



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO (PPGEEC)

STEFANE ADNA DOS SANTOS

CLASSIFICAÇÃO DE REGIÕES CÍSTICAS BUCAIS UTILIZANDO REDES NEURAIAS
CONVOLUCIONAIS

SOBRAL

2024

STEFANE ADNA DOS SANTOS

CLASSIFICAÇÃO DE REGIÕES CÍSTICAS BUCAIS UTILIZANDO REDES NEURAIS
CONVOLUCIONAIS

Dissertação apresentada PROGRAMA DE
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉ-
TRICA E DE COMPUTAÇÃO (PPGEEC) da
Universidade Federal do Ceará, Campus de
Sobral, como requisito parcial à obtenção
do título de mestre em Engenharia Elétrica
e de Computação. Área de Concentração:
Engenharia Elétrica e de Computação

Orientador: Prof. Dr. Iális Cavalcante de
Paula Júnior

SOBRAL

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S238c Santos, Stefane Adna dos.
Classificação de regiões císticas bucais utilizando Redes Neurais Convolucionais / Stefane Adna dos Santos. – 2025.
83 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Sobral, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Sobral, 2025.
Orientação: Prof. Dr. Iális Cavalcante de Paula Júnior.
1. Cistos odontogênicos. 2. Redes Neurais Convolucionais. 3. Técnicas de Realce de Imagens. I.
Título.

CDD 621.3

STEFANE ADNA DOS SANTOS

CLASSIFICAÇÃO DE REGIÕES CÍSTICAS BUCAIS UTILIZANDO REDES NEURAIS
CONVOLUCIONAIS

Dissertação apresentada PROGRAMA DE
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉ-
TRICA E DE COMPUTAÇÃO (PPGEEC) da
Universidade Federal do Ceará, Campus de
Sobral, como requisito parcial à obtenção
do título de mestre em Engenharia Elétrica
e de Computação. Área de Concentração:
Engenharia Elétrica e de Computação

Aprovada em: 29 de Outubro de 2024

BANCA EXAMINADORA

Prof. Dr. Iális Cavalcante de Paula
Júnior (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Reuber Regis de Melo
Universidade Federal do Ceará (UFC)

Prof. Dr. Flávio Henrique Duarte de Araújo
Universidade Federal do Ceará (UFC)

A minha mãe por sempre ter me influenciado a
seguir os meus sonhos.

AGRADECIMENTOS

Início a dedicatória deste trabalho agradecendo ao Papai do Céu que sempre cuidou de mim e me deu forças para percorrer toda essa trajetória. Agradeço a minha família, ao apoio a minha mãe Marta Inês, às orações do meu Pai Herneval, às perturbações do meu irmão Stevão, e ao meu cachorro Café por todo apoio e amor que foi me dado ao longo desses anos.

Gostaria de agradecer ao meu professor e orientador Ialis Cavalcante que me orientou durante toda a graduação e mestrado. Sem ele eu não seria a estudante e profissional que eu sou hoje. Agradeço a minha psicóloga Tatiana Medeiros pelo seu excelente trabalho e por ter me ajudado a enfrentar o mestrado e a vida com um olhar mais leve. Quero agradecer também aos meus amigos, Carlos Augusto e Andressa Gomes por todo o companheirismo da minha trajetória de UFC.

Agradeço ao meu amigo, João Pedro Alcântara por todas as xícaras de café que compartilhamos acompanhadas de longas e incríveis conversas. A Carline Tavares por todas as porções de batata frita que compartilhamos e por todo o amor me dado. A Vitoria Santos por todo amor, carinho e companheirismo. A Iris Costa por no início dessa trajetória ter me guiado para o caminho da luz, quando tudo que eu via era a escuridão. Agradeço também a Mauro Furtado, João Paulo Militão, Lara Lima, Jeferson Pontes, Geronimo Aguiar, Vitória Freitas, Matheus Soares e Candido Athaide por todo apoio e carinho.

Por fim, gostaria de agradecer a todos os professores da UFC Campus Sobral que me ajudaram a ser a profissional que eu sou hoje, foram 7 anos de UFC, sendo 5 anos de graduação e 2 anos de mestrado, entrei na UFC uma menina e sai uma mulher e sou muito grata a todos que ajudaram na minha trajetória. Ademais, gostaria de agradecer a mim mesma por toda força que tive, mesmo diante das adversidades enfrentadas consegui chegar até aqui.

Obrigada a todos!!

RESUMO

As patologias dentárias, como lesões maxilares e cistos odontogênicos, apresentam dificuldades de diagnóstico devido à sua localização de difícil visualização. Além disso, o crescimento de regiões císticas pode desencadear um processo de alteração na estrutura óssea de um paciente. Logo, é de suma importância que essas regiões sejam diagnosticadas o mais rápido possível. Sabendo disso, diversos sistemas de inteligência artificial (IA) são utilizados para auxiliar dentistas na detecção destas regiões. Com isso, este trabalho oferece uma análise abrangente das técnicas de processamento de imagens, arquiteturas de redes neurais convolucionais (CNNs), métodos de otimização e taxas de aprendizado para melhorar o diagnóstico de patologias císticas odontológicas. Foram aplicadas as seguintes técnicas de realce de imagem: Filtro Laplaciano, Filtro de Prewitt, Filtro de Roberts e Morfologia Black-hat. Também foram testados os otimizadores Adam, Adagrad e RMSProp em conjunto com as arquiteturas AlexNet, InceptionV3, DCNN e CNNGen. Identificando a taxa de otimização de 0,0001, o otimizador Adam e o filtro Blackhat como técnicas mais eficazes na melhoria da capacidade de classificação das CNNs, destacou-se a rede CNNGen, proposta nesta pesquisa, como especializada na classificação de imagens císticas odontológicas. O modelo desenvolvido com a rede DCNN e o filtro Blackhat alcançou uma acurácia de 88,50% e *f1-score* de 89,01%, destacando a importância da escolha adequada de arquiteturas e parâmetros para o treinamento dos modelos.

Palavras-chave: Cistos odontogênicos; Redes Neurais Convolucionais; Técnicas de Realce de Imagens.

ABSTRACT

Dental pathologies, such as maxillary lesions and odontogenic cysts, pose diagnostic challenges due to their difficult-to-visualize locations. Additionally, the growth of cystic regions can trigger changes in a patient's bone structure. Therefore, it is crucial to diagnose these regions as quickly as possible. In response, various artificial intelligence (AI) systems are employed to assist dentists in detecting these regions. This work provides a comprehensive analysis of image processing techniques, convolutional neural network (CNN) architectures, optimization methods, and learning rates to enhance the diagnosis of dental cystic pathologies. The following image enhancement techniques were applied: Laplacian Filter, Prewitt Filter, Roberts Filter, and Black-hat Morphology. Additionally, the Adam, Adagrad, and RMSProp optimizers were tested alongside the AlexNet, InceptionV3, DCNNDan, and CNNGen architectures. Identifying an optimization rate of 0,0001, the Adam optimizer, and the Black-hat filter as the most effective techniques for improving CNN classification performance, the CNNGen network proposed in this research was highlighted as specialized in classifying dental cystic images. The model developed with the DCNNGen network and the Black-hat filter achieved an accuracy of 88,50% and an f1-score of 89,01%, underscoring the importance of selecting appropriate architectures and parameters for model training.

Palavras-chave: Odontogenic cysts; Convolutional Neural Networks; Image Enhancement Techniques.

LISTA DE FIGURAS

Figura 1 – Exemplo de radiografias intrabuciais.	27
Figura 2 – Exemplo de radiografia panorâmica.	28
Figura 3 – Região com lesão radiolúcida em imagem de radiografia panorâmica.	28
Figura 4 – Fluxo do sistema de Visão Computacional.	29
Figura 5 – Processo de quantização da imagem.	30
Figura 6 – Operação de convolução.	32
Figura 7 – Exemplo da estrutura de uma CNN.	39
Figura 8 – Arquitetura da rede InceptionV3.	40
Figura 9 – Arquitetura da rede AlexNet.	40
Figura 10 – Arquitetura da rede DCNNDAN	41
Figura 11 – Metodologia do Projeto.	47
Figura 12 – Exemplo de radiografias panorâmicas.	49
Figura 13 – Tela inicial do <i>software</i> LabelImg.	50
Figura 14 – Exemplo da aplicação dos filtros utilizados em uma sub-imagem cística.	51
Figura 15 – Exemplo da aplicação das técnicas de aumento de dados.	53
Figura 16 – Arquitetura da CNNGen	56
Figura 17 – Diagrama do K-fold.	56
Figura 18 – Matriz de confusão.	58
Figura 19 – Matrizes de Confusão da CNN treinada com o Adagrad.	61
Figura 20 – Matrizes de Confusão da CNN treinada com o RMSProp.	63
Figura 21 – Matrizes de Confusão da CNN treinada com o Adam.	65
Figura 22 – Matrizes de Confusão da CNN sem aplicação de pré-processamento.	69
Figura 23 – Matrizes de Confusão da CNN com aplicação do Filtro Laplaciano	70
Figura 24 – Matrizes de Confusão da CNN com aplicação do Filtro de Prewitt	71
Figura 25 – Matrizes de Confusão da CNN com aplicação do Filtro de Roberts	73
Figura 26 – Matrizes de Confusão da CNN com aplicação da morfologia BlackHat	74

LISTA DE TABELAS

Tabela 1 – Relação da quantidade de imagens utilizada de cada base.	48
Tabela 2 – Relação da quantidade de sub-imagens por rótulos.	49
Tabela 3 – Acurácias médias para as taxas de otimização do Adagrad.	60
Tabela 4 – Acurácias médias para as taxas de otimização do Adagrad.	61
Tabela 5 – Acurácias médias para as taxas de otimização do RMSProp.	62
Tabela 6 – Acurácias médias para as taxas de otimização do RMSProp.	63
Tabela 7 – Acurácias médias para as taxas de otimização do Adam.	64
Tabela 8 – Acurácias médias para as taxas de otimização do Adam.	65
Tabela 9 – Kfold e CNNs	67
Tabela 10 – Métricas de avaliação das CNNs sem aplicação de pré-processamento. . . .	68
Tabela 11 – Métricas de avaliação das CNNs com aplicação do Filtro Laplaciano. . . .	69
Tabela 12 – Métricas de avaliação das CNNs com aplicação do Filtro de Prewitt. . . .	71
Tabela 13 – Métricas de avaliação das CNNs com aplicação do Filtro de Roberts. . . .	72
Tabela 14 – Métricas de avaliação das CNNs com aplicação da morfologia BlackHat. . .	73

LISTA DE ABREVIATURAS E SIGLAS

TOG	Tumores Odontogênicos
IA	Inteligência Artificial
DCNN	Redes Neurais Convolucionais Profundas
COs	Cistos Odontogênicos
OMS	Organização Mundial da Saúde
A/D	Analógico/Digital
CNN	<i>Convolutional Neural Network</i>
FC	<i>Fully Connected</i>
UFC	Universidade Federal do Ceará
VP	Verdadeiro Positiva
VN	Verdadeiro Negativa
FP	Falso Positivo
FN	Falso Negativo
ACC	Acurácia

LISTA DE SÍMBOLOS

y	Distribuição de rótulos verdadeiros.
s	Distribuição dos valores de saída da camada de ativação.
C	Número de classes de um banco de imagens.
θ	Parâmetros do modelo de rede neural.
θ_i	Parâmetros i do rede neural.
$J(\theta)$	Função objetivo.
∇	Vetor gradiente.
η	Taxa de aprendizado.
$\mathbb{R}^d$	Espaço real com d dimensões.
$\in$	Pertencente.
v_t	Vetor de velocidade e direção para otimizadores com momentum.
γ	Taxa de decaimento do momentum.
$g_{t,i}$	Gradiente de uma função objetivo

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	16
1.2	Objetivos	17
1.3	Organização do trabalho	18
2	TRABALHOS RELACIONADOS	19
2.1	Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network	19
2.2	Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture	19
2.3	Development of an automatic detection model using artificial intelligence for the detection of cyst-like radiolucent lesions of the jaws on panoramic radiographs with small training datasets	20
2.4	Deep Learning Based Detection Tool for Impacted Mandibular Third Molar Teeth	20
2.5	Diffusion-Based Hierarchical Multi-Label Object Detection to Analyze Panoramic Dental X-rays	21
2.6	Machine learning in the detection of dental cyst, tumor, and abscess lesions	21
2.7	Automatic diagnosis of retention pseudocyst in the maxillary sinus on panoramic radiographs using a convolutional neural network algorithm	22
2.8	Analysis of the influence of pre-processing techniques with convolutional neural networks for automatic detection of cysts in wisdom teeth.	23
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	Cisto odontogênicos	24
3.2	Imagens odontológicas	25
3.2.1	<i>Regiões císticas em imagens odontológicas</i>	26
3.3	Sistemas de Visão Computacional	27
3.3.1	<i>Aquisição da Imagem</i>	29
3.3.2	<i>Pré-Processamento</i>	30
3.3.2.1	Filtros de imagens	32
3.3.2.1.1	Filtro Laplaciano	33

3.3.2.1.2	Filtro de Prewitt	34
3.3.2.1.3	Filtro de Roberts	34
3.3.2.2	Morfologia	34
3.3.3	Segmentação	35
3.4	Aprendizado de Máquina	36
3.4.1	Redes Neurais Convolucionais	38
3.4.1.1	InceptionV3	39
3.4.1.2	AlexNet	39
3.4.1.3	DCNNDAN	40
3.4.1.4	Função de perda	41
3.4.1.5	Otimizadores	42
3.4.1.5.1	Adagrad	43
3.4.1.5.2	RMSprop	44
3.4.1.5.3	Adam	44
4	METODOLOGIA	46
4.1	Formação do Banco de Imagens	48
4.2	Extração da Região de Interesse	48
4.3	Aplicação do filtro de Realce	50
4.4	Aumento de dados	52
4.5	Arquiteturas de CNNs utilizadas	52
4.5.1	CNNGen	55
4.6	Validação cruzada	55
4.7	Métricas de avaliação	56
5	RESULTADOS	59
5.1	Resultados da validação cruzada das CNNs variando os otimizadores	59
5.1.1	Adagrad	60
5.1.2	RMSProp	62
5.1.3	Adam	64
5.1.4	Análise dos resultados das funções de otimização	66
5.2	Resultados dos experimentos com CNN	66
5.2.1	Resultado sem aplicação de pré-processamento	67
5.2.2	Resultado com aplicação do Filtro Laplaciano	68

5.2.3	<i>Resultado com aplicação do Filtro de Prewitt</i>	70
5.2.4	<i>Resultado com aplicação do Filtro de Roberts</i>	72
5.2.5	<i>Resultado com aplicação da morfologia BlackHat</i>	72
5.3	Considerações finais dos resultados	74
6	CONCLUSÕES	77
6.1	Trabalhos Futuros	78
	REFERÊNCIAS	79

1 INTRODUÇÃO

Entende-se saúde bucal como o estado da boca, dentes e estruturas orofaciais que habilita as pessoas a efetuarem funções essenciais, como, por exemplo, comer, falar e respirar. Tal conceito varia ao longo da vida, da infância à velhice, e compreende dimensões psicossociais, como o bem-estar individual, autoconfiança e a capacidade de trabalhar e socializar sem desconfortos, dores ou constrangimento. No caso, apesar das doenças orais serem em sua maioria evitáveis, elas são um fardo considerável à saúde mundial porque causam sintomas físicos, limitações funcionais e impactos emocionais.

A radiografia panorâmica é uma das técnicas de diagnóstico mais utilizadas por dentistas, permitindo a visualização de toda a área maxilo-mandibular em uma única imagem. Tais radiografias são úteis para detectar cistos odontológicos e tumores. (SIVASUNDARAM; PANDIAN, 2021a). Viabilizam uma melhor visão das estruturas internas e externas da região bucal, como canais radiculares dos dentes e região da gengiva (LAKHANI *et al.*, 2016). Por exemplo, entre as patologias que podem prejudicar a saúde da boca, têm-se: cistos, pseudocistos, neoplasias malignas, lesões maxilares, que podem ser ocorrências de Tumores Odontogênicos (TOG) e outras patologias ósseas (LOUREDO *et al.*, 2017).

Os cistos odontogênicos são lesões benignas dos maxilares, geradas nas células que originam as estruturas dentárias. Eles são clinicamente definidos como cavidades patológicas, com cápsula, epitélio e conhecidos por possuírem preenchimento líquido ou mole, semelhantes a sacos ou bolsas. Quando um cisto está em processo de crescimento, pode causar uma desordem na estrutura dentária, pressionando outras estruturas anatômicas e acarretando o surgimento de dor ao paciente (MIKULKA *et al.*, 2012). Deste modo, como o crescimento de regiões císticas pode desencadear um processo de alteração na estrutura óssea de um paciente, é de suma importância que essas regiões sejam diagnosticadas o mais rápido possível.

Também, mesmo não sendo o objeto de estudo do trabalho, vale mencionar que a necessidade de ferramentas de diagnóstico rápidas é ainda mais justificada devido às doenças resultantes de um cisto odontogênico poderem gerar um TOG. Tal risco ocorre devido a esses tumores derivarem-se de células de aparatos odontogênicos e seus restos, e como os cistos em questão podem ser recorrentes, aumentando os riscos dessas ocorrências (EL-NAGGAR *et al.*, 2017).

Assim, diversos sistemas de Inteligência Artificial (IA) são utilizados para auxiliar dentistas na detecção de regiões císticas (OLIVEIRA *et al.*, 2019; KWON *et al.*, 2020). Um

exemplo notável é o estudo conduzido por Oliveira *et al.* (2022) que propõe uma metodologia baseada em Redes Neurais Convolucionais Profundas (DCNN) para detectar cistos em imagens de terceiros molares, explorando diferentes técnicas de pré-processamento e comparando com arquiteturas clássicas como VGG16 e Inception-V3.

Para que um algoritmo de inteligência artificial obtenha uma boa confiabilidade, realizando o maior índice de previsões corretas, é necessário que ele seja treinado com dados de qualidade, não contendo muitos ruídos, informações irrelevantes e *outliers*. Com isso, se faz necessário o uso de um tratamento prévio nas imagens que serão utilizadas para o treinamento dos modelos de IA. Este tratamento pode ser realizado a partir da utilização de técnicas de pré-processamento de imagens, que visa realçar as regiões de interesse das imagens e eliminar objetos e ruídos não desejados.

Sendo assim, este estudo visa identificar a metodologia mais eficaz para classificar imagens dessas lesões, buscando aprimorar os métodos de diagnóstico por meio de avanços em inteligência artificial e processamento de imagens. Para atingir esse objetivo, examinamos o impacto dos filtros de realce de imagem na classificação por redes neurais, bem como desenvolvemos uma rede especializada em classificar imagens císticas odontológicas.

1.1 Justificativa

O crescimento de regiões císticas pode desencadear alterações na estrutura óssea de um paciente, pressionando outras estruturas anatômicas e causando dor. Portanto, é crucial que essas regiões sejam diagnosticadas o mais cedo possível. Uma das ferramentas mais comumente utilizadas pelos dentistas para esse diagnóstico é a radiografia panorâmica, que permite a visualização de toda a área maxilo-mandibular em uma única imagem.

Neste contexto, algoritmos de aprendizado de máquina vêm sendo utilizados para auxiliar os dentistas no diagnóstico dessas regiões císticas. O trabalho Ha *et al.* (2023) desenvolve um modelo de aprendizado profundo capaz de classificar automaticamente os pseudocistos de retenção nos seios maxilares em radiografias panorâmicas. Em Kwon *et al.* (2020) é utilizado aprendizado profundo para diagnosticar automaticamente os cistos odontológicos. Por fim, o trabalho Tajima *et al.* (2022) realiza o desenvolvimento de um sistema de IA para detectar lesões radiolúcidas semelhantes a cistos nas mandíbulas em radiografias panorâmicas.

Além disso, é de suma importância que as imagens utilizadas pelos algoritmos de IA sejam de alta qualidade. Para alcançar isso, é necessário aplicar técnicas de processamento de

imagens que destaquem as principais características das radiografias, facilitando a classificação das regiões císticas pelos algoritmos. Alguns trabalhos na literatura utilizam essa metodologia com o objetivo de melhorar os resultados de classificação dos algoritmos. O estudo Kumar *et al.* (2023) examina e compara a precisão de diversas técnicas de análise de textura em radiografias panorâmicas dentárias para identificar lesões de cisto dentário, tumor e abscesso. No trabalho Sivasundaram e Pandian (2021b) é apresentada uma arquitetura modificada da LeNet em uma rede neural convolucional para classificar imagens de cistos orais, acompanhada de um método de segmentação morfológica para identificar regiões de cistos nas imagens classificadas.

Além disso, é crucial desenvolver modelos de classificação com a menor taxa de erro possível, utilizando os parâmetros mais adequados. A análise comparativa da influência de otimizadores no desempenho de Redes Neurais Convolucionais é essencial para criar modelos com alta capacidade preditiva. Em dados médicos, mesmo pequenas diferenças nos resultados dos métodos de otimização são significativas, pois previsões errôneas podem atrasar os diagnósticos dos pacientes.

Ao analisar esses fatores, torna-se plausível realizar experimentos para detectar a melhor metodologia para a classificação das regiões císticas bucais. Uma vez que, um modelo treinado utilizando imagens realçadas, a rede adequada e os parâmetros de otimização corretos terá uma maior taxa de confiabilidade na classificação dessas regiões císticas.

1.2 Objetivos

Este trabalho tem como objetivo geral realizar uma série de experimentos a fim de detectar a melhor metodologia para a classificação de regiões císticas bucais em imagens de radiografias odontológicas. Além disso, os objetivos específicos deste trabalho são:

- Utilizar técnicas de realce de imagens a fim de destacar as principais características das radiografias.
- Realizar uma análise comparativa de diversas taxas de aprendizado para vários métodos de otimização, a fim de selecionar a taxa de aprendizado e o método de otimização mais adequado para ser utilizado no modelo de classificação.
- Analisar o desempenho de várias arquiteturas de CNN a fim de selecionar a melhor rede para classificação de imagens císticas.
- Desenvolver uma rede neural especializada em classificar regiões císticas odontológicas.

- Utilizar métricas da avaliação nos modelos desenvolvidos a fim de analisar os resultados obtidos nos experimentos realizados.
- Identificar a metodologia mais promissora, que inclui a técnica de realce, o método de otimização e a rede neural mais adequada para a classificação das regiões císticas bucais.

1.3 Organização do trabalho

A estrutura deste trabalho está dividida em cinco capítulos. O capítulo 2 contém os trabalhos relacionados. O capítulo 3 explica a base teórica, abordando os aspectos conceituais relacionados às imagens de radiografia e as técnicas empregadas na pesquisa. O capítulo 4 descreve a metodologia proposta, sua implementação e as métricas utilizadas para avaliar os resultados. No capítulo 5, são apresentados e discutidos os resultados das investigações. Finalmente, o capítulo 6 traz as conclusões, considerações finais, contribuições e sugestões para trabalhos futuros.

2 TRABALHOS RELACIONADOS

Nesta seção, são descritos alguns trabalhos relacionados ao projeto. Destaca-se, em particular, o trabalho de Oliveira *et al.* (2023), considerado aqui o mais relevante, pois possui objetivos semelhantes ao deste trabalho.

2.1 Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network

O trabalho Kwon *et al.* (2020) utiliza aprendizado profundo para diagnosticar automaticamente os cistos odontológicos. Para isso, foi desenvolvida uma CNN modificada na YoloV3 e baseada na Darknet-53. A rede YOLOv3 previu caixas delimitadoras de cistos em três escalas diferentes usando recursos de escalas extraídas usando um conceito semelhante para apresentar redes de pirâmide (FPNs), onde a caixa delimitadora tinha um valor de confiança calculado a partir da regressão logística. Após o treinamento, a rede produziu limites de caixas e a confiança de que a caixa delimitadora incluía uma lesão para entrada de imagens panorâmicas. O desempenho geral da classificação para as doenças melhorou de 78,2% de sensibilidade, 93,9% de especificidade, 91,3% de precisão e 0,86 AUC usando a CNN com dados não aumentados para 88,9% de sensibilidade, 97,2% de especificidade, 95,6% de precisão e 0,94 AUC usando o CNN com conjunto de dados aumentado.

