



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Aprendizado de Máquina como auxílio na detecção
de sinais de osteoporose analisando radiografias
panorâmicas bucais**

Yuri B. Galli

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Orientador

Prof. Dr. Bruno Luigi Macchiavello Espinoza

Brasília
2023



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Aprendizado de Máquina como auxílio na detecção
de sinais de osteoporose analisando radiografias
panorâmicas bucais**

Yuri B. Galli

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Prof. Dr. Bruno Luiggi Macchiavello Espinoza (Orientador)
CIC/UnB

Prof. Dr. Fábio Wildson Gurgel Costa Dr. Pedro de Azevedo Berger
Universidade Federal do Ceará (UFC) Universidade de Brasília (UnB)

do Programa de Pós-graduação em Informática

Brasília, 09 de fevereiro de 2023

Dedicatória

Dedico este trabalho à minha esposa Julia, que é a pessoa com a qual decidi estruturar a minha vida, e que sempre me impulsiona a conhecer mais e trabalhar melhor. Ela sempre me motiva a ir além. Obrigado.

Sapere Aude! (Ouse Saber!)

Agradecimentos

A finalização desta etapa da minha vida, o Mestrado, tem muitos nomes relacionados de pessoas que ajudaram e contribuíram de várias formas. Agradeço à minha esposa Julia, que segurou todas as barras comigo. Agradeço também à minha família, que me deu bastante suporte na decisão de seguir esse caminho. Em especial, agradeço ao meu sogro Jose Umberto que foi quem me despertou para a possibilidade de buscar um Mestrado como caminho para o crescimento profissional e pessoal. Agradeço ao meu professor orientador Bruno, que conduziu todo o processo e me ensinou bastante e transmitiu seu conhecimento. Também agradeço aos professores da área da Odontologia que nos ajudaram com dados/imagens para o projeto e conhecimento da área biomédica: Nilce, André, Paulo e Carla.

Resumo

A osteoporose é sinônimo de fragilidade óssea, e é uma doença silenciosa que só é detectada comumente após já ter causado danos à pessoa que a possui. Essa doença de fragilidade óssea torna a fratura mais comum e mais danosa aos seus portadores, e por isso é uma questão de saúde pública. Identificar a doença em estágio inicial é essencial para ajudar a evitar seus danos, e nessa tarefa a inteligência artificial e o aprendizado de máquina têm demonstrado grande ajuda nos últimos anos. Os algoritmos de aprendizado de máquina podem prever o risco de osteoporose analisando as imagens do paciente provenientes de exames de rotina, como radiografias panorâmicas.

A metodologia proposta conta com um processo de dois passos, que é composto por pré-processamento de imagens e aprendizado de máquina. O pré-processamento de imagens consistiu em transformar as imagens panorâmicas bucais originais brutas em regiões de interesse reduzidas, mais específicas e mais claras para a classificação. A etapa de aprendizado de máquina consistiu em fornecer essas imagens pré-processadas para algoritmos computacionais as classificarem. Este trabalho propõe uma arquitetura de Rede Neural Convolutiva (CNN), comparada com Support Vector Machine (SVM) e Random Forest (RF), que visa identificar sinais de osteoporose nesse tipo de imagem, com objetivo de melhorar os resultados de técnica de referência, a CNN do artigo [1], ao utilizar a estrutura CNN modificada para realizar uma detecção automática de osteoporose com alta sensibilidade e utilizar o método RF para um sistema com especificidade alta.

Na sensibilidade, a CNN customizada proposta obteve 77,19%, enquanto que a CNN de referência obteve 70,18%. Na especificidade, a Random Forest proposta obteve 75,95%, enquanto que a CNN de referência obteve 22,78%. Combinando esses resultados, conseguimos melhorar o que é obtido pela técnica de referência para o conjunto de dados apresentado, que é bastante desafiador, de pacientes idosos de comunidades carentes, em imagens analógicas com diversos artefatos e características que dificultam a classificação.

Palavras-chave: osteoporose, redes neurais convolucionais, cnn, aprendizado de máquina, inteligência artificial, random forest, svm

Abstract

Osteoporosis is synonymous with bone fragility, and it is a silent disease that is only commonly detected after it has already caused damage to the person who has it. This disease of bone fragility makes fractures more common and more harmful to patients, and therefore it is a public health issue. Identifying the disease at an early stage is essential to help prevent its damage, and in this task artificial intelligence and machine learning have shown great help in recent years. Machine learning algorithms can predict osteoporosis risk by analyzing patient images from routine exams such as panoramic radiographs.

The proposed methodology has a two-step process, which consists of image preprocessing and machine learning. Image preprocessing consisted of transforming the raw original oral panoramic images into reduced, more specific, and clearer regions of interest for classification. The machine learning step consisted of providing these preprocessed images for computational algorithms to classify them. This work proposes a Convolutional Neural Network (CNN) architecture, compared with Support Vector Machine (SVM) and Random Forest (RF), which aims to identify signs of osteoporosis in this type of image, with the aim of improving the results of the reference technique, the CNN from the article [1], using the modified CNN structure to perform an automatic detection of osteoporosis with high sensitivity and using the RF method for a system with high specificity.

In sensitivity, the proposed customized CNN obtained 77.19%, while the reference CNN obtained 70.18%. In specificity, the proposed Random Forest obtained 75.95%, while the reference CNN obtained 22.78%. Combining these results, we were able to improve what is obtained by the reference technique for the presented data set, which is quite challenging, from elderly patients from poor communities, in analog images with several artifacts and characteristics that make classification difficult.

Keywords: osteoporosis, convolutional neural networks, cnn, machine learning, artificial intelligence, random forest, svm

Sumário

1	Introdução	2
1.1	Objetivos	3
1.2	Justificativa	4
1.3	Contribuições	4
2	Fundamentação Teórica	6
2.1	Preprocessamento de Imagens	6
2.1.1	Filtragem	7
2.1.2	Histogramas de Imagens	8
2.1.3	Método de Otsu para Binarização	9
2.1.4	Segmentação Automática de Imagens	10
2.2	Aprendizado de máquina	10
2.2.1	Tipos de Aprendizado de Máquina	11
2.2.2	Redes Neurais Artificiais	12
2.2.3	Redes Neurais Convolucionais (<i>Convolutional Neural Network</i> (CNN)s)	13
2.2.4	<i>Support Vector Machine</i> (SVM)	16
2.2.5	<i>Random Forest</i> (RF)	16
3	Revisão de Literatura	18
3.1	Identificando a osteoporose usando imagens médicas	18
3.2	Abordagens Estado da Arte na Identificação Computacional da Osteoporose	19
3.2.1	Abordagens de Análise de Textura	19
3.2.2	Abordagens de Aprendizado de Máquina	20
3.2.3	Revisões sobre Identificação Computacional da Osteoporose	21
3.2.4	Classificação da Osteoporose	22
3.2.5	Revisões sobre Identificação Computacional da Osteoporose	22
4	Metodologia	25
4.1	Preprocessamento de Imagens	26
4.1.1	Aquisição de Imagens	27

4.1.2 Segmentação de imagem	29
4.2 Metodologia Proposta	38
4.2.1 CNN Proposta	39
4.2.2 Treinando e Testando a Rede	42
4.2.3 SVM Proposta	43
4.2.4 Random Forest (RF) Proposta	44
5 Resultados	45
5.1 Conjunto de Dados	46
5.2 Melhorias propostas	47
5.3 Resultados Comparativos	48
5.4 Discussão dos Resultados	48
6 Conclusão	50
Referências	51

Lista de Figuras

4.1	Fluxograma completo do trabalho.	25
4.2	Fluxograma do Processamento de Imagens.	26
4.3	Fluxograma do Aprendizado de Máquina.	27
4.4	Exemplo de uma imagem radiográfica panorâmica digitalizada.	28
4.5	Example of a Digital Panoramic Radiographic Image.	28
4.6	Exemplos de aparência radiográfica do córtex mandibular inferior em radiografias panorâmicas.	30
4.7	Imagens convertidas em tons de cinza e originais lado a lado	32
4.8	imagem rotacionada (Etapa 2).	33
4.9	Imagem sem etiqueta do paciente (Etapa 3).	33
4.10	Imagem com bordas aprimoradas (Etapa 4).	34
4.11	Imagem com contraste aprimorado (Etapa 5).	34
4.12	Imagem convertida em binário (Etapa 6).	35
4.13	Gráfico contendo a soma dos pixels pretos binários de cada linha na imagem (Etapa 7).	36
4.14	Imagem binária com a região de recorte marcada por linhas (Etapa 8). . .	36
4.15	Imagem binária cortada (resultado da Etapa 8).	37
4.16	Imagem em tons de cinza com a região de recorte marcada por linhas (Etapa 9).	37
4.17	Imagem em tons de cinza cortada - Corte Final (resultado da Etapa 9). . .	38
4.18	Imagem dividida em tons de cinza com região de corte marcada por linhas. . .	38
4.19	Imagem dividida em tons de cinza.	38
4.20	Estrutura da CNN proposta, inspirada na referência [1].	39
5.1	Diferentes problemas nas imagens da base de dados.	47

Lista de Tabelas

4.1 CNN Referência	40
5.1 Resultados CNN Referência no conjunto de dados do trabalho	48
5.2 Resultados Comparativos no conjunto de dados do trabalho	48

Lista de Abreviaturas e Siglas

bagging seleção aleatória de pontos de dados da amostra com substituição.

CNN *Convolutional Neural Network*.

cross-validation reamostragem com o objetivo de tornar os resultados estatísticos mais genéricos.

Data augmentation Aumento de Dados feito gerando cópias das imagens originais alterando-as computacionalmente.

DL *Deep Learning*.

DMO Densidade Mineral Óssea.

dropout técnica de regularização que reduz *overfitting*.

DT *Decision Trees*.

DXA Densitometria Óssea.

Ensemble Agrupamento de técnicas para obtenção de melhores resultados em conjunto.

FIR *Finite Impulse Response*.

IA Inteligência Artificial.

kernel núcleo.

log-softmax função exponencial normalizada em logaritmo.

Loss Functions Funções de Perda.

NLL *Negative Log Likelihood*.

overfitting excesso de adequação de um modelo aos dados utilizados para gerá-lo.

overtraining treinamento excessivo do modelo que pode levar ao *overfitting*.

pooling janelamento.

RBF Gaussiana de Base Radial.

ReLU Função de ativação *Rectified Linear Unit*.

RF *Random Forest*.

RGB *Red Green Blue* (Sistema de cores Vermelho Verde Azul).

RN Redes Neurais.

RNA Redes Neurais Artificiais.

RNC Redes Neurais Convolucionais.

ROI Region Of Interest.

softmax função exponencial normalizada.

SVM *Support Vector Machine*.

Capítulo 1

Introdução

A osteoporose é uma doença silenciosa, ou seja, as pessoas com essa doença não apresentam nenhum sinal físico. Uma maneira, e a mais simples, de identificar a osteoporose é quando ocorre uma fratura óssea, mas tal fratura pode ser de alto risco, principalmente em pessoas idosas. O grande desafio é identificar a doença antes que ocorra a fratura. Nesse caso, a osteoporose pode ser identificada pelo indicativo de baixa Densidade Mineral Óssea (DMO). [2]

A forma mais utilizada para medir a DMO é o exame de Densitometria Óssea (DXA), com o intuito de diagnosticar a osteoporose e prevenir fraturas ósseas. [1] No entanto, o custo financeiro de tal exame é relativamente alto e o acesso a máquinas capazes de realizá-lo é reduzido. [3] Nesse contexto, estudos recentes mostram que as radiografias panorâmicas bucais podem ser uma forma acessível de diagnosticar sinais de osteoporose, pois são exames mais baratos e possuem algumas características fortemente relacionadas à baixa DMO. [4]

O diagnóstico de osteoporose por imagens de radiografias panorâmicas bucais pode servir como uma espécie de primeira triagem que separa os pacientes que apresentam fortes sinais de osteoporose daqueles cuja saúde óssea é muito boa a ponto de descartar essa hipótese a priori. [5] Estudos indicam que os profissionais da odontologia que analisam essas imagens radiográficas orais são capazes de encaminhar seus pacientes com evidência de osteoporose a profissionais especializados para serem tratados por eles, e isso aumenta significativamente a chance de melhor tratamento, atenuação da ocorrência de osteoporose e problemas de fratura nesses pacientes. [6]

A detecção de sinais de osteoporose por imagem radiográfica oral é uma solução eficiente em relação ao custo-benefício para o tratamento precoce desta doença óssea, mas o grande problema dessa solução é que normalmente os dentistas que avaliam radiografias panorâmicas não o fazem procurando especificamente por esses índices, mas por outros problemas bucais. Por isso, técnicas computacionais de diagnóstico assistido e avaliação

automática têm sido cada vez mais utilizadas nesse nicho da saúde.

Estudos têm demonstrado que quanto mais precoce a disfunção causada pela osteoporose é diagnosticada, menores são seus impactos e maiores são as chances de tratamento. [7]

Técnicas de aprendizado de máquina são comumente utilizadas para identificar estruturas e padrões que o olho humano dificilmente perceberia, com maior precisão e também com menos flutuações em diversos sinais [8]. Esses algoritmos dificilmente podem substituir o especialista humano, mas podem ser de grande ajuda para aprimorar certas tarefas como o diagnóstico médico. No problema em questão, um sistema computacional capaz de indicar a existência de indícios de osteoporose através da radiografia panorâmica pode levar a uma detecção precoce.

No Brasil e países da América do Sul, ainda é comum achar nas comunidades carentes aparelhos de radiografia analógicos. Estes aparelhos normalmente apresentam uma imagem mais ruidosa que a dos aparelhos digitais mais recentes. Nestas comunidades carentes também existe uma escassez de acompanhamento médico preventivo. Sendo assim, o presente trabalho tem como objetivo propor um método para identificação de osteoporose em imagens panorâmicas analógicas obtidas em comunidades carentes focado em pessoas idosas, para realizar um encaminhamento mais eficaz para o especialista médico em osteoporose e evitar fraturas de alto risco.

1.1 Objetivos

O objetivo geral deste trabalho é propor uma nova técnica que auxilie computacionalmente a detecção de sinais de osteoporose a partir de imagens radiográficas bucais, utilizando aprendizado de máquina.

O trabalho também objetiva especificamente:

- Verificar a viabilidade de utilizar imagens obtidas de aparelhos analógicos;
- Focar em pacientes idosos de comunidades de baixa renda;
- Realizar uma segmentação automática da região de interesse;
- Comparar três métodos de aprendizado de máquina supervisionados: *Convolutional Neural Network* (CNN), *Random Forest* (RF) e *Support Vector Machine* (SVM), aplicados ao problema apresentado;
- Obter uma alta acurácia, sensibilidade e/ou especificidade.

1.2 Justificativa

Em resumo, a motivação para este trabalho é que a osteoporose é uma doença silenciosa e de difícil detecção antes que ocorra um problema grave, como uma fratura óssea, pois não apresenta sintomas nos pacientes que a possuem. Grandes gastos têm sido gerados na área da saúde em países ao redor de todo o mundo por causa dessa doença, pois suas consequências são onerosas tanto para os pacientes quanto para o sistema de saúde. Uma forma de identificar a doença é o exame de densitometria óssea, porém esse exame é caro e na maioria das vezes inacessível, o que torna necessário buscar alternativas que auxiliem no diagnóstico. Uma dessas alternativas é o exame de radiografia bucal panorâmica, cuja análise pode fornecer evidências de osteoporose e, assim, possibilitar que os pacientes que passam por essa triagem inicial sejam encaminhados para profissionais especializados e recebam tratamento precocemente, aumentando as chances de evitar os riscos que a doença oferece. No entanto, os profissionais que normalmente avaliam radiografias panorâmicas não observam ou priorizam imediatamente os sinais de osteoporose e, como se sabe neste meio, algumas avaliações são subjetivas e feitas apenas a partir do olho humano, contando apenas com a expertise do profissional. Para auxiliar na detecção eficaz de sinais de osteoporose em radiografias bucais panorâmicas, técnicas computacionais relacionadas ao aprendizado de máquina e inteligência artificial têm entrado muito em cena, tornando os diagnósticos mais rápidos, precisos e acessíveis.

