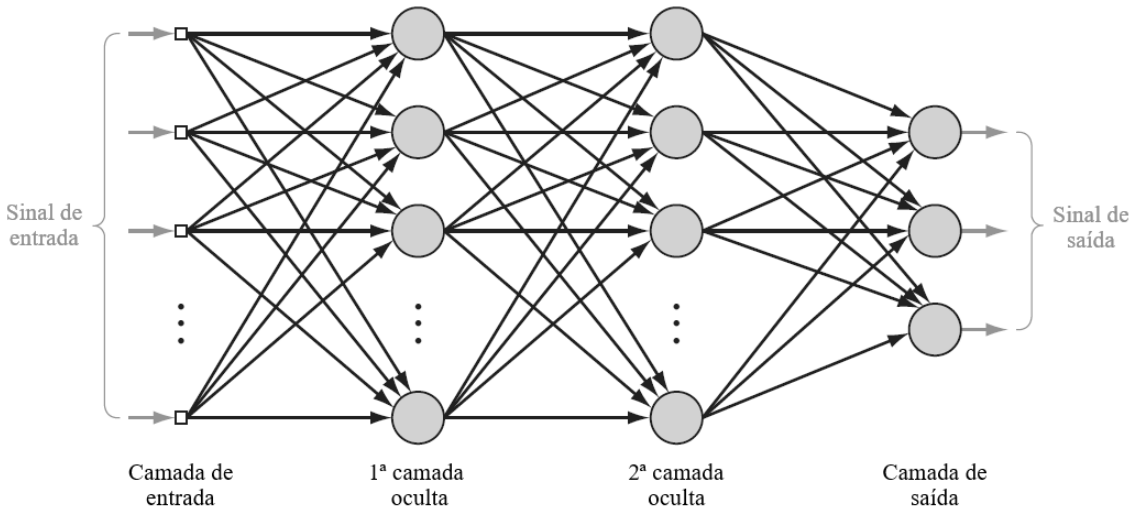

Figura 17 – Gráfico arquitetônico de um MLP com duas camadas ocultas. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

3.2.2 CNN - Convolutional Neural Network

Convolutional Neural Networks (Redes Neural Convolucionais) (CNNs) são baseadas em MLPs projetado especificamente para reconhecer e avaliar formas bidimensionais. Essas redes empregam uma arquitetura única que é adequada para a classificação de

imagens. CNNs podem ser treinadas rapidamente usando esta arquitetura. As CNNs utilizam três conceitos fundamentais: convolução, pesos e *bias* compartilhados e *pooling*.

Convolução: Considere a entrada de um quadrado de neurônios de 28 por 28, cujos valores correspondem às intensidades de 28 x 28 pixels usadas como entradas. Conectaremos os pixels de entrada a uma camada oculta normalmente, mas apenas em partes pequenas e restritas da imagem de entrada. Cada neurônio na primeira camada oculta será vinculado a uma pequena região de neurônios de entrada, como uma zona de 5×5. Essa região é chamada de campo receptivo local (NIELSEN, 2015).Cada conexão recebe um peso e o neurônio oculto recebe um *bias* geral. O campo receptivo local move-se por toda a imagem de entrada, neste caso, um pixel por vez, mas ocasionalmente é utilizado um comprimento de passo diferente. A Figura 18 mostra um campo receptivo local no canto superior esquerdo:

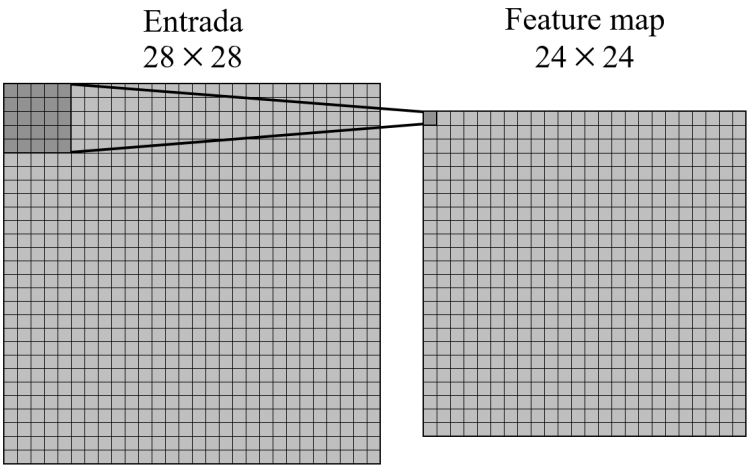

Figura 18 – *Feature map* criado a partir da convolução da primeira camada. Elaborado pelo autor.

Pesos e *bias* compartilhados: pesos e *bias* idênticos são utilizados para todos os neurônios ocultos de tamanho 24 x 24 criados. Ou seja, todos os neurônios da primeira camada oculta identificam a mesma característica.Como resultado, o mapa da camada de entrada para a camada oculta é conhecido como *feature map* (ou mapa de características). Os pesos e *bias* que definem o *feature map* são chamados de pesos e *bias* compartilhados. A Figura 18 descreve uma estrutura de rede que pode detectar apenas um tipo de característica. Para fazer o reconhecimento de imagem é necessário mais de um *feature map*, e assim uma camada convolucional completa consiste em vários *feature maps* diferentes.

Camadas de *pooling* Etapa normalmente empregada imediatamente após as camadas convolucionais para selecionar e filtrar as informações adquiridas, diminuindo o tamanho do modelo, acelerando o cálculo e aumentando a robustez dos recursos extraídos (HOU et al., 2021). Uma camada de *pooling* prepara um *feature map* condensado. Cada unidade na camada de *pooling*, por exemplo, pode resumir uma região de 2×2 neurônios na camada anterior. Depois de agrupar os 24 x 24 neurônios produzidos a partir da

camada convolucional ilustrada no exemplo, temos 12 x 12 neurônios. O agrupamento é aplicado separadamente à cada *feature map*.

A Figura 19 representa a arquitetura de uma CNN, que consiste em uma camada de entrada, quatro camadas ocultas e uma camada de saídaA camada de entrada, que consiste em 28 x 28 nós sensoriais, recebe imagens de vários caracteres que foram aproximadamente centralizados e normalizados em tamanho. Em seguida, os layouts computacionais alternam entre convolução e *pooling*. A camada de saída executa um estágio adicional de convolução, remodelando o *feature map* recebido em um vetor antes de envolver uma ou mais camadas totalmente conectadas para executar a tarefa de classificação final (SALVI et al., 2021). A camada de saída da amostra representada é composta por 26 neurônios, cada um dos quais é alocado para um dos 26 caracteres potenciais.

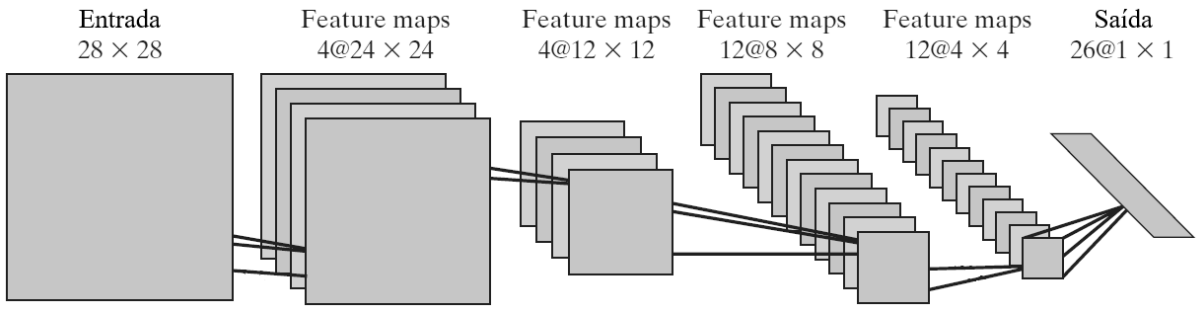

Figura 19 – Rede convolucional para processamento de imagemFonte: (HAYKIN, 2009). Adaptado pelo autor.

3.2.3 Definições e Parâmetros

Os hiper parâmetros representam variáveis que influenciam a topologia da rede e como a rede é treinada, e assim, devem ser definidos antes do treinamento. A seleção de hiper parâmetros tem um impacto significativo no desempenho da CNN, e qualquer pequena alteração em seus valores pode afetar seu desempenho geral. Como resultado, a seleção adequada de parâmetros é um tópico extremamente importante que deve ser considerado durante a criação de esquemas de otimização (ALZUBAIDI et al., 2021)Para prever com precisão, conjuntos de dados distintos exigem conjuntos distintos de hiper parâmetros.

a) Regressão e Classificação

Regressão e classificação são formas de métodos de aprendizado de máquina nos quais um modelo é ensinado usando um modelo existente e dados rotulados corretamente.

Correlações entre variáveis dependentes e independentes são descobertas por meio de regressão. Como resultado, os algoritmos de regressão auxiliam na previsão de variáveis contínuas, como preços de imóveis, movimentos de mercado, padrões climáticos, preços de petróleo e gás, e assim por diante. O objetivo do algoritmo de regressão é identificar a função de mapeamento que nos permitirá mapear a variável de entrada x para a variável de saída contínua y.

A classificação, por outro lado, é um algoritmo que identifica funções que ajudam a separar o conjunto de dados em classes, dependendo de vários fatores. Os algoritmos de classificação determinam a função de mapeamento que converterá a entrada x na saída discreta y. Em outras palavras, os algoritmos de classificação preveem a probabilidade de ocorrência de um evento. Os algoritmos de classificação são usados para uma variedade de propósitos, incluindo classificação de e-mail e spam, previsão do desejo dos clientes do banco de pagar empréstimos e identificação de células tumorais malignas.

Fundamentalmente, a classificação se preocupa em prever um rótulo, enquanto a regressão se preocupa em prever uma quantidade.

b) Validação, treino e conjunto de teste

Em geral, um conjunto de dados maior é necessário para treinar uma rede neural mais profunda e ampla. Assim, para aplicar métodos de aprendizado profundo, amostras suficientes devem ser coletadas e um grande conjunto de dados deve ser preparado antes do processo de treinamento (HOU et al., 2021). Os conjuntos de treinamento, validação e teste são especificados da seguinte forma (RIPLEY, 2007):

- Conjunto de treinamento: Um conjunto de exemplos usados para aprendizado, ou seja, para se adequar aos parâmetros do classificador.

- Conjunto de validação: Uma coleção de amostras usadas para ajustar os parâmetros de um classificador, como o número de unidades ocultas em uma rede neural.

- Conjunto de teste: Um conjunto de exemplo usado apenas para avaliar o desempenho de um modelo já treinado.

Um conjunto de dados de validação é uma amostra de dados do treinamento do modelo que é usado para estimar a habilidade do modelo enquanto ajusta os hiper parâmetros do modelo. O conjunto de dados de validação é diferente do conjunto de dados de teste, que também é retirado do treinamento do modelo, mas é utilizado para fornecer uma avaliação final imparcial da qualidade do modelo ajustado final para comparação ou seleção entre os modelos finais (KUHN; JOHNSON et al., 2013).

A taxa de divisão do conjunto de dados é determinada pela quantidade de amostras no conjunto de dados e no modelo. Ao ajustar vários hiper parâmetros, o modelo de aprendizado de máquina requer um conjunto de validação maior para otimizar o desempenho do modelo. Da mesma forma, se o modelo tiver poucos ou nenhum hiper parâmetro, seria simples validar o modelo com uma coleção pequena de dados. Não existe uma porcentagem de divisão apropriada. Deve-se chegar a um percentual de divisão que atenda às especificações e necessidades do modelo.

c) Número de neurônios e camadas

O número de neurônios na camada de entrada é igual ao número de características nos dados. O número de neurônios na saída é determinado pelo tipo de modelo. Para modelos de classificação a camada de saída terá um único neurônio ou vários neurônios,

dependendo do rótulo de classe do modelo. Já para modelos de regressão, a camada de saída conterá apenas um único neurônio.