2.2 Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture

O trabalho Sivasundaram e Pandian (2021b) apresenta uma arquitetura modificada da LeNet em uma rede neural convolucional para classificar imagens de cistos orais, acompanhada de um método de segmentação morfológica para identificar regiões de cistos nas imagens classificadas. Além disso, são aplicadas funções morfológicas de erosão e dilatação à imagem classificada, gerando versões erodida e dilatada, às quais a lógica de componentes conectados é aplicada para remover objetos menores. A imagem erodida é então subtraída da dilatada para detectar e segmentar padrões anormais nas imagens classificadas. Para aumentar o número de amostras de entrada e avaliar os resultados com precisão, são empregadas uma abordagem tradicional de aumento de dados e um método de validação cruzada em três partes, respectivamente. A metodologia proposta é testada em imagens de cistos obtidas de um hospital

odontológico, resultando em uma taxa de classificação de 99,63% para cistos e uma sensibilidade de aproximadamente 98,3% para a segmentação dos mesmos.

2.3 Development of an automatic detection model using artificial intelligence for the detection of cyst-like radiolucent lesions of the jaws on panoramic radiographs with small training datasets

O trabalho Tajima *et al.* (2022) realiza o desenvolvimento de um sistema de IA para detectar lesões radiolúcidas semelhantes a cistos nas mandíbulas em radiografias panorâmicas com pequenos conjuntos de dados de treinamento. Foram utilizadas 7160 radiografias panorâmicas, mostrando lesões radiolúcidas semelhantes a cistos nas mandíbulas como dados para treinar um algoritmo de aprendizado profundo usando transferência de aprendizado para construir uma DCNN que detectasse automaticamente as lesões. Em seguida, foram usadas 100 radiografias panorâmicas mostrando lesões radiolúcidas semelhantes a cistos como dados de teste e foram definidas quatro áreas-alvo de avaliação: (1) uma área dos dentes anteriores superiores aos pré-molares, (2) uma área do corpo mandibular (incluindo o queixo), (3) uma área do ramo mandibular direito e (4) uma área do ramo mandibular esquerdo. A DCNN foi avaliada em 400 locais-alvo e obteve uma precisão de 98,3%, sensibilidade de 94,4%, especificidade de 99,7%, precisão de 99,0%, recall de 94,4% e F1-Score de 96%.

2.4 Deep Learning Based Detection Tool for Impacted Mandibular Third Molar Teeth

O trabalho proposto em Celik (2022), desenvolve um sistema de detecção assistido por computador, baseado em redes neurais convolucionais profundas, para a detecção de dentes de terceiro molar impactados, utilizando diferentes arquiteturas, e avaliar a utilidade potencial e a precisão das soluções propostas em radiografias panorâmicas. Um total de 440 radiografias panorâmicas de 300 pacientes foram divididas aleatoriamente. Como técnica de duas etapas, foram utilizados Faster-RCNN com ResNet50, AlexNet e VGG16 como backbone e a técnica de uma etapa YOLOv3. O Faster-RCNN, como detector, apresentou a taxa de mAP@0.5 de 0,91 com o backbone ResNet50, enquanto VGG16 e AlexNet apresentaram desempenhos ligeiramente inferiores: 0,87 e 0,86, respectivamente. O outro detector, YOLOv3, proporcionou a maior eficácia de detecção com um mAP@0.5 de 0,96. Recall e precisão foram 0,93 e 0,88, respectivamente, o que apoiou seu alto desempenho. Considerando os achados das diferentes

arquiteturas, observou-se que o detector de uma etapa proposto, YOLOv3, teve excelente desempenho na detecção de dentes de terceiro molar mandibular impactados em radiografias panorâmicas.

2.5 Diffusion-Based Hierarchical Multi-Label Object Detection to Analyze Panoramic Dental X-rays

Embora diversos modelos de ML tenham sido desenvolvidos para interpretar radiografias panorâmicas, ainda não existia um modelo de ponta a ponta capaz de identificar dentes problemáticos com enumeração dentária e diagnósticos simultaneamente. Para criar tal modelo, os autores de Hamamci *et al.* (2023) organizaram três tipos de dados anotados hierarquicamente conforme o sistema FDI: quadrante, quadrante-enumeração e quadrante-enumeração-diagnóstico. Para aprender com essas hierarquias, introduziram uma nova estrutura hierárquica de detecção de objetos multi-rótulo baseada em difusão, que usa o processo de difusão para converter caixas ruidosas em caixas de objetos. Essa abordagem inclui uma técnica inovadora de manipulação de caixas ruidosas, adaptando o processo de remoção de ruído da rede de difusão com a inferência de um modelo previamente treinado. Além disso, utilizaram um método de detecção de objetos multi-rótulo para aprender eficientemente com anotações parciais e fornecer todas as informações necessárias sobre dentes anormais para o planejamento do tratamento. Resultados experimentais mostraram que o método desenvolvido superou significativamente os métodos avançados de detecção de objetos, como RetinaNet, Faster R-CNN, DETR e DiffusionDet, na análise de radiografias panorâmicas, demonstrando grande potencial para conjuntos de dados hierarquicamente e parcialmente anotados.

2.6 Machine learning in the detection of dental cyst, tumor, and abscess lesions

O estudo Kumar *et al.* (2023) teve como objetivo examinar e comparar a precisão de diversas técnicas de análise de textura em radiografias panorâmicas dentárias para identificar lesões de cisto dentário, tumor e abscesso. Foram analisadas 172 radiografias, selecionadas retrospectivamente, que incluíam essas lesões. Radiografias que não atenderam aos critérios técnicos de qualidade diagnóstica foram excluídas. A metodologia do estudo consistiu em cinco etapas. Primeiro, as radiografias foram aprimoradas e a área de interesse foi segmentada manualmente. Em seguida, várias técnicas de extração de características foram aplicadas,

incluindo a Matriz de Co-ocorrência de Níveis de Cinza (GLCM), a Matriz de Comprimento de Corrida de Níveis de Cinza (GLRLM) e a análise de wavelet. As lesões foram então classificadas como cisto, tumor ou abscesso usando um classificador de máquina de vetor de suporte (SVM). Os dados resultantes foram inseridos em uma planilha do Microsoft Excel e analisados estatisticamente com o software SPSS (versão 21). A análise estatística incluiu a computação de estatísticas descritivas e inferenciais, determinando a significância estatística com um valor de $p < 0,05$. Sensibilidade, especificidade e precisão foram utilizadas para comparar o diagnóstico avaliado com o real. Os resultados mostraram que a técnica GLCM alcançou uma precisão de 98% na distinção entre cistos, tumores e abscessos, enquanto a análise de wavelet atingiu 91% e a GLRLM 95%. A alta precisão do GLCM foi destacada pela área sob a curva (AUC), indicando um excelente desempenho na identificação de lesões.

2.7 Automatic diagnosis of retention pseudocyst in the maxillary sinus on panoramic radiographs using a convolutional neural network algorithm

O objetivo do trabalho Ha *et al.* (2023) foi desenvolver um modelo de aprendizado profundo capaz de classificar automaticamente os pseudocistos de retenção nos seios maxilares em radiografias panorâmicas. Foram incluídos no estudo um total de 426 seios maxilares de radiografias panorâmicas de 213 pacientes. Entre esses seios maxilares, 86 apresentavam pseudocistos de retenção, 261 eram saudáveis e 79 possuíam cistos ou tumores. Um modelo EfficientDet, introduzido por Tan, foi desenvolvido para detectar e classificar os seios maxilares. O modelo desenvolvido foi treinado 200 vezes nos conjuntos de dados de treinamento e validação (342 seios) e seu desempenho foi avaliado em termos de precisão, sensibilidade e especificidade no conjunto de testes (21 pseudocistos de retenção, 43 seios saudáveis e 20 cistos ou tumores). A precisão do modelo para classificar pseudocistos de retenção foi de 81%, e o modelo também demonstrou maior precisão para classificar seios saudáveis e cistos ou tumores (98% e 90%, respectivamente). Um dos 21 pseudocistos de retenção no conjunto de testes foi diagnosticado erroneamente como cisto ou tumor. O modelo proposto para classificar automaticamente pseudocistos de retenção nos seios maxilares em radiografias panorâmicas apresentou excelente desempenho diagnóstico.

2.8 Analysis of the influence of pre-processing techniques with convolutional neural networks for automatic detection of cysts in wisdom teeth.

O trabalho apresentado por Oliveira *et al.* (2023) visa apresentar uma metodologia que utiliza DCNNs para identificar a presença de cistos em imagens de terceiros molares, investigando a influência das técnicas de pré-processamento e comparando-as com as arquiteturas clássicas de VGG16 e Inception-V3. A pesquisa consiste em quatro etapas: aquisição, pré-processamento, experimentação com as arquiteturas e análise dos resultados. As imagens utilizadas foram obtidas por meio de uma parceria com o Curso de Odontologia da Universidade Federal do Ceará, campus Sobral, Brasil. Nesses registros, os cistos foram identificados por dois professores especialistas na área. Foram utilizadas 333 imagens de dentes do siso de tamanhos variados, com 163 imagens contendo cistos e 170 saudáveis. Na etapa de pré-processamento, foi inicialmente aplicado um filtro gaussiano às imagens para reduzir possíveis ruídos de alta frequência adquiridos durante a aquisição. Em seguida, quatro técnicas de aumento de contraste foram testadas separadamente: ajustes de brilho e contraste, operações morfológicas, correção gamma e equalização de histograma adaptativo limitado por contraste. Os modelos de DCNN utilizados foram o InceptionV3 e VggNet16, além de duas arquiteturas denominadas CNN1 e CNN2. O uso dessas duas arquiteturas visa investigar a influência no resultado da detecção, realizando uma redução e aumento no número de filtros ao longo das camadas. Estas arquiteturas foram experimentadas com as técnicas de pré-processamento, e os resultados foram comparados com os de VGG16 e Inception-V3 aplicados às imagens originais. A investigação das técnicas de contraste identificou a correção gamma como o melhor método para esta aplicação. Já a análise da influência da quantidade de filtros ao longo das camadas de convolução mostrou que a proposta obteve melhores resultados com a arquitetura que apresentava um aumento no número de filtros. Em conjunto, essas arquiteturas proporcionaram o melhor desempenho, com uma acurácia de 81,99%, precisão de 78,90%, sensibilidade de 86,17% e F1-score de 82,19%.

3 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são detalhadas as descrições das imagens de radiografias odontológicas, incluindo suas características físicas e o processo de identificação dos cistos. Em seguida, são delineadas as etapas de um sistema de visão computacional, destacando as técnicas empregadas. Também são abordadas as técnicas de realce de imagens e os algoritmos de aprendizado de máquina, incluindo seus parâmetros e métodos de construção.

3.1 Cisto odontogênicos

Os Cistos Odontogênicos (COs) são patologias geralmente assintomáticas, exceto quando ocorrem infecções que provocam inflamação, tornando-os sintomáticos. Esses cistos podem causar expansão e erosão da cortical, apresentando um crescimento lento, mas podendo alcançar grandes dimensões. Esse processo de crescimento pode desorganizar a estrutura mandibular, já que o cisto, ao crescer, pode pressionar outras estruturas odontogênicas, resultando em dor para o paciente (MALLYA; LAM, 2018).

Os cistos odontogênicos são cavidades patológicas revestidas por epitélio odontogênico e localizadas nos ossos gnáticos. A relação anatômica desses cistos com a estrutura dentária é crucial para sua detecção, destacando-se no exame radiográfico por sua aparência radiolúcida com limites bem definidos, podendo ser uniloculares ou multiloculares (WHITE; PHAROAH, 2013). Eles são classificados em dois grupos principais conforme sua origem: inflamatórios e de desenvolvimento (NEVILLE *et al.*, 2009).

Os cistos inflamatórios surgem de inflamações que resultam no desenvolvimento de regiões císticas. Originam-se de restos epiteliais de Malassez, epitélio crevicular e revestimento sinusal ou de trajetos fistulosos. As respostas inflamatórias nessas áreas aumentam a produção de queratinócitos, levando à proliferação do epitélio. Esses cistos são classificados como cistos radiculares e cistos colaterais inflamatórios.

Ademais, os cistos odontogênicos inflamatórios são patologias que surgem da proliferação do epitélio em resposta à inflamação. Bactérias ou traumatismos que causam alterações pulpares podem infectar os canais radiculares e originar essas lesões. Isso ocorre principalmente quando o tecido tenta eliminar a infecção dos agentes externos, resultando em reabsorção e perda óssea.

Os cistos de desenvolvimento têm origem desconhecida e estão associados a reações

inflamatórias decorrentes de infecções e traumas maxilares. Essas respostas inflamatórias permitem a proliferação de remanescentes epiteliais, formados pela inclusão de epitélio odontogênico de revestimento nas linhas de fusão dos processos embrionários, provenientes do ectoderma. A Organização Mundial da Saúde (OMS) categoriza os cistos odontogênicos de desenvolvimento em várias subcategorias, incluindo cisto dentígero, cisto de erupção, cisto odontogênico orto ceratinizado, cisto gengival do recém-nascido, cisto gengival do adulto, cisto periodontal lateral e cisto odontogênico glandular.

Além disso, os cistos odontogênicos de desenvolvimento são clinicamente definidos como cavidades patológicas compostas por cápsula e epitélio. Apresentam preenchimento líquido ou mole, semelhantes a sacos ou bolsas. A parede do cisto é semelhante a uma membrana semipermeável, permeável ao líquido e não permeável à direção externa sólida. Essa característica contribui para que seu crescimento seja expansivo. Quando um cisto está em processo de crescimento, pode causar desordem na estrutura dentária, pressionando outras estruturas anatômicas e provocando dor no paciente.

Cistos odontogênicos costumam apresentar sintomas como tumefação, descarga purulenta, parestesia, mobilidade dentária, dor e assimetria facial (VARINAUSKAS *et al.*, 2006). A detecção precoce é crucial para evitar a propagação da infecção, que pode atingir a corrente sanguínea e ameaçar a vida do paciente (SOUSA *et al.*, 2003). Obter este diagnóstico prévio pode reduzir custos e simplificar o tratamento (NURTANIO *et al.*, 2011). As radiografias são essenciais para detectar cistos, pois estes estão localizados em tecidos difíceis de visualizar em exames comuns (LI *et al.*, 2007). Contudo, essa avaliação requer habilidades específicas e pode prevenir procedimentos desnecessários, auxiliando significativamente no diagnóstico.

3.2 Imagens odontológicas

A radiografia foi a precursora das imagens médicas, permitindo a visualização de estruturas internas do corpo em duas dimensões. O raios-X, foi desenvolvido no ano de 1885 pelo físico Wilhelm Röntgen, onde foi gerado a partir de estudos com raios catódicos e tubos de Hittorf-Crookes. Sendo assim, Röntgen aplicou essa energia no corpo humano e capturou a primeira radiografia em uma placa fotográfica (FEJERSKOV *et al.*, 2018).

Os raios-X são produzidos quando elétrons são liberados a partir de um cátodo aquecido e colidem com o ânodo, gerando radiação controlada pela tensão e corrente aplicadas. A descoberta de Röntgen permitiu uma série de avanços médicos, sendo premiada com o Nobel

de Física em 1901. A odontologia logo incorporou essa tecnologia, e a primeira radiografia dentária foi realizada 14 dias após a descoberta, pelo Dr. Frederic Otto Walkhoff (KAPILA *et al.*, 2013).

Ao longo dos anos, apesar dos avanços tecnológicos, a radiografia de filmes foi amplamente utilizada na prática clínica, permanecendo comum na radiografia oral e maxilofacial. A evolução para a radiografia digital gerou uma série de vantagens como a redução da dose de radiação ao paciente e a eliminação do processo químico (LIMA *et al.*, 2018). A radiografia digital faz uso de sensores de raios-X responsáveis por converter a radiação em sinais elétricos digitais (GONZALEZ; WOODS, 2010). As imagens de raios-X dentais em duas dimensões são chamadas de radiografias planares, classificadas em extrabucais e intrabucais.

As intrabucais são divididas em periapical, interproximal e oclusal. As radiografias periapicais mostram todo o dente e o osso alveolar ao redor. As radiografias interproximais focam nas coroas dos dentes e na crista alveolar adjacente. Já as radiografias oclusais capturam uma área maior dos dentes e do osso alveolar do que as periapicais (ALVARES *et al.*, 1965). A Figura 1 exibe as radiografias intrabucais, onde o equipamento é colocado na parte interna da boca do paciente.

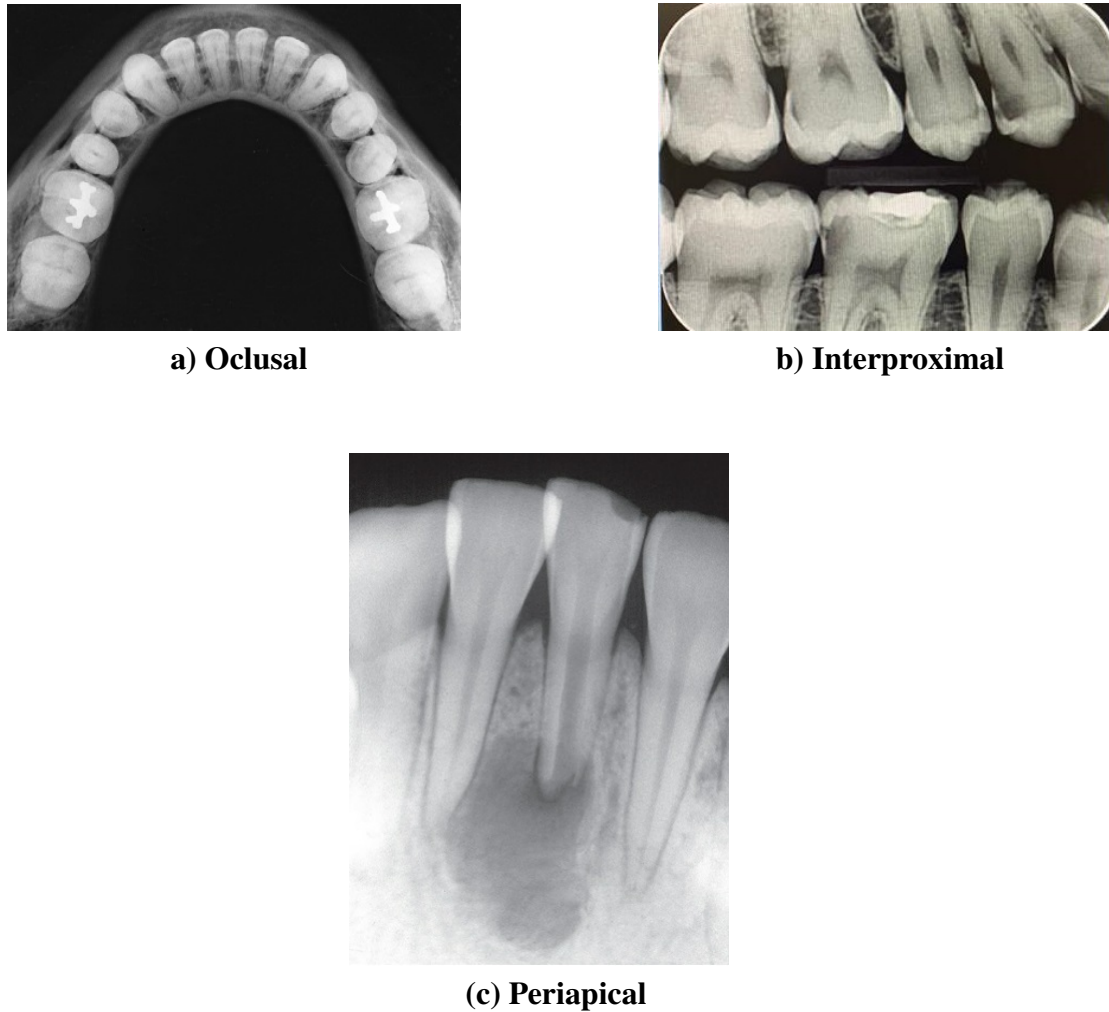
Além disso, entre os tipos de radiografias extrabucais estão: Telerradiografia Lateral, Radiografia Panorâmica, Radiografia Posteroanterior, Radiografia de Articulação Temporomandibular e Tomografia Computadorizada de Feixe Cônico. Destas citadas, a mais comum é a panorâmica. A Figura 2 exibe uma radiografia panorâmica obtida mediante uma parceria com o curso de Odontologia da Universidade Federal do Ceará no Campus de Sobral.

3.2.1 Regiões císticas em imagens odontológicas

A maioria dos cistos odontogênicos são gerados a partir de restos epiteliais da formação dos dentes, ocorrendo frequentemente nos ossos maxilares, principalmente na região dentada acima do canal do nervo alveolar inferior, podendo ser mais raro no côndilo e no processo coronóide (MALLYA; LAM, 2018). Sendo assim, a radiografia panorâmica é o exame mais comum para a detecção dessas regiões, pois permite visualizar toda a região mandibular, identificando cistos em estágios iniciais, antes de causarem transtornos maiores.

Ademais, nas radiografias, os cistos são apresentados como lesões radiolúcidas, normalmente são em formato oval ou redondo, preenchidos por fluido (COSTA *et al.*, 2016). Cistos de longa duração podem exibir calcificação distrófica, podendo resultar em uma imagem

Figura 1 – Exemplo de radiografias intrabucais.



Fonte: Figura 1a é adaptada de White e Pharoah (2013). A Figura 1b foi obtida em parceria com o curso de Odontologia da UFC. Figura 1c é adaptada de Lopes e Junior (1999).

interna mais esparsa. Eles tem crescimento lento, podendo causar deslocamento de estruturas, reabsorção de dentes e expansão da mandíbula, desenvolvendo um limite cortical fino nas áreas atingidas (SANTOS *et al.*, 2007). A Figura 3 exibe um exemplo de região com lesões radiolúcidas em imagens de radiografias panorâmicas.

3.3 Sistemas de Visão Computacional

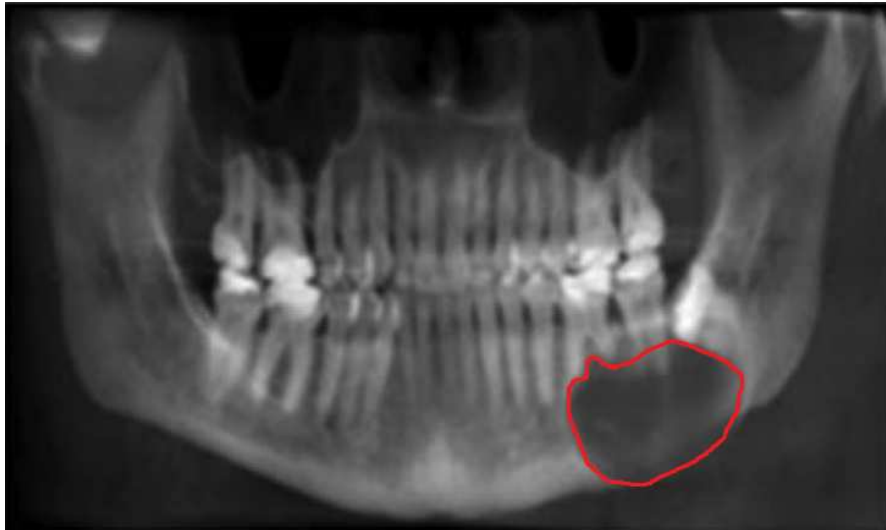
A Visão Computacional é a ciência que se dedica a programação de computadores para enxergar e extrair informações de objetos no ambiente, simplificando processos complexos com a ajuda de bibliotecas e linguagens de programação (QUEIROZ; GOMES, 2006). Definida por Ballard e Brown em 1982, essa ciência estuda tecnologias que permitem a máquinas verem e

Figura 2 – Exemplo de radiografia panorâmica.