1.3 Contribuições

Apresentado este cenário gerado pela doença de osteoporose e as atuais iniciativas para identificá-la cada vez mais cedo e de forma mais precisa, as melhorias nas técnicas computacionais apresentadas por esse trabalho contribuem para que:

- Os impactos dessa doença sejam reduzidos, e por isso o trabalho é tão relevante;
- Se abra caminho para viabilizar a utilização de imagens obtidas por aparelhos analógicos, que é o mais comum no Brasil;
- Se possa abranger também pacientes idosos de comunidades de baixa renda;
- Se possa ter um algoritmo que realiza uma segmentação automática de regiões de interesse em imagens semelhantes;
- Sejam comparados três métodos de aprendizado de máquina supervisionados: CNN, RF e SVM, aplicados ao problema apresentado, para obter a melhor acurácia, sensibilidade e/ou especificidade.

Só este trabalho não vai resolver sozinho tão grande questão, porém, somando cada contribuição científica nessa área, é possível que em breve os pacientes que possuem osteoporose sofram menos os danos de sua condição com esse auxílio da computação.

Capítulo 2

Fundamentação Teórica

Neste trabalho, a identificação dos sinais de osteoporose em radiografias panorâmicas consistiu em duas fases principais: a primeira sendo o processamento das imagens e a segunda a aplicação de uma técnica de aprendizado de máquina para realizar a classificação dessas imagens. Este capítulo fornece a fundamentação teórica usada em ambas as fases, primeiro explicando os tratamentos de imagem realizados no conjunto de dados e, posteriormente, fornecendo os detalhes necessários sobre a estrutura do algoritmo de aprendizado de máquina proposto.

A fase de processamento de imagem é essencial para fornecer melhores entradas para o algoritmo de aprendizado de máquina, de modo que ambas as fases estão vitalmente ligadas. Quanto melhores as imagens fornecidas, melhores resultados podem ser alcançados, dependendo, é claro, da eficiência dos algoritmos computacionais. As seções a seguir descrevem alguns princípios que ajudam a definir quão bem uma imagem é pré-processada ou segmentada e, posteriormente, quão bem o algoritmo de aprendizado de máquina executa a tarefa de classificação ou identificação de imagens.

2.1 Preprocessamento de Imagens

Esta seção tem como objetivo descrever os princípios das técnicas empregadas para aprimorar - ou seja, filtrar - as imagens do conjunto de dados, o que as torna propensas a uma melhor classificação, de forma matemática bastante objetiva.

Uma imagem digital é essencialmente uma matriz de números dentro de um intervalo especificado. Essa matriz é interpretada pelo computador como um arquivo e depois mostrada na tela com características que dependem de algumas informações fornecidas pelo cabeçalho interno desse arquivo. [9] Dependendo do formato do arquivo, a imagem pode ser retratada de maneiras diferentes, mas os princípios são os mesmos. Para todos

os propósitos deste trabalho, as imagens podem ser interpretadas como matrizes 2-D de números.

2.1.1 Filtragem

Filtragem é o processo de remover características de um sinal, e este processamento é feito com cálculos matemáticos em cima dos dados que se possui. No caso de imagens, dependendo de quais características se deseja alterar ou aprimorar, é feito um tipo diferente de cálculo, e isso é chamado de filtragem de imagem. [10] Por exemplo, pode-se filtrar uma imagem para enfatizar determinados recursos ou remover outros. As operações de filtragem em imagens incluem suavização, nitidez, aprimoramento de bordas e conversão de cores.

Neste trabalho vamos focar nos filtros *Finite Impulse Response* (FIR), que são uns dos mais utilizados no processamento de sinais digitais. [11] Em geral, o resultado da filtragem com filtros FIR pode ser representada pela Equação 2.1. Este tipo de filtro remove frequências específicas do sinal em sua saída, e também pode ser aplicado a sinais 2-D.

$$Y(n) = \sum_{l=0}^l h(l)x(n-l), n = 1, 2, 3... \infty \quad (2.1)$$

Neste trabalho, utiliza-se também a filtragem espacial, que se refere às operações feitas com relação à posição dos pixels na imagem. Isso se opõe à filtragem de frequência, por exemplo, que considera parâmetros de frequência. A filtragem de domínio espacial considera a posição absoluta do pixel na imagem, e as operações são feitas na vizinhança desse pixel, incluindo os pixels ao seu redor na operação. [12]

Os filtros são a maneira matemática de lidar com os valores dos pixels nas matrizes da imagem. Dependendo da saída desejada, a operação varia. Por exemplo, no realce de borda de uma imagem, o filtro tem como objetivo deixar aparecer apenas as linhas de divisão da imagem, o que é feito por meio de um filtro passa-alta. Um filtro passa-alta permite que apenas frequências superiores a uma frequência de corte estejam presentes na saída, processo que realça a borda quando aplicado com um filtro laplaciano.

O Laplaciano de uma imagem destaca regiões de rápida mudança de intensidade e, portanto, é frequentemente usado para detecção de bordas¹. O $L(x, y)$ Laplaciano de uma imagem com valores de intensidade de pixel $I(x, y)$ é dado pela Equação 2.2, que se trata de um filtro passa-alta.

¹Mais informações sobre o filtro laplaciano *neste site*.

$$L(x, y) = \frac{\partial^2 I}{\partial^2 x^2} + \frac{\partial^2 I}{\partial^2 y^2} \quad (2.2)$$

Esses cálculos dependem, é claro, do formato dos números que estão presentes nas matrizes. Por exemplo, a comumente chamada *imagem em preto e branco* pode ser descrita de várias maneiras para o computador, alguns exemplos sendo imagens binárias, escala de cinza de 8 bits, RGB, etc.

2.1.2 Histogramas de Imagens

O histograma de um sinal é a probabilidade que um valor tem de pertencer a uma certa faixa de intensidades, e é calculado segundo a Equação 2.3.

$$P_k = n_k/N, \quad (2.3)$$

em que n_k é o número de ocorrências da k -ésima intensidade do valor no sinal, k é o valor de intensidade que vai de 0 a $L - 1$ e N é o tamanho total de sinal [13].

O histograma pode ser aplicado a imagens, e neste caso, representa a probabilidade que um pixel tem de pertencer a uma certa faixa de intensidades, é calculado segundo a Equação 2.3 com as seguintes adaptações: n_k é o número de ocorrências da k -ésima intensidade na imagem, k é o valor de intensidade dos pixels que vai de 0 a $L - 1$ e N é o número total de pixels da imagem.

A menor unidade digital de uma imagem é um pixel, e cada pixel é representado por um número, que pode variar em faixa de valores dependendo do formato da imagem. As imagens binárias usam 1 bit para descrever cada pixel da imagem, o que as torna reais *imagens em preto e branco*. No entanto, as imagens em tons de cinza têm uma faixa mais ampla de valores para cada pixel. Em tons de cinza de 8 bits, por exemplo, cada pixel pode ter valores que variam entre 0 (zero) e 255. O número, via de regra, representa a intensidade da luz naquela imagem. Zero representa ausência total de luz, que é preta; o número mais alto representa a presença total de luz, que é branca. Em uma imagem binária, branco é o número 1 (um); em uma imagem em tons de cinza de 8 bits, o branco é 255.

Preto e branco podem também ser representados por imagens RGB, que possuem três canais de imagem, um para o vermelho, outro para o verde e o último para o azul. A adição desses três canais produz a imagem final. Portanto, há mais de uma matriz nesse caso.

Esta discussão sobre o espaço de cores serviu para introduzir a questão da representação de imagens em um computador. Neste trabalho, todas as imagens recebidas são convertidas para escala de cinza de 8 bits, e alguns procedimentos de tratamento de ima-

gem são feitos com elas em modo binário. No entanto, a escala de cinza pode ser vista como uma ampla gama de tons de cinza que se aproximam do preto à medida que o valor do pixel diminui em direção a zero e se aproximam do branco à medida que o valor aumenta em direção ao máximo - 255 para escala de cinza de 8 bits. Porém, alguns casos podem se tornar ambíguos, e a conversão de uma imagem da convenção de tons de cinza para binário, pode depender do critério que se utiliza para considerar um pixel preto ou branco na binarização.

Existem alguns algoritmos e técnicas para realizar tal conversão, e para essa operação uma estratégia vital é fazer um histograma da imagem. O histograma da imagem é a representação gráfica de sua probabilidade de ocorrência de intensidades versus valores de intensidade reais na imagem dada. [13] Em outras palavras, um histograma mostra quantos pixels de cada intensidade estão presentes na imagem, fornecendo uma maneira de analisar comparativamente a distribuição de intensidades.

Como método para realçar as bordas de uma imagem, para aumentar o contraste, uma estratégia é operar no histograma da imagem, usando a equalização do histograma. A equalização do histograma de uma imagem é um método para equalizar a referida distribuição de probabilidades de valores de intensidade na imagem. [13]

2.1.3 Método de Otsu para Binarização

Para converter a imagem de outro formato de cor para binário, neste trabalho o método selecionado foi o proposto por Nobuyuki Otsu [14]. O Método de Otsu é uma solução utilizada para resolver o problema de binarização, que consiste em achar o melhor limiar para dividir os pixels entre pretos e brancos vindo de um sistema de cores não-binário.

Com base no histograma cinza 1-D da imagem, este método busca a maior variância interclasse entre alvo e fundo como princípio para escolher o melhor limiar que os separa no histograma. Em outras palavras, ele escolhe o limiar que melhor divide as cores extremas dependendo do histograma da própria imagem, ou probabilidade de ocorrência de cor. Tal tarefa se torna mais difícil porque a imagem alvo tem cores semelhantes às do fundo, ou se tem muito ruído que borra o objeto e o fundo juntos. Simplificando, o algoritmo de Otsu realiza uma busca sequencial completa nos índices do histograma procurando o limiar k que minimiza a variância intraclasse, definida como uma soma ponderada das variâncias das classes analisadas. [14] Esta equação intravariância é representada pela Equação 2.4.

$$\sigma_{\omega}^2(k) = \omega_0(k)\sigma_0^2(k) + \omega_1(k)\sigma_1^2(k), \quad (2.4)$$

onde os pesos ω_0 e ω_1 são as probabilidades das duas classes $\{0, 1\}$ separadas por um limite k e σ_0^2 σ_1^2 são variações dessas duas classes. No caso da imagem, as classes são as cores preto e branco nas quais os valores de tons de cinza serão definidos.

2.1.4 Segmentação Automática de Imagens

A partir das várias operações de pré-processamento realizadas sobre uma imagem digital, torna-se mais fácil realizar a identificação e a seleção de regiões de interesse dentro dessa mesma imagem. Em inglês, as regiões de interesse são chamadas de Region Of Interest (ROI).

No caso deste trabalho, são realizadas uma sequência de operações para se chegar à segmentação automática da imagem, melhor detalhadas na Seção 4.1.2.

Com isso, tendo um conjunto original de imagens originais sem tratamento, após o pré-processamento e a segmentação automática, obtém-se um conjunto com aparência mais adequada ao uso computacional e já segmentada em regiões de interesse mais eficazes para a análise.

2.2 Aprendizado de máquina

A proposta principal deste trabalho é desenvolver um algoritmo de aprendizado de máquina que receba como entrada uma imagem radiográfica bucal panorâmica e forneça como saída a indicação se há sinais de osteoporose no paciente retratado por tal imagem. Nenhum diagnóstico assertivo pode ser feito nesse caso, pois a precisão da rede pode ainda não ser tão conclusiva como o diagnóstico de um médico [8], mas o que esse tipo de algoritmo pode fazer é fornecer uma primeira triagem do paciente, separando dois grandes grupos: pacientes livres de sinais de osteoporose e pacientes com sinais de osteoporose. Esse segundo grupo de pacientes, que apresentam sinais de osteoporose, quando encaminhados aos profissionais adequados para tratamento precoce, pode reduzir o risco de fratura óssea, que é o principal dano conhecido advindo da disfunção da osteoporose. [1]

Antes de fornecer uma imagem digitalizada ao algoritmo, este trabalho propôs um pré-processamento e uma segmentação automática que melhora a imagem e a reduz a uma região de interesse menor e mais específica, processo descrito na Subseção 4.1.2. Isso é necessário para ajudar o computador no processo de aprendizagem. Com um campo de dados muito amplo, é mais difícil detectar padrões, mas restringindo esse campo a uma região mais específica essa tarefa será mais eficiente. Agora, nesta seção, é descrito o algoritmo de aprendizado de máquina que recebe a ROI da imagem segmentada e tenta detectar sinais de osteoporose.

O aprendizado de máquina é o que possibilita aos humanos ensinar as máquinas a reconhecer padrões úteis e aprender com as informações resultantes de processos computacionais. Em outras palavras, aprender significa converter informação em conhecimento, e aprendizado de máquina é ensinar máquinas a realizar essa conversão. [15]

O aprendizado de máquina é uma aplicação de Inteligência Artificial (IA) que ajuda programas e sistemas de computador a aprender com a experiência, que é feita processando dados e adquirindo conhecimento desses dados. Cálculos matemáticos e estatísticos são feitos por computadores em um tempo muito mais rápido do que os humanos, quando eles são programados corretamente para isso. Assim, o aprendizado de máquina prospera principalmente quando o volume de dados a ser analisado é considerável, devido à sua capacidade de realizar tais cálculos rapidamente com grandes conjuntos de dados. [16] Nosso mundo moderno vem produzindo cada vez mais dados a cada dia, e isso faz com que seja cada vez mais necessário ter uma poderosa capacidade de processamento feita por computadores. [17]

No entanto, os computadores e seus programas precisam ser ensinados por humanos, eles não podem criar um programa sozinhos, eles sempre precisam da inteligência humana por trás apontando na direção certa com um objetivo claro. Se os computadores estiverem apontados na direção errada, ou sem um objetivo claro, toda a poderosa capacidade de cálculo pode ser desperdiçada. Nesse sentido, os computadores são grandes aliados da ciência, por possuírem meios de realizar cálculos muito mais rápidos do que nós humanos; no entanto, eles precisam de nossa inteligência para definir a direção ou o objetivo de seu cálculo.

2.2.1 Tipos de Aprendizado de Máquina

Um dos resultados do aprendizado de máquina é a classificação de dados. Quando definimos a direção e o objetivo de um algoritmo e também rotulamos os dados que fornecemos a ele para que ele também tenha um parâmetro um tanto confiável para fazer seus julgamentos, chamamos isso de *aprendizado supervisionado*. No aprendizado supervisionado ajudamos o computador a treinar sua capacidade de julgamento ou classificação.

Quando não fornecemos essa orientação completa e não fornecemos informações mais profundas sobre os dados para o algoritmo, chamamos isso de *aprendizagem não supervisionada*. No aprendizado não supervisionado, deixamos o computador livre para identificar padrões e agrupar informações de acordo com sua própria percepção. [15] Nesse sentido, existem dois caminhos principais para realizar o aprendizado de máquina: ajudar o computador a treinar ou não, que correspondem ao aprendizado supervisionado e não supervisionado, respectivamente.

Existe também a *aprendizagem por reforço*, que é basicamente o treinamento de algoritmos de aprendizado de máquina para chegar a uma sequência de decisões baseado em premiações ou punições. Essa técnica é utilizada em ambientes incertos onde não se sabe o resultado certo das ações, portanto, o algoritmo é levado a construir um critério de decisão recebendo premiações pelos seus acertos na direção certa e seus erros quando vai na direção errada. Com isso, no final, procura obter a sequência de decisões que maximiza as premiações, ou que minimiza as punições.

Alguns algoritmos de aprendizado supervisionado muito utilizados são *Support Vector Machine* (SVM), *Random Forest* (RF), Redes Neurais (RN), *Naive Bayes*, *Decision Tree* e *K-Nearest Neighbor*. [17] Este trabalho foca em técnicas de aprendizado supervisionado, utilizando os métodos de CNN e SVM mais especificamente, descritas nas subseções seguintes.

2.2.2 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA), ou simplesmente Redes Neurais (RN), de forma geral, são uma técnica de aprendizado de máquina que procura imitar os princípios de funcionamento e a estrutura do cérebro humano. Faz isso utilizando uma rede de nós análoga à rede de neurônios cerebral, que se conectam e se comunicam, transmitindo informações e detectando padrões pertinentes a essas informações passadas. A técnica já é antiga, com décadas de sua idealização, e com o avanço da tecnologia já deu origem a diversas variações e melhorias.

Até alguns anos atrás, conjuntos de dados compostos por imagens rotuladas para aprendizado supervisionado tendiam a consistir na ordem de dezenas de milhares de imagens, como apresenta Alex Krizhevsky et al em 2017 [18], citando alguns conjuntos de dados bem conhecidas e amplamente utilizados disponíveis para uso público, como NORB [19], Caltech-101/256 [20] [21] e CIFAR-10/100 [22].

