**Classificação de úlceras venosas dermatológicas para apoio a
consultas por similaridade utilizando superpixels e
aprendizado profundo**

Gustavo Blanco

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências
de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gustavo Blanco

Classificação de úlceras venosas dermatológicas para apoio a consultas por similaridade utilizando superpixels e aprendizado profundo

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Agma Juci Machado Traina

USP – São Carlos
Maio de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

B641c Blanco, Gustavo
 Classificação de úlceras venosas dermatológicas
 para apoio à consultas por similaridade utilizando
 superpixels e aprendizado profundo / Gustavo
 Blanco; orientadora Agma Juci Machado Traina. --
 São Carlos, 2019.
 72 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
 em Ciências de Computação e Matemática
 Computacional) -- Instituto de Ciências Matemáticas
 e de Computação, Universidade de São Paulo, 2019.

 1. Recuperação de imagens por conteúdo. 2.
 Recuperação de imagens médicas. 3. Consultas por
 similaridade. 4. Redes convolucionais. 5.
 Processamento de imagens. I. Traina, Agma Juci
 Machado, orient. II. Título.

Gustavo Blanco

Classification of venous dermatological ulcers to support similarity queries using superpixels and deep learning

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Agma Juci Machado Traina

USP – São Carlos
May 2019

*Dedico este trabalho a todas as pessoas que me motivaram
a sonhar e acreditar na ciência como profissão.
Em especial, aos meus pais que fizeram o possível e o impossível para que eu chegasse até aqui*

AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar aos meus pais Antonio e Roseli. Foram eles os responsáveis por eu sempre buscar algo além do que existia ao meu redor, a encarar o caminho mais difícil e recompensador em função do simples e imediatista. Minha formação e quaisquer outros títulos que eu venha a conquistar são tão meus quanto deles.

Agradeço também às minhas irmãs Beatriz e Bianca. Nossas discussões infinitas sobre os mais diversos assuntos me auxiliaram a ser uma pessoa crítica e disposta a ouvir outros pontos de vista.

Agradeço à minha namorada e eterna companheira Thalyne. Foi seu apoio e infinita compreensão que me fizeram insistir e ter coragem para me arriscar e colher os frutos que essas decisões me proporcionaram. Ter você ao meu lado faz toda a diferença.

Agradeço à minha orientadora Agma pelo carinho e compreensão demonstrada em momentos difíceis de minha vida. Tive a honra de aprender com você como um verdadeiro mentor deve agir, não só considerando aspectos científicos mas principalmente em como auxiliar as pessoas ao seu entorno a lidarem com as dificuldades e seguirem em frente.

Agradeço muito meu co-orientador de longa data Marcos Bedo pela constante ajuda na minha vida acadêmica e por me direcionar inúmeras vezes para o caminho certo.

Agradeço também os membros do GBdI, em especial o Willian, Luiz, Lúcio, Christian. Todos demonstraram muita paciência para me auxiliar e incentivar desde que eu era apenas um aluno de iniciação científica.

Gostaria de agradecer também a todos os membros do Laboratório de Inovação do Hospital Israelita Albert Einstein, em especial meus companheiros de trabalho do Varstation.

Por fim, agradeço ao CNPQ E Fapesp por financiar e apoiar parte desta pesquisa.

*“All sorts of computer errors are now turning up.
You’d be surprised to know the number of doctors who claim they are treating pregnant men.”
(Isaac Asimov)*