Fonte: Obtida em parceria com o curso de Odontologia UFC-Campus Sobral.

Figura 3 – Região com lesão radiolúcida em imagem de radiografia panorâmica.



Fonte: Adaptado de Sivasundaram e Pandian (2021b).

processarem imagens de sensores diversos (BALLARD; BROWN, 1982).

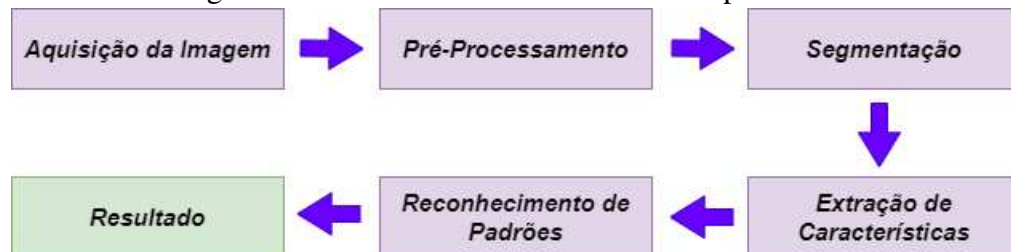
A Visão Computacional tem buscado imitar a visão humana, mas enfrenta desafios devido à complexidade da interpretação visual pelo cérebro. Hoje, é uma área interdisciplinar, integrando neurobiologia, inteligência artificial, processamento de imagens e reconhecimento de padrões (NEVES *et al.*, 2012). Aplicações práticas incluem monitoramento por imagem (SALTORATTO *et al.*, 2013; RIOS; NETTO, 2016; TETILA *et al.*, 2023), robótica industrial (CARRARA, 2015), veículos autônomos (AVILA *et al.*, 2023) e diagnósticos médicos (SILVA *et al.*, 2019; VAZ, 2023; JUNIOR, 2023), onde sistemas baseados em Visão Computacional podem detectar anomalias precocemente, melhorando a qualidade de vida dos pacientes.

A Figura 4 se baseia no trabalho Barelli (2018) que apresenta o fluxo base de um

sistema de Visão Computacional. Esse sistema começa com a aquisição de imagem, utilizando sensores para obter e digitalizar imagens em 2D, 3D ou vídeos. A segunda etapa é o pré-processamento, onde técnicas são aplicadas para destacar bordas e formas e reduzir ruídos, facilitando a identificação de regiões ou objetos de interesse.

Na etapa de segmentação, os objetos de interesse são isolados da imagem original. Em seguida, características desses objetos, como área e diâmetro, são extraídas. Por fim, ocorre o processamento de alto nível, onde os objetos segmentados são reconhecidos e classificados, validando os resultados e garantindo sua precisão. Além disso, algoritmos de aprendizado de máquina estão sendo amplamente utilizados na etapa de extração de características de reconhecimento de padrões para a classificação de imagens.

Figura 4 – Fluxo do sistema de Visão Computacional.



Fonte: Elaborado pela autora. Baseado em Barelli (2018).

3.3.1 Aquisição da Imagem

Na etapa de aquisição de imagens são utilizados sensores que convertem a luz capturada em dados digitais. As câmeras analógicas costumavam utilizar filmes químicos sensíveis à luz, enquanto as câmeras digitais atuais utilizam sensores eletrônicos, principalmente CCD e CMOS.

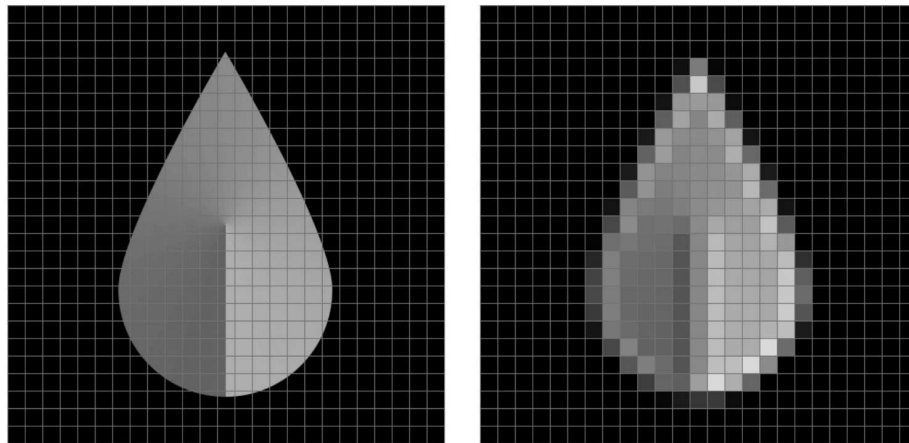
Os sensores CCD utilizam uma matriz de capacitores fotossensíveis que capturam a luz e a transformam em imagem digital, que é então convertida para um sinal digital por um conversor Analógico/Digital (A/D). Os sensores CMOS realizam essa conversão diretamente em cada célula, sem a necessidade de um conversor A/D. Ambos os tipos de sensores utilizam filtros coloridos para representar imagens em cores. Os CMOS são geralmente mais baratos, mas podem produzir mais ruído e serem menos sensíveis à luz (GONZALEZ; WOODS, 2010).

Além disso, outro ponto importante sobre o processo de aquisição está na resolução da imagem que vai ser obtida. A resolução refere-se a quantidade de detalhes que aquela imagem irá possuir (FILHO; NETO, 1999). Nas imagens digitais, a resolução pode ser definida pelo

processo de amostragem e quantização do sinal analógico, reduzindo os pontos infinitos da imagem analógica para um número finito de pixels (QUEIROZ; GOMES, 2006). A amostragem determina quantos detalhes são representados, enquanto a quantização define a cor de cada pixel com base na intensidade da luz incidente em cada célula do sensor (BALAN; INSTRUCIONAL, 2009).

A Figura 5 ilustra o processo de quantização. A imagem à esquerda mostra o sinal amostrado em um número finito de pontos, enquanto a imagem à direita exibe o resultado após o processo de quantização. A amostragem e a quantização do sinal determinam a quantidade de pixels e cores necessárias para representar a imagem em um arquivo digital, definindo assim suas resoluções.

Figura 5 – Processo de quantização da imagem.



Fonte: Barelli (2018).

A resolução das imagens digitais pode ser especificada por duas grandezas principais: a resolução espacial que relaciona a quantidade de pixels em linhas e colunas, e a resolução de bit que é a quantidade de cores ou tons de cinza que um pixel pode representar (BARELLI, 2018). Por exemplo, uma imagem de 128x64 pixels tem uma resolução espacial de 128 por 64. A resolução de bit, ou intensidade, determina quantos valores de cor um pixel pode ter, variando de 1 bit para imagens binárias a 24 bits para imagens coloridas.

3.3.2 *Pré-Processamento*

O pré-processamento digital de imagens é utilizado para aprimorar determinados aspectos estruturais de uma imagem, com o intuito de prepará-la para a realização de outras técnicas como o aprendizado de máquina ou o reconhecimento de padrões. Além disso, ele pode

ser composto por várias técnicas que devem ser utilizadas de acordo com as necessidades de cada projeto.

Diversos procedimentos podem ser realizados na etapa de pré-processamento de imagens, incluindo operações aritméticas e geométricas, ajuste de contraste e tratamento de ruído. Transformações geométricas como inversão horizontal ou vertical, rotação, ajuste de escala e correção de perspectiva são comuns (GONZALES; WINTZ, 1987). A translação desloca a imagem, a rotação envolve girar a imagem em um ângulo predeterminado, e o ajuste de escala altera suas dimensões proporcionalmente sem distorção. Corrigir distorções de perspectiva é crucial, especialmente em fotografias com linhas horizontais, verticais ou formas geométricas.

As operações aritméticas em imagens permitem adicionar o conteúdo de uma imagem em outra ou combiná-los, contanto que ambas as imagens tenham a mesma dimensão e tipo de píxel (JÄHNE, 2005). A adição de imagens envolve somar os valores dos píxeis de uma imagem com outra, podendo criar uma nova imagem ou ajustar o contraste ao adicionar valores positivos, ou negativos aos píxeis. Além disso, operações como subtração, multiplicação e divisão podem ser aplicadas para diversos propósitos.

A divisão entre imagens pode ajustar o contraste ou verificar sua similaridade, sendo que imagens idênticas resultam em valores de píxel igual a 1 após sofrer a operação de divisão. A subtração destaca as diferenças entre duas imagens, enquanto a multiplicação permite combinar imagens com pesos diferentes, criando desta forma efeitos de mistura ou transparência. No entanto, ajustes de perspectiva e escala, juntamente com operações aritméticas e equalização de histogramas, são técnicas essenciais para realçar objetos de interesse em imagens, mas podem acabar intensificando ruídos nelas.

Os ruídos são variações aleatórias e indesejáveis nos valores dos píxeis que podem obscurecer ou distorcer a qualidade da imagem (SAXENA; KOURAV, 2014). Eles podem ser gerados por fatores como poeira, vibração da câmera, distorção da lente, iluminação inadequada e ruído elétrico no sensor durante a captura e processamento das imagens. Diferentes tipos de ruídos incluem o de amostragem, de processamento, de codificação de imagem, de oclusão de cena, sal e pimenta, e gaussiano.

Ademais, algumas técnicas de filtragem também podem ser utilizadas com o intuito de realçar os objetos de interesse de uma imagem, como também de minimizar ou remover ruídos das imagens. Sendo assim, os tópicos seguintes irão descrever algumas técnicas de filtragem que podem ser utilizadas na etapa de pré-processamento de imagens.

3.3.2.1 Filtros de imagens

Os métodos de filtragem fazem parte do conjunto de técnicas disponíveis no processamento digital de imagens. Em geral, esses métodos podem ser divididos em duas categorias, sendo as técnicas de filtragem espacial e as técnicas de filtragem no domínio da frequência. Os métodos que trabalham no domínio espacial operam diretamente sobre a matriz de pixels que é a imagem digitalizada, normalmente utilizando operações de convolução com máscaras. Os métodos que atuam no domínio da frequência se baseiam na modificação da transformada de Fourier da imagem.

Os filtros são matrizes que percorrem uma imagem, alterando neste processo os valores dos píxeis para corrigir, suavizar ou realçar regiões específicas (ABDURRAZZAQ *et al.*, 2020). Essas matrizes podem ser chamadas de máscaras ou núcleos, atuam modificando os valores de cada píxel da imagem. O procedimento de deslocamento da máscara sobre a imagem para realizar essas operações é chamado de convolução.

Na operação de convolução, a máscara sobrepõe um píxel-alvo da imagem, realizando um produto ponto a ponto entre os elementos da região sobreposta e os elementos da máscara correspondente (KIM; CASPER, 2013). A Figura 6 exemplifica a operação de convolução. A primeira matriz representa uma imagem de entrada em tons de cinza, a segunda matriz é a máscara (filtro) aplicada à imagem de entrada, e a terceira matriz mostra o resultado da convolução, onde o píxel-alvo na linha 3 e coluna 3 teve seu valor alterado de 11 para 2. O píxel-alvo é o píxel que está sendo manipulado durante o processo de filtragem, e os píxeis ao redor são chamados de píxeis de vizinhança.

Figura 6 – Operação de convolução.

10	14	5	7	12	4
8	13	3	1	8	5
12	9	11	14	2	16
4	5	16	9	11	15
6	18	7	13	12	10
11	3	10	9	18	5

0	-1	0
-1	4	-1
0	-1	0

Fonte: Barelli (2018).

Além disso, os filtros podem ser lineares ou não lineares (GUPTA *et al.*, 2011). Os filtros lineares utilizam máscaras para realizar somas ponderadas das intensidades dos píxeis, suavizando, realçando detalhes e minimizando ruído sem alterar o nível médio de cinza. Já os filtros não lineares atuam realizando somas não ponderadas e são utilizados para destacar bordas, linhas e manchas, podendo alterar o nível médio de cinza da imagem.

Os filtros também podem ser classificados como passa-baixas, passa-altas ou passa-faixas, conforme o tipo de frequência filtrada (MAKANDAR; HALALLI, 2015). Os filtros denominados como passa-baixas atenuam ou removem componentes da imagem de alta frequência no domínio das transformadas de Fourier. Com isso, eles possuem um efeito de suavização, pois atua removendo ou diminuindo as regiões de bordas da imagem, provocando um efeito de borramento.

Em contrapartida, os filtros passa-altas atenuam ou removem as componentes de baixa frequência, realizando o processo de realce das bordas e aumentando o contraste da imagem. Outrossim, os filtros passa-faixa possuem um funcionamento um pouco diferenciado em comparação com os outros filtros, uma vez que ele atenua e remove os componentes acima de sua frequência de corte superior e abaixo de sua frequência de corte inferior.

3.3.2.1.1 Filtro Laplaciano

O operador laplaciano é amplamente usado para realçar bordas em imagens (BHAI-RANNAWAR, 2018). Ele utiliza uma máscara de ordem 3 que percorre a imagem, alterando o pixel-alvo pela média ponderada dos pixels vizinhos e elevando ao quadrado o valor obtido. No entanto, é muito sensível a ruídos. Para resolver esse problema, um filtro passa-baixas, como o gaussiano, pode ser aplicado antes do filtro laplaciano. Essa combinação é tão eficaz que levou à criação do filtro laplaciano de gaussiano, que unifica essas operações em uma única máscara, reduzindo o número de operações.

Em resumo, o operador laplaciano é uma derivada de segunda ordem de uma função bidimensional, realça bordas e atenua variações suaves, mas pode aumentar os ruídos na imagem. O laplaciano de uma função de duas variáveis $f(x, y)$ é definido como:

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (3.1)$$

Onde:

- $\frac{\partial^2 f}{\partial x^2}$ é a segunda derivada parcial de f em relação a x .
- $\frac{\partial^2 f}{\partial y^2}$ é a segunda derivada parcial de f em relação a y .

3.3.2.1.2 Filtro de Prewitt

As máscaras de detecção de borda do operador Prewitt são um dos métodos mais antigos e mais bem compreendidos de detecção de bordas em imagens. Sendo um operador de diferença, calcular uma aproximação do gradiente da imagem de intensidade de função (PREWITT *et al.*, 1970). Esse filtro pode detectar bordas horizontalmente e verticalmente.

As matrizes do filtro de Prewitt são definidas como:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (3.2)$$

Onde G_x é a matriz para detectar bordas na direção horizontal e G_y é a matriz para detectar bordas na direção vertical.

3.3.2.1.3 Filtro de Roberts

O filtro de Roberts é outro operador comumente usado para detecção de bordas em imagens. Ele é bastante simples e consiste em duas matrizes de convolução, sendo uma para a detecção de bordas na diagonal e outra para a detecção de bordas na direção oposta (ROBERTS, 1963). As matrizes do filtro de Roberts são definidas como:

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad G_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (3.3)$$

Onde G_x é a matriz para detectar bordas na direção horizontal e G_y é a matriz para detectar bordas na direção vertical.

3.3.2.2 Morfologia

As operações morfológicas são utilizadas com o intuito de alterar o formato ou a estrutura dos objetos em uma imagem (GONZALEZ; WOODS, 2010). Elas costumam serem utilizadas em imagens binárias para realçar objetos de interesse ou tratar ruídos da binarização.

Essas operações utilizam um elemento estruturante, que percorre toda a imagem aplicando uma técnica específica de morfologia.

Alguns elementos estruturantes comuns possuem formatos de elipse, quadrado, cruz e reta. Elementos estruturantes elípticos são melhores para objetos circulares, enquanto elementos retangulares são ideais para objetos com formato retangular ou mais retas que curvas.

A operação de erosão é caracterizada pela corrosão das bordas dos objetos na imagem, resultando em uma redução do tamanho dos objetos, ou seja, objetos "corroídos". Em contrapartida, a operação de dilatação expande os objetos de interesse, aumentando a área dos mesmos.

A operação de abertura consiste na aplicação da erosão seguida pela dilatação. Essa técnica é frequentemente utilizada para o tratamento de ruídos, pois a erosão remove muitos ruídos da imagem, mas pode também eliminar algumas bordas desejadas. A dilatação subsequente visa reconstruir as bordas perdidas durante a erosão. Por outro lado, a operação de fechamento ocorre de forma inversa à abertura, começando com a dilatação seguida pela erosão.

Atualmente a literatura abrange uma ampla variedade de operações morfológicas, incluindo o Black-Hat, que se destaca como uma técnica definida pela diferença entre o fechamento da imagem de entrada utilizando um elemento estruturante e a própria imagem original.

3.3.3 Segmentação

A segmentação de objetos em imagem é uma etapa fundamental em sistemas de Visão Computacional, que consiste em isolar a área que representa o objeto de interesse, excluindo o segundo plano (KAUR; KAUR, 2014). O objeto de interesse é considerado o primeiro plano da imagem, enquanto os píxeis restantes são considerados o segundo plano da imagem e descartados. A segmentação é crucial para a extração de características do objeto, permitindo a obtenção de informações e detalhes necessários para sua classificação.

A segmentação por binarização, também conhecida como aplicação de limiar de intensidade, define um valor de limiar para separar o objeto de interesse do segundo plano. Píxeis com valores maiores que o limiar representam o objeto de interesse e são redefinidos para preto ou branco, enquanto píxeis com valores menores são estabelecidos como segundo plano e recebem a cor oposta. Esse método resulta em uma imagem binária, geralmente com o objeto em branco e o fundo em preto.

A segmentação por cor utiliza o espaço de cor HSV para separar objetos com uma

cor específica. A segmentação por movimento detecta objetos que se movem entre capturas de uma câmera fixa. A segmentação por borda aplica filtros detectores de bordas, enquanto a segmentação por localização utiliza coordenadas predefinidas para isolar o objeto de interesse.

3.4 Aprendizado de Máquina

Aprendizado de máquina (em inglês, Machine Learning) é um campo dedicado ao estudo e desenvolvimento de algoritmos de computador que têm a habilidade de aprender a partir de dados de treinamento (GÉRON, 2019). Na década de 1950, Alan Turing, reconhecido como o pai da computação, concebeu o teste de Turing, considerado um dos primeiros marcos do aprendizado de máquina. Este teste avalia a capacidade das máquinas de desenvolver um raciocínio lógico semelhante ao dos seres humanos.

Ademais, o progresso das pesquisas nessa área teve um crescimento gradual ao longo das décadas seguintes. Somente na década de 1980, novos estudos e pesquisas começaram a ganhar destaque novamente, como o livro *Machine Learning: An Artificial Intelligence Approach*, que apresenta uma série de trabalhos e pesquisas promissoras sobre aprendizado de máquina (MICHALSKI *et al.*, 1983).

Os algoritmos de aprendizado de máquina podem ser classificados de acordo com suas características de construção e treinamento (GÉRON, 2019). Além disso, alguns algoritmos precisam ser treinados apenas uma vez, enquanto outros necessitam de um treinamento incremental, o que é definido pelo aprendizado em lote ou online.

Outrossim, eles também podem ser treinados com ou sem supervisão humana, o que determina se o aprendizado é supervisionado, não supervisionado ou semi-supervisionado. Por fim, é essencial determinar como será o processo de aprendizado do algoritmo: se ele irá prever os dados identificando suas semelhanças ou através de um modelo de padrões dos dados, o que pode ser definido pelo aprendizado baseado em instância ou em modelo.

No aprendizado em lote, também chamado de aprendizado offline, o algoritmo é treinado apenas uma vez. Se for necessário treinar o algoritmo com novos dados, ele deve ser retreinado do zero. Já no aprendizado online, o algoritmo é treinado utilizando minilotes de dados e pode ser treinado quantas vezes for necessário, permitindo a adição de novos minilotes de dados em cada novo treinamento.

O aprendizado supervisionado utiliza dados rotulados para treinamento, onde o rótulo é criado por um ser humano e se refere ao significado do dado. No aprendizado não

supervisionado, esses dados não são rotulados e sua classificação é realizada através de padrões analisados nos dados de treinamento. Já o aprendizado semi-supervisionado utiliza tanto dados rotulados quanto não rotulados para o treinamento dos algoritmos.

Os algoritmos de aprendizado de máquina precisam aprender com os dados de treinamento e ter boas previsões em dados desconhecidos. No aprendizado baseado em instância, o algoritmo memoriza os dados de treinamento e realiza a previsão de novos dados por meio de uma medida de similaridade com os dados de treinamento. No aprendizado baseado em modelo, é utilizado um modelo para realizar a previsão dos dados, sendo que esse modelo pode ser criado usando funções matemáticas para a detecção de padrões nos dados de treinamento.

Algumas características dos dados podem influenciar a qualidade do treinamento dos algoritmos de *machine learning*. Por exemplo, é essencial haver uma quantidade adequada de dados para treinamento e que esses dados sejam de qualidade, livres de informações duvidosas ou irrelevantes, outliers, ruídos e erros.

Outro fator crucial é a qualidade do algoritmo desenvolvido. Alguns algoritmos podem sofrer de sobreajuste (*overfitting*, em inglês), que ocorre quando o algoritmo não generaliza bem e, conseqüentemente, não obtém bons resultados na previsão de dados desconhecidos. Outro problema é o subajuste (*underfitting*, em inglês), que acontece quando o algoritmo é simplório demais para o problema a ser solucionado, resultando em uma incapacidade de aprender com os dados de treinamento.

Entre os algoritmos de aprendizado de máquina, se destacam aqueles com arquiteturas baseadas em Redes Neurais Convolucionais (*Convolutional Neural Network* (CNN)). Uma CNN é composta por múltiplas funções e camadas de convolução, além de apresentar uma grande quantidade de parâmetros e diversas arquiteturas.

O otimizador é um dos parâmetros utilizados por uma CNN durante o processo de treinamento dos modelos. Ele é responsável pela atualização dos pesos da rede e por minimizar a função de custo. Seu objetivo é reduzir os erros da rede neural para melhorar seu desempenho preditivo.

Nas sub-seções seguintes, será apresentado um resumo sobre a arquitetura de uma CNN e as funções otimizadoras empregadas neste trabalho. Ademais, um maior aprofundamento sobre o funcionamento dos algoritmos de otimização pode ser encontrado em Ruder (2016).

3.4.1 Redes Neurais Convolucionais

As Redes Neurais são algoritmos de aprendizado de máquina que possuem estrutura é inspirada na conectividade dos neurônios do cérebro humano. Sua configuração padrão é composta por três camadas principais: a camada de entrada, algumas camadas ocultas e a camada de saída, que contém a classificação dos dados de entrada (GÉRON, 2022). Elas recebem os dados de entrada de forma vetorizada e unidimensional. O problema disso é que, em imagens muito grandes, esse vetor seria gigantesco, o que poderia tornar o processo de treinamento da rede extremamente pesado.

Para enfrentar o problema da dimensionalidade dos dados de entrada das redes neurais, foram desenvolvidas as CNNs, mencionadas pela primeira vez em um artigo de Kunihiko Fukushima, onde foi definido um modelo de rede neural como mecanismo de reconhecimento de padrões visuais (FUKUSHIMA; MIYAKE, 1982). Desde então, essa área de aprendizado de máquina tem crescido gradativamente, sendo amplamente utilizada para o reconhecimento e classificação de imagens.

A CNN atribui pesos e vieses de importância a determinados objetos e características correspondentes da imagem, possuindo, assim, a capacidade de realizar a classificação e reconhecimento de padrões em imagens (FERLITSCH, 2021). Sua estrutura é composta por camadas de convolução, ativação, pooling e camadas densas, também conhecidas como camadas totalmente conectadas (*Fully Connected* (FC), do inglês).

As camadas de convolução são utilizadas para extrair um mapa contendo as principais características da imagem (CHOLLET, 2021). Além disso, um mapa de ativação, também chamado de mapa de recursos, é gerado após a aplicação repetida desses filtros convolucionais aos parâmetros de entrada da rede.