Quando se trata de camadas ocultas, as redes neurais com 1 a 2 camadas ocultas podem funcionar bem se os dados forem menos complexos e tiverem menos dimensões ou características. Se os dados tiverem várias dimensões ou características, 3 a 5 camadas ocultas podem ser empregadas para obter melhores resultados.

Existem vários métodos de regra prática para estimar o número ideal de neurônios a serem utilizados nas camadas ocultas, incluindo os seguintes: O número de neurônios ocultos deve ser proporcional ao tamanho das camadas de entrada e saída; O número de neurônios ocultos deve ser 2/3 do tamanho da camada de entrada mais o tamanho da camada de saída; O número de neurônios ocultos deve ser menor que o dobro do tamanho da camada de entrada.

d) Função de ativação

Nas RNAs, as funções de ativação são usadas para transformar um sinal de entrada em um sinal de saída, que é então enviado como entrada para a próxima camada.A soma dos produtos das entradas e seus pesos correspondentes é calculada, depois aplicamos uma função de ativação a ela para obter a saída dessa camada e a fornecemos como entrada para a próxima camada (SHARMA; SHARMA; ATHAIYA, 2017).

Existem diferentes tipos de funções de ativação. Atualmente, as mais utilizados nas CNNs são (NARANJO-TORRES et al., 2020):

Função Unidade Linear Retificada (ReLU): ReLU é a função de ativação mais utilizada para camadas de convolução.É uma função meio-retificada.É matematicamente definida como:

$$f(x) = max(0, x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{10}$$

Função sigmoide: sua curva parece uma forma de S. A função varia entre 0 e 1, portanto é usada para prever uma probabilidade como saída.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

Função tangente hiperbólica (tanh): a função tanh tem forma semelhante à função sigmoide, mas o intervalo está entre -1 e 1. A vantagem é que os valores zero serão mapeados próximos de zero e os valores negativos serão mapeados fortemente negativos. Sua definição matemática é:

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{12}$$

A Figura 20 demonstra os gráficos das funções de ativação.

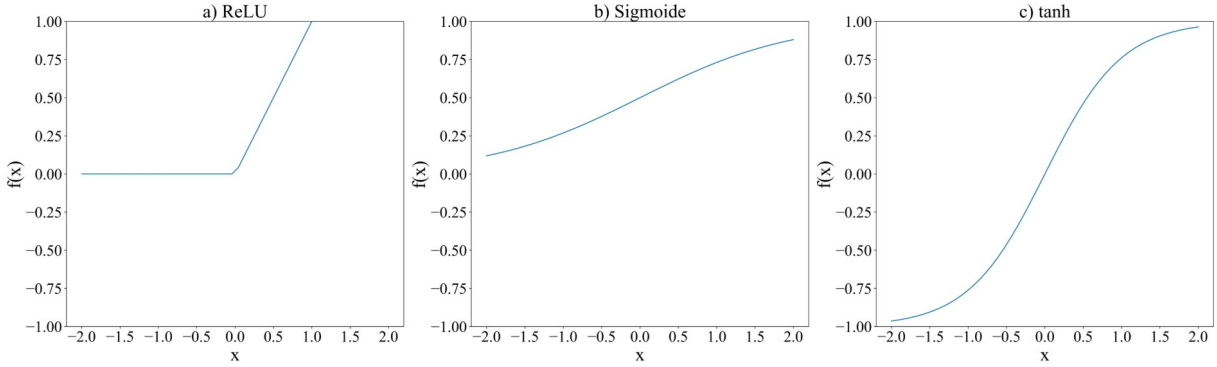

Figura 20 – Gráfico das funções de ativação. Elaborado pelo autor.

e) ***Batch*** e Épocas

O *batch size* especifica quantas imagens serão utilizadas no procedimento de treinamento. Definir esse hiper parâmetro muito alto pode fazer com que a rede demore muito para atingir a convergência, sem aumento adicional de precisão, e enquanto isso, configurá-lo muito baixo pode fazer com que a rede salte para frente e para trás sem atingir um desempenho aceitável (KANDEL; CASTELLI, 2020). Além disso, devido à complexidade dos conjuntos de dados médicos, a natureza do conjunto de dados pode afetar o tamanho do *batch*. *Batch sizes* populares incluem 32, 64 e 128 amostras. Se o conjunto de dados não for dividido uniformemente pelo *batch size*, significa simplesmente que o *batch* final tem menos amostras do que os outros.

A quantidade total de épocas é um hiper parâmetro que especifica quantas vezes o algoritmo de aprendizado será executado em todo o conjunto de dados de treinamento. Uma época significa que cada amostra no conjunto de dados de treinamento teve uma oportunidade de atualizar os parâmetros do modelo interno. Cada época é composta por um ou mais *batches*. Tradicionalmente, o número de épocas é enorme, frequentemente centenas ou milhares, permitindo que o procedimento de aprendizado continue até que o erro do modelo seja adequadamente minimizado. Na literatura, são vistos exemplos do número de épocas definidas como 10, 100, 500, 1000 e maiores.

É comum criar gráficos de linha que mostram épocas ao longo do eixo x e o erro do modelo no eixo y. Esses gráficos às vezes são chamados de curvas de aprendizado, e podem ajudar a diagnosticar se o modelo aprendeu demais, aprendeu pouco ou se está adequadamente ajustado ao conjunto de dados de treinamento.

O erro diminui à medida que o número de épocas de treinamento aumenta. Começa com um valor alto, diminui rapidamente e depois diminui progressivamente à medida que a rede se aproxima de um mínimo local na superfície de erro. Se olharmos apenas para a curva de aprendizado do treinamento, é bastante difícil determinar quando é ideal interromper o treinamento. A estratégia de treinamento de *early-stopping* é então aplicada.O processo de treinamento é interrompido regularmente e a rede é testada no subconjunto

de validação após cada intervalo de treinamento. O treinamento é encerrado após o erro nos dados de validação atingir a saturação.

A Figura 21 mostra formas de duas curvas de aprendizado, uma pertencente às medições no subconjunto de treino e a outra pertencente ao subconjunto de validação.

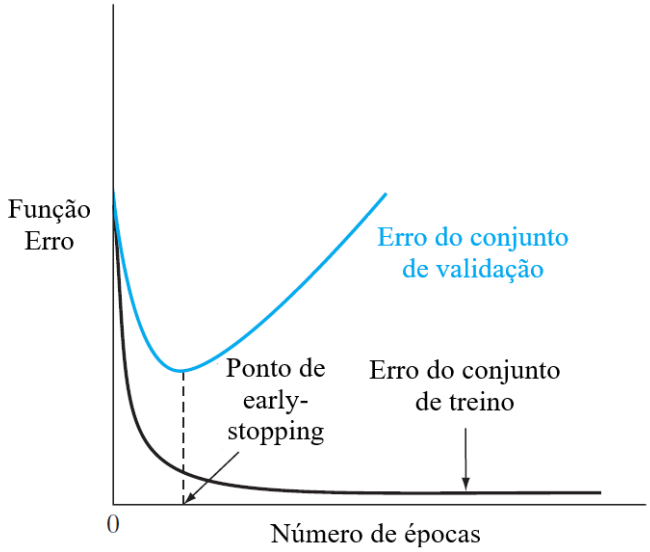

Figura 21 – Ilustração da regra de early-stopping. Fonte: (HAYKIN, 2009). Adaptado pelo autor.

f) Taxa de aprendizado

A taxa de aprendizado é um hiper parâmetro que especifica quanto o modelo deve mudar em resposta ao erro previsto cada vez que os pesos do modelo são atualizados, que tem um pequeno valor positivo, geralmente entre 0,0 e 1,0.Determina a rapidez com que o modelo se adapta à situação. Taxas de aprendizado menores requerem mais épocas de treinamento devido às mudanças menores nos pesos a cada atualização, enquanto taxas de aprendizado maiores produzem mudanças rápidas e requerem menos épocas de treinamento.

Escolher a taxa de aprendizado é difícil, pois um número muito baixo pode resultar em um longo processo de treinamento que pode travar, mas um valor muito alto pode resultar no aprendizado de um conjunto de pesos abaixo do ideal muito rapidamente ou em um processo de treinamento instável.

Ao construir uma rede neural, a taxa de aprendizado pode ser o hiper parâmetro mais importante (GOODFELLOW; BENGIO; COURVILLE, 2016). Como resultado, é fundamental entender como examinar os efeitos da taxa de aprendizado no desempenho do modelo e desenvolver a intuição sobre a dinâmica da taxa de aprendizado no comportamento do modelo.

g) Métricas de performance

O desempenho de redes profundas para classificação de imagens geralmente é avaliado pelo cálculo da acurácia. A acurácia geral é uma métrica comum utilizada em problemas de classificação e é definida como a razão entre as imagens classificadas corretamente e o número total de imagens (SALVI et al., 2021), conforme a equação (13).

$$Acurácia = \frac{\text{Número de classificações corretas}}{\text{Número de amostras}} \tag{13}$$

Como um modelo preditivo de regressão prevê uma quantidade, o desempenho do modelo deve ser relatado como um erro nessas previsões. Existem muitas maneiras de estimar o erro de um modelo preditivo de regressão, mas as mais comuns são:*Root mean squared error* (Raiz quadrada do erro-médio) (RMSE), *Mean absolute percentage error* (Erro percentual absoluto médio) (MAPE) e *Mean absolute error* (Erro médio absoluto) (MAE), definido pelas equações (14), (15) e (16).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\text{Valor previsto} - \text{Valor real})^2}{N}} \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\text{Valor real} - \text{Valor previsto}}{\text{Valor real}} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\text{Valor real} - \text{Valor previsto}| \tag{16}$$

Capítulo 4

Metodologia do Trabalho

4.1 Introdução

Este capítulo descreve o método escolhido para realizar a predição da idade óssea das imagens de radiografia, detalhando os recursos utilizados, escolhas de pré-processamento e arquitetura das RNAs desenvolvidas. A Figura 22 retrata resumidamente as etapas desenvolvidas.

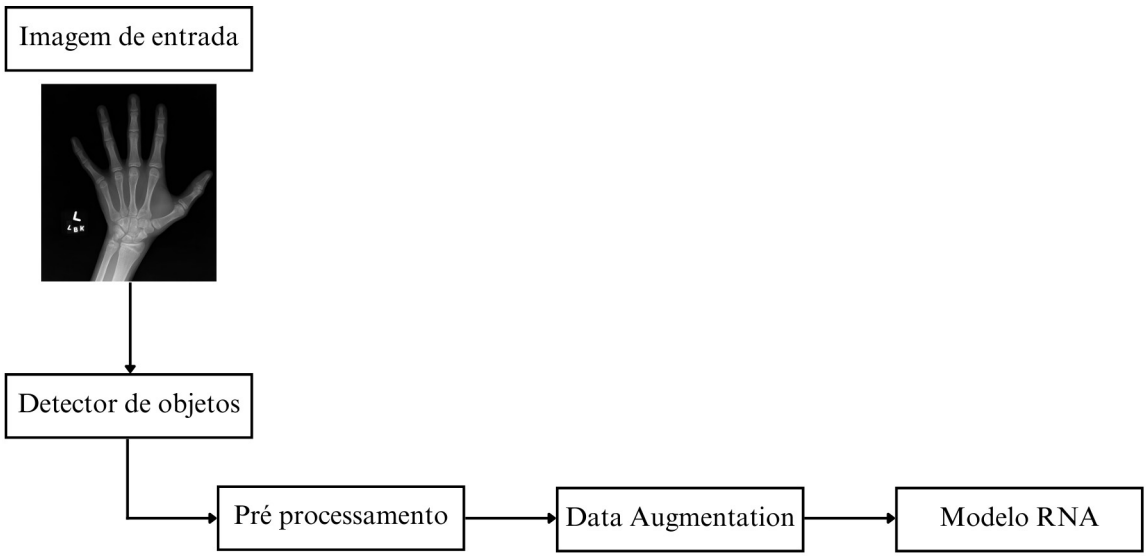

Figura 22 – Etapas de desenvolvimento do algoritmo. Elaborado pelo autor.

