
**Divergências de Bregman Aplicadas na
Recuperação de Imagens por Conteúdo em
Displasias Orais**

Tiago Rosa Marques Soares



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia
2024

Tiago Rosa Marques Soares

**Divergências de Bregman Aplicadas na
Recuperação de Imagens por Conteúdo em
Displasias Orais**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência de Dados

Orientador: Prof. Humberto Luiz Razente

Coorientador: Prof. Marcelo Zanchetta do Nascimento

Uberlândia

2024

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S676d
2024 Soares, Tiago Rosa Marques, 1996-
Divergências de Bregman aplicadas na recuperação de imagens por conteúdo em displasias orais [recurso eletrônico] / Tiago Rosa Marques Soares. - 2024.

Orientador: Humberto Luiz Razente.

Coorientador: Marcelo Zanchetta do Nascimento.

Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-graduação em Ciência da Computação.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.di.2024.5124>

Inclui bibliografia.

Inclui ilustrações.

1. Computação. I. Razente, Humberto Luiz, 1977-, (Orient.). II. Nascimento, Marcelo Zanchetta do, 1976-, (Coorient.). III. Universidade Federal de Uberlândia. Programa de Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

André Carlos Francisco
Bibliotecário Documentalista - CRB-6/3408



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Ciência da Computação

Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902

Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Dissertação de Mestrado, 20/2024, PPGCO				
Data:	31 de julho de 2024	Hora de início:	14:05	Hora de encerramento:	16:40
Matrícula do Discente:	12122CCP016				
Nome do Discente:	Tiago Rosa Marques Soares				
Título do Trabalho:	Divergências de Bregman Aplicadas na Recuperação de Imagens por Conteúdo em Displasias Orais				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	CNPq projeto número 311404/2021-9				

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Marcelo Zanchetta do Nascimento - FACOM/UFU (Coorientador), Bruno Augusto Nassif Travençolo - FACOM/UFU, Joaquim Cezar Felipe - FFCLRP/USP e Humberto Luiz Razente - FACOM/UFU, orientador do candidato.

Os examinadores participaram desde as seguintes localidades: Joaquim Cezar Felipe - Ribeirão Preto/SP . Os outros membros da banca e o aluno participaram da cidade de Uberlândia.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Humberto Luiz Razente, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir ao candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos,

conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Humberto Luiz Razente, Professor(a) do Magistério Superior**, em 01/08/2024, às 09:29, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo, Professor(a) do Magistério Superior**, em 01/08/2024, às 10:04, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 01/08/2024, às 13:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Joaquim Cezar Felipe, Usuário Externo**, em 14/08/2024, às 16:31, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5534783** e o código CRC **B6B045F6**.

*Agradeço a Deus por sempre me amparar,
aos meus pais que me guiaram pelo caminho correto
e à minha esposa pelo apoio incondicional.*

Agradecimentos

Primeiramente, agradeço a Deus por sempre me amparar e me dar força ao longo desta jornada. Sua presença constante me guiou e me deu a coragem necessária para enfrentar todos os desafios.

Aos meus pais, Gesse James e Maria Soares, expresso minha eterna gratidão. Vocês me proporcionaram uma base sólida de valores e ética, e sempre me apoiaram em cada decisão. Sem o amor, a paciência e o encorajamento de vocês, esta conquista não seria possível. Obrigado por acreditarem em mim e me guiarem pelo caminho correto.

À minha esposa, Ana Clara, meu mais profundo agradecimento. Seu apoio incondicional, sua paciência infinita e seu amor constante foram fundamentais para que eu pudesse chegar até aqui. Você esteve ao meu lado em todos os momentos, compartilhando tanto as alegrias quanto as dificuldades. Sou eternamente grato por ter você ao meu lado.

Gostaria de expressar minha sincera gratidão aos meus professores, Humberto Luiz Razente e Marcelo Zanchetta do Nascimento. Sua orientação, conhecimento e apoio foram essenciais para o desenvolvimento deste trabalho. Agradeço pelas inúmeras horas dedicadas, pelas valiosas sugestões e pela paciência em compartilhar seu vasto conhecimento. Vocês foram mais que mentores, foram verdadeiros exemplos a serem seguidos.

Agradeço também a todos os meus amigos e colegas que de alguma forma contribuíram para a realização deste trabalho. As discussões, o apoio emocional e as palavras de incentivo foram fundamentais para que eu pudesse seguir em frente. Muito obrigado a todos!

Epígrafe

*"Ebenézer! Até aqui nos ajudou o Senhor."
– 1 Samuel 7:12*

Resumo

Os sistemas CBIR têm como objetivo recuperar imagens com base em seu conteúdo visual, utilizando algoritmos que descrevem características visuais em forma de recursos numéricos. Esses recursos permitem encontrar e exibir imagens semelhantes, buscando replicar ou superar o julgamento humano sobre a similaridade. O desafio de lidar com dados não estruturados em grandes conjuntos de imagens cria oportunidades para desenvolver técnicas que identifiquem similaridades ou dissimilaridades entre imagens, dependendo do método de comparação aplicado. A maior parte do trabalho dedicado ao desenvolvimento de estruturas e algoritmos que aplicam medidas de similaridades utiliza espaços métricos, isto porque as medidas métricas satisfazem as propriedades básicas do espaço métrico. Para casos que não seguem estas propriedades, utilizar medidas de similaridade não métricas contribui para a solução de problemas em tipos de dados mais complexos. A Divergência de Bregman permite o desenvolvimento de métodos eficazes e aplicáveis que utilizam um conjunto de funções de desigualdade, em vez de se basearem em um único objetivo. Isso pode ser bem-sucedido quando os recursos métricos não correspondem ao julgamento humano. Obter resultados superiores reside na habilidade de escolher a medida que melhor se adapta ao problema em questão, sendo possível utilizar generalizações como por exemplo: Mahalanobis e Kullback-Leibler (KL). Neste estudo, o objetivo é investigar a similaridade ou dissimilaridade de imagens relacionadas ao câncer da cavidade oral, onde características foram extraídas a partir de lâminas que continham lesões induzidas experimentalmente na língua de camundongos da linhagem C57Bl/6, utilizando o carcinógeno 4NQO. A utilização de imagens médicas e recuperação de imagens desempenha um papel crucial no diagnóstico das displasias, especialmente em campos como radiologia, patologia e medicina. Para o estudo, foi realizada a comparação entre distâncias métricas e Divergências de Bregman (KL e Mahalanobis), utilizando vetores de características morfológicas e não morfológicas. Os resultados do sistema CBIR utilizando diferentes funções de distância, foram avaliadas pelas medidas de Precisão e Revocação. Através dos resultados, as Divergências de Bregman, principalmente a KL, demonstram melhor desempenho comparado com as distâncias métricas na maioria dos casos testados.

Palavras-chave: Kullback-Leibler, Divergência de Bregman, Displasia, Processamento Digital de Imagem, Recuperação de Imagens.

Abstract

CBIR systems aim perform image retrieval based on their visual content, using algorithms that describe visual characteristics in the form of numerical features. These features allow finding and displaying similar images, seeking to replicate or surpass human judgment about similarity. The challenge of dealing with unstructured data in large sets of images creates opportunities to develop techniques that identify similarities or dissimilarities between images, depending on the comparison method applied. Most of the work dedicated to the development of structures and algorithms that apply similarity measures uses metric spaces, because metric measures satisfy the basic properties of the metric space. For cases that do not follow these properties, using non-metric similarity measures contributes to the solution of problems in more complex types of data. Bregman Divergence allows the development of effective and applicable methods that use a set of inequality functions, instead of relying on a single objective. This can be successful when metric features do not match human judgment. Obtaining superior results lies in the ability to choose the measure that best fits the problem in question, being possible to use generalizations such as: Mahalanobis and Kullback-Leibler (KL). In this study, the objective is to investigate the similarity or dissimilarity of images related to oral cavity cancer, where features were extracted from slides containing experimentally induced lesions in the tongue of mice of the C57Bl/6 strain, using the carcinogen 4NQO. The use of medical images and image retrieval plays a crucial role in the diagnosis of dysplasias, especially in fields such as radiology, pathology and medicine. For the study, a comparison between metric distances and Bregman Divergences (KL and Mahalanobis) was performed, using vectors of morphological and non-morphological features. The results of the CBIR system using different distance functions were evaluated by the measures of Precision and Recall. Through the results, Bregman Divergences, especially KL, demonstrate better performance compared to metric distances in most of the tested cases.

Palavras-chave: Kullback-Leibler, Bregman Divergence, Dysplasia, Digital Image Processing, Image Retrieval.

Lista de ilustrações

Figura 1	– Ilustração de amostras histológicas da cavidade bucal: (a) tecido saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa (SILVA et al., 2022a).	40
Figura 2	– Uma ilustração de um neurônio biológico em Redes Neurais Artificiais (RNA).	41
Figura 3	– Modelo matemático de um neurônio. (FERNEDA, 2006)	41
Figura 4	– Ilustração simplificada de uma rede neural artificial. (FERNEDA, 2006)	42
Figura 5	– Exemplo de uma operação de convolução: (a) uma imagem de tamanho 7×5 , (b) um filtro de convolução de tamanho 3×3 e (c) uma matriz de características resultante. (SANTOS et al., 2021; SILVA et al., 2020)	43
Figura 6	– Representação gráfica da operação de <i>max-pooling</i> em uma matriz com dimensões de 4×4	44
Figura 7	– Representação gráfica da operação de <i>average-pooling</i> em uma matriz com dimensões de 4×4	44
Figura 8	– Unidade Residual do Modelo ResNet. (HE et al., 2015)	47
Figura 9	– Diagrama representando a utilização da arquitetura da rede neural ResNet50 na fase de segmentação dos núcleos do tecido histológico (SILVA et al., 2022a).	47
Figura 10	– Estrutura da rede Mask R-CNN para classificação dos dados coletados (SILVA et al., 2022a).	49
Figura 11	– Demonstração das caixas delimitadoras criadas para identificar os núcleos de um tecido, utilizando <i>Mask R-CNN</i> (SILVA et al., 2022a).	50
Figura 12	– Processo de segmentação: (a) Operação de dilatação binária, (b) Operação de preenchimento de regiões, (c) Operação de erosão binária. (SILVA et al., 2022a)	51
Figura 13	– Arquitetura do modelo AlexNet utilizado para extrair características de imagens de tecidos. (ALOM et al., 2018)	55
Figura 14	– Fluxo de funcionamento de um sistema CBIR	56
Figura 15	– Representação de um Conjunto de Documentos na Recuperação de Informação. (PEDROSA et al., 2011)	59
Figura 16	– Representação de um gráfico que ilustra a relação entre precisão e revocação. (ROCHA, 2015)	60

Figura 17 – Fluxograma dos principais estágios empregados para recuperação de imagem por conteúdo em imagens histológicas.	61
Figura 18 – Representação da similaridade humana (SONG et al., 2022).	62
Figura 19 – Representação geométrica da Divergência de Bregman no \mathbb{R}^2 (FERREIRA; BACKES; BARCELOS, 2015).	62
Figura 20 – Análise dos modelos em relação as métricas precisão e revogação (AlexNet) .	69
Figura 21 – Análise de precisão e revogação para ResNet50 com diferentes níveis de displasia	73
Figura 22 – Análise de precisão e revogação para ResNet50 + AlexNet com diferentes níveis de displasia	77
Figura 23 – Análise de precisão e revogação para Engenharia de Características com diferentes níveis de displasia	79
Figura 24 – Análise de precisão e revogação para Engenharia de Características + AlexNet com diferentes níveis de displasia	83
Figura 25 – Análise de precisão e revogação para Engenharia de Características + ResNet50 com diferentes níveis de displasia	86
Figura 26 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet com diferentes níveis de displasia	89
Figura 27 – Análise de precisão e revogação para AlexNet (50+) com diferentes níveis de displasia	101
Figura 28 – Análise de precisão e revogação para AlexNet (40+) com diferentes níveis de displasia	102
Figura 29 – Análise de precisão e revogação para AlexNet (30+) com diferentes níveis de displasia	103
Figura 30 – Análise de precisão e revogação para AlexNet (20+) com diferentes níveis de displasia	104
Figura 31 – Análise de precisão e revogação para AlexNet (10+) com diferentes níveis de displasia	105
Figura 32 – Análise de precisão e revogação para ResNet50 (50+) com diferentes níveis de displasia	106
Figura 33 – Análise de precisão e revogação para ResNet50 (40+) com diferentes níveis de displasia	107
Figura 34 – Análise de precisão e revogação para ResNet50 (30+) com diferentes níveis de displasia	108
Figura 35 – Análise de precisão e revogação para ResNet50 (20+) com diferentes níveis de displasia	109
Figura 36 – Análise de precisão e revogação para ResNet50 (10+) com diferentes níveis de displasia	110
Figura 37 – Análise de precisão e revogação para ResNet50 + AlexNet (50+) com diferentes níveis de displasia	111
Figura 38 – Análise de precisão e revogação para ResNet50 + AlexNet (40+) com diferentes níveis de displasia	112

Figura 39 – Análise de precisão e revogação para ResNet50 + AlexNet (30+) com diferentes níveis de displasia	113
Figura 40 – Análise de precisão e revogação para ResNet50 + AlexNet (20+) com diferentes níveis de displasia	114
Figura 41 – Análise de precisão e revogação para ResNet50 + AlexNet (10+) com diferentes níveis de displasia	115
Figura 42 – Análise de precisão e revogação para Engenharia de Características (50+) com diferentes níveis de displasia	116
Figura 43 – Análise de precisão e revogação para Engenharia de Características + AlexNet (50+) com diferentes níveis de displasia	117
Figura 44 – Análise de precisão e revogação para Engenharia de Características + AlexNet (40+) com diferentes níveis de displasia	118
Figura 45 – Análise de precisão e revogação para Engenharia de Características + AlexNet (30+) com diferentes níveis de displasia	119
Figura 46 – Análise de precisão e revogação para Engenharia de Características + AlexNet (20+) com diferentes níveis de displasia	120
Figura 47 – Análise de precisão e revogação para Engenharia de Características + AlexNet (10+) com diferentes níveis de displasia	121
Figura 48 – Análise de precisão e revogação para Engenharia de Características + ResNet50 (50+) com diferentes níveis de displasia	122
Figura 49 – Análise de precisão e revogação para Engenharia de Características + ResNet50 (40+) com diferentes níveis de displasia	123
Figura 50 – Análise de precisão e revogação para Engenharia de Características + ResNet50 (30+) com diferentes níveis de displasia	124
Figura 51 – Análise de precisão e revogação para Engenharia de Características + ResNet50 (20+) com diferentes níveis de displasia	125
Figura 52 – Análise de precisão e revogação para Engenharia de Características + ResNet50 (10+) com diferentes níveis de displasia	126
Figura 53 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet (50+) com diferentes níveis de displasia	127
Figura 54 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet (40+) com diferentes níveis de displasia	128
Figura 55 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet (30+) com diferentes níveis de displasia	129
Figura 56 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet (20+) com diferentes níveis de displasia	130
Figura 57 – Análise de precisão e revogação para Engenharia de Características + ResNet50 + AlexNet (10+) com diferentes níveis de displasia	131

Lista de tabelas

Tabela 1	–	Comparação entre métodos de identificação, classificação de displasias e divergências empregadas no contexto da recuperação de imagens.	31
Tabela 2	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) de descritores obtidos do modelo AlexNet e distâncias para as displasias nos tecidos de graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	68
Tabela 3	–	Resultados do Teste de <i>Kruskal-Wallis</i> (Modelo AlexNet)	68
Tabela 4	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) do descritor ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	72
Tabela 5	–	Resultados do Teste de <i>Kruskal-Wallis</i> (Modelo ResNet-50)	72
Tabela 6	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) do descritor AlexNet + ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	76
Tabela 7	–	Resultados do Teste de <i>Kruskal-Wallis</i> (AlexNet + ResNet-50)	76
Tabela 8	–	Resultados do Teste de <i>Kruskal-Wallis</i> (Engenharia de Características) . . .	78
Tabela 9	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) dos descritores morfológicos e não morfológicos e distâncias para as displasias nos graus de tecido leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	79
Tabela 10	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) do descritor Engenharia de Características + AlexNet e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	82
Tabela 11	–	Resultados do Teste de <i>Kruskal-Wallis</i> (Engenharia de Características + AlexNet)	82
Tabela 12	–	Valores das métricas Precisão (Prec.) e Revocação (Rev.) do descritor Handcraft + ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	85
Tabela 13	–	Resultados do Teste de <i>Kruskal-Wallis</i> (Engenharia de Características + ResNet-50)	85

Tabela 14 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor Engenharia de Características + ResNet-50 + AlexNet e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).	88
Tabela 15 – Resultados do Teste de <i>Kruskal-Wallis</i> (Engenharia de Características + AlexNet + RestNet-50)	89

Sumário

Agradecimentos	5
Epígrafe	7
Lista de ilustrações	13
Lista de tabelas	17
1 INTRODUÇÃO	21
1.1 Motivação	23
1.2 Objetivos	24
1.3 Contribuições	25
1.4 Desafios da Pesquisa e Hipóteses	25
1.5 Organização da Dissertação	26
2 TRABALHOS RELACIONADOS: ESTADO DA ARTE	27
2.1 As divergências empregadas no contexto da recuperação de imagens.	27
2.2 Identificação e Classificação de Displasias	29
2.3 Considerações Finais	30
3 FUNDAMENTAÇÃO TEÓRICA	33
3.1 Câncer	33
3.2 Câncer da cavidade oral	34
3.3 Displasia Epitelial	35
3.4 Histologia	36
3.5 Imagens Histológicas	38
3.5.1 Preparação da Base de Imagens	39
3.6 Redes Neurais Artificiais	40
3.7 Redes Neurais Convolucionais	42
3.8 Operações Morfológicas	45
3.9 Segmentação	46

3.9.1	Aplicação da Arquitetura ResNet50 e Mask R-CNN na Segmentação de Núcleos em Tecido Histológico	46
3.10	Pós Processamento	51
3.11	Extração de Características	52
3.11.1	Extração de Características em Imagens Segmentadas	52
3.11.2	Extração de Características com Redes Neurais Convolucionais	54
3.12	Processo de Recuperação de Imagens Baseado em Conteúdo	56
3.13	Avaliação do Desempenho de um CBIR	58
4	METODOLOGIA	61
4.1	Divergências de Bregman	61
4.1.1	Tratamentos para Kullback Leibler	63
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	65
5.1	Investigação dos Descritores Obtidos com as Redes de Aprendizagem Profunda	66
5.1.1	Modelo AlexNet	66
5.1.2	Modelo ResNet-50	70
5.1.3	Associação dos descritores obtidos com AlexNet e ResNet-50	74
5.1.4	Descritores Obtidos por Engenharia de Características	78
5.1.5	Descritores Obtidos por Engenharia de Características e Modelo AlexNet	80
5.1.6	Descritores Obtidos por Engenharia de Características e Modelo ResNet-50	84
5.1.7	Descritores Obtidos por Engenharia de Características e Modelos CNNs	87
6	CONCLUSÃO	91
6.1	Trabalhos Futuros	93
6.2	Contribuições em Produção Bibliográfica	93
	REFERÊNCIAS	95
	APÊNDICE	101

Introdução

Em conjuntos de imagens o desafio de lidar com dados não estruturados se torna frequente, e com o avanço da internet e a disseminação de dispositivos para a captura de imagens, como redes sociais, scanners, imagens compartilhadas por aparelhos celulares a disponibilidade de coleções de imagens aumentou significativamente nos últimos anos (SILVA; MAGALHÃES; FALCÃO, 2010).

Diante desse cenário, são criadas oportunidades de desenvolvimento de estudos e aplicações para a comparação de uma imagem com outra, visando encontrar similaridades entre elas. Resultando no aumento da busca de usuários por ferramentas eficientes que possibilitem a pesquisa, navegação e recuperação de informações em diversos domínios, tais como sensoriamento remoto, moda, prevenção de crimes, publicidade, medicina, arquitetura, entre outros (SILVA; MAGALHÃES; FALCÃO, 2010).

Na área da computação gráfica, o *Content Based Image Retrieval* (CBIR), ou recuperação de imagem por conteúdo, refere-se a um sistema de banco de dados que recebe uma imagem como entrada e fornece uma lista de imagens semelhantes como saída. As imagens empregadas podem ser provenientes de diversas fontes e modelos. Avaliar a similaridade entre as imagens geralmente envolve a aplicação de técnicas que utilizam relações de distâncias entre as imagens para identificar as mais semelhantes através da descrição do conteúdo visual de cada imagem (LI; YANG; MA, 2021; SILVA; MAGALHÃES; FALCÃO, 2010).

O CBIR têm se destacado, sobretudo, devido à subjetividade envolvida na caracterização de uma imagem pelo seu conteúdo. Isso ocorre porque distintos usuários podem manifestar interesse em aspectos diversos de uma mesma imagem, resultando em percepções distintas (SONG et al., 2022). A identificação de características distintas nas imagens, em CBIR, é diretamente impactada pela métrica de distância adotada.

A escolha da métrica de distância desempenha um papel crucial no reconhecimento de imagens, afetando diretamente a eficácia e a precisão dos algoritmos utilizados. Diferentes métricas de distância são aplicadas com base nas características das imagens e nas exigências específicas do problema em questão. A métrica de distância representa uma ferramenta fundamental em diversas subáreas da matemática e da computação. Destacam-se a Geometria, Probabilidade e Estatística, Teoria dos Grafos, Redes e Agrupamentos (*Clustering*), Sistemas de Recomendação, Reconhecimento de Padrões, Visão Computacional, Computação Gráfica, Astronomia, Biologia

Molecular, Física, e várias outras. Embora se tenha uma compreensão consolidada do conceito de distância como uma medida do espaço entre dois objetos, a aplicação desse entendimento à similaridade entre os mesmos pode não ser tão intuitiva (DEZA; DEZA, 2016).

A similaridade pode ser definida basicamente como um valor numérico que representa o quanto dois objetos ou dados são semelhantes (FERREIRA; RIBEIRO; BARCELOS, 2017). Em outras palavras se este valor for alto entre imagens ou objetos, mais semelhantes serão. Por outro lado a dissimilaridade relaciona este valor como a diferença entre imagens ou objetos, relacionando de forma oposta a similaridade. Tais medidas surgem quando se comparam distribuições de probabilidade, séries temporais, imagens e matrizes entre outros (SKOPAL, 2006).

As distâncias métricas e não métricas desempenham um papel essencial na análise de dados e na quantificação da similaridade entre diferentes conjuntos de informações. Duas distâncias comuns são a Euclidiana e a Manhattan, enquanto duas divergências Bregman não métricas notáveis são a KL (Kullback-Leibler) e Mahalanobis. Em resumo, as distâncias métricas, como a Euclidiana e a Manhattan, são ferramentas fundamentais para medir a distância física entre pontos em espaços euclidianos. Por outro lado, as divergências de Bregman, como a KL e a Mahalanobis, desempenham um papel crucial em contextos mais complexos como não métricos.

Estudos anteriores utilizaram as divergências de Bregman em sistemas CBIR buscando melhorar a precisão. Em (LIU et al., 2012) foi introduzida a Divergência Total Bregman para recuperar formas de objetos. Em (ROCHA et al., 2021) os resultados da métrica Euclidiana e distância do Cosseno são comparados com as divergências de Bregman (Kullback-Leibler e Generalizada) na recuperação de imagens nas abordagens BSM (*Binary Saliency Map*), FISM (*Fuzzy Image Descriptor Based on Saliency Map*), BoVW (*Bag-of-Visual-Words*) e BoVW-SPM (*Bag-of-Visual-Words with Spatial Pyramid Matching*), com das divergências de Bregman apresentando melhores resultados. Em (ZHENG et al., 2018) são utilizadas imagens de tecidos corados com Hematoxilina e Eosina, que destacam respectivamente o núcleo e o citoplasma das células, em processos de classificação e recuperação de imagens, que em (TOSTA et al., 2019) foram utilizadas na investigação da evolução dos estágios de um câncer. No entanto, o emprego de abordagens CBIR ainda é limitada em patologia computacional (TOMMASINO et al., 2023).

Estima-se que na última década houve um aumento de 20% na incidência de casos de câncer. O câncer da cavidade oral, foco de estudo desta pesquisa, está entre os dez tipos de cânceres mais comuns e o sexto com maior mortalidade (SANTOS et al., 2023). Em escala global ocorreram aproximadamente 19,3 milhões de novos casos de câncer e quase 10 milhões de mortes em 2020 (SUNG et al., 2021). O carcinoma de células escamosas em cavidade oral é uma lesão que ocorre na região da boca e pode ser associado ao aparecimento de lesões malignas nos tecidos. Dentre os principais estágios da evolução das lesões, há as lesões denominadas displasias, que estão em estágios pré-câncer. Essas lesões se desenvolvem a partir da progressão de uma hiperplasia epitelial, passando por diferentes níveis que podem ser categorizadas em níveis leve, moderada e severa. Quando diagnosticadas em estágios iniciais podem ser tratadas (SILVA et al., 2024).

Este estudo aborda o desafio de realizar recuperação de imagem por conteúdo de lesões com potencial maligno obtidas de imagens histológicas da cavidade oral. Utilizando técnicas avançadas de segmentação com uma rede Mask R-CNN, foram identificadas e delimitadas as

regiões de núcleos relevantes. A partir dessas regiões segmentadas, foram extraídos descritores de textura e utilizados modelos de redes neurais convolucionais para compor vetores representativos de características.

A extração de características é essencial em tarefas de processamento de imagens, como classificação e reconhecimento de padrões. Redes neurais convolucionais (CNNs) automatizam esse processo, aprendendo representações hierárquicas das imagens que capturam desde padrões simples até características complexas, sendo eficazes em diversas aplicações, como o diagnóstico médico. Em contraste, a engenharia de características envolve a extração manual de descritores, utilizando operações morfológicas (dilatação, erosão, etc.) para realçar formas e estruturas, além de técnicas não morfológicas. A combinação dessas abordagens pode aumentar a precisão na análise e recuperação de imagens.

Além disso, o estudo incluiu etapas de processamento que transformaram esses vetores em espaços de alta dimensionalidade, essenciais para a avaliação da qualidade da recuperação de imagens. Durante este processo, foram aplicadas diferentes medidas de dissimilaridade para comparar e selecionar imagens de forma eficiente.

Este trabalho também contribui para o campo ao introduzir o uso das divergências de Bregman Kullback-Leibler e Mahalanobis como métricas de similaridade no âmbito do CBIR. Essas métricas foram especialmente integradas na fase crucial de recuperação de imagens, permitindo uma avaliação mais precisa e eficaz da similaridade entre imagens. A adoção dessas divergências pode melhorar os resultados obtidos em espaços não métricos tradicionais, aumentando tanto a precisão quanto a eficiência do CBIR.

1.1 Motivação

A maior parte do trabalho dedicado ao desenvolvimento de estruturas e algoritmos que aplicam medidas de similaridade utiliza espaços métricos como o espaço euclidiano. Isto porque as medidas métricas satisfazem as propriedades básicas do espaço métrico, como não negatividade, simetria e desigualdade triangular, que permite a criação de estruturas de dados para acelerar a execução de consultas por similaridade. Entretanto, para casos que não seguem estas propriedades, utilizar medidas de similaridade não métricas contribuem para solução de problemas visando a qualidade semântica dos resultados.

Em problemas de recuperação de imagens, busca-se avaliar a similaridade entre uma parte da imagem e o todo, já que a parte reflete características presentes no conjunto completo. Embora as figuras de um homem e de um cavalo possam compartilhar algumas características visuais, como forma e estrutura, elas claramente diferem em outros aspectos. Isso demonstra que, em muitos cenários práticos, o uso da distância euclidiana (ou outras medidas métricas) pode não ser a abordagem mais adequada para avaliar similaridades complexas, já que essas medidas podem não capturar nuances mais sutis presentes nas imagens (SONG et al., 2022).

Assim existem situações para as quais as funções de similaridade métricas apresentam limitações, surgindo as divergências de Bregman como alternativas para medir a similaridade e proximidade nestas aplicações.

A generalização dada pela definição de divergências de Bregman permite também o desenvolvimento de métodos que são eficazes e aplicáveis em termos de um conjunto de funções de desigualdade ao invés de um único objetivo. A vantagem está em escolher a medida de acordo com o problema, sendo possível utilizar generalizações como por exemplo: Distância de Kullback-Leibler e Mahalanobis.

No campo de busca de imagens, existe o problema focado inicialmente em comparar duas imagens e calcular suas semelhanças (similaridade). O problema geralmente não é uma tarefa complexa para os seres humanos, pois é possível reconhecer estruturas e semântica de imagens através do nosso campo de visão, mesmo que represente uma grande desafio para os computadores. Características locais têm sido estudadas desde o início década de 1980 (MORAVEC, 1981), mas somente amplamente utilizadas em busca de imagens após o ano 2000 por (LOWE, 2004) através da solução SIFT para detectar pontos de interesses locais.

Inicialmente, nas décadas de 1970 a 1990, a análise de imagens médicas envolvia a aplicação sequencial de processamento de pixel em níveis mais baixos (como filtros detectores de borda e linhas, crescimento de região) e modelagem matemática (ajuste de linhas, círculos e elipses). Esse método visava construir regras compostas com base em sistemas que conseguissem lidar com tarefas específicas. Na virada do milênio, as técnicas supervisionadas, que envolvem a utilização de dados de treinamento para o desenvolvimento de sistemas, ganhavam crescente popularidade na área da análise de imagens médicas.

Desde 2000, o crescimento explosivo de imagens da Web reformulou muito a pesquisa de recuperação de imagens. Dados os bilhões de imagens prontamente disponíveis na Web, a utilização de conjuntos de dados de imagens da Web tornou-se muito importante para vários problemas de análise de conteúdo multimídia. No entanto, no início dos anos 2000, foi difícil construir um índice eficiente para imagens de alta dimensão porque a maioria dos sistemas não pode ser dimensionada para milhões ou bilhões de imagens (ZHANG; RUI, 2013).

Alguns algoritmos e estruturas importantes foram generalizadas utilizando as divergências de Bregman. A generalização possibilita a escolha da medida de dissimilaridade tornando flexível a aplicação dos métodos que envolvem a busca por similaridade (FERREIRA, 2017). Dado a aplicação da generalização oferecida pela definição das divergências de Bregman é possível realizar o desenvolvimento de métodos que sejam mais eficientes e aplicáveis considerando um conjunto de funções de dissimilaridades ao invés de uma única medida, tornando vantajosa a utilização de tais técnicas e sendo possível a adequação a cada problema enfrentado.

1.2 Objetivos

Espaços métricos, como o Euclidiano, são frequentemente usados em algoritmos de dissimilaridade. No entanto, em situações que envolvem dados complexos, como imagens, as métricas tradicionais podem não ser adequadas. As Divergências de Bregman, que não seguem a desigualdade triangular, oferecem maior adaptabilidade. Este trabalho utiliza as Divergências de Kullback-Leibler e Mahalanobis para recuperação de imagens de displasias bucais, visando melhorar o desempenho em relação às métricas convencionais, auxiliando no diagnóstico precoce e

avaliação da severidade das lesões.

O presente trabalho apresenta um estudo de recuperação de imagens baseada em conteúdo em imagens histológicas da cavidade oral baseado em métricas da divergência de Bregman. Assim, os principais objetivos específicos desse trabalho são:

- Investigação de métricas das divergências de Bregman no processo de recuperação de conteúdo para avaliação dos graus de displasia em imagens histológicas;
- Estudo comparativo entre as divergências de Bregman (KL e Mahalanobis) e as medidas tradicionais (Euclidiana e Manhattan) em imagens da cavidade oral;
- Análise do impacto na eficiência do processo de recuperação de conteúdo na avaliação de descritores obtidos por meio de redes neurais convolucionais e características extraídas por meio de técnicas de engenharia de características.

1.3 Contribuições

Este estudo almeja contribuir para o avanço do conhecimento ao introduzir o uso das divergências de Bregman Kullback-Leibler e Mahalanobis como medidas de similaridade no contexto de CBIR, especialmente durante a etapa de recuperação de imagens. A relevância dessa abordagem reside na capacidade de avaliar a similaridade de maneira mais eficaz, uma vez que essas divergências apresentam características que podem aprimorar os resultados em comparação com o uso de espaços não métricos. Este conceito busca aprimorar tanto a precisão quanto a eficiência do CBIR, apresentando uma proposta inovadora para enfrentar os desafios associados à avaliação de similaridade em imagens, concentrando-se, neste caso, em imagens relacionadas ao câncer da cavidade oral. Assim, este estudo contribui não apenas para o avanço do CBIR, mas também para pesquisas relacionadas à identificação do câncer da cavidade oral.

1.4 Desafios da Pesquisa e Hipóteses

A aplicação das Divergências de Bregman, especificamente as de Kullback-Leibler e Mahalanobis, é utilizada durante a fase de avaliação de similaridade em sistemas de recuperação de imagens médicas baseados em conteúdo. Espera-se que as Divergências de Bregman sejam mais eficientes em comparação com as métricas convencionais, como as distâncias Euclidiana e Manhattan. A eficácia desse enfoque é sensivelmente influenciada pelos procedimentos implementados para garantir a adequada manipulação das funções logarítmicas. Esses tratamentos específicos desempenham um papel crucial na otimização do desempenho dessas divergências, destacando a pertinência dessa estratégia para aprimorar a precisão e a eficiência na recuperação de imagens médicas, alinhando-se de forma mais adaptada às complexidades dos dados envolvidos.

1.5 Organização da Dissertação

Este texto foi organizado em 6 capítulos para apresentar o estudo de maneira clara e compreensível. Cada capítulo aborda os conceitos e métodos empregados, os experimentos realizados e os resultados obtidos.

- ❑ **Capítulo 1:** inclui as considerações iniciais e o contexto em que está inserido pesquisa, motivação para seu desenvolvimento e objetivos a atingir.
- ❑ **Capítulo 2:** revisa trabalhos correlatos que utilizam não somente divergências de Bregman, mas também outras técnicas, focando em aplicações CBIR e em outras áreas aprendizado de máquina. A análise destaca as contribuições e limitações desses estudos, proporcionando uma base teórica para o desenvolvimento deste trabalho.
- ❑ **Capítulo 3:** aborda os conceitos fundamentais necessários para o entendimento do estudo. São discutidos tópicos sobre câncer, com ênfase no câncer da cavidade oral e displasia epitelial, além de princípios de histologia e análise de imagens histológicas. Também são introduzidos métodos avançados como redes neurais artificiais, redes neurais convolucionais e operações morfológicas, que são essenciais para o processamento e análise das imagens estudadas.
- ❑ **Capítulo 4:** detalha os procedimentos metodológicos empregados no estudo. Abrange a preparação da base de imagens, segmentação utilizando redes neurais, pós-processamento, extração de características, processo de recuperação, avaliação do desempenho de sistemas CBIR, e o uso de divergências de Bregman. Este capítulo fornece uma descrição completa das etapas experimentais e das técnicas aplicadas para alcançar os resultados propostos.
- ❑ **Capítulo 5:** apresenta os experimentos realizados com diferentes combinações de características extraídas usando redes neurais AlexNet e ResNet-50, bem como engenharia de características (*handcraft*). As combinações analisadas incluem AlexNet + ResNet-50, Handcraft + AlexNet, Handcraft + ResNet-50, e Handcraft + AlexNet + ResNet-50. O capítulo avalia o desempenho dessas combinações utilizando distâncias métricas e não métricas na base de características extraídas.
- ❑ **Capítulo 6:** este capítulo oferece uma visão geral dos principais resultados obtidos e explora a aplicação dos métodos propostos, com ênfase em suas contribuições e limitações dentro do contexto da recuperação de imagens por conteúdo, utilizando divergências de Bregman.

Trabalhos Relacionados: Estado da Arte

2.1 As divergências empregadas no contexto da recuperação de imagens.

Em (LIU et al., 2012) ao aumentar o número de imagens em sistemas biométricos, bases de dados públicas e na internet, algoritmos de pesquisa eficientes e precisos para encontrar as melhores correspondências tornaram-se essenciais para uma variedade de tarefas. Como resultado, a recuperação de imagens está se tornando cada vez mais essencial para a visão computacional e é essencial para uma ampla gama de aplicações potenciais. Os algoritmos de recuperação de imagens baseados em palavras-chave e baseados em conteúdo são os dois tipos mais comuns encontrados na literatura atual. Palavras-chave são ferramentas cruciais e fáceis de usar para representar e recuperar imagens. Ainda assim, as palavras-chave são muito subjectivas, pois diferentes pessoas podem indexar a mesma imagem usando palavras-chave diferentes, apesar de serem eficientes. Como resultado, a noção de recuperação baseada em elementos da imagem, como textura, cor e forma, desperta interesse. Os estudos mostram que a forma mais genérica é uma das melhores para reconhecimento. As áreas de detecção de objetos e reconhecimento de ações usam frequentemente a comparação e classificação de formas. Os algoritmos de recuperação mais rápidos são necessários devido ao tamanho considerável dos bancos de dados de imagens e formas atualmente disponíveis. Esses algoritmos, como a pesquisa reversa de imagens do TinEye e a recuperação de imagens do Google, requerem respostas em tempo real. Este artigo apresenta um método de recuperação de formas rápido e preciso que usa modelos de mistura gaussiana (GMMs) para representar formas. Um algoritmo de agrupamento suave de divergência total de Bregman é usado para dividir as formas em grupos menores. No algoritmo, cada cluster é representado por um centro t , que é um centro de l_1 norma baseado em total de Bregman. O centro t é uma combinação de todos os elementos do conjunto ponderados. Além disso, tem uma expressão fechada e é resistente a ruídos e valores inconsistentes. O total de mínimos quadrados que foi usado na regressão linear para ajustar uma linha ou plano a um conjunto de pontos é chamado de total de Bregman. O objetivo da regressão linear de mínimos quadrados ordinários e da regressão linear baseada em mínimos quadrados totais é reduzir a distância entre cada ponto e uma linha ou plano. Foi examinado o algoritmo de clustering suave

em bancos de dados de imagens reais e conjuntos de dados sintéticos. Também examinaram o clustering rígido total de Bregman e o clustering suave e rígido de Bregman. Além disso, realizaram a tarefa de recuperação de formas usando o agrupamento suave total de Bregman para vários bancos de dados de formas. Para fazer isso, foi dividido o banco de dados em clusters menores e usaram o centro t como representação para cada cluster. Como o centro do cluster é fechado e depende apenas das médias e variâncias do GMM, o agrupamento não custa muita computação. A árvore k é um método de recuperação extremamente eficaz, o que significa que as comparações são logarítmicas. Além disso, como a totalização de Bregman entre os dois GMMs também tem uma forma explícita, todas as comparações são feitas muito rapidamente. Por fim, o método é eficiente, invariante a transformações rígidas, resistente a outliers e produz resultados superiores ou comparáveis aos dos métodos do estado da arte.

De acordo com o artigo de (SAKTEL, 2014) com a crescente utilização de imagens médicas em pesquisas clínicas relacionadas a medicamentos, doenças e educação, torna-se evidente a necessidade premente de métodos eficazes para armazenar, consultar e recuperar tais imagens com base em seu conteúdo. Este artigo delinea o design e as aprimorações de um CBIR em múltiplos níveis destinado a imagens microscópicas. Esse sistema faz uso de um banco de dados de localização contendo imagens relacionadas a diversas doenças cancerígenas. A abordagem do CBIR neste sistema opera em diversos níveis, permitindo a recuperação e classificação do principal tipo de câncer e de seus subtipos, tarefas frequentemente desafiadoras em termos de discriminação e categorização. A flexibilidade do sistema possibilita tanto a consulta de várias imagens quanto a recuperação de imagens no nível de slide, buscando preservar a coesão semântica entre as imagens recuperadas. Além disso, o sistema incorpora termos ponderados, derivados da teoria da recuperação de informações, para otimizar consultas e recuperações envolvendo múltiplas imagens. Adicionalmente, o sistema demonstra eficiência ao integrar o CBIR com histograma e dicionário de aprendizado de objetos. Neste artigo, é realizado um levantamento de várias técnicas de detecção de câncer que se baseiam na metodologia CBIR. Essa abordagem CBIR apresenta uma inovação significativa no âmbito dos algoritmos de recuperação, tanto no nível de imagem quanto no nível de slide. Para o algoritmo de recuperação em nível de imagem é utilizado Support Vector Machine (SVM) na identificação do tipo predominante de doença cancerígena. Por sua vez, o algoritmo CBIR é utilizado para avaliar a similaridade entre as imagens, atuando como um mecanismo de recuperação de imagem. Esse algoritmo realiza a categorização das imagens e localiza o vizinho mais próximo por meio do classificador KNN. Posteriormente, fornece uma pontuação de saída para as imagens. O mecanismo de recuperação em nível de slide opera com base no algoritmo de nível de imagens. Este algoritmo não apenas identifica a similaridade entre as imagens de maneira convencional, mas também realiza a classificação dessas imagens semelhantes para oferecer uma consulta definida. Ele atribui pontuações mais altas às imagens mais similares e determina o subtipo da principal doença do câncer.

No trabalho de (BOYTSOV; NYBERG, 2019) o foco está nas pesquisas de baixa dimensão não métricas, onde estratégias baseadas em árvores oferecem recuperação eficiente e precisa, mantendo um tempo de indexação reduzido. Estes métodos dependem do particionamento do espaço e requerem uma regra de poda para evitar a exploração de áreas pouco promissoras. Investigaram duas abordagens bem conhecidas para estender essas regras a espaços não métricos:

TriGen é uma aproximação linear por partes da regra de poda. Foram propostos e avaliados duas adaptações do TriGen para lidar com semelhanças não simétricas (uma limitação presente na versão original do TriGen que não suporta distâncias não simétricas). O método TriGen consiste em "esticar" a função de distância por meio de uma transformação côncava monotônica, reduzindo a não-metricidade da distância. Essa abordagem foi desenvolvida especificamente para distâncias semimétricas limitadas, as quais são essencialmente simétricas, não negativas e assumem valor zero apenas para pontos de dados idênticos. Até o momento, não possuem conhecimento de quaisquer extensões anteriores para distâncias não simétricas, com exceção de uma abordagem direta de filtragem e refinamento. Na pesquisa não métrica de baixa dimensão utilizaram uma pesquisa do vizinho mais próximo (k-NN), usada em muitos domínios, incluindo aprendizado de máquina (ML), mineração de dados, recuperação de informações, e processamento de linguagem natural. Informalmente, a pesquisa k-NN é uma tarefa de recuperação de dados com k entradas mais próximas de um ponto de consulta em relação a alguma distância ou função de similaridade. Este problema originou-se da busca espacial do mundo real. Com o desenvolvimento constante da abstração do espaço vetorial, o problema foi generalizado para um espaço vetorial multidimensional e/ou métrico genérico, onde o último pode não ter a estrutura do espaço vetorial. Motivando o uso de distâncias não métricas completas como divergências de Bregman.

2.2 Identificação e Classificação de Displasias

(SILVA et al., 2022a) propuseram um método para quantificação de displasia em imagens histopatológicas da cavidade oral utilizando modelos de aprendizado de máquina. A metodologia abrange etapas de segmentação de núcleos, pós-processamento, extração de características e classificação. Para a segmentação, foi utilizada a rede neural Mask R-CNN. O pós-processamento envolveu operações morfológicas para eliminar áreas de falsos positivos e negativos. Em seguida, foram extraídas 23 características morfológicas e não morfológicas, como área, orientação, solidez e entropia, e um classificador polinomial foi empregado para distinguir os diferentes graus da lesão. Essa abordagem foi aplicada a um conjunto de dados contendo 296 regiões de interesse, onde 9.155 núcleos celulares foram identificados e analisados. A etapa de segmentação atingiu acurácias entre 88,92% e 90,35%, enquanto a classificação obteve uma área sob a curva ROC de 0,970. Os autores concluíram que a metodologia proposta contribui para o estado da arte e pode ser utilizada como ferramenta para auxiliar patologistas com valores precisos na investigação de lesões displásicas.

No estudo realizado por (SILVA et al., 2024), os autores disponibilizaram um conjunto de imagens públicas relacionadas à displasia epitelial oral. Este conjunto de dados contém 456 imagens obtidas de 30 línguas de camundongos. As imagens foram classificadas de acordo com os graus de lesão, com as estruturas nucleares marcadas manualmente por um especialista treinado e validadas por um patologista. Além disso, os autores conduziram experimentos para demonstrar o potencial do conjunto de dados na segmentação semântica e de instância, bem como na avaliação de métodos de normalização de coloração H&E durante o pré-processamento. Na

etapa de classificação, diferentes arquiteturas de CNN e algoritmos de aprendizado de máquina foram testados. O classificador *Random Forest* obteve o melhor desempenho, com uma acurácia de 94,22%. Os autores enfatizam que, embora os resultados sejam promissores, novos estudos são necessários para melhorar essas etapas de diagnóstico automatizado. A disponibilização deste banco de imagens permite que outros pesquisadores explorem novas técnicas de classificação, aprimorando o uso de CAD em aplicações clínicas de imagens de tecido de displasia epitelial oral.

O trabalho de (KRISHNAN et al., 2009) se concentrou em apresentar uma abordagem para a classificação automatizada de células de tecidos com fibrose submucosa oral (FSO) usando características geométricas. Neste estudo, dois especialistas independentes identificaram e classificaram manualmente 2.894 células. Como classificador, o algoritmo SVM usou 1.054 células no conjunto de treinamento e 1.840 células no conjunto de teste, usando validação cruzada. Entre as células analisadas, 730 eram de amostras comuns e 1.110 estavam ligadas à FSO. Mesmo concentrando-se em características geométricas, a técnica alcançou uma taxa de precisão de 88,69%. Os autores chegaram à conclusão de que é possível usar métodos matemáticos para identificar divergências biológicas. Eles também pretendem estudar outras características morfológicas no futuro.

2.3 Considerações Finais

Os estudos comparados (Tabela 1) mostram uma variedade de abordagens e técnicas para a identificação e classificação de displasias, bem como para a recuperação de imagens. Cada método oferece vantagens específicas e contribui de maneira significativa para o estado da arte em suas respectivas áreas. As técnicas de segmentação e classificação avançadas, como as empregadas por (SILVA et al., 2022a) e (SILVA et al., 2024), demonstram altos níveis de precisão e eficácia, indicando o potencial de aplicação clínica prática. Por outro lado, os estudos focados em recuperação de imagens, como os de (LIU et al., 2012) e (SAKTEL, 2014), mostram inovações significativas em algoritmos de recuperação, essenciais para a gestão eficiente de grandes bases de dados de imagens. Em conclusão, a aplicação das divergências de Bregman na recuperação de imagens de displasias orais oferece uma oportunidade inexplorada que pode complementar e aprimorar os métodos atuais. Essa técnica tem o potencial de contribuir significativamente para o estado da arte, fornecendo uma ferramenta adicional para patologistas e pesquisadores na análise e diagnóstico de displasias orais. A integração das divergências de Bregman com técnicas de aprendizado de máquina e processamento de imagem pode levar a avanços importantes na recuperação de imagens médicas.

Tabela 1 – Comparação entre métodos de identificação, classificação de displasias e divergências empregadas no contexto da recuperação de imagens.