Tarefas simples de reconhecimento podem ser resolvidas muito bem com conjuntos de dados desse tamanho, especialmente se houver a possibilidade de aumentar os dados preservando as propriedades do rótulo com as transformações. No entanto, para dados mais complexos, com maior variabilidade, é necessário usar conjuntos de dados maiores [18], e os problemas de usar conjuntos de dados pequenos para dados complexos e altamente variáveis são bem conhecidos no campo computacional [23].

Para conjuntos de dados maiores, os métodos de aprendizado de máquina precisam ter maior capacidade de aprendizado, e aqui as Redes Profundas aparecem como uma opção mais apropriada, devido à sua característica de amplitude e profundidade variáveis, mantendo menos conexões e parâmetros [24] [18] [1]. Esses atributos tornam as redes mais

adaptáveis de problema a problema e mais fáceis de treinar comparativamente a outras técnicas.

Quanto maiores os conjuntos de dados, mais difícil é analisar os dados contidos neles e extrair informações úteis, que são o cerne de todo processo de aprendizagem. Conjuntos de dados menores, por outro lado, podem não fornecer dados suficientes para concluir as características de todo o universo de onde esses dados vêm. Por exemplo, se um pesquisador tem apenas 100 imagens de boca para realizar uma pesquisa, esse pesquisador pode não ter informações que caracterizem corretamente as bocas de cada ser humano, que em número são mais de 7 bilhões hoje. Outra maneira de dizer isso é que a pesquisa pode ser muito específica para seu conjunto de dados.

Esta é uma das motivações para usar redes com mais camadas, [24] tornando-as convolucionais em vez de totalmente conectadas. Redes com muitas camadas são frequentemente chamadas de redes profundas, e são a base do Aprendizado Profundo ou *Deep Learning* (DL).

2.2.3 Redes Neurais Convolucionais (CNNs)

Existem muitos estudos que buscam encontrar a melhor relação entre o tamanho do conjunto de dados e o número de camadas de uma rede. [25]. Uma conclusão comum é que para cada problema existe uma configuração específica que funciona melhor. No entanto, o ponto principal das Redes Neurais Convolucionais é a possibilidade de lidar com grandes quantidades de dados em menos tempo computacional, [18] porque uma parte das camadas não está totalmente conectada, e o número de pesos e parâmetros a serem atualizados em cada iteração é então menor.

Entre e fora das camadas, existe a possibilidade de incluir algumas outras técnicas de processamento que auxiliam na classificação, como *janelamento* (*pooling*), *retificadores* (ex. *Função de ativação Rectified Linear Unit (ReLU)*), *dropout*, *log-softmax*, apenas para citar alguns, que foram utilizados neste trabalho.

Pooling fornece uma abordagem para reduzir o tamanho dos mapas de recursos sintetizando a presença de recursos em patches da entrada. Os métodos de pooling mais comuns são o pooling médio e o pooling máximo, que percorrem a entrada nas janelas resumindo, respectivamente, a presença média de um recurso e a presença mais ativada de um recurso. Camadas de pooling podem ser descritas pelas seguintes relações na Equação 2.5.

$$\begin{aligned}
\text{Input layer} &= W_1 \times H_1 \times D_1, \\
\text{Output layer} &= W_2 \times H_2 \times D_2, \\
W_2 &= [(W_1 - F)/S] + 1, \\
H_2 &= [(H_1 - F)/S] + 1, \\
D_2 &= D_1,
\end{aligned} \tag{2.5}$$

onde W é a largura da camada, H é sua altura e D é sua profundidade ou quantidade de canais. Além disso, F é a extensão espacial do *pooling*, ou seja, o tamanho do núcleo (kernel), e S é o passo, ou seja, o número de unidades que o filtro de pooling pula antes de avançar.

Existem apenas duas variações comumente vistas de uma camada de pooling máxima: uma camada de pooling com $F = 3$, $S = 2$ (ou seja, *pooling* sobreposto) e ainda mais comumente $F = 2$, $S = 2$ configuração, usada neste projeto. Tamanhos de *pool* com tamanhos de kernel maiores tendem a ser muito destrutivos para a entrada.

Funções de Ativação são um método para acionar os neurônios em redes neurais. Como nem todos os nós ou neurônios em uma rede neural convolucional são ativados, deve haver uma maneira de saber quando esse nó deve estar ativado ou não. A função de ativação é responsável por traduzir a entrada ponderada somada de um determinado nó na ativação desse nó. Algumas técnicas de retificação são sigmóide, tangente hiperbólica e função de ativação linear retificada. Uma unidade que usa a função de ativação linear retificada é chamada ReLU para abreviar. As duas primeiras abordagens mencionadas não podem ser usadas em redes multicamadas por causa de sua matéria gradiente evanescente; enquanto o ReLU supera esse problema de gradiente de fuga, de uma maneira que permite que os modelos tenham um desempenho melhor e aprendam mais rápido. Por essas razões, ReLU é o principal método de ativação usado no desenvolvimento de perceptrons multicamadas e CNNs. [1] A função ReLU de um array de entrada x pode ser descrita pela Equação 2.6.

$$ReLU(x) = (x)^+ = \max(0, x). \tag{2.6}$$

Dropout é uma técnica que reduz aleatoriamente o número de nós ativos em uma rede em tempo de execução, o que faz com que a rede varie de iteração para iteração, proporcionando mais variância e reduzindo o excesso de adequação de um modelo aos dados utilizados para gerá-lo (overfitting). As redes neurais profundas tendem a se ajustar muito rapidamente aos dados quando o conjunto de dados é baixo em número. *Dropout* oferece um método computacionalmente barato de regularização que reduz o overfitting

e também melhora o erro de generalização em redes neurais profundas.

Funções de Perda (Loss Functions) são funções que comparam os valores de saída de um algoritmo computacional, calculando a diferença ou o erro entre o valor previsto pelo algoritmo e o valor correto que deveria ser obtido; com isso, mede o quão bem a rede modela os dados fornecidos como entrada. Ao treinar uma rede, pretende-se minimizar essa perda entre os resultados previstos e os desejados.

Log-softmax é uma adaptação logarítmica do *Softmax*, um algoritmo que ajuda a classificar dados de entrada em classes de saída. *Softmax* é uma função matemática que converte uma matriz de números em uma matriz de probabilidades, na qual as probabilidades de cada valor são proporcionais à escala relativa de cada valor na matriz. Essa função é usada no aprendizado de máquina aplicado também como uma função de ativação. A equação para *Log-softmax* de um array de entrada x é mostrada na Equação 2.7.

$$\text{LogSoftmax}(x_i) = \log(\text{Softmax}(x_i)) = \log\left(\frac{\exp(x_i)}{\sum_j \exp(x_j)}\right). \quad (2.7)$$

Log-softmax é necessário para realizar o cálculo da probabilidade de que o resultado fornecido pela rede esteja correto usando a função de perda chamada *Negative Log Likelihood (NLL)*. Com *Log-softmax*, a saída da rede é transformada em um array de probabilidades. Quanto maior a probabilidade de uma classe na matriz, mais provável é para a rede que a entrada seja um exemplo dessa classe. Em seguida, calcula-se o logaritmo negativo na maior probabilidade desse array e, conseqüentemente, quanto menor for esse número final, mais preciso será o resultado obtido para o esperado no rótulo do exemplo. O cálculo de NLL é descrito na Equação 2.8.

$$\text{NLL} = -\log(\max(\text{LogSoftmax}(x_i))). \quad (2.8)$$

A função de perda da arquitetura da CNN deste trabalho foi implementada tomando os meios simples do NLL de ambas as classes de cada dado. Este processo para encontrar o NLL de classes c é descrito na Equação 2.9.

$$\text{NLLLoss} = \sum_{i=1}^c \frac{-\log(\text{LogSoftmax}(x_i))}{c}. \quad (2.9)$$

A CNN deste trabalho foi inspirada no artigo [1]. A arquitetura de implementação e as modificações serão descritas na Subseção 4.2.1.

2.2.4 *Support Vector Machine (SVM)*

Uma máquina de vetor de suporte (SVM) é uma técnica de aprendizado de máquina supervisionado que usa algoritmos de classificação para problemas de classificação de dois grupos. Fornecendo a uma SVM conjuntos de dados de treinamento rotulados para cada categoria, eles podem categorizar novos dados após treinamento. [26]

A técnica SVM é bastante conhecida por sua capacidade de classificar tanto dados linearmente separáveis e dados não linearmente separáveis ao aplicar uma função kernel à equação do hiperplano [26]. Neste método de classificação, queremos atingir um máximo de distância da margem definida entre dois hiperplanos paralelos, como mostra a Equação 2.10.

$$f(x_k) = \sum_{i \in SV} \alpha_i \omega_i k(x_i, x), \quad (2.10)$$

com $k = \{1, \dots, K\}$, x como vetores de características no espaço n-dimensional, e a função kernel $k(x_i)$, em que se define um rótulo de classe $f(x_k) \in \{1, 1\}$ para os pontos da amostra x_k , para obter um hiperplano classificador.

Os limites para a tomada de decisão são definidos como $\{\alpha_i \omega_i k(x_i, x) + b\} = 0$. Uma SVM pega pontos de dados espalhados em um plano e gera um hiperplano, que em duas dimensões é simplesmente uma linha, que melhor separa os rótulos. Essa linha é o limite de decisão: qualquer coisa que caia de um lado dela, é classificado como Classe 1 e qualquer coisa que caia do outro lado, como Classe 2, por exemplo.

As principais funções kernel utilizadas podem ser Gaussiana, Gaussiana de Base Radial (RBF), Sigmoid, Polinomial e Linear.

Em comparação com algoritmos mais recentes, como Redes Neurais (RN)s, eles têm duas vantagens principais: maior velocidade e melhor desempenho com um número limitado de amostras (na casa dos milhares). Isso torna o algoritmo muito adequado para problemas com conjuntos de dados menores.

2.2.5 *Random Forest (RF)*

Uma floresta aleatória, ou *Random Forest* (RF), é um classificador que consiste em uma coleção de classificadores estruturados em árvore de decisão, ou *Decision Trees* (DT), de acordo com a Equação 2.11.

$$h(x, \Theta_k), k = 1, \dots, \quad (2.11)$$

onde os $\{\Theta_k\}$, são vetores aleatórios identicamente distribuídos independentes e cada árvore lança um voto unitário para a classe mais popular na entrada x . [27] O erro de

generalização para florestas converge até um limite à medida que o número de árvores na floresta aumenta. O erro de generalização de uma floresta de classificadores de árvores depende da força das árvores individuais na floresta e da correlação entre elas.

A RF estende as ideias de uma árvore de classificação e regressão, na qual o conjunto de dados é recursivamente dividido de acordo com o limiar da variável de decisão que melhor separa os dados de treinamento em nós filhos de diferentes classes. [28] Em uma floresta aleatória, muitas dessas árvores são construídas por meio da agregação bootstrap (seleção aleatória de pontos de dados da amostra com substituição, também conhecida como bagging), o que ajuda a evitar o treinamento excessivo do modelo que pode levar ao *overfitting* (overtraining).

Além disso, em cada nó de decisão, o melhor ramo de um subconjunto selecionado aleatoriamente das possíveis variáveis de decisão é tomado. Isso aumenta a independência das árvores na floresta. Cada árvore individual pode então produzir uma estimativa (possivelmente fraca) da probabilidade de classificação do objeto, tomando o conjunto de variáveis de decisão através dos nós de ramificação da árvore. A probabilidade de saída dessa árvore é então a proporção de tipos de classe no nó de decisão final. [28]

Portanto, a RF é uma técnica de aprendizado de máquina que consiste na construção de um conjunto de árvores de decisão em tempo de treinamento. Para tarefas de classificação, a saída da floresta aleatória é a classe selecionada pela maioria das árvores. São utilizadas em problemas de classificação e regressão. RFs são frequentemente chamadas de modelos "caixa preta", pois geram previsões razoáveis em uma ampla gama de dados, exigindo pouca configuração. [28]

A técnica constrói árvores de decisão em diferentes amostras e leva sua maioria de votos para classificação e média em caso de regressão. Uma das características mais importantes do RF é que ele pode lidar com o conjunto de dados contendo variáveis contínuas como no caso de regressão e variáveis categóricas como no caso de classificação, mas costuma apresentar melhores resultados para problemas de classificação.

Capítulo 3

Revisão de Literatura

Neste capítulo é apresentada uma revisão da literatura, retratando algumas das abordagens atuais para a questão da identificação da osteoporose por meio de imagens médicas. Em seguida, os artigos foram selecionados relativos ao uso da inteligência artificial no auxílio de um primeiro diagnóstico de osteoporose em radiografias panorâmicas bucais. E, por fim, o artigo principal utilizado como guia para o direcionamento deste trabalho foi explorado com mais detalhes, incluindo todos os parâmetros de sua proposta de estrutura de Rede Neural Convolutacional. [1]

3.1 Identificando a osteoporose usando imagens médicas

A osteoporose é uma disfunção na densidade mineral do osso que leva a uma tendência maior do que o normal de fratura em caso de acidente, como queda ou esbarrar em um objeto. [29] Além disso, é uma doença silenciosa, o que dificulta a detecção antes que tal evento ocorra, tornando-o conhecido apenas pela pessoa que a possui após a fratura. [30] Assim, a disfunção da osteoporose tem dois problemas importantes, segundo o consenso médico o primeiro problema é a sua detecção; por se tratar de uma doença silenciosa, ou seja, sem sintomas apresentados pela pessoa que a tem, ela é descoberta principalmente quando já existem problemas concretos. O que leva ao segundo problema, que é o seu impacto. Pessoas com osteoporose tendem a sofrer muito mais danos em um acidente do que pessoas sem ela. [29]

No campo da informática, o objetivo dos trabalhos não diz respeito à segunda parte do problema, que é o dano causado pela presença da osteoporose em um acidente. No entanto, muito pode ser - e já está sendo - feito em relação à primeira parte do problema: a detecção da osteoporose. Esta primeira parte da revisão da literatura mostra

algumas das abordagens atuais da ciência da computação para enfrentar o problema da detecção da osteoporose. Estudos têm demonstrado que quanto mais precoce a disfunção é diagnosticada, menores são seus impactos e maiores são as chances de tratamento. [7]

As áreas de Inteligência Artificial com Aprendizado de Máquina e Visão Computacional são especialmente úteis nesta tarefa de detecção, pois através de algoritmos computacionais é possível identificar estruturas e padrões que o olho humano dificilmente perceberia, com maior precisão e também com menos flutuações. [8] Essas identificações já estão possibilitando que os médicos recebam um grande auxílio na identificação não só da osteoporose, mas de um grande número de doenças. Esses algoritmos dificilmente podem substituir todo o conhecimento humano sobre o universo biomédico, mas podem ser de grande ajuda para buscar detalhes e diagnósticos que poderiam passar despercebidos se um médico não os procurasse especificamente. Eles também podem ser ótimos primeiros ensaios para direcionar ainda mais os pacientes para tratamentos específicos. É importante conhecer o que está sendo pesquisado atualmente no campo biomédico-computacional, e os problemas que estão à frente de suas descobertas, que são o tema das Seções seguintes.

3.2 Abordagens Estado da Arte na Identificação Computacional da Osteoporose

Algumas das abordagens atuais para a questão da detecção automática de osteoporose em levante são apresentadas nesta seção. Foram escolhidos artigos e trabalhos escritos até 2017, porque esse campo de pesquisa está emergindo rapidamente e suas descobertas estão substituindo rapidamente as alcançadas antes delas. Ainda não é um campo de pesquisa amplamente escolhido, mas já existem alguns trabalhos importantes, que melhoram em grande porcentagem a acurácia do diagnóstico de osteoporose, principalmente quando comparada à acurácia humana. [12]

3.2.1 Abordagens de Análise de Textura

A seguir estão alguns dos trabalhos que abordam o tema do diagnóstico de patologias dentárias por meio da análise de textura.

O artigo *Caracterização de Patologias Dentárias Utilizando Imagens de Raios-X Panorâmicos Digitais com base na Análise de Textura* [31] sugere que o exame radiográfico de rotina da boca seja feito com equipamentos que geram muito ruído, e que muitas vezes leva a uma inúmeras dificuldades na detecção de patologias dentárias. A detecção de algumas patologias dentárias pode ser melhorada com a ajuda de computadores. O primeiro

algoritmo apresentado pelo artigo é uma nova abordagem chamada *transformação negativa hibridada*. A segunda parte do artigo apresenta uma *análise estatística de textura*. A textura das imagens é caracterizada pelo que é chamado de *Gray Level Cooccurrence Matrix (GLCM)*. As características de textura obtidas a partir desta análise são energia, entropia, homogeneidade, contraste e correlação. Esses recursos podem ser usados para encontrar a textura limites para segmentar melhor as imagens. Os modelos resultantes do trabalho foram validados com o diagnóstico de radiologistas maxilofaciais.