4.2 Banco de dados

Para fins de pesquisa, a RSNA - Radiological Society of North America (Sociedade de Radiologia da América do Norte) (RSNA, 2017) publicou um conjunto de dados de fotografias da mão direita de jovens entre 0 e 20 anos. O acervo contém 12611 imagens de radiografia. Junto à base de imagens, é fornecido também uma tabela informando a

idade óssea aferida por radiologistas e o gênero. Na Figura 23 é apresentado um gráfico de frequência da idade óssea em anos para melhor visualização.

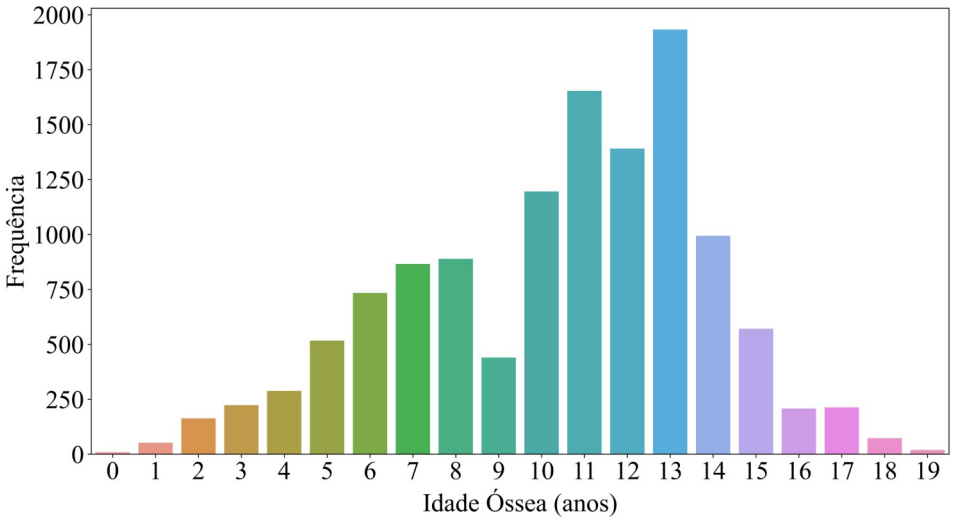

Figura 23 – Distribuição da base considerando a idade óssea. Elaborado pelo autor.

A aparência das radiografias no banco de dados varia muito devido aos diferentes métodos de aquisição utilizados na captura. O brilho, o contraste, a resolução e até a proporção das imagens podem diferir umas das outras. A Figura 24 mostra algumas das imagens de amostra do conjunto de dados.

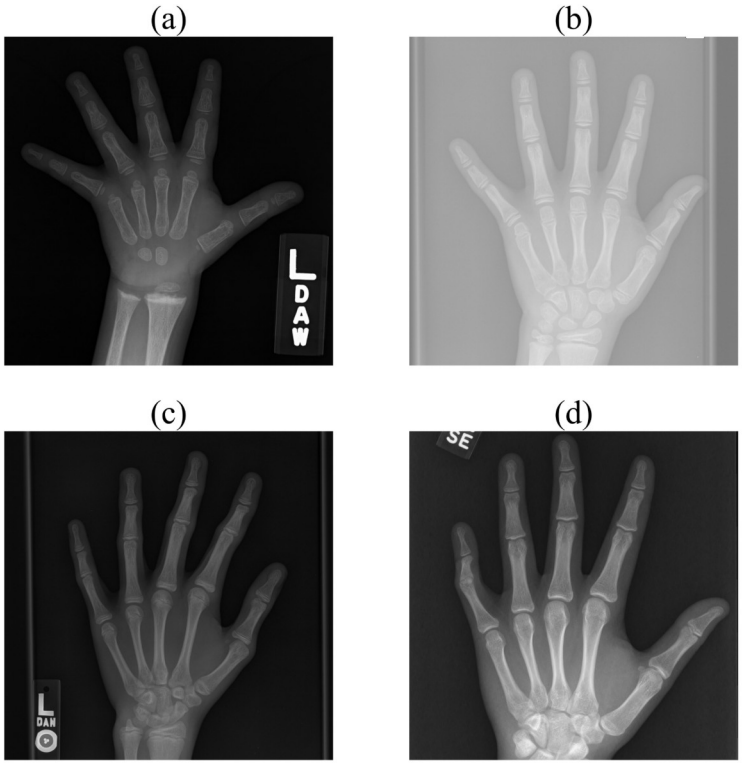

Figura 24 – Exemplares de radiografias disponíveis na base. a) Paciente com 2 anos. b) Paciente com 7 anos. c) Paciente com 14 anos. d) Paciente com 18 anos. Elaborado pelo autor.

Os dados de imagem de raios-X usados para apoiar as conclusões deste artigo foram depositados no repositório RSNA em doi:10.1148/radiol.2018180736.

4.3 Softwares e Bibliotecas

4.3.1 Python

Python é uma linguagem de programação orientada a objetos com alto nível de abstração, sendo facilmente aprendida por iniciantes. Tem uma sintaxe simples que imita a linguagem natural, e é versátil, podendo ser usado desde para o desenvolvimento de aplicativos Web até aprendizado de máquina. É de código aberto, o que significa que é gratuito para usar e distribuir, mesmo para fins comerciais. Seu arquivo de módulos e bibliotecas - pacotes de código que usuários de terceiros criaram para expandir os recursos do Python - é vasto e crescente.

4.3.2 Anaconda

O Anaconda é uma plataforma gratuita que permite escrever e executar código na linguagem de programação Python. Simplifica a implantação e o gerenciamento de pacotes, além de portar muitas bibliotecas e pacotes desde sua instalação. Como o Anaconda é gratuito e de código aberto, qualquer pessoa pode contribuir para o seu desenvolvimento.

4.3.3 Jupyter Notebook

Jupyter Notebook é um IDE - *Integrated Development Environment* (Ambiente de Desenvolvimento Integrado) de código aberto que permite criar e compartilhar documentos que contêm código ativo, equações, visualizações e texto, que já vem instalado junto ao Anaconda.

4.3.4 Numpy

NumPy - Numerical Python, é uma biblioteca de código aberto, e é o padrão universal para trabalhar com dados numéricos em Python. A biblioteca NumPy contém matrizes multidimensionais e estruturas de dados de matrizes, podendo ser usado para executar uma ampla variedade de operações matemáticas, fornecendo uma enorme biblioteca de funções matemáticas de alto nível. Como as imagens também podem ser consideradas compostas de matrizes, o NumPy pode ser usado para executar diferentes tarefas de processamento de imagens.

4.3.5 Matplotlib

Matplotlib é uma biblioteca para visualização de dados e plotagem gráfica, tais como histogramas, gráficos de linha, dispersão, barras entre outros, além de gráficos 3D, para Python e sua extensão numérica NumPy. É excepcionalmente rápido em uma variedade de operações. Além disso, ele pode exportar visualizações para todos os formatos de imagem populares.

4.3.6 Glob

O módulo Glob é uma parte útil da biblioteca padrão do Python. Abreviação de global, é usado para retornar todos os caminhos de arquivo que correspondem a um padrão específico, sendo capaz de ler pastas inteiras e retornar todos os arquivos contidos.

4.3.7 OpenCV

OpenCV - *Open Source Computer Vision Library* (Biblioteca de Visão Computacional de Código Aberto) - é uma biblioteca Python que permite executar tarefas de processamento de imagem e vídeo e visão computacional. Ele fornece uma ampla gama de recursos, incluindo detecção de objetos, reconhecimento facial e rastreamento.

4.3.8 Pandas

Pandas é uma biblioteca amplamente usada para manipulação, pré-processamento de dados e aprendizado de máquina. A biblioteca oferece alto desempenho, facilita o uso de estruturas de dados, ferramentas de análise e manipulação de tabelas numéricas e séries temporais. Permite importar e exportar dados de vários formatos de arquivos de tabelas e consultas à banco de dados.

4.3.9 Tensorflow

TensorFlow é uma biblioteca de aprendizado de máquina de código aberto desenvolvida pelo Google, e facilita o processo de aquisição de dados, treinamento de modelos e refinamento de resultados, além de possuir uma ampla biblioteca de modelos pré-treinados que podem ser usados. O TensorFlow permite que os desenvolvedores criem gráficos de fluxo de dados - estruturas que descrevem como os dados se movem por um gráfico ou uma série de nós de processamento - onde cada nó no gráfico representa uma operação matemática.

4.3.10 Keras

Keras é uma ferramenta de aprendizado profundo de alto nível incorporado no TensorFlow usado para facilitar a implementação de redes neurais.Contém inúmeras implementações de blocos de construção de rede neural comumente usados, como camadas, funções de ativação, otimizadores e uma série de ferramentas para facilitar o trabalho com dados de imagem e texto para simplificar a codificação necessária para escrever código de RNA.

4.3.11 Scikit-learn

Scikit-learn é uma biblioteca em Python que fornece muitos algoritmos de aprendizado não supervisionados e supervisionados, construído sobre NumPy, pandas e Matplotlib. Inclui funcionalidades de regressão (linear e logística), classificação, seleção de modelos, pré-processamento e normalização de dados, além de agrupamento automático de dados semelhantes em conjuntos de dados.

4.3.12 CVAT

O CVAT é uma ferramenta de anotação de imagens de código aberto originalmente desenvolvida pela Intel e agora mantida pela OpenCV. O CVAT fornece recursos de rotulação para tarefas de detecção, classificação, rastreamento e segmentação de objetos. A variedade de opções de rotulagem do CVAT oferece diferentes maneiras de rotulação seus dados para uma determinada finalidade do projeto.

4.3.13 Roboflow

Roboflow é uma estrutura de desenvolvedor de visão computacional para melhor coleta de dados para pré-processamento e técnicas de treinamento de modelos.O Roboflow tem conjuntos de dados públicos prontamente disponíveis para os usuários e também permite que os usuários carreguem seus próprios dados personalizados, sendo uma ferramenta prática para realizar treinamento de RNAs de detecção. Roboflow aceita vários formatos de anotação. No pré-processamento de dados, há etapas disponíveis como reorientações de imagem, redimensionamento, contraste e *Data Augmentation*.

4.4 Pré Processamento

4.4.1 Segmentação de regiões

A primeira etapa do processo de pré-processamento será a segmentação das regiões de interesse, importante para eliminar estruturas indesejáveis e fornecer como entrada do modelo de RNA apenas dados considerados importantes. Para esse processo, primeiro foi

desenvolvido com o auxílio de softwares um modelo de detecção de objetos, onde foram rotuladas 80 imagens com as classes dedo e pulsoA rotulação foi realizada com o software CVAT, sendo possível anotar as imagens de maneira simples e rápida, exportando ao final as anotações no formato desejado. O processo de rotulação manual e geração de caixas delimitadoras pode ser visto na Figura 25.

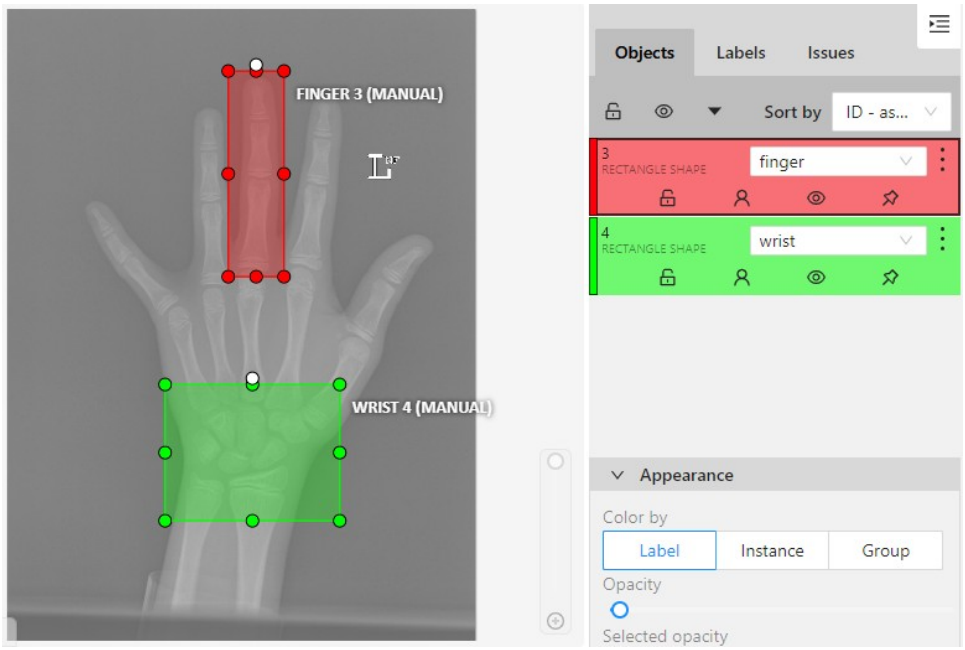

Figura 25 – Interface do CVAT para rotulação de imagens. Elaborado pelo autor.