Estudo	Objetivo	Métodos Utilizados	Resultados Principais
(SILVA et al., 2022a)	Quantificação de displasia em imagens histopatológicas da cavidade oral	Segmentação de núcleos com Mask R-CNN, pós-processamento, extração de características (23 características morfológicas e não morfológicas) e classificação com classificador polinomial	Segmentação: acurácia entre 88,92% e 90,35%; Classificação: AUC de 0,970
(SILVA et al., 2024)	Publicação de um conjunto de imagens públicas de displasia epitelial oral	Segmentação semântica e de instância, normalização de coloração H&E, classificação com CNN e Random Forest	Classificação com Random Forest: acurácia de 94,22%
(KRISHNAN et al., 2009)	Classificação automatizada de células de tecidos com Fibrose Submucosa Oral (FSO)	Identificação e classificação manual por especialistas, características geométricas, classificador SVM, validação cruzada	Taxa de precisão de 88,69%
(LIU et al., 2012)	Recuperação de formas rápidas e precisas	Modelos de mistura gaussiana (GMMs), algoritmo de agrupamento suave de divergência total de Bregman	Método eficiente, invariante a transformações rígidas, resistente a outliers
(SAKTEL;, 2014)	Recuperação de imagens microscópicas de doenças cancerígenas	CBIR com histograma, dicionário de aprendizado de objetos, SVM, KNN	Classificação de tipos de câncer e subtipos, consultas definidas
(BOYTISOV; NYBERG, 2019)	Pesquisas de baixa dimensão não métricas	TriGen, aproximação linear por partes da regra de poda, k-NN	Redução da não-metricidade da distância, uso em domínios como ML e recuperação de informações

Fundamentação Teórica

3.1 Câncer

O câncer, uma complexa questão genética, resulta de mutações intrínsecas ao DNA, algumas ocorrendo de forma espontânea, enquanto outras são induzidas por fatores ambientais. Essas variações genéticas são por vezes herdadas, transmitidas de geração em geração, moldando o destino das células descendentes. Nesse complexo processo biológico, células com essas mutações são submetidas à rigorosa seleção natural, a qual favorece o crescimento e a sobrevivência, potencialmente tornando-se dominantes na população celular. Definido como um conjunto de mais de 100 enfermidades, o câncer é marcado pelo descontrolado crescimento de células, capacitadas à se espalhar entre os tecidos e órgãos próximos à região, inicialmente, comprometida no organismo humano (SILVA et al., 2022a; TEIXEIRA et al., 2012).

A partir da segunda metade do século XX, as preocupações com o câncer aumentaram significativamente na sociedade brasileira. Esse fenômeno o tornou uma presença constante no dia a dia e uma preocupação social de grande relevância. Recentemente, seu impacto epidemiológico crescente e a demanda por uma cobertura mais abrangente das intervenções, especialmente através do Sistema Único de Saúde (SUS), têm colocado a doença no epicentro dos debates sobre medicina e saúde pública no Brasil. Essa situação tem promovido uma mudança no panorama epidemiológico da sociedade, tanto devido à crescente exposição a agentes cancerígenos, ao envelhecimento da população, ao avanço das tecnologias de diagnóstico, quanto ao aumento da taxa de mortalidade relacionada ao câncer (NETO; TEIXEIRA, 2017).

O carcinoma *in situ*, também conhecido como câncer não invasivo, representa o estágio inicial no processo de classificação do câncer (com exceção dos cânceres do sistema sanguíneo). Neste estágio, as células cancerosas estão confinadas à camada de tecido onde se originaram, sem se espalhar para outras camadas do órgão afetado. Geralmente, os cânceres *in situ* podem ser tratados com sucesso e têm potencial de cura se forem identificados e tratados antes de progredirem para uma fase invasiva (TEIXEIRA et al., 2012; VISENTIN; LENARDT, 2010).

De modo geral, sabe-se que, quanto antes o câncer for detectado e tratado, mais efetivo o tratamento tende a ser, maior a possibilidade de cura e melhor será a qualidade de vida do paciente. A seguir, serão discutidas as medidas que compõem a detecção precoce. Nessa etapa

do cuidado, o objetivo é detectar lesões pré-cancerígenas ou cancerígenas quando ainda estão localizadas no órgão de origem e antes que invadam os tecidos circundantes ou outros órgãos (YOUNG; HOBBS; KERR, 2011).

3.2 Câncer da cavidade oral

O câncer na cavidade oral representa um desafio importante para a saúde pública, sendo classificado como uma das formas mais comuns de neoplasias malignas em todo o mundo. Muitas dessas neoplasias se originam de lesões orais potencialmente malignas, as quais podem persistir por longos períodos, em parte devido à falta de conscientização da população sobre essas lesões e os fatores que contribuem para o seu desenvolvimento (MAIA et al., 2016).

Lesões pré-cancerosas e condições pré-cancerosas representam estágios críticos no desenvolvimento do câncer, fornecendo *insights* valiosos para a prevenção e o diagnóstico precoce da doença. Essas lesões são caracterizadas por alterações morfológicas nos tecidos que indicam uma predisposição significativa ao desenvolvimento de câncer em comparação com tecidos normais (MAIA et al., 2016). É fundamental ressaltar que a maioria das lesões desse tipo não progride naturalmente para o câncer. No entanto, ao estimular a proliferação regenerativa e expor as células a condições que podem desencadear inflamações, essas lesões aumentam as chances de evolução para formas malignas (SILVA et al., 2022a).

De acordo com os conceitos de causa necessária e causa suficiente, a presença de um agente específico em um indivíduo não é suficiente para o desenvolvimento de uma doença (THULER; SANT'ANA; REZENDE, 2011). São necessárias outras influências (ou causas) que, combinadas com o agente específico, levem ao surgimento da doença. O agente específico é considerado a causa necessária, enquanto as outras influências são chamadas de causas predisponentes. Quando combinadas, a causa necessária e as causas predisponentes formam a causa suficiente para o desenvolvimento da doença. Dessa forma, doenças multifatoriais, como o câncer, podem ter diferentes conjuntos de causas suficientes (YOUNG; HOBBS; KERR, 2011).

Lesões pré-cancerosas se concentram em mudanças específicas observadas nos tecidos, onde o câncer é mais propenso a ocorrer do que em tecidos saudáveis. Essas alterações podem incluir anomalias celulares, como hiperplasia, displasia ou metaplasia, que representam estágios iniciais de transformação celular maligna. É importante ressaltar que nem todas as lesões pré-cancerosas progredem para câncer, mas muitas delas carregam um risco significativo de evolução maligna.

A metaplasia é um processo adaptativo que pode ser revertido, onde um tipo de célula (seja epitelial ou mesenquimal) é substituído por outro tipo celular da mesma linhagem. O crescimento celular em tecidos normais pode ser regulado ou não, limitado a uma região específica ou autolimitado, e podem ocorrer pequenas variações nas células normais durante esse processo (MOURA et al., 2022).

Hiperplasia refere-se ao aumento localizado e autolimitado no número de células de um órgão ou tecido. Essas células mantêm sua forma normal e função característica do tecido original. A hiperplasia pode ser classificada como fisiológica, quando ocorre para atender às demandas normais do organismo, como durante a gestação, quando a glândula mamária aumenta

em resposta aos estímulos hormonais. Por outro lado, na forma patológica, a hiperplasia é desencadeada por estímulos excessivos, como o excesso de estrogênio que pode estimular a hiperplasia endometrial. Uma característica importante da hiperplasia é que, uma vez que os estímulos cessam, a proliferação celular também cessa (YOUNG; HOBBS; KERR, 2011).

A displasia é um fenômeno de crescimento celular que implica modificações nas características celulares. Isso envolve alterações na forma e no tamanho das células, frequentemente acompanhadas por um aumento na ocorrência de mitoses (divisões celulares). A displasia geralmente ocorre em um contexto de crescimento desordenado do epitélio, muitas vezes associado a um estado metaplásico do epitélio. No entanto, é importante ressaltar que nem todos os epitélios metaplásicos são necessariamente displásicos (DIAS, 2021; YOUNG; HOBBS; KERR, 2011).

3.3 Displasia Epitelial

Lesões potencialmente malignas são distúrbios que carregam uma maior probabilidade de se transformarem em câncer ao longo do tempo. Quando observamos o tecido epitelial, podemos identificar alterações morfológicas, como a displasia, que é um sinal claro de uma maior tendência à progressão para um estado maligno, especialmente para o *Carcinoma* de Células Escamosas (CEC). A displasia representa uma anomalia na estrutura das células, indicando um desvio do padrão normal de crescimento e diferenciação celular. Essas células displásicas possuem uma capacidade aumentada de proliferação e uma maior propensão à acumulação de mutações genéticas, aumentando assim o risco de desenvolvimento de câncer. Em comparação com o epitélio saudável, o tecido com displasia apresenta uma predisposição significativamente maior para a transformação maligna, destacando a importância da detecção precoce e da intervenção adequada para prevenir a progressão para o câncer invasivo (WAAL, 2009; DIAS, 2021).

A Organização Mundial de Saúde (OMS) estabeleceu uma classificação histológica composta por três níveis - leve, moderado e severo - para essa condição. Essa categorização é fundamental para uma avaliação abrangente da gravidade das lesões displásicas e para orientar o plano de tratamento mais adequado para cada paciente. As chances de malignidade variam dependendo do grau de displasia: lesões leves ou moderadas têm uma probabilidade de malignidade entre 4% e 11%, enquanto alterações displásicas graves apresentam uma faixa de 2% a 35% de chance de evolução para câncer (JÚNIOR; OCANHA-XAVIER, 2018; DIAS, 2021; SHIRANI et al., 2014).

As displasias leves, consistem em um distúrbio arquitetônico que afeta apenas o terço inferior do epitélio, juntamente com uma leve atipia citológica. Em relação as displasias moderadas, o primeiro critério para identificar essa categoria é a presença de uma perturbação arquitetônica que abrange até o terço médio do epitélio. Em seguida, avalia-se o grau de atipia citológica. Para identificar displasia grave, é necessário que mais de dois terços do epitélio apresentem distúrbio arquitetônico juntamente com atipia citológica. Entretanto, como mencionado anteriormente, se houver distúrbio arquitetônico até o terço médio do epitélio com atipia citológica significativa, o diagnóstico pode ser atualizado de displasia moderada para grave (WARNAKULASURIYA et al., 2008).

Um número crescente de pesquisas está se dedicando às mudanças moleculares em lesões

pré-malignas, com a esperança de que, eventualmente, possam oferecer uma base mais sólida para prever o prognóstico com maior precisão. No entanto, até o momento, esses métodos ainda não alcançaram um estágio em que possam ser amplamente utilizados para diagnóstico de rotina, mesmo em instituições com acesso à tecnologia necessária. Em muitos laboratórios de diagnóstico, é provável que a avaliação tradicional por patologistas de biópsias de tecido continue sendo essencial por vários anos. É importante ressaltar que a avaliação de displasia pelos patologistas é subjetiva, e há evidências de uma considerável variação entre eles, tanto em avaliações individuais quanto em avaliações repetidas do mesmo material em ocasiões diferentes (WARNAKULASURIYA et al., 2008; SILVA, 2019).

3.4 Histologia

A histologia é uma prática essencial na medicina, consistindo na análise minuciosa de tecidos e órgãos por meio de procedimentos como seccionamento, coloração e observação microscópica das amostras resultantes. Também conhecida como anatomia microscópica e histoquímica, essa abordagem permite não apenas a visualização detalhada da estrutura dos tecidos, mas também a detecção de quaisquer alterações específicas que possam ter ocorrido, sejam elas decorrentes de doenças, lesões ou outros fatores. A aplicação da histologia é vasta, sendo utilizada em diagnósticos clínicos, investigações científicas, autópsias forenses e em pesquisas acadêmicas, desempenhando um papel fundamental na compreensão e tratamento de uma variedade de condições médicas (EVERS III et al., 2022).

Diversas técnicas histológicas permitem a visualização de quatro tipos fundamentais de tecido humano. Embora o epitélio, o tecido conjuntivo, o tecido muscular e o tecido nervoso compartilhem algumas características, eles se apresentam estruturalmente distintos após a coloração. Cada método de coloração destaca uma característica ou componente específico dentro de um tipo de tecido. Por exemplo, a hematoxilina, um corante básico comum, confere uma coloração azul às proteínas, enquanto a eosina, outra substância corante, as colore em rosa (EVERS III et al., 2022). Antes que a coloração específica possa ocorrer, as amostras de tecido devem ser preparadas através dos seguintes estágios:

- ❑ **Fixação:** O processo inicia-se com a etapa de fixação dos tecidos, com o propósito de prevenir a deterioração das células causada por suas próprias enzimas ou por agentes bacterianos. Essa fase é essencial para garantir a preservação da estrutura celular e a manutenção das características morfológicas durante as etapas subsequentes de preparação e análise histológica. Decorrido o tempo da fixação, aconselha-se aparar as amostras para que eventuais resíduos sejam removidos na lavagem com água corrente (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020).
- ❑ **Lavagem:** Durante esta etapa, as amostras são submetidas a uma lavagem contínua em água corrente por um período de 24 horas, visando remover completamente o fixador utilizado anteriormente, como o formol (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020).

- ❑ **Desidratação:** Para facilitar a infiltração subsequente da resina, que não é miscível com a água presente nos tecidos, é realizada a desidratação, removendo a água do tecido. O recipiente utilizado deve comportar um volume de álcool cerca de dez vezes maior que o da amostra, permitindo seu armazenamento. A concentração inicial de álcool varia de acordo com a técnica adotada. Opta-se por iniciar com 30% de álcool, pois isso minimiza a contração dos tecidos, evitando distorções. Cada graduação alcoólica subsequente, 40%, 70%, 95% e 100%, deve ser mantida por dois dias antes da troca. Ao final, a amostra deve ser imersa novamente em álcool a 100% por mais dois dias (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020).
- ❑ **Diafanização:** A etapa seguinte é a diafanização, que consiste na completa remoção dos resíduos de água e álcool presentes na amostra até o momento. As amostras devem ser mantidas em xilol por dois dias e, em seguida, submetidas a uma nova troca de xilol por mais dois dias. Essa técnica não apenas conclui a desidratação, mas também cria condições ideais para a subsequente infiltração da resina (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020).
- ❑ **Infiltração:** A infiltração é quando a resina metil metacrilato (MMA) se espalha pelo material e preenche os espaços no tecido (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020).
- ❑ **Inclusão:** A inclusão é o estágio em que a resina é derramada sobre a peça, cobrindo-a completamente com uma quantidade excessiva de solução. Em seguida, o tecido é removido do recipiente e a resina se solidifica, formando um bloco que preserva a estrutura do tecido (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020; SILVA et al., 2020).
- ❑ **Microtomia:** A microtomia envolve a obtenção de cortes finos do bloco de resina, possibilitando a preparação de lâminas histológicas com o auxílio de um micrótomo. Esse dispositivo apresenta uma lâmina e um mecanismo que movimenta o bloco em intervalos regulares, permitindo cortes precisos e finos, geralmente com espessuras variando de 1 a 10 micrômetros. Após essa fase, os cortes são transferidos para lâminas de vidro para serem posteriormente corados (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020; SILVA et al., 2020).
- ❑ **Acabamento, polimento e coloração:** O próximo passo consiste no desbaste e polimento, visando alcançar a espessura desejada, minimizar a presença de arranhões e transparência. Após tornarem-se transparentes, os tecidos são tratados com corantes como Hematoxilina e Eosina (H&E), que distinguem os elementos básicos e ácidos das células. A Eosina colore o citoplasma e o colágeno em rosa, enquanto a Hematoxilina colore os núcleos celulares e outros componentes ácidos em roxo ou azul. Uma lâmina de vidro é colocada sobre o corte de tecido para protegê-lo durante a manipulação e facilitar a visualização pelo microscópio (DOMINGUEZ et al., 2012; GURINA; SIMMS, 2020; SILVA et al., 2020).

3.5 Imagens Histológicas

A microscopia óptica é o principal método para análise de estruturas celulares por meio da utilização da luz, sendo empregada há mais de 350 anos. Sua eficácia e versatilidade tornaram-na uma ferramenta fundamental na investigação e compreensão dos componentes celulares (ALVES et al., 2021).

A utilização da análise computadorizada de imagens é comum em uma variedade de aplicações do mundo real, incluindo a área médica, onde é empregada para auxiliar no diagnóstico por meio da interpretação de imagens (SOUSA, 2021). Automatizar o processo de análise de imagens envolve desenvolver algoritmos que possam conduzir o processamento digital e também extrair características relevantes das imagens. Isso permite não apenas uma análise mais eficiente, mas também abre caminho para a aplicação de técnicas avançadas de aprendizado de máquina e inteligência artificial na interpretação e tomada de decisões baseadas em imagens (GONZALEZ; WOODS, 2009). As etapas comumente empregadas no processamento de imagens histológicas são pré-processamento, segmentação, pós-processamento, extração de características e classificação.

Durante o pré-processamento, ocorre uma otimização da qualidade da imagem, a qual se desdobra em duas categorias principais: métodos que atuam no domínio espacial e métodos que atuam no domínio de frequência. As técnicas de processamento no domínio espacial utilizam filtros que operam diretamente na estrutura da imagem, ao passo que as técnicas de processamento no domínio da frequência empregam filtros que afetam o espectro de frequência da imagem (ALBUQUERQUE; ALBUQUERQUE, 2000).

Na fase de segmentação, a imagem é dividida em componentes distintas que compõem a totalidade dela, porém com características diferentes entre si. Os pixels de interesse são comumente referidos como "objetos", enquanto os pixels irrelevantes ou sem utilidade são classificados como "fundo". No conjunto de tarefas do processamento de imagens, a segmentação emerge como uma etapa crítica, fundamental para a identificação e definição das áreas de interesse que serão objeto de análises e procedimentos futuros. Falhas ou imprecisões nessa fase reverberam em todas as etapas subsequentes, podendo comprometer os resultados finais e minar a eficiência do processo como um todo (ALBUQUERQUE; ALBUQUERQUE, 2000; GONZALEZ; WOODS, 2009).

Um procedimento de segmentação frequentemente empregado é a limiarização de Otsu, que determina o limiar de modo a otimizar a diferença de variância entre as classes, enquanto minimiza a variância interna de cada uma delas, facilitando assim a separação das regiões (GONZALEZ; WOODS, 2009). Um método adicional empregado é o K-means, que consiste em três etapas. Na primeira etapa, o algoritmo inicializa um número de centros de cluster aleatoriamente. Na segunda etapa, os pixels que são semelhantes ao centróide de cada cluster são atribuídos a esse cluster. Na terceira etapa, os centróides de cada cluster são recalculados. Estas etapas são iteradas até que não ocorram mais mudanças nos pixels atribuídos aos clusters (IRSHAD et al., 2014).

Sendo imprescindível a aplicação de técnicas de pós-processamento para aprimorar o resultado da segmentação, realçando os dados essenciais para uma extração mais precisa. Isso visa aproximar as informações obtidas durante a segmentação daquelas interpretadas por espe-

cialistas Para esse fim, é frequente empregar operações morfológicas, as quais visam eliminar pequenos objetos indesejados ou realçar áreas cruciais para a análise da imagem.

Após as etapas anteriores, a aplicação da extração de características possibilita a computação dos atributos dos núcleos segmentados, permitindo a identificação de anomalias. Tipicamente, são exploradas características relacionadas à morfologia e textura dos núcleos. Essas características dos núcleos celulares são valiosas para detectar o surgimento de lesões e o avanço do câncer. Portanto, a extração de características de imagens histológicas desempenha um papel crucial na análise quantitativa das estruturas celulares (IRSHAD et al., 2014; GONZALEZ; WOODS, 2009).

Durante essa etapa, é viável empregar técnicas de aprendizado profundo, como as Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks* - CNN), embora a extração de características seja abordada de forma distinta. As CNNs possuem uma arquitetura composta por diversas camadas, cada uma encarregada de extrair características específicas do dado de entrada (GONÇALVES et al., 2016).

Durante o processo de classificação, o objetivo é realizar de maneira automática a identificação dos objetos segmentados na imagem. Nesse procedimento, há a colaboração de um especialista, que, em uma fase inicial, realiza a classificação manual de algumas imagens e identifica quais características podem ser empregadas para determinar se apresentam alguma patologia. Em seguida, as técnicas classificam os demais objetos com base na sua semelhança com as classes já estabelecidas (SILVA, 2019).

3.5.1 Preparação da Base de Imagens

A avaliação dos tecidos histológicos é amplamente considerada como uma abordagem confiável para a detecção de doenças. Especificamente, no campo da saúde bucal, são conduzidas pesquisas sobre lesões, chamadas de pré-câncer, com o objetivo de esclarecer as características e anomalias que desempenham um papel significativo no surgimento de patologias. A histologia compreende a investigação das estruturas, funções e organização celular para a formação de órgãos. Além disso, envolve a análise das dimensões celulares por meio do uso de microscópio, sendo necessário a preparação do material fatiado em cortes histológicos, resultando em lâminas obtidas por etapas que compõem o processo de preparação dos tecidos. Essas etapas podem ser denominadas por: fixação, desidratação e diafanização, inclusão, microtomia, montagem e coloração.

Para este trabalho as imagens histológicas foram coletadas, tratadas e fornecidas por (SILVA et al., 2022a). Foi criado um repositório de imagens a partir de lâminas que continham lesões induzidas experimentalmente na língua de camundongos da linhagem C57Bl/6, utilizando o carcinógeno 4NQO. O experimento *in vivo* foi previamente autorizado pelo Comitê de Ética na Utilização de Animais, sob o protocolo número 038/09, e conduzido no período compreendido entre 2009 e 2010 (FARIA et al., 2011; MENDONÇA et al., 2012; SOUZA et al., 2014). No protocolo de indução de lesões na língua de camundongos, foram usados 30 animais (20-23g, 6-8 semanas), distribuídos aleatoriamente em dois subgrupos: semana 16 e semana 32. Os camundongos foram mantidos em gaiolas de plástico com serragem, a 22°C e ciclo claro/escuro

de 12h, no Centro de Bioterismo da Universidade Federal de Uberlândia. O protocolo seguiu os critérios de (TANG et al., 2004): o carcinógeno 4NQO foi diluído em propilenoglicol (5 mg/ml) e depois em água filtrada (100 µg/ml). Os camundongos consumiram essa solução por 16 semanas. No grupo "semana 16", foram sacrificados imediatamente após o tratamento; no grupo "semana 32", beberam apenas água filtrada por mais 16 semanas antes do sacrifício. Após o sacrifício, as línguas foram fixadas em formalina tamponada a 4% por 24 horas, processadas, incluídas em parafina, cortadas e coradas com H&E para estudo histopatológico. As lâminas foram classificadas como tecido saudável, displasia leve, moderada ou severa, conforme (LUMERMAN; FREEDMAN; KERPEL, 1995).

As imagens histológicas digitais foram obtidas com um microscópio óptico Leica DM500 em magnificação de 400x e armazenadas em formato TIFF (RGB, 8 bits, 2048×1536 pixels). Foram digitalizadas 43 lâminas, catalogadas como tecido saudável, displasia leve, moderada e severa, com a ajuda de um patologista. As imagens foram recortadas em regiões de interesse (do inglês, *Region of Interest* - ROI) de 450×250 pixels, totalizando 296 ROI (74 por classe de tecido). Na Figura 1, exemplifica-se cada tipo de tecido. Para cada classe de tecido, 30 ROI foram selecionadas como padrão-ouro, com os núcleos marcados manualmente pelo especialista, gerando máscaras binárias com núcleos em branco e fundo em preto.

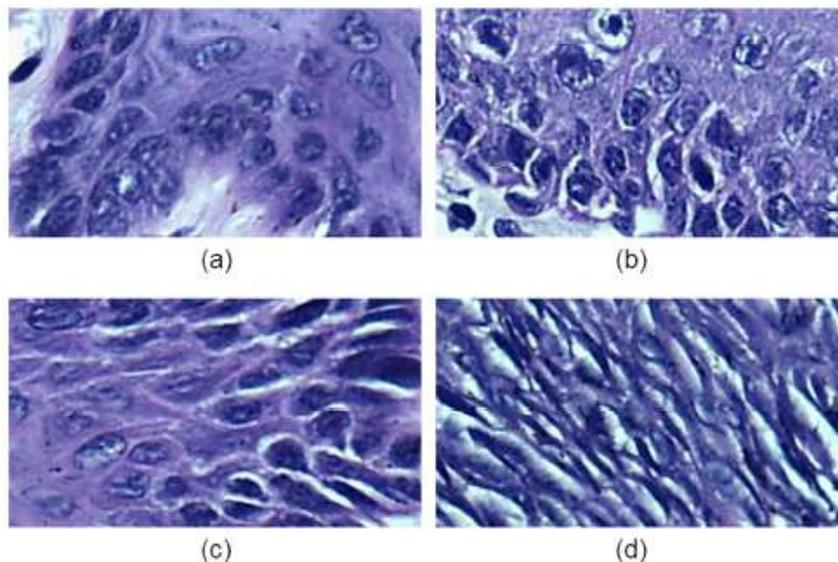


Figura 1 – Ilustração de amostras histológicas da cavidade bucal: (a) tecido saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa (SILVA et al., 2022a).

Essas imagens foram empregadas em um estudo apresentado por (SILVA et al., 2022a) para criação de um sistema de apoio a tomada de decisão constituído das etapas: segmentação, pós-processamento, extração de características e classificação. No entanto, neste estudo são considerados as etapas de segmentação, pós-processamento e extração de características.

3.6 Redes Neurais Artificiais

É conhecido que o cérebro é composto por bilhões de neurônios. Um neurônio, por sua vez, é uma célula composta por três partes com funções específicas e complementares: o corpo

celular, os dendritos e o axônio. Os dendritos têm a função de captar estímulos recebidos em um intervalo de tempo determinado e transmiti-los ao corpo celular, onde são processados. Quando esses estímulos atingem um determinado limiar, o corpo celular do neurônio gera um novo impulso que se propaga ao longo do axônio e é transmitido às células vizinhas por meio de sinapses. Esse processo pode se repetir em diversas camadas de neurônios. Como consequência, a informação de entrada é processada, potencialmente levando o cérebro a gerar respostas físicas (FERNEDA, 2006). A Figura 2 apresenta um exemplar de neurônio biológico.

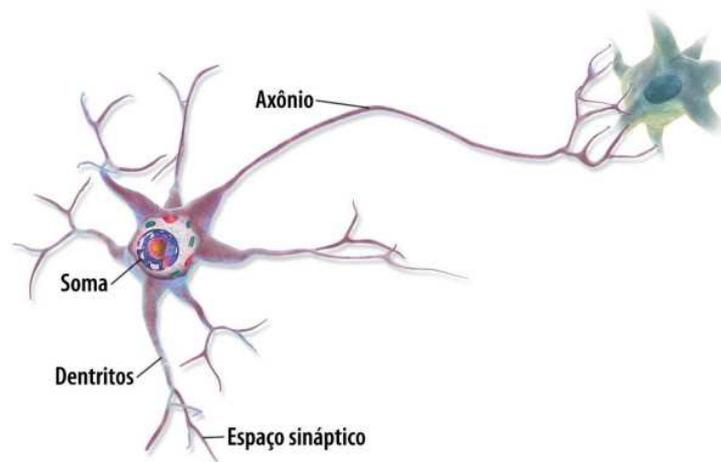


Figura 2 – Uma ilustração de um neurônio biológico em Redes Neurais Artificiais (RNA).

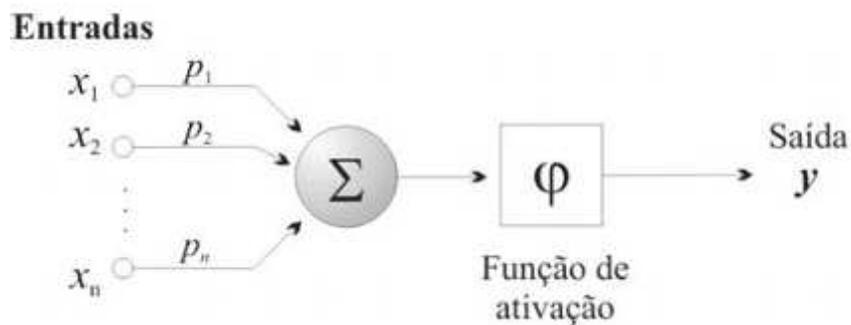


Figura 3 – Modelo matemático de um neurônio. (FERNEDA, 2006)

As Redes Neurais Artificiais (RNA), por sua vez, são sistemas de aprendizado que se inspiram na dinâmica de conexões neurais do cérebro humano. Elas replicam, em nível computacional, a interação entre neurônios biológicos, os quais possuem componentes essenciais, como dendritos, corpo celular, axônios e sinapses citados anteriormente. Essa abordagem visa gerar conhecimento e insights sobre conjuntos de objetos, simulando o funcionamento orgânico do sistema neural

Este modelo é composto por três elementos básicos (Figura 3):

- um conjunto de n conexões de entrada (x_1, x_2, \dots, x_n) , caracterizadas por pesos (p_1, p_2, \dots, p_n) ;
- um somador (Σ) para acumular os sinais de entrada;

- uma função de ativação (ϕ) que limita o intervalo permissível de amplitude do sinal de saída (y) a um valor fixo.

O funcionamento das interações entre os neurônios é emulada através de seus pesos. Esses valores podem ser tanto negativos quanto positivos, refletindo se as conexões são inibitórias ou excitatórias. A influência de um sinal originário de outro neurônio é calculada pela multiplicação da intensidade do sinal recebido pelo peso da conexão correspondente ($x_i \times p_i$). A soma dos produtos entre os valores x_i e os pesos p_i de todas as conexões é calculada, e o valor resultante é então encaminhado para a função de ativação, responsável por determinar a saída (y) do neurônio (FERNEDA, 2006).

Ao agrupar múltiplos neurônios, cria-se uma rede neural artificial. Esses modelos buscam emular o processamento de informações do cérebro humano e consistem em unidades de processamento elementares, os neurônios, que se conectam por meio de sinapses (FERNEDA, 2006).

Em uma abordagem simplificada, uma rede neural artificial pode ser concebida como um grafo, onde os nós denotam os neurônios e as conexões entre eles simulam as sinapses, como ilustrado na Figura 3 (FERNEDA, 2006).

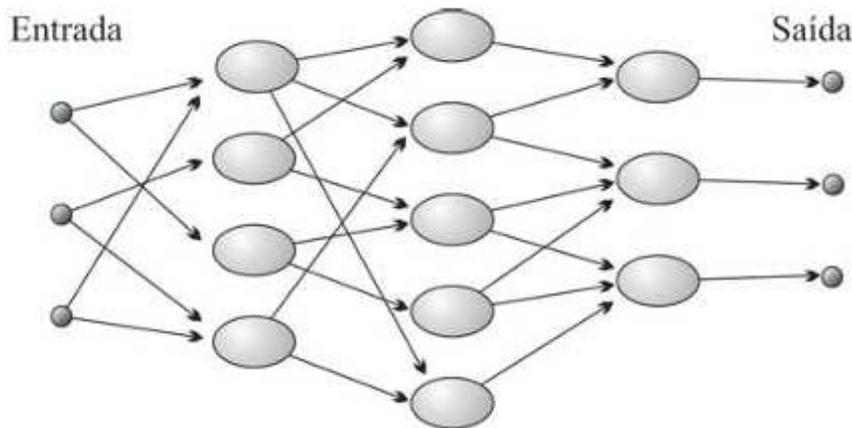


Figura 4 – Ilustração simplificada de uma rede neural artificial. (FERNEDA, 2006)

Para que a rede possa aprender os padrões presentes no conjunto de dados, é necessário submetê-la a uma fase de treinamento. Durante esse estágio, os dados são apresentados a todos os neurônios da rede, e a precisão do modelo é avaliada. Esse processo é repetido várias vezes, sendo cada ciclo conhecido como época. Após cada época, os dados são novamente propagados pela rede, e os pesos das conexões sinápticas são ajustados com base nos erros cometidos, utilizando algoritmos de otimização como o gradiente descendente. Esses ajustes visam melhorar gradativamente a precisão da rede. O treinamento continua até que se alcance uma métrica de desempenho satisfatória ou até que um número pré-definido de épocas seja completado (FERNEDA, 2006).

3.7 Redes Neurais Convolucionais

As CNNs, introduzidas por (CUN et al., 1989), representam um tipo específico de arquitetura de rede neural projetada principalmente para lidar com tarefas de visão computacional, incluindo

o reconhecimento de padrões em imagens e vídeos. O termo “convolucional” é atribuído a essas redes devido à operação essencial chamada convolução, empregada para remodelar e dimensionar imagens, preservando suas características essenciais para uma representação significativa, sem perda de informação relevante, desempenhando um papel crucial nesse tipo de arquitetura. As CNNs são altamente eficazes na extração de características relevantes de imagens complexas, permitindo o reconhecimento preciso de padrões visuais em diversos contextos (BRITO, 2023).

Na primeira camada da rede, realiza-se a operação de convolução. Essa operação implica o uso de uma matriz pequena chamada filtro ou kernel, que percorre todas as camadas de uma imagem de tamanho $x \times y \times z$ para extrair características locais. À medida que o filtro se desloca pela imagem, executa operações de multiplicação e soma para gerar um mapa de características bidimensional (BRITO, 2023). Esses mapas podem ter suas dimensões de largura (x) e altura (y) reduzidas em relação à imagem original, porém todos os mapas produzidos têm o mesmo tamanho. Conforme pode ser observado na Figura 5.

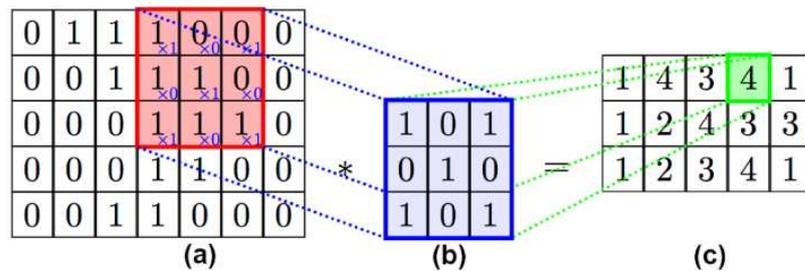


Figura 5 – Exemplo de uma operação de convolução: (a) uma imagem de tamanho 7×5 , (b) um filtro de convolução de tamanho 3×3 e (c) uma matriz de características resultante. (SANTOS et al., 2021; SILVA et al., 2020)

As dimensões x e y do mapa de características são determinadas pelo tamanho do filtro convolucional e pelo número de passos realizados em cada direção. Um filtro maior resulta em menos deslocamentos na imagem, enquanto um maior número de passos por deslocamento diminui o número total de passos. Por sua vez, a dimensão z é definida pelo número de filtros utilizados nas operações convolucionais, pois cada filtro gera uma camada no mapa de características.

Nas CNNs, cada camada tem um número específico de filtros predefinidos aplicados às matrizes de características. As convoluções nas camadas iniciais capturam informações de baixo nível, como bordas, cor e gradiente. À medida que mais camadas são adicionadas, a rede consegue extrair informações de nível mais alto. Cada filtro é projetado para calcular características distintas. Portanto, quanto mais filtros são empregados em cada camada, maior é a variedade de características extraídas. Isso permite que a rede aprenda representações cada vez mais complexas e abstratas dos dados de entrada ao longo das diferentes camadas.

Em certas CNNs, é comum ser necessário reduzir as dimensões do mapa de atributos para manter apenas as informações relevantes. Esse processo é realizado através das operações de *pooling*, que subdividem o mapa de atributos em um número $2n$ de regiões, e então, extraem valores dessas regiões para criar um novo mapa de atributos com dimensões $n \times n$. As duas operações de *pooling* mais comuns são: *max-pooling* e *average-pooling*.

Na operação de *max-pooling*, o mapa de características é subdividido em regiões, e o maior valor presente em cada região é mapeado para formar um novo mapa. Esta operação pode ser ilustrada considerando um mapa inicial de dimensões 4×4 , dividido em 4 regiões, e os maiores valores de cada região são então utilizados para construir um novo mapa de dimensões 2×2 . Esse procedimento é repetido para cada conjunto de regiões, conforme especificado por $n = 2$, resultando em um mapa final de dimensões 2×2 . A operação de *max-pooling* pode ser observado na Figura 6.

Na etapa de *average-pooling*, o mapa de características é igualmente dividido em um número determinado de regiões, e a média dos valores em cada região é calculada para gerar os valores do novo mapa. Considerando um exemplo em que um mapa inicial de dimensões 4×4 é processado, a aplicação do *average-pooling* resultaria em um novo mapa de características de tamanho 2×2 , onde cada valor é a média dos valores de uma região específica do mapa original. A operação de *average-pooling* pode ser observado na Figura 7.

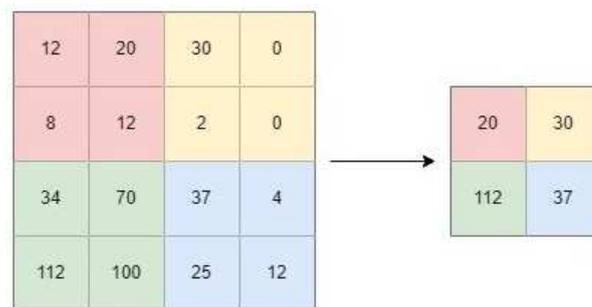


Figura 6 – Representação gráfica da operação de *max-pooling* em uma matriz com dimensões de 4×4 .

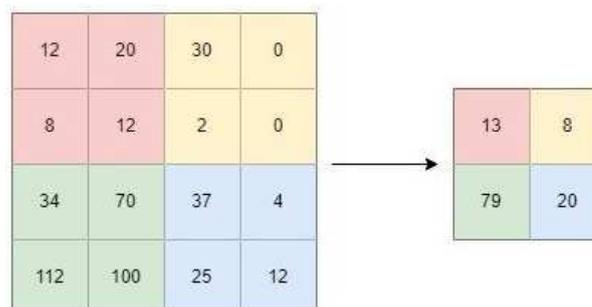


Figura 7 – Representação gráfica da operação de *average-pooling* em uma matriz com dimensões de 4×4 .

Por último, nas CNNs, são incluídas camadas totalmente conectadas para a etapa de classificação dos objetos em categorias pré-estabelecidas. Essas camadas são denominadas totalmente conectadas devido ao fato de que todos os neurônios de uma camada estão conectados a todos os neurônios da camada subsequente. Na camada final, é comum utilizar a função de ativação *softmax*. Essa função recebe como entrada os resultados dos neurônios da camada anterior e calcula as probabilidades de uma imagem pertencer a cada uma das classes definidas durante o treinamento da RNC. Dessa forma, a camada de saída fornece a distribuição de probabilidade

sobre as possíveis classes para uma determinada entrada, permitindo assim que a rede faça suas previsões com base nessas probabilidades.

3.8 Operações Morfológicas

A palavra "Morfologia" é um termo originário do campo da biologia, onde é empregado para estudar as formas e estruturas de animais e plantas. No âmbito da Morfologia Matemática, essa palavra assume um novo significado, sendo utilizada como uma ferramenta para extrair componentes relevantes de imagens. Esses componentes visuais, como bordas, esqueletos, realces e segmentações, são fundamentais para representar e descrever a forma de uma região em uma imagem (GONZALEZ; WOODS, 2009).

As operações morfológicas encontram aplicação em uma variedade de tipos de imagens, mas seu uso mais comum ocorre em imagens binárias, onde os pixels assumem valores de branco e preto. Essas operações são essenciais para processar imagens onde a distinção entre objetos de interesse e o fundo é fundamental. Basicamente, as operações morfológicas envolvem a aplicação de um elemento estruturante A em uma imagem B . Esse elemento estruturante é uma região binária com um tamanho e uma distribuição pré-definidos de elementos. Ao deslocar o elemento estruturante sobre a imagem, ocorre sua aplicação. As principais operações morfológicas incluem a dilatação, que expande regiões de interesse, e a erosão, que contrai essas regiões. Essas operações desempenham um papel fundamental em muitas aplicações de processamento de imagem, como segmentação e reconhecimento de objetos. A morfologia matemática é considerada um pouco à parte no que diz respeito a análise de imagens, pois é baseada em operadores não lineares que operam sobre a forma do objeto (SOLOMON; BRECKON, 2013; GONZALEZ; WOODS, 2009; SILVA et al., 2020).

A dilatação dos conjuntos A e B é determinada pela operação de adição de Minkowski, expressa como (SOLOMON; BRECKON, 2013; GONZALEZ; WOODS, 2009; SILVA et al., 2020):

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\}, \quad (1)$$

$(\hat{B})_z$ denota a reflexão de A em relação à origem, seguida de uma translação por z , e o símbolo \ominus é chamado de "operador morfológico de erosão". É comumente utilizado em operações de processamento de imagem, especialmente na morfologia matemática, para reduzir ou erodir os contornos dos objetos em uma imagem binária.

A interseção é a operação que identifica os elementos compartilhados entre a imagem e o elemento estruturante. O elemento estruturante é movido sobre a imagem, verificando se seu ponto de origem coincide com um pixel da imagem. Se o pixel tiver valor 1, a região correspondente na imagem é preenchida com o elemento estruturante. Se o pixel tiver valor 0, a imagem permanece inalterada nessa região.

A erosão entre uma imagem A e um elemento estruturante B é definida pela operação de subtração Minkowski, representada por (SOLOMON; BRECKON, 2013; GONZALEZ; WOODS, 2009; SILVA et al., 2020):

$$A \ominus B = \{z | (\hat{B})_z \cap A^c \neq \emptyset\}, \quad (2)$$

Onde A^c denota o complemento do conjunto A . O processo de erosão implica em deslocar o elemento estruturante sobre a imagem. Durante esse deslocamento, é crucial verificar se todos os elementos do elemento estruturante estão presentes na região do objeto. Se essa condição for atendida, os pixels na imagem que coincidem com o elemento estruturante são removidos, deixando apenas a origem do elemento intacta. Um exemplo prático de erosão pode ser visualizado na Figura 8, onde um elemento estruturante em forma de cruz de dimensões 3×3 é aplicado à imagem. Quando o elemento estruturante está completamente contido dentro da imagem, apenas a origem é mantida, enquanto o restante dos elementos é removido.

3.9 Segmentação

Durante a fase de segmentação de núcleos displásicos foi implementado um método de segmentação por instância de objetos baseado na rede neural Mask R-CNN. A Mask R-CNN é uma rede neural convolucional orientada por regiões, projetada para analisar agrupamentos de pixels semelhantes organizados em regiões para fins de classificação.

3.9.1 Aplicação da Arquitetura ResNet50 e Mask R-CNN na Segmentação de Núcleos em Tecido Histológico

A ResNet (Rede Residual) representa uma arquitetura consagrada de rede convolucional, frequentemente utilizada como estrutura principal em diversas tarefas de visão computacional. Este modelo foi distinguido como vencedor do desafio ImageNet em 2015. Sua contribuição revolucionária reside na capacidade de treinar redes neurais excepcionalmente profundas, com mais de 150 camadas. Antes do surgimento da ResNet, a construção de redes neurais profundas era uma tarefa desafiadora devido ao problema dos gradientes desaparecidos: após um certo número de camadas, adicionar mais camadas não resultava em melhorias significativas na qualidade do modelo. A ResNet superou essa limitação, introduzindo blocos residuais que permitem o treinamento eficiente de redes extremamente profundas, mantendo a capacidade de aprendizado e evitando a degradação do desempenho à medida que a profundidade da rede aumenta. Essa abordagem inovadora revolucionou o campo das redes neurais convolucionais, abrindo caminho para avanços significativos na capacidade de representação e generalização de modelos de aprendizado profundo (HE et al., 2015).

Nesta arquitetura, a ideia central é que as camadas continuem a receber não apenas os valores resultantes das funções de ativação ReLu (Unidade Linear Retificada) - uma função não linear simples que retorna zero para valores negativos e mantém o valor de entrada para valores positivos ou zero ($F(x)$) - da camada anterior, mas também recebam diretamente os valores de entrada (x) dessas funções. Isso pode ser visualizado na Figura 8. Essa abordagem permite que as camadas subsequentes tenham acesso direto tanto aos valores transformados pela função

de ativação quanto aos valores de entrada originais, possibilitando uma melhor propagação e utilização das informações ao longo da rede (HE et al., 2015).

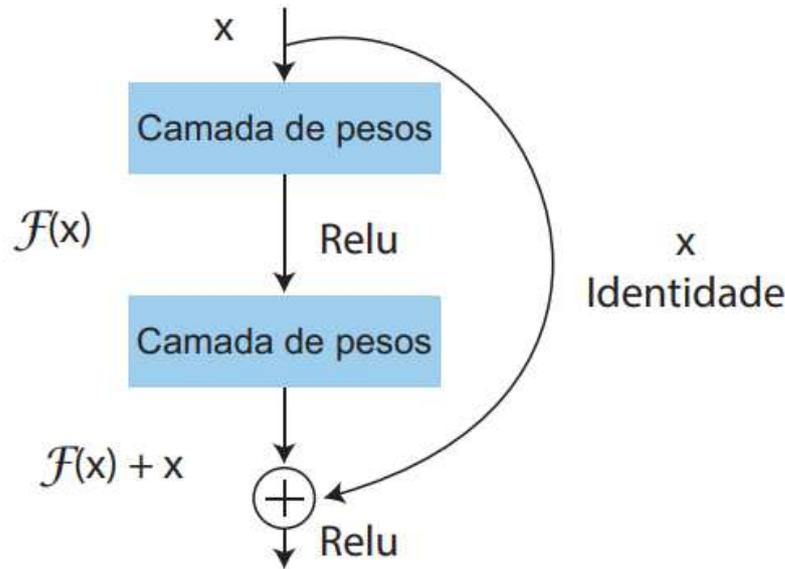


Figura 8 – Unidade Residual do Modelo ResNet. (HE et al., 2015)

Neste estudo, a rede convolucional ResNet50 foi adotada como a espinha dorsal para a extração de características dos núcleos celulares. A estrutura dessa rede está representada na Figura 11. A arquitetura utilizada é composta por uma camada convolucional de entrada, seguida por 16 conjuntos de camadas convolucionais organizadas em blocos 1, 2, 3 e 4, culminando em uma camada de classificação. Na camada inicial, são empregados 64 filtros convolucionais de tamanho 7x7 pixels, operando sobre a imagem original com uma janela deslizante de 2 pixels de deslocamento. Posteriormente, uma etapa de subamostragem é realizada por meio de um filtro de max pooling de tamanho 3x3 pixels, com um deslocamento de 2 pixels. Após a operação de max pooling, o Bloco 1 (B1) é aplicado, consistindo em três camadas convolucionais. A primeira camada do bloco contém 64 filtros convolucionais de tamanho 1x1 pixels, a segunda camada possui 64 filtros de tamanho 3x3 pixels, enquanto a terceira camada apresenta 256 filtros de tamanho 1x1 pixel. Este bloco é repetido três vezes, totalizando 9 camadas convolucionais (HE et al., 2015; SILVA et al., 2022a).

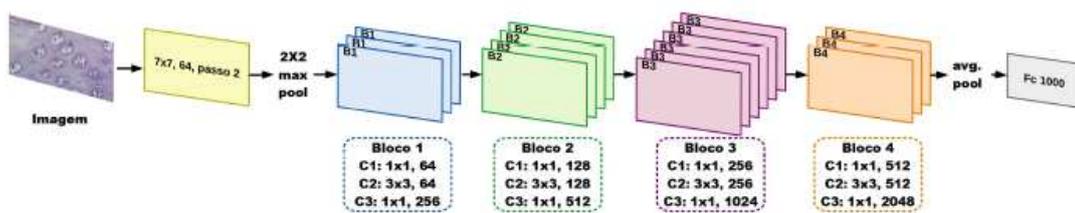


Figura 9 – Diagrama representando a utilização da arquitetura da rede neural ResNet50 na fase de segmentação dos núcleos do tecido histológico (SILVA et al., 2022a).