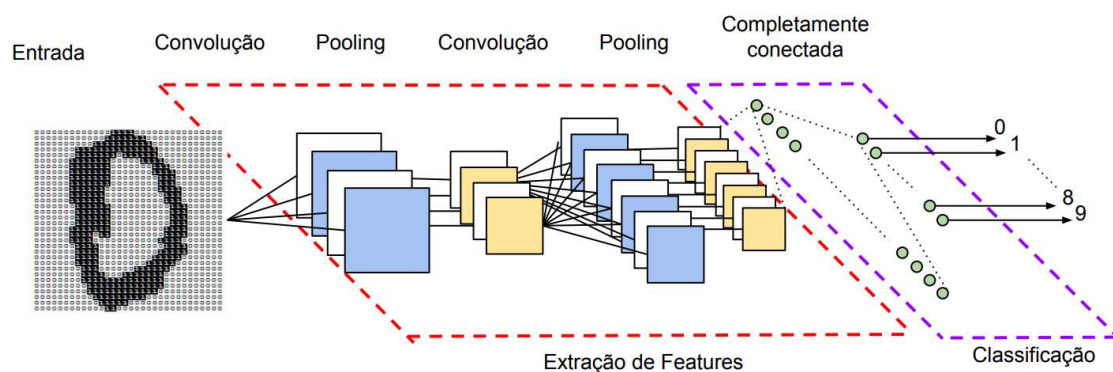
As camadas de *pooling* realizam a sumarização do mapa de características da imagem (SUN *et al.*, 2017), com o objetivo de reduzir a quantidade de pesos para o treinamento e evitar o sobreajuste (*overfitting*). Além disso, algumas arquiteturas de CNN utilizam camadas de *Dropout*, responsáveis por regularizar a rede, desligando alguns neurônios durante o processo de treinamento para evitar o *overfitting*.

As camadas de ativação são utilizadas para evitar linearidades no treinamento. Elas são compostas por funções de ativação, que podem ser de diversos tipos (STEVENS *et al.*, 2020). A função de ativação ReLU (*Rectified Linear Unit*) retorna 0 para valores de entrada negativos

e o próprio valor de entrada para valores de entrada positivos. Essa função possui uma boa convergência, porém pode gerar problemas se os neurônios só receberem valores negativos, já que retorna 0 para todos esses valores. Outra função de ativação é a Softmax, que produz uma distribuição de probabilidades para cada classe da imagem durante o processo de classificação, sendo principalmente usada como saída da camada totalmente conectada.

Além disso, antes da camada totalmente conectada, existe a camada *Flatten*, responsável por transformar a matriz resultante de saída em um vetor unidimensional. Por último, as camadas totalmente conectadas são utilizadas para compilar os recursos extraídos e gerar uma saída que dependerá do número de classes do problema (GÉRON, 2022).

Figura 7 – Exemplo da estrutura de uma CNN.



Fonte: (VARGAS *et al.*, 2016)

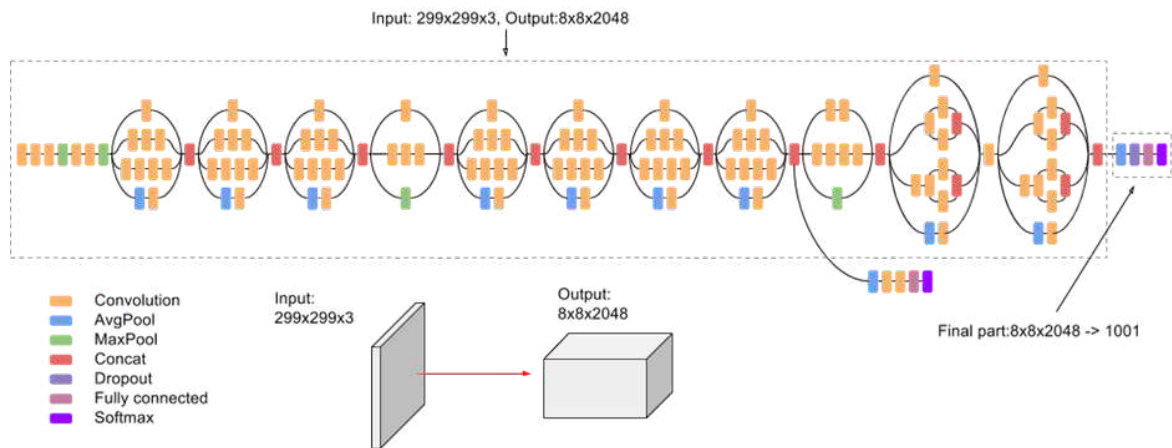
3.4.1.1 InceptionV3

A arquitetura *Inception V3* (SZEGEDY *et al.*, 2016), também conhecida como GoogleNet, é a terceira versão da rede neural convolucional produzida pela Google. Ela possui 48 camadas de profundidade, contendo componentes assimétricos e simétricos, com camadas de convoluções, agrupamentos médios e máximos (*average pooling*, *max pooling*), concatenações (*concats*), desistências (*dropout*) normalização em lote, e camadas totalmente conectadas. A Figura 8 apresenta a estrutura da arquitetura da rede InceptionV3.

3.4.1.2 AlexNet

Outra arquitetura de rede neural muito conhecida é a AlexNet (ALOM *et al.*, 2018). Ela possui 8 camadas, sendo 5 camadas convolucionais, com filtros de tamanhos

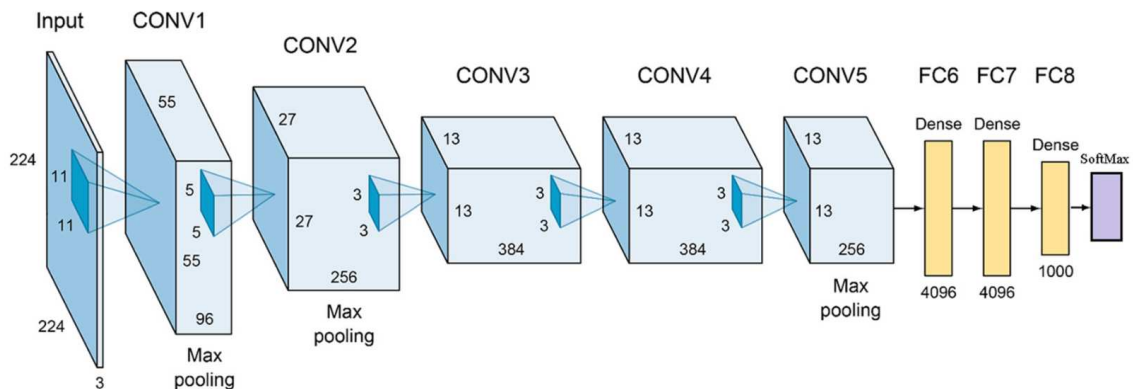
Figura 8 – Arquitetura da rede InceptionV3.



Fonte: (DUONG-TRUNG *et al.*, 2019)

11x11, 5x5, 3x3, 3x3, 3x3 respectivamente, e função de ativação Relu, seguidas por 2 camadas totalmente conectadas com 4096 neurônios ocultos e uma camada densa com 2 neurônios ocultos e função de ativação Softmax. A Figura 10 apresenta a estrutura da arquitetura da rede AlexNet.

Figura 9 – Arquitetura da rede AlexNet.



Fonte: (KARSAZ, 2022)

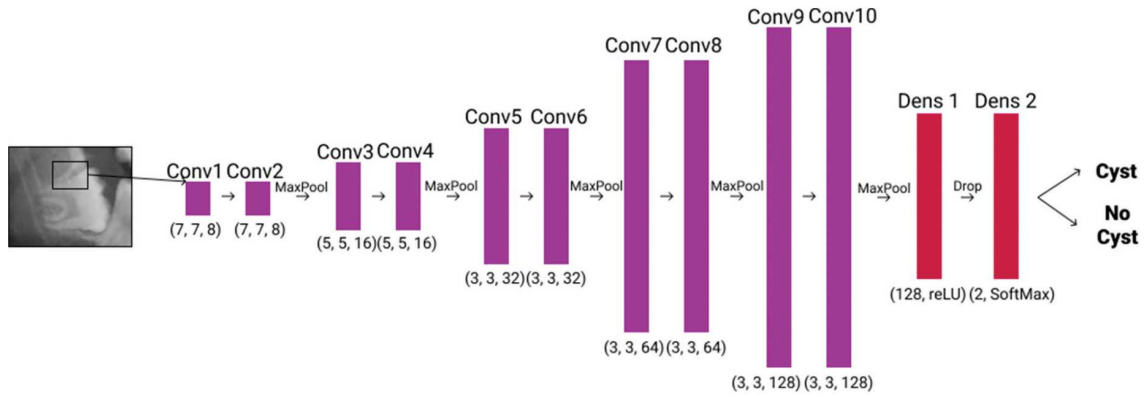
3.4.1.3 DCNNDAN

O estudo realizado por (OLIVEIRA *et al.*, 2022) apresenta duas Redes Neurais Convolucionais Profundas (DCNNs), denominadas CNN1 e CNN2. Neste trabalho, optamos por utilizar a CNN2, a qual nomeamos como DCNNDan. Nas convoluções, a DCNNDan utiliza camadas duplas para cada quantidade de filtros, com quantidades de 8, 16, 32, 64 e 128 filtros, também aplicando preenchimento nos inputs de cada convolução e passos com um valor de 1. Os tamanhos de janela usados nos filtros são de 3×3 , 5×5 e 7×7 .

Além disso, em todas as camadas de ativação, é utilizada a função ReLU. Camadas

de *pooling* são aplicadas após as camadas de convolução dupla, também utilizando *MaxPooling* com tamanho 2×2 . Após as convoluções, são utilizadas duas camadas totalmente conectadas, com 128 e 2 neurônios (número de classes). A primeira usa a função de ativação ReLU e a segunda usa *softmax*. Entre essas camadas, a técnica de dropout com 0,2 é aplicada.

Figura 10 – Arquitetura da rede DCNNDAN



Fonte: (KARSAZ, 2022)

3.4.1.4 Função de perda

As funções de perda desempenham o papel de calcular os erros cometidos durante o processo de treinamento dos algoritmos de aprendizado de máquina. O objetivo é que o valor de saída da função de perda seja menor a cada época de aprendizado; portanto, quanto menor for esse valor, melhor será o desempenho do algoritmo na realização de predições corretas.

Um exemplo de função de perda é a entropia cruzada, que calcula o erro através da diferença entre duas distribuições de probabilidade. Nela, são comparados os termos da distribuição de rótulos verdadeiros com os valores de predição da camada de saída da rede neural. Abaixo, segue a equação da função de perda definida como CE:

$$CE = \sum_i^C y_i \log(s_i) \quad (3.4)$$

onde C é o número de classes, y é a distribuição de rótulos corretos e s é a distribuição de saída da função de ativação da última camada da rede neural.

3.4.1.5 Otimizadores

Para minimizar os valores de saída da função de perda, os hiperparâmetros e pesos do modelo são ajustados por funções otimizadoras durante o processo de treinamento. Essas funções são utilizadas pelos algoritmos de aprendizado de máquina com o propósito de melhorar o desempenho do treinamento e, conseqüentemente, aumentar sua velocidade de aprendizado. Assim, elas atualizam os parâmetros do modelo de acordo com a saída da função de perda.

Os otimizadores são baseados no método do gradiente descendente. Este método realiza iterações com o objetivo de encontrar os valores dos parâmetros que irão minimizar uma função objetivo $J(\theta)$, chamada de função de custo no nosso problema. Onde θ são os parâmetros do modelo, dados para todo $\theta \in \mathbb{R}^d$, e os parâmetros são atualizados na direção oposta ao gradiente da função objetivo $\nabla J(\theta)$. Além disso, ele usa uma taxa de aprendizado η , que controla a rapidez com que o otimizador irá encontrar o mínimo local da função de custo, ou seja, controla a distância que o algoritmo deve percorrer em cada iteração (RUDER, 2016). Sabendo que o número da iteração é dado por t , segue abaixo a função que calcula a descida do gradiente para a função de custo:

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t) \quad (3.5)$$

É crucial encontrar um valor ideal para a taxa de aprendizado, uma vez que isso determina quanto tempo o algoritmo levará para encontrar o mínimo local da função. Deve-se ter cuidado na escolha desse valor, pois uma taxa de aprendizado alta pode fazer o modelo convergir rapidamente para um resultado inferior ao ideal, enquanto uma taxa baixa pode travar o aprendizado do modelo, gerando *underfitting*.

Taxas de aprendizado maiores exigem menos épocas de treinamento, devido às grandes mudanças nos parâmetros durante o treinamento. Por outro lado, taxas de aprendizado menores exigem mais épocas de treinamento, sendo consequência das pequenas alterações nos parâmetros ao longo do treinamento.

Além disso, para melhorar o desempenho do otimizador, pode-se utilizar o gradiente descendente com *momentum*, também conhecido como impulso. Como o próprio nome sugere, o impulso é utilizado para acelerar o otimizador, sendo calculado a partir da multiplicação de uma taxa de decaimento γ com o vetor de velocidade utilizando os parâmetros da iteração anterior.

O vetor de impulso, somado à multiplicação da taxa de aprendizado com o vetor gradiente, gera um vetor de velocidade v_t que indica a direção e a velocidade para ajuste dos novos parâmetros. Dessa forma, seu tamanho é maior para direções que apontam no mesmo sentido que o gradiente, e menor para direções em que o gradiente muda de direção. Isso faz com que o otimizador convirja mais rapidamente para o mínimo local da função de perda. A função $v(t)$ utilizando o *momentum* é dada por:

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta_{t-1}) \quad \theta_t = \theta_{t-1} - v_t \quad (3.6)$$

O otimizador com *momentum* calcula a cada iteração o vetor de velocidade com o gradiente acumulado e depois dá um passo em sua direção. O problema disso é que muitas vezes sua velocidade pode ser muito alta, fazendo com que ele passe do mínimo local, da mesma forma sua velocidade pode ser muito baixa e isso pode gerar uma estagnação e parada em um local indevido.

3.4.1.5.1 Adagrad

O Adagrad é um otimizador que possui taxas de aprendizado específicas, adaptadas aos parâmetros ao longo do treinamento. Ele realiza atualizações de acordo com a frequência dos parâmetros, aplicando maiores atualizações a parâmetros mais frequentes (DUCHI *et al.*, 2011). Por esse motivo, ele é mais eficiente para trabalhar com dados esparsos. Pode-se definir o gradiente da função objetivo como $g_{t,i}$, para cada parâmetro θ_i , na iteração t .

$$g_{t,i} = \nabla_{\theta} J(\theta_{t-1,i}) \quad (3.7)$$

O cálculo para os parâmetros que minimizam a função de custo é dado abaixo.

$$\theta_{t,i} = \theta_{t-1,i} - \eta \cdot g_{t,i} \quad (3.8)$$

A atualização dos parâmetros utilizando o Adagrad modifica a taxa de aprendizado a cada iteração t , para cada parâmetro θ_i , com base nos gradientes anteriores para θ_i . A equação abaixo mostra o cálculo dos parâmetros utilizando o Adagrad:

$$\theta_{t,i} = \theta_{t-1,i} - \frac{\eta}{\sqrt{G_{t-1,ii} + \epsilon}} \cdot g_{t,i} \quad (3.9)$$

Sendo $G_t \in \mathbb{R}^{d \times d}$ uma matriz diagonal, onde cada elemento i, i é a soma dos quadrados dos gradientes para o parâmetro θ_i até a iteração t . Além disso, ε é um termo de suavização para impedir a divisão por zero.

O benefício do Adagrad é que ele elimina a necessidade de ajuste manual da taxa de aprendizado. Em contrapartida, ele acumula os gradientes quadrados no denominador, fazendo com que esse valor tenda a aumentar a cada iteração e continue crescendo ao longo do treinamento. Isso faz com que a taxa de aprendizado calculada pelo algoritmo diminua, podendo se tornar tão pequena que o algoritmo pode perder a capacidade de continuar aprendendo.

3.4.1.5.2 RMSprop

O RMSprop foi desenvolvido para reduzir o problema da taxa de aprendizado monotonicamente decrescente gerado pelo Adagrad (RUDER, 2016). Ele é um método de taxa de aprendizagem adaptativa que realiza a divisão da taxa de aprendizado pela média exponencialmente decrescente de gradientes quadrados.

Este método é bem parecido com o Adadelta, porém ele não utiliza a média exponencialmente decrescente dos quadrados da atualização dos parâmetros, sendo dependente desta forma de um valor de taxa de aprendizado. Abaixo segue a equação para atualização dos parâmetros utilizando o RMSprop.

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{E[g^2]_{t-1} + \varepsilon}} g_{t-1} \quad (3.10)$$

3.4.1.5.3 Adam

O otimizador de estimativa de momento adaptativo, também conhecido como Adam, é um método que calcula taxas de aprendizado adaptativas para cada parâmetro (KINGMA; BA, 2014). Ele é simples de implementar e amplamente utilizado para problemas com grandes volumes de dados e parâmetros. Adam utiliza m_t para a média exponencialmente decrescente de gradientes passados e v_t para o cálculo da média exponencialmente decrescente dos quadrados dos gradientes passados. Os momentos de primeira e segunda ordem são dados por:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \quad (3.11)$$

onde β_1 e β_2 são as taxas de decaimento exponencial para as estimativas de primeiro e segundo momento, respectivamente. As estimativas de primeiro e segundo momento são inicializadas com valor zero, o que pode enviesar os valores para zero ao longo do treinamento, especialmente quando as taxas de decaimento exponencial são muito baixas. Para resolver esse problema, utilizam-se as equações abaixo:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.12)$$

A equação para a atualização dos parâmetros utilizando o otimizador Adam é dada por:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{\hat{v}_{t-1} + \epsilon}} \hat{m}_{t-1} \quad (3.13)$$

4 METODOLOGIA

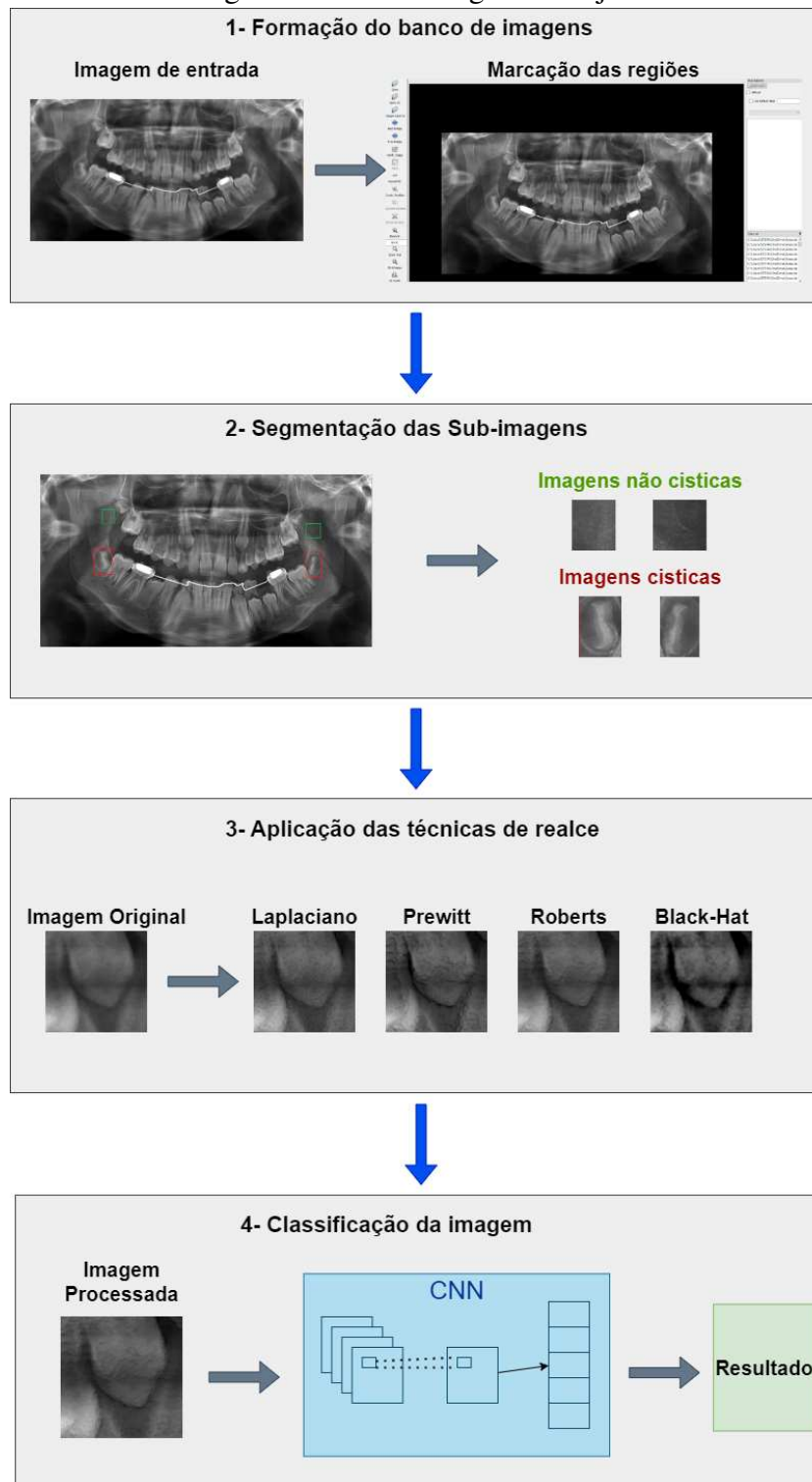
As etapas necessárias para a detecção automática de cistos em imagens de radiografia panorâmica, descritas neste capítulo, fazem uso das técnicas apresentadas no capítulo 3. A Figura 11 ilustra o fluxograma da metodologia adotada, detalhando cada uma das etapas envolvidas. A primeira etapa consiste na formação do banco de imagens e na marcação das regiões císticas e não císticas, conforme descrito na seção 4.1.

Na etapa 2, detalhada na seção 4.2, são aplicados pré-processamentos às imagens para prepará-las para as fases subsequentes. Nesse momento, ocorre a segmentação das regiões identificadas na etapa anterior. Em seguida, na etapa 3, discutida na seção 4.3, são aplicados filtros de realce nas imagens para destacar as principais características relevantes para a classificação.

O estágio 4, descrito na seção 4.5, envolve o uso de algoritmos de CNN para classificar as imagens em císticas ou não císticas. Durante o processo de treinamento, a etapa de aumento de dados é incorporada, como descrito na seção 4.4, para expandir a base de dados e aumentar a confiabilidade do algoritmo.

Por fim, a técnica de validação cruzada, explicada na seção 4.6, foi utilizada para mensurar a performance dos modelos de aprendizado de máquina. Além disso, essa seção também apresenta as métricas utilizadas para avaliar o desempenho dos modelos quando aplicados a dados desconhecidos.

Figura 11 – Metodologia do Projeto.



Fonte: Elaborado pela autora.

4.1 Formação do Banco de Imagens

Os exames utilizados neste trabalho consistem em radiografias panorâmicas completas da região bucal, obtidas por meio da digitalização de exames de radiografia. A base de imagens utilizada foi formada a partir de duas fontes de dados. A primeira fonte foi desenvolvida com a colaboração de dois dentistas especialistas da Universidade Federal do Ceará (UFC) e um estudante do curso de Engenharia da Computação, que identificaram e marcaram as regiões císticas e não císticas das imagens. Sendo assim, foram utilizadas 604 imagens de radiografias provindas desta base.

A segunda fonte utilizada foi a *Tufts Dental Database* (PANETTA *et al.*, 2021), que contém mais de 1.000 exemplos de radiografias panorâmicas completas. No entanto, neste trabalho, foram utilizadas apenas 24 imagens dessa base, pois estas continham as regiões císticas de interesse. A Figura 12 ilustra duas das radiografias utilizadas: a Figura 12a foi extraída da base da UFC, enquanto a Figura 12b é proveniente da base Tufts. Ademais, a Tabela 1 mostra a quantidade de imagens utilizadas em cada base.

Tabela 1 – Relação da quantidade de imagens utilizada de cada base.