Esses arquivos de rotulação junto com suas radiografias correspondentes são usados para treinar a rede neural de detecção. Para treinar o modelo, é utilizado o software Roboflow. A Figura 26 apresenta sua interface.

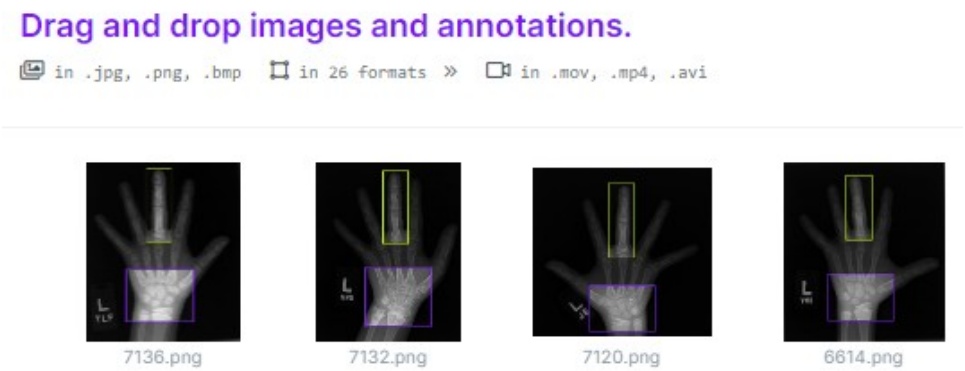

Figura 26 – Interface do Roboflow para entrada de imagens e anotações. Elaborado pelo autor.

Este permite realizar pré-processamento nos dados, tais como redimensionamento para treinamento mais eficaz e técnicas de *Data Augmentation*. Para redimensionamento, o tamanho escolhido foi de 640 x 640, e para a geração de novas imagens, cada uma das 80

imagens rotuladas irá gerar mais 2, com variações em rotação de até $\pm 20^{o}$ e ajuste no brilho de até $\pm 10\%$, sendo uma etapa muito importante devido à alta variabilidade do conjunto de dados. A Figura 27 mostra o resultado do modelo final de detecção e a opção de já realizar testes fazendo o upload de radiografias, e a Figura 28 mostra o resultado de uma das detecções.

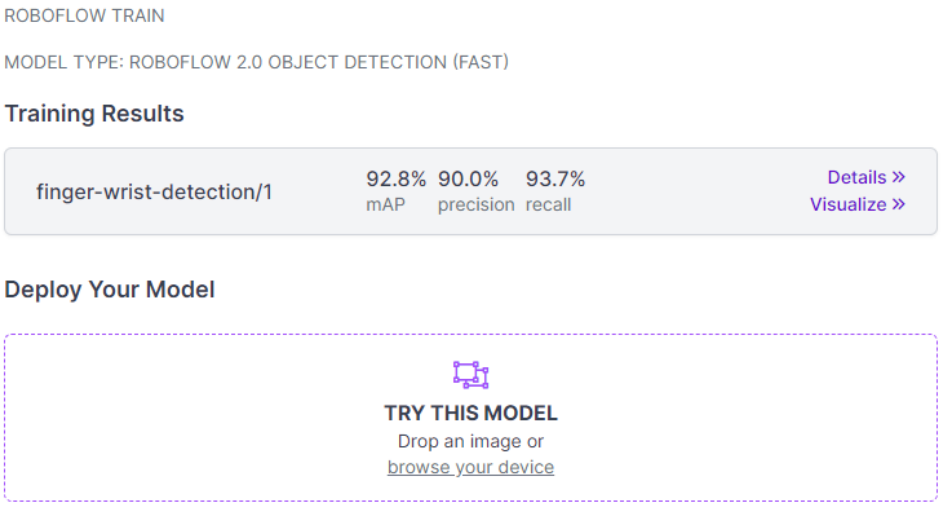

Figura 27 – Interface do Roboflow para resultado do modelo de detecção de objetos.Elaborado pelo autor.

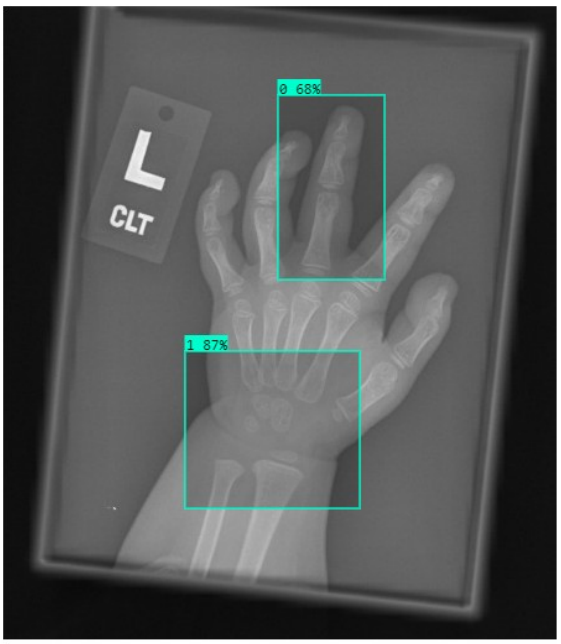

Figura 28 – Detecção de regiões do modelo. Elaborado pelo autor.

Após a finalização do modelo, este fica armazenado na plataforma do Roboflow, e é facilmente implementado no Jupyter Notebook, onde será então usado para realizar a detecção de todas as imagens presentes no banco de dados utilizada nesse trabalho.Para

cada detecção, o modelo retorna a classe e a informações como localização dos pontos x e y, que representam o centro da caixa delimitadora, e a altura e largura da caixa, permitindo assim extrair as regiões de interesse das imagens, como visto na Figura 29.


```
{'predictions': [{'x': 787,
   'y': 1273,
   'width': 656,
   'height': 588,
   'confidence': 0.8228829503059387,
   'class': '1',
   'image_path': '12905.png',
   'prediction_type': 'ObjectDetectionModel'},
  {'x': 799,
   'y': 403,
   'width': 261,
   'height': 700,
   'confidence': 0.7510044574737549,
   'class': '0',
   'image_path': '12905.png',
   'prediction_type': 'ObjectDetectionModel'}],
 'image': {'width': '1522', 'height': '1712'}}
```


Figura 29 – Saída de dados do modelo de detecção. Elaborado pelo autor.

O conjunto de treinamento e validação inicial contém um total de 12611 imagens. Algumas destas foram eliminadas devido à erros de detecçãoA Tabela 4 mostra a relação entre a base inicial e a base final considerando os erros e a acurácia da detecção. Para o modelo de mosaico, juntou-se os erros da segmentação de dedo e de pulso para filtrar as imagens detectadas erroneamente, resultando em 180 eliminadas, pois haviam apenas 2 erros na base do Modelo 2 que não havia errado no Modelo 1 também.

Tabela 4 – Relação de detecção para a base de treino e validação

Modelo	Base inicial	Erros de detecção	Base final	Acurácia
Modelo 1 - Dedo	12611	178	12433	98,58%
Modelo 2 - Pulso	12611	16	12595	99,87%
Modelo 3 - Mosaico	12611	-	12431	-

No conjunto de teste, existem 1423 imagens, e seguindo a mesma dinâmica para a base de treinamento, a Tabela 5 detalha os resultados de detecção.

Tabela 5 – Relação de detecção para a base de teste

Modelo	Base inicial	Erros de detecção	Base final	Acurácia
Modelo 1 - Dedo	1423	21	1402	98,52%
Modelo 2 - Pulso	1423	0	1423	100%
Modelo 3 - Mosaico	1423	-	1402	-

Cada região segmentada do conjunto de treinamento é submetida a algumas técnicas de aprimoramento de imagem, que serão descritas nos subitens em sequência.

4.4.2 Alteração nos canais de cor

As imagens do banco de dados possuem canal de cor *Red - Green - Blue* (Vermelho - Verde - Azul) (RGB), então se realiza a conversão para canal de tons de cinza, variando em 256 níveis, de 0 – preto total – à 255 – branco total. A importância dessa transformação se dá pelas seguintes vantagens:

- Redução de dimensão - imagens RGB possuem três canais de cores e três dimensões, enquanto as imagens em tons de cinza são unidimensionais;
- Redução da complexidade do modelo: Para imagens RGB de 10x10x3 pixels, a camada de entrada terá 300 nós de entrada, mas a mesma rede neural precisará de apenas 100 nós de entrada para imagens em tons de cinza;
- Utilização de algoritmos: muitos algoritmos são personalizados para funcionar apenas em imagens em tons de cinza, por exemplo, equalização de histograma e binarização.

4.4.3 Ajuste de contraste

Uma das técnicas mais aplicadas para ajuste de contraste e brilho em imagens é a equalização de histograma, que pode ser realizado de forma global ou adaptativa. A equalização de histograma adaptativa gera muitos histogramas, cada um correspondendo a uma área diferente da imagem, e os utiliza para redistribuir os valores de brilho da imagem. Como resultado, é apropriado para fortalecer o contraste local e aprimorar as definições de borda em cada seção de uma imagem, sendo muitas vezes mais efetivo do que a equalização global. Porém, tem propensão a amplificar o ruído em áreas relativamente homogêneas de uma imagem.

A utilização de *Contrast Limited Adaptive Histogram Equalization* (Equalização de Histograma Adaptativa Limitada por Contraste) (CLAHE) supera esse efeito restringindo a amplificação. O algoritmo CLAHE é dividido em três seções: geração de blocos, equalização de histograma e interpolação bilinear. A imagem de entrada é primeiro seccionada. O histograma é então equalizado para cada parte usando um limite de corte predefinido. Os valores de intensidade no histograma que excedem o limite de corte são agregados e

dispersos para outros valores. As seções geradas são unidas usando interpolação bilinear para fornecer uma imagem de saída mais contrastada.

4.4.4 Filtro de suavização

Método de filtragem aplicado às imagens, no sentido de se excluir os ruídos presentes na imagem, com a finalidade de diminuir a diferença entre a intensidade dos pixels, de modo que apresentem uma uniformidade. É aplicado antes da binarização para garantir uma separação do fundo mais uniformizada. A técnica de suavização escolhida para o tratamento das imagens foi o Filtro Gaussiano, baseado no GLPF, de tamanho 25 x 25.

4.4.5 Binarização

A binarização é uma das mais importantes abordagens para a segmentação de imagens, visando a extração de objetos que se encontram em uma imagem, mesclada com o fundo. Neste trabalho a binarização será utilizada para separar a mão do fundo da imagem. Como as imagens variam em relação aos níveis de cinza do fundo, o método de limiarização aplicado foi o de Otsu, que encontra o valor do limiar automaticamente.

4.4.6 Processamento Morfológico

Realiza-se a subtração entre a imagem da região segmentada pré-processada e a binarizada, também expressa pela operação booleana AND, atuando como um filtro, fazendo com que apenas as áreas da imagem original correspondente às regiões em branco da imagem binarizada sejam conservadas na imagem final.A Figura 30 é uma representação das técnicas descritas nas últimas seções.