Após a aplicação do Bloco 1 (B1), prosseguimos com a utilização do Bloco 2 (B2), composto por três camadas distintas. Nessa etapa, cada camada do Bloco 2 contém uma sequência espe-

cífica de filtros convolucionais. Inicialmente, a primeira camada do Bloco 2 contém 128 filtros de tamanho 1x1 pixel, seguida por uma camada com 128 filtros de tamanho 3x3 pixels e, por fim, uma camada com 512 filtros de tamanho 1x1 pixel. Esse processo é repetido quatro vezes, totalizando 12 camadas convolucionais para o Bloco 2. Em seguida, avançamos para a aplicação do Bloco 3 (B3), também composto por três camadas. Na primeira camada desse bloco, são utilizados 256 filtros de tamanho 1x1 pixel, seguidos por 256 filtros de tamanho 3x3 pixels na segunda camada e, por fim, 1024 filtros de tamanho 1x1 pixel na terceira camada. Essa sequência é repetida seis vezes, resultando em um total de 18 camadas convolucionais para o Bloco 3. Na fase seguinte, ocorre uma transição para o Bloco 4 (B4), que também possui três camadas. Na primeira camada deste bloco, são aplicados 512 filtros de tamanho 1x1 pixel, seguidos por 512 filtros de tamanho 3x3 pixels na segunda camada e, finalmente, 2048 filtros de tamanho 1x1 pixel na terceira camada. Esse bloco é repetido três vezes, totalizando 9 camadas convolucionais. Após o processamento pelos blocos convolucionais, os dados passam por uma camada de média de agrupamento, resultando em um vetor de características. Por fim, esse vetor é inserido em uma camada totalmente conectada com 1000 neurônios, onde é aplicada a função softmax para classificar os objetos entre as classes de núcleo ou região de fundo da imagem (HE et al., 2015; SILVA et al., 2022a).

Durante a transição entre os blocos B1, B2, B3 e B4, é necessário reduzir o tamanho da matriz de características em uma proporção de 2. Para realizar essa redução, empregamos uma técnica que utiliza uma janela deslizante com um deslocamento de 2 pixels durante as convoluções de transição. Um exemplo claro desse processo pode ser observado após a aplicação dos 256 filtros convolucionais do último B1, onde os 128 filtros de tamanho 1x1 do primeiro B2 operam com uma janela deslizante com esse deslocamento específico de 2 pixels. Dessa forma, as convoluções de tamanho 1x1 desempenham um papel crucial na redução e na restauração das dimensões dos dados, permitindo que as convoluções 3x3 operem com dimensões de entrada e saída menores. Essa estratégia contribui para a eficiência do processo de transição entre os blocos, garantindo uma progressão suave no processamento das características ao longo da rede convolucional (HE et al., 2015; SILVA et al., 2022a).

Como parte de sua natureza residual, a saída de cada bloco de camadas é adicionada à saída do bloco seguinte, produzindo um mapeamento residual. Esse mapeamento captura a discrepância observada nessa camada em todas as operações convolucionais. Assim, a cada trio de camadas convolucionais, é aplicada a seguinte equação:

$$y = f(x, W_i) + x, \quad (3)$$

Nesta equação, x e y denotam os vetores de saída e entrada das camadas, respectivamente. A função $f(x, W_i)$ simboliza o mapeamento residual a ser aprendido, permitindo assim a prevenção da degradação na precisão da rede.

3.9.1.1 Mask R-CNN

A rede *Mask R-CNN*, abreviação de *Mask Region-based Convolutional Neural Network*, é uma extensão do algoritmo *Faster R-CNN*, que por sua vez é uma evolução do *R-CNN*. O *Mask*

R-CNN é uma arquitetura de rede neural convolucional profundamente enraizada na detecção e segmentação de objetos em imagens (HE et al., 2015).

Uma das principais características do *Mask* R-CNN é sua capacidade de segmentar os objetos detectados, ou seja, não apenas identificar a presença de objetos em uma imagem, mas também delinear com precisão suas formas exatas, gerando máscaras detalhadas para cada objeto identificado, conforme pode ser observado na Figura 10.

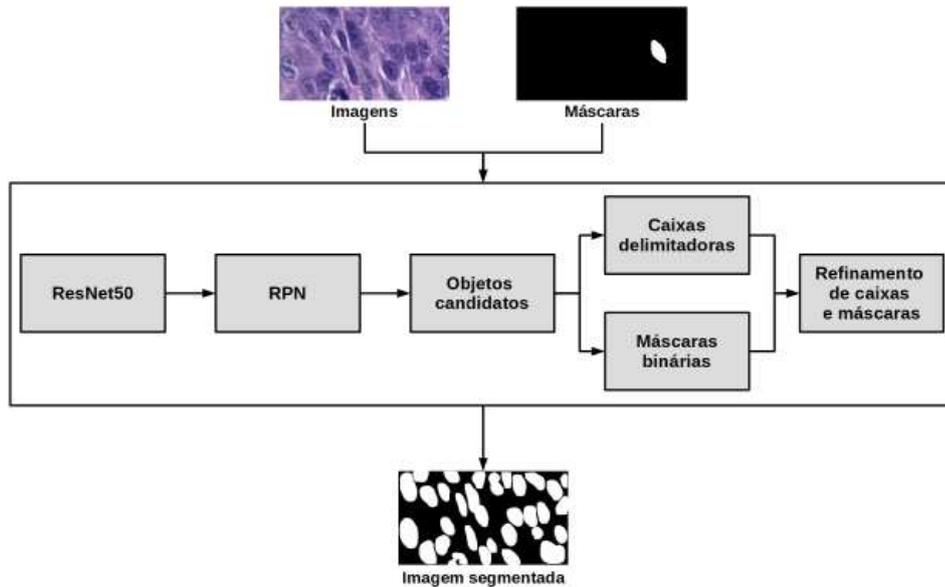


Figura 10 – Estrutura da rede *Mask* R-CNN para classificação dos dados coletados (SILVA et al., 2022a).

A fase de segmentação neste método compreende três etapas distintas. As duas primeiras etapas são parte de uma rede denominada *Region Proposal Network* (RPN), a qual emprega a ResNet 50 para a detecção de objetos na imagem, conforme ilustrado na Figura 10. Para isso, utiliza-se as camadas convolucionais da ResNet 50 para criar uma pirâmide de características conhecida como *Feature Proposal Network* (FPN). As camadas da FPN contêm informações de diferentes níveis de abstração da imagem, que são aproveitadas para identificar os objetos presentes, utilizando o cálculo de similaridade entre regiões.

Na segunda fase, uma técnica de varredura foi aplicada às regiões sugeridas pela RPN usando uma matriz de dimensão 3×3 para extrair um conjunto de características. Essas características foram então alimentadas em duas camadas totalmente conectadas, uma para a regressão das caixas delimitadoras e outra para a classificação das regiões. A cada passo da varredura, várias regiões candidatas (k) foram processadas simultaneamente. A camada de regressão teve uma saída de dimensão $4k$, que representava as coordenadas x e y do canto superior esquerdo, juntamente com a altura h e a largura w de cada região k . Enquanto isso, a camada de classificação gerou uma saída de dimensão $2k$, abordando a probabilidade de a região pertencer à classe de núcleo e à classe de fundo. Na Figura 11, pode se observado uma representação das caixas delimitadoras que foram geradas.

Na terceira fase, a rede gera uma máscara binária para cada objeto identificado na imagem. Para isso, utiliza-se uma Rede Totalmente Convolutiva (do inglês, *Fully Convolutional Network*

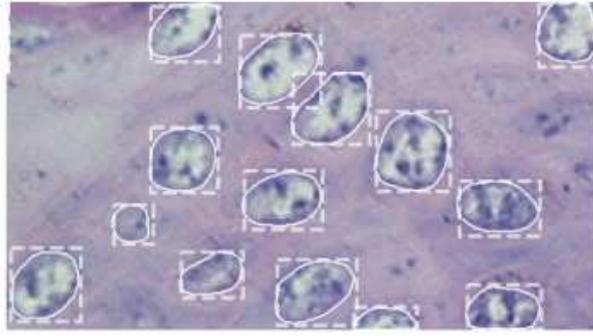


Figura 11 – Demonstração das caixas delimitadoras criadas para identificar os núcleos de um tecido, utilizando *Mask R-CNN* (SILVA et al., 2022a).

- FNC) em conjunto com os mapas de características. Essa abordagem permite estimar uma máscara de tamanho $m \times n$ para cada núcleo identificado na imagem, preservando o formato espacial em relação à imagem original. Um exemplo dessas máscaras geradas pode ser visualizado na Figura 11.

Durante o treinamento da rede, calcula-se uma taxa de perda multitarefa para cada região. A função de perda utilizada é especificada por:

$$L = L_{cls} + L_{box} + L_{mask}. \quad (4)$$

Nesse contexto, tem-se que $L_{cs} = -\log p_u$, em que p_u é a probabilidade de ocorrência do padrão-ouro u relativo a cada uma das $K + 1$ classes.

A taxa de perda das caixas delimitadores foi definida por:

$$L_{box}(t^u, v) = \sum_{i \in (x, y, w, h)} L_1(t_i^u, v_i), \quad (5)$$

onde:

$$L_1(x) = \begin{cases} 0.5x^2, & \text{se } |x| < 1, \\ |x| - 0.5, & \text{c.c.} \end{cases} \quad (6)$$

T_u representa a regressão das caixas, v é a caixa delimitadora calculada com base no padrão de referência, x e y são as coordenadas do canto superior esquerdo de cada objeto, e os parâmetros h e w denotam a altura e a largura da região, respectivamente. Para este estudo (SILVA et al., 2022a), a função de perda L_{mask} adotada é a média binária de entropia cruzada (HE et al., 2015).

Durante o processo de treinamento conduzido em (SILVA et al., 2022a), foram utilizadas dez imagens para cada categoria de displasia. Com a colaboração de um patologista, foram criadas máscaras binárias para cada núcleo identificado nas 40 imagens, totalizando 1220 máscaras individuais de núcleos. A rede foi, inicialmente, treinada com o conjunto de dados ImageNet, o qual abrange milhões de imagens, incluindo amostras histológicas. Posteriormente, empregou-se o método de transferência de aprendizado, realizando um segundo treinamento no qual os pesos de todas as camadas foram ajustados com base nos resultados do treinamento com o ImageNet. Esse procedimento foi conduzido utilizando o modelo do segmentador por instância, com 40 épocas e 150 iterações por época, conforme determinado por experimentações empíricas com

o conjunto de dados. A taxa de aprendizado adotada foi de 0,001, utilizando um otimizador *Stochastic Gradient Descent* (SGD) com um momentum de 0,9, em conformidade com a proposta de (HE et al., 2015). Esse modelo de rede para segmentação foi avaliado nos estudos de (SILVA et al., 2022b) e (SILVA et al., 2024). No trabalho de (SILVA et al., 2024) outros modelos foram comparados e a Mask R-CNN trouxe resultados relevantes para o processo de segmentação assim como também contribuições nas etapas subsequentes de extração de características para classificação.

3.10 Pós Processamento

Em (SILVA et al., 2022a), após a obtenção das imagens binárias para cada núcleo, é comum ocorrerem imperfeições nos contornos, áreas não completamente preenchidas e pequenos artefatos que podem afetar a precisão da segmentação. Com o intuito de corrigir essas irregularidades e aprimorar os núcleos identificados, empregaram-se operações morfológicas durante a fase de pós-processamento. Inicialmente, buscou-se completar os contornos dos núcleos e eliminar áreas de falsos-negativos, isto é, preencher os núcleos. Para isso, recorreu-se à técnica de dilatação morfológica, a qual expande os objetos em uma imagem binária, preenchendo os espaços vazios e reforçando as linhas e contornos das bordas.

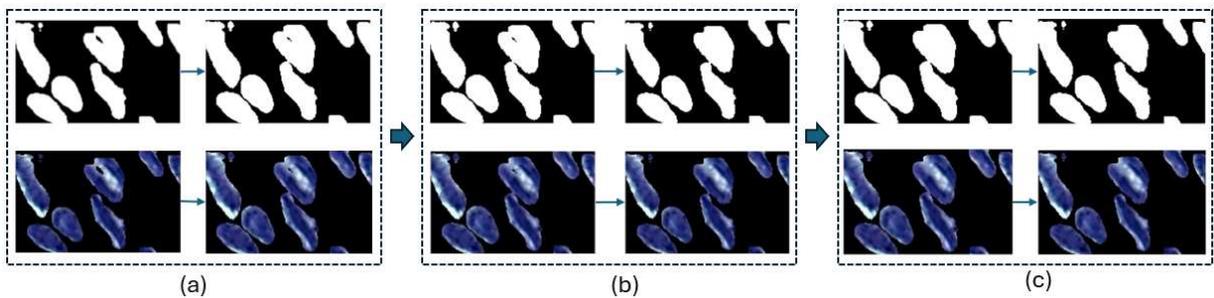


Figura 12 – Processo de segmentação: (a) Operação de dilatação binária, (b) Operação de preenchimento de regiões, (c) Operação de erosão binária. (SILVA et al., 2022a)

Após a etapa anterior, é possível que permaneçam áreas de falsos-negativos, como núcleos não completamente preenchidos, nas imagens. Portanto, para lidar com essa situação, foi empregada uma função de preenchimento de regiões. Essa função identifica regiões de contorno fechadas e atribui a todos os pixels dentro delas o mesmo valor de intensidade do contorno. Além disso, um filtro de erosão, utilizando um elemento estruturante em forma de cruz com dimensões 3×3 pixels, foi aplicado para remover ruídos das imagens e restaurar o tamanho original dos núcleos. Após testes empíricos, constatou-se que aplicar o filtro de erosão após o preenchimento de regiões resultou em menos falsos-negativos do que aplicar antes do preenchimento. Por fim, todos os objetos na imagem com área menor que 30 pixels foram rotulados como região de fundo, determinado empiricamente como um tamanho de área adequado para eliminar falsos-positivos e manter pequenos núcleos. Os resultados de cada etapa conduzida por (SILVA et al., 2022a) estão representados na Figura 12.

3.11 Extração de Características

Essa seção traz o processo de extração de características em modelos baseados em técnicas de engenharia de características e descritores baseados em mapas de características obtidos de CNNs.

3.11.1 Extração de Características em Imagens Segmentadas

O reconhecimento de padrões em imagens é altamente dependente do nível de detalhamento das características conhecidas de cada objeto. Assim, a extração de informações torna-se uma fase essencial no processo de reconhecimento ou análise de similaridade (KHODASKAR; LADHAKE, 2014).

As características morfológicas são atributos quantitativos que descrevem a forma, tamanho e estrutura de objetos em uma imagem. Essas características são frequentemente utilizadas em análises de imagem e processamento de visão computacional para extrair informações úteis sobre objetos e regiões de interesse. Em (SILVA, 2019), as análises dos tecidos para definir os níveis de displasia envolveram a obtenção de características morfológicas e não morfológicas das imagens. Para as características morfológicas foram obtidos os atributos: Área (A), Excentricidade (E), Orientação (O_R), Perímetro (P_E) e Solidez (S). Além disso, foram consideradas as características não morfológicas, como Entropia (E_N) e Índice de Moran (I_M), que exploram informações internas das regiões em imagens histológicas.

A métrica de (A) é uma medida quantitativa que descreve a extensão ou tamanho de um objeto em uma imagem. Ela é calculada contando o número de pixels que compõem a região delimitada pelo objeto. Quanto maior a área, maior é o tamanho do objeto na imagem. Essa métrica é amplamente utilizada em análises de imagem e processamento de visão computacional para caracterizar e quantificar objetos de interesse. Por exemplo, em análises médicas, a área pode ser usada para determinar o tamanho de uma lesão ou tumor em uma imagem de exame de imagem. Em aplicações de reconhecimento de padrões e classificação de objetos, a (A) pode ser uma característica importante para distinguir entre diferentes tipos de objetos com base em seu tamanho relativo. Essa métrica é definida pela Equação 7:

$$A = \sum_{i=1}^N A_i, \quad (7)$$

onde A_i é a área de cada pixel dentro do objeto e N é o número total de pixels na região de interesse.

A (E) é uma medida que descreve a forma de um objeto em relação a uma forma geométrica ideal, geralmente uma circunferência. Sua equação é expressa por:

$$E = \frac{a}{b}, \quad (8)$$

onde a representa a distância entre o centro do objeto e um ponto fixo, enquanto b é a distância do centro para um vértice. A métrica de orientação (O_R) determina a relação entre o maior eixo

de um objeto e um eixo de referência, fornecendo informações sobre a direção predominante do objeto em relação a esse referencial. Sua equação é expressa por:

$$O_R = \arctan \left(\frac{D - B + \sqrt{(D - B)^2 + C^2}}{C} \right), \quad (9)$$

na expressão acima, B , C e D denotam, respectivamente, a variância dos pixels ao longo do eixo y , do eixo x e do eixo principal. A métrica de (P_E) é uma medida importante que quantifica o comprimento da borda de um objeto em uma imagem. Essa medida é fundamental em operações morfológicas e análise de imagens, pois fornece informações sobre a forma e a estrutura dos objetos. Em essência, (P_E) representa a extensão da fronteira entre os pixels de objeto e os pixels de fundo em torno do objeto. Sua equação é expressa por:

$$P_E = 2\pi \sqrt{\frac{c^2 + d^2}{2}}, \quad (10)$$

de tal forma que c e d são os eixos maior e menor.

A métrica (S) é uma medida que quantifica a regularidade ou a "solidez" da forma de um objeto em relação à sua área convexa. A área convexa é o menor polígono convexo que contém completamente o objeto. A (S) é calculada como a razão entre a área do objeto e a área de seu polígono convexo circundante.

A métrica de solidez é definida como a proporção entre a área do objeto e um polígono que abrange sua região convexa, ou seja, o menor polígono que abrange a área do objeto. Quanto mais próximo o valor da solidez estiver de 1, mais regular é a forma do objeto, indicando que o objeto preenche bem seu envelope convexo. Por outro lado, valores de (S) menores do que 1 indicam formas mais irregulares ou fragmentadas. Sua equação é expressa por:

$$S = \frac{A}{R_a}, \quad (11)$$

onde A representa a área do núcleo e R_a é a região convexa que abrange o objeto.

As medidas não morfológicas são características que descrevem objetos em uma imagem de forma quantitativa, mas não estão relacionadas diretamente à estrutura ou forma dos objetos. Entre esses tipos de métricas, tem-se a entropia. A entropia é uma medida que avalia a variação nos níveis de intensidade dentro de uma região da imagem, oferecendo *insights* sobre a textura da área considerada. Essa métrica é particularmente útil para caracterizar a complexidade da textura, pois quantifica o grau de incerteza ou desordem nos níveis de intensidade dos pixels. Quanto maior a entropia, maior a variabilidade na textura da região. Isso significa que áreas com alta entropia tendem a ter uma textura mais complexa, com variações significativas na intensidade dos pixels. Por outro lado, regiões com baixa entropia têm uma textura mais homogênea, com menos variações na intensidade dos pixels. Essa medida é fundamental em várias aplicações de processamento de imagem, como análise de textura, segmentação de regiões e reconhecimento de padrões. Sua equação é expressa por:

$$E_N = - \sum_{i=1}^n p_i \log_2 p_i, \quad (12)$$

na fórmula, cada p_i indica a probabilidade associada à ocorrência do nível de intensidade i , enquanto n representa a quantidade total de tons de cinza contidos na imagem. Para extrair os tons de cinza das imagens, (SILVA, 2019) utilizou uma conversão para o modelo de cores HSL. Durante esse processo, os canais de matiz(h) e saturação foram descartados, preservando apenas o canal de brilho, que é crucial para representar os diferentes níveis de intensidade das imagens. Em (SILVA, 2019) a extração de entropia foi realizada utilizando sete tamanhos de vizinhança: 3x3, 5x5, 7x7, 9x9, 11x11, 13x13 e 15x15 pixels.

O *Índice de Moran*, por outro lado, é uma medida que avalia a autocorrelação espacial, considerando o desvio da intensidade de um pixel em relação à média da região circundante. Isso significa que ele analisa como a variação de intensidade de um pixel está relacionada à média das intensidades dos pixels vizinhos. A Equação 13 representa essa métrica:

$$I = \frac{N \sum_i \sum_j w_{ij} Z_i Z_j}{w \sum_i Z_i^2}, \quad (13)$$

onde N é o número de pixels, w_{ij} representa a vizinhança de pixel ij , W é a soma de todos os w_{ij} e Z_i e Z_j representam os desvios dos pixels z_i e z_j em relação à média z , sendo calculado por $Z_k = z_k - z$ (SILVA, 2019).

Em (SILVA, 2019), foram extraídas essas características apresentadas anteriormente individualmente para cada núcleo n em uma imagem I_k , onde k representa o número da imagem. Cada imagem I_k é representada por uma matriz de tamanho $13 \times n$. Em seguida, medidas estatísticas de média (M) e desvio padrão (D) foram calculadas para cada característica dentro de I , resultando em dois vetores, M_k e D_k , totalizando 13 descritores cada grupo. Esses vetores foram concatenados para formar um vetor obtido por métodos baseados em engenharia de características, $I_k = [M_k, D_k]$, utilizado na etapa de análise de similaridade das imagens.

3.11.2 Extração de Características com Redes Neurais Convolucionais

As Redes Neurais Convolucionais são empregadas na extração de características de imagens histológicas podendo gerar descritores relevantes até mesmo a partir de imagens que não foram segmentadas. Ao contrário de métodos tradicionais que demandam a identificação prévia de regiões de interesse por estratégias de testes, as CNNs são capazes de aprender automaticamente padrões e características diretamente a partir dos dados brutos da imagem. Em (SILVA et al., 2022a), a obtenção das características das imagens foi realizada por meio da aplicação de dois modelos CNNs. Nos experimentos, foram empregados as camadas de convolução, de agrupamento máximo e precedendo a camada de achatamento. As saídas resultantes do processo de achatamento foram armazenadas em vetores de características. Para treinamento dos modelos foi empregado a técnica de transferência de aprendizagem com os pesos extraídos da base ImageNet.

Além do modelo ResNet, um dos modelos utilizados foi AlexNet, o qual foi desenvolvido por (KRIZHEVSKY; SUTSKEVER; HINTON, 2017), considerado o predecessor da CNN e um impulsionador em aprendizagem profunda. Na história do concurso *ImageNet Large Scale Visual*

Recognition Challenge (ILSVRC), AlexNet definitivamente ganhou popularidade. O ImageNet contém cerca de 22.000 categorias e mais de 15 milhões de fotos marcadas em alta resolução. Os acadêmicos desejam testar seus modelos de classificação de imagens devido à alta qualidade do conjunto de dados ImageNet. Em 2012, com uma ampla margem de aproximadamente 11,3%, o AlexNet tornou-se a primeira arquitetura de aprendizagem profunda a vencer o desafio ImageNet.

O modelo AlexNet é composto por cinco camadas convolucionais intercaladas com três operações de pooling, seguidas por duas camadas totalmente conectadas e uma função softmax (Figura 13).

Assim, o modelo ResNet-50 consiste em uma camada de entrada seguida por 48 camadas convolucionais organizadas em blocos repetidos, apresentando 16 conexões de salto e finalizando com a camada de classificação (Figura 9) (SILVA et al., 2022a). As camadas iniciais dessas arquiteturas são responsáveis por extrair atributos que quantificam as bordas, formas e cores dos objetos nas imagens. Além disso, as camadas finais são usadas para identificar padrões de textura globais. As cinco camadas convolucionais do modelo AlexNet, excluindo a softmax e as camadas totalmente conectadas, forneceram 616.032 características. Considerando as duas primeiras camadas e as três últimas camadas do modelo, um total de 1.095.488 características foram obtidos para a arquitetura ResNet-50 (SILVA et al., 2022a).

Nas imagens submetidas às arquiteturas foi feita a extração de características e seleção de características relevantes por meio do algoritmo ReliefF, permitindo a seleção dos 50/40/30/20/10 atributos mais significativos. O algoritmo calcula a diferença entre os valores de atributos de instâncias próximas (vizinhas) e distantes no espaço de atributos, permitindo identificar atributos que são relevantes e que podem contribuir para melhorar a precisão do processo de recuperação de imagens (SILVA et al., 2022a).

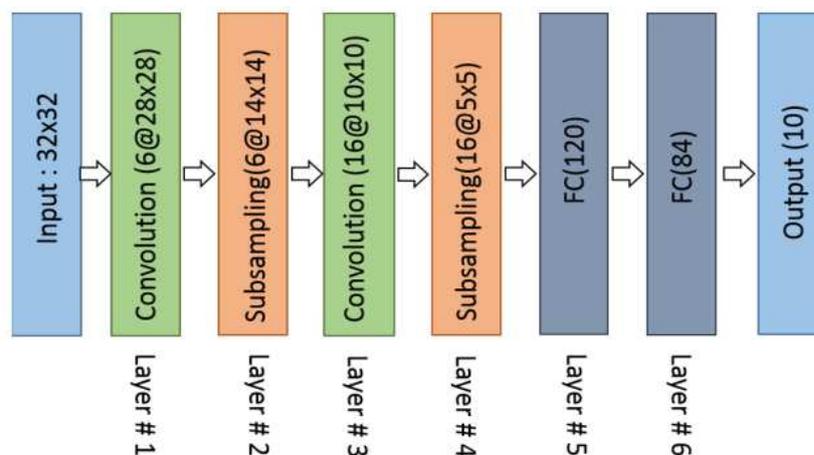


Figura 13 – Arquitetura do modelo AlexNet utilizado para extrair características de imagens de tecidos. (ALOM et al., 2018)

3.12 Processo de Recuperação de Imagens Baseado em Conteúdo

Atualmente, as descrições textuais não são mais necessárias para as técnicas de recuperação de imagens baseadas em conteúdo (*Content-Based Image Retrieval - CBIR*). Estas abordagens podem recuperar imagens específicas de uma coleção significativa de bancos de dados de imagens. A recuperação de imagens no CBIR depende de aspectos do conteúdo, como texturas, cores ou formas, que são características de baixo nível de imagem (ATLAM GAMAL ATTIYA, 2013).

A aplicação da visão computacional ao problema de recuperação de imagens é conhecida como recuperação de imagens baseada em conteúdo, também conhecida como consulta por imagem de conteúdo ou recuperação de informações visuais baseadas em conteúdo (ATLAM GAMAL ATTIYA, 2013).

O sistema CBIR básico é mostrado como um diagrama de blocos na Figura 14. A extração e a correspondência de recursos são os dois estágios do CBIR. A primeira fase consiste em extrair as características da imagem, enquanto a segunda fase consiste em ajustar essas características. Simplesmente, as características de cada imagem armazenada no banco de dados são extraídas e comparadas com as características da imagem da consulta.

A recuperação baseada em conteúdo utiliza descritores (vetores de características) extraídos automaticamente das imagens. Durante uma consulta o usuário fornece uma imagem, e o sistema deve extrair o seu vetor de características que será utilizado para computação da similaridade com os vetores de características das imagens presentes no banco de dados. Uma estrutura de indexação pode ser utilizada para melhorar o desempenho do procedimento, reduzindo o espaço de busca das mais similares, como ilustrado na Figura 14.

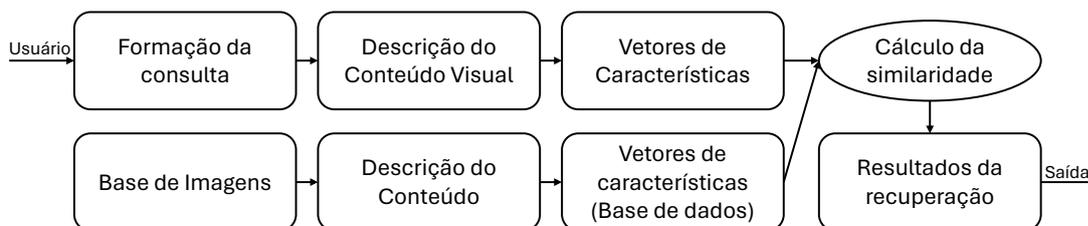


Figura 14 – Fluxo de funcionamento de um sistema CBIR

A similaridade é uma fase do processo de recuperação de imagens por conteúdo. A medida de similaridade e a consulta de similaridade são as duas etapas da etapa de similaridade. Na primeira etapa, chamada de medida de similaridade, as funções são usadas para determinar quanto as imagens são semelhantes entre as representações da imagem de consulta e as representações da imagem do banco. Para cada comparação, é dado um valor e , com base nesse valor, é possível determinar o quão semelhantes são as imagens comparadas. Posteriormente, a segunda rotina é a escolha do operador de consulta a ser usado (PATEL; YADAV; GHOSH, 2020).

Ao longo do processo de recuperação de imagens baseadas em conteúdo, cada imagem é representada como um ponto n -dimensional, onde n é a quantidade de características da imagem. Em seguida, uma medida de similaridade, também conhecida como função de distância, é aplicada para determinar a quantidade de diferença entre as representações das imagens entre si,

resultando em uma lista ranqueada. Um valor real positivo é o resultado da função da distância d entre dois conjuntos de imagens.

De tal forma onde duas imagens são mais similares quanto maior o valor retornado pela função de comparação; caso as imagens sejam idênticas, este valor é igual a zero, e o valor aumenta de acordo com a dissimilaridade entre esses objetos. Em muitos casos, essas funções são normalizadas para resultarem em valores entre 0 e 1, facilitando sua interpretação.

Sendo X um conjunto qualquer e considerando que $s(x, y)$ é a similaridade entre os elementos x e y , então, s deve satisfazer as propriedades:

1. **Identidade:** $s(x, y) = 1 \iff x = y$
2. **Simetria:** $s(x, y) = s(y, x), \forall x, y \in X$.

Medidas de similaridade são amplamente utilizadas em conjuntos de dados que apresentam valores esparsos e podem conter atributos que não seguem uma distribuição simétrica. Essas medidas ignoram as correspondências entre dois valores iguais a zero. Isso mostra conceitualmente que a similaridade de um par de objetos complexos não depende do número de características que os separam, mas sim do número de características que ambos compartilham. A maioria dos objetos tem apenas algumas das características descritas pelos atributos em dados dispersos e assimétricos, portanto, são extremamente semelhantes em termos das características que não têm (TAN; STEINBACH; KUMAR, 2006).

É importante salientar que, tecnicamente, mais de uma métrica pode ser usada para avaliar a similaridade dos vetores de características obtidos por um extrator de características específico; cada métrica pode produzir um resultado distinto.

As métricas da família de Minkowski, também conhecidas como métricas L_p , são usadas em domínios multidimensionais e são amplamente utilizadas. Essas métricas são definidas pela variação do parâmetro $p \in \mathbb{R} | p \geq 1$, sendo x e y vetores em \mathbb{R}^n , ou seja, $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$.

$$d(x, y) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p}, \quad (14)$$

a distância de Manhattan, também chamada de *City Block*, é para $p = 1$, enquanto a distância Euclidiana é para $p = 2$ (SHAHID et al., 2009).

A distância de Mahalanobis, representa a relação de covariância entre os atributos, deste modo, é calculada a matriz de covariância do conjunto. Essa medida de distância é utilizada para calcular a similaridade entre os vetores x e y (MCLACHLAN, 1999):

$$d(x, y) = \sqrt{(x - y)^T V^{-1} (x - y)}, \quad (15)$$

A dissimilaridade pode ser obtida por meio de funções de distância (métricas) ou por funções que não garantem uma ou mais propriedades das métricas, visando melhorar a qualidade semântica (SANTINI; JAIN, 1999). Formalmente, uma métrica é uma função $d : X \cdot X \rightarrow \mathbb{R}$ que satisfaz as propriedades:

1. **Identidade:** $d(x, y) = 0 \iff x = y$
2. **Não-negatividade:** $d(x, y) \geq 0 \forall x, y \in X$
3. **Simetria:** $d(x, y) = d(y, x), \forall x, y \in X$
4. **Desigualdade Triangular:** $d(x, z) \leq d(x, y) + d(y, z) \forall x, y, z \in X$

Dissimilaridade é usado mais frequentemente para descrever a proximidade entre atributos contínuos; além disso, as distâncias fornecem uma maneira precisa de combinar essas diferenças em uma medida geral de proximidade. Mas muitas dissimilaridades não satisfazem uma ou mais das propriedades métricas, como a logistic loss, as divergências Kullback-Leibler (KL) ou Itakura Saito (JIAWEI; MICHELINE, 2006; ROCHA, 2015).

Uma medida conhecida como divergência de Kullback-Leibler, ou simplesmente divergência KL, tem sido amplamente usada na literatura de mineração de dados para calcular a diferença entre duas distribuições de probabilidade sobre a mesma variável x . A teoria da informação e a teoria da probabilidade forneceram as bases para esse conceito (JIAWEI; MICHELINE, 2006; ROCHA, 2015).

A divergência KL é uma medida não simétrica da diferença entre duas distribuições de probabilidade $p(x)$ e $q(x)$. É muito relacionada à entropia relativa, divergência de informação e divergência de informação para discriminação. A distribuição de probabilidade de uma variável aleatória discreta x , é $p(x)$ e $q(x)$. Isso significa que tanto $p(x)$ quanto $q(x)$ somam 1, e para qualquer x em X sera, $p(x) > 0$ e $q(x) > 0$ (JIAWEI; MICHELINE, 2006; ROCHA, 2015). $D_{KL}(p(x)||q(x))$ é definida na Equação (16).

$$D_{KL}(p(x)||q(x)) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)} \quad (16)$$

3.13 Avaliação do Desempenho de um CBIR

Quando um usuário realiza uma consulta por meio do processo de recuperação de imagens, ele espera que o sistema retorne apenas documentos ou imagens que sejam pertinentes aos seus interesses (ou seja, relevantes). A avaliação da eficácia desse processo geralmente se concentra em dois aspectos principais: precisão e revocação. A precisão refere-se à capacidade do sistema de recuperar apenas os documentos relevantes, minimizando a quantidade de documentos irrelevantes apresentados ao usuário. Por outro lado, a revocação, também conhecida como cobertura, mede a capacidade do sistema de recuperar todos os documentos relevantes disponíveis no conjunto de dados, sem deixar nenhum de fora. Enquanto a precisão enfatiza a proporção de documentos relevantes entre todos os documentos recuperados, a revocação considera a proporção de documentos relevantes que foram recuperados em relação ao total de documentos relevantes existentes. Assim, uma avaliação abrangente da eficácia de um sistema de recuperação de imagens deve levar em consideração tanto a precisão quanto a revocação, buscando um equilíbrio entre esses dois aspectos para garantir uma experiência satisfatória ao usuário.

Assim, deve-se saber com antecedência quais documentos o usuário considera relevantes. Considerando uma coleção à qual uma consulta é enviada, vários conjuntos de documentos

estão envolvidos. O conjunto de documentos pertinentes à consulta (conjunto de documentos relevantes R) e o conjunto de documentos recuperados (conjunto de documentos resposta A) são os mais significativos. Neste último, há documentos relevantes e não relevantes, sendo os documentos relevantes recuperados R_a os mais importantes (PEDROSA et al., 2011).

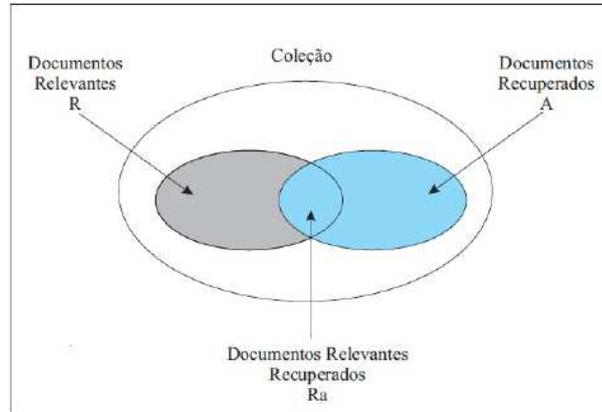


Figura 15 – Representação de um Conjunto de Documentos na Recuperação de Informação. (PEDROSA et al., 2011)

Dada uma consulta x , a precisão é definida como a fração entre o número de imagens relevantes recuperadas, representado por $R_a(x)$, e o número total de imagens recuperadas A :

$$P_r(x) = \frac{R_a(x)}{A}. \quad (17)$$

A revocação é definida com a fração entre o número de imagens relevantes recuperadas, $R_a(x)$, e o número total de imagens relevantes na coleção R :

$$R_e(x) = \frac{R_a(x)}{R}. \quad (18)$$

Essas duas medidas estão ligadas umas às outras, pois dependem da quantidade de documentos pertinentes recuperados, R_a . Em geral, aumentar a revocação reduz a precisão. Isso ocorre porque a quantidade de documentos relevantes recuperados aumenta à medida que documentos adicionais são inseridos na resposta a uma consulta específica. No entanto, documentos não relevantes são inseridos em maior quantidade que documentos relevantes, o que resulta em uma resposta menos precisa (PEDROSA et al., 2011).

A Figura 16 ilustra um gráfico de precisão e revocação, com duas curvas representadas por X e Y . A Figura 16 mostra que as curvas têm comportamentos diferentes, indicando que diferentes algoritmos de recuperação foram usados. Ao examinar a curva X , é evidente que ela apresenta valores altos de precisão em conjunto com níveis baixos de revocação. Isso indica que a busca do usuário retorna as imagens pertinentes nas primeiras posições, o que pode ser intrigante quando apenas as vinte ou trinta primeiras imagens são consideradas importantes.

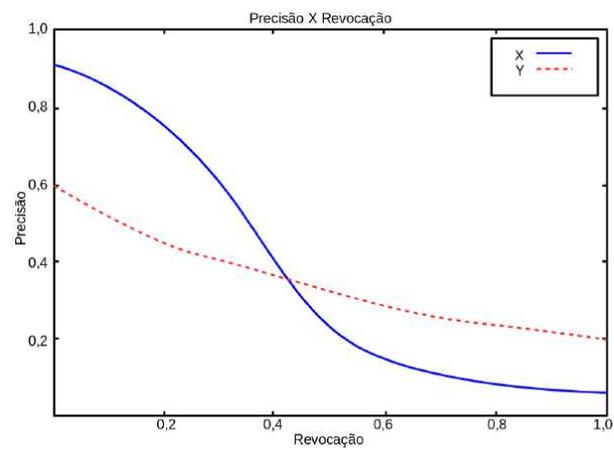


Figura 16 – Representação de um gráfico que ilustra a relação entre precisão e revocação. (ROCHA, 2015)

Metodologia

O método proposto envolve um processo sistemático de segmentação, extração de características e avaliação de métricas, conforme ilustrado na Figura 17. O sistema foi desenvolvido utilizando Python. Para extrair características morfológicas das regiões de interesse nas imagens histológicas, foram empregadas técnicas de segmentação e pós-processamento.

O capítulo também explora a aplicação de diversos métodos para avaliar o desempenho, incluindo a utilização de divergências de Bregman para medir a eficiência dos processos de recuperação. As etapas descritas fornecem uma base sólida para compreender as técnicas e procedimentos empregados no desenvolvimento e análise dos modelos propostos.

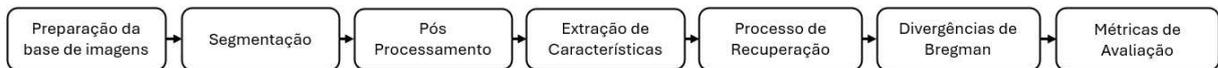


Figura 17 – Fluxograma dos principais estágios empregados para recuperação de imagem por conteúdo em imagens histológicas.

4.1 Divergências de Bregman

No aprendizado de máquina, as divergências de Bregman surgem frequentemente, para agrupamento e otimização, e muitas áreas são essenciais para divergências, como a divergência KL e a distância euclidiana quadrada. O aprendizado envolve muitos problemas que exigem divergências diferentes das distâncias euclidianas. Em situações em que há uma divergência entre duas distribuições, por exemplo, as divergências de Bregman são naturais nesses ambientes (BREGMAN, 1967; BANERJEE et al., 2005; BANERJEE et al., 2004).

Muitas aplicações, como a análise e recuperação de imagens e vídeos, o reconhecimento de fala e a análise de séries temporais, utilizam das divergências de Bregman, que generalizam uma ampla gama de medidas de dissimilaridade não métricas. Isso ocorre porque, diferentemente das medidas métricas que satisfazem as propriedades básicas de um espaço métrico, como não negatividade, simetria e desigualdade triangular, as divergências de Bregman não necessariamente seguem essas propriedades. Apesar disso, as medidas métricas representadas pela distância euclidiana não são coerentes com as percepções humanas de semelhança (SONG et al., 2022).

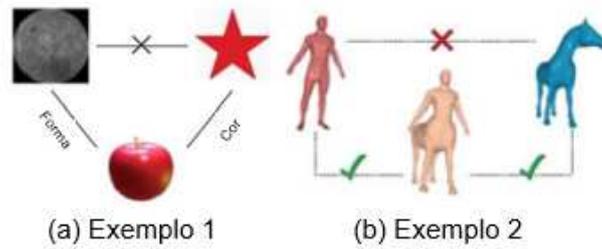


Figura 18 – Representação da similaridade humana (SONG et al., 2022).

Conforme demonstrado no Exemplo 1 da Figura 18, a lua e a maçã compartilham semelhanças em sua forma, enquanto a estrela e a maçã compartilham características de cor. No entanto, não há uma relação de similaridade entre a lua e a estrela. Esse cenário evidencia como nossa percepção de semelhança desafia a noção de desigualdade triangular, indicando uma tendência dos seres humanos em preferir o uso de medidas de dissimilaridade não métricas em vez de métricas, especialmente em conjuntos de dados complexos. Igualmente, conforme ilustrado no Exemplo 2 da Figura 18, tanto o homem quanto o cavalo apresentam semelhanças perceptíveis em sua estrutura, porém claramente se distinguem um do outro. Assim, torna-se inadequado utilizar a distância euclidiana como medida de distância em muitos contextos práticos.

A definição da divergência de Bregman deriva de uma função convexa $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ correspondente à divergência de Bregman entre x e $y \in \text{Dom}(\varphi)$ é expressa pela Equação 19 (SONG et al., 2022), onde $\nabla\varphi(y)$ é o vetor gradiente de φ em y (FERREIRA; BACKES; BARCELOS, 2015):

$$d_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x - y \rangle. \quad (19)$$

A Figura 19 mostra essa definição, embora com pouco rigor matemático.

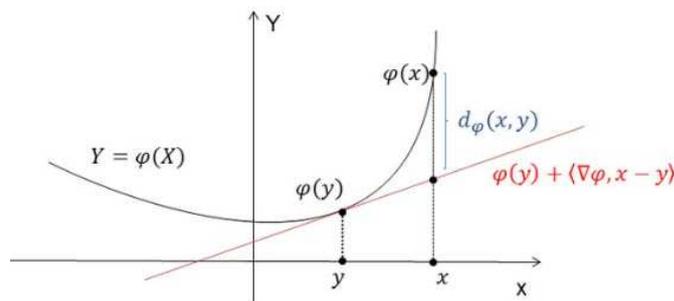


Figura 19 – Representação geométrica da Divergência de Bregman no \mathbb{R}^2 (FERREIRA; BACKES; BARCELOS, 2015).

A soma das funções geradoras unidimensionais aplicadas em cada coordenada de $x \in \mathbb{R}^d$ pode ser usada para construir divergências entre x e y pertencentes a $X \subset \mathbb{R}^d$ tal que $d > 1$ (FERREIRA; BACKES; BARCELOS, 2015). Assim,

$$\varphi(x) = \sum_{i=1}^d \phi(x_i). \quad (20)$$

Quando diferentes funções convexas são empregadas, as distâncias de Bregman definem várias funções de distância conhecidas, entre elas (SONG et al., 2022):

1. **Distância de Mahalanobis:** $\varphi(x) = \frac{1}{2}x^T Q x \rightarrow d_\varphi(x, y) = \frac{1}{2}(x - y)^T Q (x - y)$.

Q representa a matriz de covariância inversa do conjunto de dados, também conhecida como a matriz de precisão. Esta matriz ajusta a escala e a orientação dos dados, considerando as correlações entre os atributos. Na equação da distância de Mahalanobis, Q permite calcular a distância ponderada entre os vetores x e y , levando em conta a estrutura interna dos dados através de suas relações de covariância.

2. **Divergência de Kullback-Leibler:** Quando $\varphi(x) = \sum_{j=1}^d x_j \log_2 x_j \rightarrow d_\varphi(x, y) = \sum_{j=1}^d x_j \log_2 \left(\frac{x_j}{y_j}\right)$. A divergência KL é uma medida comum usada para comparar distribuições.

A seguir estão algumas propriedades relevantes das divergências de Bregman (BANERJEE et al., 2005; BANERJEE et al., 2004):

1. **Não negatividade:** $d_\varphi(x, y) \geq 0, \forall x \in S, y \in ri(S)$, e $d_\varphi(x, y) = 0$
2. **Convexidade:** $d_\varphi(x, y)$ é convexo no primeiro argumento, mas não necessariamente no segundo. A distância *Squared Loss* e *KL-Divergence* são exemplos de divergências de Bregman que são convexas em seus dois argumentos.
3. **Linearidade:** A divergência de Bregman é um operador linear, ou seja $\forall x \in S, y \in ri(S)$,

$$\begin{aligned} d_{\varphi_1 + \varphi_2}(x, y) &= d_{\varphi_1}(x, y) + d_{\varphi_2}(x, y), \\ d_{c\varphi}(x, y) &= c d_\varphi(x, y), (parac \geq 0). \end{aligned} \quad (21)$$

4. **Separação linear:** A localização de todos os pontos $x \in S$ que são equidistantes de dois pontos fixos $\mu_1, \mu_2 \in ri(S)$ é um hiperplano, ou seja, as partições induzidas pela divergência de Bregman têm separadores lineares dados por:

$$\begin{aligned} d_\varphi(x, \mu_1) &= d_\varphi(x, \mu_2) \\ \varphi(x) - \varphi(\mu_1) - \langle x - \mu_1, \nabla \varphi(\mu_1) \rangle &= \varphi(x) - \varphi(\mu_2) - \langle x - \mu_2, \nabla \varphi(\mu_2) \rangle \\ \langle x, \varphi(\mu_2) - \varphi(\mu_1) \rangle &= (\varphi(\mu_1) - \varphi(\mu_2)) - (\langle \mu_1, \nabla \varphi(\mu_1) \rangle - \langle \mu_2, \nabla \varphi(\mu_2) \rangle) \end{aligned} \quad (22)$$

5. **Teorema de Pitágoras generalizado:** Para qualquer $x_1 \in S$ e $x_2, x_3 \in ri(S)$, a seguinte propriedade entre três pontos se estabelece: $d_\varphi(x_1, x_2) + d_\varphi(x_2, x_3) \leq d_\varphi(x_1, x_3)$.

4.1.1 Tratamentos para Kullback Leibler

Em (ROCHA et al., 2021) foram propostos três tratamentos para a divergência KL com o objetivo de manter as características dentro do domínio *d-Simplex*. Para esse estudo foi utilizado somente o tratamento baseado em Teoria da Informação (TI).