Base de dados	Quantidade Imagens
Imagens da UFC	604
Imagens da Tufts Dental Database	24
Total	628

Fonte: Elaborado pela autora.

A marcação das regiões de interesse foi realizada com o software LabelImg, uma ferramenta de código aberto que pode ser instalada via o gerenciador de pacotes do Python. Este software permite salvar os rótulos e as localizações das regiões marcadas em três formatos de arquivo: .txt, .json e .xml. A Figura 13 exibe a tela inicial do LabelImg preparado para iniciar as marcações.

4.2 Extração da Região de Interesse

As marcações das imagens geraram arquivos JSON contendo as coordenadas x e y de cada uma das áreas selecionadas. Com essas informações, foi possível avançar para a etapa de segmentação automatizada, que separou as regiões de interesse de cada imagem original.

Figura 12 – Exemplo de radiografias panorâmicas.



a) Radiografia da Base da UFC



b) Radiografia da Base da Tufts

Fonte: Elaborado pela autora.

Essas sub-imagens segmentadas constituem a base para as análises subsequentes. Ao associar cada sub-imagem ao seu rótulo correspondente, foi possível estruturar um banco de dados que servirá de insumo para o treinamento e teste das redes neurais convolucionais, permitindo avaliar o desempenho de diferentes filtros de realce na classificação das imagens císticas. A Tabela 2 exibe a relação da quantidade de sub-imagens e seus rótulos.

Tabela 2 – Relação da quantidade de sub-imagens por rótulos.

Tipo de imagem	Quantidade Imagens
Imagens Císticas	1219
Imagens Não Císticas	1212

Fonte: Elaborado pela autora.

Figura 13 – Tela inicial do *software* LabellImg.



Fonte: Elaborado pela autora.

4.3 Aplicação do filtro de Realce

Após a segmentação das imagens panorâmicas em sub-imagens, foi desenvolvido um algoritmo específico para aplicar técnicas de realce visando destacar as principais características dessas imagens. A segmentação inicial permitiu dividir as imagens maiores em partes menores e mais gerenciáveis, facilitando a aplicação subsequente das técnicas de processamento de imagem. Essa etapa é crucial para garantir que as características relevantes das imagens, como bordas e contrastes, sejam enfatizadas adequadamente.

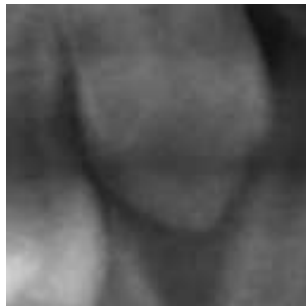
O algoritmo foi desenvolvido em Python com o auxílio da biblioteca OpenCV. Este algoritmo utiliza uma combinação de diferentes técnicas de realce para melhorar a qualidade visual e a precisão da identificação das estruturas nas imagens. As técnicas empregadas foram: filtro laplaciano, que realça regiões de transição rápida de intensidade; filtro de Prewitt, que detecta bordas ao calcular a mudança de intensidade; filtro de Roberts, que também é um detector de bordas, mas que se concentra em transições diagonais; e a técnica Black-hat, que evidencia detalhes mais escuros em fundos claros, realçando as estruturas que poderiam passar despercebidas em métodos tradicionais.

A Figura 14 exibe a aplicação dos filtros utilizados em uma subimagem de uma lesão cística. A Figura 14a apresenta a imagem original, sem a aplicação de nenhuma técnica de realce. A Figura 14b mostra a imagem após a aplicação do filtro laplaciano, onde é possível observar o destaque das áreas escuras e a criação de ruído do tipo sal e pimenta. Na Figura 14c, a aplicação

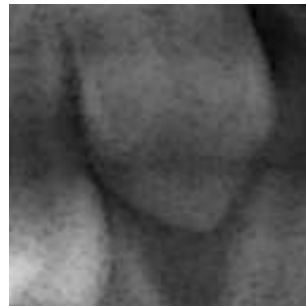
do filtro de Prewitt também realça as regiões escuras, criando uma textura densa e escura. A Figura 14d exibe o resultado do filtro de Roberts, que gera uma textura escura, porém mais suave em comparação ao filtro de Prewitt. Já a Figura 14e apresenta a aplicação da morfologia black-hat, que foi a técnica que mais intensificou as áreas escuras, além de adicionar um ruído sal e pimenta mais denso.

Ao aplicar esses filtros, o algoritmo conseguiu aprimorar as bordas e contrastes das imagens, proporcionando uma base mais robusta para as etapas subsequentes de análise e classificação das imagens. Porém, é importante ter cuidado ao utilizá-los, pois, além de realçar detalhes, esses filtros podem intensificar ou gerar ruídos indesejados nas imagens.

Figura 14 – Exemplo da aplicação dos filtros utilizados em uma sub-imagem cística.



a) Imagem Original



b) Filtro Laplaciano



c) Filtro de Prewitt



d) Filtro de Roberts



e) Black-hat

Fonte: Elaborado pela autora.

4.4 Aumento de dados

A eficácia dos algoritmos de aprendizado de máquina está diretamente ligada à quantidade de dados disponíveis para o treinamento. Portanto, para garantir maior confiabilidade nos modelos desenvolvidos, torna-se essencial utilizar técnicas que ampliem a base de dados. Para isso, foi empregada a função *ImageDataGenerator()* ² da biblioteca Keras do Python.

A ampliação da base de imagens é realizada por meio da aplicação de um conjunto de transformações nas imagens originais, gerando assim inúmeras variações de uma mesma imagem. Neste trabalho, as seguintes transformações foram aplicadas ao conjunto de imagens de treinamento:

- Zoom: Ampliação da imagem original.
- Espelhamento Vertical e Horizontal: Inversão da imagem nas direções vertical ou horizontal, de forma aleatória.
- Rotação: Gira a imagem com base em um ângulo especificado como parâmetro da função.
- Cisalhamento: Aplica um corte angular na imagem, distorcendo-a.
- Deslocamento Vertical e Horizontal: Move a imagem ao longo dos eixos vertical ou horizontal.

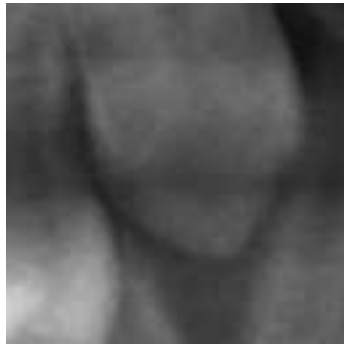
A Figura 15 exibe exemplos dos resultados obtidos com o aumento de dados a partir de uma imagem cística. Na Figura 15a, observa-se a imagem original. A Figura 15b mostra a imagem após a aplicação da técnica de cisalhamento, enquanto a Figura 15c apresenta o deslocamento horizontal. Já a Figura 15d exibe o deslocamento vertical, e a Figura 15e ilustra a rotação da imagem. A Figura 15f apresenta o efeito do zoom, seguida pela Figura 15g, que demonstra o espelhamento horizontal. Por fim, a Figura 15h exibe o espelhamento vertical. Essas visualizações evidenciam como as técnicas de aumento de dados geram diversas variações a partir de uma única imagem, gerando desta forma um *dataset* ampliado.

4.5 Arquiteturas de CNNs utilizadas

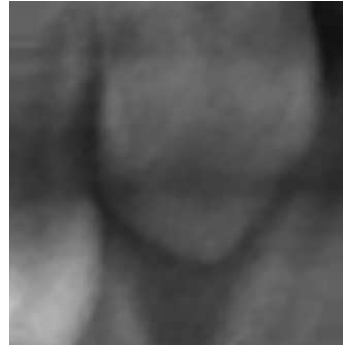
Para o desenvolvimento dos algoritmos de classificação, a linguagem Python foi escolhida como base, utilizando as bibliotecas TensorFlow e Keras. Essas bibliotecas são amplamente reconhecidas na comunidade de aprendizado de máquina por suas capacidades

² Disponível em: <<https://www.tensorflow.org/apidocs/python/tf/keras/preprocessing/image/ImageDataGenerator>>. Acesso em: 31 de julho de 2024.

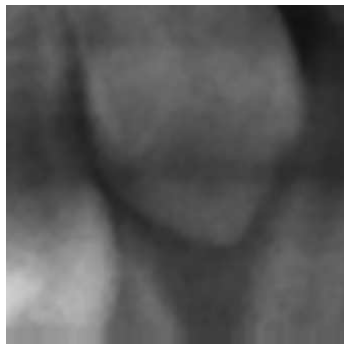
Figura 15 – Exemplo da aplicação das técnicas de aumento de dados.



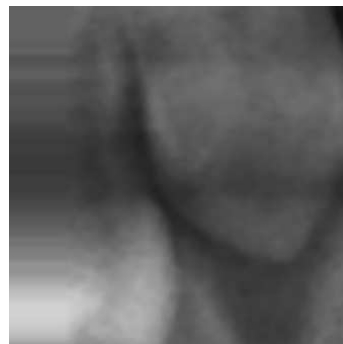
a) Imagem Original



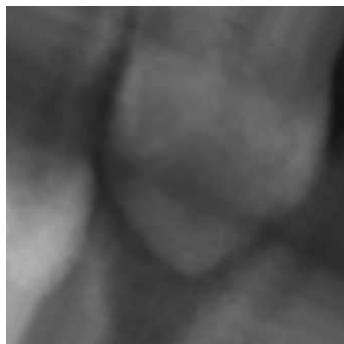
b) Cisalhamento



c) Deslocamento Horizontal



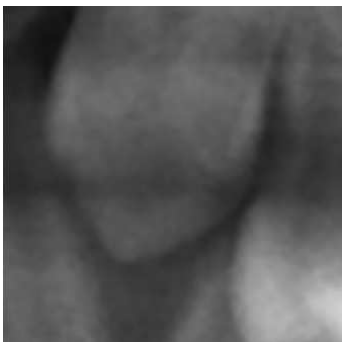
d) Deslocamento Vertical



e) Rotação



f) Zoom



g) Espelhamento Horizontal



h) Espelhamento Vertical

Fonte: Elaborado pela autora.

robustas no desenvolvimento de arquiteturas de redes neurais. O Keras, oferece uma interface de alto nível que simplifica a criação de modelos complexos, enquanto o TensorFlow fornece a infraestrutura necessária para o processamento eficiente desses modelos. Por esses motivos, essas ferramentas foram fundamentais para a construção e experimentação das redes neurais propostas neste estudo.

Para o treinamento dos modelos, foi adotado imagens de 128x128 pixels nos três canais de cores RGB. O tamanho do lote (*batch size*) foi fixado em 32, utilizando a função de perda de entropia cruzada categórica (*Categorical Crossentropy*). Além disso, foram utilizadas 300 épocas e a técnica de "EarlyStopping" para monitorar as métricas de desempenho do modelo e interromper o treinamento quando o desempenho começar a degradar, prevenindo assim o *overfitting*.

Quatro redes neurais foram implementadas durante esta fase do projeto, cada uma com características e objetivos específicos. A InceptionV3 foi a primeira rede criada, carregada já pré-treinada com os pesos da ImageNet. O uso de pesos pré-treinados é uma técnica comum em aprendizado por transferência, permitindo que modelos previamente treinados em grandes *datasets* sejam ajustados para outras tarefas.

As redes AlexNet e DCNNDan foram construídas utilizando o módulo "layers" do Keras. Esse módulo permite a implementação flexível de várias camadas, como convolução, densas, *flatten*, *dropouts* e *max-pooling*, que são os blocos fundamentais de qualquer CNN. Ambas as redes foram escolhidas para serem utilizadas nesta pesquisa por serem utilizadas nos trabalhos citados no tópico de trabalhos relacionados. A rede DCNNDan foi desenvolvida pelos seus criadores justamente para a classificação de imagens císticas, por isso decidimos utilizá-la neste trabalho.

A CNNGen foi a última rede neural desenvolvida e seguiu uma abordagem mais experimental e personalizada. Utilizando também o módulo "layers" do Keras, foi realizado uma série de experimentos para identificar a arquitetura mais eficaz para a classificação das imagens císticas. Esses experimentos foram guiados pela técnica de *grid search*, que consiste em testar sistematicamente diferentes combinações de hiperparâmetros. Foram variáveis a profundidade da rede (quantidade de camadas), a largura (quantidade de neurônios por camada), os valores de *max-pooling* e de *dropout*, além da quantidade de neurônios e camadas densas na camada totalmente conectada. O tópico a seguir desenvolve a explicação desta arquitetura criada neste trabalho.

4.5.1 CNNGen

Após uma análise dos resultados obtidos nos experimentos, foi identificada uma configuração de rede que apresentou um desempenho superior na tarefa de classificação. Esta arquitetura foi nomeada CNNGen. A CNN foi concebida com base em uma abordagem de aumento progressivo de complexidade, incorporando múltiplas camadas convolucionais e de pooling para capturar as características mais importantes nas imagens odontológicas.

Inicialmente, é aplicada uma camada convolucional com 32 filtros, utilizando uma janela de convolução 3x3 e ativação ReLU. Em seguida, são acrescentadas duas camadas convolucionais consecutivas, cada uma com 64 filtros, mantendo a mesma configuração de janela e ativação ReLU. Para regularizar o modelo e facilitar o treinamento, é incluída uma camada de normalização em lote após estas camadas, seguida por uma camada de *max pooling* para redução da dimensionalidade.

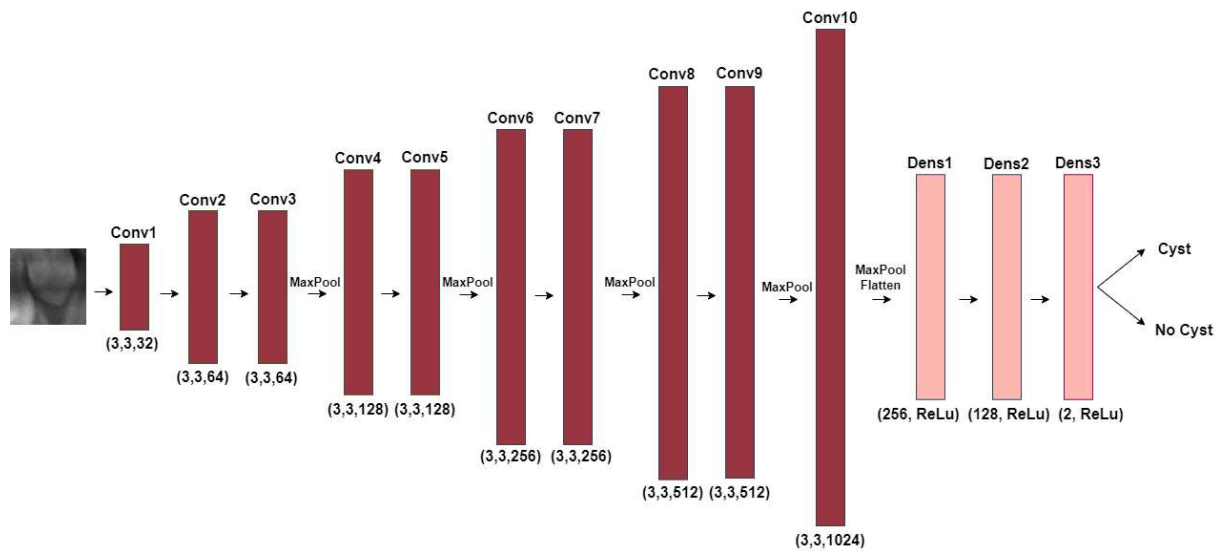
Este padrão é repetido com camadas convolucionais subsequentes, aumentando progressivamente o número de filtros para 128, 256 e 512, respectivamente, em cada conjunto de camadas convolucionais. Após cada conjunto, é aplicada normalização em lote e *max pooling* para controle de *overfitting* e redução da resolução espacial.

Por fim, após a última camada de *max pooling*, é aplicada uma camada de convolução adicional com 1024 filtros seguida por normalização em lote e *max pooling*. Em seguida, a saída é achatada (*flattened*) para alimentar uma sequência de camadas totalmente conectadas. São utilizadas duas camadas densas com ativação ReLU, com tamanhos de 256 e 128 unidades, respectivamente, antes de uma camada de saída com ativação softmax, contendo dois neurônios para representar as classes de saída. A Figura 16 exibe a arquitetura da CNNGen desenvolvida.

4.6 Validação cruzada

A validação cruzada é uma técnica usada para mensurar a performance de modelos de aprendizado de máquina. Um método comum para isso é o algoritmo K-fold, no qual o conjunto de dados é dividido em K partes, também conhecidas como *folds*. Nessa abordagem, K-1 partes são usadas para o treinamento do modelo, enquanto o subconjunto restante serve para avaliar a capacidade do modelo em classificar dados novos. Esse procedimento é repetido K vezes, garantindo que a cada iteração um *fold* diferente seja utilizado para teste. O K-fold foi implementado utilizando a biblioteca *Scikit-Learn*. A Figura 17 ilustra o diagrama de

Figura 16 – Arquitetura da CNNGen



Fonte: Elaborado pela autora.

funcionamento do K-fold, com K igual a 4. No final do treinamento, o melhor modelo gerado pelo K-fold é avaliado utilizando dados que ainda não foram utilizados anteriormente.

Figura 17 – Diagrama do K-fold.



Fonte: Elaborado pela autora.

4.7 Métricas de avaliação

As métricas de avaliação são utilizadas com o intuito de analisar o desempenho de modelos na predição correta das classes, utilizando dados desconhecidos, que são as imagens que não foram utilizadas durante o processo de treinamento. Os modelos utilizados neste trabalho são os gerados através do treinamento das redes InceptionV3, DCNNDan, CNNGen e AlexNet com as técnicas de realce abordadas na pesquisa.

O *dataset* utilizado neste trabalho possui imagens odontológicas classificadas com

regiões císticas e não císticas. Considerando como positiva as imagens com regiões císticas e negativas as imagens com regiões não císticas. Pode-se classificar as previsões do modelo em Verdadeiro Positiva (VP) uma previsão correta da região cística, Verdadeiro Negativa (VN) uma previsão correta da região não cística, Falso Positivo (FP) quando uma região não cística é classificada como cística e Falso Negativo (FN) quando uma imagem cística é classificada como não cística. As métricas de avaliação utilizadas nesse trabalho estão no modulo *Metrics*⁵ da Biblioteca *Scikit-Learn*. São elas:

- Acurácia (ACC): É considerada a métrica mais simples de avaliação, ela está relacionada com o número de previsões corretas do modelo.

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.1)$$

- Precisão: Percentual de imagens classificadas como positivas e que são realmente positivas.

$$VPP = \frac{VP}{VP + FP} \quad (4.2)$$

- *Recall*: Calcula a capacidade classificatória correta dos modelos na predição das regiões císticas.

$$SEN = \frac{VP}{VP + FN} \quad (4.3)$$

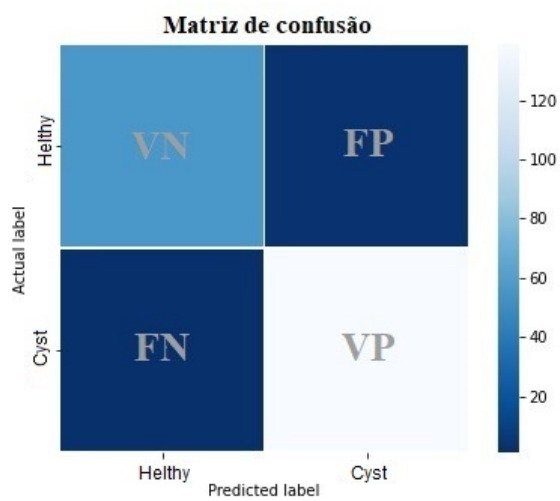
- F1-Score: É a média harmônica entre o *recall* e a precisão.

$$F1 = \frac{2 \cdot (VPP \cdot SEN)}{VPP + SEN} \quad (4.4)$$

- Matriz de confusão: É uma tabela que realiza a exibição da relação entre os falsos positivos e negativos e os verdadeiros positivos e negativos. A Figura 18 exibe uma matriz de confusão.

⁵ Disponível em: <https://scikit-learn.org/stable/modules/model_evaluation.html>. Acesso em: 30 de julho de 2024.

Figura 18 – Matriz de confusão.



Fonte: Elaborado pela autora.

5 RESULTADOS

Este capítulo irá apresentar os resultados obtidos nos experimentos realizados a partir da metodologia especificada no Capítulo 4. Os experimentos foram executados utilizando um notebook Dell G3 processador Intel core i7, 16 Gb de memória RAM e placa gráfica da NVIDIA Gerfoce GTX 1040ti. Além disso, também foram executados experimentos utilizando o Google Colab Pro, um servidor em nuvem pago da Google.

Inicialmente, foram extraídas sub-imagens das radiografias panorâmicas, formando uma nova base de dados. Em seguida, essas sub-imagens passaram por técnicas de realce, aprimorando a qualidade das regiões de interesse. Diversos experimentos foram conduzidos utilizando diferentes otimizadores e taxas de aprendizado, visando identificar a melhor abordagem para otimizar as CNNs. A validação cruzada foi aplicada para garantir a qualidade do modelo, e as métricas de desempenho foram calculadas utilizando o melhor modelo identificado durante esse processo.

Ademais, este trabalho também teve como objetivo testar diferentes arquiteturas de CNNs, avaliando tanto redes sem pesos pré-treinados quanto redes com pesos já pré-treinados, além de arquiteturas desenvolvidas especificamente para a classificação de imagens císticas odontológicas. A rede InceptionV3 foi utilizada com seus pesos pré-treinados, enquanto a AlexNet foi implementada sem o uso de pesos pré-treinados. A rede DCNNDan, criada por outros autores com o propósito de classificar imagens císticas, foi testada para comparação. Por fim, a rede CNNGen foi desenvolvida especificamente neste estudo, com o objetivo de criar uma arquitetura mais especializada e eficiente para a classificação dessas imagens odontológicas. Os resultados e os detalhes desses experimentos serão apresentados nas subseções seguintes.

5.1 Resultados da validação cruzada das CNNs variando os otimizadores

Os primeiros experimentos utilizando CNNs nesta pesquisa foram desenvolvidos visando identificar a melhor taxa de aprendizado e o otimizador mais adequado para cada rede. Neste capítulo, a taxa de aprendizado será chamada de lr , que vem da abreviação de *learning rate*, do inglês. Atualmente, a literatura possui uma enorme variedade de funções de otimização, porém os experimentos deste trabalho utilizaram as funções Adam, Adagrad e RMSProp.

Ademais, foi aplicada a técnica de validação cruzada com $k = 4$, utilizando as redes InceptionV3, DCNNDan, AlexNet e CNNGen. Em cada execução do k -fold, uma rede neural

foi testada com um otimizador e uma taxa de aprendizado distinta. As taxas avaliadas foram 0,1, 0,01, 0,001 e 0,0001. As subseções a seguir apresentarão os resultados obtidos a partir desses experimentos.

5.1.1 Adagrad

O primeiro otimizador utilizado nos experimentos foi o Adagrad. A Tabela 3 apresenta as médias das acurácias e os desvios padrões obtidos ao validar o desempenho do modelo com a base de imagens de teste. Com relação a InceptionV3 os piores resultados foram obtidos utilizando o lr de 0,01 e os melhores com o lr de 0,0001, tendo acurácias médias de 50.86% e 67.80%, respectivamente. A rede AlexNet teve os piores resultados utilizando o lr de 0,1, com média de acurácia de 48.86% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 78.68%.

A rede DCNNDan teve os piores resultados utilizando o lr de 0,1, com média de acurácia de 50.40% e os melhores resultados utilizando o lr de 0,01 com média de acurácia de 82.38%. Por fim, a A rede CNNGen obteve os piores resultados utilizando o lr de 0,0001, com média de acurácia de 57.74% e os melhores resultados utilizando o lr de 0,01 com média de acurácia de 64.78%.

Tabela 3 – Acurácias médias para as taxas de otimização do Adagrad.