Técnicas de diagnóstico por imagem atuais e emergentes para avaliação de osteoporose e risco de fratura [12] é uma revisão feita em 2018 que apresenta alguns dos avanços mais recentes e de última geração nas técnicas de imagem para diagnóstico osteoporose. Apresenta também alguns métodos que têm sido utilizados para segmentar regiões de interesse de imagens, um passo inicial frequentemente utilizado antes de qualquer algoritmo ser aplicado a conjuntos de dados. Esta é uma etapa que também foi implementada neste trabalho, explicada com mais detalhes no Capítulo 3. Além disso, algumas análises de textura também são apresentadas.

O próximo artigo desta revisão de literatura, *Detecção Automática Baseada em Computador e Classificação da Osteoporose em Radiografias Ósseas* [32] apresenta uma nova abordagem para ajudar a diagnosticar osteoporose em radiografias panorâmicas, imagens que normalmente não têm muita definição visual. O algoritmo proposto é um método totalmente automático para seleção, classificação e detecção de recursos usando classificadores de aprendizado de máquina, como *Support Vector Machines (SVMs)* e *Nearest Neighbors (NNs)*, resultando em uma precisão relevante de 95 % no diagnóstico.

Por fim, o artigo *Classificação da Osteoporose Usando Recursos de Textura* [33] apresenta uma abordagem para ajudar a diagnosticar a osteoporose usando características de textura dos ossos da boca. Apenas a olho nu, algumas características podem ser difíceis de discernir, é quando surgem os computadores, com sua eficiente capacidade de identificar padrões em imagens. O método consiste em filtros Gabor multi-resolução e recursos de padrão binário local 1D (1DLBP). Os resultados mostram uma acurácia diagnóstica de 72,71%, que é superior a outros resultados contemporâneos em comparação.

3.2.2 Abordagens de Aprendizado de Máquina

O artigo *Classificação da estrutura óssea trabecular de pacientes osteoporóticos usando visão de máquina* [34] propõe um método baseado no processamento automático de imagens de raios-X. Consistiu na validação de desempenho de um método de extração de características através de quatro classificadores: *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *Naive Bayes* e *Artificial Neural Networks* em imagens do osso calcâ-

neo, com baixo tempo complexidade. Seus resultados experimentais indicam uma taxa de classificação de 98% usando o classificador SVM, que foi o mais promissor dos quatro.

Este outro artigo *Machine Learning Principles Can Improve Hip Fracture Prediction* [8] apresenta uma abordagem de aprendizado de máquina que comprova a importância de prever fraturas de quadril e estima a eficiência dessa previsão em absorciometria de raios X de dupla energia (DXA) - escaneou homens e mulheres. Os dados DXA vindos da Dinamarca entre 1996 e 2006 foram usados para criar 24 modelos estatísticos que foram construídos em 75% dos pontos de dados por meio de k-5, validação cruzada de 5 repetições e, em seguida, validados nos 25% restantes dos pontos de dados para calcular a área sob a curva (AUC) e calibrar as estimativas de probabilidade. Para as mulheres, o melhor modelo foi "*bagFDA*" (análise discriminante flexível agregada de bootstrap) com área sob curva (AUC) de 0,91; para os homens, o melhor foi "*xgbTree*" (eXtreme Gradient Boosting) com AUC de 0,89. A conclusão apresentada pelo artigo é que o aprendizado de máquina pode melhorar a previsão de fratura de quadril além da regressão logística usando modelos de conjunto.

Publicado em 2018, o artigo *Octuplet Siamese Network For Osteoporosis Analysis On Dental Panoramic Radiographs* [35], propõe a estrutura para uma *Deep Octuplet Siamese Network (OSN)* para criar modelos que predizem osteoporose por meio de recursos extração, aumento de dados online e ajuste fino com pesos pré-treinados AlexNet. Algumas estatísticas são mostradas nos resultados deste artigo em que a precisão de algumas outras técnicas de identificação são avaliadas usando critérios de validação cruzada *leave-one-out*.

3.2.3 Revisões sobre Identificação Computacional da Osteoporose

A revisão crítica fornecida por *Inteligência artificial na identificação de grupos de risco para osteoporose, uma revisão geral* [36] analisa os maiores sistemas que usam inteligência artificial para identificar grupos de risco para osteoporose ou fratura. Algumas características foram estudadas na revisão: abrangência no diagnóstico, baixo custo e capacidade de identificar fatores somáticos mais significativos. Foi realizada uma pesquisa bibliográfica nas seguintes bases de dados: PubMed, IEEE Explorer Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (LILACS), Sistema de Análise e Recuperação de Literatura Médica Online (MEDLINE), Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, Web of Science e Science Direct pesquisando os termos "Neural Network", "Osteoporosis Machine Learning" e "Osteoporosis Neural Network".

O artigo *Inteligência Artificial Aplicada à Osteoporose: Uma Comparação de Desempenho de Algoritmos de Aprendizado de Máquina na Previsão de Fraturas de Fragilidade a*

partir de *Dados de Ressonância Magnética* [37] identifica os classificadores de aprendizado de máquina que melhor predizem fraturas ósseas originadas da osteoporose e, por meio dos dados, destacar os recursos de imagem e regiões anatômicas que mais contribuem para o desempenho da previsão. Os três classificadores mais bem qualificados foram *árvores impulsionadas por RUS, regressão logística e discriminante linear*.

Oriundo do mesmo autor deste último, esta publicação *Inteligência artificial, osteoporose e fraturas de fragilidade* [38] teve como objetivo avaliar as últimas pesquisas feitas com a aplicação de inteligência artificial na previsão de osteoporose entre janeiro de 2017 e março 2019. A maior parte dos artigos prevê por classificação ou regressão algum indicador de osteoporose, como massa óssea ou fraturas por fragilidade. Outra grande parte deles estuda ferramentas usadas para segmentar automaticamente imagens provenientes de pacientes com risco de osteoporose. A fonte dos dados de imagem vem de várias fontes hoje em dia: acústica, Ressonância Magnética (RM), Tomografia Computadorizada (TC) e as radiografias panorâmicas comuns.

3.2.4 Classificação da Osteoporose

Cita-se ainda dois artigos-base deste trabalho, do professor André Leite da Faculdade de Saúde da UnB, com seus professores colegas, *Radiografia panorâmica: instrumento auxiliar no diagnóstico da osteoporose e Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women*. [39, 40] Nestes trabalhos, procura-se uma correlação entre o Índice Mandibular Cortical, e outros índices bucais, e a presença da osteoporose no paciente. Esse índice, como sendo uma medida objetiva de distância, por poder ser analisado computacionalmente, foi o ponto de partida e a premissa do presente trabalho, por se tratar também de radiografias panorâmicas bucais, porém, visto de forma médica. Portanto, deu-se uma abordagem computacional como continuação deste trabalho importante na área de biomédica.

3.2.5 Revisões sobre Identificação Computacional da Osteoporose

Para concluir esta revisão de literatura, apresenta-se o seguinte artigo, que é a abordagem que mais se aproxima da proposta deste trabalho. A linha principal deste trabalho é projetar a estrutura de uma *Convolutional Neural Network (CNN)* para auxiliar o diagnóstico da osteoporose através da avaliação de radiografias panorâmicas. As métricas para validar os resultados foram os diagnósticos anteriores provenientes de radiologistas bucomaxilofaciais. Esses especialistas em odontologia foram integrados ao projeto em

parceria com o Departamento de Odontologia da Universidade de Brasília (UnB), e as imagens do conjunto de dados foram gentilmente cedidas também por eles.

A estrutura inicial da CNN deste trabalho foi estudada e adaptada a partir do trabalho descrito no parágrafo seguinte. No entanto, alguns dos parâmetros da estrutura não puderam ser reproduzidos, e essas adaptações e detalhes estão comentados na Seção 4.

O referido artigo é *Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study* [1], que é um artigo de pesquisa que propõe a estrutura para um *Deep Convolutional Neural Network (DCNN)* para auxiliar o diagnóstico de osteoporose com computadores. Os diagnósticos automatizados foram avaliados através de uma comparação com os diagnósticos de radiologistas bucomaxilofaciais, assim como no conjunto de dados do presente trabalho.

Primeiramente, as imagens foram feitas e, em seguida, especialistas diagnosticados através da observação da região cortical do osso maxilar, um método comum para verificar a osteoporose. O número final de imagens foi de 1.268, provenientes de mulheres com idade média de 52 anos. Entre essas pacientes do sexo feminino, 635 não apresentavam osteoporose e 633 apresentavam. Para obter um diagnóstico auxiliado por computador, as imagens tiveram que ser segmentadas e reduzidas, este é um primeiro passo. O artigo compara então três estruturas semelhantes, uma treinada com imagens monocentradas da boca (SC-DCNN), uma também com imagens monocentradas, mas com aumento de dados (SC-DCNN (Augment)), e finalmente uma treinada com duas regiões laterais de interesse da mesma imagem em vez de uma única imagem centrada (MC-DCNN). Entre todo o conjunto de dados, 200 imagens foram usadas para testar a estrutura da rede após o treinamento.

Os valores da área sob a curva (AUC) obtidos usando SC-DCNN, SC-DCNN (Augment) e MC-DCNN foram 0,9763, 0,9991 e 0,9987, respectivamente. Isso mostra uma alta concordância com os diagnósticos fornecidos pelos profissionais de radiologia. A principal conquista dessa abordagem é a possibilidade de detectar uma doença de difícil diagnóstico por meio de um exame comum e rotineiro, que são as radiografias panorâmicas. Essa abordagem não tem a intenção de substituir os odontologistas em seu trabalho, mas sim fornecer um método fácil e de baixo custo para identificar precocemente a doença de alto impacto que é a osteoporose.

Utilizando uma estrutura semelhante, o presente trabalho visa obter resultados aproximados em comparação com este artigo de estado da arte de 2018, mas com a adição de uma segmentação automática de imagem feita previamente, e com algumas adaptações em relação aos hiperparâmetros.

Além disso, o presente trabalho busca comparar a estrutura proposta neste referido artigo com outras configurações de hiperparâmetros e com algumas adaptações relevantes:

- Aplica-se a um conjunto bastante heterogêneo de dados;
- Utiliza radiografias adquiridas de forma analógica, com menor qualidade;
- Aplica-se a pessoas idosas de comunidades carentes, então tem um impacto não só tecnológico, mas social;
- Compara outras técnicas computacionais à CNN, como SVM e RF, procedimento que é descrito no Capítulo 4.

Capítulo 4

Metodologia

A metodologia proposta por este trabalho conta com um processo de dois passos, que é composto por Processamento de Imagens e Aprendizado de Máquina. O fluxograma de tais passos é mostrado na Figura 4.1.

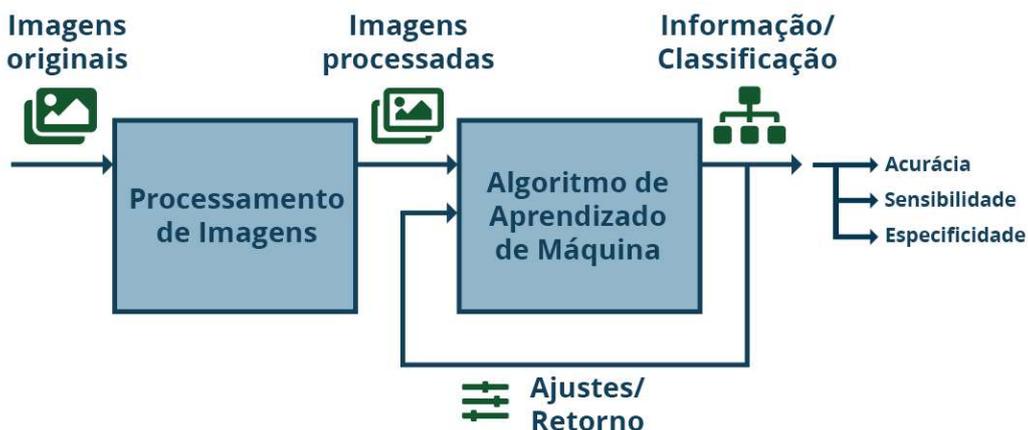


Figura 4.1: Fluxograma completo do trabalho.

O Fluxograma pode ser descrito com os seguintes passos:

- Preprocessamento de imagens:
 - Primeiramente, são recebidas imagens originais "brutas" da boca provenientes de radiografias panorâmicas;
 - Com o processamento de imagens, essas imagens brutas são transformadas em imagens processadas e segmentadas para conter apenas a ROI;
- Aprendizado de Máquina:
 - As imagens segmentadas são fornecidas ao Algoritmo de Aprendizado de Máquina (CNN, RF ou SVM), que extrai informações sobre elas;

- Essa informação serve como meio de classificação se o paciente representado pela imagem apresenta ou não sinais da doença osteoporose;
- A cada interação do algoritmo, novas informações são geradas sobre seu desempenho na classificação, e esse feedback é fornecido à rede, possibilitando ajustes em sua estrutura e seus parâmetros;
- Ao final, o algoritmo fornece os resultados de classificação do conjunto de dados, que são comparados com os rótulos originais, e dessa comparação derivam as métricas de acurácia, especificidade e sensibilidade (ver Capítulo 5).

A etapa de Processamento de Imagem, descrita com mais detalhes na Seção 4.1, é retratada na Figura 4.2.

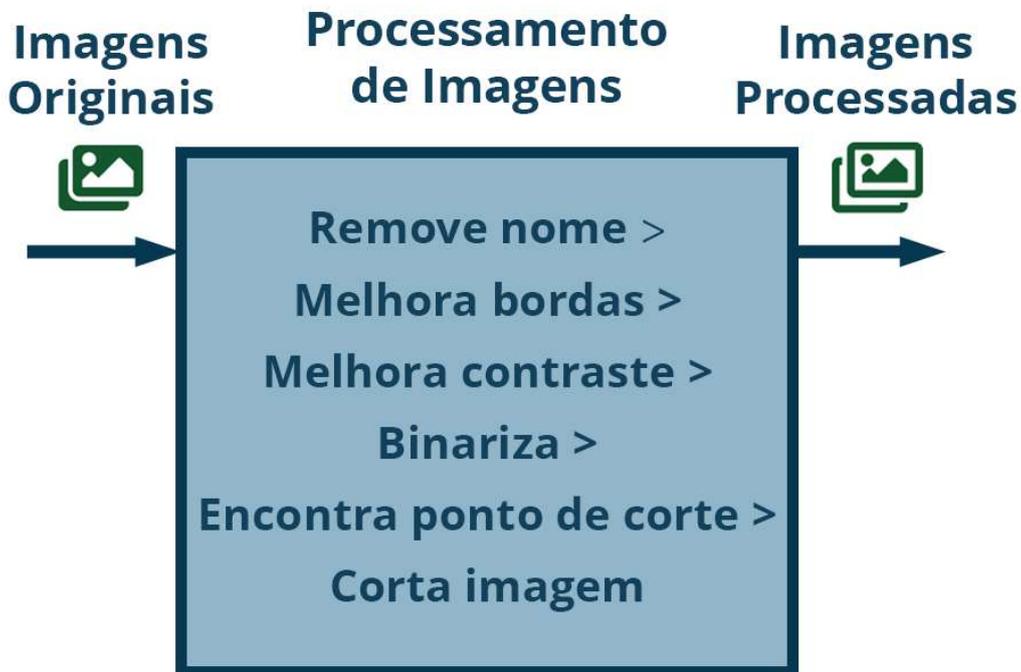


Figura 4.2: Fluxograma do Processamento de Imagens.

A arquitetura do algoritmo de aprendizado de máquina, descrita com mais detalhes na seção 4.2, é retratada na Figura 4.3.