RESUMO

BLANCO, G.. **Classificação de úlceras venosas dermatológicas para apoio a consultas por similaridade utilizando superpixels e aprendizado profundo**. 2019. 71 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Sistemas de recuperação de imagens por conteúdo (do inglês “Content-based Image Retrieval” - CBIR) têm sido cada vez mais utilizados em diversas aplicações de tratamento e análise de imagens, devido a dois fatores: CBIR é um procedimento que pode ser feito automaticamente, permitindo tratar o grande volume de imagens adquiridos em hospitais, e também é a base para o processamento de consultas por similaridade. No contexto médico tais sistemas auxiliam em diversas tarefas, desde treinamento de profissionais até em sistemas de auxílio a diagnóstico (do inglês “Computer-Aided Diagnosis” - CAD). Um sistema computacional capaz de comparar e classificar imagens obtidas em exames de pacientes utilizando uma base prévia de conhecimento poderia agilizar o atendimento da população e fornecer aos especialistas informações relevantes de forma rápida e simples. Neste trabalho, o foco foi na análise de imagens de úlceras venosas. Foram desenvolvidas duas técnicas para classificação dessas imagens. A primeira, denominada Counting-Labels Similarity Measure (CL-Measure) possui a vantagem de lidar com imagens segmentadas de forma automática, por superpixels, e ser versátil o suficiente para permitir adaptação para outros domínios. A ideia principal do CL-Measure consiste na criação de sub-imagens baseadas em uma classificação prévia, calcular a distância entre elas e agregar as distâncias parciais obtidas a partir de uma função apropriada. A segunda técnica, denominada Quality of Tissues from Dermatological Ulcers (QTDU), faz uso de redes convolucionais (CNNs) para rotulação dos superpixels com a vantagem de compor todo o processo de identificação de características e classificação, dispensando a necessidade de identificar qual o extrator de características mais adequado para o contexto em questão. Experimentos realizados sobre a base de imagens analisada, utilizando 179572 superpixels divididos em 4 classes, indicam que a QTDU é a abordagem mais eficaz até o momento para o contexto de classificação de imagens dermatológicas, com médias de AUC=0,986, sensibilidade=0,97, e especificidade=0,974 superando as abordagens anteriores baseadas em aprendizado de máquina em 11,7% e 8,2% considerando o coeficiente KAPPA e F-Measure, respectivamente.

Palavras-chave: CBIR, Processamento de imagens, Consultas por similaridade, CNN.

ABSTRACT

BLANCO, G.. **Classification of venous dermatological ulcers to support similarity queries using superpixels and deep learning**. 2019. 71 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Content-based Image Retrieval (CBIR) systems have been increasingly used in many image processing and analysis applications because of two factors: CBIR is a procedure that can be done automatically, allowing to handle the large volume of images acquired in hospitals, and it is also the basis for processing similarity queries. In the medical context, such systems assist in various tasks, from training of professionals to develop Computer-Aided Diagnosis – CAD systems. A computer system capable of comparing and classifying images obtained from patient exams using a prior knowledge base could expedite the care of the population and provide specialists with relevant information quickly. In this study, the focus was on the analysis of images of venous ulcers. Two techniques were developed to classify these images. The first, called Counting-Labels Similarity Measure (CL-Measure) has the advantage of dealing with automatically segmented images by superpixels, and is versatile enough to allow adaptation to other domains. The main idea of CL-Measure is to create sub-images based on a previous classification, calculate the distance between them and add the partial distances obtained from an appropriate function. The second technique, called Quality of Tissues from Dermatological Ulcers (QTDU), makes use of convolutional networks (CNNs) for superpixels labeling, with the advantage of encompassing the whole process of identification of features and classification, without the need of identifying which extractor would be the best for the context in question. Experiments carried out on the image database using 179,572 superpixels divided into 4 classes, indicate that the QTDU is the most effective approach to date for the context of classification of dermatological ulcer images, with averages of $AUC = 0.986$, sensitivity = 0.97, and specificity = 0.974, surpassing previous approaches based on machine learning in 11.7% and 8.2% considering the KAPPA and F-Measure coefficients, respectively.