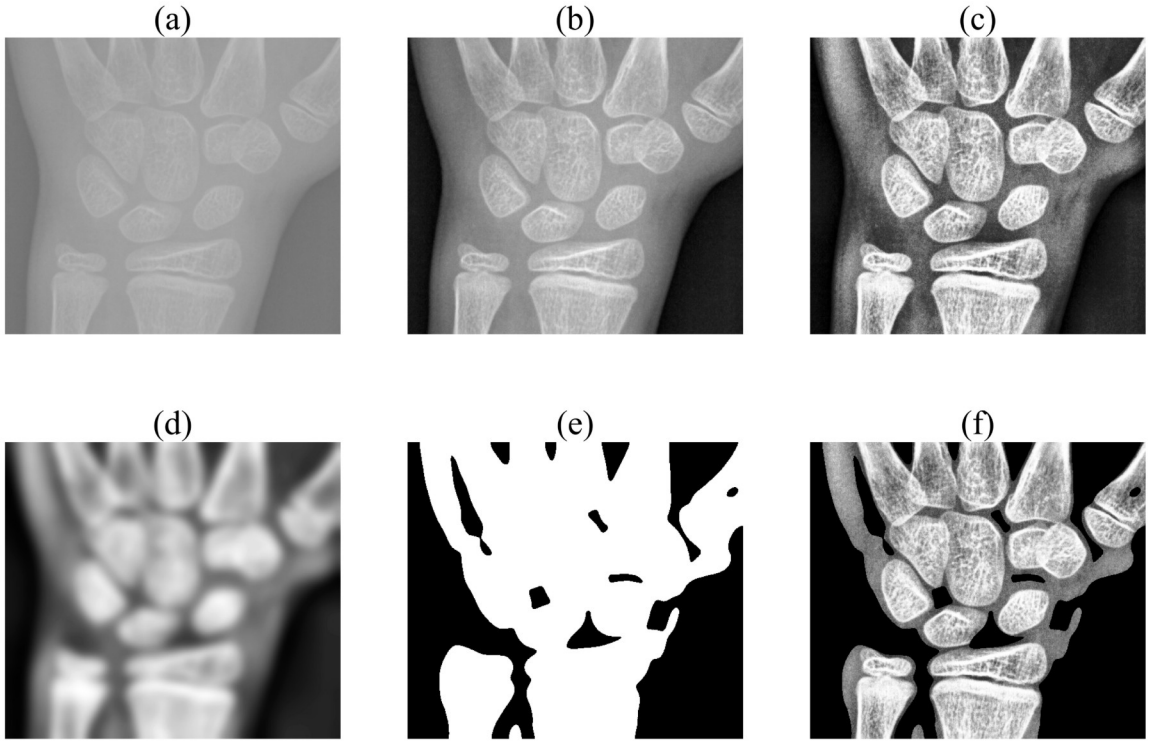

Figura 30 – Etapas de pré processamento. a) Região segmentada original. b) Representação em níveis de cinza. c) Ajuste de contraste. d) Aplicação de filtro de suavização. e) Binarização. f) Processamento morfológico entre a imagem com brilho ajustado e a imagem binarizada. Elaborado pelo autor.

4.4.7 Redimensionamento

As radiografias nem sempre têm a mesma altura e largura, e é fundamental não alterar a proporção de uma radiografia porque isso pode afetar as propriedades avaliadas pelos radiologistas. Como resultado, reduzimos nossa imagem de radiografia enquanto mantemos a proporção para que a maior dimensão seja igual ao tamanho pretendido, e um preenchimento de zeros são então usados para completar a dimensão menor. Embora o redimensionamento implique em perda de informações, ele também minimiza o número de pesos a serem aprendidos, facilitando a convergência.Quanto maiores as dimensões, mais lenta se dá a implantação, porém melhor é o desempenho.Quanto menores as dimensões, o tempo de implantação é reduzido enquanto o pré-processamento é prejudicado. Logo, deve-se alcançar um equilíbrio buscando melhor performance. Como resultado, o tamanho final da imagem para o primeiro modelo é altura de 150 pixels x largura de 50 pixels, 100 x 100 para o segundo modelo e 100 x 150 para o terceiro modelo.

4.4.8 Data Augmentation

Data Augmentation é comumente empregado para expandir o conjunto de dados de imagem, ao mesmo tempo em que torna a rede mais resistente à variâncias.É o processo de duplicar coleções de imagens originais invertendo, girando, ampliando e alterando o brilho. Neste trabalho, primeiro fazemos rotação aleatória na faixa de 30 graus e zoom de 20%. Além disso, usamos inversões horizontais aleatórias e deslocamentos de largura e/ou altura de 10% das dimensões da imagem. Os procedimentos de *Data Augmentation* é realizado apenas para amostras de idade óssea em meses que possuem menos de 80 exemplares, dessa forma, produzindo cópias variadas até que se atinja o valor de 80 imagens. A Tabela 6 mostra o número de imagens distintas produzidas pela combinação de distorções nas imagens de treinamento. Podemos ver que as distorções aleatórias aumentam significativamente o tamanho da base para treinamento. A Figura 31 exibe uma imagem original seguida de três imagens obtidas artificialmente.

Tabela 6 – Base final para treino com Data Augmentation.

Modelo	Base inicial	Imagens geradas	Base final
Modelo 1 - Dedo	12433	9219	21652
Modelo 2 - Pulso	12595	9228	21823
Modelo 3 - Mosaico	12431	9218	21649

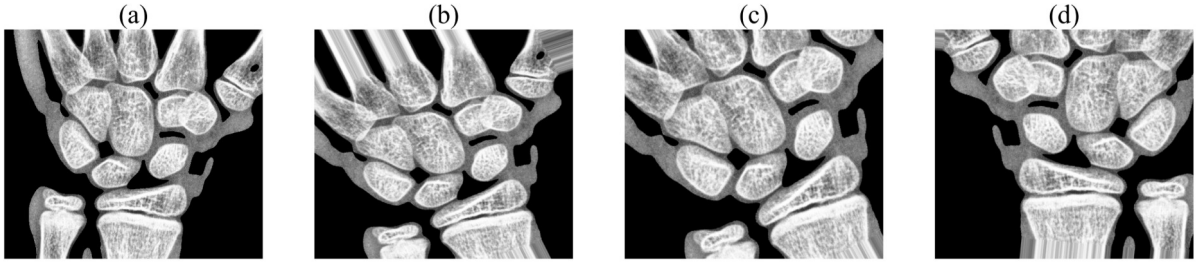

Figura 31 – Resultados do uso de Data Augmentation. a) Imagem original. b), c) e d) Imagens geradas automaticamente. Elaborado pelo autor.

4.5 Arquitetura da Rede Neural

Dados mistos, que incluem dados numéricos, categóricos e visuais, podem ser processados usando uma variedade de modelos de aprendizado de máquina. Em dados mistos, dados numéricos ou categóricos são utilizados para análise de regressão, enquanto dados de imagem são usados para classificação, o que ajuda na previsão precisa. Devido à natureza heterogênea dos dados, combinar dados mistos em um único modelo é muito mais difícil, e diferentes pré-processamentos são necessários para uma execução de rede única de ponta a ponta. Uma RNA complexa é necessária para lidar com o modelo de múltiplas entradas.

Para dados de imagem, a CNN é particularmente eficiente, enquanto conjuntos de dados categóricos ou numéricos são mais bem tratados por MLP simples. O banco de dados contém imagens segmentadas de radiografias de mãos, e a característica categórica de sexo é reunido em um formato de arquivo de dados tabulares, que é posteriormente associado junto com as imagens. O gênero foi considerado relevante, tendo visto que a maturação óssea é mais avançada nas mulheres do que nos homens da mesma idade cronológica (CAVALLO et al., 2021).

O processo de análise é triplo: imagem, características e concatenação. A CNN manipula a seção de processamento de imagens, enquanto a MLP lida com a seção de categorização. No estágio final, os resultados são integrados usando uma rede neural para fornecer a saída final prevista. A Figura 32 mostra em resumo o fluxo da rede completa.

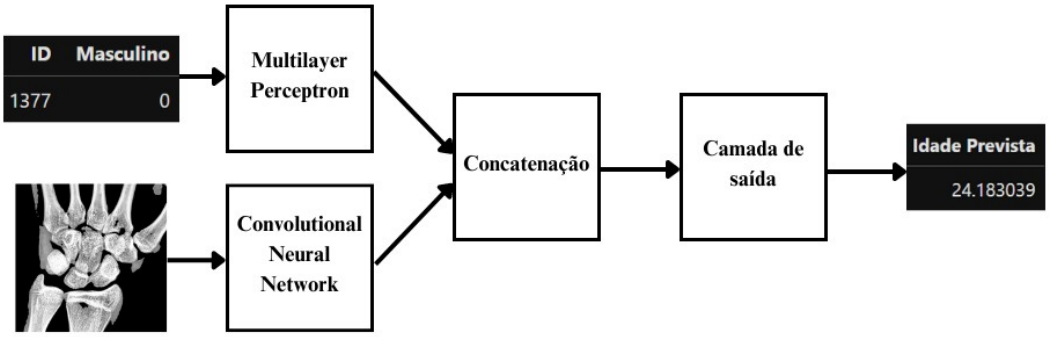

Figura 32 – Fluxo da RNA. Elaborado pelo autor.

A rede MLP lida com o dado categórico de sexo, dado por 0 ou 1, sendo 0 feminino e 1 masculino. A ativação ReLU ativa uma camada densa totalmente conectada, seguida por uma camada oculta totalmente conectada, de dimensões arbitrárias para cada modelo que será definida pelo algoritmo otimizador detalhado mais à frente, enquanto a ativação linear é usada para a camada de saída de regressão ideal final.

CNN é um bom classificador preditivo, mas deve enquadrar-se em um certo intervalo de valores para ser usado para regressão. Todos os valores de intensidade da imagem são manipulados por várias redes neurais convolucionais e o conjunto de imagens é dividido por 255, o valor máximo da imagem, para dimensioná-lo entre 0 e 1.

A rede aplica uma convolução em três fases e performa a função de ativação ReLU antes de entrar em cada camada de *pooling*. Cada convolução tem três tipos diferentes de filtro, ou seja, 16, 32 e 64, onde cada filtro tem o mesmo tamanho de 3x3 em *loop*. O resultado é alimentado em uma camada de ativação de ReLU seguida pela camada de *pooling* máximo, aplicada para reduzir a dimensão espacial do tamanho de entrada. O tamanho de *pooling* é usado duas vezes do tamanho 2x2 para diminuir o tamanho de entrada das imagens e acrescentar *feature maps*.

O vetor de saída da CNN é acoplado com o vetor de saída do MLP no último estágio para formar um vetor de estrutura regressiva. A estrutura é processada usando a arquitetura de CNN + MLP, utilizando ativação ReLU para obter uma saída linear.A última camada linear é aplicada para fornecer previsão de saída numéricaA Figura 33 apresenta as camadas da rede desenvolvida.