4.1 Preprocessamento de Imagens

Para fornecer melhores entradas para algoritmos de aprendizado de máquina, as imagens biomédicas devem primeiro ser geradas por meio de um exame médico e depois segmentadas com algum tipo de método computacional, para que o resultado final do processo

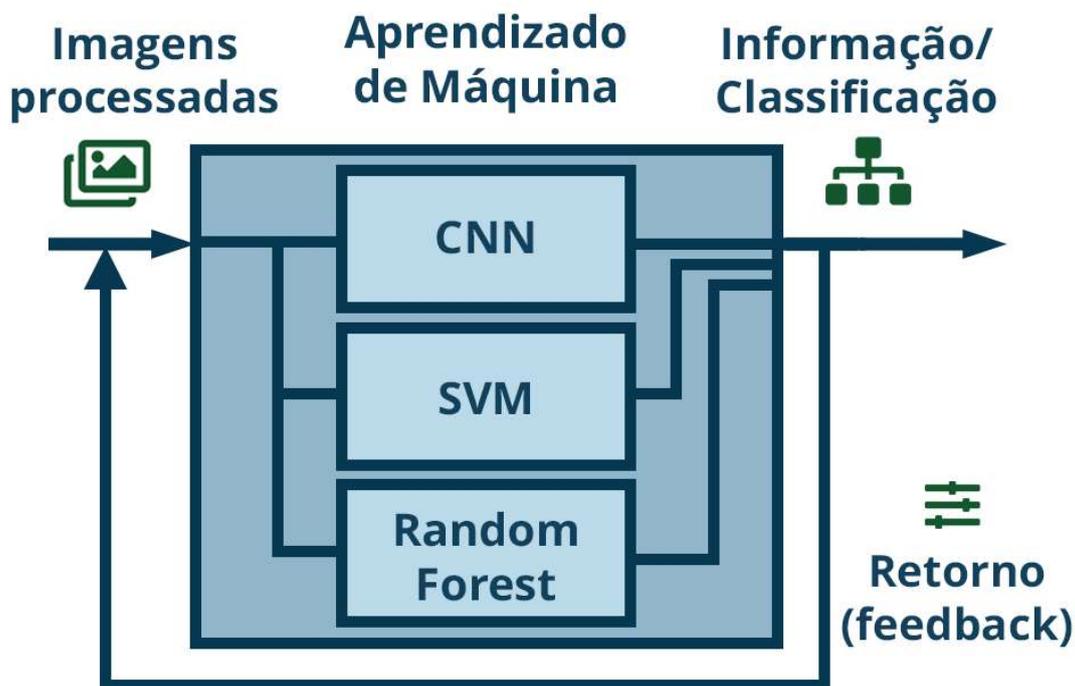


Figura 4.3: Fluxograma do Aprendizado de Máquina.

seja uma seção que melhor represente a região de interesse dentro da imagem. [12] A primeira parte deste processo será apresentada na Subseção 4.1.1, e a segunda será melhor explicada na Subseção 4.1.2.

4.1.1 Aquisição de Imagens

As imagens utilizadas neste trabalho foram as radiografias panorâmicas bucais, comumente solicitadas como exame de rotina para avaliar a saúde da boca de um paciente. Um exemplo da imagem deste exame é mostrado em Figura 4.4.

A Figura 4.4 é um exemplo comum de radiografia panorâmica antiga, quando as imagens eram impressas em um material especial sensível aos raios X. Nessas imagens anteriores havia até uma etiqueta com o nome do paciente e outras informações pertinentes (vista no lado esquerdo da referida imagem), o que nas imagens mais recentes não é mais feito. Para tornar esta imagem digital, é necessário digitalizar a imagem com um dispositivo. As imagens do conjunto de dados deste trabalho foram digitalizadas usando o dispositivo EPSON Expression 1680¹, usando uma configuração de 600 dpi.

¹Mais informações sobre o dispositivo de digitalização em *Epson Expression 1680 website*.

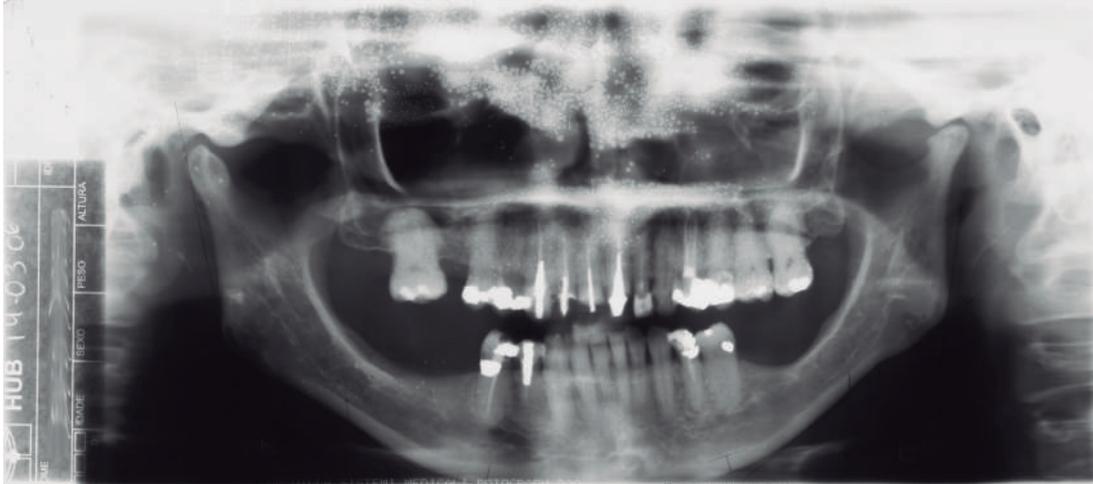


Figura 4.4: Exemplo de uma imagem radiográfica panorâmica digitalizada.

Nas técnicas recentes de aquisição de imagens, as radiografias já são geradas em formato digital, dispensando o processo de digitalização. Um exemplo de uma imagem de radiografia já digital é mostrado em Figura 4.5.



Figura 4.5: Example of a Digital Panoramic Radiographic Image.

Como regra geral, essas imagens já digitais têm melhor definição do que as originalmente analógicas que passam por um processo adicional de digitalização. E não há necessidade de um rótulo na própria imagem, porque os arquivos finais do computador podem ser nomeados ou rotulados para essa finalidade. Porém, mesmo com todas as vantagens das radiografias panorâmicas já digitais, devido ao fato de todas as imagens do conjunto de dados deste trabalho já estarem impressas, e devido à dificuldade de aquisição de dados de pacientes biomédicos nesta localidade, foi utilizada a primeira técnica de

aquisição. Ou seja, todas as imagens do conjunto de dados deste trabalho são analógicas posteriormente digitalizadas.

4.1.2 Segmentação de imagem

Existem muitas abordagens a respeito da imagem biomédica adquirida para analisá-la e extrair dela informações de saúde. Várias dessas abordagens foram detalhadas em Capítulo 2.

Uma etapa antes de inserir a imagem no algoritmo de aprendizado de máquina é comum entre muitas dessas abordagens, que é primeiro segmentar a imagem para que o algoritmo tenha informações focadas nas quais confiar. [31, 35]

Além disso, a maioria dos artigos e artigos sobre o campo de pesquisa mesclado computação-biomédica credita a eficiência da saída do algoritmo em quão bem a imagem adquirida inicialmente é segmentada. [1]

Há muitas maneiras de determinar qual parte de uma imagem será cortada dela para ser analisada mais detalhadamente por um programa de computador. Normalmente, uma região é selecionada por causa de sua significativa congregação de informações sobre o tópico em foco. Comumente esta parte da imagem é chamada de Region Of Interest (ROI).

Alguns trabalhos contam com a textura da imagem para fazer essa segmentação, [12] outros a realizam definindo uma porção fixa do tamanho inicial. [1] Mas um princípio comum é frequentemente adotado: que a imagem deve ser segmentada automaticamente, devido ao fato de que os conjuntos de dados raramente são pequenos. Em vez de cortar conjuntos de dados que às vezes ultrapassam milhares de imagens uma a uma, é mais eficiente criar algoritmos de computador que fazem o mesmo trabalho, com menos esforço bruto.

Assim, a primeira parte deste trabalho foi, após adquirir as imagens para formar o conjunto de dados, criar um algoritmo computacional que segmente automaticamente as radiografias panorâmicas para auxiliar a rede de aprendizado de máquina reduzindo seu campo de possibilidades em cada imagem. Ou seja, um programa que extrai automaticamente a ROI da radiografia.

Outra vantagem importante do processo automático de segmentação de imagens é que, para os propósitos sociais deste trabalho e de muitos outros, é de suma importância que todas as verificações sejam feitas por computador. O objetivo deste trabalho é criar um método automatizado para detectar sinais de osteoporose, nas ocasiões em que o dentista ou radiologista especialista não estaria procurando especificamente por ela. Por exemplo, uma pessoa entra em uma clínica de radiologia com a intenção de verificar sua cárie, mas ao executar automaticamente o programa de computador com a imagem de seu

exame como entrada, ele também pode ser verificado automaticamente quanto a sinais de osteoporose. Isso poderia evitar a detecção da doença apenas quando seus danos já foram causados.

Muitos softwares/programas podem realizar uma tarefa de segmentação de imagem, mas *MathWorks MATLAB*² foi utilizado neste trabalho, na versão 2019-b.

A ROI visada para este trabalho é a faixa cortical do osso mandibular inferior. A espessura dessa região é considerada um dos principais indicadores ou sinais de osteoporose, comumente chamada de *índice cortical mandibular*. [39, 1]

A Figura 4.6 mostra as aparências comuns desta ROI em radiografias panorâmicas. A imagem foi emprestada de um trabalho feito sobre o assunto [40] pelos especialistas em radiologia que ajudaram este trabalho como parceiros, como fornecedores das imagens para o conjunto de dados e também como diagnosticadores das imagens para a fase de treinamento da rede.

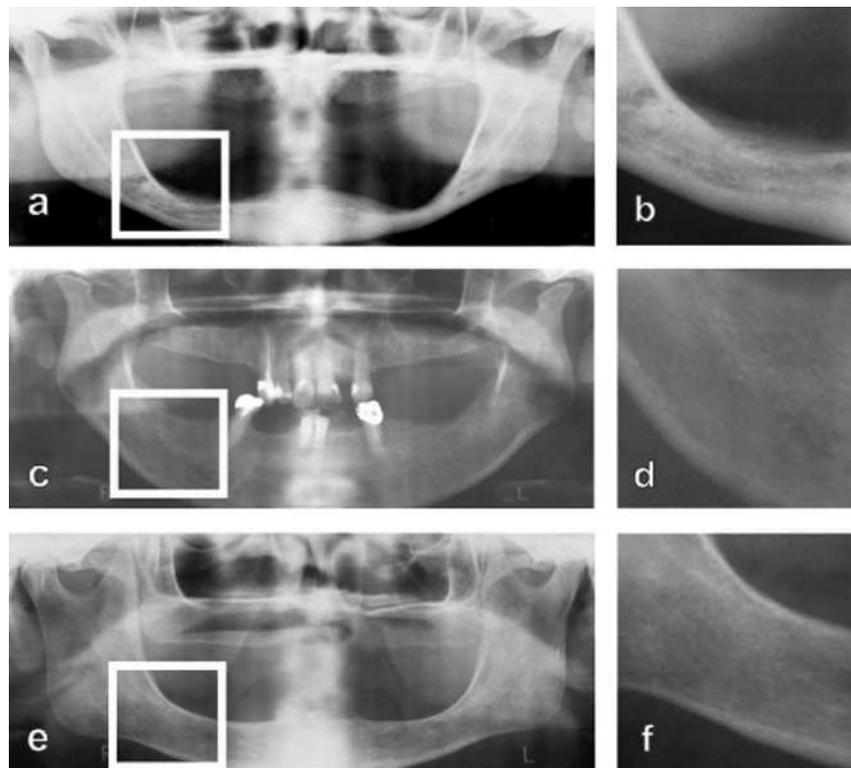


Figura 4.6: Exemplos de aparência radiográfica do córtex mandibular inferior em radiografias panorâmicas (Fonte: [40]).

Uma primeira explicação do procedimento de segmentação automática para obter esta ROI é apresentada a seguir:

1. Converte a imagem para tons de cinza (se já não estiver);

²Para mais informações, visite *MathWorks MATLAB*.

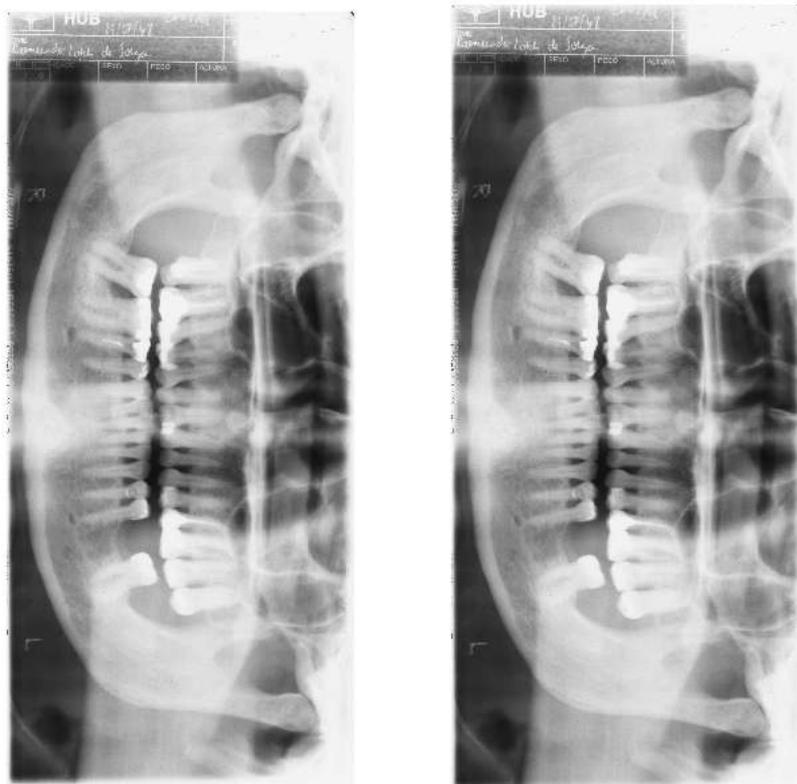
2. Gira a imagem para as proporções horizontal/paisagem (se já não estiver);
3. Remove a etiqueta do paciente (no caso deste trabalho cortando a parte esquerda da imagem em tamanho fixo);
4. Realça as bordas da imagem usando o filtro laplaciano;
5. Melhora o contraste da imagem usando equalização de histograma;
6. Converte imagem em binário usando o limite de tons de cinza do método de Otsu [14];
7. Calcula a soma dos pixels pretos binários em cada coluna. $x_{bordas} = \max(\sum_{i=0}^l g(y_i))$, em que $g(y_i)$ é a quantidade de pixels pretos na coluna i com x_{bordas} , sendo que a coluna com mais pixels binários que deve representar as bordas externas da mandíbula x_{bordas} (o ponto de referência do corte vertical). Ao calcular isso, o meio das bordas da mandíbula deve coincidir com o centro da boca, de modo que a imagem recortada fique centralizada;
8. Calcule a soma dos pixels pretos binários em cada linha da imagem: a linha com mais pixels pretos binários deve representar a parte mais baixa do queixo (o ponto de referência do recorte horizontal);
9. Do queixo para cima, 30% da altura da imagem é selecionada para cortar a faixa horizontal da imagem final;
10. A última parte da segmentação é cortar cada imagem verticalmente mantendo a mesma proporção em todas elas. Isso é feito escolhendo a largura do corte que termina com imagens de mesma proporção, que é 5:1.

Detalhando que: Todos os cortes são feitos com dimensões proporcionais à imagem original, ou seja, não com tamanho de pixel fixo. Isso torna o algoritmo livre do formato de imagem específico que foi usado. A partir dessa rotina de segmentação, a imagem de saída sempre tem uma proporção de 5:1 (largura:altura). Para o algoritmo de aprendizado de máquina, posteriormente, todas as imagens são reduzidas para ter o tamanho real de 1000x200 pixels, portanto, manter a mesma proporção era uma forma de manter a proporção no algoritmo. Este tamanho foi usado no artigo de referência [1].

Esses passos listados foram retratados como imagens para melhor representar o processo de segmentação. Após as imagens a seguir, há uma descrição mais detalhada de cada etapa, com números e justificativas quanto às decisões tomadas ao desenhar a rotina de segmentação.