Keywords: CBIR, Image processing, Similarity search, CNN.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de arquitetura de um sistema <i>CBIR</i>	26
Figura 2 – Processo de extração de características.	28
Figura 3 – (a) Processo de extração global gerando um único vetor de características e (b) extração local onde se extrai informações de sub-regiões da imagem de entrada.	29
Figura 4 – Exemplo de um histograma de cores a partir de uma imagem em escala de cinza.	29
Figura 5 – (a) Exemplo de histograma original e (b) histograma métrico calculado sobre o original.	30
Figura 6 – Exemplo de identificação dos pontos chaves via SIFT.	31
Figura 7 – Exemplo de identificação de <i>superpixels</i> utilizando o <i>SLIC</i>	33
Figura 8 – Arquitetura da AlexNet.	35
Figura 9 – Modulo inception.	36
Figura 10 – Modulo residual.	37
Figura 11 – Exemplos de distâncias (a) <i>City-block</i> , (b) <i>Euclidiana</i> , (c) <i>Chebyshev</i> e (d) <i>Norma fracionada</i> com $p = 0,5$. O ponto preto indica o centro a partir do qual todos os elementos que estão na borda possuem mesma distância.	39
Figura 12 – Exemplo de (a) consulta por abrangência, (b) aos k -vizinhos mais próximos ($k = 4$) e (c) diversidade aos k vizinhos mais próximos de acordo com (SANTOS <i>et al.</i> , 2013). A estrela representa o elemento de consulta s_q e os quadrados em verde consistem nos elementos retornados em cada uma das consultas enquanto ε representa o raio da consulta por abrangência.	40
Figura 13 – Interface do Kundaha para realização de consultas por similaridade e diversidade.	45
Figura 14 – Taxonomia empregada para classificação de imagens dermatológicas em (ESTEVA <i>et al.</i> , 2017).	46
Figura 15 – Exemplos de imagens de úlceras venosas.	47
Figura 16 – <i>Pipeline</i> de execução do <i>CL-Measure</i> para comparação de duas imagens dermatológicas.	51
Figura 17 – Resultado da consulta aos 5 vizinhos mais próximos considerando os cenários (a) <i>CL-Measure</i> não métrico, (b) <i>CL-Measure</i> métrico, e (c) <i>MEDIDA_PADRAO</i>	54

Figura 18 – Curvas de precisão e revocação para as medidas de similaridade <i>MEDIDA_-PADRAO</i> , <i>CL-Measure</i> não métrico e <i>CL-Measure</i> métrico utilizando a <i>BASE_DERMA</i>	54
Figura 19 – Análise visual da <i>BASE_DERMA</i> . (a) Visualização <i>T-SNE</i> das características do <i>Color-structure</i> . (b) <i>PCA Scree plot</i> . (c) Variância individual do <i>PCA</i>	58
Figura 20 – Análise dos extratores do <i>MPEG-7</i> e classificadores considerando <i>ROC-AUC</i> e coeficiente <i>KAPPA</i>	59
Figura 21 – Arquitetura geral do QTDU. Na fase de aprendizado, a imagem da úlcera é dividida em um conjunto de superpixels para ajustar a CNN e suas seis novas camadas finais.	60
Figura 22 – Média e desvio padrão do tempo necessário para treinamento da (i) <i>RandomForest</i> com <i>Color Structure and Scalable Color-KG</i> e (ii) QTDU com <i>Inception V3</i> and <i>ResNet</i>	62
Figura 23 – F-Measures para todas as classes de lesões (maior o valor, melhor). (a) <i>Color Structure and RandomForest</i> . (b) <i>Scalable Color-KG and RandomForest</i> . (c) <i>ResNet QTDU</i> . (d) <i>Inception V3 QTDU</i>	63
Figura 24 – Exemplo de (a) imagem original, (b) segmentação manual, e (c) quantificação da área via QTDU.	64

LISTA DE TABELAS

Tabela 1 – Resumo de estudos focados na segmentação de lesões dermatológicas.	47
Tabela 2 – Comparação de extratores para superpixels de úlceras venosas.	52
Tabela 3 – Comparação do tempo de execução das funções de distância.	53
Tabela 4 – Comparação das abordagens machine learning e QTDU.	62

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Motivação	21
1.2	Objetivos do Trabalho	22
1.3	Organização do Texto	23
2	CONCEITOS E BASE TEÓRICA PARA A RECUPERAÇÃO DE IMAGENS POR CONTEÚDO	25
2.1	Visão geral	25
2.2	Extratores de características	27
2.2.1	<i>Extratores globais</i>	28
2.2.2	<i>Extratores locais</i>	30
2.2.3	<i>Segmentação para extração de características locais</i>	31
2.3	Extração de características por meio de redes neurais	33
2.3.1	<i>AlexNet</i>	34
2.3.2	<i>Inception</i>	35
2.3.3	<i>Resnet</i>	36
2.4	Consultas por similaridade	37
2.4.1	<i>Funções de distância</i>	38
2.4.2	<i>Consultas em espaços métricos</i>	39
2.5	Considerações finais	41
3	ASPECTOS DA RECUPERAÇÃO DE IMAGENS MÉDICAS POR CONTEÚDO	43
3.1	Visão geral	43
3.2	Sistemas CBMIR	44
3.3	Análise de imagens de úlceras venosas	46
3.4	Considerações finais	47
4	O MÉTODO CL-MEASURE (COUNTING-LABELS SIMILARITY MEASURE)	49
4.1	Visão geral	49
4.2	A abordagem CL-Measure	49
4.3	Experimentos	51
4.3.1	<i>Configurações para o CL-Measure</i>	52