Os vetores de características podem conter valores iguais a zero. Esta questão exige um tratamento, uma vez que a divergência KL é definida utilizando a função logarítmica cujo domínio é $x > 0$. A divergência KL entre dois vetores x e y , $x = (x_1, \dots, x_d)$ e $y = (y_1, \dots, y_d)$ é definida pela Equação 23, onde x e y pertencem ao conjunto d -Simplex definido na Equação 24.

$$d_\varphi(x, y) = \sum_{j=1}^d x_j \log_2\left(\frac{x_j}{y_j}\right) \quad (23)$$

$$\Delta^d = \{x = (x_1, x_2, \dots, x_d); 0 \leq x_j \leq 1, \text{ tal que } \sum_{j=1}^d x_j = 1\} \quad (24)$$

Assim, se para algum j , $x_j = 0$ e $y_j \neq 0$, a parcela $0 \log_2(0/y_j)$ tende a nula, uma vez que o limite $\lim_{x_j \rightarrow 0} x_j \log_2 x_j = 0$. Entretanto se $x_j \neq 0$ e $y_j = 0$, pode-se assumir que $x_j \log_2(x_j/0) = \infty$, como apresentado na Equação 25. É possível ter duas opções de tratamento para os dados quando os vetores x e y apresentarem coordenadas onde $x_j \neq 0$ e $y_j = 0$. A primeira opção, denominada Kullback-Leibler baseado na Teoria da Informação (KL/TI), é assumir $x_j \log_2(x_j/0) = \gamma$, onde γ é um valor suficientemente grande. A segunda, denominada denominado Kullback-Leibler com o tratamento do ϵ (KL/ ϵ), considera que se $y_j \rightarrow 0$, então $y_j = \epsilon$, onde ϵ é um valor próximo de zero (ROCHA et al., 2021). Neste trabalho foi utilizada a opção Kullback-Leibler com o tratamento do ϵ (KL/ ϵ).

$$x_j \log_2\left(\frac{x_j}{y_j}\right) = \begin{cases} 0, & \text{se } x_j = 0 \\ \infty, & \text{se } x_j \neq 0 \text{ e } y_j = 0 \end{cases} \quad (25)$$

Experimentos e Análise dos Resultados

Neste estudo, foram empregados os descritores baseados em modelos de redes neurais convolucionais, especificamente das renomadas arquiteturas AlexNet e ResNet-50, para a extração de características das imagens de cavidade bucal em relação aos graus de displasia. Para complementar essa abordagem, utilizou-se a segmentação das imagens por meio de uma rede Mask R-CNN, visando a obtenção de informações obtidas com os descritores morfológicos e não morfológicos. Foram incluídas análises de textura, através da entropia e do índice de Moran, para enriquecer o conjunto de características obtidas.

Ao explorar a composição dos vetores com os descritores não morfológicos e informações morfológicas, como área, excentricidade, orientação, perímetro e solidez, de descritores obtidos por engenharia de características. A associação da extração de características por meio de redes neurais e informações de descritores obtidos por engenharia de características, permitiu a obtenção de um conjunto abrangente de informações descritivas, que foram fundamentais para a análise e interpretação dos graus de displasia presentes nas amostras estudadas.

Para a análise de recuperação de imagens utilizando as divergências de Bregman, foram selecionadas as duas abordagens métricas, Euclidiana e Manhattan, e as duas não-métricas, divergência KL e Mahalanobis. A partir destas distâncias, foram calculadas as médias de precisão e revocação para as 10 primeiras imagens recuperadas do banco de dados de displasias e posteriormente feito o gráfico de precisão \times revocação para todas as imagens, conforme pode ser observado em 6.2.

Para que fosse possível mensurar as diferenças entre as divergências de Bregman e as distâncias métricas, a metodologia foi aplicada em uma série de associações investigadas na composição dos vetores obtidos pelas técnicas de extração das características:

- AlexNet
- ResNet-50
- AlexNet + ResNet-50
- Engenharia de Características
- Engenharia de Características + AlexNet

- ❑ Engenharia de Características + ResNet-50
- ❑ Engenharia de Características + AlexNet + ResNet-50

Para avaliar a precisão e revocação foi realizado a comparação entre os graus de displasia:

- ❑ Saudável x Leve
- ❑ Saudável x Moderado
- ❑ Saudável x Severo
- ❑ Leve x Moderado
- ❑ Leve x Severo
- ❑ Moderado x Severo

Nestes experimentos, foram aplicados as distâncias propostas nos grupos de (50/40/30/20/10) características mais relevantes das imagens utilizando o algoritmo ReliefF. Essa abordagem de separação do número de características foram definidas nos estudos (SILVA et al., 2022a) e (OLIVEIRA et al., 2023).

5.1 Investigação dos Descritores Obtidos com as Redes de Aprendizagem Profunda

Nesta seção, são os experimentos e os resultados do estudo comparativo aos modelos de redes neurais convolucionais. São apresentados as funções de similaridade investigadas e os resultados no processo de recuperação de imagem.

5.1.1 Modelo AlexNet

A Tabela 2 apresenta os resultados obtidos utilizando as comparações propostas em seus níveis de displasias para precisão e revocação. A partir dos resultados médios das 10 primeiras imagens recuperadas, é possível destacar alguns resultados para os graus de displasia.

No grupo de comparação “saudável × leve”, a métrica Mahalanobis (ReliefF 30+) destacou-se com uma precisão de 94,4% e revocação de 12,8%, demonstrando alta eficácia na diferenciação entre tecido saudável e displasia leve. Para o grupo “saudável × moderada”, a métrica Manhattan (ReliefF 50+) obteve uma precisão de 91,6% e revocação de 12,4%, sendo a mais eficaz para distinguir entre tecido saudável e displasia moderada. Já no grupo “saudável × severa”, a métrica KL (ReliefF 40+) também apresentou bons resultados com uma precisão elevada de 92,1% e revocação de 8,0% na média para às 10 primeiras imagens, porém graficamente a métrica Mahalanobis foi mais eficiente com o aumento de imagens recuperadas.

Em comparações entre diferentes níveis de displasia, os resultados variaram. Para o grupo “leve × moderada”, as métricas KL (ReliefF 50+) e Mahalanobis (ReliefF 50+) apresentaram

boas precisões, com destaque para a métrica KL que obteve uma precisão de 86,1% e revocação de 8,3%. No grupo “leve × severa”, a métrica KL (ReliefF 40+) destacou-se com a maior precisão de 93,2% e revocação de 9,5%, indicando sua eficácia na diferenciação entre displasia leve e severa. Por fim, no grupo “moderada × severa”, a métrica KL (ReliefF 50+) alcançou uma precisão de 85,7% e revocação de 9,3%, mostrando-se eficaz para essa comparação.

Na Figura 20(a), que aborda o grupo de comparação “saúdável × leve”, observa-se que a distância Euclidiana e a distância Mahalanobis mantêm um desempenho consistentemente alto. Ambas apresentam precisão entre 85% e 95% ao longo de quase toda a extensão da revocação, indicando uma eficácia robusta na distinção entre estas classes. A distância Manhattan também demonstra um desempenho semelhante, sustentando uma alta precisão. Em contraste, a distância KL exibe uma queda acentuada na precisão com o aumento da revocação, começando em torno de 80% e caindo para cerca de 50% quando a revocação atinge 100%. Isso sugere que a distância KL é menos eficaz para esta tarefa específica, resultando em uma performance inferior comparada às outras distâncias.

Na Figura 20(b), que trata o grupo “saúdável × moderada”, a distância Euclidiana novamente demonstra um desempenho robusto, com precisão alta que começa próxima a 95% e se mantém elevada até cerca de 80% de revocação antes de diminuir gradualmente. A distância Mahalanobis apresenta um comportamento semelhante, mantendo uma precisão alta e estável até aproximadamente 80% de revocação, indicando eficácia comparável à da distância Euclidiana. A distância Manhattan também mantém uma boa performance inicial, com alta precisão até cerca de 80% de revocação antes de começar a cair, mostrando-se eficaz e comparável às distâncias Euclidiana e Mahalanobis. Em contraste, a distância KL novamente apresenta uma performance inferior, com uma queda significativa na precisão à medida que a revocação aumenta, começando em torno de 70% e caindo para aproximadamente 50%.

A análise da Figura 20(c) para o grupo “saúdável × severa”, mostra que a distância Mahalanobis oferece o melhor desempenho, mantendo alta precisão até cerca de 60% de revocação. A distância Euclidiana começa com alta precisão, mas cai para cerca de 50% com revocação total. A distância Manhattan tem desempenho intermediário, também caindo para cerca de 50%. Por outro lado, a distância KL exibe a performance mais fraca entre as quatro distâncias. A precisão começa em torno de 80% e diminui significativamente, caindo para 50% com a revocação total.

A Figura 20(d) para o grupo “leve × moderada”, mostra que a distância KL inicia com uma alta precisão de cerca de 90%, mantendo-se acima de 70% até aproximadamente 50% de revocação e caindo para entre 50% e 70% de revocação total. A distância Mahalanobis, Euclidiana e Manhattan começa com precisão em torno de 70%, decrescendo gradualmente para aproximadamente 55% à medida que a revocação atinge 100%.

Observando a Figura 20(e) o grupo “leve × severa”, mostra que a distância KL inicia com uma alta precisão de acima de 90%, mantendo-se acima de 70% até aproximadamente 60% de revocação e caindo para cerca de 60% com revocação total. A distância Mahalanobis começa com precisão em torno de 60%, decrescendo gradualmente para aproximadamente 55% à medida que a revocação atinge 100%. A distância Euclidiana apresenta um desempenho estável similar à Manhattan, iniciando com cerca de 70% de precisão e diminuindo para aproximadamente 55% no final.

Para o grupo “moderada \times severa”, Figura 20(f), a distância KL apresenta a melhor performance inicial com precisão de próximo aos 90%, caindo gradualmente para 60%. As distâncias Euclidiana, Mahalanobis e Manhattan começam próximas de 70% e caem para cerca de 55–60%.

Portanto utilizando os descritores do modelo AlexNet, as distâncias Euclidiana, Mahalanobis e Manhattan são mais eficazes para a classificação de displasias orais nos contextos de comparação entre “saudável \times leve”, “saudável \times moderada”, e “saudável \times severa”. Para as comparações entre displasias “leves \times moderadas”, “leves \times severas”, e “moderadas \times severas”, a distância KL consistentemente apresenta a melhor performance inicial, mantendo alta precisão nos primeiros níveis de revocação.

Tabela 2 – Valores das métricas Precisão (Prec.) e Revocação (Rev.) de descritores obtidos do modelo AlexNet e distâncias para as displasias nos tecidos de graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

CNN Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
AlexNet	50+	Euclideana	90,5	12,2	91,0	12,3	84,9	11,5	70,1	9,5	75,5	10,2	74,9	10,1
		Manhattan	91,4	12,3	91,6	12,4	84,0	11,3	70,1	9,5	75,6	10,2	75,3	10,2
		Kullback-Leibler (KL)	82,4	9,4	60,1	8,1	89,3	7,7	86,1	8,3	92,2	9,6	85,7	9,3
		Mahalanobis	94,5	12,8	90,5	12,2	88,5	12,0	70,5	9,5	58,9	8,0	72,6	9,8
	40+	Euclideana	91,2	12,3	91,3	12,3	84,3	11,4	70,7	9,6	74,3	10,0	74,0	10,0
		Manhattan	90,7	12,3	91,4	12,3	82,8	11,2	71,5	9,7	75,3	10,2	73,9	10,0
		Kullback-Leibler (KL)	87,9	9,6	60,2	8,4	92,1	8,0	85,5	8,4	93,2	9,5	82,4	8,7
		Mahalanobis	94,1	12,7	90,2	12,2	87,7	11,9	69,2	9,3	62,9	8,5	72,9	9,9
	30+	Euclideana	92,0	12,4	91,5	12,4	83,7	11,3	70,5	9,5	73,2	9,9	74,1	10,0
		Manhattan	92,4	12,5	91,2	12,3	82,4	11,1	70,3	9,5	73,6	10,0	73,6	10,0
		Kullback-Leibler (KL)	90,1	9,9	61,4	8,4	88,3	8,5	83,0	8,2	90,6	9,3	80,5	8,7
		Mahalanobis	94,4	12,8	90,5	12,2	84,5	11,4	69,1	9,3	63,2	8,5	70,6	9,5
	20+	Euclideana	92,8	12,5	90,8	12,3	83,0	11,2	70,9	9,6	71,6	9,7	71,1	9,6
		Manhattan	92,8	12,5	90,9	12,3	83,2	11,2	71,3	9,6	71,1	9,6	70,8	9,6
		Kullback-Leibler (KL)	87,0	10,9	64,9	8,7	79,1	8,7	83,5	8,3	92,0	8,5	81,1	8,3
		Mahalanobis	93,2	12,6	90,1	12,2	82,3	11,1	70,0	9,5	64,9	8,8	60,9	8,2
	10+	Euclideana	93,3	12,6	90,8	12,3	75,5	10,2	70,6	9,5	67,6	9,1	68,4	9,2
		Manhattan	93,7	12,7	90,7	12,3	74,6	10,1	70,4	9,5	68,1	9,2	68,6	9,3
		Kullback-Leibler (KL)	61,8	8,5	67,8	9,3	80,1	6,4	72,4	8,7	78,6	8,8	81,4	8,6
		Mahalanobis	93,4	12,6	90,9	12,3	79,1	10,7	63,9	8,6	61,4	8,3	55,1	7,4

Tabela 3 – Resultados do Teste de *Kruskal-Wallis* (Modelo AlexNet)

Métrica	Estatística	$p - Valor$
Precisão	0,3003	0,9600
Revocação	8,6099	0,0350

Atráves do resultado médio para as métricas de precisão e revocação e os dados não seguirem uma distribuição normal, foi realizado o teste não paramétrico de *Kruskal-Wallis*, com o objetivo de determinar se há diferenças estatisticamente significativas entre as medições de desempenho para as diferentes distâncias (Euclidiana, Manhattan, Kullback-Leibler, Mahalanobis).

Observando a Tabela 3, para a métrica de Precisão, o $p - valor$ do teste é muito maior que 0,05, indicando que não há diferenças estatisticamente significativas entre as diferentes distâncias utilizadas. Isso sugere que, com relação à precisão, todas as distâncias se comportam de maneira semelhante. E para a métrica de Revocação, o $p - valor$ é menor que 0,05, indicando

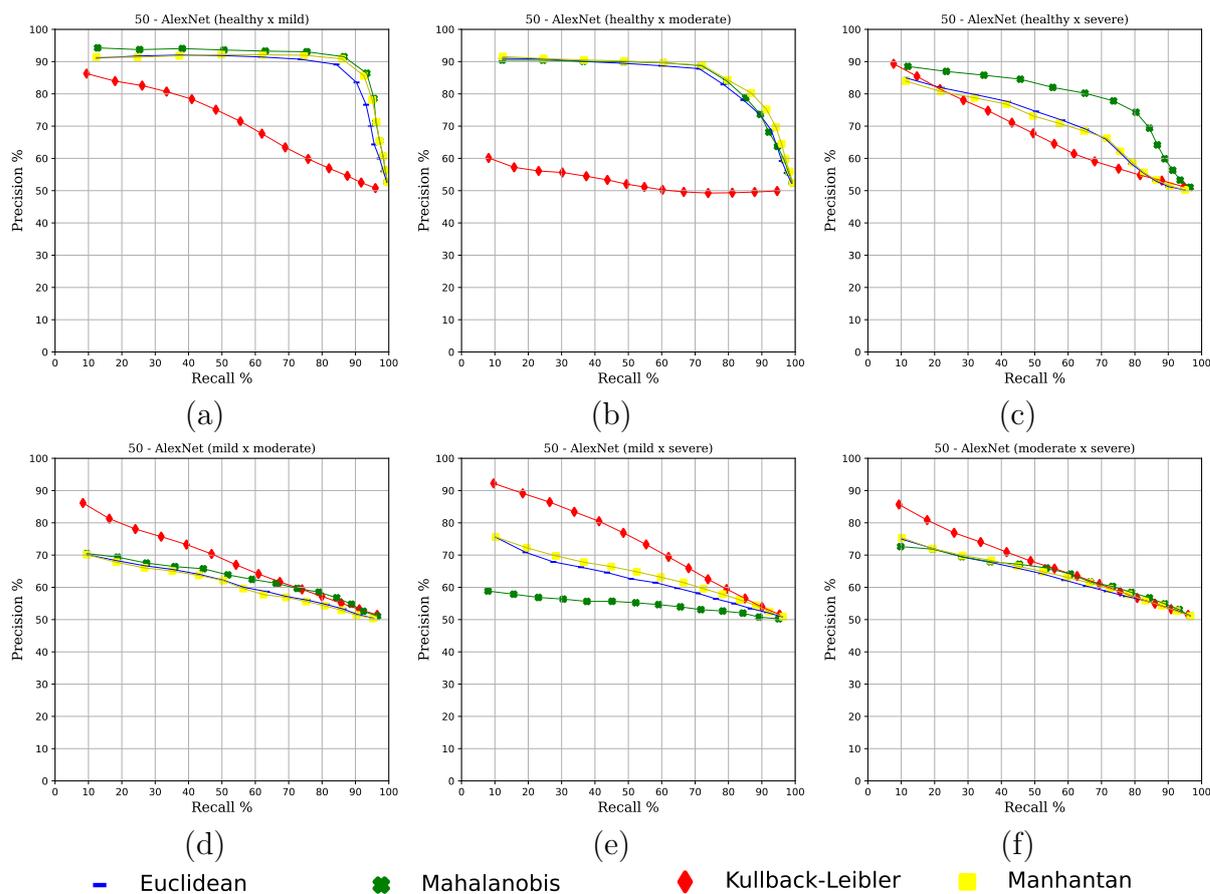


Figura 20 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - AlexNet (saudável x leve) (b) 50 - AlexNet (leve x moderada) (c) 50 - AlexNet (moderada x severa) (d) 50 - AlexNet (leve x moderada) (e) 50 - AlexNet (leve x severa) (f) 50 - AlexNet (moderada x severa)

que há diferenças estatisticamente significativas entre as diferentes distâncias utilizadas. Isso sugere que, com relação a revogação, algumas distâncias podem ser melhores que outras.

5.1.2 Modelo ResNet-50

Os resultados da avaliação das métricas de revogação e precisão usando os descritores obtidos do modelo ResNet-50 são apresentados na Tabela 4.

Realizando a média dos grupos de características mais relevantes e a média dos resultados para cada grau de displasia, utilizando esse descritor, os valores das métricas precisão e desvio padrão foram de $79,0 \pm 1,0$ Euclidiana, $80,6 \pm 1,3$ Manhathan, $77,9 \pm 6,2$ KL e $71,8 \pm 3,7$ Mahalanobis, respectivamente.

Observando a Tabela 4, é possível observar para o grupo “saudável \times leve”, o melhor desempenho com a métrica KL (ReliefF 20+), que alcançou uma precisão de 95.0% e revogação de 10.9%. No grupo “saudável \times moderada”, o desempenho mais elevado foi novamente obtido com a métrica KL (ReliefF 50+), que apresentou uma precisão de 97.1% e revogação de 12.0%. Para o grupo “saudável \times severa”, a métrica KL (ReliefF 50+) destacou-se com a maior precisão de 96.8% e revogação de 9.1%. No grupo “leve \times moderada”, a métrica KL (ReliefF 10+) teve o melhor desempenho, com uma precisão de 71.2% e revogação de 7.3%. Para o grupo “leve \times severa”, a métrica Manhathan (ReliefF 20+) obteve a maior precisão de 75.0% e revogação de 10.1%. Finalmente, no grupo “moderada \times severa”, o melhor desempenho foi alcançado com a métrica KL (ReliefF 10+), que obteve uma precisão de 85.1% e revogação de 8.8%.

Para a comparação entre o grupo “saudável \times leve”, Figura 21(a), a distância KL inicia com uma precisão alta de aproximadamente 90%, mas apresenta uma queda acentuada conforme a revogação aumenta, atingindo cerca de 55% de precisão em revogação total. As distâncias Euclidiana, Mahalanobis e Manhattan mantêm uma alta precisão, variando entre 85% e 95%, até cerca de 90% de revogação, demonstrando uma maior consistência e robustez neste cenário.

A análise do gráfico para o grupo “saudável \times moderada”, Figura 21(b), mostra que a distância KL tem o melhor desempenho, iniciando com cerca de 95% de precisão e mantendo-se acima de 70% até 90% de revogação. A distância Manhattan também performa bem, começando com 95% e caindo gradualmente. A distância Euclidiana inicia em torno de 90% e diminui para 60%. A Mahalanobis começa com cerca de 60% e mantém-se estável, indicando ser menos adequada.

O gráfico para o grupo “saudável \times severa”, Figura 21(c), indica que a distância KL apresenta o melhor desempenho. Ela inicia com cerca de 95% de precisão e mantém-se acima de 70% até aproximadamente 90% de revogação. A distância Manhattan também tem um bom desempenho, começando em 95% e caindo gradualmente. A Euclidiana começa com 90% e diminui para 60%, enquanto a Mahalanobis inicia e permanece em torno de 60%. Conclui-se que a distância KL é a mais eficaz para esta classificação.

Na Figura 21(d) para o grupo “leve \times moderada”, as distâncias Manhattan e Euclidiana mantêm desempenho similar, iniciando com cerca de 70% de precisão e caindo gradualmente para 55 – 60%. A Mahalanobis começa próxima de 65% e permanece estável. KL tem o pior desempenho, mantendo-se em torno de 50%.

No grupo “leve \times severa”, Manhattan e Euclidiana novamente se destacam como observado na Figura 21(e), iniciando com cerca de 70 – 75% de precisão e caindo de forma estável. Mahalanobis é ligeiramente inferior, começando em torno de 65%. KL mantém cerca de 50% de

precisão.

Na Figura 21(f), “moderada × severa”, todas as distâncias começam em torno de 70% e caem gradualmente para 55 – 60%, com Manhattan e Euclidiana mantendo uma ligeira vantagem. É observado que a distância KL (ReliefF 10+) apresenta o melhor desempenho inicial, começando com cerca de 90% de precisão e decrescendo para aproximadamente 55%. Conclui-se que a distância KL é mais eficaz através das 10 características mais relevantes para este grupo.

Em resumo, a distância KL é inicialmente eficaz na comparação entre “saudável” e outros níveis de displasias, mostrando-se superior em relação às outras distâncias. No entanto, nas comparações entre diferentes níveis de displasias, seu desempenho é inferior, demonstrando eficácia apenas quando utilizadas as 10 características mais relevantes para esses grupos.

Tabela 4 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

CNN Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
ResNet-50	50+	Euclideana	92,9	12,6	88,5	12,0	86,2	11,7	66,7	9,0	71,6	9,7	67,4	9,1
		Manhattan	93,7	12,7	91,5	12,4	92,5	12,5	68,2	9,2	73,6	10,0	69,5	9,4
		Kullback-Leibler (KL)	93,9	9,3	97,1	12,0	96,8	9,1	57,8	7,6	59,1	8,5	68,9	9,2
		Mahalanobis	93,9	12,7	55,9	7,6	77,7	10,5	63,4	8,6	72,0	9,7	67,2	9,1
	40+	Euclideana	93,0	12,6	86,5	11,7	89,5	12,1	66,6	9,0	73,0	9,9	67,4	9,1
		Manhattan	93,9	12,7	90,0	12,2	92,8	12,5	67,0	9,0	74,3	10,0	67,9	9,2
		Kullback-Leibler (KL)	94,7	9,5	96,5	11,6	95,4	9,2	53,6	7,6	50,5	8,6	71,7	9,0
		Mahalanobis	93,2	12,6	66,5	9,0	85,5	11,6	62,6	8,5	70,7	9,6	64,6	8,7
	30+	Euclideana	92,9	12,6	87,5	11,8	88,3	11,9	65,9	8,9	72,5	9,8	69,6	9,4
		Manhattan	94,0	12,7	89,4	12,1	91,9	12,4	66,4	8,9	74,3	10,0	70,1	9,5
		Kullback-Leibler (KL)	95,0	10,0	96,6	12,0	88,9	8,3	55,9	8,1	52,7	8,6	78,5	8,9
		Mahalanobis	94,5	12,8	54,6	7,4	83,4	11,3	62,6	8,5	73,0	9,9	65,3	8,8
	20+	Euclideana	91,9	12,4	87,9	11,9	87,9	11,9	65,5	8,8	72,4	9,8	67,2	9,1
		Manhattan	92,8	12,5	90,0	12,2	90,3	12,2	65,7	8,9	75,0	10,1	67,6	9,1
		Kullback-Leibler (KL)	95,0	10,9	95,9	12,1	86,6	9,1	70,8	8,1	51,3	8,5	83,0	8,7
		Mahalanobis	92,3	12,5	56,0	7,6	78,4	10,6	61,5	8,3	71,5	9,7	62,1	8,4
	10+	Euclideana	92,6	12,5	87,2	11,8	83,0	11,2	66,1	8,9	73,4	9,9	68,6	9,3
		Manhattan	92,9	12,6	88,9	12,0	83,2	11,2	67,6	9,1	74,5	10,1	68,5	9,3
		Kullback-Leibler (KL)	80,2	7,3	93,9	12,1	81,6	7,2	71,2	7,3	38,1	8,7	85,1	8,8
		Mahalanobis	92,9	12,6	69,5	9,4	73,4	9,9	64,6	8,7	65,3	8,8	60,3	8,2

Tabela 5 – Resultados do Teste de *Kruskal-Wallis* (Modelo ResNet-50)

Métrica	Estatística	Valor p
Precisão	1,686	0,6399
Revocação	5,3151	0,1501

Para a métrica de Precisão (Tabela 5), o p -valor é 0,9600, que é muito maior do que o nível de significância comum de 0,05. Isso indica que não há diferenças estatisticamente significativas entre as diferentes distâncias utilizadas. Em outras palavras, as precisões medidas com as distâncias Euclidiana, Manhattan, KL e Mahalanobis são, em média, indistinguíveis umas das outras. Portanto, a escolha da distância não afeta significativamente a precisão dos resultados. Para a métrica de Revocação (Tabela 5), o p -valor é de 0,0350. Isso indica que há diferenças estatisticamente significativas entre as diferentes distâncias utilizadas. Ou seja, os recalls medidos com as distâncias Euclidiana, Manhattan, KL e Mahalanobis não são iguais, e pelo menos uma dessas distâncias se comporta de maneira diferente das outras em termos de revocação.

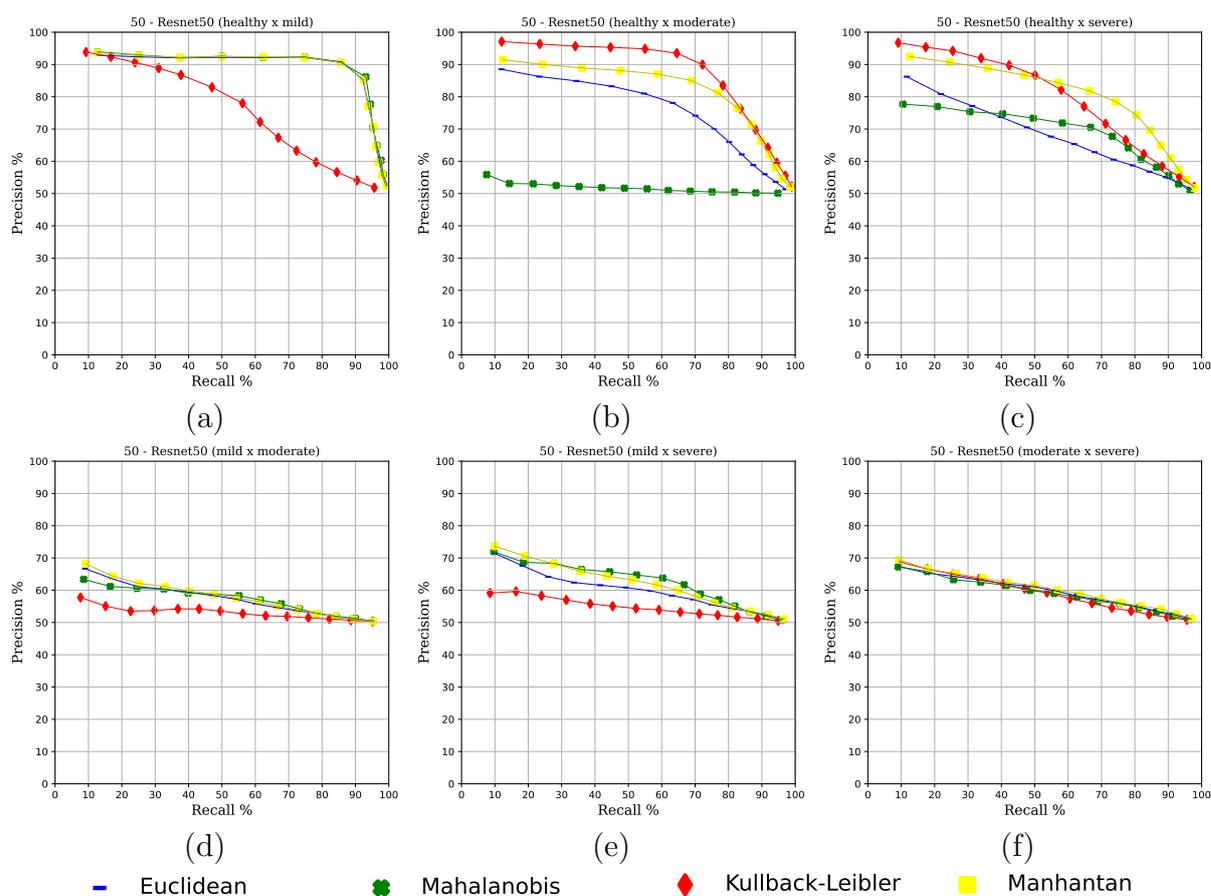


Figura 21 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - ResNet50 (saudável x leve) (b) 50 - ResNet50 (leve x moderada) (c) 50 - ResNet50 (moderada x severa) (d) 50 - ResNet50 (leve x moderada) (e) 50 - ResNet50 (leve x severa) (f) 50 - ResNet50 (moderada x severa)

5.1.3 Associação dos descritores obtidos com AlexNet e ResNet-50

Neste experimento foi realizada a união das características obtidas utilizando as redes neurais AlexNet e ResNet-50. Utilizar diferentes redes neurais para a extração de características de imagens e combinar as informações obtidas pode ajudar a melhorar a precisão na recuperação de imagens. Este processo é conhecido como “*Ensemble Learning*” e pode aproveitar as forças complementares de diferentes arquiteturas de redes neurais para obter representações mais robustas e discriminativas das imagens.

Com a união das duas redes é possível adquirir uma diversidade de características, onde diferentes arquiteturas de redes neurais têm diferentes estruturas e profundidades, o que leva à extração de diferentes tipos de características das imagens. AlexNet, sendo uma rede mais simples e menos profunda, pode capturar características mais gerais, enquanto ResNet-50, com sua profundidade e blocos residuais, pode capturar características mais complexas e detalhadas. Combinando características de diferentes redes, é possível também reduzir o risco de overfitting, pois os modelos individuais podem complementar as falhas uns dos outros. E as características combinadas de várias redes podem tornar o sistema de recuperação de imagens mais robusto a variações e ruídos nos dados de entrada.

As divergências de KL obteve resultados favoráveis em todos os grupos de graus de displasia em comparação com outras distâncias. Isso é demonstrado pelos resultados médios das 10 primeiras imagens recuperadas (Tabela 6). Para o grupo “saudável × leve”, o melhor desempenho foi obtido com a métrica KL (ReliefF 50+), que alcançou uma precisão de 98.8% e revocação de 11.3%. No grupo “saudável × moderada”, a métrica KL (ReliefF 50+) novamente se destacou com a maior precisão de 97.6% e revocação de 11.3%. Para o grupo “saudável × severa”, a métrica KL (ReliefF 40+) obteve a maior precisão de 99.9% e revocação de 11.9%. No grupo “leve × moderada”, a métrica KL (ReliefF 10+) apresentou o melhor desempenho com uma precisão de 77.7% e revocação de 8.4%. Para o grupo “leve × severa”, a métrica KL (ReliefF 50+) obteve a maior precisão de 81.6% e revocação de 10.0%. Finalmente, no grupo “moderada × severa”, o melhor desempenho foi alcançado com a métrica KL (ReliefF 10+), que obteve uma precisão de 86.4% e revocação de 8.8%.

Os valores altos de precisão para a métrica KL ocorre devido a união das redes complementar as falhas uns dos outros em resultados de descritores iguais a zero. É possível notar que para alguns grupos de de displasia a divergência de Mahalanobis não obteve resultados satisfatórios, efeito direto da matriz de covariância uma vez que não conseguiu capturar as correlações e variações entre as variáveis. Para efeito de comparação se utilizarmos a matriz identidade assumindo que as características entre as imagens são independentes, a divergência de Mahalanobis resulta na distância Euclidiana.

Realizando a média dos grupos de características mais relevantes e a média dos resultados para cada grau de displasia, utilizando esses descritores, resultaram em uma precisão e desvio padrão respectivamente de $81,0 \pm 0,9$ Euclidiana, $82,3 \pm 1,0$ Manhattan, $86,9 \pm 3,8$ KL e $70,4 \pm 3,4$ Mahalanobis.

A Figura 22(a) para o grupo “saudável × leve”, revela que a distância KL apresenta o

melhor desempenho inicial, com precisão próxima de 95%. Embora a precisão diminua significativamente após 80% de revocação, caindo para cerca de 70%, é importante destacar que as primeiras imagens recuperadas garantem alta precisão. As distâncias Manhattan e Euclidiana também apresentam alta precisão inicial, em torno de 95%, mas caem mais acentuadamente após 80% de revocação, atingindo cerca de 60 – 65%. A distância Mahalanobis mantém uma precisão mais estável, começando em torno de 85% e caindo gradualmente para cerca de 75% com o aumento da revocação.

Para o grupo “saudável \times moderada”, Figura 22(b), mostra que a distância KL tem o melhor desempenho inicial, com precisão próxima de 95%, caindo para cerca de 70% com o aumento da revocação. As distâncias Manhattan e Euclidiana também iniciam com alta precisão de 95%, mas caem para cerca de 60% ao final. A distância Mahalanobis (verde) é a menos eficaz, começando com precisão de 85% e caindo para 60%. Apesar da queda na precisão, a KL mantém alta precisão nas primeiras imagens recuperadas.

Observando a Figura 22(c) para o grupo “saudável \times severa”, mostra o alto desempenho que a distância KL obteve, com precisão próxima de 99%. Apesar de a precisão diminuir para cerca de 70% à medida que a revocação aumenta, a KL garante alta precisão nas primeiras imagens recuperadas. As distâncias Manhattan e Euclidiana também começam com alta precisão de 90 – 95%, mas caem mais acentuadamente para cerca de 60%. A distância Mahalanobis é a menos eficaz, com precisão inicial em torno de 70%, mantendo-se estável mas mais baixa ao longo da curva. Assim, destaca-se o alto desempenho da KL nas etapas iniciais de recuperação.

Para os grupos, “leve \times moderada”, “leve \times severa”, “moderada \times severa”, mostram que a distância KL consistentemente apresenta a melhor precisão inicial, embora todas as distâncias exibam uma tendência de declínio com o aumento da revocação. No cenário “leve \times moderada”, Figura 22(d), todas as distâncias iniciam com precisão em torno de 70%, caindo gradualmente para cerca de 55 – 60%. No cenário “leve \times severa”, Figura 22(e), a KL começa com aproximadamente 85% de precisão, mantendo-se superior às outras distâncias até cerca de 70% de revocação, enquanto as demais caem para cerca de 55%. No cenário “moderada \times severa”, Figura 22(f), todas as distâncias começam com precisão em torno de 70%, decrescendo de maneira mais uniforme para cerca de 55%. Em resumo, a KL destaca-se inicialmente em todos os cenários, mas todas as distâncias convergem para desempenhos semelhantes com revocação total.

Em resumo, a distância KL, apesar de sua precisão diminuir rapidamente, garante alta precisão para as primeiras imagens recuperadas. As distâncias Manhattan e Euclidiana também oferecem alta precisão inicial com queda acentuada posteriormente, enquanto a Mahalanobis mostra-se mais estável ao longo da curva. Também é possível observar que, com a seleção das 10 variáveis mais relevantes, a métrica KL apresenta uma performance superior em comparação com as outras métricas.

Observando a Tabela 7, para a métrica de Precisão, o p – Valor é de 0.0589. Este valor é ligeiramente maior que o nível de significância de 0,05. Portanto, embora esteja próximo do limite, ainda não há evidências suficientes para rejeitar a hipótese nula de que não há diferenças significativas nas precisões entre as diferentes distâncias utilizadas. Porém se considerarmos o nível de significância de 0,1, há evidências suficientes para rejeitar a hipótese nula de que não

Tabela 6 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor AlexNet + ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

CNN Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
AlexNet + ResNet-50	50+	Euclidean	90,7	12,3	91,6	12,4	90,2	12,2	71,2	9,7	73,4	9,9	68,2	9,2
		Manhattan	91,7	12,4	92,7	12,5	92,8	12,5	71,4	9,7	76,8	10,4	71,6	9,7
		Kullback-Leibler (KL)	98,8	11,3	97,6	11,3	99,8	11,9	73,9	8,0	81,6	10,0	72,0	9,4
		Mahalanobis	82,4	11,1	85,2	11,5	54,6	7,4	69,1	9,4	70,4	9,5	69,5	9,4
	40+	Euclidean	91,4	12,3	91,1	12,3	90,7	12,3	70,9	9,6	74,1	10,0	68,6	9,3
		Manhattan	92,3	12,5	92,2	12,5	92,0	12,4	69,9	9,5	77,5	10,5	71,7	9,7
		Kullback-Leibler (KL)	98,6	11,2	97,1	10,6	99,9	11,9	69,8	8,0	78,7	10,0	74,7	9,1
		Mahalanobis	79,9	10,8	89,7	12,1	54,3	7,3	69,1	9,4	73,5	9,9	68,6	9,3
	30+	Euclidean	91,7	12,4	90,9	12,3	88,9	12,0	70,8	9,6	75,1	10,1	70,5	9,5
		Manhattan	92,6	12,5	91,6	12,4	90,6	12,2	71,1	9,7	77,2	10,4	72,5	9,8
		Kullback-Leibler (KL)	98,6	11,0	97,2	11,0	99,5	11,7	71,4	8,1	77,2	9,9	81,2	9,3
		Mahalanobis	75,0	10,1	90,8	12,3	54,4	7,4	67,0	9,1	68,2	9,2	67,0	9,0
	20+	Euclidean	91,8	12,4	90,3	12,2	89,2	12,1	71,0	9,7	74,2	10,0	68,8	9,3
		Manhattan	92,6	12,5	91,4	12,3	89,9	12,2	71,1	9,7	77,6	10,5	70,6	9,5
		Kullback-Leibler (KL)	97,8	10,9	96,5	11,3	99,1	11,4	76,5	8,1	78,9	9,7	83,2	9,0
		Mahalanobis	76,3	10,3	90,4	12,2	56,1	7,6	67,7	9,2	65,9	8,9	65,3	8,8
	10+	Euclidean	93,0	12,6	90,5	12,2	85,5	11,6	70,1	9,5	74,6	10,1	69,8	9,4
		Manhattan	93,3	12,6	91,8	12,4	85,7	11,6	69,9	9,5	76,4	10,3	70,1	9,5
		Kullback-Leibler (KL)	97,0	12,1	93,9	11,8	96,7	10,9	77,0	8,4	56,1	9,6	86,4	8,8
		Mahalanobis	76,6	10,3	85,5	11,6	58,6	7,9	65,7	8,9	52,8	7,1	61,7	8,3

Tabela 7 – Resultados do Teste de *Kruskal-Wallis* (AlexNet + ResNet-50)

Métrica	Estatística	$p - Valor$
Precisão	7,4467	0,0589
Revocação	5,5738	0,1343

há diferenças significativas nas precisões entre as diferentes distâncias utilizadas. Mostrando que a escolha da distância afeta significativamente a precisão dos resultados. Na Revocação, o $p - Valor$ é de 0,1343, que é maior do que o nível de significância comum de 0,05. Isso indica que não há diferenças estatisticamente significativas entre a revocação das diferentes distâncias utilizadas. Em outras palavras, os recalls medidos com as distâncias Euclidiana, Manhattan, Kullback-Leibler e Mahalanobis são, em média, indistinguíveis uns dos outros.

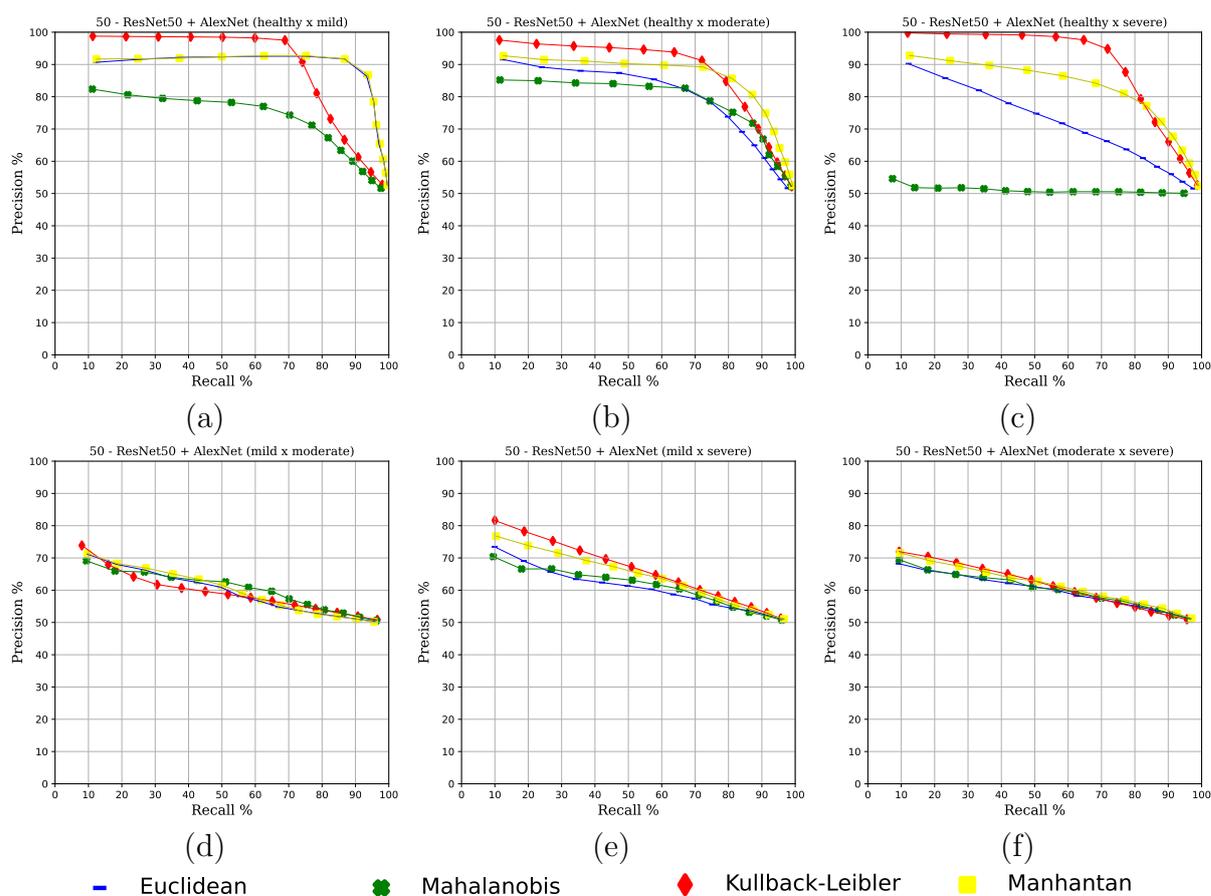


Figura 22 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - ResNet50 + AlexNet (saúdável x leve) (b) 50 - ResNet50 + AlexNet (leve x moderada) (c) 50 - ResNet50 + AlexNet (moderada x severa) (d) 50 - ResNet50 + AlexNet (leve x moderada) (e) 50 - ResNet50 + AlexNet (leve x severa) (f) 50 - ResNet50 + AlexNet (moderada x severa)

5.1.4 Descritores Obtidos por Engenharia de Características

Neste experimento foram extraídos descritores baseados em engenharia de características. Essas características foram extraídas após as imagens serem segmentadas com a rede Mask R-CNN. A Tabela 9 traz os valores das métricas em relação ao conjunto de informações morfológicas e não morfológicas obtidas após a etapa de segmentação. As métricas utilizadas para as características morfológicas foram: área (A), excentricidade (E), orientação (O_R), perímetro (P_E) e solidez (S). E para as características não morfológicas são: entropia (E_N) e índice de moran (I_M). Essas características exploram informações internas das regiões em imagens histológicas.

Observando os grupos comparados de displasias na Figura 23, fica visualmente nítido que a divergência KL foi superior a outras distâncias utilizadas em todas as comparações.

Nesta etapa, dada pelos descritores obtidos por engenharia de características, não há grupos de características mais relevantes. Desta forma, a média dos resultados para cada grau de displasia, resultaram em uma precisão média de 60,9 Euclidiana, 63,4 Manhattan, 71,0 KL e 59,4 Mahalanobis.

A métrica KL destacou-se por apresentar o melhor desempenho geral entre as métricas avaliadas. Para a distinção entre grupo “saudável \times leve”, a KL obteve a maior precisão de 74,5% com uma revocação de 8,6%. Na comparação entre o grupo “saudável \times moderada”, a KL também apresentou o melhor resultado, com precisão de 76,2% e revocação de 9,5%. Em comparações envolvendo “saudável \times severa”, a métrica KL alcançou a maior precisão de 74,0% e revocação de 9,2% e “leve \times moderada”, precisão de 61,4% e revocação de 7,8%. Para os grupos “leve \times severa”. “moderada \times severa” a métrica KL também alcançou os maiores valores de precisão sendo 67,0% e 73,0% de precisão respectivamente.

A análise dos gráficos na Figura 23, mostra que, em todos os cenários de recuperação utilizando a métrica KL, apresenta uma precisão inicial relativamente alta, especialmente quando comparando amostras saudáveis com outros níveis de displasia. No cenário “saudável \times leve”, a precisão inicial é próxima de 70%, mas cai gradualmente para cerca de 50% com o aumento da revocação. Em “saudável \times moderada” e “saudável \times severa”, a métrica KL começa com precisão em torno de 70% e 65%, respectivamente, e também diminui para cerca de 50%. Nos cenários “leve \times moderada”, “leve \times severa” e “moderada \times severa”, a métrica KL inicia com uma precisão um pouco menor, por volta de 60 – 65%, e mantém-se próxima de 50% com o aumento da revocação. Em resumo, embora a métrica KL apresente uma queda na precisão à medida que a revocação aumenta, ela mantém uma performance inicial superior em comparação com outras métricas, especialmente nas primeiras imagens recuperadas.

Tabela 8 – Resultados do Teste de *Kruskal-Wallis* (Engenharia de Características)

Métrica	Estatística	$p - Valor$
Precisão	11,3598	0,0099
Revocação	8,2625	0,0409

Para as métricas de Precisão e Revocação (Tabela 8, os $p - Valor$ são 0,0099 e 0,0409, respectivamente, ambos menores que o nível de significância de 0,05. Isso indica que há diferen-

ças estatisticamente significativas tanto nas precisões quanto na revocação entre as diferentes distâncias utilizadas. Em outras palavras, as medidas de precisão e revocação variam de forma significativa com a escolha da distância, impactando os resultados. Portanto, a escolha da distância afeta significativamente tanto a precisão quanto a revocação dos resultados. De fato, observando os resultados do experimento, a métrica KL mantém uma performance superior em comparação com outras métricas, especialmente nas primeiras imagens recuperadas, garantindo a diferença significativa com as distâncias comparadas.