Em primeiro lugar, a rotina de segmentação começa com o carregamento de uma imagem da base de dados usando o MATLAB. Todos os demais procedimentos são feitos

considerando que esta imagem está no modo de cores em tons de cinza, ou seja, uma matriz $m \times n$ que contém valores na faixa de 0 (preto) a 255 (branco). A imagem original pode estar no formato de cor RGB, por exemplo. Assim, essa transformação é necessária para que as etapas a seguir funcionem. Uma forma de verificar se a imagem está em tons de cinza ou RGB é verificar quantas dimensões tem a imagem que representa a matriz. Se tiver três dimensões, está no modo RGB, então a imagem é convertida em escala de cinza. Se a matriz tiver apenas duas dimensões, ela será preservada. A conversão da imagem para tons de cinza corresponde ao Passo 1 da rotina de segmentação. Se a imagem já estiver no modo de cores em tons de cinza, essa etapa é desnecessária.



(a) Imagem original

(b) imagem convertida em tons de cinza (Etapa 1)

Figura 4.7: Imagens convertidas em tons de cinza e originais lado a lado

A etapa seguinte é girar a imagem para garantir que ela esteja na orientação horizontal/paisagem. Isso é feito por uma simples verificação de qual dimensão da imagem é maior, altura ou largura. Se a altura for maior, a imagem é girada em 90 graus. Isso corresponde ao Passo 2 da rotina de segmentação.

Em seguida, retire a faixa da imagem que contém a etiqueta do paciente, muito comum em imagens analógicas. Nas imagens utilizadas neste trabalho, esses rótulos estavam todos



Figura 4.8: imagem rotacionada (Etapa 2).

na parte mais à esquerda da imagem. Isso corresponde ao Passo 3.



Figura 4.9: Imagem sem etiqueta do paciente (Etapa 3).

O passo 4 corresponde ao realce das bordas da imagem utilizando um filtro laplaciano multidimensional 3×3 que aproxima a forma do operador Laplaciano bidimensional.

O passo 5 corresponde ao aumento do contraste da imagem utilizando a equalização do histograma.

O passo 6 corresponde à binarização da imagem em poucos passos. Primeiro, a imagem é convertida para precisão dupla, então, um limiar de imagem global em tons de cinza é definido utilizando o método de Otsu [14]. Em seguida, a binarização é realizada através



Figura 4.10: Imagem com bordas aprimoradas (Etapa 4).



Figura 4.11: Imagem com contraste aprimorado (Etapa 5).

do limiar de Otsu, considerando todos os pixels acima dele como brancos, e abaixo como pretos.

Todos os cálculos posteriores são feitos em relação à imagem binária. Na Etapa 7, para encontrar o queixo para o recorte horizontal, a soma dos pixels pretos binários é calculada para cada linha da imagem. Os pixels pretos neste tipo de imagem representam uma parte sem osso e os pixels brancos representam partes com osso. Apenas o último sexto da imagem na horizontal é analisado, pois é onde ficará a região do queixo. Neste cálculo da soma não são considerados os primeiros cinco sextos da imagem. A região onde o máximo de pixels pretos é calculado é mostrada na Figura 4.13 dentro das linhas verdes.

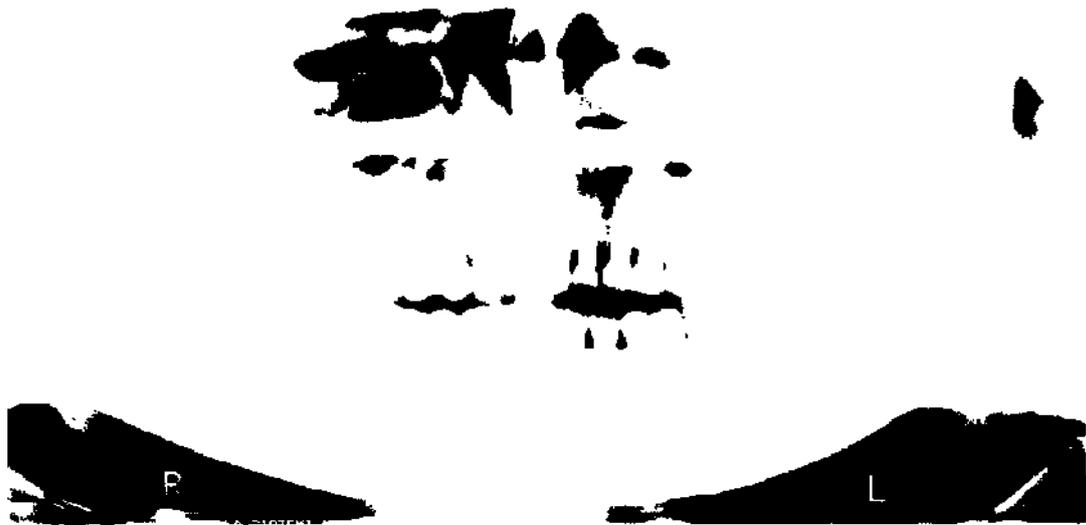


Figura 4.12: Imagem convertida em binário (Etapa 6).

A primeira linha horizontal com o maior número de pixels pretos binários nesta região deve representar o queixo. As próximas linhas com a mesma quantidade de pixels pretos devem estar acima do queixo, portanto, uma maneira de identificar o queixo é obter a primeira linha com a soma horizontal máxima de pixels pretos.

Há um problema com este método: as radiografias panorâmicas são uma imagem tirada normalmente através da varredura do paciente de lado a lado da face. Comumente, alguma porção do pescoço também é mostrada na imagem final. Esses ossos laterais que aparecem podem afetar o cálculo da soma de pixels, mas esse foi o melhor e mais simples método testado até agora para definir o queixo.

A ROI da imagem é uma porção definida do queixo para cima, correspondendo a uma faixa com aproximadamente 30% da altura total. Esse percentual foi a primeira proposta do nosso algoritmo, mas outros percentuais poderão ser testados posteriormente. Decidiu-se que uma porcentagem era melhor do que um número fixo de pixels, para reprodução do experimento para outros conjuntos de imagens. Este corte horizontal corresponde ao Passo 8, e é representado por Figura 4.14.

A última parte da segmentação é calcular onde cortar a imagem verticalmente com a preocupação de manter a mesma proporção em todo o conjunto de dados. Essa foi uma escolha feita para este trabalho, pois existem maneiras de usar ROIs de diferentes tamanhos no algoritmo de aprendizado de máquina por meio de downsampling de imagens primeiro ou pré-processamento delas. O artigo de referência [1] que orientou este trabalho utilizou uma proporção de pixels de 1000 x 200 para seu conjunto de dados. Por esta razão, a proporção da imagem deste trabalho foi de 5:1, para que o down-sample feito posteriormente fosse menos distorcido, e para que ficasse em escala em todas as imagens.

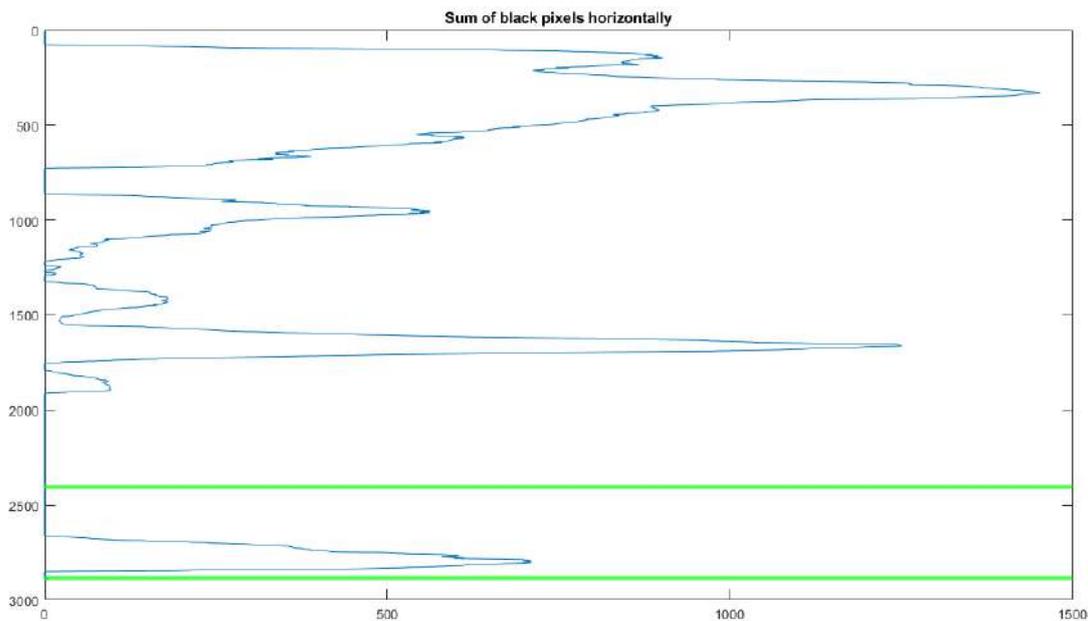


Figura 4.13: Gráfico contendo a soma dos pixels pretos binários de cada linha na imagem (Etapa 7).

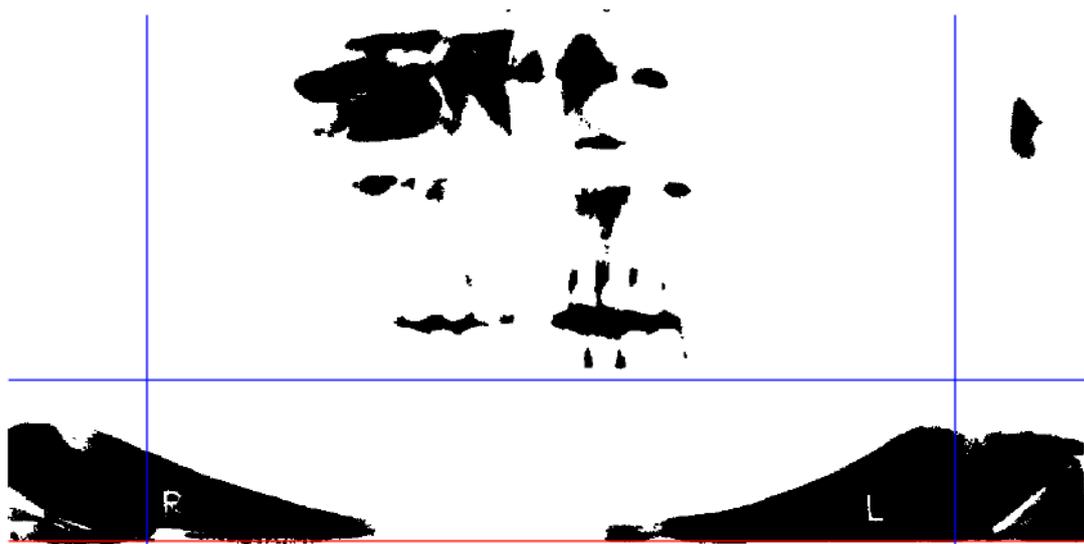


Figura 4.14: Imagem binária com a região de recorte marcada por linhas (Etapa 8).

Assim, a partir do centro da imagem, a região de corte vertical é cinco vezes o tamanho do corte horizontal. Foi decidido que as referências de corte seriam colocadas em relação ao centro da imagem. Algumas tentativas de definir o centro da imagem automaticamente foram feitas, mas o fato é que pouco se ganhou com a outra abordagem que é apenas obter a metade da dimensão horizontal, em relação ao fato de as radiografias panorâmicas virem na maioria das vezes já centralizadas. As ROIs cortadas que vieram do cálculo automático

foram piores do que as que usaram o ponto médio real da imagem.



Figura 4.15: Imagem binária cortada (resultado da Etapa 8).

Em outras palavras, a altura de corte horizontal é multiplicada por cinco, este é o tamanho de corte vertical. Metade dele será usado para cortar à esquerda do centro e metade à direita. Isso corresponde ao Passo 9.



Figura 4.16: Imagem em tons de cinza com a região de recorte marcada por linhas (Etapa 9).

Após a Etapa 9, a imagem completa da boca tirada do dispositivo de escaneamento representado pela Figura 4.7 é transformada na ROI representada pelo corte final da imagem na Figura 4.17. Isso conclui o algoritmo de segmentação que foi usado até agora neste trabalho, que é transformar a imagem da boca cheia em uma porção da boca representada por uma faixa horizontal logo acima do queixo. Essa faixa é o que alimenta a entrada do algoritmo de aprendizado de máquina.

Outras etapas podem incorporar este algoritmo, a fim de testar o algoritmo de aprendizado de máquina não usando uma faixa totalmente centrada, mas dividida com uma ROI menor de cada lado do osso da mandíbula, mais semelhante à região representada na Figura 4.6.



Figura 4.17: Imagem em tons de cinza cortada - Corte Final (resultado da Etapa 9).

O artigo usado como referência para este trabalho [1] já testou a rede com listras totalmente centradas e divididas, e em seus resultados, as totalmente centradas com aumento de dados e as divididas sem essa técnica obtiveram resultados semelhantes. Para ilustrar como serão as imagens divididas nas próximas fases deste trabalho, duas imagens usando o algoritmo de segmentação aprimorada já implementado são mostradas a seguir, Figuras 4.18 a 4.19.

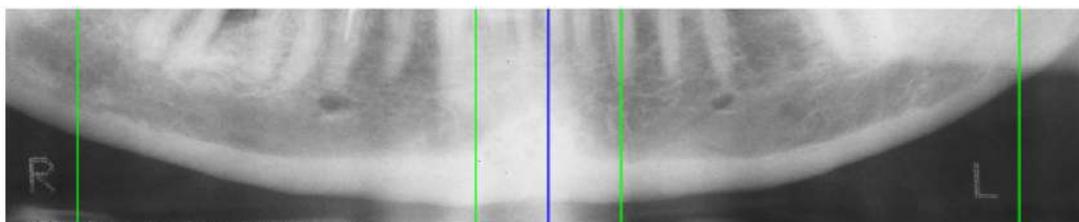


Figura 4.18: Imagem dividida em tons de cinza com região de corte marcada por linhas.

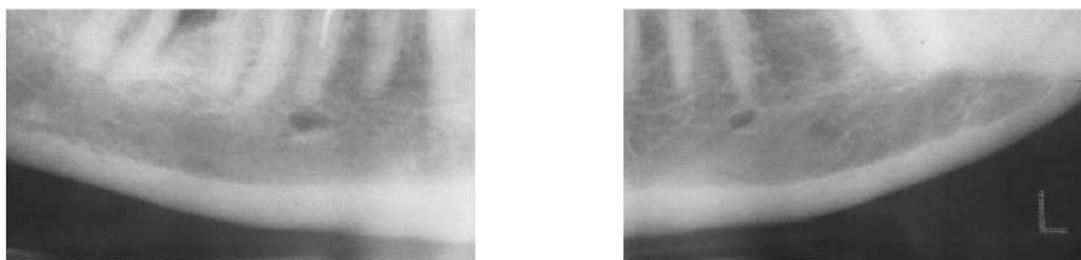


Figura 4.19: Imagem dividida em tons de cinza.

4.2 Metodologia Proposta

Assim como foi mencionado na Seção 2.2, a proposta deste trabalho é desenvolver um algoritmo de aprendizado de máquina que receba como entrada uma imagem radiográfica panorâmica da boca e forneça como saída a indicação se houver sinais de osteoporose

no paciente retratado por tal imagem. A estrutura para tal algoritmo é descrita nos parágrafos seguintes.

O conjunto de dados utilizado inicialmente neste trabalho consiste em 340 radiografias panorâmicas fornecidas e previamente diagnosticadas por especialistas em saúde do Departamento de Odontologia da Universidade de Brasília (UnB).

Todas as imagens foram submetidas ao algoritmo de segmentação de imagens descrito na Sub-subseção 4.1.2. Esta é a entrada das técnicas de aprendizado de máquina propostas. Como formas de Aprendizado de Máquina, foram utilizadas a Rede Neural Convolutiva (*Convolutional Neural Network*, CNN em inglês), e a Floresta Aleatória (*Random Forest*, RF em inglês), e Support Vector Machina (SVM) como técnica comparativa. A linguagem utilizada foi Python.

4.2.1 CNN Proposta

Primeiramente, será descrita a CNN principal deste projeto. O objetivo primário de uma CNN é receber dados e informações de saída sobre esses dados. No caso de um problema de classificação de imagens, uma imagem é fornecida como entrada e uma classe ou a probabilidade de a imagem pertencer a uma classe é fornecida como saída. Isso é feito enviando os dados brutos dessa imagem por meio de várias camadas que processam matematicamente esses dados e fornecem um resultado matemático no final.

Assim como descrito na Subseção 2.2.3, entre as camadas deste trabalho, alguns processamentos foram feitos entre e fora das camadas, tais como: *pooling* com tamanho de janela 2, *ReLU*, *dropout* de 50% e *log-softmax*.