4.3.2	<i>Avaliação do CL-Measure</i>	53
4.4	Considerações finais	55
5	O MÉTODO QTDU (QUALITY OF TISSUES FROM DERMATO- LOGICAL ULCERS)	57
5.1	Visão geral	57
5.2	A abordagem QTDU	57
5.2.1	<i>Experimentos</i>	58
5.2.2	<i>Configurações para o QTDU</i>	58
5.2.3	<i>Avaliação do QTDU</i>	61
5.2.3.1	<i>Métricas de tempo</i>	61
5.2.3.2	<i>Métricas de acurácia</i>	61
5.2.3.3	<i>Quantificação de áreas lesionadas a partir do QTDU</i>	63
5.3	Considerações finais	64
6	CONCLUSÕES	65
6.1	Considerações finais	65
6.2	Contribuições	65
6.3	Trabalhos futuros	66
	REFERÊNCIAS	67

INTRODUÇÃO

1.1 Motivação

A diminuição no custo de aquisição de imagens de alta qualidade, devido em maior parte à grande quantidade de dispositivos disponíveis como máquinas fotográficas, smartphones, webcams, além de máquinas de tomografia e ressonância magnética, motivaram o desenvolvimento de sistemas especialistas para armazenamento e manipulação deste tipo de dado. Aplicações baseadas na análise de imagens estão sendo cada vez mais utilizadas em diversos domínios de aplicação, como em mídias sociais, segurança e em contextos clínicos, quase sempre utilizando um sistema gerenciador de dados para armazenar as imagens em questão.

Usualmente, as imagens são inseridas na base de dados do hospital juntamente com o diagnóstico do especialista, para que posteriormente possa ser parte de uma base de conhecimento útil para a realização de futuras consultas. Estudos já demonstraram o potencial do uso desses tipos de bases de dados para apoiar especialistas na elaboração de novos diagnósticos por meio da recuperação e identificação de similaridades em casos anteriores ([MURAMATSU SHUNICHI HIGUCHI, 2018](#)). Por exemplo, um dermatologista pode procurar por padrões de úlceras venosas diagnosticadas anteriormente a partir de uma hipótese inicial, auxiliando assim o diagnóstico de novas lesões ([CHINO *et al.*, 2018](#)). Considerando o contexto citado, a utilização de um sistema de apoio pode auxiliar os médicos na identificação de tecidos afetados por meio da classificação automática das regiões analisadas em classes como saudável, granulação, fibrina ou pele necrosada. A partir desta classificação é possível diagnosticar o estágio no qual a lesão se encontra e indicar o tratamento mais eficaz. Outro exemplo é o uso de um sistema de apoio para identificação de fraturas em estruturas vertebrais capaz de fornecer informações como localização da lesão, densidade e dimensão da mesma ([LING *et al.*, 2017](#)). Ambos os processos são realizados por meio de um sistema especialista que recebe o nome de sistema de auxílio a diagnóstico ou *CAD* (do inglês, *Computer-Aided Diagnosis*).

Sistemas CAD tipicamente realizam consultas sobre a base de conhecimento utilizando dados de pacientes como imagens, anotações, entre outros, para auxiliarem futuras consultas realizadas por especialistas. Esta abordagem tem sido extensivamente utilizada para diagnósticos de diferentes patologias como melanomas e identificação de tumores (MISHRA; CELEBI, 2016). Uma forma alternativa de consultas em CAD é a realização de consultas baseadas nas características visuais. Esses sistemas são um importante complemento das buscas baseadas em textos, principalmente em casos com diagnósticos similares ou quando a percepção visual imediata é similar mesmo sendo diferentes tipos de lesões. Sistemas baseados na recuperação de imagens a partir de suas características visuais são chamados de sistemas CBIR, ou no caso de imagens médicas sistemas CBMIR (do inglês, *Computer-Based Medical Image Retrieval*).