Tabela 9 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) dos descritores morfológicos e não morfológicos e distâncias para as displasias nos graus de tecido leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
Euclidean	57,4	8,1	63,4	8,9	61,3	8,5	55,2	7,6	62,9	8,7	65,5	8,8
Manhattan	59,6	8,4	64,9	9,1	64,1	8,9	57,3	7,8	65,9	9,1	68,7	9,1
Kullback-Leibler (KL)	74,5	8,6	76,2	9,5	74,0	9,2	61,4	7,8	67,0	9,2	73,0	8,9
Mahalanobis	57,0	8,1	61,0	8,5	58,3	8,1	57,3	7,6	60,2	7,7	62,6	8,3

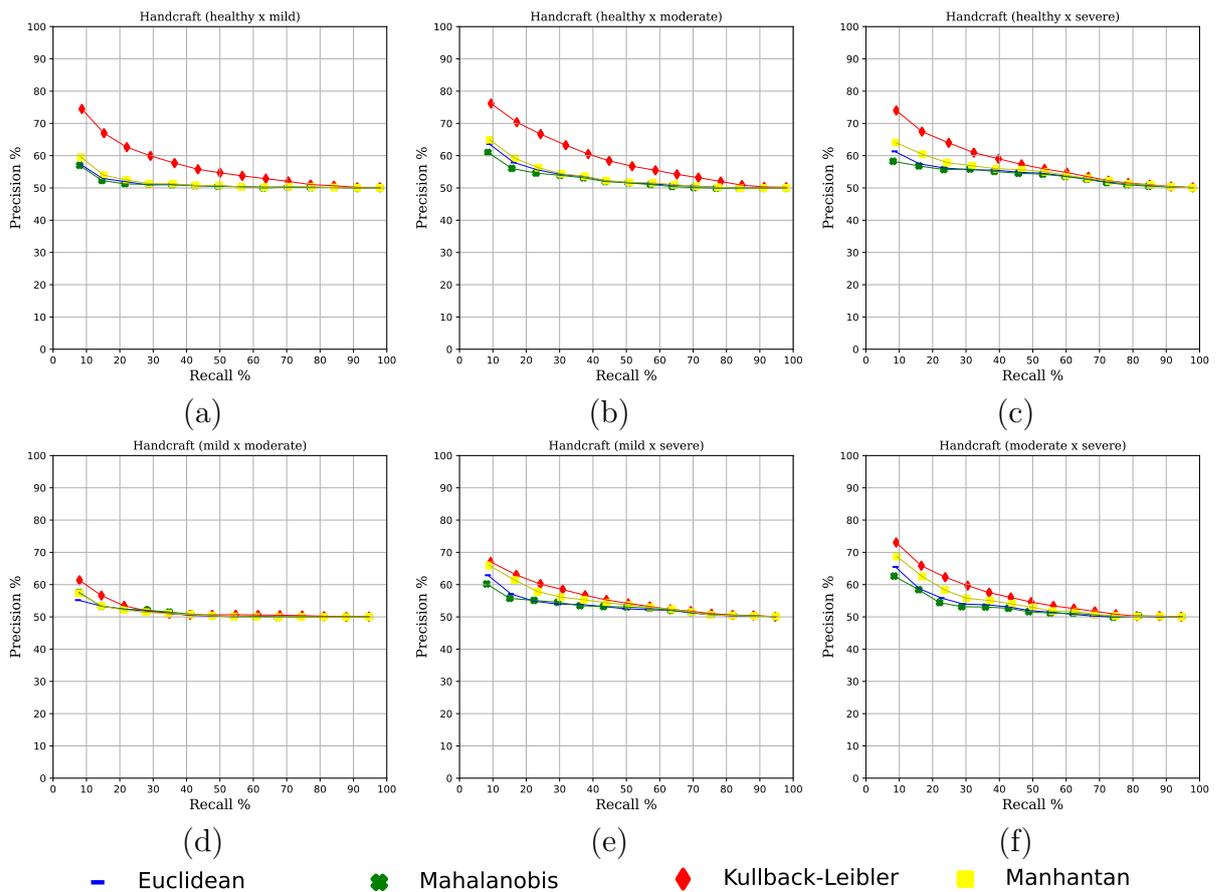


Figura 23 – Análise dos modelos em relação as métricas precisão e revocação: (a) Engenharia de Características (saudável x leve) (b) Engenharia de Características (leve x moderada) (c) Engenharia de Características (moderada x severa) (d) Engenharia de Características (leve x moderada) (e) Engenharia de Características (leve x severa) (f) Engenharia de Características (moderada x severa)

5.1.5 Descritores Obtidos por Engenharia de Características e Modelo AlexNet

Nessa etapa, uma avaliação foi realizada com os descritores baseados em engenharia de característica e a rede AlexNet. A Tabela 10 apresenta os valores das medidas com as distâncias e intervalos de valores das características. Os grupos são apresentados em uma investigação binária entre as classes de lesões. Os valores médios e desvio padrão foram $82,1 \pm 1,2$, Euclidiana $83,3 \pm 0,8$ Manhattan, $90,6 \pm 2,8$ KL e $79,9 \pm 2,8$ Mahalanobis.

O melhor desempenho observado para o grupo “saudável \times leve” foi obtido com a combinação da métrica KL (ReliefF 30+), alcançando uma precisão de 97,2% e revocação de 11,2%. No grupo “saudável \times moderada”, a combinação da métrica KL (ReliefF 20+) apresentou o melhor desempenho para este grupo, com uma precisão de 95,2% e revocação de 11,8%. Para a comparação entre o grupo “saudável \times severa”, a métrica KL (ReliefF 40+) obteve a melhor precisão de 98,1% e revocação de 8,6%. O grupo “leve \times moderada” teve seu melhor desempenho com a métrica KL (ReliefF 20+), alcançando uma precisão de 88,9% e revocação de 8,6%. O melhor resultado para o grupo “leve \times severa” foi com a métrica KL (ReliefF 40+), apresentando precisão de 93,9% e revocação de 9,5%. Por fim, a comparação “moderada \times severa” teve também o melhor desempenho com a métrica KL (ReliefF 50+), alcançando uma precisão de 85,0% e revocação de 9,5%.

Os gráficos para a combinação de descritores baseados em engenharia de característica e a rede AlexNet na classificação entre amostras “saudável \times leve”, Figura 24(a) e “saudável \times moderada”, Figura 24(b), mostram que a distância KL tem o melhor desempenho inicial, com precisão próxima de 95%. No entanto, a precisão diminui significativamente após 80% de revocação, caindo para cerca de 70% em “saudável \times leve” e para cerca de 60% em “saudável \times moderada”. As distâncias Manhattan e Euclidiana mantêm alta precisão inicial de 90 – 95%, mas caem de forma acentuada após 90% de revocação. A distância Mahalanobis mantém uma precisão mais estável, começando em torno de 85 – 90% e caindo gradualmente para cerca de 75 – 80%. Portanto, a KL é mais eficaz inicialmente, mas as outras distâncias mantêm uma precisão mais estável à medida que a revocação aumenta.

Na Figura 24(c), “saudável \times severa”, mostra que a distância KL inicia com alta precisão, próxima de 90%, mas diminui gradualmente para cerca de 50% com o aumento da revocação. A distância Mahalanobis também apresenta uma precisão inicial alta, mas mantém-se mais estável, caindo para cerca de 60% no final. As distâncias Euclidiana e Manhattan começam com precisões similares, em torno de 85 – 90% mas apresentam quedas mais acentuadas, chegando a cerca de 50 – 55% ao longo da curva. Isso indica que, embora a KL tenha uma performance inicial superior, a Mahalanobis oferece uma precisão mais estável à medida que a revocação aumenta, enquanto Euclidiana e Manhattan têm desempenho intermediário.

Para o grupo “leve \times moderada”, Figura 24(d), a distância KL tem a melhor precisão inicial, próxima de 90%, mas diminui gradualmente para cerca de 60% à medida que a revocação aumenta. As distâncias Mahalanobis, Euclidiana e Manhattan apresentam desempenhos similares, começando com precisão em torno de 75 – 80% e caindo de forma mais linear para cerca de 55 – 60%. Isso indica que a KL oferece a maior precisão inicial, mas todas as distâncias tendem

a convergir para valores semelhantes com o aumento da revocação, porém com a KL garantindo a maior precisão nas primeiras imagens recuperadas.

É possível observar em “leve \times severa”, Figura 24(e) que a distância KL apresenta a melhor precisão inicial, próxima de 90%. Embora a precisão diminua gradualmente para cerca de 60% à medida que a revocação aumenta, é crucial notar que a KL mantém uma alta precisão nas primeiras imagens recuperadas, o que é vital para a eficácia inicial da recuperação. A distância Manhattan também mostra um bom desempenho inicial, com precisão em torno de 85%, caindo para cerca de 60% ao longo da revocação. A distância Euclidiana começa com cerca de 80% de precisão e diminui para aproximadamente 55% a 60%. A distância Mahalanobis apresenta o desempenho inicial mais baixo, começando em torno de 70% e caindo para cerca de 50%. Em resumo, a distância KL não só tem a melhor precisão inicial, mas também garante que as primeiras imagens recuperadas sejam altamente precisas, o que é essencial para a eficácia deste grupo de comparação.

Por fim no grupo “moderada \times severa”, Figura 24(f), mostra que a distância KL apresenta a melhor precisão inicial, próxima de 90%, mas diminui gradualmente para cerca de 60% com o aumento da revocação. As distâncias Manhattan e Euclidiana começam com precisão em torno de 80 – 85% e caem de forma mais linear para cerca de 60%, enquanto Mahalanobis inicia com 75% e diminui para cerca de 55 – 60%.

Os resultados demonstram que a combinação da engenharia de características com AlexNet melhorou significativamente os resultados para a métrica KL, garantindo uma maior precisão nas primeiras imagens recuperadas. Esta combinação se mostrou eficaz ao manter uma precisão inicial elevada, o que é crucial para a eficácia do sistema de recuperação de imagens. Além disso, com a seleção das características mais relevantes, observou-se um aumento na precisão e uma maior sustentabilidade desse valor de precisão à medida que a revocação progride. De certa forma os descritores da engenharia de características também conseguiu suprir parte dos descritores iguais a zero do modelo AlexNet.

Os p -Valor dos testes de *Kruskal-Wallis* (Tabela 11 para as métricas de precisão e revocação são 0,2328 e 0,1272, respectivamente, sendo ambos maiores que o nível de significância de 0,05. Isso indica que as precisões e revocação das diferentes distâncias não diferem estatisticamente significativamente. Portanto, as medidas de revocação e precisão são, em média, iguais entre todas as distâncias avaliadas; isso indica que a escolha da distância não teve um impacto significativo nos resultados dessas métricas.

Tabela 10 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor Engenharia de Características + AlexNet e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.		
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	
Engenharia de Características + AlexNet	50+	Euclideana	93,1	13,0	93,1	13,5	87,0	12,1	70,1	9,7	77,1	10,4	75,3	10,5	
		Manhattan	95,0	13,3	93,9	13,6	87,2	12,1	70,8	9,8	79,4	10,7	75,8	10,5	
		Kullback-Leibler (KL)	96,9	10,6	94,6	10,4	97,2	8,2	86,6	8,7	93,4	9,6	85,0	9,5	
			Mahalanobis	94,8	13,3	92,6	13,4	89,4	12,5	71,1	9,9	64,6	8,7	75,4	10,5
	40+	Euclideana	93,4	13,1	93,6	13,6	86,5	12,1	71,0	9,9	76,3	10,3	74,2	10,3	
		Manhattan	94,4	13,2	93,9	13,6	86,4	12,0	71,3	9,9	79,7	10,8	75,5	10,5	
		Kullback-Leibler (KL)	96,6	10,9	94,5	10,8	98,1	8,6	87,6	8,7	93,9	9,5	82,2	8,9	
		Mahalanobis	94,7	13,2	92,2	13,4	89,7	12,5	70,5	9,9	64,0	8,6	77,1	10,8	
	30+	Euclideana	94,3	13,2	93,3	13,5	85,8	12,0	70,6	9,8	75,5	10,2	74,8	10,4	
		Manhattan	95,4	13,3	94,2	13,7	85,7	11,9	71,0	9,9	78,7	10,6	75,7	10,5	
		Kullback-Leibler (KL)	97,2	11,2	94,6	11,3	95,8	9,4	85,6	8,5	90,9	9,2	82,1	8,9	
			Mahalanobis	95,2	13,3	92,4	13,4	85,9	12,0	68,9	9,6	66,7	9,0	72,9	10,2
	20+	Euclideana	94,8	13,3	92,7	13,4	85,7	12,0	70,8	9,9	75,1	10,2	71,7	10,0	
		Manhattan	96,0	13,4	93,8	13,6	86,9	12,1	71,3	9,9	76,8	10,4	74,3	10,4	
		Kullback-Leibler (KL)	95,5	12,5	95,2	11,8	95,1	9,7	88,9	8,6	92,2	8,4	81,1	8,5	
		Mahalanobis	94,1	13,2	92,0	13,3	84,1	11,8	71,7	10,0	70,3	9,5	66,7	9,3	
	10+	Euclideana	96,1	13,4	92,8	13,4	83,0	11,6	71,1	9,9	72,1	9,7	71,3	9,9	
		Manhattan	96,8	13,5	94,1	13,6	85,5	11,9	71,0	9,9	75,1	10,1	74,3	10,4	
		Kullback-Leibler (KL)	94,2	12,9	94,1	12,6	95,0	10,5	74,9	9,1	78,0	8,8	80,9	9,0	
		Mahalanobis	90,0	12,6	89,3	12,9	83,5	11,7	65,4	9,1	68,7	9,3	64,6	9,0	

Tabela 11 – Resultados do Teste de *Kruskal-Wallis* (Engenharia de Características + AlexNet)

Métrica	Estatística	$p - Valor$
Precisão	4,2800	0,2328
Revocação	5,6991	0,1272

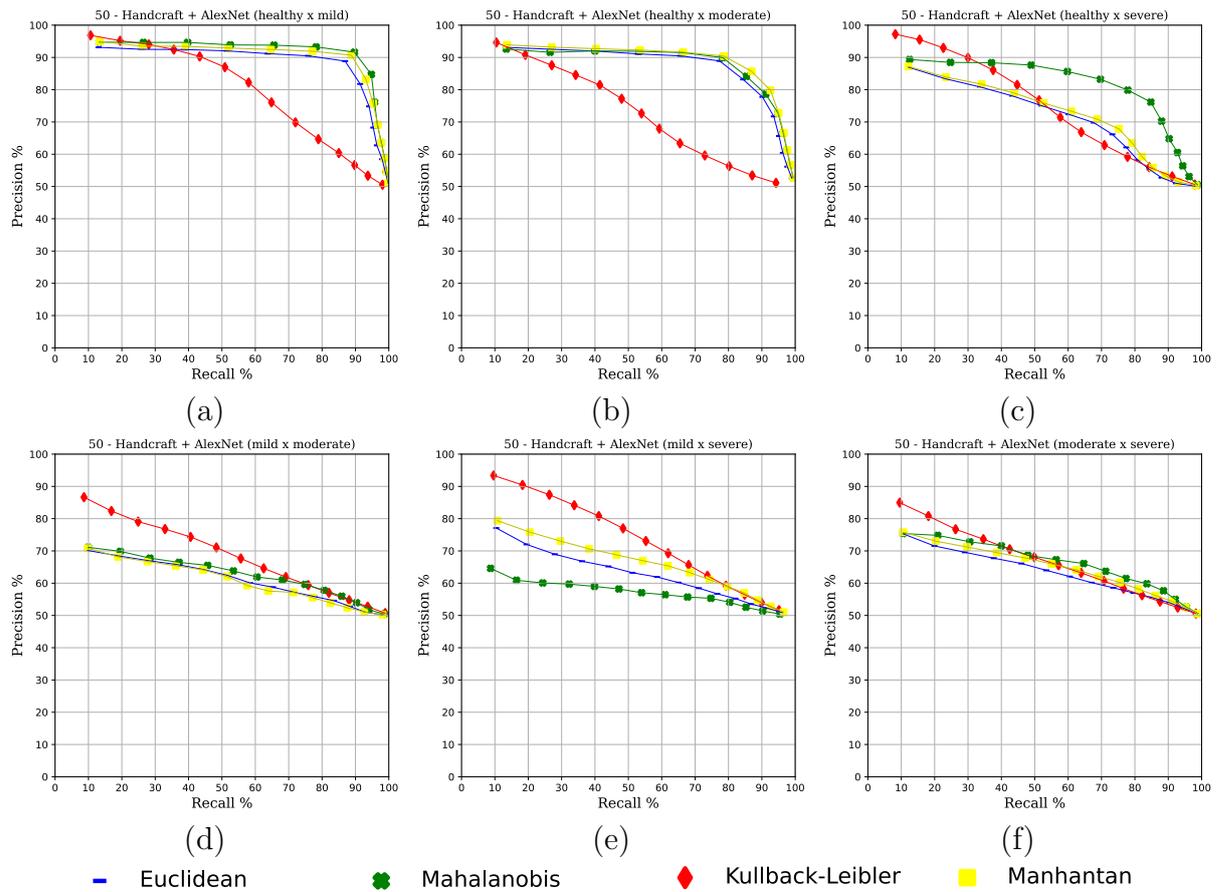


Figura 24 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + AlexNet (saudável x leve) (b) 50 - Engenharia de Características + AlexNet (leve x moderada) (c) 50 - Engenharia de Características + AlexNet (moderada x severa) (d) 50 - Engenharia de Características + AlexNet (leve x moderada) (e) 50 - Engenharia de Características + AlexNet (leve x severa) (f) 50 - Engenharia de Características + AlexNet (moderada x severa)

5.1.6 Descritores Obtidos por Engenharia de Características e Modelo ResNet-50

Nessa etapa, uma avaliação foi realizada com os descritores baseados em engenharia de característica e a rede ResNet-50. A Tabela 12 apresenta os valores das medidas com as distâncias e intervalos de valores das características. Os grupos são apresentados em uma investigação binária entre as classes de lesões. Os valores médios e desvio padrão foram $79,8 \pm 1,0$, Euclidiana $82,2 \pm 1,4$ Manhattan, $83,7 \pm 4,4$ KL e $70,5 \pm 4,0$ Mahalanobis.

Para o grupo “saudável \times leve”, o melhor desempenho foi obtido com a combinação da métrica KL (ReliefF 50+), alcançando uma precisão de 99,0% e revocação de 9,7%. A combinação da métrica KL (ReliefF 50+), também apresentou o melhor desempenho para o grupo “saudável \times moderada”, com uma precisão de 97,3% e revocação de 13,0%. Para a comparação entre o grupo “saudável \times severa”, a métrica KL (ReliefF 50+) obteve a melhor precisão de 98,3% e revocação de 9,5%. O grupo “leve \times moderada”, teve seu melhor desempenho com a métrica KL (ReliefF 10+), alcançando uma precisão de 78,9% e revocação de 8,8%. O melhor resultado para o grupo “leve \times severa”, foi com a métrica Manhattan (ReliefF 10+), apresentando precisão de 80,1% e revocação de 10,8%. Por fim, a comparação “moderada \times severa”, teve o melhor desempenho com a métrica KL (ReliefF 10+), alcançando uma precisão de 84,9% e revocação de 9,1%.

Para “saudável \times leve”, Figura 25(a), a KL mantém uma precisão inicial próxima de 95%, que se sustenta até cerca de 80% de revocação, onde começa a cair rapidamente para cerca de 60%. As distâncias Manhattan, Euclidiana e Mahalanobis também apresentam alta precisão inicial, mas caem de forma mais acentuada após 90% de revocação.

Para “saudável \times moderada”, Figura 25(b), a KL inicia com alta precisão em torno de 95% e mantém esse desempenho até aproximadamente 70% de revocação, caindo para cerca de 60%. As distâncias Manhattan e Euclidiana começam com alta precisão, mas caem mais rapidamente após 80% de revocação, enquanto Mahalanobis tem a precisão inicial mais baixa, em torno de 75%, e permanece estável, mas inferior às outras distâncias.

A Figura 25(c), para o grupo “saudável \times severa”, mostra que a métrica KL mantém a melhor precisão inicial, próxima de 95%, caindo gradualmente para cerca de 65% à medida que a revocação aumenta. As distâncias Manhattan e Euclidiana também apresentam boa precisão inicial, em torno de 90%, mas caem de forma mais acentuada após 80% de revocação, chegando a cerca de 60%. A distância Mahalanobis tem a menor precisão inicial, começando em torno de 75% e diminuindo para cerca de 55% ao longo da revocação.

No Figura 25(d), “leve \times moderada”, a precisão inicial é aproximadamente 70% para todas as distâncias, com a KL apresentando uma leve vantagem inicial. À medida que a revocação aumenta, todas as distâncias convergem para cerca de 55–60% de precisão. Para “leve \times severa”, Figura 25(e), a precisão inicial é mais alta, cerca de 80% para KL, mas todas as distâncias começam a declinar consistentemente, com a KL mantendo uma ligeira superioridade até cerca de 60% de revocação, onde se iguala às demais distâncias em torno de 55%. No grupo “moderada \times severa”, Figura 25(f), a precisão inicial para todas as distâncias é em torno de 70 – 75%, com uma queda gradual e uniforme para cerca de 50 – 55% à medida que a revocação aumenta. Em

resumo, a métrica KL mantém uma vantagem inicial em todos os grupos, especialmente notável nas primeiras imagens recuperadas, mas todas as distâncias tendem a convergir para valores semelhantes de precisão com o aumento da revocação.

Os resultados indicam que a integração da engenharia de características com a ResNet-50 aprimorou de maneira significativa os resultados para a métrica KL, assegurando uma maior precisão nas primeiras imagens recuperadas. Essa combinação mostrou-se eficaz em manter uma alta precisão inicial, aspecto crucial para a eficácia do sistema de recuperação de imagens. Além disso, ao selecionar as características mais relevantes, houve um aumento na precisão e uma maior sustentabilidade desse valor à medida que a revocação avança. Ademais, os descritores da engenharia de características conseguiram compensar parcialmente os descritores nulos presentes no modelo ResNet-50.

Tabela 12 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor Handcraft + ResNet-50 e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
Engenharia de Características + ResNet-50	50+	Euclideana	94,3	13,2	89,9	13,0	86,0	12,1	67,6	9,5	71,6	9,7	67,2	9,4
		Manhattan	95,2	13,3	93,7	13,6	92,7	13,0	69,4	9,7	75,3	10,2	69,7	9,7
		Kullback-Leibler (KL)	99,0	9,7	97,3	13,0	98,3	9,5	67,7	8,1	68,8	8,5	69,3	9,5
		Mahalanobis	95,0	13,3	57,1	8,3	75,7	10,6	64,1	9,0	69,9	9,4	68,1	9,5
	40+	Euclideana	94,5	13,2	87,8	12,7	89,4	12,5	67,5	9,5	72,8	9,8	67,1	9,4
		Manhattan	95,2	13,3	92,2	13,4	93,6	13,1	68,0	9,5	76,8	10,4	68,3	9,5
		Kullback-Leibler (KL)	99,0	9,9	96,9	12,4	97,9	9,7	65,7	7,6	62,3	8,6	71,5	9,2
		Mahalanobis	93,5	13,1	67,1	9,7	82,8	11,6	62,7	8,8	69,5	9,4	65,6	9,2
	30+	Euclideana	94,7	13,3	89,4	13,0	88,3	12,4	66,9	9,4	72,8	9,8	69,5	9,7
		Manhattan	95,0	13,3	93,0	13,5	92,7	13,0	67,3	9,4	76,6	10,3	70,5	9,8
		Kullback-Leibler (KL)	98,7	10,5	96,5	12,9	98,0	9,4	72,7	8,1	64,2	8,6	78,5	9,1
		Mahalanobis	93,3	13,0	53,8	7,8	80,0	11,2	64,2	9,0	71,7	9,7	65,9	9,2
	20+	Euclideana	94,4	13,2	89,6	13,0	89,0	12,5	66,8	9,4	72,8	9,8	67,1	9,4
		Manhattan	94,5	13,2	92,9	13,5	91,2	12,8	67,1	9,4	77,4	10,5	68,7	9,6
		Kullback-Leibler (KL)	98,7	11,5	95,9	13,0	96,4	10,5	78,1	8,7	65,3	8,7	82,4	8,9
		Mahalanobis	93,2	13,0	55,9	8,1	72,3	10,1	61,3	8,6	69,4	9,4	62,9	8,8
	10+	Euclideana	95,0	13,3	89,1	12,9	85,2	11,9	64,9	9,1	74,7	10,1	68,9	9,6
		Manhattan	95,0	13,3	91,6	13,3	86,0	12,1	66,6	9,3	80,1	10,8	71,3	10,0
		Kullback-Leibler (KL)	97,2	12,5	94,1	13,1	89,3	11,1	78,9	8,8	48,7	9,0	84,9	9,1
		Mahalanobis	91,7	12,8	64,3	9,3	63,4	8,9	61,2	8,6	57,0	7,7	63,6	8,9

Tabela 13 – Resultados do Teste de *Kruskal-Wallis* (Engenharia de Características + ResNet-50)

Métrica	Estatística	p – Valor
Precisão	4,6667	0,1979
Revocação	4,9848	0,1729

Os p – Valor dos testes de *Kruskal-Wallis* (Tabela 13 para as métricas de precisão e retenção são 0,1979 e 0,1729, respectivamente, maiores que o nível de significância comum de 0,05. Isso indica que não há diferenças estatisticamente significativas nas precisões e recalls entre as várias distâncias utilizadas. Em outras palavras, as medidas de precisão e revocação entre todas as distâncias avaliadas são, em média, comparáveis. As métricas de precisão e o revocação dos resultados não são significativamente afetados pela escolha da distância.

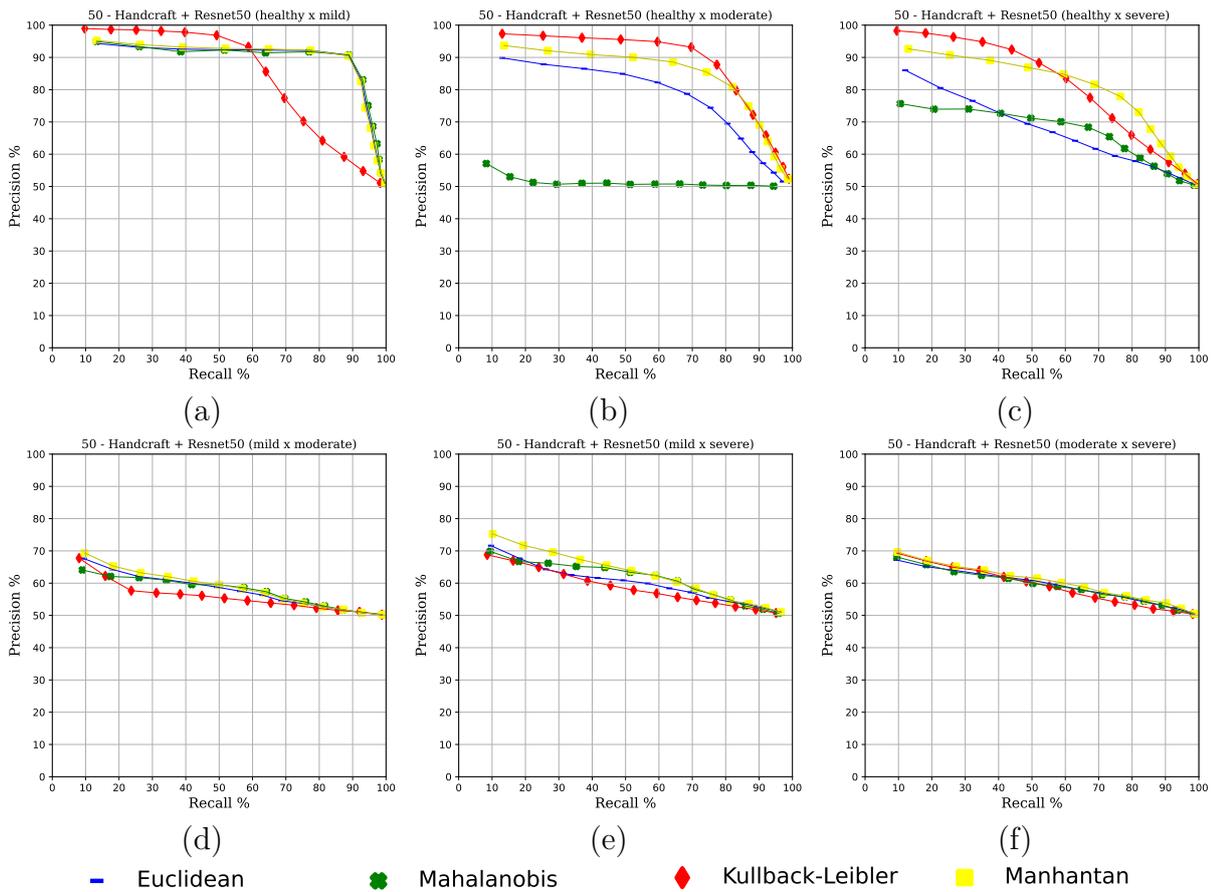


Figura 25 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 50 - Engenharia de Características + ResNet50 (leve x moderada) (c) 50 - Engenharia de Características + ResNet50 (moderada x severa) (d) 50 - Engenharia de Características + ResNet50 (leve x moderada) (e) 50 - Engenharia de Características + ResNet50 (leve x severa) (f) 50 - Engenharia de Características + ResNet50 (moderada x severa)

5.1.7 Descritores Obtidos por Engenharia de Características e Modelos CNNs

Nessa etapa, uma avaliação foi realizada com os descritores baseados em engenharia de característica, AlexNet e ResNet-50. A Tabela 14 apresenta os valores das medidas com as distâncias e intervalos de valores das características. Os grupos são apresentados em uma investigação binária entre as classes de lesões. Os valores médios e desvio padrão foram $86,1 \pm 1,0$, Euclidiana $83,9 \pm 1,0$ Manhattan, $88,3 \pm 3,5$ KL e $70,9 \pm 3,4$ Mahalanobis.

Para o grupo “saudável \times leve”, a combinação da métrica KL (ReliefF 50+) obteve o desempenho mais elevado, com uma precisão de 98,7% e revocação de 11,5%. No grupo “saudável \times moderada”, a métrica KL (ReliefF 50+) também liderou com precisão de 97,8% e revocação de 12,6%, evidenciando sua robustez na identificação de displasias moderadas. Para as comparações entre “saudável \times severa”, a métrica KL (ReliefF 40+) obteve resultados impressionantes, atingindo uma precisão de 99,9% e revocação de 12,3%. No grupo “leve \times moderada”, a combinação de KL (ReliefF 20+) continuou a mostrar superioridade com uma precisão de 83,4% e revocação de 8,5%. Para o grupo “leve \times severa”, a métrica KL (ReliefF 20+) destacou-se, atingindo precisão de 82,9% e revocação de 10,0%. Finalmente, no grupo “moderada \times severa”, a combinação KL (ReliefF 10+) demonstrou a melhor performance, com precisão de 85,7% e revocação de 9,0%.

A combinação das características permitiu capturar uma gama mais ampla de informações relevantes das imagens, melhorando a discriminação entre diferentes classes. A AlexNet, com sua arquitetura menos profunda, foi eficaz na captura de padrões mais globais e texturais das imagens. Em contraste, a ResNet-50, com suas camadas residuais mais profundas, foi capaz de extrair características mais detalhadas e específicas e a Engenharia de Características adicionaram uma camada de informação baseada em conhecimento prévio sobre as imagens segmentadas.

Na Figura 26(a), o grupo “saudável \times leve”, mostra que a métrica KL apresenta a melhor precisão inicial, próxima de 99%. Embora a precisão da KL diminua gradualmente com o aumento da revocação, ela ainda se mantém acima de 70% até cerca de 80% de revocação, caindo para aproximadamente 60% no final. As distâncias Manhattan e Euclidiana também iniciam com alta precisão, cerca de 90 – 95%, mas sofrem uma queda mais acentuada após 80% de revocação, chegando a 60%. A distância Mahalanobis, por outro lado, começa com uma precisão mais baixa, em torno de 85%, e diminui de forma mais estável para cerca de 55 – 60% ao longo da revocação.

No grupo “saudável \times moderada”, Figura 26(b), revela que a métrica KL mantém a melhor precisão inicial, próxima de 95%. À medida que a revocação aumenta, a precisão da KL diminui gradualmente, caindo para cerca de 70% aos 80% de revocação e atingindo aproximadamente 60% no final. As distâncias Manhattan e Euclidiana começam com precisão inicial elevada, em torno de 90 – 95%, mas também apresentam quedas acentuadas após 80% de revocação, convergindo para cerca de 60%. A distância Mahalanobis inicia com uma precisão mais baixa, em torno de 85%, e diminui de forma mais gradual, alcançando cerca de 60% ao final da revocação. Embora todas as distâncias mostrem uma diminuição na precisão com o aumento da revocação, a KL destaca-se por manter um desempenho superior ao longo de um maior intervalo de revocação.

No grupo “saudável \times severa”, Figura 26(c), indica que a métrica KL mantém a maior precisão inicial, próxima de 99%. No entanto, a precisão da KL diminui gradualmente para cerca de 70% aos 80% de revocação, e cai para aproximadamente 60% no final. A distância Manhattan também inicia com uma precisão elevada, em torno de 90%, mas apresenta uma queda mais acentuada após 70% de revocação, terminando em cerca de 60%. A distância Euclidiana começa com aproximadamente 85% de precisão, mas sua curva é mais íngreme, caindo para 50% ao final. A distância Mahalanobis, por sua vez, mantém uma precisão estável, começando em torno de 70% e permanecendo próxima de 50% ao longo da revocação. Embora todas as distâncias mostrem uma diminuição na precisão com o aumento da revocação, a KL destaca-se por manter um desempenho superior ao longo de um maior intervalo de revocação, enquanto as distâncias Euclidiana e Mahalanobis apresentam quedas mais acentuadas.

No Figura 26(d), “leve \times moderada”, a precisão inicial para todas as distâncias está em torno de 70 – 80%, com a métrica KL mostrando um leve superioridade inicial. À medida que a revocação aumenta, todas as distâncias convergem para cerca de 55 – 60% de precisão. Para “leve \times severa”, Figura 26(e), a precisão inicial é maior, com a KL começando em torno de 85%, mas todas as distâncias apresentam um declínio gradual. A KL mantém uma vantagem inicial, mas todas as distâncias convergem para cerca de 50 – 60% ao final da revocação. No grupo “moderada \times severa”, Figura 26(f), a precisão inicial para todas as distâncias está entre 70 – 75%, com uma queda gradual e uniforme para cerca de 50 – 55% à medida que a revocação aumenta.

Em resumo, a métrica KL mantém uma vantagem inicial em todos os gráficos, destacando-se nas primeiras imagens recuperadas. No entanto, à medida que a revocação aumenta, todas as distâncias tendem a convergir para valores de precisão semelhantes, mostrando uma robustez na combinação de descritores Engenharia de Características, ResNet50 e AlexNet.

Tabela 14 – Valores da métricas Precisão (Prec.) e Revocação (Rev.) do descritor Engenharia de Características + ResNet-50 + AlexNet e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

Descritores	ReliefF	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
			Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
Engenharia de Características + AlexNet + ResNet-50	50+	Euclidean	91,5	12,8	92,2	13,4	89,9	12,5	71,8	10,0	73,6	9,9	68,0	9,5
		Manhattan	93,7	13,1	94,3	13,7	93,0	13,0	72,2	10,1	78,0	10,5	71,4	10,0
		Kullback-Leibler (KL)	98,7	11,5	97,8	12,6	99,7	12,3	77,5	8,3	82,9	10,0	72,9	9,6
		Mahalanobis	82,4	11,5	88,4	12,8	54,4	7,6	67,7	9,5	70,7	9,6	68,4	9,6
	40+	Euclidean	92,2	12,9	92,0	13,3	90,8	12,7	71,3	10,0	74,2	10,0	68,2	9,5
		Manhattan	94,1	13,2	94,1	13,6	92,7	13,0	71,3	9,9	78,9	10,7	71,5	10,0
		Kullback-Leibler (KL)	98,6	11,4	97,3	11,5	99,9	12,3	75,0	8,3	81,0	10,0	74,9	9,3
		Mahalanobis	80,6	11,3	92,3	13,4	55,4	7,7	68,7	9,7	71,4	9,6	69,2	9,7
	30+	Euclidean	92,9	13,0	92,8	13,4	89,4	12,5	71,6	10,0	75,2	10,2	70,5	9,8
		Manhattan	94,8	13,3	94,3	13,7	91,9	12,8	71,3	9,9	78,9	10,7	72,8	10,2
		Kullback-Leibler (KL)	98,6	11,3	97,0	12,0	99,7	12,1	77,5	8,5	78,9	9,9	82,4	9,6
		Mahalanobis	76,2	10,6	92,2	13,4	56,8	7,9	66,5	9,4	65,9	8,9	68,5	9,6
20+	Euclidean	93,4	13,1	91,1	13,2	89,8	12,5	71,7	10,0	74,8	10,1	68,5	9,5	
	Manhattan	95,7	13,4	93,7	13,6	91,8	12,8	72,8	10,2	80,2	10,8	71,5	10,0	
	Kullback-Leibler (KL)	97,8	11,0	96,8	12,3	99,2	11,8	83,4	8,5	80,3	9,7	83,8	9,2	
	Mahalanobis	76,9	10,8	91,6	13,3	55,3	7,7	67,4	9,5	63,9	8,6	65,0	9,1	
10+	Euclidean	95,5	13,4	91,4	13,2	87,1	12,2	70,4	9,9	76,2	10,3	70,0	9,8	
	Manhattan	96,7	13,5	93,8	13,6	88,8	12,4	70,2	9,8	80,3	10,8	72,2	10,1	
	Kullback-Leibler (KL)	97,6	12,4	95,5	12,8	96,9	11,4	81,5	8,8	59,8	9,6	85,7	9,0	
	Mahalanobis	75,7	10,6	87,6	12,7	65,2	9,1	63,9	9,0	55,2	7,5	64,3	9,0	

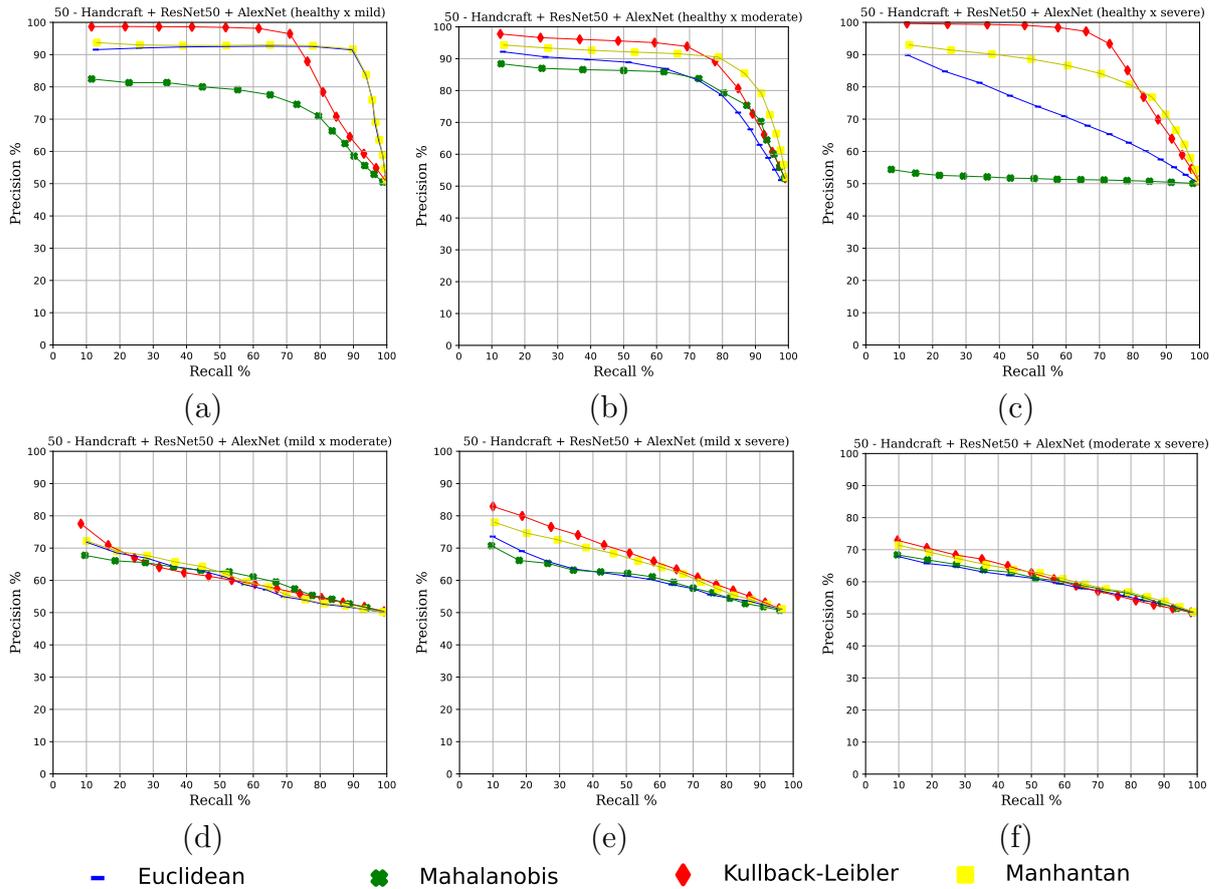


Figura 26 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + ResNet50 + AlexNet (saúdável x leve) (b) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 50 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 50 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)

Tabela 15 – Resultados do Teste de *Kruskal-Wallis* (Engenharia de Características + AlexNet + RestNet-50)

Métrica	Estatística	$p - Valor$
Precisão	8,2466	0,0411
Revocação	5,2019	0,1575

Para a métrica de Precisão (Tabela 15), o $p - Valor$ é de 0,0412, que é ligeiramente menor do que o nível de significância comum de 0,05. Isso indica que há diferenças estatisticamente significativas nas precisões entre as diferentes distâncias utilizadas. Em outras palavras, as precisões medidas com as distâncias Euclidiana, Manhattan, Kullback-Leibler e Mahalanobis não são todas iguais, e pelo menos uma dessas distâncias se comporta de maneira diferente das outras em termos de precisão. Considerando um nível de significância mais rigoroso de 0,1, o $p - Valor$ é de 0,0412 também menor do que 0,1. Isso reforça ainda mais a conclusão de que há diferenças estatisticamente significativas nas precisões entre as diferentes distâncias

utilizadas. Portanto, tanto ao nível de significância de 0,05 quanto de 0,1, a escolha da distância afeta significativamente a precisão dos resultados. Para a métrica de Revocação (Tabela 15), o p -Valor é de 0,1576, que é maior do que o nível de significância comum de 0,05 e de 0,1. Isso indica que não há diferenças estatisticamente significativas entre a revocação das diferentes distâncias utilizadas.

A avaliação dos resultados obtidos utilizando a combinação destes descritores, revela insights importantes sobre a eficácia das diferentes métricas de distância na classificação de displasias orais. Comparando com o estado da arte, os métodos de classificação de displasias orais geralmente dependem de redes neurais convolucionais (CNNs) isoladas, como AlexNet ou ResNet-50, ou de descritores baseados Engenharia de Características. Estas abordagens têm suas respectivas vantagens e limitações. Redes neurais como AlexNet e ResNet-50 são reconhecidas por suas capacidades de extração automática de características e classificação de imagens. AlexNet é conhecida por sua arquitetura relativamente simples e eficaz, enquanto ResNet-50 se destaca por sua profundidade e uso de conexões residuais, permitindo um melhor aprendizado em redes profundas. Por outro lado, a Engenharia de Características envolve a extração manual de características relevantes das imagens, baseada em conhecimentos especializados. Embora esta abordagem possa ser limitada pela subjetividade e pelo conhecimento do especialista, ela pode capturar detalhes específicos que podem não ser facilmente aprendidos por redes neurais automáticas. A combinação de Engenharia de Características com AlexNet e ResNet-50 visa unir o melhor dos dois mundos: a precisão e a riqueza das características manuais com a capacidade de generalização e extração automática das redes profundas.

Este trabalho avança o estado da arte na classificação de displasias orais ao demonstrar que a combinação de Engenharia de Características com AlexNet e ResNet-50 supera significativamente as abordagens tradicionais, oferecendo uma nova direção promissora para pesquisas futuras e aplicações clínicas na detecção e classificação de displasias orais. Os achados deste estudo têm implicações práticas significativas para o diagnóstico clínico, oferecendo uma abordagem mais precisa e confiável para a detecção e classificação de displasias orais, potencialmente melhorando os resultados dos pacientes através de diagnósticos mais precoces e precisos.

Conclusão

Os sistemas de recuperação de imagens baseados em conteúdo (CBIR) têm tido ampla aplicação na análise e recuperação de imagens, especialmente em áreas como a medicina e a biologia. No entanto, as aplicações atuais desses sistemas apresentam limitações evidentes devido à utilização de métricas tradicionais, que muitas vezes não capturam adequadamente a complexidade dos dados. Para mitigar essas limitações, alguns estudos têm explorado a aplicação de sistemas CBIR em conjunto com diversas medidas de similaridade, visando tratar a descontinuidade semântica presente nas imagens.

Neste trabalho, investigamos a eficácia das divergências de Bregman, especificamente as divergências Kullback-Leibler (KL) e Mahalanobis, para computar a dissimilaridade entre vetores de características que representam dados de imagens histológicas de displasia bucal. Essas divergências foram comparadas com as métricas tradicionais, como a distância Euclidiana e Manhattan, para avaliar sua capacidade de diferenciar entre os diversos níveis de displasia bucal (saudável, leve, moderado e severa).

Para a avaliação da metodologia proposta, foram apresentados os resultados de precisão e revocação entre os diferentes níveis de displasia bucal ("saudável × moderada", "saudável × severa", "leve × moderada", "leve × severa", "moderada × severa"), utilizando diferentes combinações de técnicas para extração das características. Os resultados indicam que, na maioria dos casos, o uso das divergências de Bregman, particularmente a divergência KL, superou as métricas Euclidiana e Manhattan. Isso mostra que as divergências de Bregman são mais eficazes na identificação de características que melhor distinguem os níveis de displasia, contribuindo para uma recuperação de imagens mais precisa e relevante.

No primeiro experimento utilizando apenas a rede neural AlexNet, o menor ganho obtido, comparando apenas a distância Euclidiana em relação a distância KL, foi de 0,7%. Isso foi determinado pela média de todas as características relevantes (50/40/30/20/10) usando Relief. No segundo experimento utilizando apenas a rede neural ResNet-50, a distância Euclidiana foi superior a distância KL em 1,4%. O baixo desempenho nos dois primeiros experimentos é reflexo direto de descritores iguais a zero, que impactaram diretamente na precisão da divergência de KL mesmo com o tratamento do ϵ . No quarto experimento (Engenharia de Características), utilizamos uma rede Mask R-CNN para extrair características morfológicas e não morfológicas de imagens segmentadas, permitindo uma análise detalhada das informações internas das regiões

nas imagens histológicas. Neste experimento a distância KL foi superior a distância Euclidiana em 16,5%, sendo o maior ganho obtido utilizando a divergência de Bregman aplicada a distância KL. No terceiro, quinto, sexto e sétimo experimento a utilização de diferentes redes neurais para a extração de características de imagens e a combinação das informações obtidas ajudou a melhorar a precisão na recuperação de imagens obtendo representações mais robustas e discriminativas das imagens. Respectivamente para cada experimento tivemos um aumento em relação a distância Euclidiana de 7,3%, 10,4%, 4,9% e 8,2%.