Um exemplo de tal configuração é mostrado na Figura 4.20. Este exemplo de configuração da CNN representa uma implementação comum, e foi utilizado como inspiração para a arquitetura proposta deste trabalho.

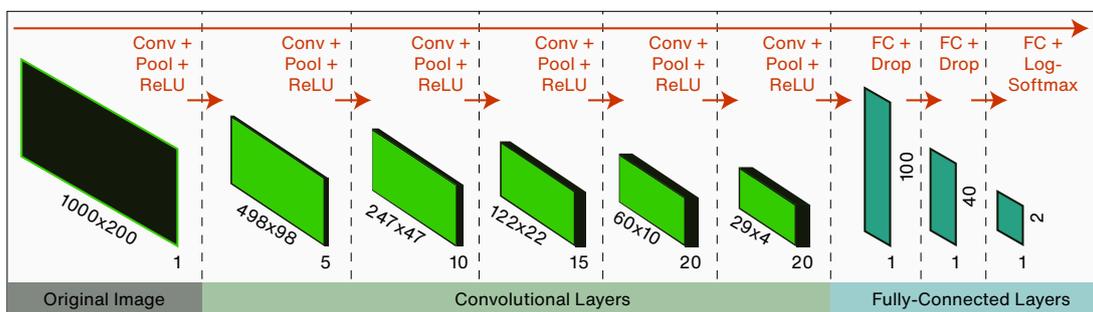


Figura 4.20: Estrutura da CNN proposta, inspirada na referência [1].

A estrutura proposta da CNN deste trabalho foi implementada em Python.

Estrutura da CNN em Python

```
class CNN(nn.Module):
    def __init__(self):
        super(CNN, self).__init__()
        self.conv1 = nn.Conv2d(1, 5, 5)
        self.conv2 = nn.Conv2d(5, 10, 5)
        self.conv3 = nn.Conv2d(10, 15, 5)
        self.conv4 = nn.Conv2d(15, 20, 3)
        self.conv5 = nn.Conv2d(20, 20, 3)
        self.pool = nn.MaxPool2d(2, 2)
        if downsample_opt == True:
            self.fc1 = nn.Linear(540, 100)
        else:
            self.fc1 = nn.Linear(63360, 100)
        self.fc2 = nn.Linear(100, 40)
        self.fc3 = nn.Linear(40, 2)
        self.dropout = nn.Dropout(0.5)
        self.logsoftmax = nn.LogSoftmax(dim = 1)

    def forward(self, x):
        x = self.pool(F.relu(self.conv1(x.float())))
        x = self.pool(F.relu(self.conv2(x.float())))
        x = self.pool(F.relu(self.conv3(x.float())))
        x = self.pool(F.relu(self.conv4(x.float())))
        x = self.pool(F.relu(self.conv5(x.float())))
        x = x.view(x.size(0), -1)
        x = self.dropout(self.fc1(x)) # dropout with p = 0.5
        x = self.dropout(self.fc2(x)) # dropout with p = 0.5
        x = self.logsoftmax(self.fc3(x.float())) # Using log softmax to match NLLLoss

    return x
```

As camadas da CNN de referência podem ser descritas como:

Tabela 4.1: CNN Referência

Camada	Entrada	Saída	Kernel
Conv1	1 canal	5 canais	5x5
Conv2	5 canais	10 canais	5x5
Conv3	10 canais	15 canais	5x5
Conv4	15 canais	20 canais	3x3
Conv5	20 canais	20 canais	3x3
FC1	100	40	-
FC2	40	10	-
FC3	10	1	-

Cada camada de convolução 2D funciona de maneira semelhante a um filtro que possui vários canais de entrada, vários canais de saída e um tamanho de kernel. Nenhum `padding` ou `stride` foram usados nas camadas convolucionais. O agrupamento entre as camadas foi de tamanho 2x2. A ativação `ReLU` também foi usada entre as camadas convolucionais. Cada camada nesta customização apresentada é descrita como:

1. Camada de Convolução 1: 1 canal de entrada, 5 canais de saída, tamanho do kernel de 5x5, `ReLU 2D`, `Maxpooling` com janela 2x2;
2. Camada de Convolução 2: 5 canais de entrada, 10 canais de saída, tamanho do kernel de 5x5, `ReLU 2D` `Maxpooling` com janela 2x2;
3. Camada de Convolução 3: 10 canais de entrada, 15 canais de saída, tamanho do kernel de 5x5, `ReLU 2D` `Maxpooling` com janela 2x2;
4. Camada de Convolução 4: 15 canais de entrada, 20 canais de saída, tamanho do kernel de 3x3, `ReLU 2D` `Maxpooling` com janela 2x2;
5. Camada de Convolução 5: 20 canais de entrada, 20 canais de saída, tamanho do kernel de 3x3, `ReLU 2D` `Maxpooling` com janela 2x2.

Após essas camadas, camadas totalmente conectadas foram implementadas com 50% de `dropout`. A estrutura tem um filtro `log-softmax` no final, porque a função de perda é a função *Negative Log Likelihood* (NLL), que fornece saída logarítmica.

O tamanho da imagem em pixels não importa para as camadas convolucionais, pois elas são separadas do tamanho de entrada. As camadas totalmente conectadas funcionam com matrizes 1-D. A fórmula usada para calcular o tamanho do pixel de uma imagem é:

$$h_l = [(h_{l-1} - f + 2 * p) / s] + 1, \quad (4.1)$$

onde h_f é a altura da imagem na camada l , f é o tamanho quadrado do kernel do filtro, p é o `padding` e s é o `stride`.

As camadas totalmente conectadas (`fully-connected`) nesta seção são descritas como:

1. Camada totalmente conectada 1: Pega a saída das camadas convolucionais e a transforma em um array de tamanho 100;
2. Camada totalmente conectada 2: Pega a saída de tamanho 100 da primeira camada totalmente conectada e a transforma em um array de tamanho 40;
3. Camada totalmente conectada 3: Pega a saída de tamanho 40 das camadas convolucionais e a transforma em um array de tamanho 2.

Essa matriz de tamanho 2 é então processada por meio de uma função `log-softmax` para transformar a resposta em valores da mesma dimensão e forma que a entrada no intervalo $(-\infty, 0)$. Essa transformação é necessária mais tarde no loop de treinamento para calcular a função de perda, que é NLL.

Todo esse processamento é feito através da Computação Acelerada por GPU, utilizando o toolkit CUDA. Isso é feito para acelerar os cálculos e é uma prática comum no aprendizado de máquina.

4.2.2 Treinando e Testando a Rede

O processo de treinamento da CNN consistiu em fazer um loop do conjunto de dados comparando esses dois valores para prever se o paciente tem osteoporose ou não, e comparar esse valor único posteriormente também com o rótulo de diagnóstico anterior que a imagem possui. Essa comparação origina um erro, que no aprendizado de máquina é calculado usando o que é chamado de **função de perda**. A perda acumulada no final de um loop de treinamento fornece um vislumbre de quão bem a rede está aprendendo a prever o diagnóstico com base apenas na imagem e em seus próprios parâmetros internos, sem ajuda humana.

No entanto, sempre na fase de treinamento de aprendizado de máquina, existe o perigo de que a rede esteja aprendendo a prever resultados muito específicos para esse conjunto de dados exclusivo, não para todas as imagens que ele pode receber. Isso torna necessário ter outra fase no desenvolvimento da rede que é a fase de testes. Nesta fase, a rede é submetida a um conjunto de dados diferente daquele que treinou. Isso geralmente é feito separando o conjunto de dados original em dois por uma proporção escolhida. O conjunto de dados de treinamento geralmente é maior que o de teste.

Também é possível introduzir uma etapa de validação durante o treinamento para simular a fase de teste, para períodos de treinamento mais longos. Isso torna o processo ainda mais inteligente e eficiente, e já é uma prática comum. Isso foi feito neste projeto.

A fase de treinamento consistiu em uma primeira etapa de treinamento real e uma segunda etapa de validação, que faz com que a rede avalie seu desempenho em diferentes imagens do treinamento. O treinamento aconteceu em loops por todo o conjunto de dados de validação de treinamento, um ciclo que geralmente é chamado de "época" no aprendizado de máquina. Um número anterior de épocas é definido para o treinamento e ele percorre o conjunto de dados até que esse número de épocas termine.

Após o treinamento, a rede foi submetida a teste, ou seja, outro conjunto de dados diferente das fases de treinamento e validação. Este teste é o resultado real que diz se a rede será capaz de funcionar em uma imagem geral fornecida a ela posteriormente em uma aplicação real.

A arquitetura CNN foi treinada, validada e testada usando o seguinte conjunto de dados de tamanho 324*, dividido em três grupos diferentes:

- Tmagens para a fase de treinamento;
- Imagens para a fase de validação;
- Imagens para a fase de testes.

* O tamanho original do conjunto de dados era de 340 imagens, porém, após uma triagem, 16 destas não foram consideradas apropriadas para o treinamento por conterem muitos artefatos que atrapalhariam na classificação. Isso acontece em parte por causa da má qualidade das imagens que são analógicas e posteriormente digitalizadas, e também pela má qualidade dos aparelhos comuns em nosso país. Porém, para suprir essa deficiência de imagens, o conjunto foi duplicado ao dividir as imagens em dois lados, como mostrado na Figura 4.19.

Assim como já foi descrito nas seções anteriores, as imagens foram previamente diagnosticadas por especialistas em radiologia odontológica e receberam um rótulo que as classificava entre ter ou não osteoporose.

O conjunto de dados de treinamento foi balanceado para ter o mesmo número de imagens de cada uma das duas classes. Ou seja, 100 pacientes com osteoporose e 100 pacientes sem osteoporose, escolhidos aleatoriamente. Em vez de passar por cada imagem individualmente, ou seja, uma abordagem ao vivo, este trabalho implementou a abordagem de mini-lote, sendo cada mini-lote de tamanho 4. O tamanho do mini-lote foi determinado pela capacidade computacional máxima disponível no computador que foi usado nas execuções. (Tudo foi implementado em Python, com aceleração de GPU usando o kit de ferramentas CUDA.)

Aqui estão alguns outros parâmetros de treinamento usados:

- Inicialização de pesos usando o método uniforme de Xavier;
- Otimização usando algoritmo Adam com taxa de aprendizado $lr = 0,001$.

4.2.3 SVM Proposta

A técnica SVM é consideravelmente mais fácil de configurar que a CNN computacionalmente. Para este trabalho, foram utilizadas as configurações a seguir:

```
# Estrutura da SVM em Python
x_train,y_train = acquire_data()
param_grid={
```

```
'C':[0.001,0.01,0.1,1,10,100,1000],
'gamma':[0.001,0.01,0.1,1,10,100,1000],
'kernel':['rbf','poly']}]
svc=svm.SVC(probability=True)
model=GridSearchCV(svc,param_grid)
model.fit(x_train,y_train)
```

Nestas configurações, foi utilizada a Classificação Suporte- C^3 , com busca de configurações ótimas utilizando⁴, da biblioteca `scikit learn`.

Nesta busca por hiperparâmetros ótimos, a grade de parâmetros utilizada foi o parâmetro de regulação $C = [0,001;1000]$, com passo aumentando em potências de 10; $gamma = [0,001;1000]$, com passo aumentando em potências de 10; e o `kernel` entre `rbf` (função de base radial) e `poly` (polinomial). Nesses passos consiste o "treinamento" do algoritmo SVM proposto como técnica comparativa adicional.

4.2.4 Random Forest (RF) Proposta

A técnica Random Forest também é mais fácil de configurar que a CNN computacionalmente. Para este trabalho, foi utilizado o classificador `RandomForestClassifier`⁵.

```
# Estrutura da RF em Python
x_train, y_train = acquire_data()
best_n = 0
best_acc = 0.0
best_model = RandomForestClassifier(n_estimators=10)

# Loop
for i in range(10, 201, 5):
    model = RandomForestClassifier(n_estimators=i)
    model.fit(x_train, y_train)

    current_acc = model.score(x_test, y_test)

    if(current_acc >= best_acc):
        best_acc = current_acc
        best_n = i
        best_model = model
```

Para a RF, foi feito um loop com estimadores entre 10 e 200, com passo de 5, calculando o melhor número de estimadores e salvando sua acurácia e seu modelo final.

³ Função `sklearn.svm.SVC`

⁴ Função `sklearn.model_selection.GridSearchCV`

⁵ Função `sklearn.ensemble.RandomForestClassifier`

Capítulo 5

Resultados

Neste trabalho é proposto um algoritmo de segmentação automática como etapa de pré-processamento nas imagens. Foram estudadas diversas técnicas de aprendizado de máquina supervisionado para obter os melhores resultados dado o caso específico já mencionado de imagens de baixa qualidade e pacientes carentes.

Para avaliar os algoritmos, compara-se os seus resultados com os rótulos recebidos, obtendo os valores de:

- Verdadeiros Positivos (VP);
- Verdadeiros Negativos (VN);
- Falsos Positivos (FP);
- Falsos Negativos (FN).

E então, para cada um se calculou três métricas:

- Acurácia = $(VP + VN)/(VP + VN + FP + FN)$;
- Sensibilidade = $(VP)/(VP + FN)$;
- Especificidade = $(VN)/(VN + FP)$.

A proposta deste trabalho é melhorar os resultados da técnica de referência, a CNN do artigo [1], ao utilizar a estrutura CNN para realizar uma detecção automática de osteoporose com alta sensibilidade e utilizar o método RF para um sistema com especificidade alta.

De forma similar ao artigo [1] uma das técnicas utilizadas for uma CNN A estrutura da rede proposta neste trabalho partiu dos parâmetros propostos no artigo de referência já mencionado, porém, nem todas as características eram favoráveis a essa estrutura proposta. Uma grande diferença é o conjunto de dados, que no artigo em questão era

digital, de boa qualidade e homogêneo, e neste trabalho eram imagens analógicas obtidas com aparelhos antigos e escaneadas, o que faz com que sejam bastante heterogêneas e muito pouco regulares.

Essa diferença nos dados gerou a necessidade de algumas mudanças na rede para os resultados da classificação serem melhores, porém, isso serve até como uma referência de como são as imagens no Brasil, país onde foram extraídas. Ainda existem regiões do país com métodos de obtenção das radiografias antigos como o deste trabalho, então cremos que o estudo pode ter relevância ainda maior para casos onde as imagens não tem alta qualidade.

Para avaliar os resultados do trabalho, será comentado o conjunto de dados e em seguida o resultado comparado das técnicas já descritas de aprendizado de máquina.

5.1 Conjunto de Dados

O conjunto de dados utilizado neste trabalho consiste em 340 radiografias panorâmicas fornecidas e previamente diagnosticadas por especialistas em saúde do Departamento de Odontologia da Universidade de Brasília (UnB). O conjunto de dados foi aprovado pela comitê de ética, o que é sempre um requisito para esse tipo de trabalhos com dados de pacientes reais.

Todas as imagens foram submetidas aos algoritmos de segmentação de imagens descritos anteriormente. Esta é a entrada da CNN e da RF propostas. Suas saídas são uma matriz de dois valores contendo a probabilidade de que o paciente representado por aquela imagem tenha ou não osteoporose.

As arquiteturas foram treinadas, validadas e testadas usando o seguinte conjunto de dados de tamanho 340. Como as imagens são divididas em regiões esquerda e direita, o conjunto duplica de tamanho, indo para 680 entradas, divididas em três grupos diferentes:

- 237 ($\times 2 = 474$) imagens na fase de treinamento (70%);
- 35 ($\times 2 = 70$) imagens na fase de validação (10%);
- 68 ($\times 2 = 136$) imagens na fase de testes (20%).

O conjunto de dados de treinamento foi balanceado para ter o mesmo número de imagens de cada uma das duas classes. Ou seja, 100 pacientes com osteoporose e 100 pacientes sem osteoporose, escolhidos aleatoriamente. Este trabalho implementou a abordagem de mini-lotes, (*mini-batches*), sendo cada mini-lote de tamanho 4. O tamanho do mini-lote de processamento foi determinado pela capacidade computacional máxima disponível no computador que foi usado nas execuções. Também foram utilizadas técnicas de aumento

de dados nas execuções em tempo de treinamento para aumentar a variabilidade dos dados.