CNN	0,1	0,01	0,001	0,0001
InceptionV3	51.23% $\pm$ 0.42	50.86% $\pm$ 0.26	63.64% $\pm$ 7.42	67.80% $\pm$1.95
AlexNet	48.86% $\pm$ 21.31	59.85% $\pm$ 11.20	71.50% $\pm$ 1.70	78.69% $\pm$1.54
DCNNDan	50.40% $\pm$ 0.65	82.38% $\pm$0.94	78.58% $\pm$ 0.51	72.58% $\pm$ 0.51
CNNGen	62.02% $\pm$ 8.15	64.78% $\pm$6.28	60.41% $\pm$ 6.91	57.74% $\pm$ 4.50

Fonte: Elaborado pela autora.

A Tabela 4 exhibe a relação das melhores taxas e a acurácia e f1-score do melhor modelo desenvolvido durante a validação cruzada utilizando a rede e a taxa especificada. Sendo assim, a melhor taxa para as redes InceptionV3 e AlexNet foi com lr de 0,0001 e para as redes DCNNDan e CNNGen foi com o lr de 0,01. Além disso, A rede DCNNDan obteve o melhor desempenho utilizando o otimizador Adagrad, alcançando o valor de 83,98% de acurácia e 85,17% de f1-score. Os piores resultados utilizando esse otimizador foram obitidos com a rede InceptionV3, onde teve apenas 69,19% de acurácia e 69,87% de f1-score.

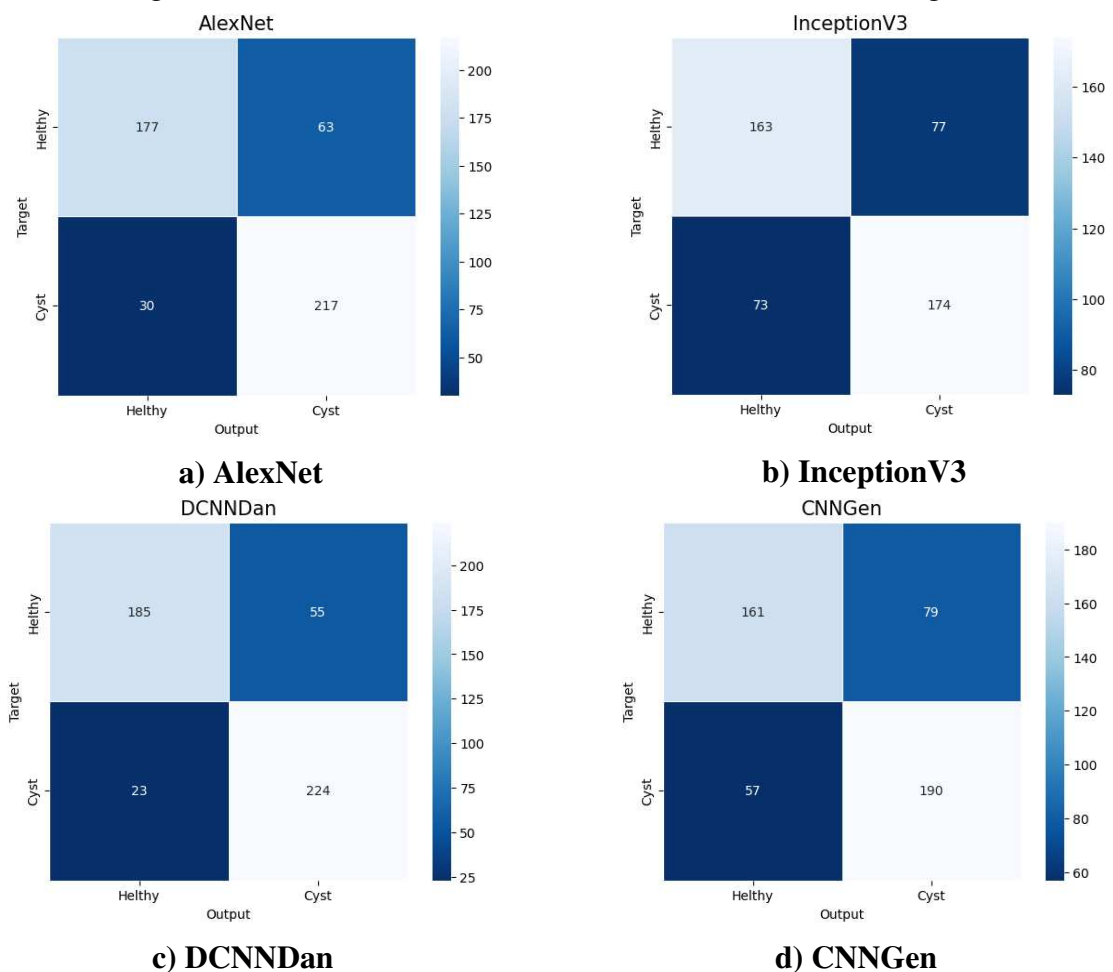
Tabela 4 – Acurácias médias para as taxas de otimização do Adagrad.

CNN	LR	Acurácia	F1-Score
InceptionV3	0,0001	69,19%	69,87%
AlexNet	0,0001	80,90%	82,35%
DCNNDan	0,01	83,98%	85,17%
CNNGen	0,01	72,07%	73,64%

Fonte: Elaborado pela autora.

A Figura 19 mostra as matrizes de confusão dos melhores modelos desenvolvidos durante a validação cruzada para cada rede. A CNNGen apresentou o maior número de falsos positivos, classificando incorretamente 79 imagens como císticas. A InceptionV3, por sua vez, teve o maior número de falsos negativos, com 73 imagens erroneamente classificadas como não císticas. A DCNNDan foi a mais eficiente na identificação de todas as classes, com 185 verdadeiros negativos e 224 verdadeiros positivos.

Figura 19 – Matrizes de Confusão da CNN treinada com o Adagrad.



Fonte: Elaborado pela autora.

5.1.2 RMSProp

O segundo otimizador testado foi o RMSProp. A Tabela 5 apresenta as médias das acurácias e os desvios padrões obtidos ao validar o desempenho do modelo com a base de imagens de teste. Com relação a InceptionV3 os piores resultados foram obtidos utilizando o lr de 0,01 e os melhores com o lr de 0,0001, tendo acurácias médias de 50,20% e 81,05%, respectivamente. A rede AlexNet teve os piores resultados utilizando o lr de 0,1, com média de acurácia de 50,04% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 61,08%.

A rede DCNNDan teve os piores resultados utilizando o lr de 0,1, com média de acurácia de 50,71% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 79,10%. Por fim, a rede CNNGen obteve os piores resultados utilizando o lr de 0,0001, com média de acurácia de 50,81% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 73,40%.

Tabela 5 – Acurácias médias para as taxas de otimização do RMSProp.

CNN	0,1	0,01	0,001	0,0001
InceptionV3	50,61% $\pm$ 1,01	50,20% $\pm$ 0,53	65,39% $\pm$ 12,52	81,05% $\pm$ 2,35
AlexNet	50,71% $\pm$ 0	50,04% $\pm$ 0,77	50,76% $\pm$ 0,36	61,08% $\pm$ 3,04
DCNNDan	50,71% $\pm$ 0	55,69% $\pm$ 9,47	55,43% $\pm$ 3,04	79,10% $\pm$ 1,26
CNNGen	50,81% $\pm$ 0,10	53,18% $\pm$ 4,77	59,90% $\pm$ 8,48	73,40% $\pm$ 3,28

Fonte: Elaborado pela autora.

A Tabela 6 exibe a relação das melhores taxas e a acurácia e f1-score do melhor modelo desenvolvido durante a validação cruzada utilizando a rede e a taxa especificada. Com isso, a melhor taxa para todas as redes foi a de 0,0001, onde as acurácias foram de: 84,80% para InceptionV3, 65,29% para a AlexNet, 80,08% para a DCNNDan e 77,61% para a CNNGen. A rede InceptionV3 obteve o melhor desempenho utilizando o otimizador RMSProp, alcançando o valor de 84,80% de f1-score. Os piores resultados utilizando esse otimizador foram obtidos com a rede AlexNet, onde teve apenas 71,40% de f1-score.

A Figura 20 exibe as matrizes de confusão dos melhores modelos desenvolvidos durante a validação cruzada para cada rede. A AlexNet apresentou o maior número de falsos positivos, classificando incorretamente 133 imagens como císticas. A CNNGen, por sua vez,

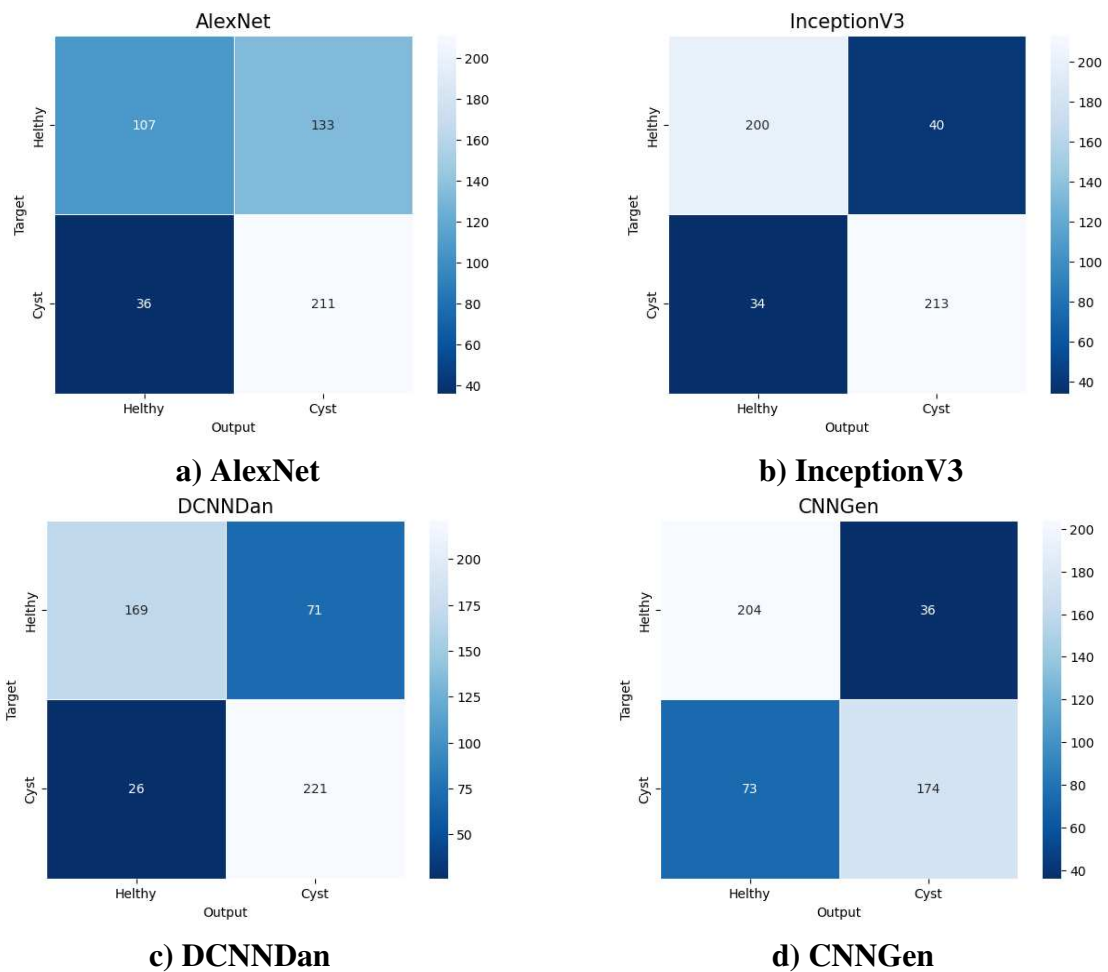
Tabela 6 – Acurácias médias para as taxas de otimização do RMSProp.

CNN	LR	Acurácia	F1-Score
InceptionV3	0,0001	84,80%	85,20%
AlexNet	0,0001	65,29%	71,40%
DCNNDan	0,0001	80,08%	82,00%
CNNGen	0,0001	77,61%	76,14%

Fonte: Elaborado pela autora.

teve o maior número de falsos negativos, com 73 imagens erroneamente classificadas como não císticas. A CNNGen foi a mais eficiente na identificação de imagens não císticas, com 204 verdadeiros negativos, enquanto a DCNNDan alcançou o melhor desempenho na classificação de imagens císticas, com 221 verdadeiros positivos.

Figura 20 – Matrizes de Confusão da CNN treinada com o RMSProp.



Fonte: Elaborado pela autora.

5.1.3 Adam

O último otimizador testado foi o Adam. A Tabela 7 apresenta as médias das acurácias e os desvios padrões obtidos ao validar o desempenho do modelo com a base de imagens de teste. Com relação a InceptionV3 os piores resultados foram obtidos utilizando o lr de 0,1 e os melhores com o lr de 0,0001, tendo acurácias médias de 50,15% e 82,44%, respectivamente. A rede AlexNet teve os piores resultados utilizando o lr de 0,1, com média de acurácia de 49,89% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 83,57%.

A rede DCNNDan teve os piores resultados utilizando o lr de 0,01, com média de acurácia de 50,35% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 83,92%. Por fim, a rede CNNGen obteve os piores resultados utilizando o lr de 0,0001, com média de acurácia de 49,63% e os melhores resultados utilizando o lr de 0,0001 com média de acurácia de 85,31%.

Tabela 7 – Acurácias médias para as taxas de otimização do Adam.

CNN	0,1	0,01	0,001	0,0001
InceptionV3	50,15% $\pm$ 0,94	53,02% $\pm$ 1,80	68,06% $\pm$ 4,04	82,44 % $\pm$ 3,00
AlexNet	49,89% $\pm$ 0,61	50,71% $\pm$ 0	50,61% $\pm$ 1,01	83,57 % $\pm$ 0,88
DCNNDan	50,71% $\pm$ 0	50,35% $\pm$ 0,61	51,68% $\pm$ 2,06	83,92 % $\pm$ 1,18
CNNGen	49,63% $\pm$ 0,63	58,36% $\pm$ 2,90	71,81% $\pm$ 4,85	85,31 % $\pm$ 1,0

Fonte: Elaborado pela autora.

A Tabela 8 exibe a relação das melhores taxas e a acurácia e f1-score do melhor modelo desenvolvido durante a validação cruzada utilizando a rede e a taxa especificada. Com isso, a melhor taxa para todas as redes foi a de 0,0001, onde as acurácias foram de: 85,83% para InceptionV3, 82,34% para a AlexNet, 85,62% para a DCNNDan e 86,03% para a CNNGen. A rede CNNGen obteve o melhor desempenho utilizando o otimizador Adam, alcançando o valor de 86,77% de f1-score.

A Figura 21 exibe as matrizes de confusão dos melhores modelos desenvolvidos durante a validação cruzada para cada rede. A AlexNet apresentou o maior número de falsos positivos e negativos, classificando incorretamente 51 imagens como císticas e 35 imagens como não císticas. A CNNGen foi a mais eficiente na identificação de imagens não císticas, com 203

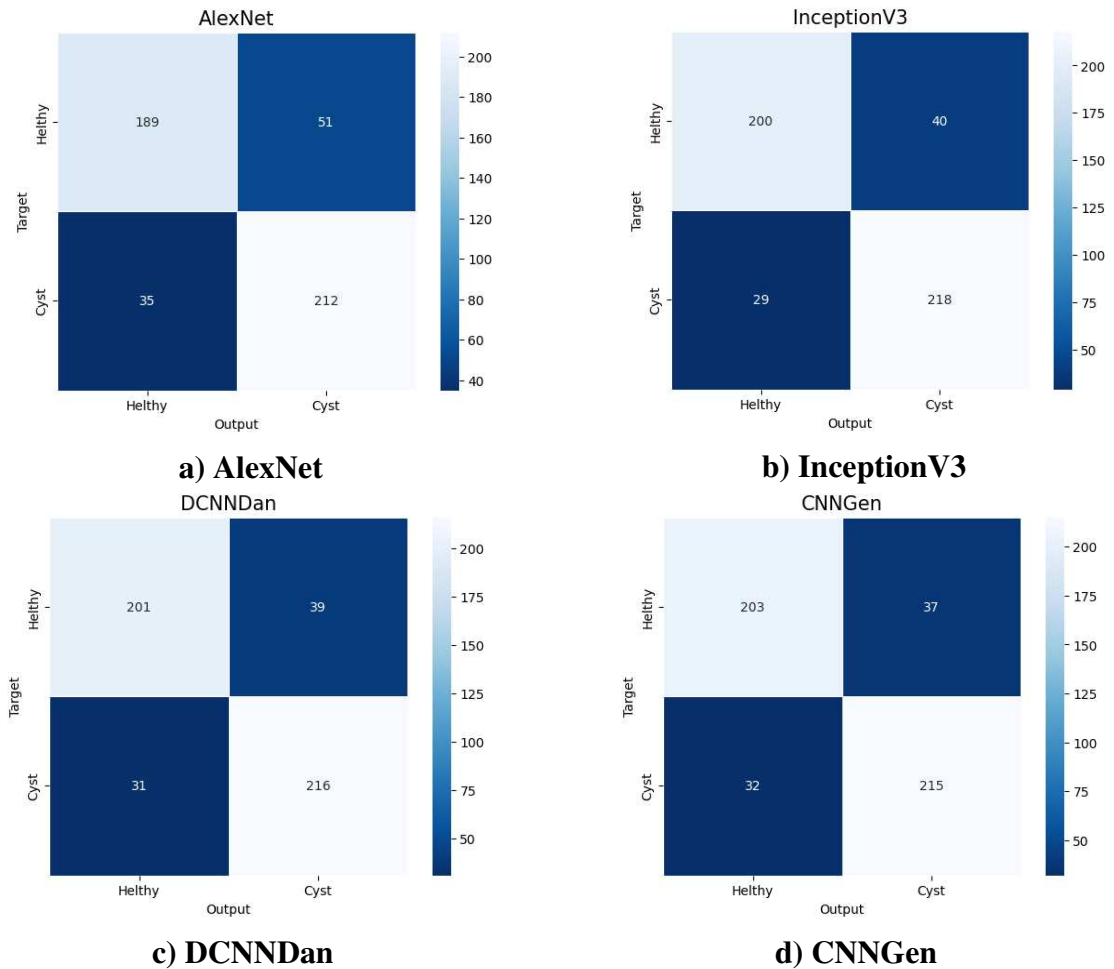
Tabela 8 – Acurácias médias para as taxas de otimização do Adam.

CNN	LR	Acurácia	F1-Score
InceptionV3	0,0001	85,83%	86,33%
AlexNet	0,0001	82,34%	86,13%
DCNNDan	0,0001	85,62%	86,05%
CNNGen	0,0001	86,03%	86,77%

Fonte: Elaborado pela autora.

verdadeiros negativos, enquanto a DCNNDan alcançou o melhor desempenho na classificação de imagens císticas, com 216 verdadeiros positivos.

Figura 21 – Matrizes de Confusão da CNN treinada com o Adam.



Fonte: Elaborado pela autora.

5.1.4 Análise dos resultados das funções de otimização

As partir dos resultados apresentados nas subseções anteriores, podemos tirar algumas conclusões. Na maioria dos resultados apresentados, os piores valores de médias de acurácia foram obtidos nos experimentos com taxas de aprendizado altas. Isso pode acontecer por muitos motivos, entre eles porque uma taxa de aprendizado alta pode acabar gerando *overshooting* que é quando o modelo ultrapassa o mínimo da função de custo e pode acabar gerando *undefitting*.

Por outro lado, na maioria dos experimentos, os melhores resultados foram obtidos utilizando uma taxa de aprendizado de 0,0001, a menor experimentada. Isso pode estar relacionado ao fato de que, com uma taxa de aprendizado tão baixa, o modelo executa atualizações menores nos pesos, permitindo uma convergência mais gradual e estabilizada para o mínimo da função de custo. Sendo assim, o modelo tem mais possibilidades de detectar os mínimos globais ou locais adequados para serem utilizados.

Além disso, as funções otimizadoras Adagrad e RMSProp obtiveram resultados intermediários de desempenho. Por outro lado, a função Adam utilizando a taxa de 0,0001 obteve os melhores resultados para todas as redes experimentadas. Sendo assim, essa função e essa taxa serão utilizadas nos experimentos das seções seguintes.

5.2 Resultados dos experimentos com CNN

A técnica de validação cruzada conhecida como *K-fold* foi utilizada para validar os modelos desenvolvidos. Com isso, em cada experimento realizado com o *k-fold*, o modelo foi treinado utilizando uma técnica de realce e uma CNN distinta. A Tabela 9 apresenta as médias de acurácia de cada modelo selecionado como o melhor em cada *fold*, avaliados utilizando as imagens de teste após o final do treinamento. Como uma forma de simplificar a análise e visualização dos resultados, a Tabela 9 não possui o resultado de acurácia de cada *fold*, possuindo apenas a média e o desvio padrão das acurácias obtidas nos 4 *folds* utilizados.

Para a CNN InceptionV3, o filtro Roberts registrou um desempenho abaixo da média, com uma acurácia média de 79,92%. Em contrapartida, o filtro BlackHat destacou-se como a melhor opção para esta rede, alcançando uma acurácia média de 84,39%. Além disso, para a arquitetura DCNNDan, o filtro Prewitt apresentou a menor média de acurácia, registrando 83,57%, uma queda em relação ao modelo que não utilizou nenhum filtro. Por outro lado, o filtro BlackHat se destacou com o melhor resultado médio, alcançando 84,49%.

Tabela 9 – Kfold e CNNs

CNN	Filtro	Média
InceptionV3	Sem filtro	82,44% $\pm$ 3,00
	Laplacian	82,49% $\pm$ 3,44
	Prewitt	82,39% $\pm$ 0,6
	Roberts	79,92% $\pm$ 4,8
	Blackhat	84,39% $\pm$ 2,97
DCNNDan	Sem filtro	83,92% $\pm$ 1,18
	Laplacian	84,28% $\pm$ 1,09
	Prewitt	83,57% $\pm$ 0,88
	Roberts	84,39% $\pm$ 0,75
	Blackhat	84,49% $\pm$ 1,60
AlexNet	Sem filtro	83,57% $\pm$ 0,88
	Laplacian	83,31% $\pm$ 1,46
	Prewitt	81,00% $\pm$ 1,65
	Roberts	75,76% $\pm$ 8,59
	Blackhat	83,92% $\pm$ 0,39
CNNGen	Sem filtro	85,31% $\pm$ 1,01
	Laplacian	85,26% $\pm$ 1,97
	Prewitt	86,08% $\pm$ 1,60
	Roberts	85,52% $\pm$ 1,44
	Blackhat	86,75% $\pm$ 1,39

Com relação à rede AlexNet o filtro de roberts obteve uma acurácia de 75,76% com desvio padrão de 8,59 possuindo resultados muito inferiores ao esperado. Além disso, os filtros Laplacian e Prewitt obtiveram acurácias médias de 83,31% e 81,00% respectivamente. O filtro BlackHat obteve o melhor resultado de acurácia e desvio padrão, sendo de 83,92% e 0,39 respectivamente. Ademais, com relação à rede CNNGen, o filtro Laplacian e o modelo sem filtro apresentaram desempenhos semelhantes, com 85,26% e 85,31%, respectivamente. Por outro lado, o filtro Blackhat se destacou como a opção mais eficaz, alcançando uma acurácia média de 86,75%.

5.2.1 Resultado sem aplicação de pré-processamento

Os resultados apresentados na Tabela 10 mostram o desempenho das quatro CNNs treinadas utilizando a base de imagens sem aplicação de filtros de realce. O pior desempenho foi

da rede AlexNet, com uma acurácia de 82,34%. Por outro lado, a rede DCNNDan se destacou, alcançando a melhor precisão, com um valor de 84,70%. A InceptionV3 apresentou resultados intermediários, não se destacando nem como a melhor, nem como a pior entre os experimentos. A CNNGen, no entanto, obteve os melhores resultados em todas as outras métricas, com acurácia de 86,03%, *f1-score* de 86,77% e *recall* de 90,28%.