Os testes de *Kruskal-Wallis* foram aplicados para avaliar a significância das diferenças nas métricas de Precisão e Revocação entre as diferentes distâncias em várias configurações de redes neurais e engenharia de características. Os principais resultados indicam que para a AlexNet, o $p - Valor$ do teste para Precisão foi de 0,9600, o que não é significativo, enquanto para o Revocação foi de 0,0350, indicando uma diferença significativa. Na ResNet-50, tanto a Precisão ($p - Valor = 0,6399$) quanto o Revocação ($p - Valor = 0,1501$) não apresentaram diferenças significativas entre as distâncias. Na configuração AlexNet + ResNet-50, o $p - Valor$ foi de 0,0589 para Precisão e 0,1343 para Revocação, ambos não significativos, embora a Precisão esteja próxima do nível de significância de 0,05 e se tornando significativa ao nível de 0,1.

Na análise da Engenharia de Características, o $p - Valor$ foi de 0,0099 para Precisão e 0,0409 para Revocação, ambos indicando diferenças significativas entre as distâncias. No entanto, quando combinada com AlexNet, $p - Valor$ foi de 0,2328 para Precisão e 0,1272 para Revocação, não apresentando significância estatística. A combinação Engenharia de Características + ResNet-50 também não apresentou diferenças significativas, com o $p - Valor$ de 0,1979 para Precisão e 0,1729 para Revocação. Finalmente, na configuração Engenharia de Características + AlexNet + ResNet-50, o $p - Valor$ para Precisão foi de 0,0412, indicando uma diferença significativa, enquanto para Revocação foi de 0,1576, não significativo.

Esses resultados sugerem que as distâncias Euclidiana, Manhattan, KL e Mahalanobis apresentam diferenças significativas na métrica de Precisão em algumas configurações, especialmente com a Engenharia de Características e sua combinação com outras redes neurais. Por outro lado, as diferenças nas distâncias não são estatisticamente significativas na maioria das configurações de Revocação, exceto para a AlexNet e a Engenharia de Características isoladamente. Mesmo que as diferenças nas médias não sejam significativamente diferentes em todas as configurações, a análise gráfica demonstra que a distância KL possui uma superioridade notável em relação às outras distâncias. Isso é evidente não apenas nas primeiras imagens recuperadas, mas também na manutenção da estabilidade da precisão com o aumento do número de imagens recuperadas nos experimentos com a união de descritores. Portanto, a distância KL se destaca como a mais eficaz, garantindo uma performance superior em termos de precisão e estabilidade em múltiplos cenários de teste.

As divergências de Bregman, especialmente quando aplicada a métrica KL, desempenham um papel crucial no processo de recuperação de imagem em displasias orais. A utilização dessas métricas permite uma avaliação mais precisa e robusta das imagens, o que facilita a identificação e classificação de diferentes níveis de displasia. Essa precisão aprimorada auxilia significativamente os profissionais de saúde na tomada de decisões, fornecendo-lhes ferramentas mais eficazes para o diagnóstico e tratamento das condições orais.

Comparado ao estado da arte, a quantidade de pesquisas que utilizam divergências de Bregman e CBIR no processo de recuperação de imagens relacionadas a displasias orais é notavelmente escassa. Essa lacuna na literatura evidencia a inovação e a forte contribuição desta abordagem dentro do campo. A aplicação das divergências de Bregman, e em particular da métrica KL, não só melhora a precisão da recuperação de imagens, mas também estabelece uma nova direção para futuras pesquisas e práticas clínicas, promovendo avanços significativos na área de saúde bucal.

6.1 Trabalhos Futuros

Para trabalhos futuros, propõe-se a investigação de outras medidas de distâncias utilizando divergências de Bregman e a divergência de Bregman generalizada, que podem oferecer diferentes perspectivas e melhorias na recuperação de imagens. Essas medidas podem ser exploradas para verificar se proporcionam ganhos adicionais de precisão e robustez na identificação e classificação de displasias orais.

Além disso, a aplicação dessas divergências em outros descritores e a utilização de arquiteturas de CNNs mais avançadas e recentes representam uma área promissora de pesquisa. A integração dessas redes mais modernas pode potencialmente aprimorar ainda mais a precisão e a eficiência dos sistemas CBIR.

Outra vertente interessante para futuros estudos é a aplicação de divergências de Bregman em descritores com imagens já segmentadas. Segmentar as áreas de interesse antes da aplicação das CNNs pode permitir uma análise mais focada e precisa, eliminando ruídos, e adquirindo informações profundas sobre os núcleos.

Além das pesquisas teóricas e experimentais, existe a possibilidade de desenvolver uma aplicação prática e utilizável para os profissionais de saúde. Essa aplicação poderia integrar os métodos avançados de recuperação de imagens utilizando divergências de Bregman com uma interface intuitiva e acessível. A ferramenta permitiria que os profissionais realizassem diagnósticos mais rápidos e precisos, facilitando a tomada de decisões clínicas com base em dados visualizados de maneira eficiente.

6.2 Contribuições em Produção Bibliográfica

O resultado desta dissertação foi publicado em:

SOARES, Tiago R. M.; SILVA, Adriano B.; LOYOLA, Adriano M.; CARDOSO, Sérgio V.; FARIA, Paulo R. de; NEVES, Leandro A.; NASCIMENTO, Marcelo Z.; RAZENTE, Humberto. Divergências de Bregman aplicadas na Recuperação por Conteúdo de Imagens de Displasias Orais. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE (SBCAS), 24. , 2024, Goiânia/GO. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 130-141. ISSN 2763-8952. DOI: <<https://doi.org/10.5753/sbcas.2024.2058>> (SOARES et al., 2024).

Referências

ALBUQUERQUE, M. P. de; ALBUQUERQUE, M. P. de. Processamento de imagens: métodos e análises. **Centro Brasileiro de Pesquisas Físicas MCT**, 2000.

ALOM, M. Z. et al. **The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches**. arXiv, 2018. Disponível em: <<https://arxiv.org/abs/1803.01164>>.

ALVES, M. C. F. B. et al. Uso de scanner de lâminas histológicas no processo de aprendizado de patologia entre acadêmicos de medicina / use of histological blade scanner in the pathology learning process among medicine students. **Brazilian Journal of Health Review**, South Florida Publishing LLC, v. 4, n. 5, p. 22098–22103, out. 2021. ISSN 2595-6825. Disponível em: <<http://dx.doi.org/10.34119/bjhrv4n5-307>>.

ATLAM GAMAL ATTIYA, N. E.-f. H. F. Comparative study on cbir based on color feature. **International Journal of Computer Applications**, v. 78, n. 16, p. 9–15, September 2013. ISSN 0975-8887. Disponível em: <<https://ijcaonline.org/archives/volume78/number16/13605-1387/>>.

BANERJEE, A. et al. Clustering with bregman divergences. In: **SIAM International Conference on Data Mining (SDM)**. [S.l.]: SIAM, 2004. p. 234–245. doi:10.1137/1.9781611972740.22.

_____. Clustering with bregman divergences. **J. Mach. Learn. Res.**, v. 6, p. 1705–1749, 2005. Disponível em: <<http://jmlr.org/papers/v6/banerjee05b.html>>.

BOYTSOV, L.; NYBERG, E. Pruning algorithms for low-dimensional non-metric k-NN search: A case study. In: **International Conference on Similarity Search and Applications (SISAP)**. Newark, NJ, USA: Springer, 2019, (LNCS, v. 11807). p. 72–85. Disponível em: <https://doi.org/10.1007/978-3-030-32047-8_7>.

BREGMAN, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. **USSR computational mathematics and mathematical physics**, Elsevier, v. 7, n. 3, p. 200–217, 1967.

BRITO, E. R. F. Aplicações de redes neurais convolucionais no reconhecimento de danos em veículos. **REVISTA FOCO**, South Florida Publishing LLC, v. 16, n. 10, p. e3223, out. 2023. ISSN 1981-223X. Disponível em: <<http://dx.doi.org/10.54751/revistafoco.v16n10-003>>.

CUN, Y. L. et al. Handwritten digit recognition: applications of neural network chips and automatic learning. **IEEE Communications Magazine**, Institute of Electrical and

Electronics Engineers (IEEE), v. 27, n. 11, p. 41–46, nov. 1989. ISSN 0163-6804. Disponível em: <<http://dx.doi.org/10.1109/35.41400>>.

DEZA, M. M.; DEZA, E. **Encyclopedia of Distances**. Springer Berlin Heidelberg, 2016. ISBN 9783662528440. Disponível em: <<http://dx.doi.org/10.1007/978-3-662-52844-0>>.

DIAS, L. L. Displasias epiteliais orais: Um estudo descritivo. **Anais dos Seminários de Iniciação Científica**, Universidade Estadual de Feira de Santana, n. 24, out. 2021. ISSN 2595-0339. Disponível em: <<http://dx.doi.org/10.13102/semic.vi24.6757>>.

DOMINGUEZ, G. C. et al. How to prepare histological sections of non-decalcified tissue with metal implants?: Modified technique. **RPG. Revista de Pós-Graduação**, v. 19, n. 2, p. 81–87, 2012.

EVERS III, C. D. et al. Histology, dimensions, and fluorescein staining characteristics of nodular and cuticular drusen in age-related macular degeneration. **Retina**, LWW, p. 10–1097, 2022.

FARIA, P. R. et al. Absence of galectin-3 does not affect the development of experimental tongue carcinomas in mice. **Experimental and Molecular Pathology**, Elsevier BV, v. 90, n. 2, p. 189–193, abr. 2011. Disponível em: <<https://doi.org/10.1016/j.yexmp.2010.12.007>>.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, IBICT, v. 35, n. 1, p. 25–30, abr. 2006. ISSN 0100-1965. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652006000100003>>.

FERREIRA, D. P. L. **Divergências de Bregman e total Bregman aplicadas na análise de imagens**. Tese (Doutorado) — Programa de Pós-graduação em Ciência da Computação, Universidade Federal de Uberlândia, 2017. Disponível em: <<http://dx.doi.org/10.14393/ufu.te.2018.791>>.

FERREIRA, D. P. L.; BACKES, A. R.; BARCELOS, C. A. Z. Bregman divergence applied to hierarchical segmentation problems. In: SPRINGER. **Iberoamerican Congress on Pattern Recognition**. [S.l.], 2015. p. 493–500.

FERREIRA, D. P. L.; RIBEIRO, E.; BARCELOS, C. A. Z. Variational non rigid registration with bregman divergences. In: **ACM Symposium on Applied Computing (SAC)**. Marrakech, Morocco: ACM, 2017. p. 158–163. Disponível em: <<https://doi.org/10.1145/3019612.3019646>>.

GONÇALVES, W. G. e. et al. Classificação de estratos florestais utilizando redes neurais artificiais e dados de sensoriamento remoto. **Revista Ambiente & Água**, SciELO Brasil, v. 11, n. 3, p. 612–624, 2016.

GONZALEZ, R. C.; WOODS, R. C. **Processamento digital de imagens**. [S.l.]: Pearson Educación, 2009.

GURINA, T. S.; SIMMS, L. Histology, staining. 2020.

HE, K. et al. **Deep Residual Learning for Image Recognition**. arXiv, 2015. Disponível em: <<https://arxiv.org/abs/1512.03385>>.

IRSHAD, H. et al. Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—current status and future potential. **IEEE Reviews in Biomedical Engineering**, Institute of Electrical and Electronics Engineers (IEEE), v. 7, p. 97–114, 2014. ISSN 1941-1189. Disponível em: <<http://dx.doi.org/10.1109/RBME.2013.2295804>>.

JIAWEI, H.; MICHELINE, K. **Data mining: concepts and techniques**. [S.l.]: Morgan kaufmann, 2006.

JÚNIOR, J. C. C. X.; OCANHA-XAVIER, J. P. What does the 4th edition of the world health organization classification of head and neck tumors (2017) bring new about mucosal melanomas? **Anais Brasileiros de Dermatologia**, FapUNIFESP (SciELO), v. 93, n. 2, p. 259–260, mar. 2018. ISSN 0365-0596. Disponível em: <<http://dx.doi.org/10.1590/abd1806-4841.20187206>>.

KHODASKAR, A.; LADHAKE, S. A. Pattern recognition: Advanced development, techniques and application for image retrieval. In: **2014 International Conference on Communication and Network Technologies**. IEEE, 2014. Disponível em: <<http://dx.doi.org/10.1109/CNT.2014.7062728>>.

KRISHNAN, M. M. R. et al. Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis—an svm based approach. **Computers in biology and medicine**, Elsevier, v. 39, n. 12, p. 1096–1104, 2009.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. **Communications of the ACM**, Association for Computing Machinery (ACM), v. 60, n. 6, p. 84–90, maio 2017. ISSN 1557-7317. Disponível em: <<http://dx.doi.org/10.1145/3065386>>.

LI, X.; YANG, J.; MA, J. Recent developments of content-based image retrieval (cbir). **Neurocomputing**, Elsevier BV, v. 452, p. 675–689, set. 2021. ISSN 0925-2312. Disponível em: <<http://dx.doi.org/10.1016/j.neucom.2020.07.139>>.

LIU, M. et al. Shape retrieval using hierarchical total bregman soft clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Institute of Electrical and Electronics Engineers (IEEE), v. 34, n. 12, p. 2407–2419, dez. 2012. ISSN 2160-9292. Disponível em: <<http://dx.doi.org/10.1109/TPAMI.2012.44>>.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **Int. J. Comput. Vis.**, v. 60, n. 2, p. 91–110, 2004. Disponível em: <<https://doi.org/10.1023/B:VISI.0000029664.99615.94>>.

LUMERMAN, H.; FREEDMAN, P.; KERPEL, S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, Elsevier BV, v. 79, n. 3, p. 321–329, mar. 1995. Disponível em: <[https://doi.org/10.1016/s1079-2104\(05\)80226-4](https://doi.org/10.1016/s1079-2104(05)80226-4)>.

MAIA, H. C. d. M. et al. Potentially malignant oral lesions: clinicopathological correlations. **Einstein (São Paulo)**, FapUNIFESP (SciELO), v. 14, n. 1, p. 35–40, mar. 2016. ISSN 1679-4508. Disponível em: <<http://dx.doi.org/10.1590/S1679-45082016AO3578>>.

MCLACHLAN, G. J. Mahalanobis distance. **Resonance**, Springer Science and Business Media LLC, v. 4, n. 6, p. 20–26, jun. 1999. ISSN 0973-712X. Disponível em: <<http://dx.doi.org/10.1007/BF02834632>>.

MENDONÇA, D. F. et al. The inactive form of glycogen synthase kinase-3 β is associated with the development of carcinomas in galectin-3 wild-type mice, but not in galectin-3-deficient mice. **Int. J. Clin. Exp. Pathol.**, v. 5, n. 6, p. 547–554, jul. 2012.

MORAVEC, H. P. Rover visual obstacle avoidance. In: HAYES, P. J. (Ed.). **7th International Joint Conference on Artificial Intelligence (IJCAI)**. [S.l.: s.n.], 1981. p. 785–790.

MOURA, E. R. et al. Metaplasia e lesões histopatológicas de mucosa esofágica. **Research, Society and Development**, Research, Society and Development, v. 11, n. 2, p. e29711225778, jan. 2022. ISSN 2525-3409. Disponível em: <<http://dx.doi.org/10.33448/rsd-v11i2.25778>>.

NETO, L. A. A.; TEIXEIRA, L. A. De doença da civilização a problema de saúde pública: câncer, sociedade e medicina brasileira no século xx. **Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas**, FapUNIFESP (SciELO), v. 12, n. 1, p. 173–188, jan. 2017. ISSN 1981-8122. Disponível em: <<http://dx.doi.org/10.1590/1981.81222017000100010>>.

OLIVEIRA, C. I. de et al. Hybrid models for classifying histological images: An association of deep features by transfer learning with ensemble classifier. **Multimedia Tools and Applications**, Springer Science and Business Media LLC, v. 83, n. 8, p. 21929–21952, ago. 2023. ISSN 1573-7721. Disponível em: <<http://dx.doi.org/10.1007/s11042-023-16351-4>>.

PATEL, B.; YADAV, k.; GHOSH, D. State-of-art: Similarity assessment for content based image retrieval system. In: **2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)**. [S.l.: s.n.], 2020. p. 1–6.

PEDROSA, G. V. et al. **Similaridade de formas via identificação e caracterização de saliências**. Dissertação (Mestrado) — Programa de Pós-graduação em Ciência da Computação, Universidade Federal de Uberlândia, 2011. Disponível em: <<https://repositorio.ufu.br/bitstream/123456789/12514/1/d.pdf>>.

ROCHA, B. **O uso da Divergência de Kullback-Leibler e da Divergência Generalizada como medida de similaridade em sistemas CBIR**. Dissertação (Mestrado) — Editora da Universidade Federal de Uberlândia, 2015. Disponível em: <<https://doi.org/10.14393/ufu.di.2015.475>>.

ROCHA, B. M. et al. I-generalized and kullback-leibler divergences for content-based image retrieval. In: **34th IEEE Canadian Conference on Electrical and Computer Engineering, (CCECE)**. virtual event: IEEE, 2021. p. 1–4. Disponível em: <<https://doi.org/10.1109/CCECE53047.2021.9569175>>.

SAKTEL, A. W. B. P. Survey on cancer detection using content based image retrieval technique. **International Journal of Science and Research (IJSR) (Vol.3, No. 11)**, Volume 3 Issue 11, p. 1019–1020, nov 2014.

SANTINI, S.; JAIN, R. C. Similarity measures. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 21, n. 9, p. 871–883, 1999. Disponível em: <<https://doi.org/10.1109/34.790428>>.

SANTOS, E. S. et al. Avaliando a eficácia de redes neurais artificiais para reconhecimento de face utilizando raspberry pi. In: **Anais da XXI Escola Regional de Computação Bahia, Alagoas e Sergipe (ERBASE 2021)**. Sociedade Brasileira de Computação - SBC, 2021. (ERBASE 2021). Disponível em: <<http://dx.doi.org/10.5753/erbase.2021.20366>>.

SANTOS, M. d. O. et al. Estimativa de incidência de câncer no brasil, 2023-2025. **Rev. Bras. Cancerol.**, INCA, v. 69, n. 1, p. 1–12, feb 2023.

SHAHID, R. et al. Comparison of distance measures in spatial analytical modeling for health service planning. **BMC Health Services Research**, Springer Science and Business Media LLC, v. 9, n. 1, nov. 2009. ISSN 1472-6963. Disponível em: <<http://dx.doi.org/10.1186/1472-6963-9-200>>.

SHIRANI, S. et al. Epithelial dysplasia in oral cavity. **Iran. J. Med. Sci.**, v. 39, n. 5, p. 406–417, set. 2014.

- SILVA, A. **Métodos Computacionais para Análise e Classificação de Displasias em Imagens da Cavidade Bucal**. Dissertação (Mestrado) — Editora da Universidade Federal de Uberlândia, 2019. Disponível em: <<http://dx.doi.org/10.14393/ufu.di.2019.2390>>.
- SILVA, A. et al. Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. In: **SIBGRAPI**. IEEE, 2022. p. 264–269. Disponível em: <<http://dx.doi.org/10.1109/SIBGRAPI55357.2022.9991758>>.
- SILVA, A. B. et al. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. **Expert Systems with Applications**, Elsevier, v. 193, p. 116456, 2022. Disponível em: <<https://doi.org/10.1016/j.eswa.2021.116456>>.
- _____. OralEpitheliumDB: A dataset for oral epithelial dysplasia image segmentation and classification. **Journal of Imaging Informatics in Medicine**, Springer, 2024. pages 1-20.
- _____. Segmentation of oral epithelial dysplasias employing mask R-CNN and color normalization. In: **IEEE International Conference on Bioinformatics and Biomedicine, (BIBM)**. virtual event, South Korea: IEEE, 2020. p. 2818–2824. Disponível em: <<https://doi.org/10.1109/BIBM49941.2020.9313101>>.
- SILVA, A. T. da; MAGALHÃES, L. P.; FALCÃO, A. X. Uma nova abordagem chir baseada em realimentação de relevância e classificação por opf. 2010. Disponível em: <https://www.dca.fee.unicamp.br/portugues/pesquisa/seminarios/2010/artigos/silva_magalhaes_falcao.pdf>.
- SKOPAL, T. On fast non-metric similarity search by metric access methods. In: **International Conference on Extending Database Technology (EDBT)**. Munich, Germany: Springer, 2006, (LNCS, v. 3896). p. 718–736. Disponível em: <https://doi.org/10.1007/11687238_43>.
- SOARES, T. R. M. et al. Divergências de bregman aplicadas na recuperação por conteúdo de imagens de displasias orais. In: **Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2024)**. Sociedade Brasileira de Computação - SBC, 2024. (SBCAS 2024). Disponível em: <<http://dx.doi.org/10.5753/sbcas.2024.2058>>.
- SOLOMON, C.; BRECKON, T. **Fundamentos de processamento digital de imagens-Uma abordagem prática com exemplos em MatLab**. [S.l.: s.n.], 2013.
- SONG, Y. et al. BrePartition: Optimized high-dimensional ik/iNN search with bregman distances. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 34, n. 3, p. 1053–1065, 2022. Disponível em: <<https://doi.org/10.1109/tkde.2020.2992594>>.
- SOUSA, R. A. d. **Processamento automático de imagens histológicas para auxílio ao diagnóstico do remodelamento cardíaco**. Tese (Doutorado) — Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA), 2021. Disponível em: <<http://dx.doi.org/10.11606/D.59.2021.tde-13072021-213138>>.
- SOUZA, M. V. R. et al. Expression of APC protein during tongue malignant transformation in galectin-3-deficient mice challenged by the carcinogen 4-nitroquiniline-n-oxide. **Int. J. Clin. Exp. Pathol.**, v. 7, n. 6, p. 3255–3263, may 2014.
- SUNG, H. et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, Wiley, v. 71, n. 3, p. 209–249, fev. 2021. ISSN 1542-4863. Disponível em: <<http://dx.doi.org/10.3322/caac.21660>>.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. Data mining introduction. **People's Posts and Telecommunications Publishing House, Beijing**, 2006.

- TANG, X.-H. et al. Oral cavity and esophageal carcinogenesis modeled in carcinogen-treated mice. **Clinical Cancer Research**, AACR, v. 10, n. 1, p. 301–313, 2004.
- TEIXEIRA, L. A. d. S. et al. **O câncer no Brasil: passado e presente**. [S.l.]: Outras Letras, 2012.
- THULER, L. C. S.; SANT'ANA, D. R.; REZENDE, M. C. R. Abc do câncer: abordagens básicas para o controle do câncer. In: **ABC do câncer: abordagens básicas para o controle do câncer**. [S.l.: s.n.], 2011. p. 127–127.
- TOMMASINO, C. et al. Histopathological image deep feature representation for cbir in smart pacs. **Journal of Digital Imaging**, Springer Science and Business Media LLC, v. 36, n. 5, p. 2194–2209, jun. 2023. ISSN 1618-727X. Disponível em: <<http://dx.doi.org/10.1007/s10278-023-00832-x>>.
- TOSTA, T. A. A. et al. Unsupervised method for normalization of hematoxylin-eosin stain in histological images. **Computerized Medical Imaging and Graphics**, Elsevier BV, v. 77, p. 101646, out. 2019. ISSN 0895-6111. Disponível em: <<http://dx.doi.org/10.1016/j.compmedimag.2019.101646>>.
- VISENTIN, A.; LENARDT, M. H. O itinerário terapêutico: história oral de idosos com câncer. **Acta paulista de enfermagem**, SciELO Brasil, v. 23, p. 486–492, 2010.
- WAAL, I. van der. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. **Oral Oncology**, Elsevier BV, v. 45, n. 4–5, p. 317–323, abr. 2009. ISSN 1368-8375. Disponível em: <<http://dx.doi.org/10.1016/j.oraloncology.2008.05.016>>.
- WARNAKULASURIYA, S. et al. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. **Journal of Oral Pathology and Medicine**, Wiley, v. 37, n. 3, p. 127–133, fev. 2008. ISSN 1600-0714. Disponível em: <<http://dx.doi.org/10.1111/j.1600-0714.2007.00584.x>>.
- YOUNG, A. M.; HOBBS, R.; KERR, D. J. **ABC of colorectal cancer**. [S.l.]: John Wiley & Sons, 2011. v. 154.
- ZHANG, L.; RUI, Y. Image search—from thousands to billions in 20 years. **ACM Transactions on Multimedia Computing, Communications, and Applications**, Association for Computing Machinery (ACM), v. 9, n. 1s, p. 1–20, out. 2013. ISSN 1551-6865. Disponível em: <<http://dx.doi.org/10.1145/2490823>>.
- ZHENG, Y. et al. Histopathological whole slide image analysis using context-based cbir. **IEEE Transactions on Medical Imaging**, Institute of Electrical and Electronics Engineers (IEEE), v. 37, n. 7, p. 1641–1652, jul. 2018. ISSN 1558-254X. Disponível em: <<http://dx.doi.org/10.1109/TMI.2018.2796130>>.

APÊNDICE

Aqui estão reunidas as imagens dos experimentos realizados no Capítulo 5. Estas imagens abrangem todos os experimentos conduzidos, utilizando as métricas de distâncias com as (50/40/30/20/10) características mais relevantes. Elas fornecem uma visualização detalhada dos resultados obtidos para cada combinação de descritores e modelos testados, permitindo uma análise comparativa das diferentes abordagens empregadas.

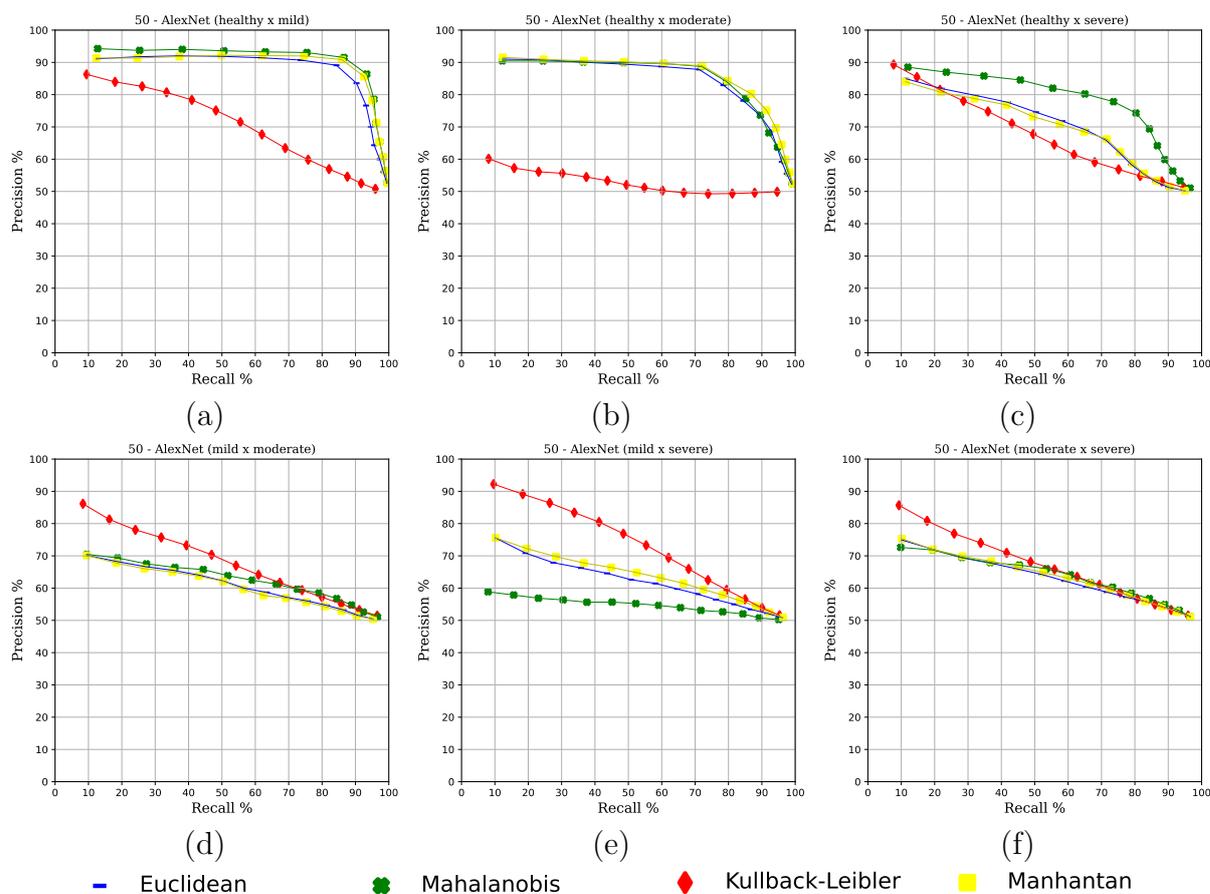


Figura 27 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - AlexNet (saudável x leve) (b) 50 - AlexNet (leve x moderada) (c) 50 - AlexNet (moderada x severa) (d) 50 - AlexNet (leve x moderada) (e) 50 - AlexNet (leve x severa) (f) 50 - AlexNet (moderada x severa)

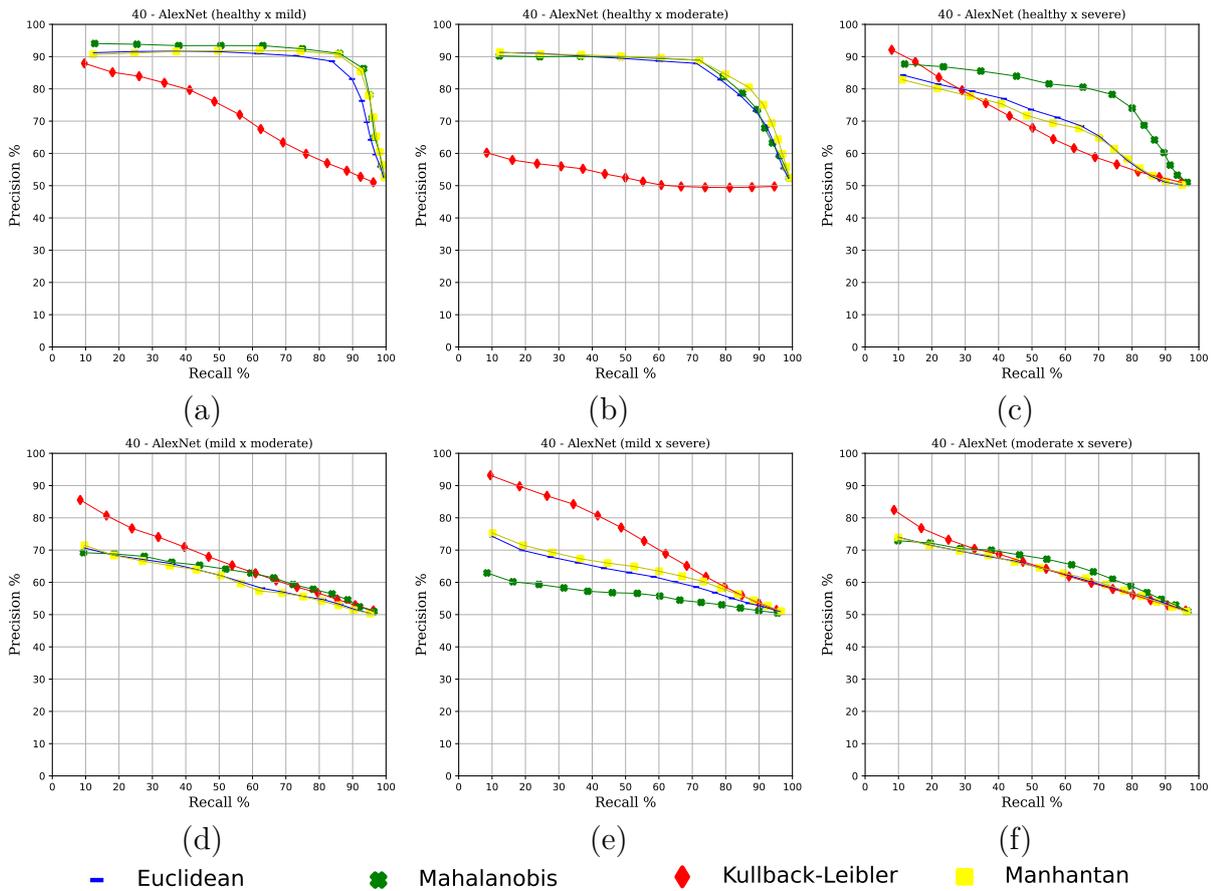


Figura 28 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - AlexNet (saúdável x leve) (b) 40 - AlexNet (leve x moderada) (c) 40 - AlexNet (moderada x severa) (d) 40 - AlexNet (leve x moderada) (e) 40 - AlexNet (leve x severa) (f) 40 - AlexNet (moderada x severa)

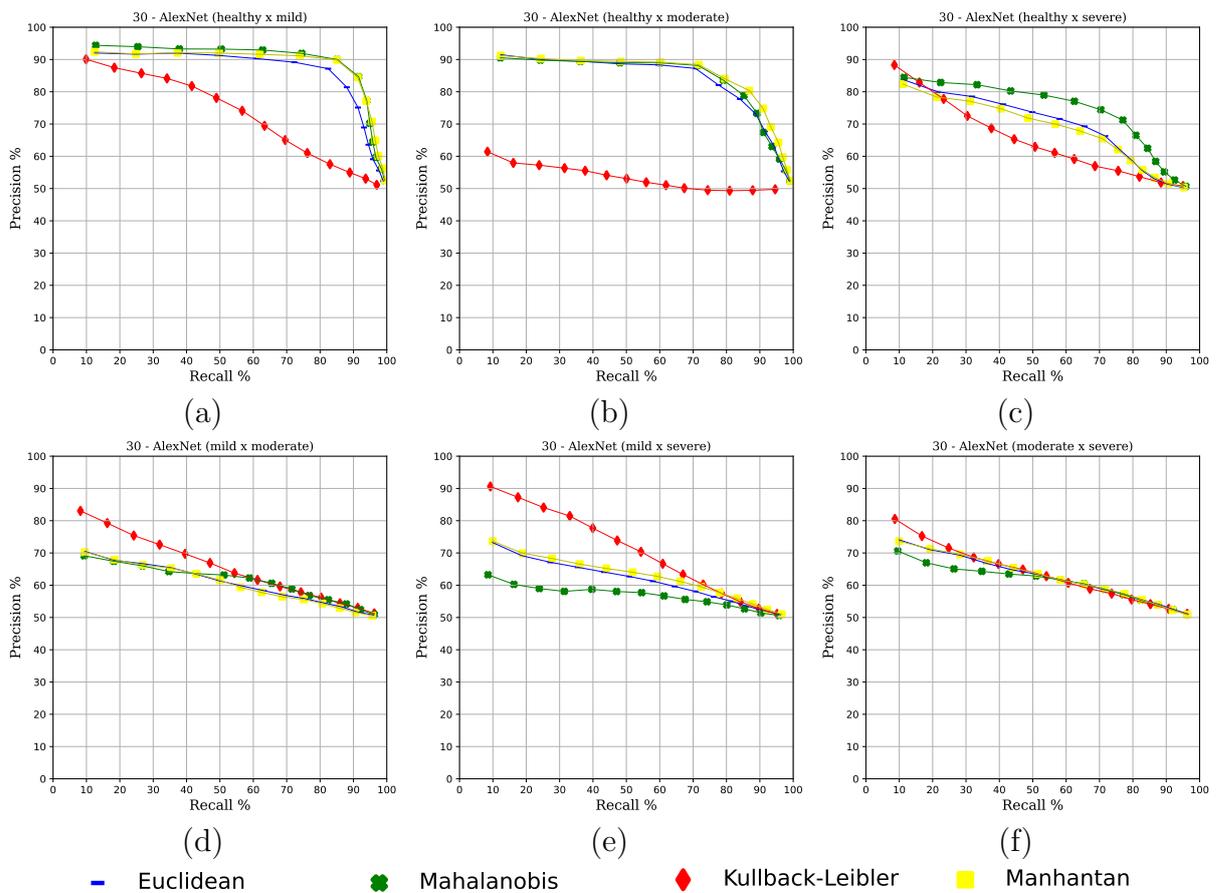


Figura 29 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - AlexNet (saudável x leve) (b) 30 - AlexNet (leve x moderada) (c) 30 - AlexNet (moderada x severa) (d) 30 - AlexNet (leve x moderada) (e) 30 - AlexNet (leve x severa) (f) 30 - AlexNet (moderada x severa)

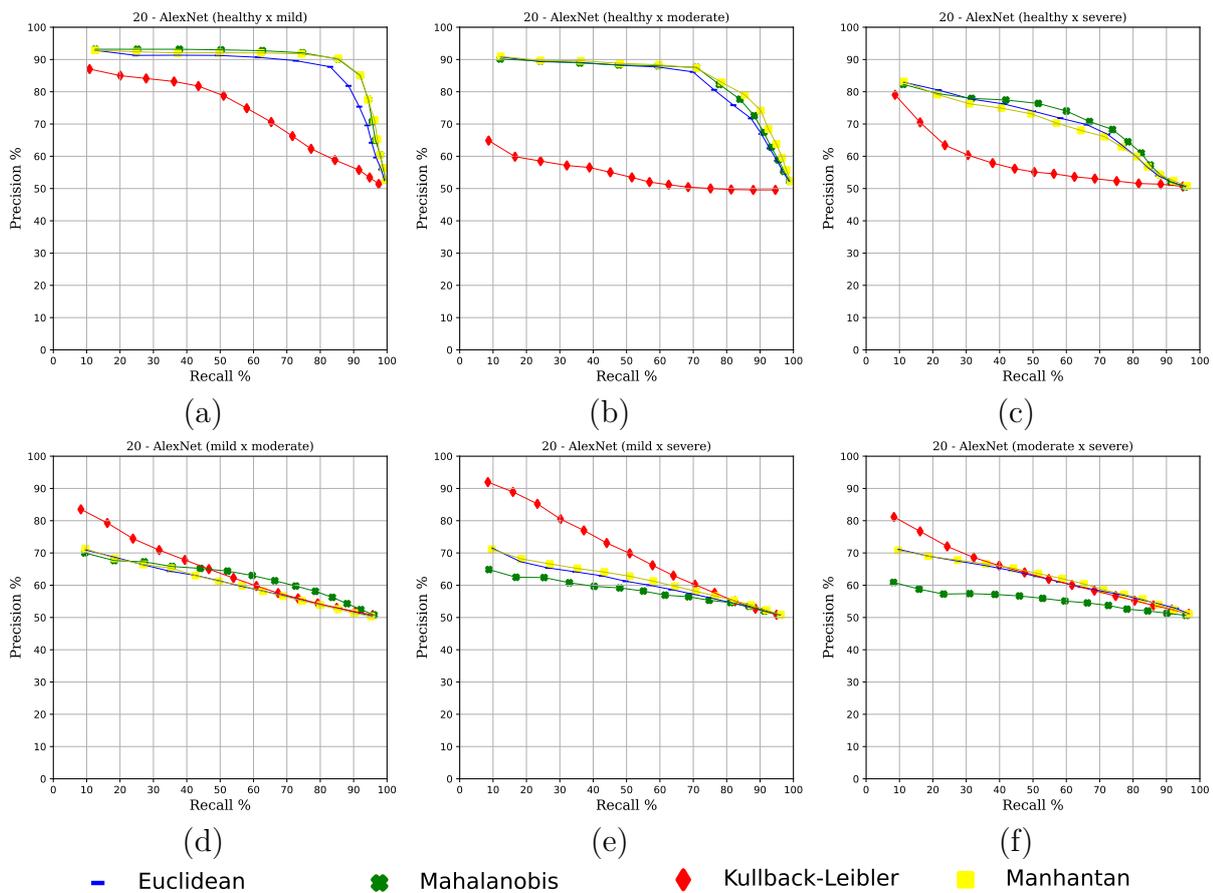


Figura 30 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - AlexNet (saudável x leve) (b) 20 - AlexNet (leve x moderada) (c) 20 - AlexNet (moderada x severa) (d) 20 - AlexNet (leve x moderada) (e) 20 - AlexNet (leve x severa) (f) 20 - AlexNet (moderada x severa)

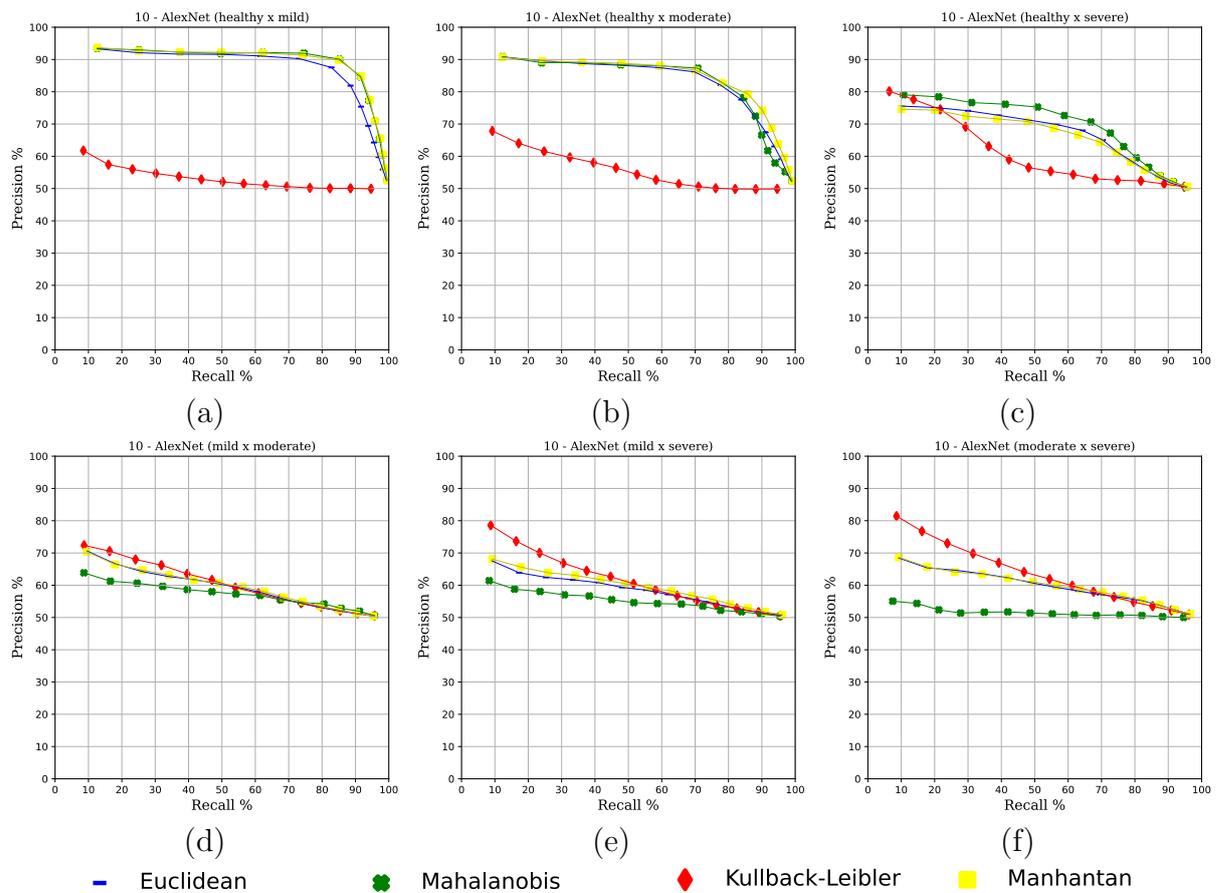


Figura 31 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - AlexNet (saudável x leve) (b) 10 - AlexNet (leve x moderada) (c) 10 - AlexNet (moderada x severa) (d) 10 - AlexNet (leve x moderada) (e) 10 - AlexNet (leve x severa) (f) 10 - AlexNet (moderada x severa)

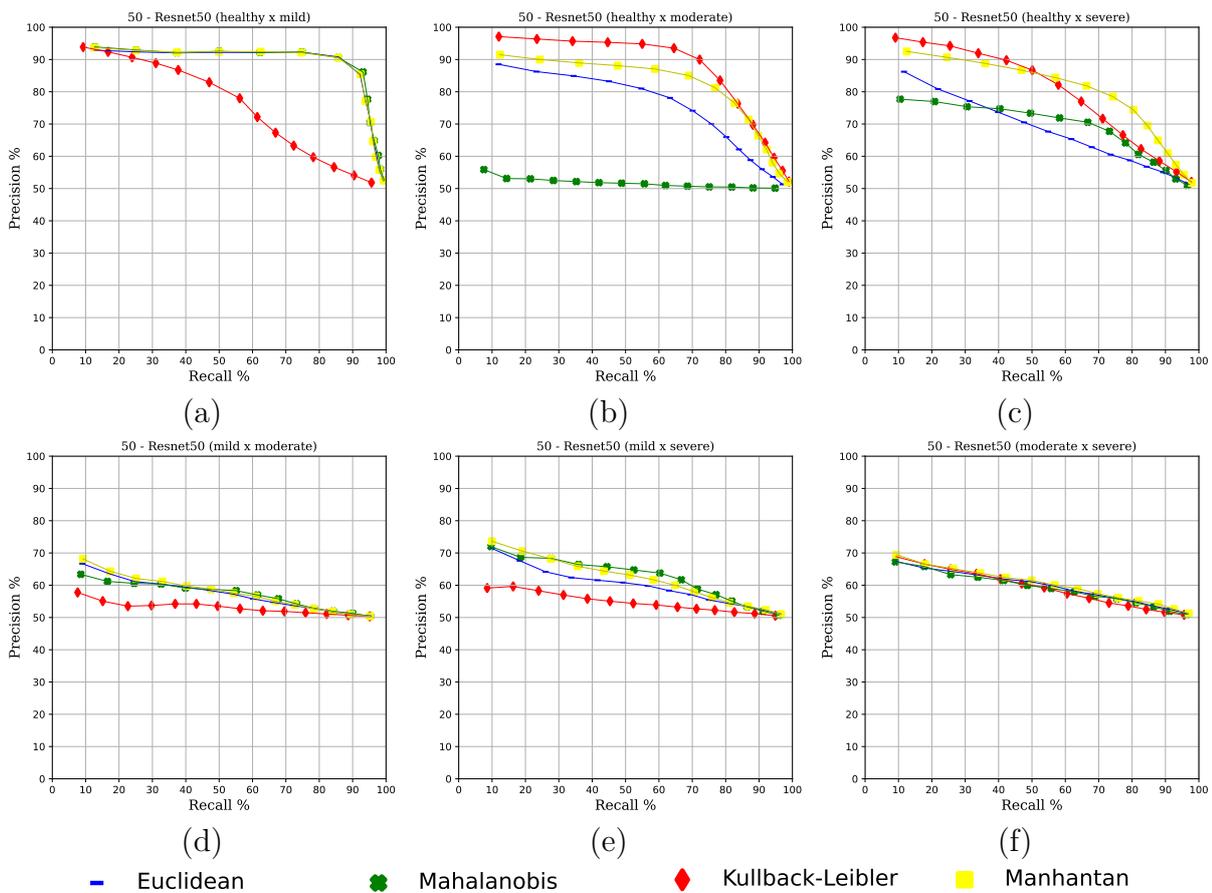


Figura 32 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - ResNet50 (saúdável x leve) (b) 50 - ResNet50 (leve x moderada) (c) 50 - ResNet50 (moderada x severa) (d) 50 - ResNet50 (leve x moderada) (e) 50 - ResNet50 (leve x severa) (f) 50 - ResNet50 (moderada x severa)