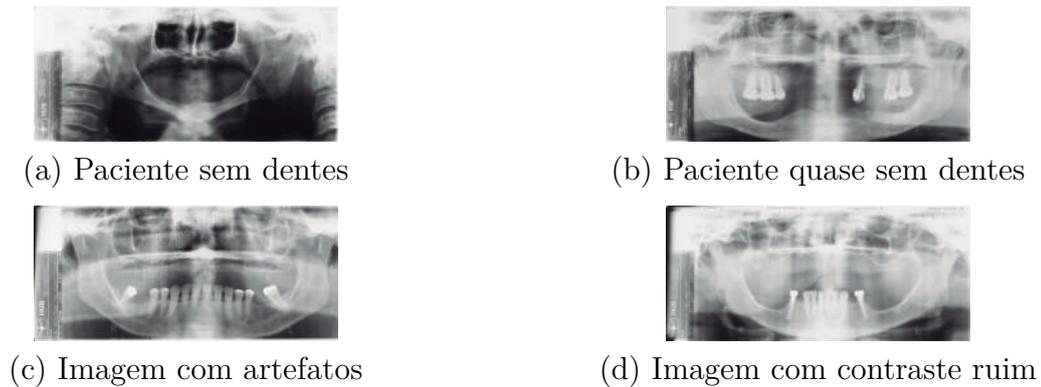


Figura 5.1: Diferentes problemas nas imagens da base de dados.

As figuras mostradas na Fig. 5.1 retratam alguns exemplos das diferentes dificuldades encontradas no conjunto de dados, tanto na variedade de pacientes quanto na técnica de aquisição. Esses problemas dificultam a identificação e a classificação das imagens pois poluem a imagem com informações indesejadas.

Porém, esse é o tipo de imagens que conseguimos obter atualmente em nosso país, considerando as regiões de baixa renda e também a população que tem essas características bucais. Portanto, esse trabalho tem uma relevância social importante por lidar com imagens mais reais ao contexto social do Brasil, quando comparado com as imagens de alta qualidade do artigo de referência, por exemplo.

5.2 Melhorias propostas

Para melhorar os resultados de acurácia obtidos com o conjunto de dados mencionado, também foram realizadas algumas tentativas:

- Data augmentation em forma de rotação, espelhamento e filtro *blur*. Essas técnicas melhoraram os resultados e foram levadas adiante até o fim dos experimentos;
- Aumento do número de classes para 3 em vez de 2. Essa mudança não melhorou os resultados e foi abandonada depois de ser testada;
- Outras configurações de hiperparâmetros na rede CNN para melhorar a estrutura de referência. Vários testes experimentais foram feitos com esses hiperparâmetros, utilizando princípios do estado da arte. O que obteve melhores resultados, e mais abrangentes, foi o apresentado no Capítulo 4;

- *Ensemble* juntando as três técnicas avaliadas (CNN, SVM e RF), utilizando *reamostragem com o objetivo de tornar os resultados estatísticos mais genéricos (cross-validation)*. Essa tentativa de juntar os resultados em uma técnica só não melhorou os resultados, mas foi implementada de forma superficial, sem aprofundamento, e foi abandonada depois de ser testada. Porém, pode ser melhorada para tentar aumentar a acurácia nos resultados.

5.3 Resultados Comparativos

Foram utilizadas as seguintes estrutura de aprendizado de máquina para este trabalho:

- A estrutura da CNN de referência [1];
- A estrutura da CNN proposta;
- Random Forest (RF);
- Support Vector Machine (SVM), para obter mais resultados comparativos.

Tabela 5.1: Resultados CNN Referência no conjunto de dados do trabalho

Algoritmo	Acurácia	Sensibilidade	Especificidade
CNN Referência	0,4265	0,7018	0,2278

Depois de obter tais resultados, foram testados também como forma de comparação outras três configurações de algoritmo, descritas na Seção 4.2. Os resultados obtidos foram os seguintes:

Tabela 5.2: Resultados Comparativos no conjunto de dados do trabalho

Algoritmo	Acurácia	Sensibilidade	Especificidade
CNN Referência	0,4265	0,7018	0,2278
CNN Customizada	0,4485	0,7719	0,2152
SVM	0,5294	0,4912	0,5570
RF	0,6397	0,4737	0,7595

5.4 Discussão dos Resultados

Analisando a Tabela 5.2, o algoritmo de melhor acurácia foi a RF, o de melhor sensibilidade foi a CNN Proposta deste trabalho (customizada), e o de melhor especificidade foi também a RF.

Ou seja, com o conjunto de imagens utilizado, as técnicas propostas obtiveram resultados mais eficazes que o método que utiliza a configuração da CNN de referência [1]. Apesar de o resultado de acurácia não ser o mesmo obtido pela técnica de referência, comparativamente é possível verificar que o método proposto tem aplicabilidade para um conjunto de dados como o descrito e utilizado neste trabalho.

Além disso, é uma boa vantagem ter um algoritmo com boa sensibilidade e um algoritmo com boa especificidade, porque ambas as métricas são vantajosas em aplicações biomédicas.

Capítulo 6

Conclusão

No trabalho foi apresentada uma técnica automática para detecção de osteoporose em imagens digitalizadas provenientes de teste de radiografias panorâmicas analógicas em pacientes idosos de comunidades carentes. O resultado não mostra uma acurácia alta, porém, conseguimos uma especificidade e sensibilidade acima de 70% utilizando os dados de imagem deste trabalho.

- Na acurácia, a Random Forest proposta obteve 63,97%, enquanto que a CNN de referência obteve 42,65%;
- Na sensibilidade, a CNN customizada proposta obteve 77,19%, enquanto que a CNN de referência obteve 70,18%;
- Na especificidade, a Random Forest proposta obteve 75,95%, enquanto que a CNN de referência obteve 22,78%.

Combinando esses resultados, conseguimos melhorar o que é obtido pela técnica de referência para o conjunto de dados apresentado.

A base de dados utilizada é bastante desafiadora, mostrado pelo baixo desempenho de uma técnica previamente apresentada na literatura. Porém, representa uma situação real no Brasil e em países em desenvolvimento. Os pacientes da base são principalmente idosos, com baixa dentição, com vários artefatos metálicos em várias imagens, sem contar que as imagens adquiridas são de baixa qualidade e originalmente analógicas. Todos esses são fatores que dificultam a classificação das imagens.

Como trabalhos futuros podemos realizar um junção (*ensemble*) destas e outras técnicas de aprendizagem de máquina para verificar a possibilidade de um aumento na acurácia e nos resultados em geral. É possível também fazer uma validação cruzada mais robusta para testar a configuração proposta com várias configurações do algoritmo. Porém, este já é um primeiro passo para atenuar os problemas de osteoporose no Brasil.

Referências

- [1] Lee JS, Adhikari S, Liu L Jeong HG Kim H Yoon SJ: *Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study*. Dentomaxillofac Radiol, página 48, 2018. <https://doi.org/10.1259/dmfr.20170344>. vi, vii, x, 2, 10, 12, 14, 15, 18, 23, 29, 30, 31, 35, 38, 39, 45, 48, 49
- [2] Qaseem, Amir, Mary Ann Forciea, Robert M. McLean, Thomas D. Denberg e for the Clinical Guidelines Committee of the American College of Physicians: *Treatment of Low Bone Density or Osteoporosis to Prevent Fractures in Men and Women: A Clinical Practice Guideline Update From the American College of Physicians*. Annals of Internal Medicine, 166(11):818–839, junho 2017, ISSN 0003-4819. <https://doi.org/10.7326/M15-1361>. 2
- [3] Lau E, Mithau ADV: *Asian audit / international osteoporosis foundation*, 2009. <http://www.iofbonehealth.org/asian-audit>. 2
- [4] Gomes, C. C., G. L. de Rezende Barbosa, R. P. Bello, F. N. Bóscolo e S. M. de Almeida: *A comparison of the mandibular index on panoramic and cross-sectional images from cbct exams from osteoporosis risk group*. Osteoporosis International, 25(7):1885–1890, Jul 2014, ISSN 1433-2965. <https://doi.org/10.1007/s00198-014-2696-3>. 2
- [5] Taguchi, A, Y Suei, M Ohtsuka, K Otani, K Tanimoto e M Ohtaki: *Usefulness of panoramic radiography in the diagnosis of postmenopausal osteoporosis in women. width and morphology of inferior cortex of the mandible*. Dentomaxillofacial Radiology, 25(5):263–267, 1996. <https://doi.org/10.1259/dmfr.25.5.9161180>, PMID: 9161180. 2
- [6] Taguchi, A., M. Tsuda e M Ohtsuka: *Use of dental panoramic radiographs in identifying younger postmenopausal women with osteoporosis*. Osteoporos Int, 2006. 2
- [7] R, Sindeaux, Figueiredo PT, de Melo NS, Guimarães AT, Lazarte L, Pereira FB, de Paula AP e Leite AF: *Fractal dimension and mandibular cortical width in normal and osteoporotic men and women*. Maturitas, 2014. 3, 19
- [8] Kruse, Christian, Pia Eiken e Peter Vestergaard: *Machine learning principles can improve hip fracture prediction*. Calcified Tissue International, 100, fevereiro 2017. 3, 10, 19, 21

- [9] Zhang, Y.: *The studies and implementation for conversion of image file format*. Em *2015 10th International Conference on Computer Science Education (ICCSE)*, páginas 190–193, 2015. 6
- [10] Suhas, S. e C. R. Venugopal: *Mri image preprocessing and noise removal technique using linear and nonlinear filters*. Em *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEEC-COT)*, páginas 1–4, 2017. 7
- [11] Kumar, D. Ashok, Maroju SaiKumar e P. Samundiswary: *Design and study of modified parallel fir filter using fast fir algorithm and symmetric convolution*. Em *International Conference on Information Communication and Embedded Systems (ICES2014)*, páginas 1–6, 2014. 7
- [12] Areeckal, A. S., M. Kocher e S. D. S.: *Current and emerging diagnostic imaging-based techniques for assessment of osteoporosis and fracture risk*. *IEEE Reviews in Biomedical Engineering*, 12:254–268, 2019. 7, 19, 20, 27, 29
- [13] Patel, S. e M. Goswami: *Comparative analysis of histogram equalization techniques*. Em *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, páginas 167–168, 2014. 8, 9
- [14] Otsu, N.: *A threshold selection method from gray-level histograms*. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. 9, 31, 33
- [15] Freeman, I. C., A. J. Haigler, S. E. Schmeelk, L. R. Ellrodt e T. L. Fields: *What are they researching? examining industry-based doctoral dissertation research through the lens of machine learning*. Em *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, páginas 1338–1340, 2018. 11
- [16] Kim, Dong Wook, Hwiyoung Kim, Woong Nam, Hyung Jun Kim e In Ho Cha: *Machine learning to predict the occurrence of bisphosphonate-related osteonecrosis of the jaw associated with dental extraction: A preliminary report*. *Bone*, 116:207–214, 2018, ISSN 8756-3282. <http://www.sciencedirect.com/science/article/pii/S8756328218301741>. 11
- [17] Gramajo, M., L. Ballejos e M. Ale: *Seizing requirements engineering issues through supervised learning techniques*. *IEEE Latin America Transactions*, 18(07):1164–1184, 2020. 11, 12
- [18] Krizhevsky, Alex, Ilya Sutskever e Geoffrey E. Hinton: *Imagenet classification with deep convolutional neural networks*. *Commun. ACM*, 60(6):84–90, maio 2017, ISSN 0001-0782. <https://doi.org/10.1145/3065386>. 12, 13
- [19] LeCun, Y., Fu Jie Huang e L. Bottou: *Learning methods for generic object recognition with invariance to pose and lighting*. Em *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, páginas II–104 Vol.2, 2004. 12

- [20] Li Fei-Fei, R. Fergus e P. Perona: *Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories*. Em *2004 Conference on Computer Vision and Pattern Recognition Workshop*, páginas 178–178, 2004. 12
- [21] Griffin, Gregory, Alex Holub e Pietro Perona: *Caltech-256 Object Category Dataset*, 2007 (accessed October 7, 2020). <http://authors.library.caltech.edu/7694>. 12
- [22] Krizhevsky, Alex: *Learning multiple layers of features from tiny images*. University of Toronto, maio 2012. 12
- [23] Pinto, Nicolas, David D Cox e James J DiCarlo: *Why is real-world visual object recognition hard?* PLOS Computational Biology, 4(1):1–6, janeiro 2008. <https://doi.org/10.1371/journal.pcbi.0040027>. 12
- [24] Silver, D., Aja Huang, Chris J. Maddison, A. Guez, L. Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, S. Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel e Demis Hassabis: *Mastering the game of go with deep neural networks and tree search*. Nature, 529:484–489, 2016. 12, 13
- [25] Gao, G., C. Dong e K. Niu: *Sparsely connected neural network for massive mimo detection*. Em *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, páginas 397–402, 2018. 13
- [26] Suprijanto, Endang Juliastuti, Yudhi Diputra, Menasita Mayantasari e Azhari: *Dental panoramic image analysis on mandibular bone for osteoporosis early detection*. Em *2013 3rd International Conference on Instrumentation Control and Automation (ICA)*, páginas 138–143, 2013. 16
- [27] Breiman, L: *Random forests*. Machine Learning, 45:5–32, outubro 2001. 16
- [28] Roberts, M. G., J. Graham e H. Devlin: *Image texture in dental panoramic radiographs as a potential biomarker of osteoporosis*. IEEE Transactions on Biomedical Engineering, 60(9):2384–2392, Sep. 2013. 17
- [29] Health, National Institutes: *Osteoporosis prevention, diagnosis, and therapy*. <https://consensus.nih.gov/2000/2000osteoporosis111html.htm>. Accessed: 2019-10-04. 18
- [30] Curtis, Elizabeth M, Rebecca J Moon, Nicholas C Harvey e Cyrus Cooper: *The impact of fragility fracture and approaches to osteoporosis risk assessment worldwide*. Bone, 104:29 – 38, 2017, ISSN 8756-3282. <http://www.sciencedirect.com/science/article/pii/S8756328217300248>, DXA: 30 Years and Counting. 18
- [31] Veena Divya K., A. Jatti, R. Joshi e Deepu Krishna S.: *Characterization of dental pathologies using digital panoramic x-ray images based on texture analysis*. Em *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, páginas 592–595, 2017. 19, 29

- [32] Bhattacharya, S., D. Nair, A. Bhan e A. Goyal: *Computer based automatic detection and classification of osteoporosis in bone radiographs*. Em *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, páginas 1047–1052, 2019. 20
- [33] Riaz, F., R. Nemati, H. Ajmal, A. Hassan, E. Edifor e R. Nawaz: *Osteoporosis classification using texture features*. Em *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, páginas 575–579, 2019. 20
- [34] Singh, Anushikha, Malay Dutta, Rachid Jennane e Eric Lespessailles: *Classification of the trabecular bone structure of osteoporotic patients using machine vision*. *Computers in Biology and Medicine*, 91, outubro 2017. 20
- [35] Chu, Peng, Chunjuan Bo, Xin Liang, Jie Yang, Vasileios Megalooikonomou, Fan Yang, Bingyao Huang, Xinyi Li e Haibin Ling: *Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs*. Volume 2018, páginas 2579–2582, julho 2018. 21, 29
- [36] Cruz, Agnaldo, Hertz Lins, Ricardo Medeiros, José Filho e Sandro Silva: *Artificial intelligence on the identification of risk groups for osteoporosis, a general review*. *BioMedical Engineering OnLine*, 17:12, dezembro 2018. 21
- [37] Ferizi, Uran, Harrison Besser, Pirro Hysi, Joseph Jacobs, Chamith Rajapakse, Cheng Chen, Punam Saha, Stephen Honig e Gregory Chang: *Artificial intelligence applied to osteoporosis: A performance comparison of machine learning algorithms in predicting fragility fractures from mri data: Artificial intelligence applied to osteoporosis*. *Journal of Magnetic Resonance Imaging*, 49, setembro 2018. 22
- [38] Ferizi, Uran, Stephen Honig e Gregory Chang: *Artificial intelligence, osteoporosis and fragility fractures*. *Current Opinion in Rheumatology*, 31:1, abril 2019. 22
- [39] Leite, André, Paulo Figueiredo, Cláudio Guia, Nilce Melo e Ana Patricia De Paula: *Radiografia panorâmica: instrumento auxiliar no diagnóstico da osteoporose*. *Revista Brasileira De Reumatologia*, 48, agosto 2008. 22, 30
- [40] Leite, André Ferreira, Paulo Tadeu de Souza Figueiredo, Cláudio Mares Guia, Nilce Santos Melo e Ana Patrícia de Paula: *Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women*. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 109(3):449 – 456, 2010, ISSN 1079-2104. <http://www.sciencedirect.com/science/article/pii/S1079210409001413>. 22, 30