Diversos sistemas CBMIR já foram propostos considerando diferentes contextos de aplicação. Um deles foi construído justamente para lidar com imagens tridimensionais obtidas a partir de ressonâncias magnéticas cerebrais (JENITTA; RAVINDRAN, 2017) com o objetivo de aproveitar também a interpretação visual da imagem ao invés de ficar restrito apenas às informações textuais. Para fazer a extração de informações, as imagens tridimensionais foram alinhadas e analisou-se a correlação entre os *pixels* de cada par de imagens obtendo assim os chamados “volumes regionais” que são utilizados para análise posterior.

Um outro domínio de imagens médicas o qual sistemas CBMIRs já foram propostos é o de imagens de mamografias, como por exemplo o sistema chamado *Kundaha* (SANTOS *et al.*, 2018). O *Kundaha* foi desenvolvido com o objetivo de auxiliar na recuperação de imagens médicas a partir de uma interface gráfica intuitiva e permitindo a escolha dos parâmetros a serem utilizados em cada consulta. As imagens em questão foram analisadas utilizando extratores de características globais e seus vetores de características comparados a partir das funções de distância tradicionais, ou seja, a partir da comparação entre pares de vetores de características.

Em ambos CBMIRs citados, os resultados preliminares demonstraram potencial de uso em contextos reais com possibilidades de aplicação em outros domínios de imagens médicas, enfatizando a importância no desenvolvimento dessas aplicações.

1.2 Objetivos do Trabalho

Este projeto de mestrado tem como objetivo desenvolver métodos e algoritmos para extração de características de imagens médicas com foco na análise de imagens de úlceras venosas e suas particularidades, permitindo que os algoritmos gerados possam ser integrados em sistemas CBMIR futuros. Mais especificamente, os objetivos a serem atingidos são:

- Desenvolvimento de uma técnica que identifique as regiões lesionadas de imagens dermatológicas de úlceras venosas e as rotule de acordo com a classificação realizada por especialistas.

- Desenvolvimento de uma nova abordagem para permitir a comparação entre imagens considerando o conteúdo de suas sub-regiões e a área que as mesmas ocupam.
- Analisar o impacto das abordagens desenvolvidas no contexto de imagens de úlceras venosas considerando trabalhos estado da arte.
- Divulgar a técnica proposta para que fique disponível à comunidade científica para apoio em novos trabalhos e futuras melhorias.

A validação dos resultados foi feita por meio da comparação da acurácia obtida pela combinação de diferentes extratores e classificadores presentes na literatura utilizando conjuntos de imagens médicas reais como o de úlceras venosas (CHINO *et al.*, 2018), obtidas em parceria com o Centro de Ciências Física Médica (CCIFM) do Hospital das Clínicas de Ribeirão Preto (HCRP-USP). Este conjunto de imagens fornecido também foi segmentado e rotulado manualmente por dermatologistas para a construção de uma base de imagens derivadas constituída por *superpixels* das imagens originais.

1.3 Organização do Texto

Este trabalho está organizado em quatro capítulos. No capítulo 2, são descritos os conceitos que envolvem o desenvolvimento de um sistema *CBIR* como extratores de características e funções de distâncias assim como as vantagens e desvantagens no uso de cada abordagem. Também são apresentadas abordagens *deep learning* que têm sido empregadas para lidar com classificações de imagens e exemplos de sistemas *CBMIR* que fazem uso desses conceitos. No Capítulo 3 são apresentados alguns dos diversos sistemas *CBMIR* existentes e quais são as particularidades que devem ser consideradas no desenvolvimento dos mesmos. Nos capítulos 4 e 5 são apresentadas as novas abordagens desenvolvidas: o *Counting-labels Similarity Measure* e o *Quality of Tissues from Dermatological ulcers*. Finalmente, no capítulo 6 são apresentadas as conclusões desta dissertação além das contribuições e dos trabalhos futuros a serem desenvolvidos.