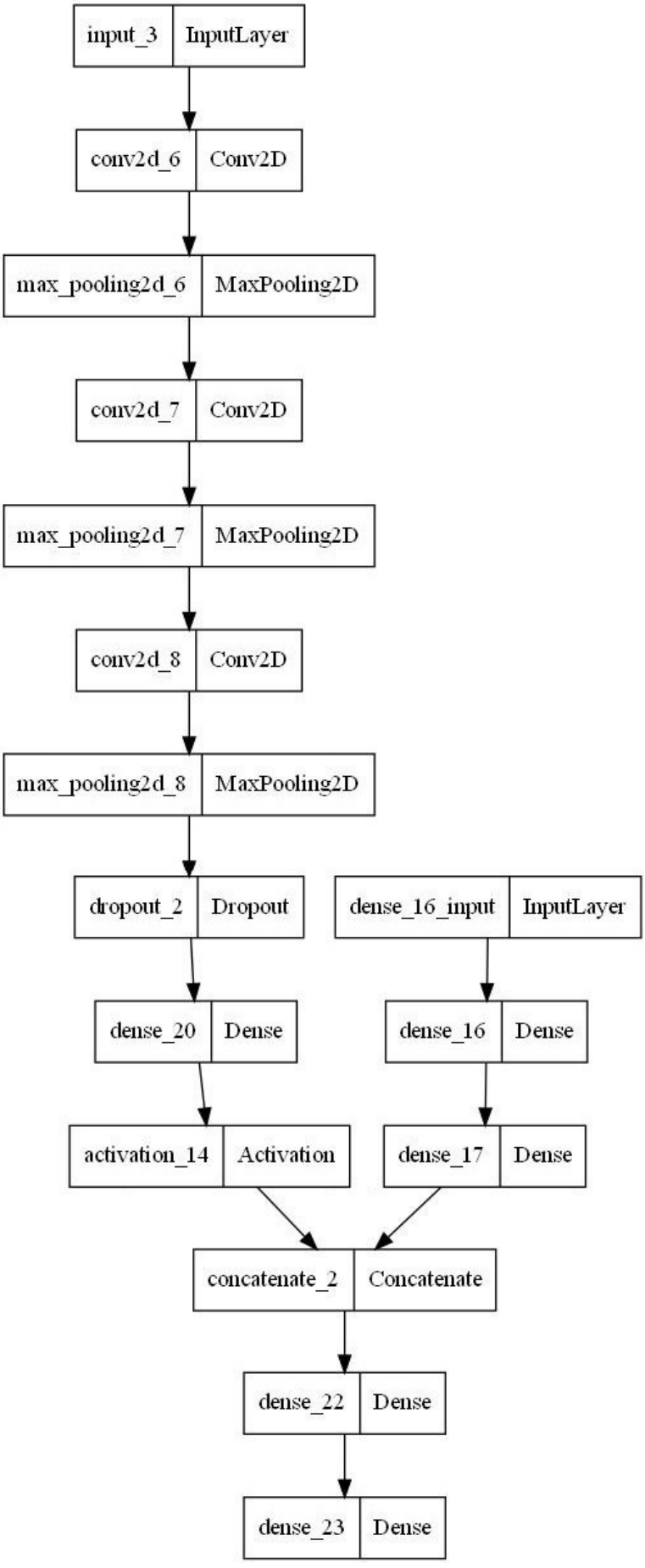

Figura 33 – Fluxo da RNA. Elaborado pelo autor.

O desempenho dos algoritmos de RNA depende especificamente da seleção de um conjunto sólido de hiper parâmetros, e automatizar a otimização desses parâmetros pode tornar a construção de algoritmos complexos mais eficiente e menos demorada. Keras inclui algumas técnicas de otimização de hiper parâmetros, das quais foi escolhida a Hyperband, um algoritmo simples, versátil e teoricamente sólido.

É um método de *early-stoppping* baseado em princípios que aloca um recurso predefinido, como iterações ou amostras de dados, para configurações amostradas aleatoriamente e, em seguida, treina o modelo com cada configuração, interrompendo as configurações de treinamento com baixo desempenho ao alocar recursos adicionais a configurações promissoras. Hyperband faz uso significativo de reduções pela metade sucessivas, que atribui uma provisão a um conjunto de configurações de hiper parâmetros. Isso é feito de forma consistente e, uma vez gasto o recurso, metade das configurações é descartada com base no desempenho. As 50% melhores são retidas e ensinadas com uma nova provisão, enquanto as 50% restantes são descartadas. O processo continua até que reste apenas uma configuração (LI et al., 2017). O número de testes em épocas cumulativas é determinado aproximadamente pela equação (17), e a configuração do otimizador é feita de acordo com a Figura 34 .

$$épocas = max\_epochs \left(\log_{f\ actor} max\_epochs\right)^2 \tag{17}$$

```
tuner = kt.Hyperband(model_builder,
                     objective='val_accuracy',
                     max_epochs=10,
                     factor=3,
                     directory='my_dir',
                     project_name='intro_to_kt')
```

Figura 34 – Inicialização do algoritmo de otimização. Fonte: (TENSORFLOW, 2022).

Como o conjunto de dados de treinamento fornecido não possui um conjunto de dados de treinamento e validação separado, 25% das imagens no conjunto de dados são escolhidas aleatoriamente e usadas como um conjunto de dados de validação. O conjunto de dados de validação ajuda a ajustar os hiper parâmetros para descobrir os melhores modelos após cada época. As épocas de treinamento (max_epochs) foram definidas em 50, o fator foi mantido como padrão e o objetivo, por se tratar de um modelo de regressão, foi val_mape, o valor do MAPE para os dados de validação de cada iteração.

4.6 Método para Avaliação das Predições

Três arquiteturas de rede profunda foram treinadas usando porções distintas de imagens - dedo, pulso e um mosaico de ambos - para avaliar como os ossos afetam o de-

sempenho do modelo. Por fim, comparamos previsões de outros modelos e avaliamos o desempenho geral de nossa abordagem. Todos os modelos treinados são testados usando o conjunto de teste para ver qual obteve melhor performanceA métrica geral de medição de erro usada será o MAE - Mean absolute error (Erro médio absoluto), a fim de melhor comparação entre os métodos de regressão avaliados na seção de trabalhos correlatos.

Será gerado também um gráfico de dispersão para cada modelo desenvolvido, útil para verificar como dois conjuntos de dados comparáveis concordam entre si.Neste caso, uma linha é desenhada como uma referência, e quanto mais os conjuntos de dados concordarem, mais os pontos dispersos tendem a se concentrar ao redor da linha.

Adicionalmente, será fornecido os valores de MAPE - Mean absolute percentage error (Erro percentual absoluto médio) dos modelos, usado para medir o processo de previsão de análise de desempenho, em que se o valor de MAPE for inferior a 10%, os resultados da previsão são considerados muito bons, se o valor de MAPE estiver entre 10% e 20%, os resultados da previsão são considerados bons (LEWIS, 1982). O valor da precisão de previsão com base no valor MAPE é descrito na Tabela 7.

Tabela 7 – Valor MAPE para avaliação de previsão.

MAPE (%)	**Avaliação**
MAPE $\preceq 10\%$	Previsão de alta precisão
$10\% <$ MAPE $\preceq 20\%$	Boa previsão
$20\% <$ MAPE $\preceq 50\%$	Previsão razoável
MAPE $> 50\%$	Previsão imprecisa

Capítulo 5

Resultados

5.1 Resultados de treinamento

Para a estimação da idade óssea foram testados três métodos de regressão:o primeiro modelo contendo imagens radiográficas de um dedo de cada mão, o segundo modelo contendo o pulso, e o terceiro contendo um mosaico dessas duas imagens, como demonstrado na Figura 35.

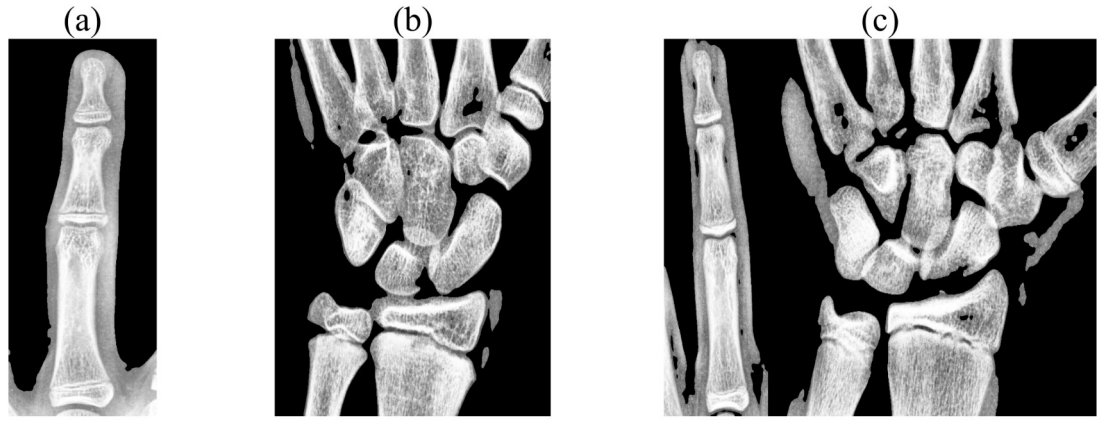

Figura 35 – Imagens de entrada dos modelos. Elaborado pelo autor.

Para cada método foram treinadas redes com conjuntos de hiper parâmetros diferentes, utilizando o algoritmo Hyperband a fim de selecionar o conjunto mais adequado, ou seja, o que fornece o menor valor de MAPE - Mean absolute percentage error (Erro percentual absoluto médio) do conjunto de validação.

Os três modelos foram programados para treinar por 50 épocas, com taxa de aprendizado de 0,001 para os Modelos 1 e 2 e 0,01 para o Modelo 3 definido pelo otimizador. *Early-stopping* com uma paciência de 5 épocas foi empregada. A Tabela 8 mostra os valores de dimensões das camadas definidos pelo otimizador.

Tabela 8 – Dimensões das redes CNN e MLP.

Modelo	Camada 1 Convolucional	Camada 2 Convolucional	Camada 3 Convolucional	Pooling Final	Saída CNN Vetorizada	MLP
Modelo 1	(150, 50, 16)	(75, 25, 32)	(37, 12, 64)	(18, 6, 64)	6912	480
Modelo 2	(100, 100, 16)	(50, 50, 32)	(25, 25, 64)	(12, 12, 64)	9216	352
Modelo 3	(100, 150, 16)	(50, 75, 32)	(25, 37, 64)	(12, 18, 64)	13824	416

A curva de convergência do treinamento da rede neural nos três métodos aproximou-se do MAPE do conjunto de validação indicado na Tabela 9. A relação entre o erro no conjunto de treino e de validação e as épocas treinadas é mostrada na Figura 36.

Tabela 9 – Resultados de validação.

Modelo	MAPE (%)
Modelo 1	8,16
Modelo 2	7,54
Modelo 3	6,58

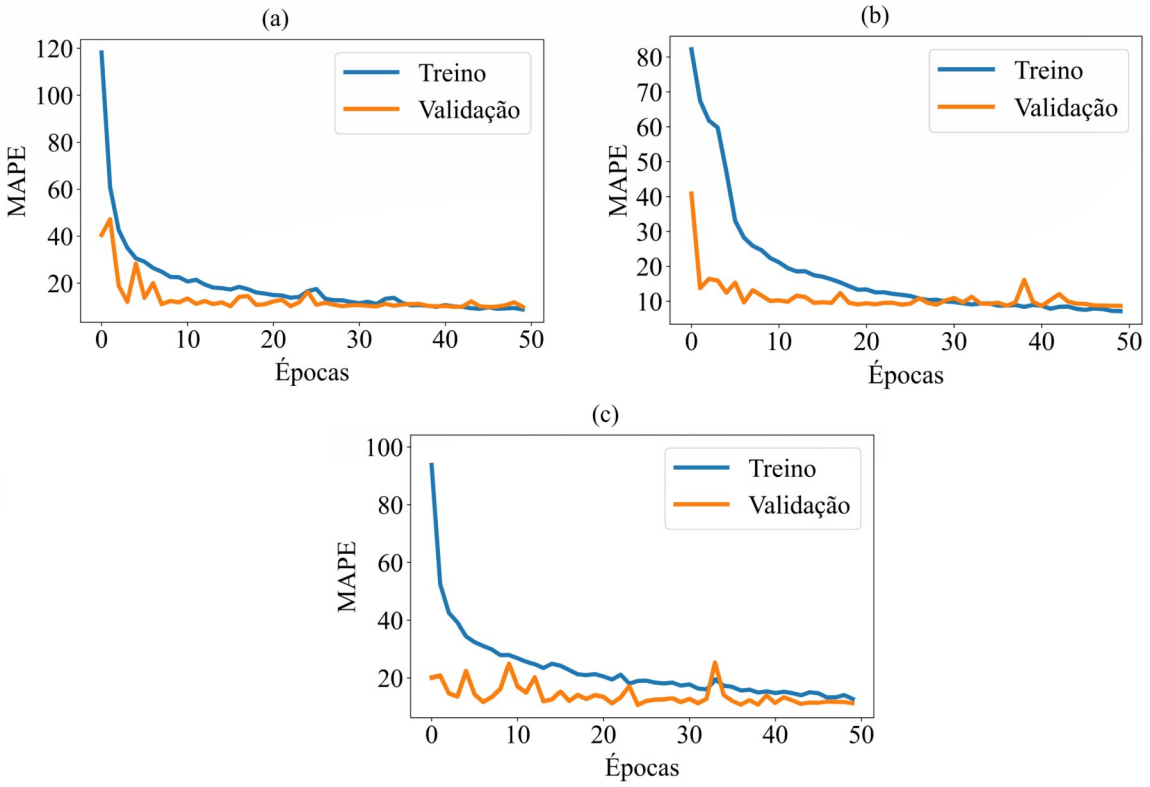

Figura 36 – Erros de treino e validação no treinamento dos modelos. a) Modelo 1. b) Modelo 2. c) Modelo 3. Elaborado pelo autor.