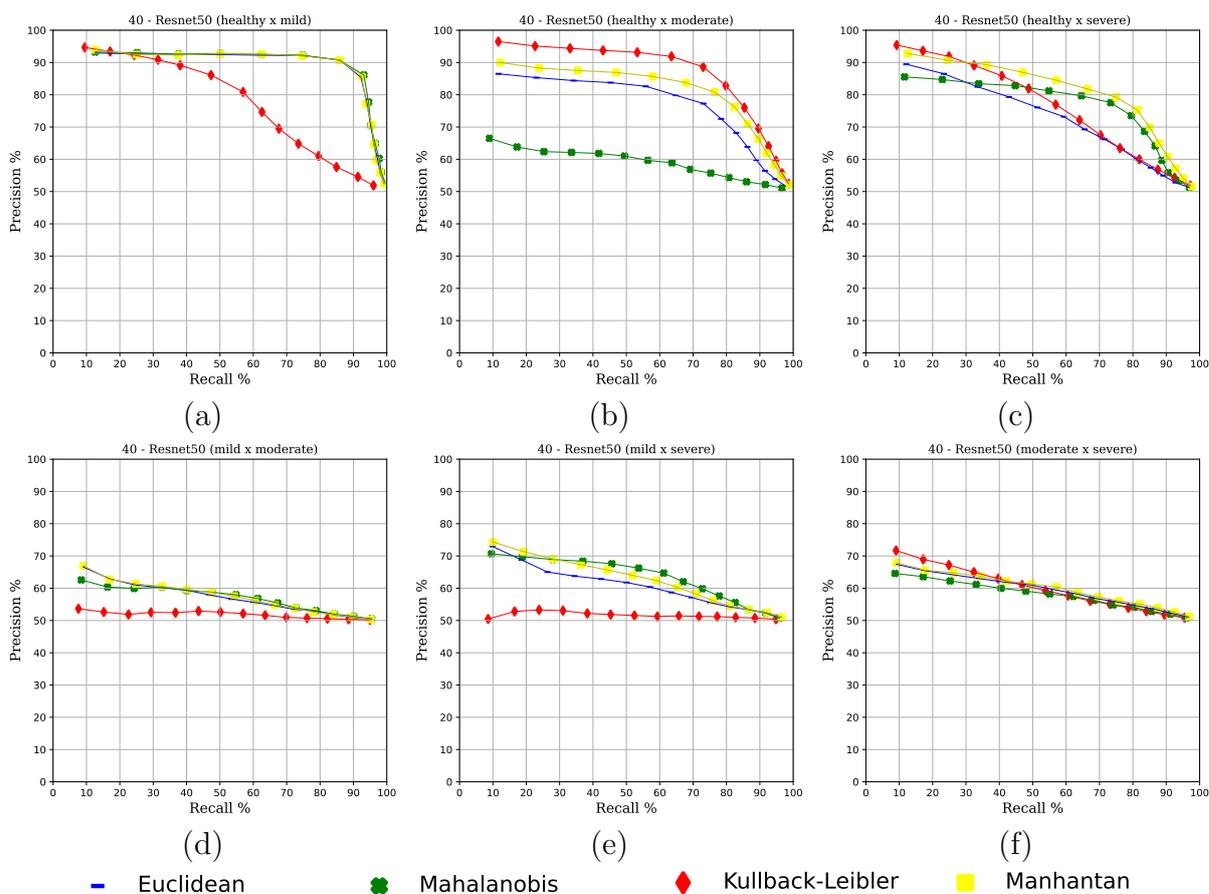


Figura 33 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - ResNet50 (saudável x leve) (b) 40 - ResNet50 (leve x moderada) (c) 40 - ResNet50 (moderada x severa) (d) 40 - ResNet50 (leve x moderada) (e) 40 - ResNet50 (leve x severa) (f) 40 - ResNet50 (moderada x severa)

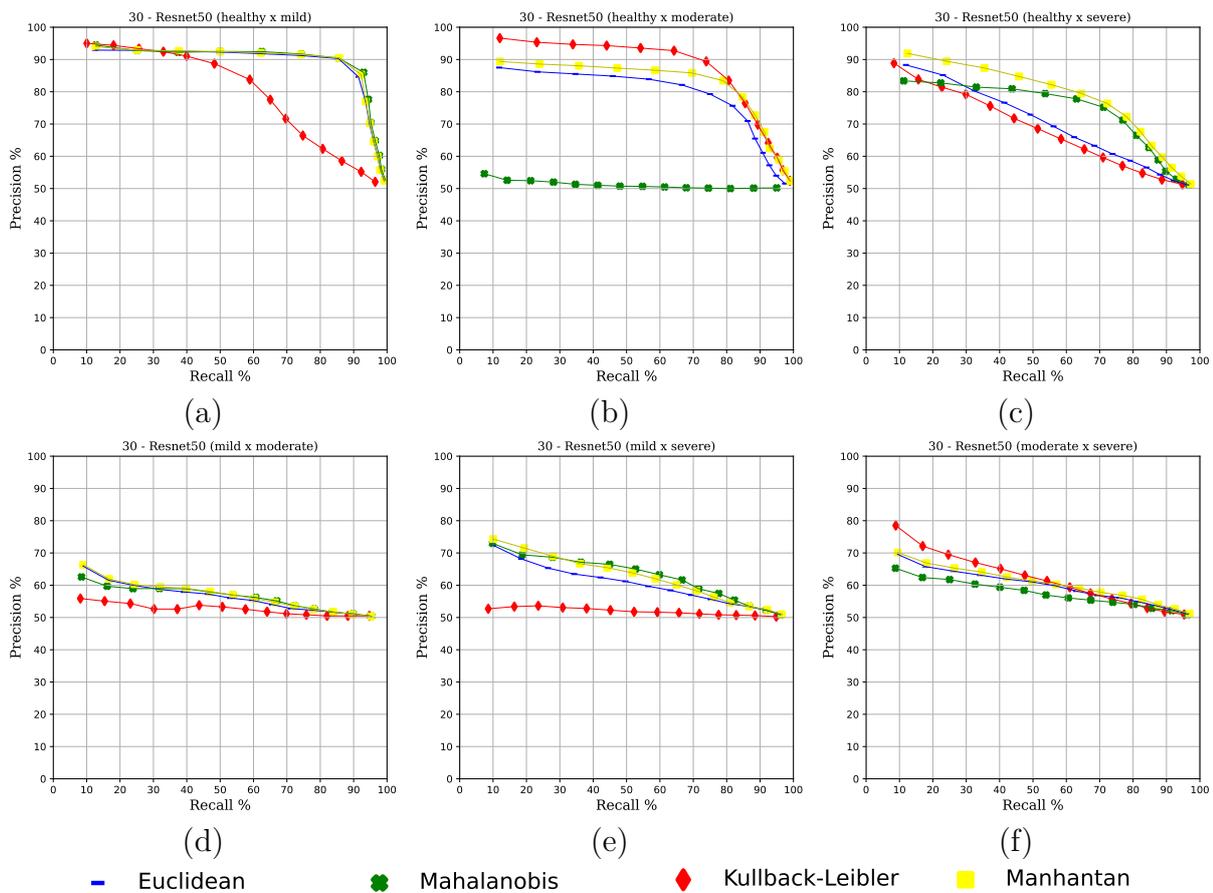


Figura 34 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - ResNet50 (saudável x leve) (b) 30 - ResNet50 (leve x moderada) (c) 30 - ResNet50 (moderada x severa) (d) 30 - ResNet50 (leve x moderada) (e) 30 - ResNet50 (leve x severa) (f) 30 - ResNet50 (moderada x severa)

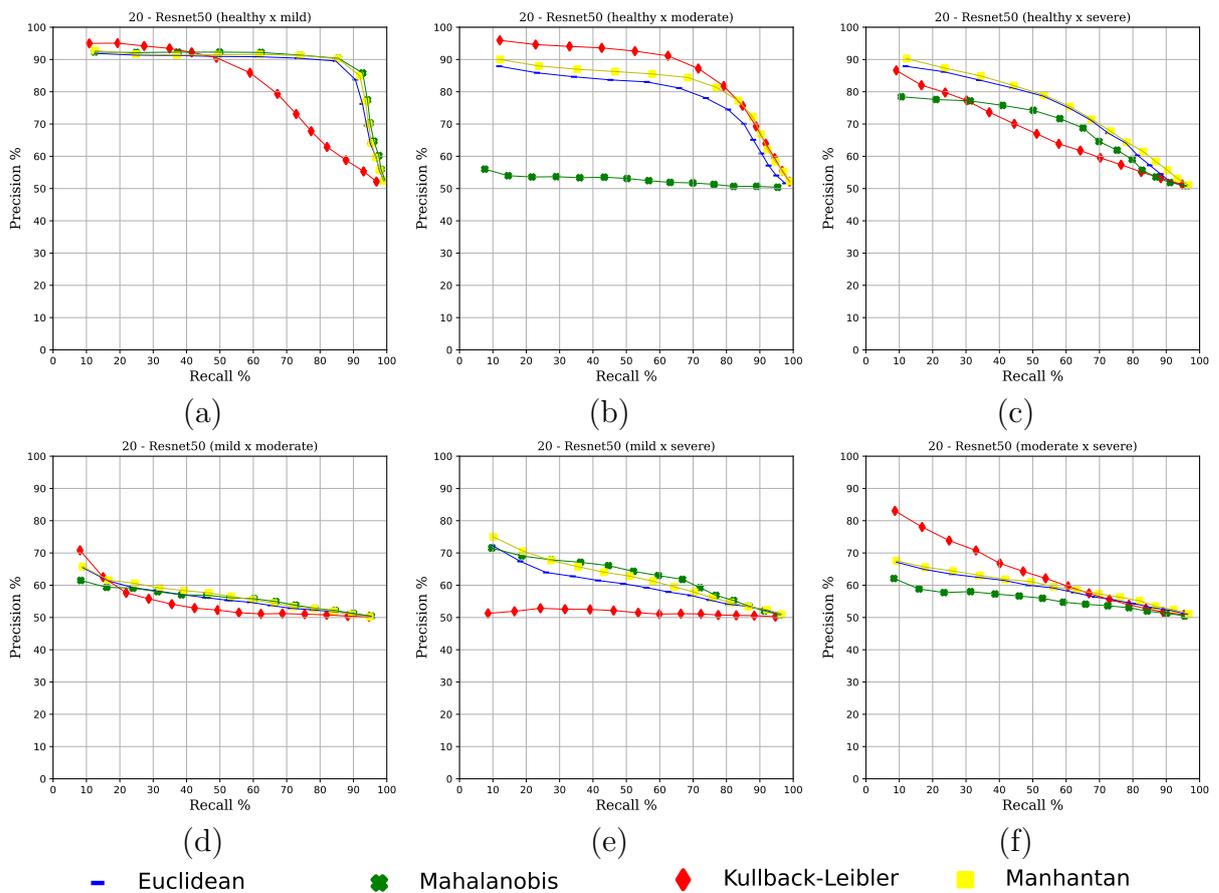


Figura 35 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - ResNet50 (saudável x leve) (b) 20 - ResNet50 (leve x moderada) (c) 20 - ResNet50 (moderada x severa) (d) 20 - ResNet50 (leve x moderada) (e) 20 - ResNet50 (leve x severa) (f) 20 - ResNet50 (moderada x severa)

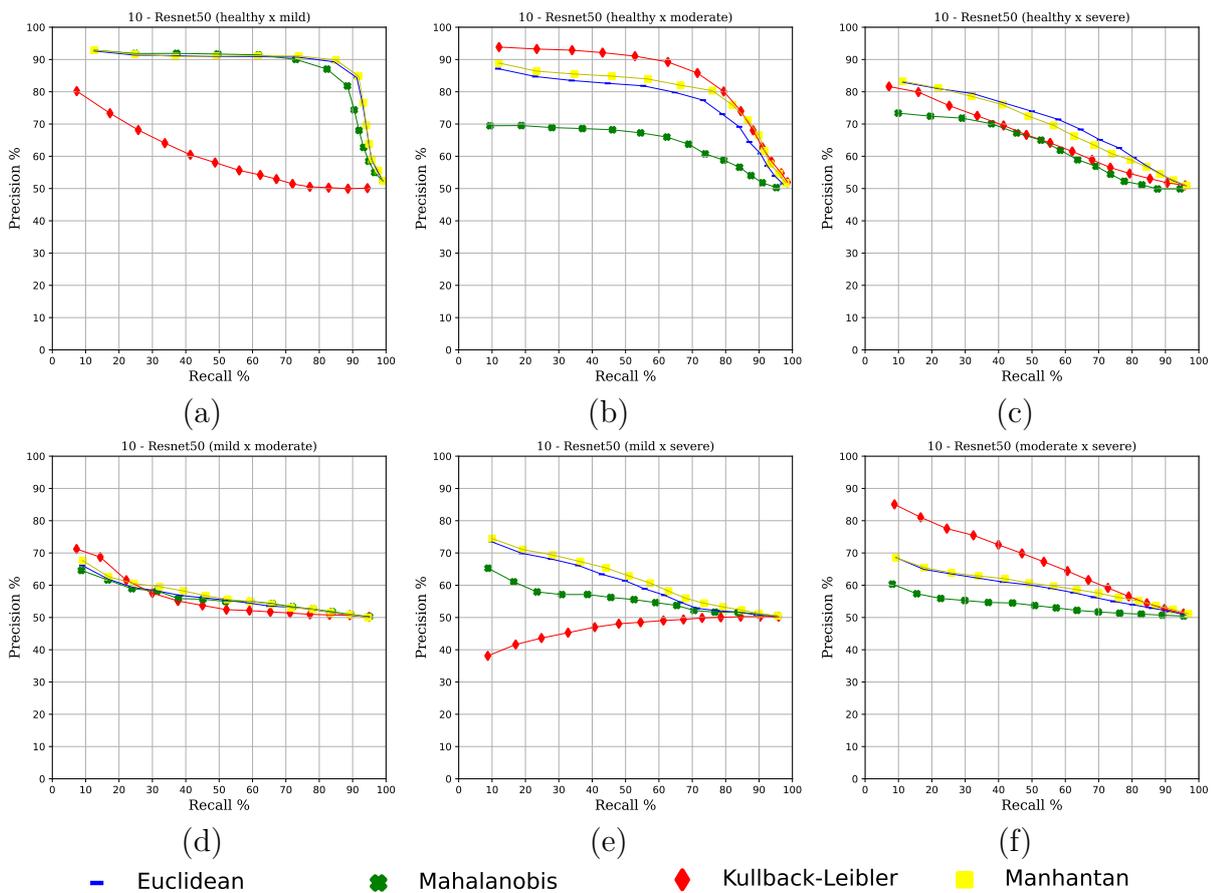


Figura 36 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - ResNet50 (saúdável x leve) (b) 10 - ResNet50 (leve x moderada) (c) 10 - ResNet50 (moderada x severa) (d) 10 - ResNet50 (leve x moderada) (e) 10 - ResNet50 (leve x severa) (f) 10 - ResNet50 (moderada x severa)

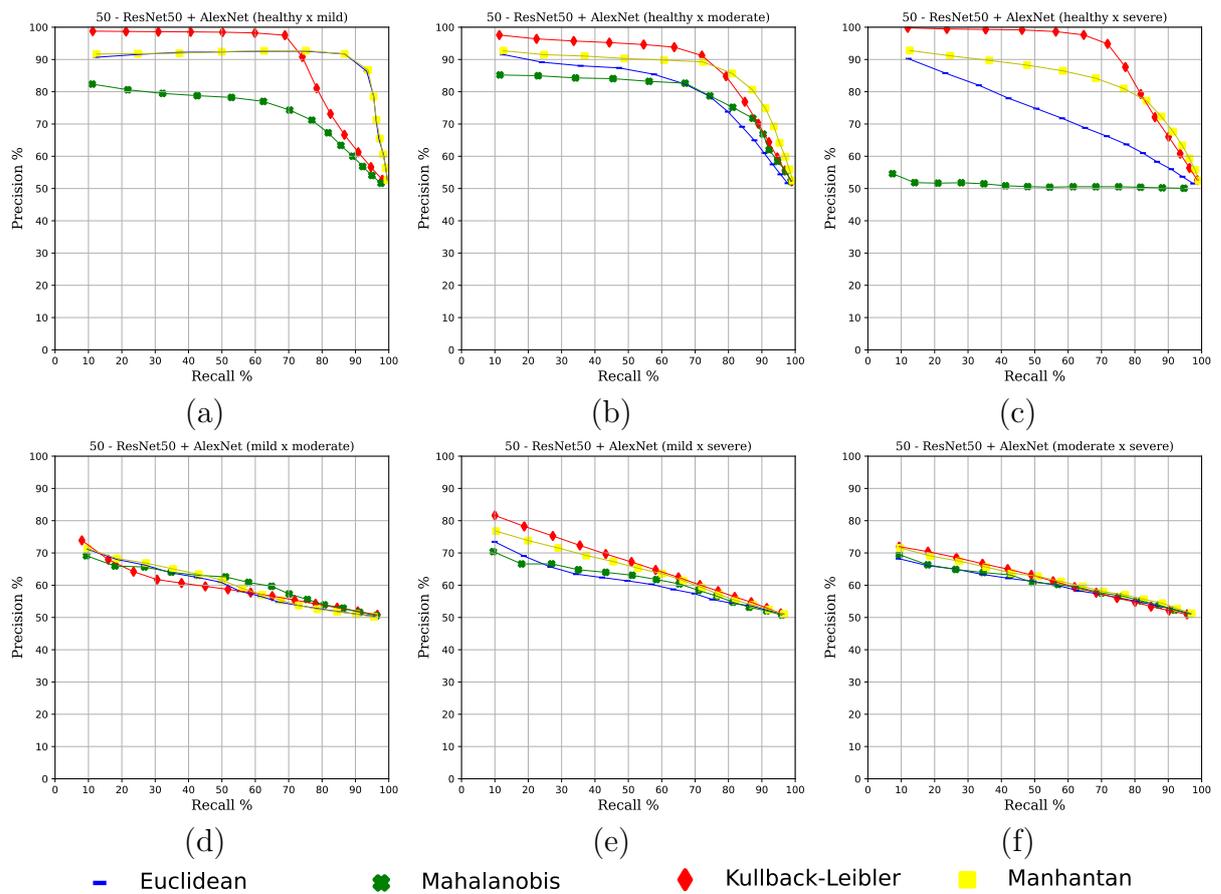


Figura 37 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - ResNet50 + AlexNet (saúdável x leve) (b) 50 - ResNet50 + AlexNet (leve x moderada) (c) 50 - ResNet50 + AlexNet (moderada x severa) (d) 50 - ResNet50 + AlexNet (leve x moderada) (e) 50 - ResNet50 + AlexNet (leve x severa) (f) 50 - ResNet50 + AlexNet (moderada x severa)

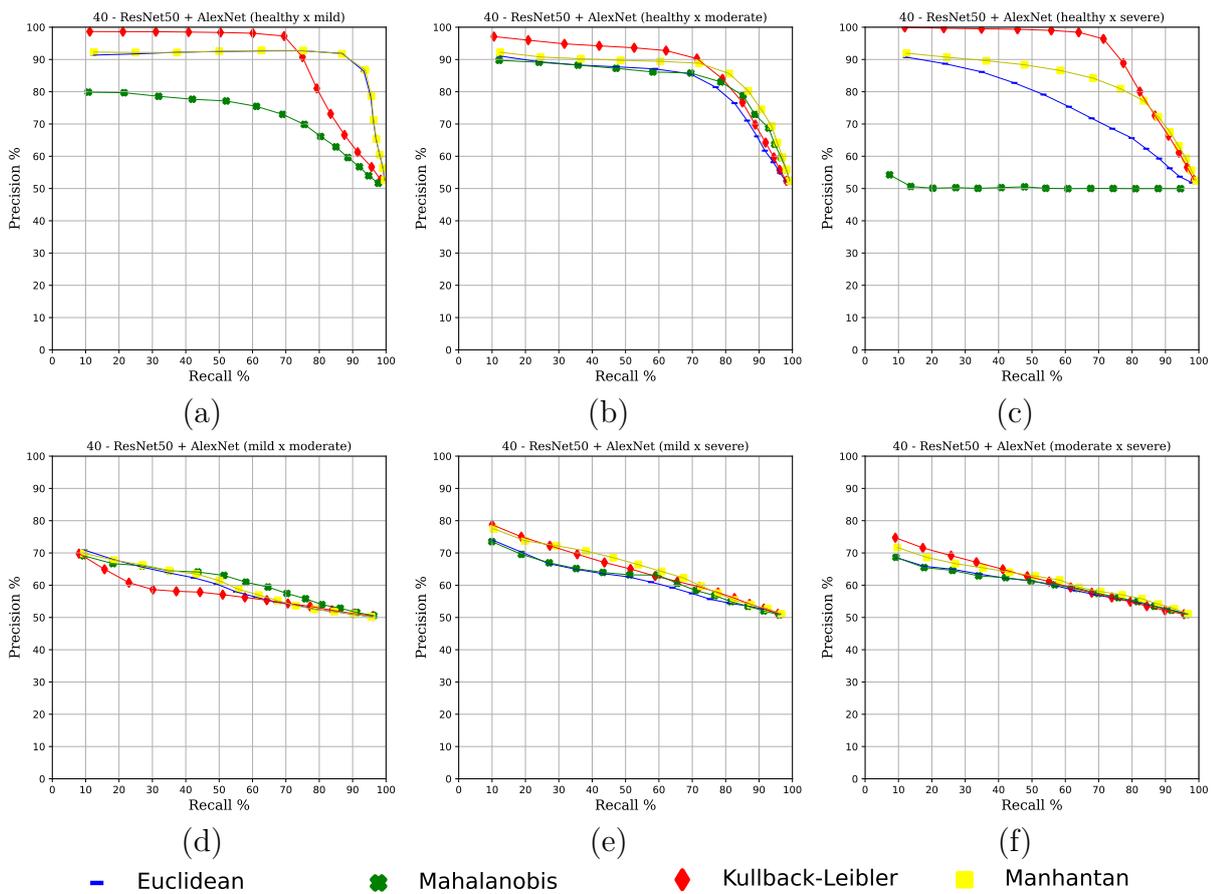


Figura 38 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - ResNet50 + AlexNet (saudável x leve) (b) 40 - ResNet50 + AlexNet (leve x moderada) (c) 40 - ResNet50 + AlexNet (moderada x severa) (d) 40 - ResNet50 + AlexNet (leve x moderada) (e) 40 - ResNet50 + AlexNet (leve x severa) (f) 40 - ResNet50 + AlexNet (moderada x severa)

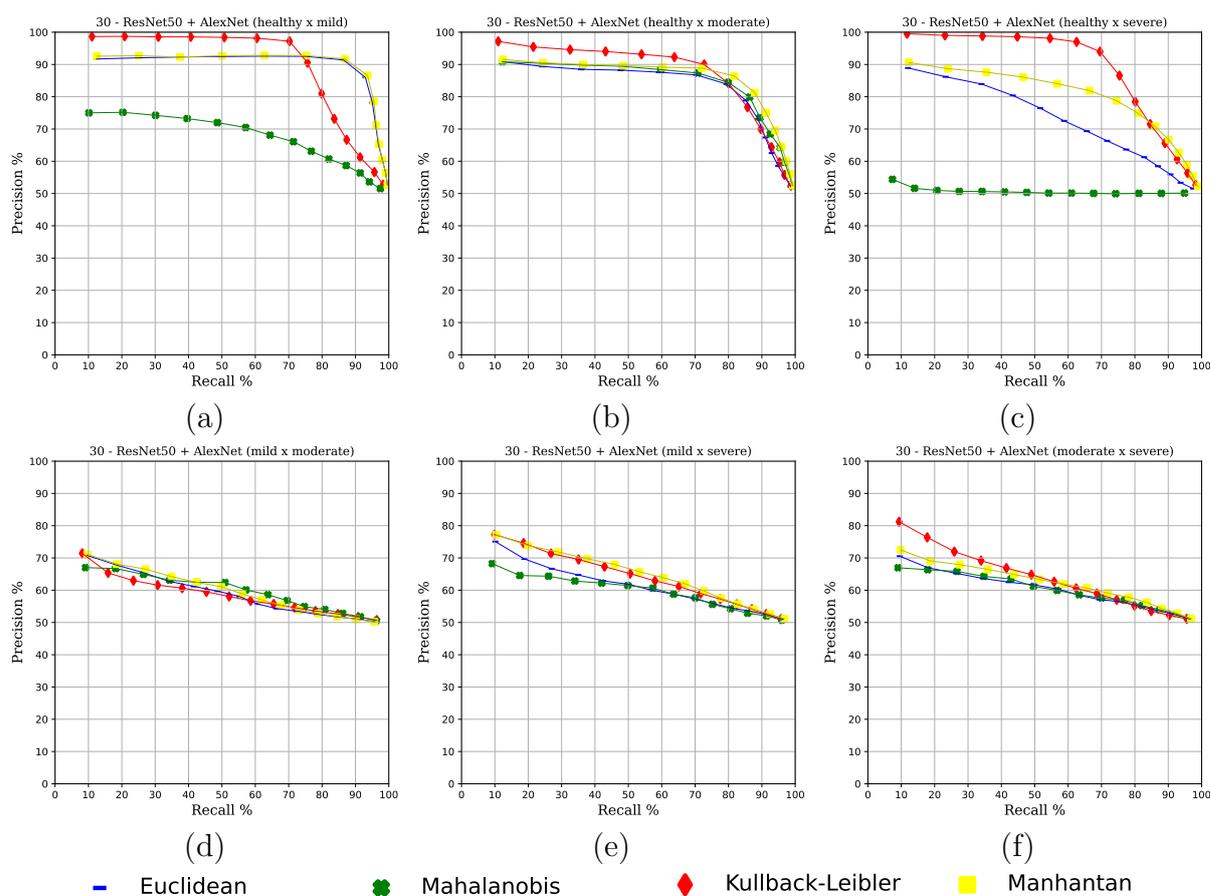


Figura 39 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - ResNet50 + AlexNet (saúdável x leve) (b) 30 - ResNet50 + AlexNet (leve x moderada) (c) 30 - ResNet50 + AlexNet (moderada x severa) (d) 30 - ResNet50 + AlexNet (leve x moderada) (e) 30 - ResNet50 + AlexNet (leve x severa) (f) 30 - ResNet50 + AlexNet (moderada x severa)

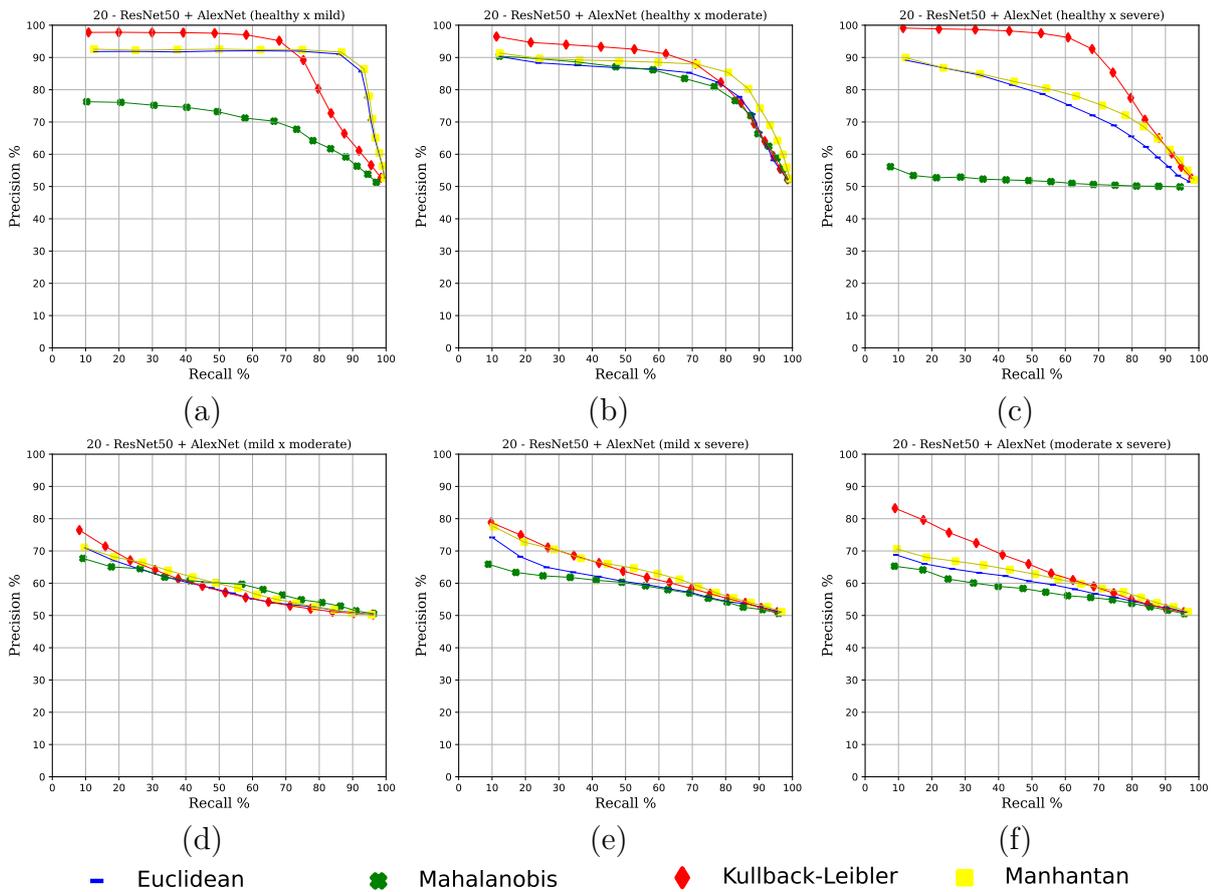


Figura 40 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - ResNet50 + AlexNet (saudável x leve) (b) 20 - ResNet50 + AlexNet (leve x moderada) (c) 20 - ResNet50 + AlexNet (moderada x severa) (d) 20 - ResNet50 + AlexNet (leve x moderada) (e) 20 - ResNet50 + AlexNet (leve x severa) (f) 20 - ResNet50 + AlexNet (moderada x severa)

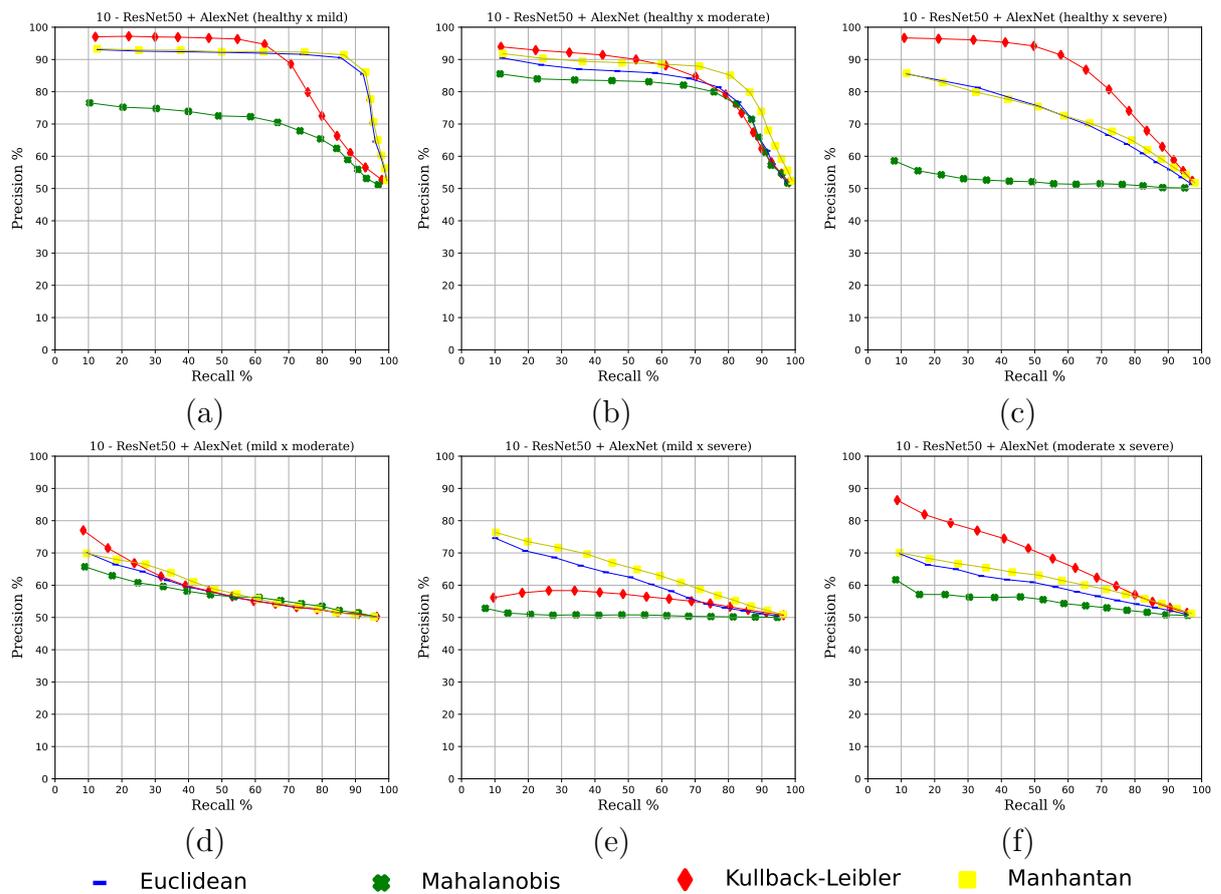


Figura 41 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - ResNet50 + AlexNet (saúdável x leve) (b) 10 - ResNet50 + AlexNet (leve x moderada) (c) 10 - ResNet50 + AlexNet (moderada x severa) (d) 10 - ResNet50 + AlexNet (leve x moderada) (e) 10 - ResNet50 + AlexNet (leve x severa) (f) 10 - ResNet50 + AlexNet (moderada x severa)

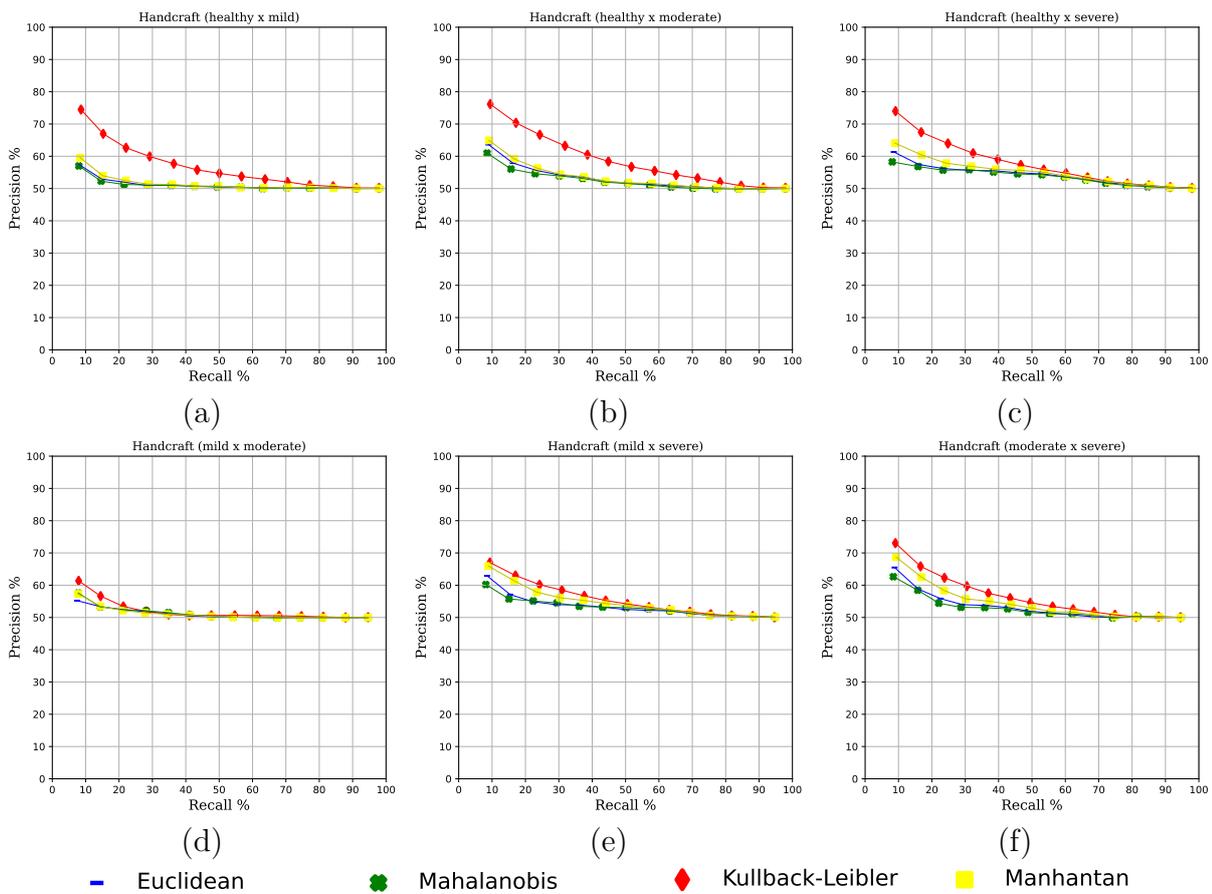


Figura 42 – Análise dos modelos em relação as métricas precisão e revogação: (a) Engenharia de Características (saúdável x leve) (b) Engenharia de Características (leve x moderada) (c) Engenharia de Características (moderada x severa) (d) Engenharia de Características (leve x moderada) (e) Engenharia de Características (leve x severa) (f) Engenharia de Características (moderada x severa)

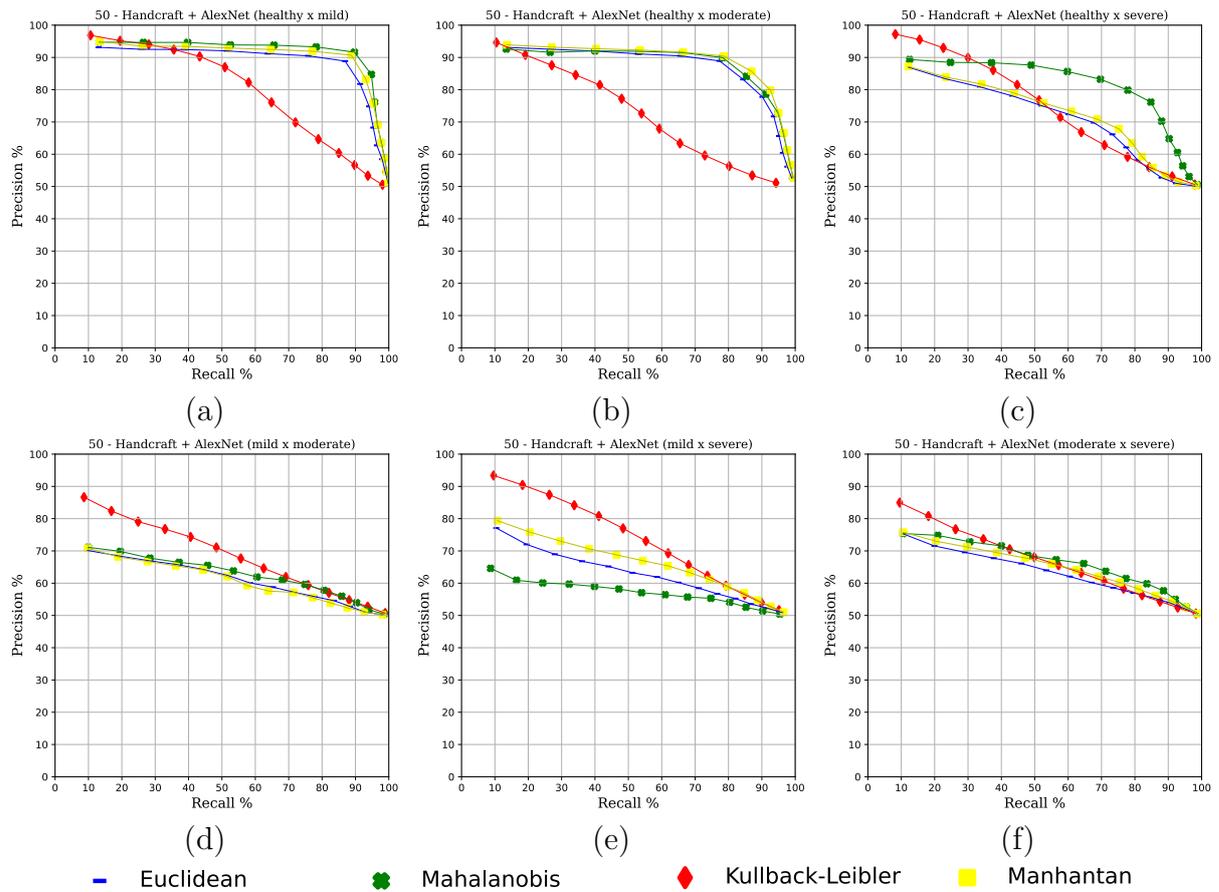


Figura 43 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + AlexNet (saudável x leve) (b) 50 - Engenharia de Características + AlexNet (leve x moderada) (c) 50 - Engenharia de Características + AlexNet (moderada x severa) (d) 50 - Engenharia de Características + AlexNet (leve x moderada) (e) 50 - Engenharia de Características + AlexNet (leve x severa) (f) 50 - Engenharia de Características + AlexNet (moderada x severa)

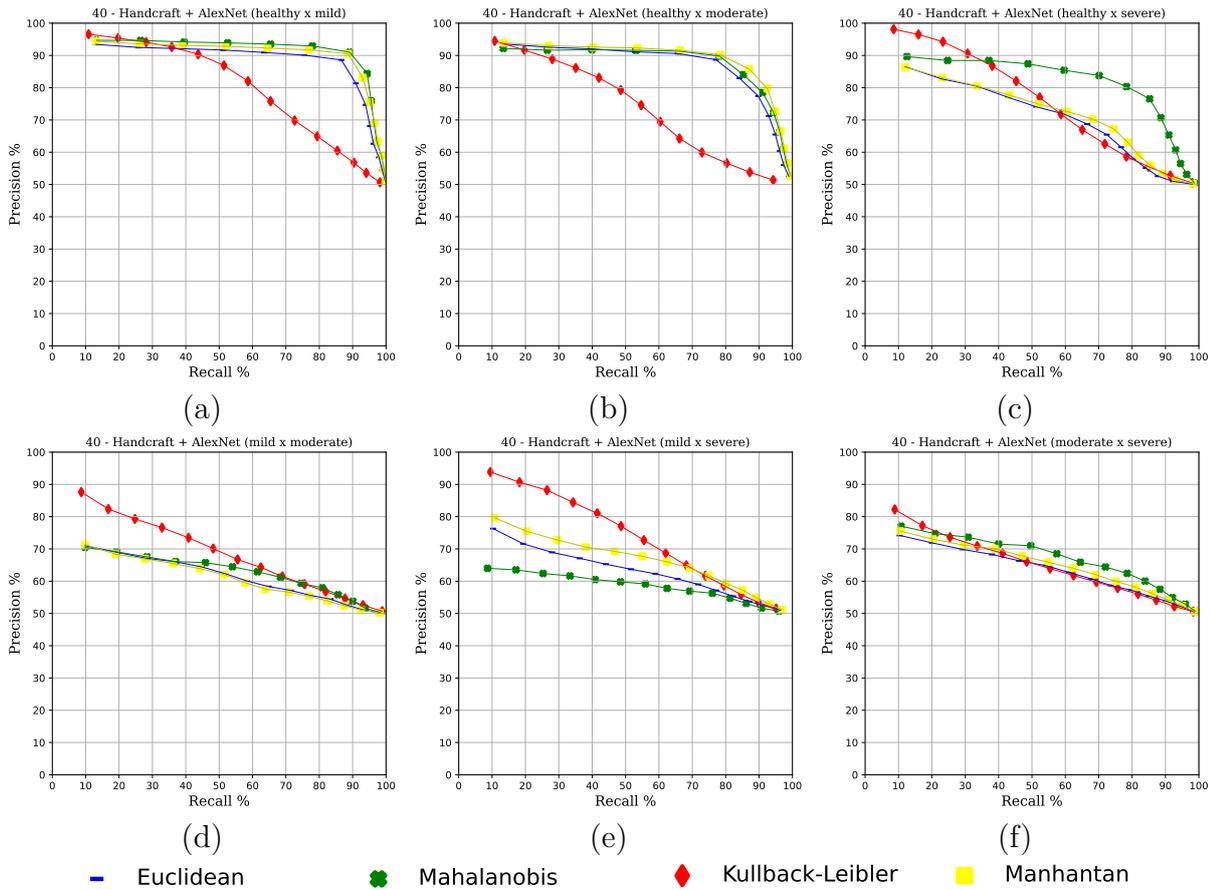


Figura 44 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - Engenharia de Características + AlexNet (saudável x leve) (b) 40 - Engenharia de Características + AlexNet (leve x moderada) (c) 40 - Engenharia de Características + AlexNet (moderada x severa) (d) 40 - Engenharia de Características + AlexNet (leve x moderada) (e) 40 - Engenharia de Características + AlexNet (leve x severa) (f) 40 - Engenharia de Características + AlexNet (moderada x severa)