Em relação às matrizes de confusão exibidas na Figura 22, a AlexNet apresentou o maior número de erros, com os maiores valores de falsos positivos e falsos negativos, classificando 51 imagens incorretamente como císticas e 35 como não císticas. A CNNGen foi a mais eficaz na classificação de imagens não císticas, com 203 verdadeiros negativos, enquanto a DCNNDan obteve o melhor resultado na classificação de imagens císticas, com 216 verdadeiros positivos.

Tabela 10 – Métricas de avaliação das CNNs sem aplicação de pré-processamento.

CNN	Acurácia	F1-Score	Precisão	Recall
InceptionV3	85,83%	86,33%	84,49%	88,25%
AlexNet	82,34%	86,13%	80,60%	85,82%
DCNNDan	85,62%	86,05%	84,70%	87,44%
CNNGen	86,03%	86,77%	83,52%	90,28%

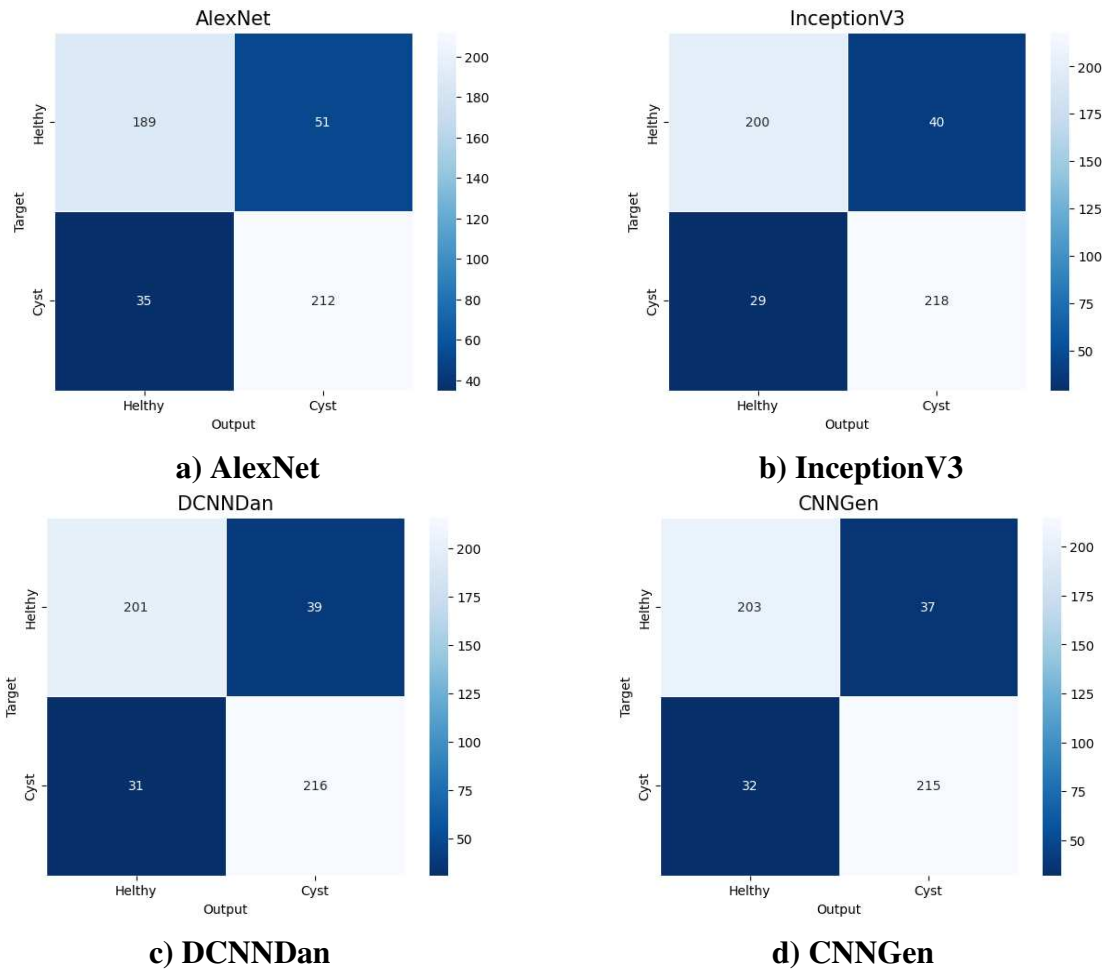
Fonte: Elaborado pela autora.

5.2.2 Resultado com aplicação do Filtro Laplaciano

Os resultados apresentados na Tabela 11 mostram o desempenho das quatro CNNs treinadas utilizando a base de imagens com aplicação do filtro laplaciano. O pior desempenho foi da rede AlexNet, com uma acurácia de 85,21% e precisão de 83,26%. A rede DCNNDan também não obteve resultados tão favoráveis, alcançando o pior *f1-score*, com um valor de 85,65%. A InceptionV3 apresentou resultados intermediários, não se destacando nem como a melhor, nem como a pior entre os experimentos. Por fim, a CNNGen obteve os melhores resultados em todas as métricas de avaliação, com acurácia de 87,06%, *f1-score* de 87,57%, *recall* de 89,87% e precisão de 88,25%.

Em relação às matrizes de confusão, exibidas na Figura 23, a AlexNet apresentou o maior número de erros, com os maiores valores de falsos positivos e falsos negativos, classificando 44 imagens erroneamente como císticas e 28 como não císticas. A CNNGen foi a mais

Figura 22 – Matrizes de Confusão da CNN sem aplicação de pré-processamento.



Fonte: Elaborado pela autora.

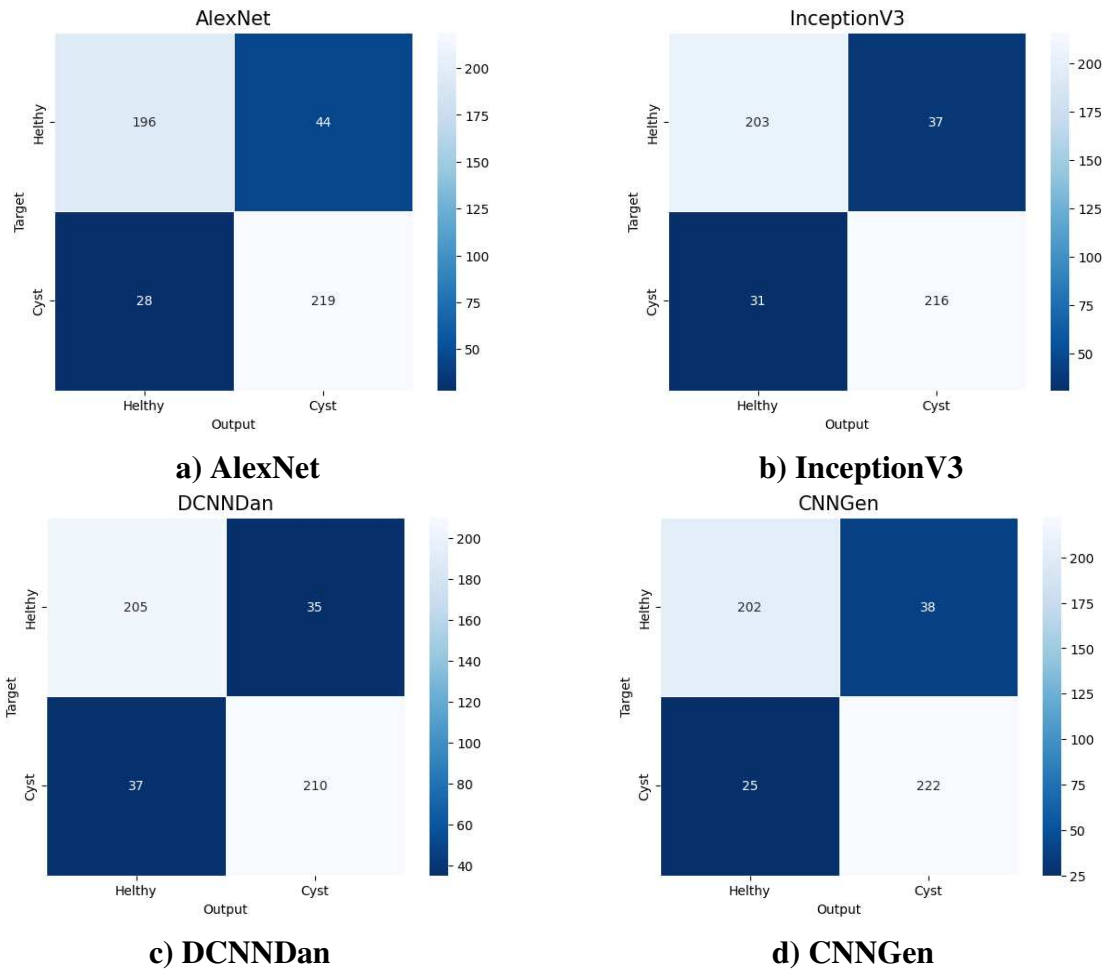
eficaz na classificação de imagens císticas e não císticas, com 202 verdadeiros negativos e 222 verdadeiros positivos.

Tabela 11 – Métricas de avaliação das CNNs com aplicação do Filtro Laplaciano.

CNN	Acurácia	F1-Score	Precisão	Recall
InceptionV3	86,03%	86,4%	85,37%	87,44%
AlexNet	85,21%	85,88%	83,26%	88,66%
DCNNDan	85,42%	85,65%	85,48%	85,82%
CNNGen	87,06%	87,57%	88,25%	89,87%

Fonte: Elaborado pela autora.

Figura 23 – Matrizes de Confusão da CNN com aplicação do Filtro Laplaciano



Fonte: Elaborado pela autora.

5.2.3 Resultado com aplicação do Filtro de Prewitt

Os resultados apresentados na Tabela 12 mostram o desempenho das quatro CNNs treinadas com a aplicação do filtro de Prewitt. A AlexNet apresentou o pior desempenho em termos de acurácia, com valor de 83,16% e precisão, com valor de 78,54%, mas obteve o maior valor de *recall*, com 91,90%. A rede InceptionV3 também apresentou resultados insatisfatórios, registrando o menor *f1-score*, com 84,70%. A DCNNdan destacou-se pela melhor precisão, atingindo 86,55%. Já a CNNGen obteve resultados de classificação mais equilibrados, com acurácia de 87,88% e *f1-score* de 88,31%.

Nas matrizes de confusão apresentadas na Figura 24, a AlexNet apresentou o maior número de falsos positivos, classificando incorretamente 62 imagens como císticas. A InceptionV3, por sua vez, teve o maior número de falsos negativos, com 42 imagens erroneamente classificadas como não císticas. A CNNGen foi a mais eficiente na identificação de imagens não císticas, com 212 verdadeiros negativos, enquanto a AlexNet alcançou o melhor desempenho na

classificação de imagens císticas, com 227 verdadeiros positivos.

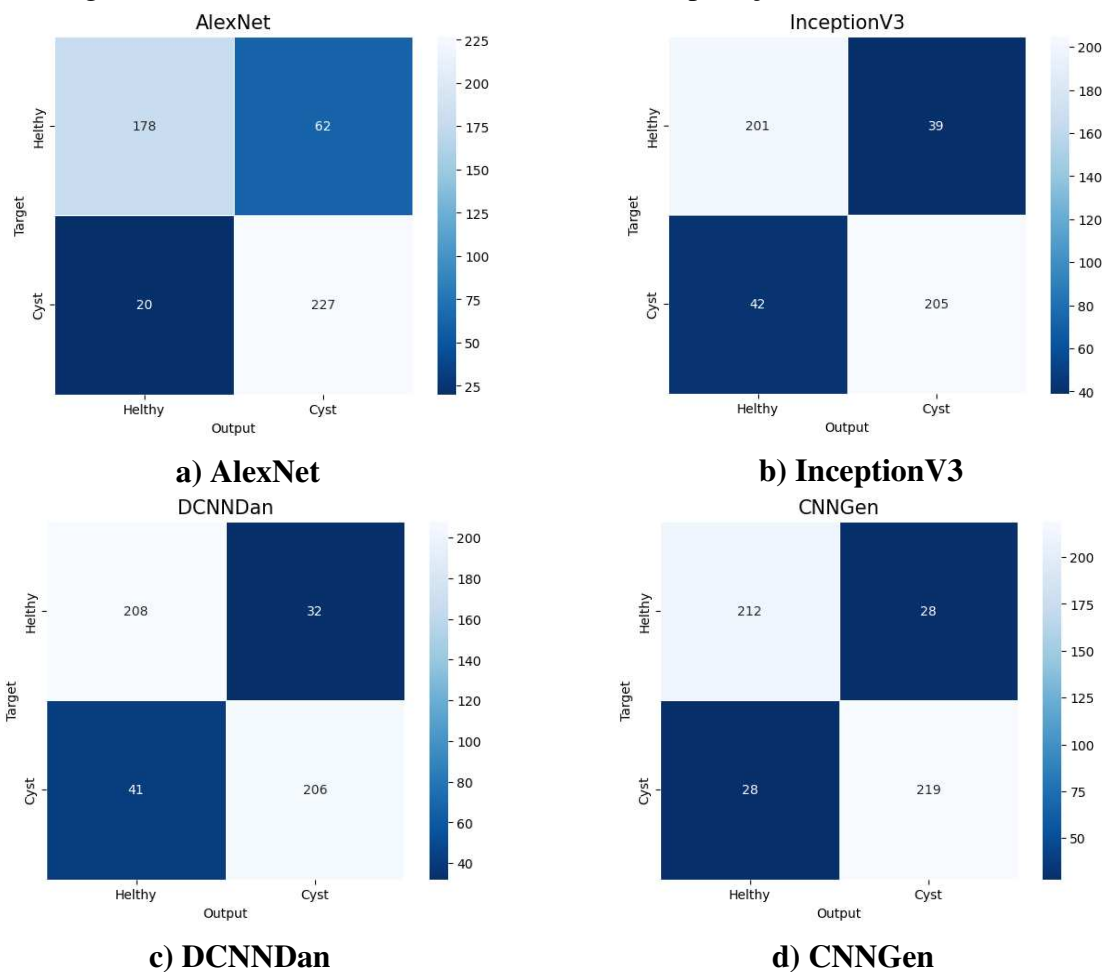
No entanto, apesar da AlexNet ter obtido resultados satisfatórios na classificação de imagens císticas, sua alta precisão combinada com o baixo *recall* sugere um desequilíbrio nas previsões, indicando que a rede tende a classificar um tipo de imagem com mais frequência do que outro, o que pode comprometer a confiança dos resultados desta rede.

Tabela 12 – Métricas de avaliação das CNNs com aplicação do Filtro de Prewitt.

CNN	Acurácia	F1-Score	Precisão	Recall
InceptionV3	83,36%	83,50%	84,01%	82,99%
AlexNet	83,16%	84,70%	78,54%	91,90%
DCNNDan	85,01%	84,94%	86,55%	83,40%
CNNGen	87,88%	88,31%	86,43%	90,28%

Fonte: Elaborado pela autora.

Figura 24 – Matrizes de Confusão da CNN com aplicação do Filtro de Prewitt



Fonte: Elaborado pela autora.

5.2.4 Resultado com aplicação do Filtro de Roberts

Os resultados apresentados na Tabela 13 mostram o desempenho das quatro CNNs treinadas com a aplicação do filtro de Roberts. A AlexNet apresentou o pior desempenho em termos de acurácia, com valor de 82,34%, *f1-score* de 78,54%, e *recall* de 81,37%. A rede InceptionV3 obteve o melhor resultado de *f1-score*, com valor de 87,57%. A DCNNDan obteve resultados insatisfatórios de precisão, com valor de 82,35%. Já a CNNGen obteve resultados de classificação mais equilibrados, com acurácia de 87,06% e *f1-score* de 87,52%.

Nas matrizes de confusão exibidas na Figura 25, a DCNNDan apresentou o maior número de falsos positivos, classificando incorretamente 48 imagens como císticas. A AlexNet, por sua vez, teve o maior número de falsos negativos, com 45 imagens erroneamente classificadas como não císticas. A CNNGen foi a mais eficiente na identificação de imagens não císticas, com 203 verdadeiros negativos, enquanto a InceptionV3 alcançou o melhor desempenho na classificação de imagens císticas, com 229 verdadeiros positivos.

Tabela 13 – Métricas de avaliação das CNNs com aplicação do Filtro de Roberts.

CNN	Acurácia	F1-Score	Precisão	Recall
InceptionV3	86,75%	87,57%	82,97%	92,71%
AlexNet	82,34%	82,37%	83,40%	81,37%
DCNNDan	85,42%	86,31%	82,35%	90,68%
CNNGen	87,06%	87,52%	86,65%	89,47%

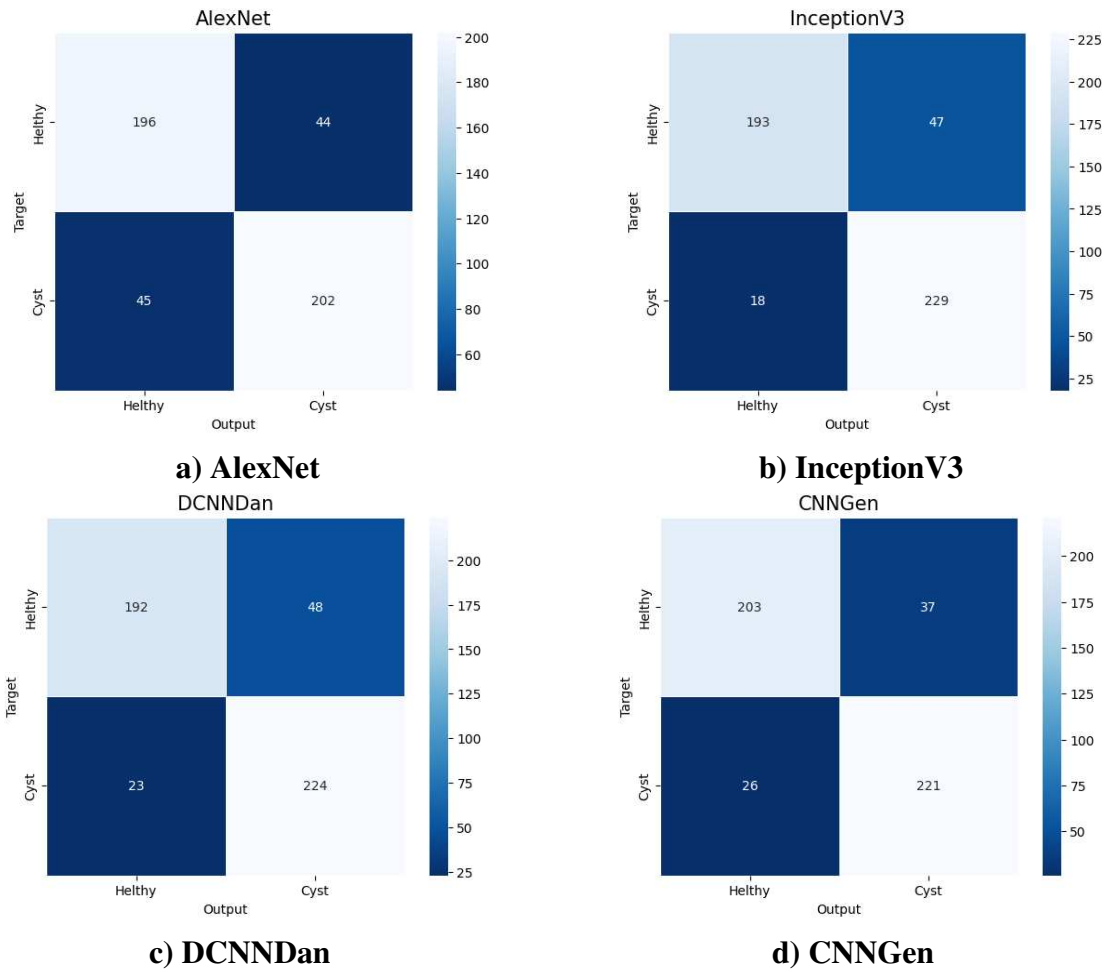
Fonte: Elaborado pela autora.

5.2.5 Resultado com aplicação da morfologia BlackHat

Os resultados apresentados na Tabela 14 mostram o desempenho das quatro CNNs treinadas com a aplicação da morfologia Black-hat. A AlexNet apresentou os piores desempenhos em todas as métricas utilizadas com 84,39% de acurácia, 85,03% de *f1-score*, 82,75% de precisão e 87,44% de *recall*. As redes InceptionV3 e DCNNDan obtiveram resultados intermediários e bem semelhantes, tendo 86,85% e 86,65% de acurácia respectivamente. A rede CNNGen obteve os melhores resultados em todas as métricas de avaliação, com acurácia de 88,50%, *f1-score* de 89,01%, precisão de 86,31% e *recall* de 91,90%.

Em relação às matrizes de confusão apresentadas na Figura 26, a AlexNet apresentou

Figura 25 – Matrizes de Confusão da CNN com aplicação do Filtro de Roberts



Fonte: Elaborado pela autora.

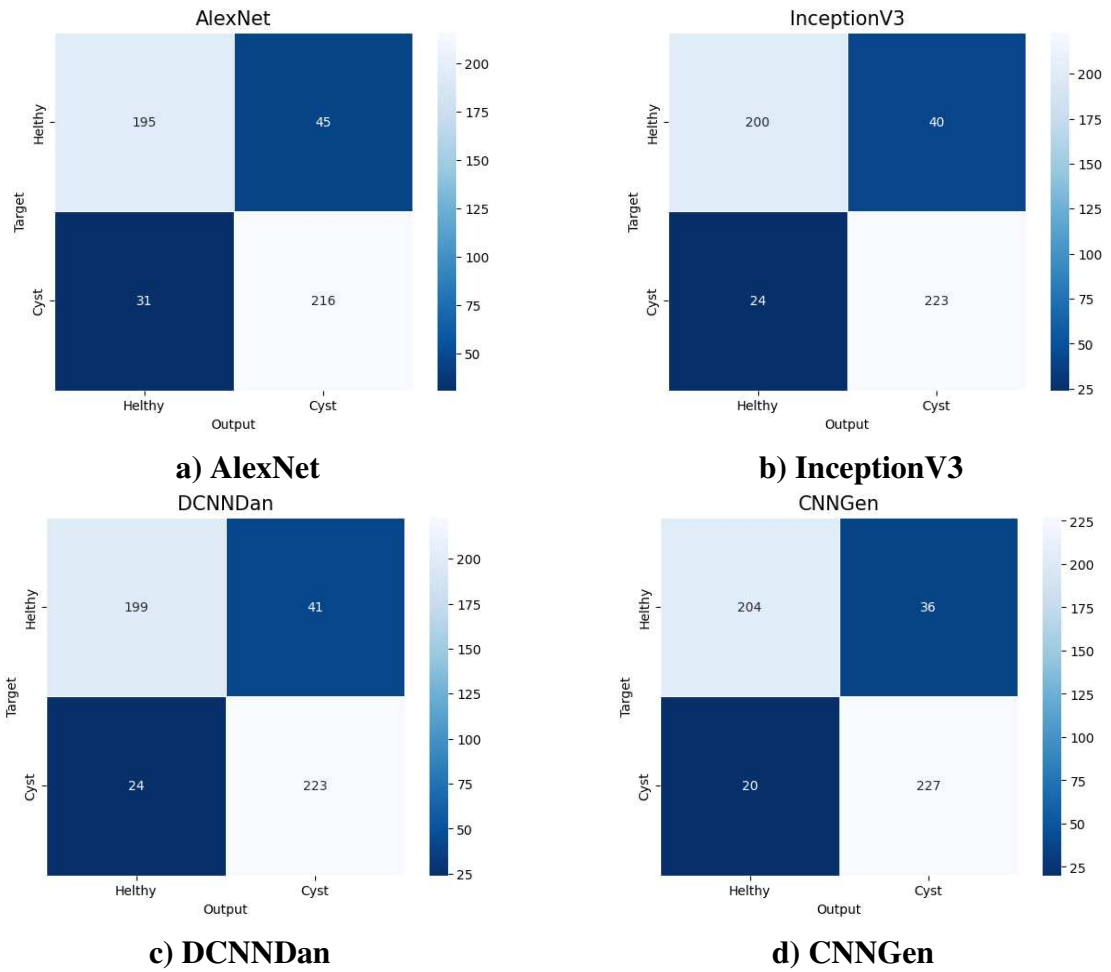
o maior número de erros, com os maiores valores de falsos positivos e falsos negativos, classificando 45 imagens erroneamente como císticas e 31 como não císticas. A CNNGen foi a mais eficaz na classificação de imagens, com 204 verdadeiros negativos e 227 verdadeiros positivos.