5.2 Validação em conjunto de teste

O valor de MAE e MAPE dos três diferentes métodos de avaliação da idade óssea está detalhado na Tabela 10. Para o primeiro modelo, o MAE do conjunto de teste foi 9,29 meses. Para o segundo modelo, o MAE do conjunto de teste foi de 8,24 meses. Comparando este método com o anterior, a taxa de concordância aumentou. Isso ocorre porque o modelo conseguiu prever melhor com regiões do pulso do que com regiões do dedo.

A fim de aumentar ainda mais a taxa de concordância, foi realizado o teste com um mosaico das duas regiões, que atingiu um MAE de 7,31 meses, apresentando um resultado ainda mais significativo quando utilizado as duas regiões combinadas.

Quando analisado os resultados de MAPE na base de teste em comparação com os valores de referência na literatura, os valores encontrados de 8,90%, 7,68% e 6,75%, para os Modelos 1, 2 e 3 respectivamente, indicam um modelo com previsão de alta precisão.

Tabela 10 – Resultados finais na base de teste.

Modelo	MAE (meses)	MAPE (%)
Modelo 1	9,29	8,90
Modelo 2	8,24	7,68
Modelo 3	7,31	6,75

A Figura 37 mostra os gráficos de dispersão dos resultados dos modelos aplicados na base de teste. Para a construção do gráfico, todos as amostras de idade óssea real na base de teste foram rearranjadas de forma crescente, representadas pela linha vermelha do gráfico. Os pontos azuis representam a idade óssea presumida para uma determinada amostra. O resultado é uma relação linear entre o valor real e o valor encontrado pelo modelo, indicando uma forte correlação entre os dois valores.

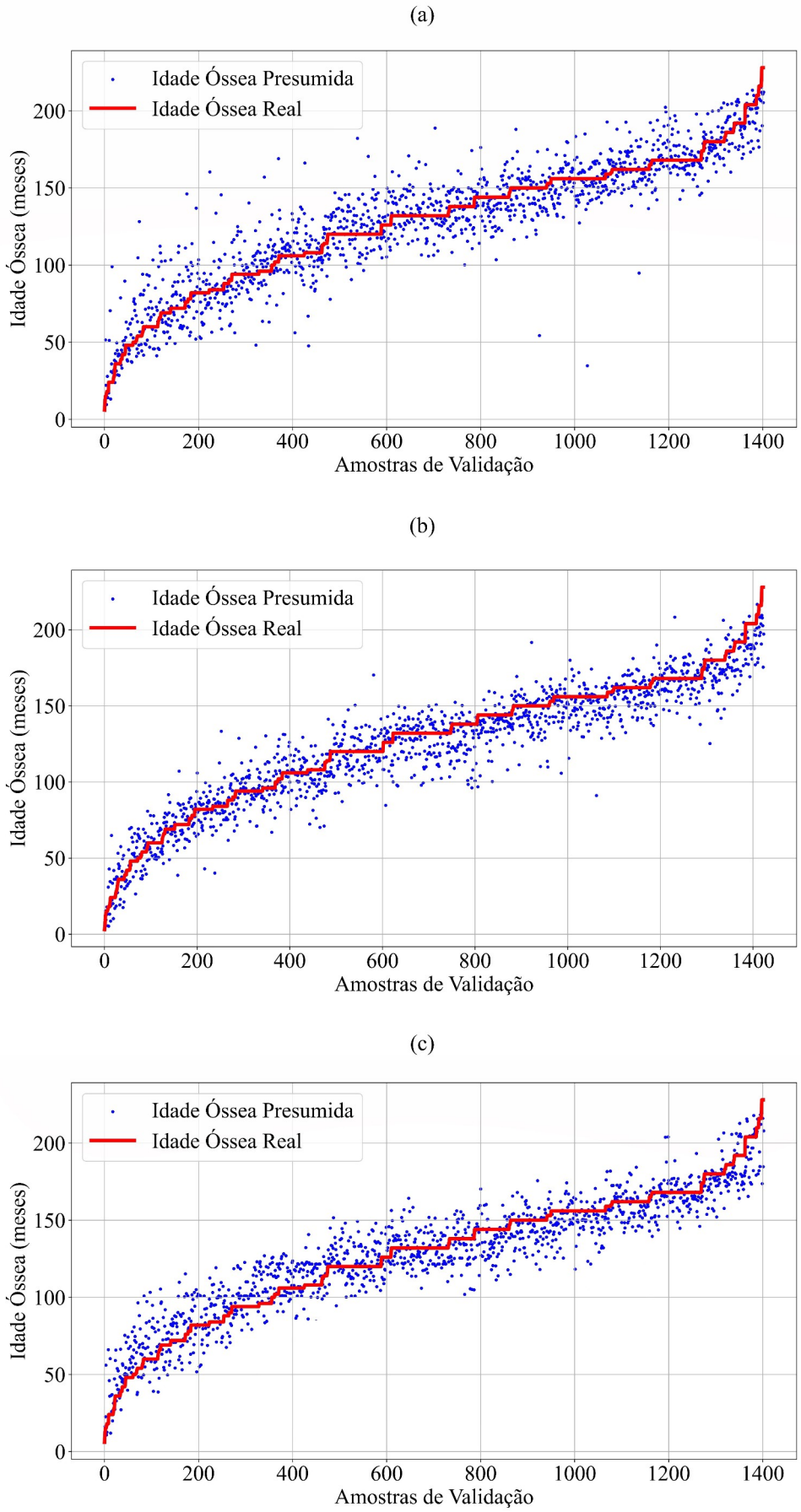

Figura 37 – Gráficos de dispersão das previsões dos modelos. a) Resultado do Modelo 1. b) Resultado do Modelo 2. c) Resultado do Modelo 3. Elaborado pelo autor.

5.3 Discussões

Neste capítulo foram apresentados os resultados da etapa de estimação da idade óssea. Para comparação direta, foram selecionados os métodos utilizando RNAs de classificação e regressão com maior e menor erro, que se encontram em resumo na Tabela 10. Os métodos utilizando cálculo de regiões não foi considerado pelo fato de os resultados serem descritos apenas considerando desvio padrão e classificação em estágios, onde 255 meses foram resumidos em 8 estágios, não sendo viável realizar paralelo com o método desenvolvido neste trabalho.

Tabela 11 – Resumo de resultados de estado da arte.

Método	Tipo	MAE	Vantagem	Desvantagem
(AN, 2017)	Classificação	19 meses	Sem Data Augmentation	945 amostras; MAE maior
(SON et al., 2019)	Classificação	5,52 meses	-	Segmentação de 13 centros de ossificação; Exclusão de estágios
(WESTERBERG, 2020)	Regressão	12 meses	Sem segmentação; Sem Data Augmentation	MAE maior
(GUO et al., 2022)	Regressão	6,07 meses	Sem Data Augmentation	Segmentação de 6 centros de ossificação; Combinação de 6 modelos no final

O método de (AN, 2017) utilizando Classificação atingiu o maior MAE de 19 meses, com a vantagem de não utilizar *Data Augmentation*, porém como foi utilizado um pequeno número de imagens, o modelo não é robusto para imagens totalmente diferentes de seu conjunto de dados interno. (SON et al., 2019) obteve o melhor resultado de Classificação com MAE de 5,52 meses, atingindo esse resultado com as desvantagens de segmentar de 13 centros de ossificação e excluir estágios justificado por poucas amostras. Comparando com o Modelo 3 desenvolvido neste trabalho, onde foi atingido 7,31 de MAE, a diferença entre os resultados seria de 1,79 meses.

Para os trabalhos utilizando Regressão, onde todos desenvolveram modelos com a mesma base de dados da RSNA, o método com maior MAE foi o elaborado por (WESTERBERG, 2020) de 12 meses. O mesmo não utilizou nenhuma segmentação para a entrada do modelo e não fez uso de técnicas de *Data Augmentation*. (GUO et al., 2022) alcançou o melhor MAE de 6,07 meses sem empregar também *Data Augmentation*, contudo foi necessário segmentar 6 centros de ossificação para cada mão e dispor de técnicas de combinações de 6 modelos diferentes. A diferença desse método com o proposto neste trabalho foi de 1,24 meses.

Os resultados obtidos mostraram-se satisfatórios quando comparados aos métodos apresentados no Capítulo 2, ficando bem próximo dos melhores valores de MAE encontrados, onde tal diferença não seria clinicamente significativa.

Capítulo 6

Conclusão

6.1 Introdução

Com o rápido desenvolvimento da tecnologia, o processamento de imagens digitais passou por uma rápida evolução e é amplamente utilizada em diversas áreas.É extremamente útil em sensoriamento remoto, medicina, reconhecimento e detecção de objetos.A profissão médica pode se beneficiar tremendamente do emprego de RNAs em imagens para colaborar com análises e diagnóstico. Neste trabalho foi estudado o uso de aprendizado de máquina em imagens médicas, principalmente para avaliação automatizada da idade óssea usando radiografias de mão. A seguir será resumido o desempenho das técnicas desenvolvidas e os resultados obtidos. Em seguida serão apresentadas as contribuições deste trabalho e, por último, serão descritas algumas possibilidades de trabalhos futuros que poderão ser desenvolvidos.

6.2 Conclusões Finais

O método desenvolvido neste trabalho foi treinado em um conjunto de dados público da RSNA contendo mais de 12500 radiografias. Este sistema padroniza automaticamente todas as radiografias manuais de diferentes formatos, aquisições e qualidade para serem usadas como um conjunto de dados de treinamento. A utilização do MLP foi adequada para previsão de regressão, em conjunto com a CNN treinada para classificação de imagens. O resultado deste estudo mostra que a saída é otimizada pela concatenação das redes acima. A RNA proposta mostrou precisão satisfatória na avaliação de imagem, onde para se treinar uma rede com precisão, muitos hiper parâmetros precisam ser ajustados, dependendo do conjunto de dados que está sendo usado.

A precisão da rede proposta foi obtida realizando a avaliação de maturidade óssea para mais de 1400 radiografias do conjunto de dados de teste. O modelo proposto alcançou um MAE de 9,29, 8,24 e 7,31 meses dos Modelos 1, 2 e 3, respectivamente, comparável ao desempenho do estado da arte. Isso indica um nível substancial de concordância entre

as revisões obtidas e a avaliação absoluta. Esses resultados foram alcançados, em grande parte, por meio do uso de técnicas de *Data Augmentation* para aumentar artificialmente o tamanho do conjunto de treinamento. A precisão da solução proposta é semelhante à obtida por radiologistas especializados e semelhante aos sistemas automatizados anteriores.

6.3 Contribuições

De acordo com os resultados relatados, pode-se afirmar que este trabalho trouxe novas contribuições para a área de processamento de imagens médicas, sendo possível auxiliar profissionais de ortopedia, pediatria, endocrinologia, entre outros, no diagnóstico de pacientes. Foi demonstrada a viabilidade de implementar um método totalmente automatizado para determinar a idade óssea, além da possibilidade de desenvolver um modelo de segmentação de imagens de radiografia com apenas 80 amostras de treinamento, rendendo ótimos resultados, mesmo com imagens apresentando posicionamento e contraste diferenciados.

As abordagens utilizadas para pré-processamento, segmentação e extração de regiões são eficientes, simples e de baixo custo computacional, podendo ser aplicadas em diversos tipos de aplicações que utilizam imagens digitalizadas de radiografias.A solução proposta foi desenvolvida do início ao fim, pode ser facilmente treinada com diferentes conjuntos de dados e requer um tempo de treinamento razoável. Na prática, o sistema proposto pode ser facilmente implantado no ambiente clínico em um computador comum. O método de concatenar MLP e CNN pode ser implementado em sistemas que consistem em formatos de dados mistos.