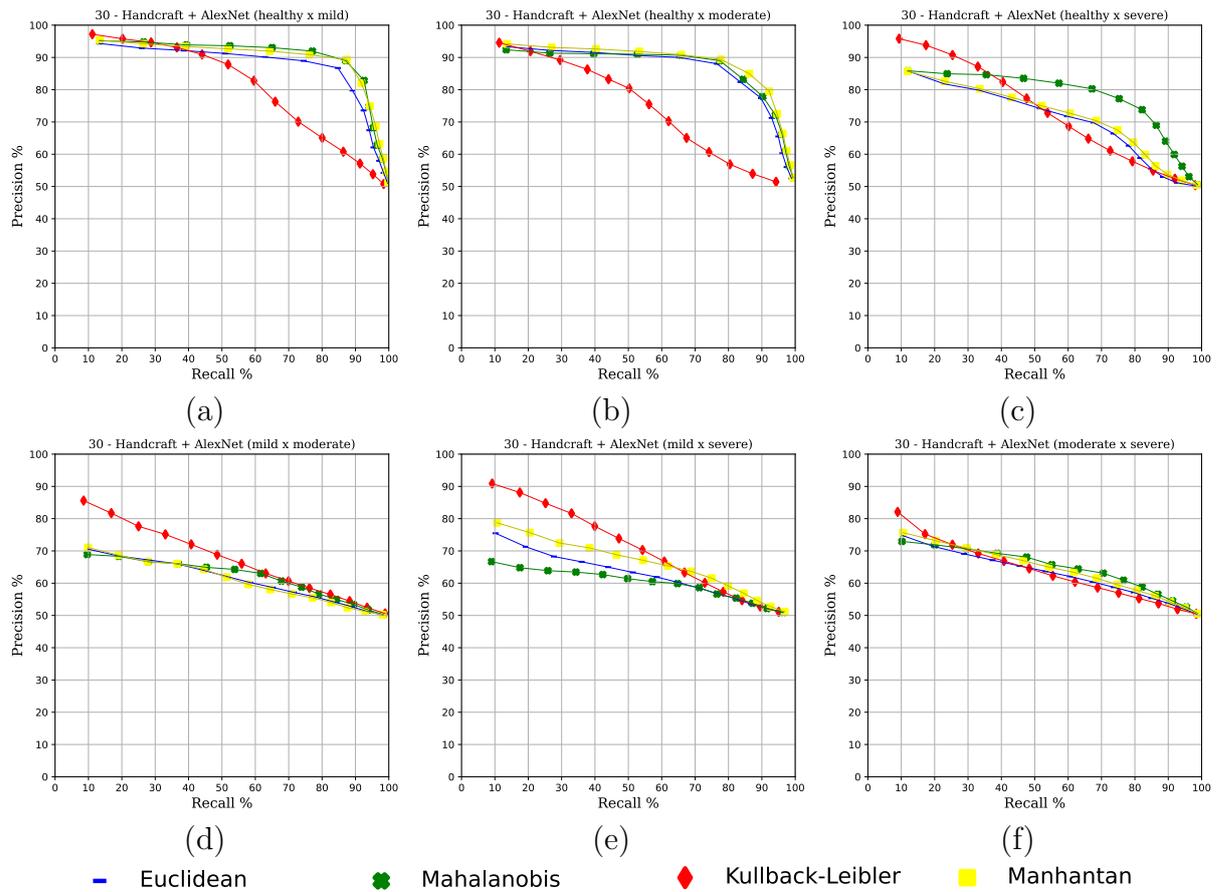


Figura 45 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - Engenharia de Características + AlexNet (saudável x leve) (b) 30 - Engenharia de Características + AlexNet (leve x moderada) (c) 30 - Engenharia de Características + AlexNet (moderada x severa) (d) 30 - Engenharia de Características + AlexNet (leve x moderada) (e) 30 - Engenharia de Características + AlexNet (leve x severa) (f) 30 - Engenharia de Características + AlexNet (moderada x severa)

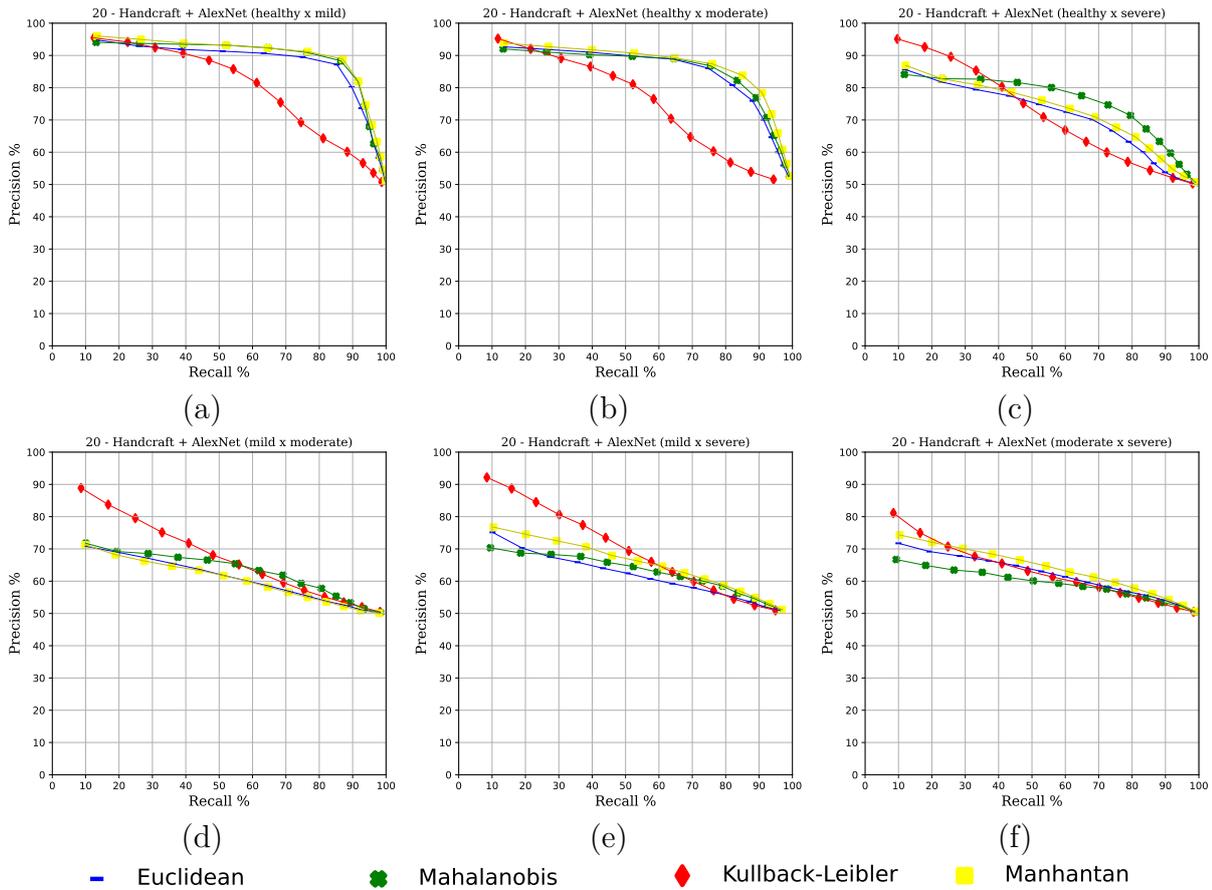


Figura 46 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - Engenharia de Características + AlexNet (saudável x leve) (b) 20 - Engenharia de Características + AlexNet (leve x moderada) (c) 20 - Engenharia de Características + AlexNet (moderada x severa) (d) 20 - Engenharia de Características + AlexNet (leve x moderada) (e) 20 - Engenharia de Características + AlexNet (leve x severa) (f) 20 - Engenharia de Características + AlexNet (moderada x severa)

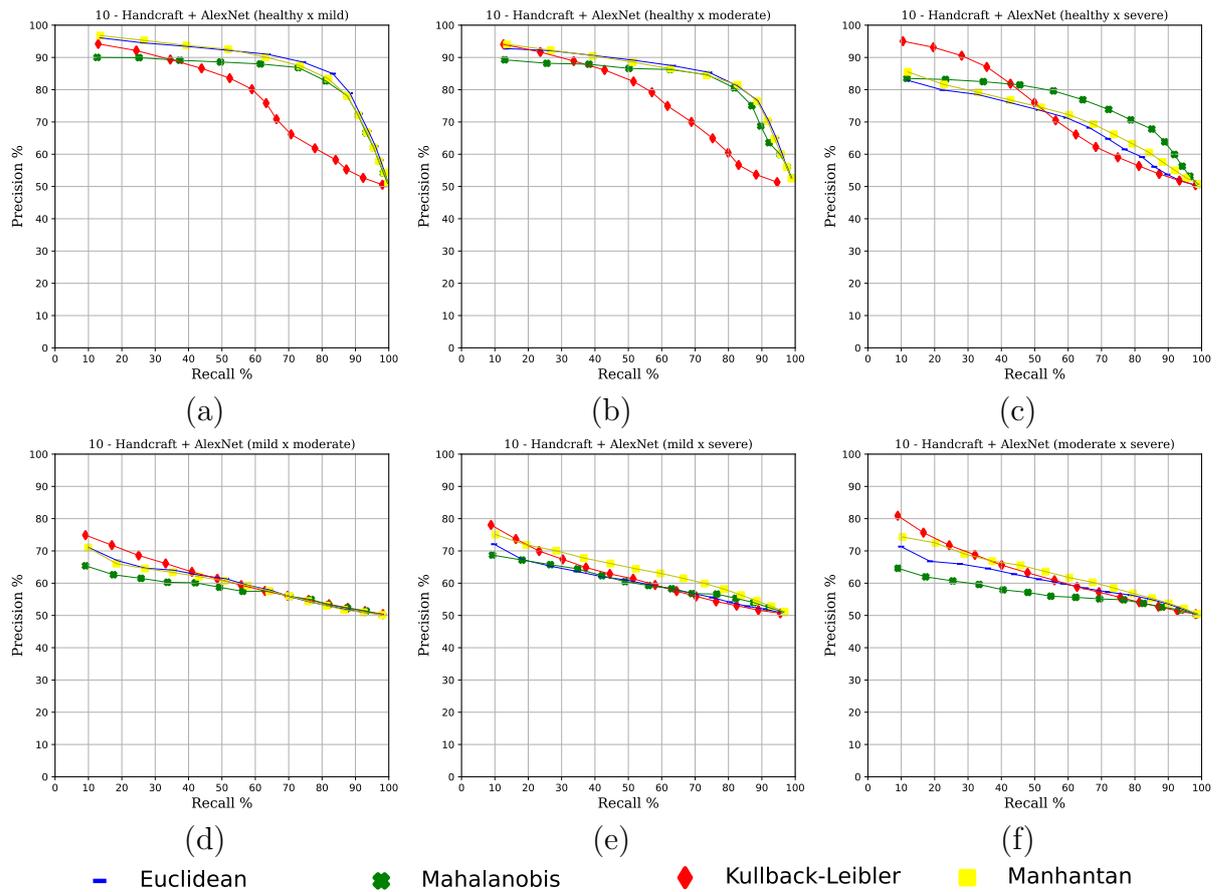


Figura 47 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - Engenharia de Características + AlexNet (saudável x leve) (b) 10 - Engenharia de Características + AlexNet (leve x moderada) (c) 10 - Engenharia de Características + AlexNet (moderada x severa) (d) 10 - Engenharia de Características + AlexNet (leve x moderada) (e) 10 - Engenharia de Características + AlexNet (leve x severa) (f) 10 - Engenharia de Características + AlexNet (moderada x severa)

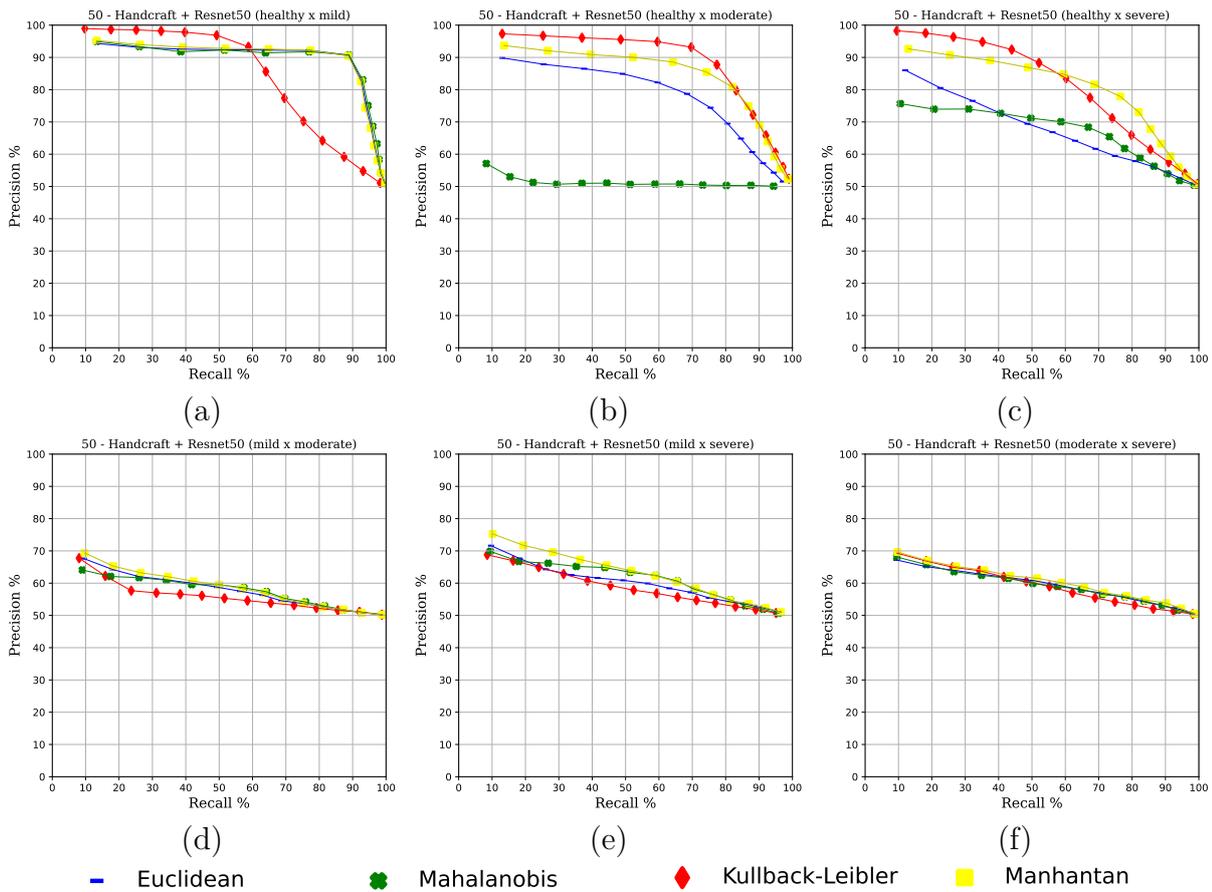


Figura 48 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 50 - Engenharia de Características + ResNet50 (leve x moderada) (c) 50 - Engenharia de Características + ResNet50 (moderada x severa) (d) 50 - Engenharia de Características + ResNet50 (leve x moderada) (e) 50 - Engenharia de Características + ResNet50 (leve x severa) (f) 50 - Engenharia de Características + ResNet50 (moderada x severa)

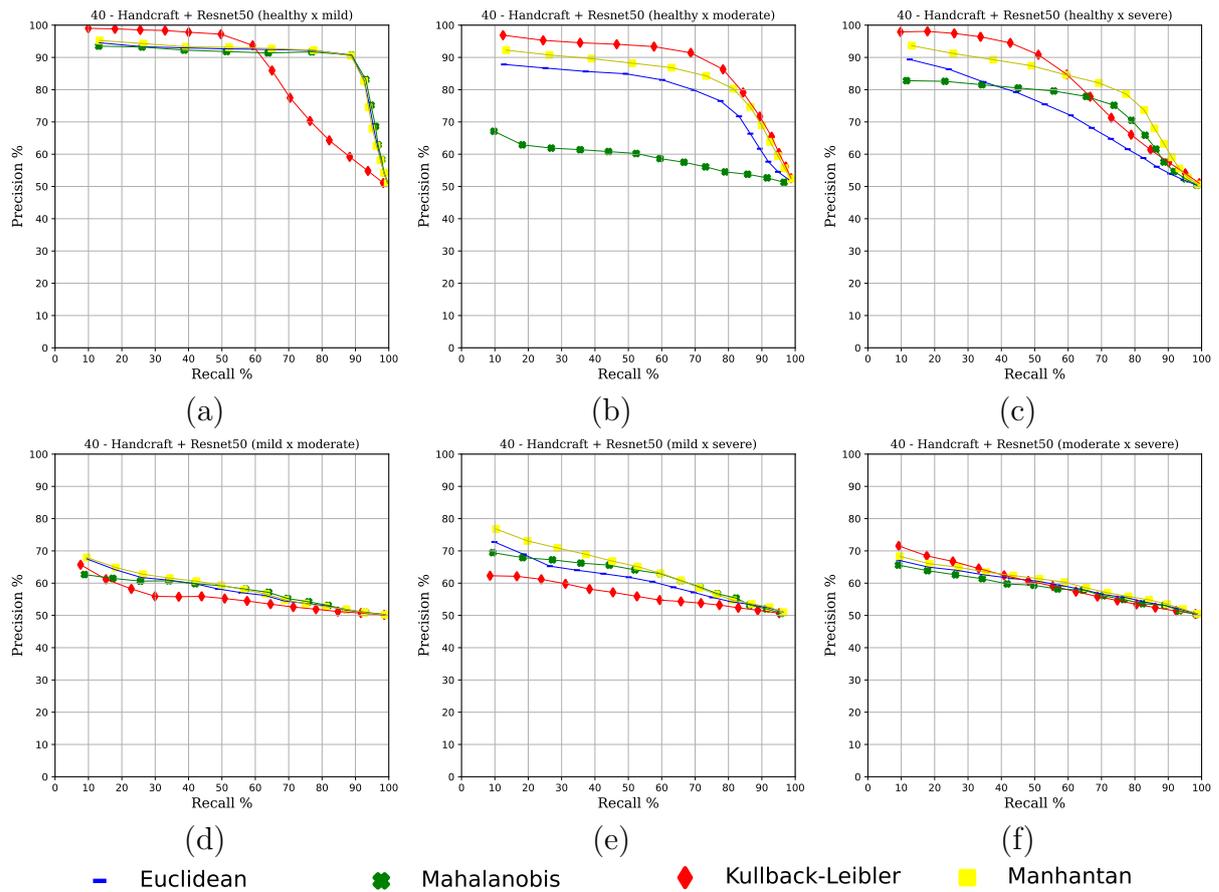


Figura 49 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 40 - Engenharia de Características + ResNet50 (leve x moderada) (c) 40 - Engenharia de Características + ResNet50 (moderada x severa) (d) 40 - Engenharia de Características + ResNet50 (leve x moderada) (e) 40 - Engenharia de Características + ResNet50 (leve x severa) (f) 40 - Engenharia de Características + ResNet50 (moderada x severa)

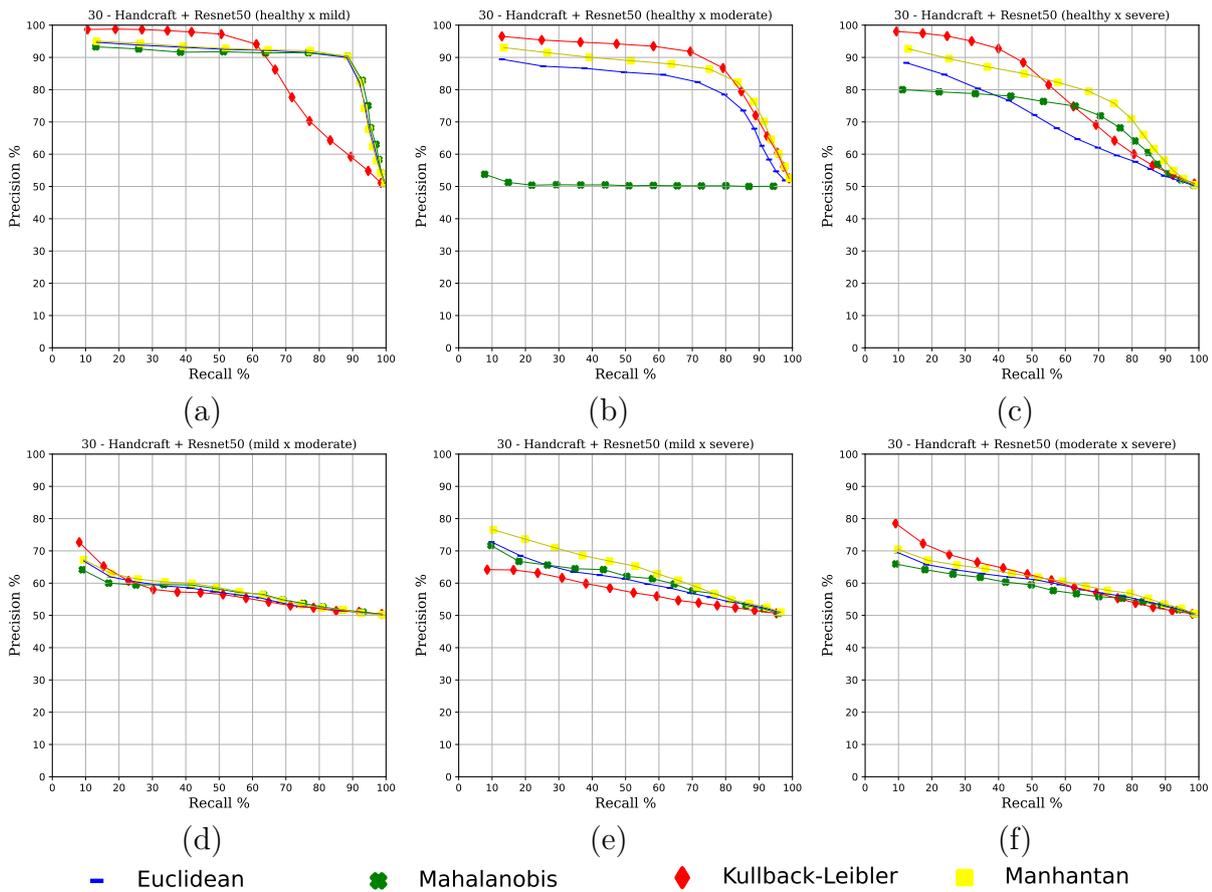


Figura 50 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 30 - Engenharia de Características + ResNet50 (leve x moderada) (c) 30 - Engenharia de Características + ResNet50 (moderada x severa) (d) 30 - Engenharia de Características + ResNet50 (leve x moderada) (e) 30 - Engenharia de Características + ResNet50 (leve x severa) (f) 30 - Engenharia de Características + ResNet50 (moderada x severa)

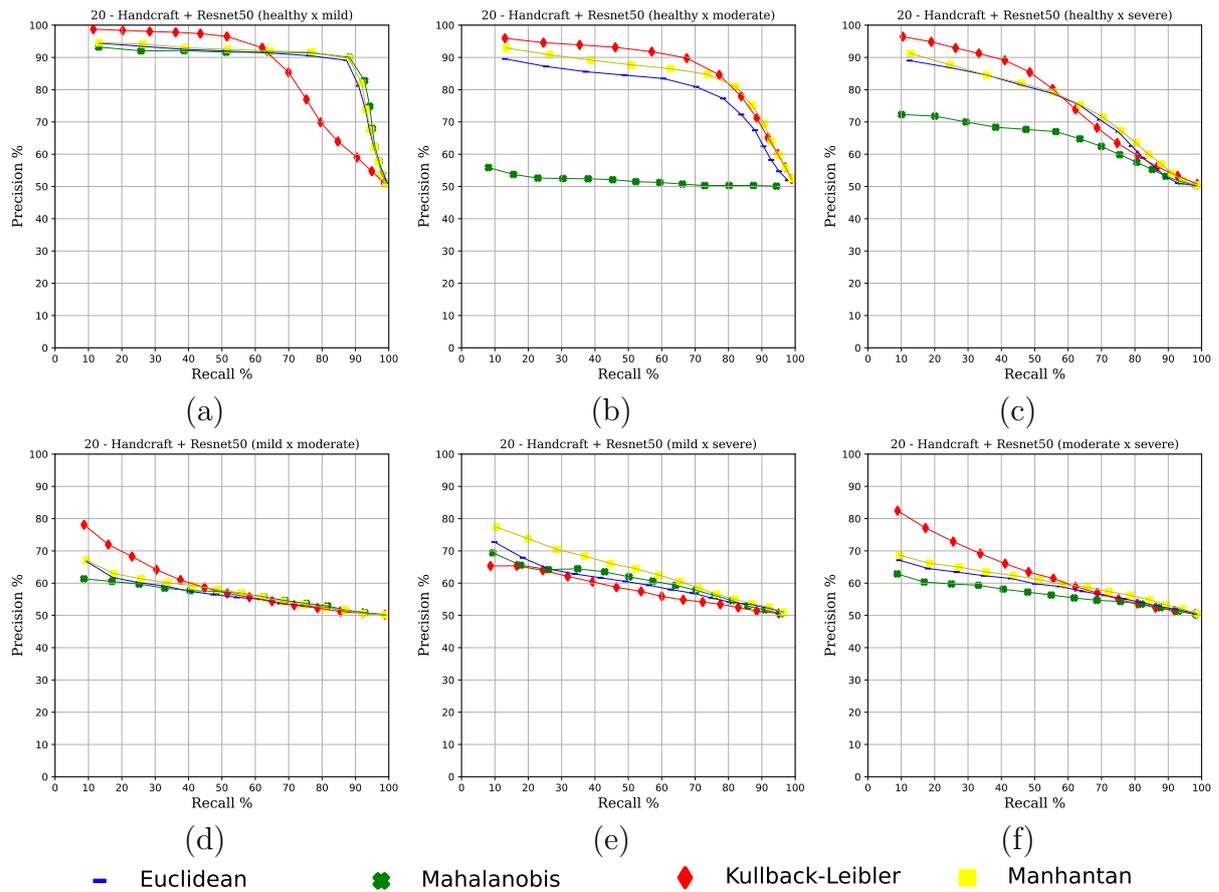


Figura 51 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 20 - Engenharia de Características + ResNet50 (leve x moderada) (c) 20 - Engenharia de Características + ResNet50 (moderada x severa) (d) 20 - Engenharia de Características + ResNet50 (leve x moderada) (e) 20 - Engenharia de Características + ResNet50 (leve x severa) (f) 20 - Engenharia de Características + ResNet50 (moderada x severa)

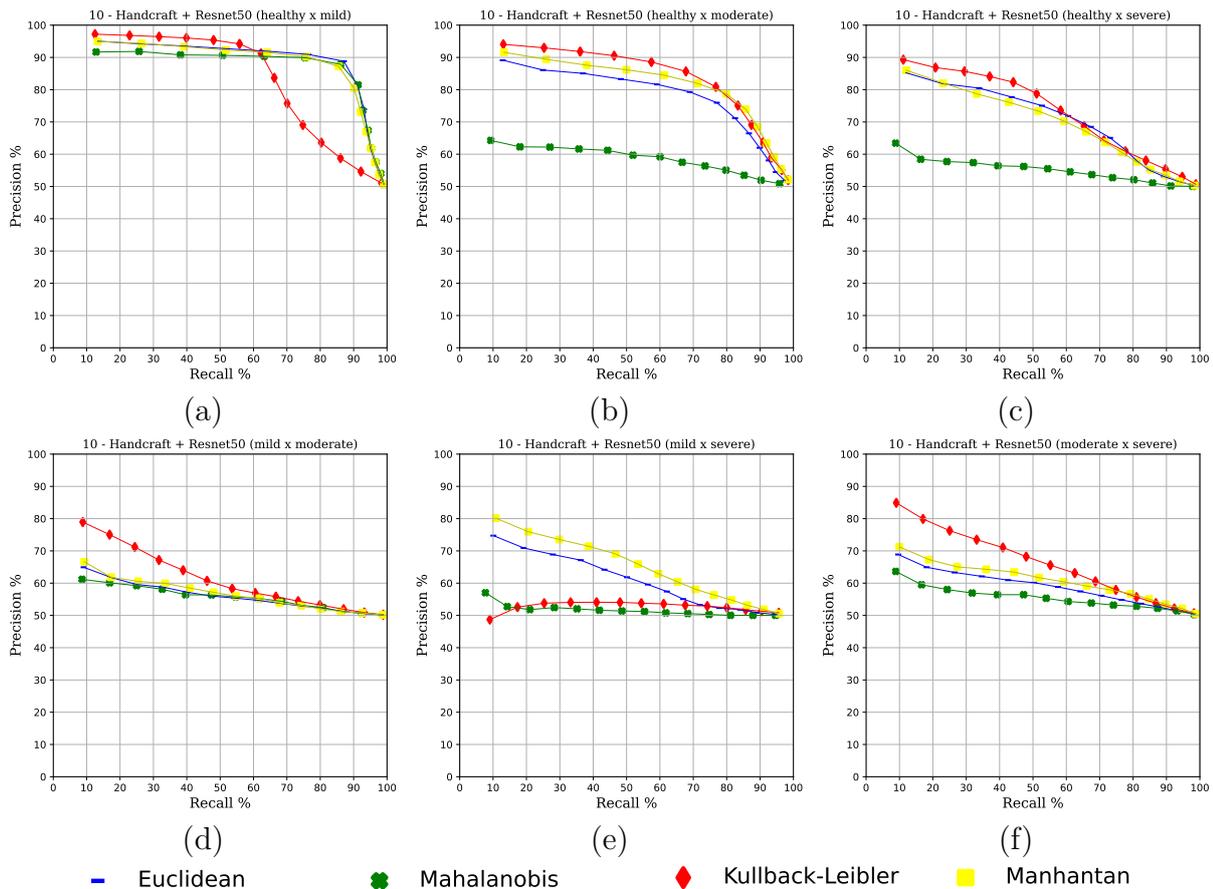


Figura 52 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - Engenharia de Características + ResNet50 (saúdável x leve) (b) 10 - Engenharia de Características + ResNet50 (leve x moderada) (c) 10 - Engenharia de Características + ResNet50 (moderada x severa) (d) 10 - Engenharia de Características + ResNet50 (leve x moderada) (e) 10 - Engenharia de Características + ResNet50 (leve x severa) (f) 10 - Engenharia de Características + ResNet50 (moderada x severa)

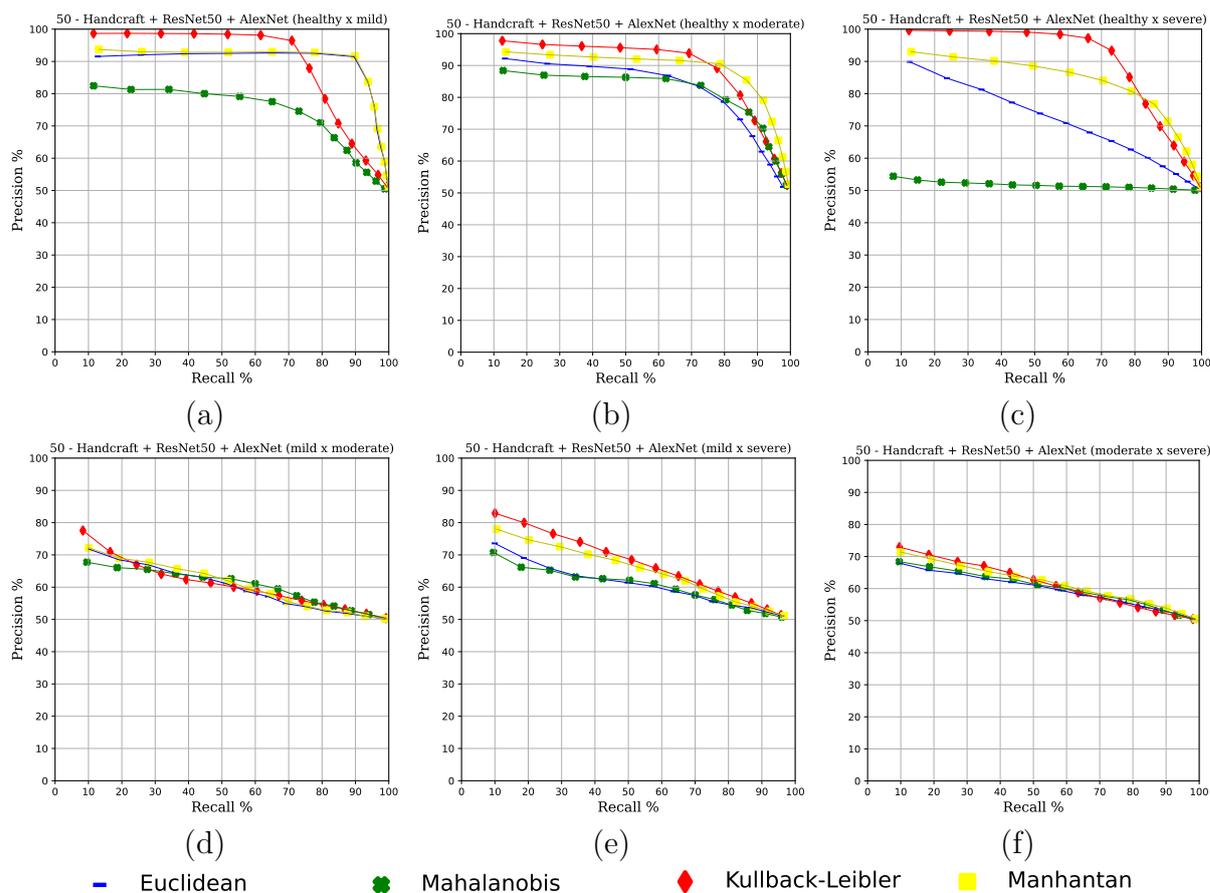


Figura 53 – Análise dos modelos em relação as métricas precisão e revogação: (a) 50 - Engenharia de Características + ResNet50 + AlexNet (saudável x leve) (b) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 50 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 50 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 50 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)

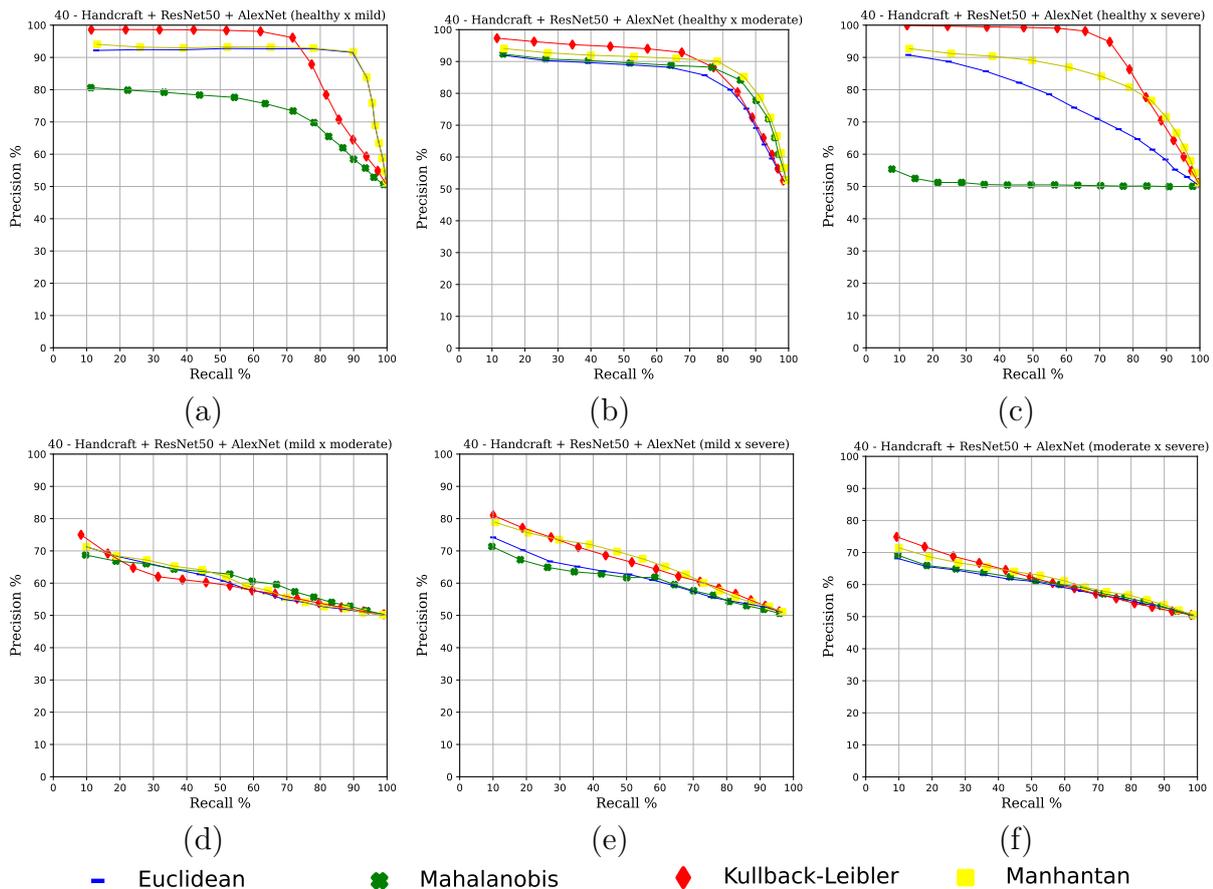


Figura 54 – Análise dos modelos em relação as métricas precisão e revogação: (a) 40 - Engenharia de Características + ResNet50 + AlexNet (saudável x leve) (b) 40 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 40 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 40 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 40 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 40 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)

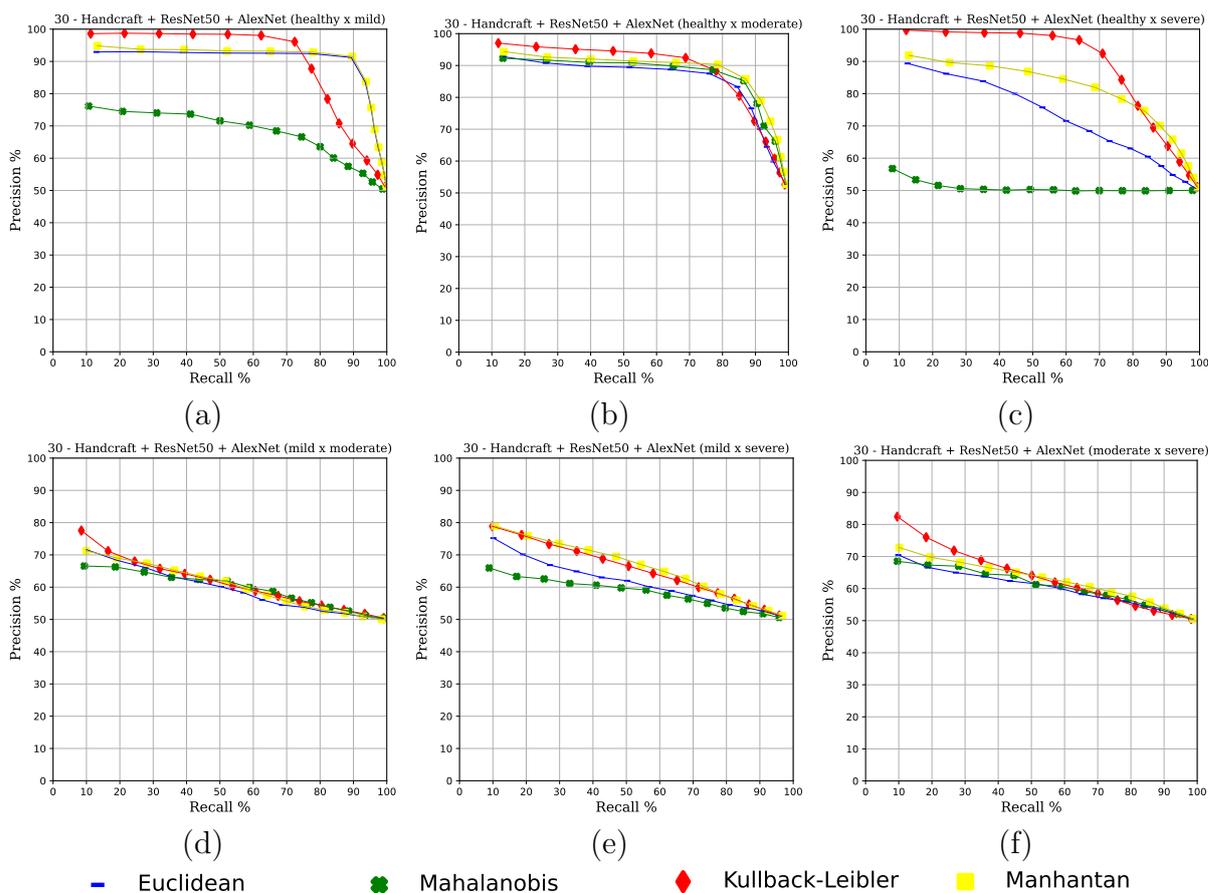


Figura 55 – Análise dos modelos em relação as métricas precisão e revogação: (a) 30 - Engenharia de Características + ResNet50 + AlexNet (saúdável x leve) (b) 30 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 30 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 30 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 30 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 30 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)

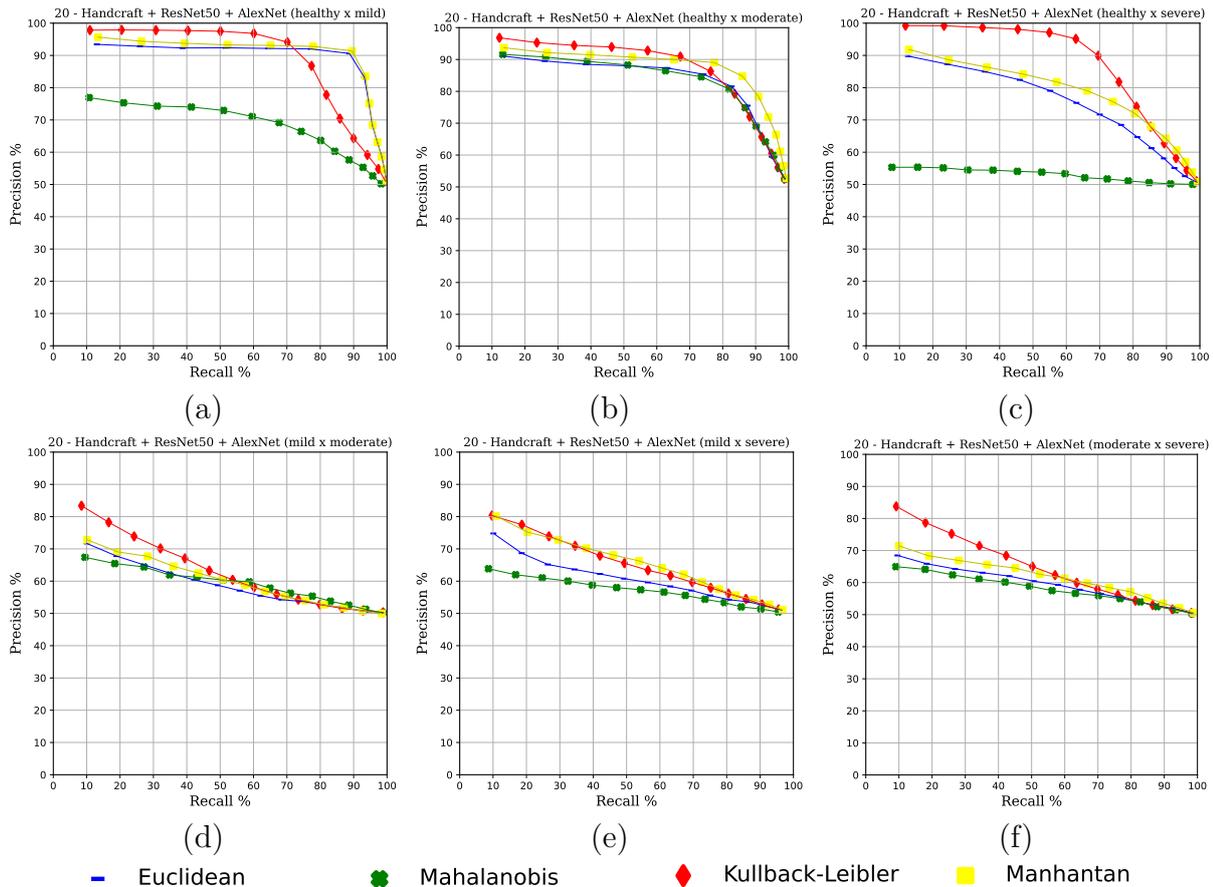


Figura 56 – Análise dos modelos em relação as métricas precisão e revogação: (a) 20 - Engenharia de Características + ResNet50 + AlexNet (saúdável x leve) (b) 20 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 20 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 20 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 20 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 20 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)

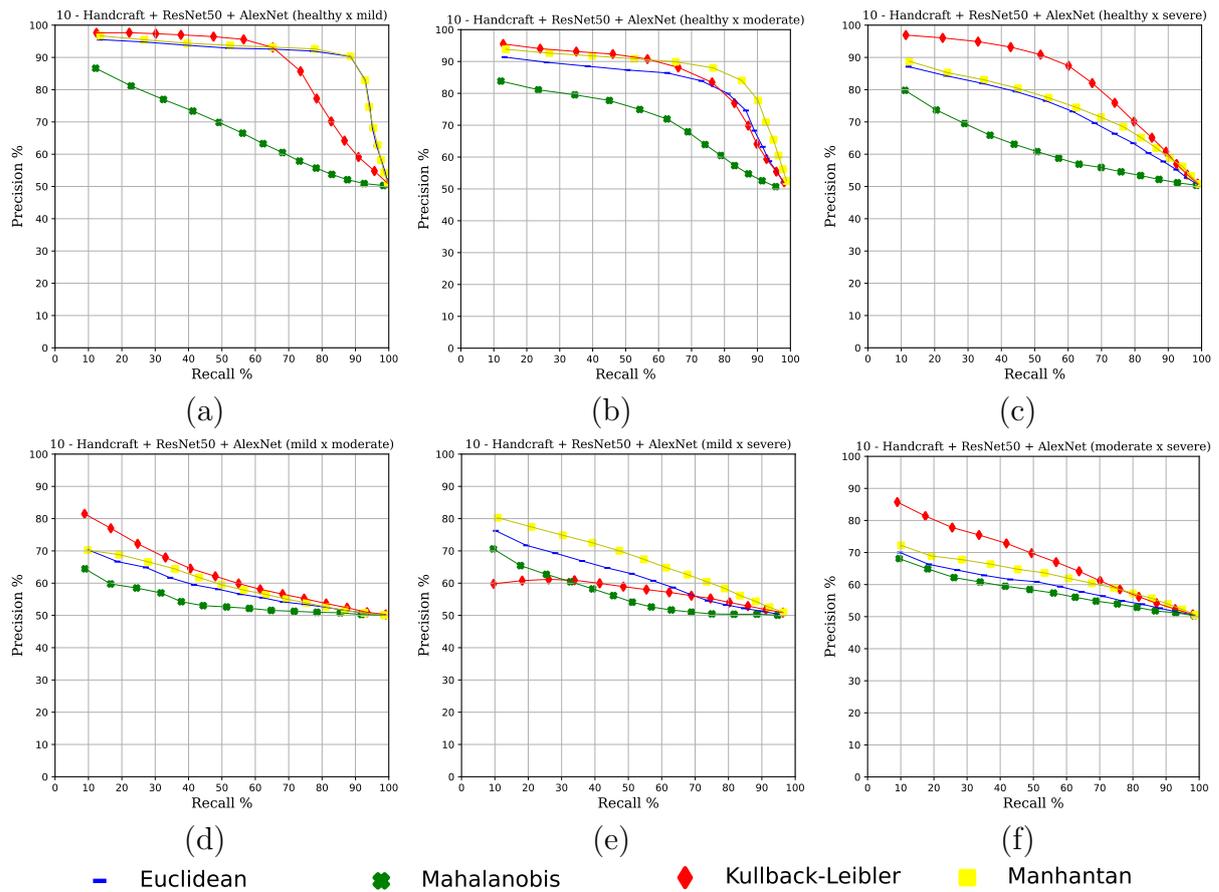


Figura 57 – Análise dos modelos em relação as métricas precisão e revogação: (a) 10 - Engenharia de Características + ResNet50 + AlexNet (saudável x leve) (b) 10 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (c) 10 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa) (d) 10 - Engenharia de Características + ResNet50 + AlexNet (leve x moderada) (e) 10 - Engenharia de Características + ResNet50 + AlexNet (leve x severa) (f) 10 - Engenharia de Características + ResNet50 + AlexNet (moderada x severa)