Tabela 14 – Métricas de avaliação das CNNs com aplicação da morfologia BlackHat.

CNN	Acurácia	F1-Score	Precisão	Recall
InceptionV3	86,85%	87,45%	84,79%	90,28%
AlexNet	84,39%	85,03%	82,75%	87,44%
DCNNDan	86,65%	87,27%	84,46%	90,28%
CNNGen	88,50%	89,01%	86,31%	91,90%

Fonte: Elaborado pela autora.

Figura 26 – Matrizes de Confusão da CNN com aplicação da morfologia BlackHat



Fonte: Elaborado pela autora.

5.3 Considerações finais dos resultados

A partir dos resultados expostos nas seções anteriores, pode-se realizar algumas conclusões e questionamentos. Sabe-se que o objetivo principal deste trabalho é o desenvolvimento da metodologia mais especializada para classificar regiões císticas. Para isso, foram realizadas uma série de experimentos, utilizando vários tipos de modelos de classificação, várias taxas de aprendizado, vários otimizadores e várias técnicas para realces das imagens.

A partir dos experimentos realizados utilizando as funções otimizadoras, podemos concluir que tanto a taxa de aprendizado, quando a função otimizadora é capaz de interferir no

resultado do modelo, podendo gerar uma classificação boa ou ruim. Logo, esses parâmetros de treinamento para os modelos precisam ser escolhidos com cautela, pois uma escolha ruim, pode interferir negativamente no resultado da classificação. Se esses experimentos não fossem realizados, não teríamos como concluir qual seria o melhor conjunto de otimização para este trabalho. Sendo assim, o conjunto que obteve o melhor resultado para os modelos foi o otimizador Adam utilizado com a taxa de aprendizado de 0,0001.

Os experimentos foram todos realizados utilizando o algoritmo de validação cruzada, isso foi feito para garantir que o modelo realmente estava dando bons resultados. É importante entender que um modelo começa a ser treinado com pesos aleatórios e isso pode interferir no seu resultado de classificação. Por esse motivo, o modelo foi treinado várias vezes utilizando a técnica de k-fold, e a média das acurácias foi calculada com base nos dados de teste. Esse procedimento é capaz de aumentar a confiabilidade dos resultados obtidos.

Com relação às técnicas utilizadas para realce das características principais das imagens. A partir dos resultados obtidos, pode-se concluir que essas técnicas são sim capazes de influenciar o resultado de classificação, podendo melhorar ou piorar o resultado. Nos experimentos realizados pode-se notar que na maioria das redes o Filtro de Prewitt obteve resultados piores do que os resultados obtidos quando a rede foi treinada sem usar nenhuma técnica de realce.

Outro exemplo de filtro que teve resultados ruins foi o Roberts, que obteve apenas 79,92% de média de acurácia na rede InceptionV3 e 75,73% de acurácia na rede AlexNet. No entanto, esse filtro chegou a alcançar o segundo melhor resultado de acurácia média na rede CNNGen, com o valor de 85,52%. Além disso, em todos os casos, a morfologia Blackhat obteve os melhores resultados das métricas de avaliação. Com isso, podemos concluir que dependendo da arquitetura de rede utilizada, a técnica de realce pode influenciar negativamente ou positivamente nos resultados.

Com relação às arquiteturas de redes neurais utilizadas, as rede InceptionV3 e AlexNet obtiveram resultados insatisfatórios e bem intermediários. A AlexNet obteve a pior precisão de todos os testes, tendo 78,54% quando o modelo foi treinado e testado utilizando imagens com o filtro de prewitt aplicado. As redes DCNNDan e CNNGen foram ambas desenvolvidas com o intuito de serem especialistas em classificação de regiões císticas e ambas obtiveram os melhores resultados. Sendo a CNNGen a rede com o melhor resultado no *pipeline* de classificação deste trabalho. Onde utilizando imagens com o filtro Blackhat, obteve 88,50%

de acurácia e 89,01 de f1-score.

Por fim, pode-se concluir que é de suma importância selecionar as redes neurais de maneira extremamente cuidadosa ao desenvolver um trabalho nessa área. Muitas vezes uma arquitetura obtém bons resultados para um tipo de imagem, mas acaba gerando resultados ruins para outros tipos de imagens. Por isso, é fundamental experimentar a maior quantidade de arquiteturas e suas variações de parâmetros, para que assim se tenha maior confiabilidade do modelo que está sendo desenvolvido.

6 CONCLUSÕES

A qualidade de um sistema de visão computacional está diretamente relacionada às técnicas utilizadas no seu desenvolvimento. Visando desenvolver a melhor metodologia para um sistema de visão computacional cujo objetivo é realizar a classificação de regiões císticas de imagens de radiografias panorâmicas odontológicas. Este trabalho realiza uma série de experimentos avaliando várias taxas de aprendizagens, métodos de otimização, técnicas de realce de imagens e arquiteturas de redes neurais.

Sendo assim, os métodos de otimização Adam, Adagrad e RMSProp foram testados com as taxas de aprendizado 0,1, 0,01, 0,001 e 0,0001. Cada método de otimização utilizado apresenta um valor de taxa de aprendizado mais adequado para ser utilizado durante a otimização do modelo. Para este trabalho, o Adam com a taxa de aprendizado de 0,0001 possui o melhor resultado para otimização de todas as redes neurais testadas.

Com relação às técnicas de realce utilizadas, pode-se notar que após o treinamento dos modelos utilizando o filtro de Prewitt e de Roberts, muitas vezes o resultado foi inferior aos modelos treinados sem utilizar nenhuma técnica para realce. O Filtro Laplaciano obteve resultados intermediários nos experimentos realizados. Além disso, em todos os casos, a morfologia Blackhat obteve os melhores resultados das métricas de avaliação, se destacando como a melhor técnica para realçar as características das imagens císticas utilizadas nesta pesquisa. Sendo assim, pode-se concluir que as técnicas de realce são capazes de influenciar o resultado da classificação, podendo tanto aprimorar quanto prejudicar o desempenho do modelo.

Em relação às arquiteturas de CNNs utilizadas, a AlexNet tem a pior precisão de todos os experimentos realizados, com o valor de 78,54% quando o modelo foi treinado e testado utilizando imagens com o filtro de Prewitt aplicado. A rede InceptionV3 apresenta resultados de classificação intermediários, não sendo nem os melhores e nem os piores resultados deste trabalho. As redes DCNNDan e CNNGen foram criadas visando se especializarem na classificação de regiões císticas, e ambas alcançaram os melhores resultados nos experimentos desenvolvidos.

Outrossim, a rede CNNGen desenvolvida neste trabalho mediante uma série de experimentos, se destaca como a rede que possui os melhores resultados de classificação utilizando todas as técnicas de realce de imagens. Em especial, o melhor resultado é gerado utilizando imagens com o filtro Blackhat, com 88,50% de acurácia e 89,01% de f1-score.

Sendo assim, esta pesquisa evidencia a importância do desenvolvimento de expe-

rimentos para detectar a melhor metodologia para um sistema de visão computacional. Fica claro que as técnicas de pré-processamento de imagens, os métodos de otimização, as taxas de aprendizado e as arquiteturas utilizadas podem influenciar no resultado de classificação da rede. Se tratando de dados médicos, é de suma importância que a rede possua a maior confiabilidade possível, pois uma detecção errônea pode acarretar sérios problemas para o paciente.

6.1 Trabalhos Futuros

Para trabalhos futuros pretende-se experimentar outras técnicas de realce de imagens, além de testar outras funções de otimização como a Nadam, Adamax e a Adadelta. Pretende-se também fazer novos testes utilizando as técnicas de *GridSearch* para detectar outras arquiteturas de CNNs que também tenham bons resultados de classificação de imagens císticas odontológicas.

Além disso, pretende-se explorar técnicas de *autoencoders*, uma vez que são capazes de aprender de forma mais eficiente as representações dos dados, melhorando o pré-processamento, atuando na remoção eficiente de ruídos e redução da dimensionalidade das imagens. Pretende-se também estudar a aplicação de modelos de difusão em imagens de radiografias dentárias, a fim de melhorar a qualidade das imagens utilizadas, ou até gerar uma nova base para ser utilizada na etapa de aumento de dados.

REFERÊNCIAS

- ABDURRAZZAQ, A.; JUNOH, A. K.; MUHAMAD, W. Z. A. W.; YAHYA, Z.; MOHD, I. An overview of multi-filters for eliminating impulse noise for digital images. **TELKOMNIKA (Telecommunication Computing Electronics and Control)**, v. 18, n. 1, p. 385–393, 2020.
- ALOM, M. Z.; TAHA, T. M.; YAKOPCIC, C.; WESTBERG, S.; SIDIKE, P.; NASRIN, M. S.; ESESN, B. C. V.; AWWAL, A. A. S.; ASARI, V. K. The history began from alexnet: A comprehensive survey on deep learning approaches. **arXiv preprint arXiv:1803.01164**, 2018.
- ALVARES, L. C.; EURICO, E.; FREITAS, J. A. d. S. Manual de técnica radiográfica: radiografia intra-orais: periapical, interproximal e oclusal. 1965.
- AVILA, R.; KITANI, E.; ZAMPIROLI, F. D. A.; YOSHIOKA, L.; CELIBERTO, L. A.; IBUSUKI, U. Revisão sistemática de visão computacional para veículos autônomos agrícolas. In: IEEE. **2023 15th IEEE International Conference on Industry Applications (INDUSCON)**. [S. l.], 2023. p. 485–492.
- BALAN, W. C.; INSTRUCIONAL, A. A imagem digital. **Apostila, UNESP, São Paulo, SP, Brasil, 14p**, 2009.
- BALLARD, D. H.; BROWN, C. M. **Computer vision**. [S. l.]: Prentice Hall Professional Technical Reference, 1982.
- BARELLI, F. **Introdução à visão computacional: Uma abordagem prática com Python e OpenCV**. [S. l.]: Editora Casa do Código, 2018.
- BHAIRANNAWAR, S. S. Efficient medical image enhancement technique using transform hsv space and adaptive histogram equalization. In: **Soft Computing Based Medical Image Analysis**. [S. l.]: Elsevier, 2018. p. 51–60.
- CARRARA, V. Introdução à robótica industrial. **São José dos Campos: Instituto Nacional de Pesquisas Espaciais-Inpe**, 2015.
- CELIK, M. E. Deep learning based detection tool for impacted mandibular third molar teeth. **Diagnostics**, MDPI, v. 12, n. 4, p. 942, 2022.
- CHOLLET, F. **Deep learning with Python**. [S. l.]: Simon and Schuster, 2021.
- COSTA, F. R.; ESTEVES, C.; BACELAR, M. T. Lesões benignas da mandíbula: uma revisão pictórica. **Acta Radiológica Portuguesa**, v. 28, n. 108, p. 25–36, 2016.
- DUCHI, J.; HAZAN, E.; SINGER, Y. Adaptive subgradient methods for online learning and stochastic optimization. **Journal of machine learning research**, v. 12, n. 7, 2011.
- DUONG-TRUNG, N.; QUACH, L.-D.; NGUYEN, C.-N. Learning deep transferability for several agricultural classification problems. **International Journal of Advanced Computer Science and Applications**, Science and Information (SAI) Organization Limited, v. 10, n. 1, 2019.
- EL-NAGGAR, A. K.; CHAN, J. K. C.; GRANDIS, J. R.; TAKATA, T.; SLOOTWEG, P. J. **WHO Classification of Head and Neck Tumours**. 4. ed. [S. l.]: World Health Organization, 2017.

FEJERSKOV, O.; URIBE, S.; MARIÑO, R. J. Dentistry in a historical perspective and a likely future of the profession. **Career paths in oral health**, Springer, p. 3–19, 2018.

FERLITSCH, A. **Deep Learning Patterns and Practices**. [S. l.]: Simon and Schuster, 2021.

FILHO, O. M.; NETO, H. V. **Processamento digital de imagens**. [S. l.]: Brasport, 1999.

FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: **Competition and cooperation in neural nets**. [S. l.]: Springer, 1982. p. 267–285.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems**. [S. l.]: "O'Reilly Media, Inc.", 2019.

GÉRON, A. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. [S. l.]: "O'Reilly Media, Inc.", 2022.

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

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagem**. **Pearson, ISBN-10: 8576054019**, v. 10, p. 11–27, 2010.

GUPTA, G. *et al.* Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter. **International Journal of Soft Computing and Engineering (IJSCE)**, v. 1, n. 5, p. 304–311, 2011.

HA, E.-G.; JEON, K. J.; CHOI, H.; LEE, C.; CHOI, Y. J.; HAN, S.-S. Automatic diagnosis of retention pseudocyst in the maxillary sinus on panoramic radiographs using a convolutional neural network algorithm. **Scientific Reports**, Nature Publishing Group UK London, v. 13, n. 1, p. 2734, 2023.

HAMAMCI, I. E.; ER, S.; SIMSAR, E.; SEKUBOYINA, A.; GUNDOGAR, M.; STADLINGER, B.; MEHL, A.; MENZE, B. Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. [S. l.], 2023. p. 389–399.

JÄHNE, B. **Digital image processing**. [S. l.]: Springer Science & Business Media, 2005.

JUNIOR, F. V. L. C. O uso de aprendizado profundo na classificação de ressonâncias magnéticas para detecção de tumor cerebral. 2023.

KAPILA, R.; DHALIWAL, A.; SINGH, N.; BV, S. History of x-rays in dentistry. **Annals of Dental Research**, v. 2, n. 1, p. 22–5, 2013.

KARSAZ, A. Diabetic retinopathy screening using improved support vector domain description: a clinical study. **Soft Computing**, Springer, v. 26, n. 19, p. 10085–10101, 2022.

KAUR, D.; KAUR, Y. Various image segmentation techniques: a review. **International Journal of Computer Science and Mobile Computing**, v. 3, n. 5, p. 809–814, 2014.

KIM, S.; CASPER, R. Applications of convolution in image processing with matlab. **University of Washington**, p. 1–20, 2013.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KUMAR, V. S.; KUMAR, P. R.; YADALAM, P. K.; ANEGUNDI, R. V.; SHRIVASTAVA, D.; ALFURHUD, A. A.; ALMAKTOOM, I. T.; ALFTAIKHAH, S. A. A.; ALSHARARI, A. H. L.; SRIVASTAVA, K. C. Machine learning in the detection of dental cyst, tumor, and abscess lesions. **BMC Oral Health**, Springer, v. 23, n. 1, p. 833, 2023.

KWON, O.; YONG, T.-H.; KANG, S.-R.; KIM, J.-E.; HUH, K.-H.; HEO, M.-S.; LEE, S.-S.; CHOI, S.-C.; YI, W.-J. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. **Dentomaxillofacial Radiology**, Oxford University Press, v. 49, n. 8, p. 20200185, 2020.

LAKHANI, K.; MINOCHA, B.; GUGNANI, N. Analyzing edge detection techniques for feature extraction in dental radiographs. **Perspectives in Science**, Elsevier, v. 8, p. 395–398, 2016.

LI, S.; FEVENS, T.; KRZYŻAK, A.; JIN, C.; LI, S. Semi-automatic computer aided lesion detection in dental x-rays using variational level set. **Pattern Recognition**, Elsevier, v. 40, n. 10, p. 2861–2873, 2007.

LIMA, F. D. S. *et al.* Segmentação de lesões periapicais em imagens odontológicas radiográficas digitais utilizando modelos de contorno ativo com inicialização automática. Universidade Federal do Ceará, 2018.

LOPES, H. P.; JUNIOR, J. F. S. Endodontia: biologia e técnica. In: **Endodontia: biologia e técnica**. [S. l.: s. n.], 1999. p. 650–650.

LOUREDO, B. V. R.; FREITAS, C. T. S. de; CÂMARA, J.; LIBÓRIO-KIMURA, T. N. Estudo epidemiológico de lesões odontogênicas provenientes do departamento de patologia e medicina legal da universidade federal do Amazonas. **Revista Brasileira de Odontologia**, v. 74, n. 2, p. 126, 2017.

MAKANDAR, A.; HALALLI, B. Image enhancement techniques using highpass and lowpass filters. **International Journal of Computer Applications**, Citeseer, v. 109, n. 14, p. 12–15, 2015.

MALLYA, S.; LAM, E. **White and Pharoah's oral radiology: principles and interpretation**. [S. l.]: Elsevier Health Sciences, 2018.

MICHALSKI, R. S.; CARBONELL, J. G.; LEARNING, T. An artificial intelligence approach. **Understanding the Nature of Learning**, Springer, v. 2, p. 3–26, 1983.

MIKULKA, J.; KABRDA, M.; GESCHIEDTOVÁ, E.; PEŘINA, V. Classification of jawbone cysts via orthopantomogram processing. In: IEEE. **2012 35th International Conference on Telecommunications and Signal Processing (TSP)**. [S. l.], 2012. p. 499–502.

NEVES, L. A. P.; NETO, H. V.; GONZAGA, A. Avanços em visão computacional. **Omnipax, Curitiba, PR**, v. 1, 2012.

NEVILLE, B. W. *et al.* **Patologia Oral e Maxilofacial**. [SI]. [S. l.]: Elsevier, 2009.

NURTANIO, I.; PURNAMA, I. K. E.; HARIADI, M.; PURNOMO, M. H. Cyst and tumor lesion segmentation on dental panoramic images using active contour models. **IPTEK The Journal for Technology and Science**, v. 22, n. 3, 2011.

OLIVEIRA, D.; BARRETO, J.; MESQUITA, I.; JR, I. P.; CHAVES, F.; SAMPIERI, M.; MADEIRO, J. Analysis of the influence of pre-processing techniques with convolutional neural networks for automatic detection of cysts in wisdom teeth. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis, p. 1–12, 2022.

OLIVEIRA, D.; BARRETO, J.; MESQUITA, I.; JR, I. P.; CHAVES, F.; SAMPIERI, M.; MADEIRO, J. Analysis of the influence of pre-processing techniques with convolutional neural networks for automatic detection of cysts in wisdom teeth. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization**, Taylor & Francis, v. 11, n. 3, p. 299–310, 2023.

OLIVEIRA, D. A.; SOUSA, L. M. R. d.; SAMPIERI, M. B. d. S.; CHAVES, F. N.; JÚNIOR, I. C. d. P. Identificação de regiões císticas dentárias em imagens de radiografia panorâmica utilizando algoritmos de segmentação e classificação por textura. **Sociedade Brasileira de Automática (SBA)**, 2019.

PANETTA, K.; RAJENDRAN, R.; RAMESH, A.; RAO, S. P.; AGAIAN, S. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. **IEEE journal of biomedical and health informatics**, IEEE, v. 26, n. 4, p. 1650–1659, 2021.

PREWITT, J. M. *et al.* Object enhancement and extraction. **Picture processing and Psychopictorics**, Academic Press New York, v. 10, n. 1, p. 15–19, 1970.

QUEIROZ, J. E. R. de; GOMES, H. M. Introdução ao processamento digital de imagens. **Rita**, v. 13, n. 2, p. 11–42, 2006.

RIOS, M.; NETTO, J. F. Uma abordagem utilizando visão computacional para monitoramento de robôs móveis em ambientes de tarefas na robótica educacional. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S. l.: s. n.], 2016. v. 27, n. 1, p. 480.

ROBERTS, L. G. **Machine perception of three-dimensional solids**. Tese (Doutorado) – Massachusetts Institute of Technology, 1963.

RUDER, S. An overview of gradient descent optimization algorithms. **arXiv preprint arXiv:1609.04747**, 2016.

SALTORATTO, A. Y. K.; SILVA, F. A. da; CAMARGO, A. C. A. C.; SILVA, P. C. G. da; SOUZA, L. F. A. de. Monitoramento de avicultura a partir de técnicas de visão computacional. In: **Colloquium Exactarum. ISSN: 2178-8332**. [S. l.: s. n.], 2013. v. 5, n. 2, p. 47–66.

SANTOS, T. d. S.; ANTUNES, A. A.; AVELAR, R. L.; ANTUNES, A. P. Cistos odontogênicos: estudo epidemiológico de 72 casos. **Rev. bras. cir. cabeça pescoço**, p. 30–32, 2007.

SAXENA, C.; KOURAV, D. Noises and image denoising techniques: A brief survey. **International journal of Emerging Technology and advanced Engineering**, v. 4, n. 3, p. 878–885, 2014.

SILVA, A. M. M. da; PATROCÍNIO, A. C.; SCHIABEL, H. Processamento e análise de imagens médicas. **REVISTA BRASILEIRA DE FÍSICA MÉDICA (ONLINE)**, 2019.

SIVASUNDARAM, S.; PANDIAN, C. Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture. **International Journal of Imaging Systems and Technology**, v. 31, n. 4, p. 2214–2225, 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ima.22625>.

SIVASUNDARAM, S.; PANDIAN, C. Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture. **International Journal of Imaging Systems and Technology**, Wiley Online Library, v. 31, n. 4, p. 2214–2225, 2021.

SOUSA, E. L. R. de; FERRAZ, C. C. R.; GOMES, B. P. F. de A.; PINHEIRO, E. T.; TEIXEIRA, F. B.; SOUZA-FILHO, F. J. de. Bacteriological study of root canals associated with periapical abscesses. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, Elsevier, v. 96, n. 3, p. 332–339, 2003.

STEVENS, E.; ANTIGA, L.; VIEHMANN, T. **Deep learning with PyTorch**. [S. l.]: Manning Publications, 2020.

SUN, M.; SONG, Z.; JIANG, X.; PAN, J.; PANG, Y. Learning pooling for convolutional neural network. **Neurocomputing**, Elsevier, v. 224, p. 96–104, 2017.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S. l.: s. n.], 2016. p. 2818–2826.

TAJIMA, S.; OKAMOTO, Y.; KOBAYASHI, T.; KIWAKI, M.; SONODA, C.; TOMIE, K.; SAITO, H.; ISHIKAWA, Y.; TAKAYOSHI, S. Development of an automatic detection model using artificial intelligence for the detection of cyst-like radiolucent lesions of the jaws on panoramic radiographs with small training datasets. **Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology**, Elsevier, v. 34, n. 5, p. 553–560, 2022.

TETILA, E. C.; MORAES, P. M. de; CONSTANTINO, M.; COSTA, R. B. da; AYRES, F. M.; REYNALDO, G. O.; COLMAN, N. A.; MACHADO, F. C. A. P.; SOARES, K. G.; GRECO, M. M. D. M. *et al.* Classificação e monitoramento de áreas urbanizadas usando técnicas de visão computacional. **Desenvolvimento e Meio Ambiente**, v. 61, 2023.

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: SN. **Proceedings of the xxix conference on graphics, patterns and images**. [S. l.], 2016. v. 1, n. 4.

VARINAUSKAS, V.; GERVICKAS, A.; KAVOLIUNIENE, O. Analysis of odontogenic cysts of the jaws. **Medicina (Kaunas)**, v. 42, n. 3, p. 201–7, 2006.

VAZ, G. C. **Aplicação de modelos de deep learning para qualificação da área da engenharia biomédica: um estudo de caso de visão computacional em imagens de raio-x da região torácica**. Tese (Doutorado) – [sn], 2023.

WHITE, S. C.; PHAROAH, M. J. **Oral radiology: principles and interpretation**. [S. l.]: Elsevier Health Sciences, 2013.