6.4 Trabalhos Futuros

As propostas para melhorias neste trabalho e sugestões para trabalhos futuros são:

• Pesquisar a utilização de outras técnicas de pré-processamento de imagens para otimizar a segmentação dos ossos dos dedos e punho;

• Aumentar a quantidade de dados no conjunto de dados original provavelmente levaria a uma melhoria adicional na precisão do classificador proposto, uma vez que existe um desbalanceamento do banco de dados usado para treinamento em relação às idades;

• Desenvolver uma CNN com ajuste fino de *Transfer Learning*, utilizando pesos de soluções relacionadas à tarefa específica de imagens médicas, por exemplo, outro sistema envolvendo classificação de imagens de raios-x;

• A elaboração de um sistema completo de determinação da maturidade óssea, com a possibilidade de desenvolver um aplicativo usando o modelo de regressão desenvolvido;

- Realizar testes com mais camadas, aumentar as dimensões das imagens de entrada e executar o processamento em computadores mais potentes.

Referências

ABDULATEEF, S. K.; SALMAN, M. D. A comprehensive review of image segmentation techniques. **Iraqi Journal for Electrical & Electronic Engineering**, v. 17, n. 2, 2021.

ALZUBAIDI, L. et al. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. **Journal of big Data**, Springer, v. 8, p. 1–74, 2021.

AN, D. Y. Estimativa da idade óssea usando mosaicos dos centros de ossificação de radiografias carpais como imagens de entrada para deep learning. 2017.

ANDRADE, C. A student's guide to the classification and operationalization of variables in the conceptualization and design of a clinical study: Part 2. **Indian Journal of Psychological Medicine**, v. 43, n. 3, p. 265–268, 2021.

BALLARD, D. H.; BROWN, C. M. **Computer Vision**. New Jersey, USA: Prentice Hall, 1982.

BANKMAN, I. **Handbook of medical image processing and analysis**. [S.l.]: Elsevier, 2008.

CASTRO, F. C. d. et al. Localização, segmentação e classificação automáticas de regiões de interesse para a determinação de maturidade óssea utilizando o método de tanner-whitehouse. Universidade Federal de Uberlândia, 2009.

CAVALLO, F. et al. Evaluation of bone age in children: a mini-review. **Frontiers in Pediatrics**, Frontiers Media SA, v. 9, p. 580314, 2021.

CHEN, M. Automated bone age classification with deep neural networks. In: **Technical Report**. [S.l.]: Stanford University, USA, 2016.

CHEN, X. et al. Automatic feature extraction in x-ray image based on deep learning approach for determination of bone age. **Future Generation Computer Systems**, Elsevier, v. 110, p. 795–801, 2020.

DIGHE, P. C.; GURU, S. K. Survey on image resizing techniques. **International Journal of Science and Research (IJSR)**, v. 3, n. 12, p. 1444–1448, 2014.

EKLÖF, O.; RINGERTZ, H. A method for assessment of skeletal maturity. In:**Annales de radiologie**. [S.l.: s.n.], 1967. v. 10, n. 3, p. 330–336.

ERICKSON, B. J.; BARTHOLMAI, B. Computer-aided detection and diagnosis at the start of the third millennium. **Journal of digital imaging**, Springer, v. 15, p. 59–68, 2002.

GEDRAITE, E. S.; HADAD, M. Investigation on the effect of a gaussian blur in image filtering and segmentation. In: IEEE. **Proceedings ELMAR-2011**. [S.l.], 2011. p. 393–396.

GILSANZ, V.; RATIB, O. **Hand bone age: a digital atlas of skeletal maturity**. [S.l.]: Springer, 2005. v. 1.

GONZALES, R. C.; WINTZ, P. **Digital image processing**. [S.l.]: Addison-Wesley Longman Publishing Co., Inc., 1987.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. (Adaptive computation and machine learning). ISBN 9780262035613. Disponível em: <https://books.google.co.in/books?id=Np9SDQAAQBAJ>.

GREULICH, W.; PYLE, S. **Radiographic Atlas of Skeletal Development of the Hand and Wrist**. Stanford University Press, 1959. ISBN 9780804703987. Disponível em: <https://books.google.com.br/books?id=olezJFYxM6oC>.

GUO, L. et al. Bone age assessment based on deep convolutional features and fast extreme learning machine algorithm. **Frontiers in Energy Research**, Frontiers, v. 9, p. 888, 2022.

HAYKIN, S. **Neural Networks and Learning Machines**. Prentice Hall, 2009. (Neural networks and learning machines, v. 10). ISBN 9780131471399. Disponível em: <https://books.google.com.br/books?id=K7P36lKzI_QC>.

HOU, Y. et al. The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis. **Engineering**, Elsevier, v. 7, n. 6, p. 845–856, 2021.

HU, B. et al. Bone age prediction method based on convolutional neural network. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S.l.], 2020. v. 1646, n. 1, p. 012065.

IGLOVIKOV, V. I. et al. Paediatric bone age assessment using deep convolutional neural networks. In: SPRINGER. **Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4**. [S.l.], 2018. p. 300–308.

JÚNIOR, C. O. **Estimativa da idade óssea através da análise carpal baseada na simplificação do método de Eklof & Ringertz**. Tese (Doutorado) — Universidade de São Paulo, 2005.

KANDEL, I.; CASTELLI, M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. **ICT express**, Elsevier, v. 6, n. 4, p. 312–315, 2020.

KHAN, W. Image segmentation techniques: A survey. **Journal of image and graphics**, v. 1, n. 4, p. 166–170, 2013.

KIM, J.-Y.; KIM, L.-S.; HWANG, S.-H. An advanced contrast enhancement using partially overlapped sub-block histogram equalization. **IEEE transactions on circuits and systems for video technology**, IEEE, v. 11, n. 4, p. 475–484, 2001.

KRIG, S. **Computer vision metrics: Survey, taxonomy, and analysis**. [S.l.]: Springer nature, 2014.

KUHN, M.; JOHNSON, K. et al. **Applied predictive modeling**. [S.l.]: Springer, 2013. v. 26.

LEE, H. et al. Fully automated deep learning system for bone age assessment. **Journal of digital imaging**, Springer, v. 30, p. 427–441, 2017.

LEE, S. U.; CHUNG, S. Y.; PARK, R. H. A comparative performance study of several global thresholding techniques for segmentation. **Computer Vision, Graphics, and Image Processing**, Elsevier, v. 52, n. 2, p. 171–190, 1990.

LEWIS, C. International and business forecasting methods butterworths: London. 1982.

LI, L. et al. Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2017.

LONGUI, C. A. **Previsão da estatura final-acertando no"alvo"?** [S.l.]: SciELO Brasil, 2003. 636–637 p.

MARROCOS, M. L. L. d. L. et al. Detecção automática de idade óssea através da radiografia de mão e punho utilizando redes neurais convolucionais. Universidade Federal de Campina Grande, 2019.

MELMED, S. et al. **Williams textbook of endocrinology E-Book**. [S.l.]: Elsevier Health Sciences, 2015.

MUGHAL, A. M.; HASSAN, N.; AHMED, A. Bone age assessment methods: a critical review. **Pakistan journal of medical sciences**, Professional Medical Publications, v. 30, n. 1, p. 211, 2014.

NARANJO-TORRES, J. et al. A review of convolutional neural network applied to fruit image processing. **Applied Sciences**, MDPI, v. 10, n. 10, p. 3443, 2020.

NIELSEN, M. A. **Neural networks and deep learning**. [S.l.]: Determination press San Francisco, CA, USA, 2015. v. 25.

PAN, X. et al. Fully automated bone age assessment on large-scale hand x-ray dataset. **International journal of biomedical imaging**, Hindawi, v. 2020, 2020.

PARSANIA, P.; VIRPARIA, P. V. et al. A review: Image interpolation techniques for image scaling. **International Journal of Innovative Research in Computer and Communication Engineering**, v. 2, n. 12, p. 7409–7414, 2014.

POOSARLA, A. **Bone age prediction with convolutional neural networks**. Tese (Doutorado) — California State University, Sacramento, 2018.

PRATT, W. K. **Digital image processing: PIKS Scientific inside**. [S.l.]: Wiley Online Library, 2007. v. 4.

PRIYA, G.; NAWAZ, K. Effective morphological image processing techniques and image reconstruction. **IJTRD, vol. abs/1703.10593**, 2017.

QUEIROZ, A. d. C. **Metodologia de extração automática de características da mão para a estimação da idade óssea utilizando redes neurais artificiais no processo de decisão**. Tese (Doutorado) — Universidade de São Paulo, 2006.

RIPLEY, B. D. **Pattern recognition and neural networks**. [S.l.]: Cambridge university press, 2007.

RSNA. 2017. Disponível em: <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017>.

SALVI, M. et al. The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. **Computers in Biology and Medicine**, Elsevier, v. 128, p. 104129, 2021.

SAUVOLA, J. et al. Adaptive document binarization. In: IEEE. **Proceedings of the fourth international conference on document analysis and recognition**. [S.l.], 1997. v. 1, p. 147–152.

SÉRGIO, H. et al. Proposta de um sistema para reconhecimento de idade óssea auxiliada por computador utilizando redes neurais artificiais. 2011.

SEZGIN, M.; SANKUR, B. l. Survey over image thresholding techniques and quantitative performance evaluation. **Journal of Electronic imaging**, Society of Photo-Optical Instrumentation Engineers, v. 13, n. 1, p. 146–168, 2004.

SHAHINFAR, S.; MEEK, P.; FALZON, G. "how many images do i need?" understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. **Ecological Informatics**, Elsevier, v. 57, p. 101085, 2020.

SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. **Towards Data Sci**, v. 6, n. 12, p. 310–316, 2017.

SILVA, T. A. M. d. et al. Avaliação do aam-active appearance model para detecção de regiões de interesse em radiografias carpais na estimativa da idade óssea através do método tw. Universidade Federal de Uberlândia, 2016.

SON, S. J. et al. Tw3-based fully automated bone age assessment system using deep neural networks. **IEEE Access**, IEEE, v. 7, p. 33346–33358, 2019.

SOUZA, D.; OLIVEIRA, M. M. End-to-end bone age assessment with residual learning. In: IEEE. **2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S.l.], 2018. p. 197–203.

SREEDHAR, K.; PANLAL, B. Enhancement of images using morphological transformation. **arXiv preprint arXiv:1203.2514**, 2012.

TAN, L. et al. Comparison of retinanet, ssd, and yolo v3 for real-time pill identification. **BMC Medical Informatics and Decision Making**, v. 21, 11 2021.

TANNER, J. et al. Prediction of adult height from height, bone age, and occurrence of menarche, at ages 4 to 16 with allowance for midparent height. **Archives of disease in childhood**, BMJ Publishing Group Ltd, v. 50, n. 1, p. 14–26, 1975.

TAVANO, O. **Radiografias Carpal e Cefalométrica como Estimadores da Idade Óssea e do Crescimento e Desenvolvimento**. [S.l.]: Bauru–Brasil, 2001.

TENSORFLOW. 2022. Disponível em: <https://www.tensorflow.org/tutorials/keras/keras_tuner>.

TOET, A.; WU, T. Efficient contrast enhancement through log-power histogram modification. **Journal of Electronic Imaging**, Society of Photo-Optical Instrumentation Engineers, v. 23, n. 6, p. 063017–063017, 2014.

WESTERBERG, E. **AI-based age estimation using X-ray hand images: A comparison of object detection and deep learning models**. 2020.

ZEFERINO, A. et al. Acompanhamento do crescimento. **Jornal de pediatria**, SciELO Brasil, v. 79, p. S23–S32, 2003.

ZULKIFLEY, M. A. et al. Intelligent bone age assessment: an automated system to detect a bone growth problem using convolutional neural networks with attention mechanism. **Diagnostics**, MDPI, v. 11, n. 5, p. 765, 2021.