CONCEITOS E BASE TEÓRICA PARA A RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

Neste capítulo é apresentada uma visão geral sobre sistemas de recuperação de imagens por conteúdo e as abordagens empregadas para tal. Deve-se ressaltar que o texto apresentado neste capítulo visa oferecer a base essencial para o desenvolvimento desse tipo de sistema, e não contempla o vasto universo de extratores de características, funções de distâncias e redes neurais que existem na literatura. Além disso serão apresentados os desafios na criação deste tipo de sistema e que devem estar sempre em foco durante todo o processo de desenvolvimento. Os algoritmos aqui descritos estão entre os mais utilizados em trabalhos da literatura e, conseqüentemente, são usados em diversos domínios de aplicação.

2.1 Visão geral

O estudo e desenvolvimento de sistemas *CBMIR* têm crescido significativamente na última década, impulsionado principalmente pelo aumento da aquisição e armazenamento de imagens digitais proporcionado por novas tecnologias. Entretanto, o problema de representação de imagens para acesso e consultas em bases de dados tem sido abordado desde o início da década de 1980 ([CHANG; KUNIL, 1981](#)), contribuindo para o acúmulo de conhecimento na área e maior maturidade nos trabalhos desenvolvidos atualmente.

O principal obstáculo deste tipo de sistema consiste em como representar o conteúdo de determinada imagem de forma a sintetizar os aspectos visuais relevantes para o usuário. Este fato torna a percepção do ser-humano indispensável durante o processo de desenvolvimento e possui grande impacto no sistema final. É necessário que a aplicação seja capaz de recuperar casos similares considerando essa percepção, automatizando assim a análise que o usuário faria para

toda uma base de imagens. Como o computador analisa características de mais baixo nível (cor, textura, forma) existe uma separação clara entre a interpretação de uma imagem considerando um ser-humano e um computador. Essa separação é chamada de *gap semântico* (HARE *et al.*, 2006) e deve ser abordada de acordo com o contexto da aplicação. Sistemas *CBIR* são utilizados nas mais diversas aplicações, como segurança (LIU *et al.*, 2018), imagens geográficas (TEKESTE; DEMIR, 2018), ambientes clínicos (ESTEVA *et al.*, 2017), entre outros. Obviamente cada uma dessas aplicações implicam em uma série de particularidades que devem ser consideradas, o que força o desenvolvimento de algoritmos específicos e dificulta a reutilização de abordagens anteriores em novos domínios. A figura 1 apresenta uma arquitetura simplificada de um sistema *CBIR*.

Figura 1 – Exemplo de arquitetura de um sistema *CBIR*.

Fonte: Elaborada pelo autor.

Usualmente sistemas *CBIR* são organizados em 3 estágios: (i) Extração de características; (ii) Aplicação de uma função de distância; e (iii) Realização de consulta por similaridade, para obter o ranqueamento das imagens mais semelhantes que atendem à consulta. O primeiro estágio utiliza algoritmos de extração de características para analisar a imagem e representá-la numericamente permitindo que seja aplicada uma função de avaliação de distância. A partir de uma função de distância obtém-se usualmente um vetor que será utilizado para a realização da consulta por similaridade que permitirá a recuperação das imagens mais similares. Para que seja possível realizar uma recuperação eficaz, deve-se caracterizar e quantificar a similaridade perceptual de acordo com o especialista da área em questão. Dessa forma é crucial identificar o perfil do usuário que fará uso do sistema e traduzir sua percepção em parâmetros mensuráveis e configuráveis dentro do sistema *CBIR*. O conjunto de extrator de características e função de avaliação de distância é chamado de descritor (TORRES; FALCAO, 2006). Como a escolha do descritor reflete no tipo de conteúdo que será consultado e potencialmente traduzem a visão do usuário sobre as imagens, podemos fazer um paralelo dessa configuração com os chamados **parâmetros perceptuais** (BUGATTI *et al.*, 2011).

É possível também utilizar o *feedback* do usuário como forma de melhorar a consulta

realizada utilizando um processo que recebe o nome de *realimentação de relevância*. Neste processo, o usuário identifica em cada uma das consultas as imagens que ele considera relevantes, permitindo o uso de técnicas para que o sistema se adapte constantemente, potencialmente atingindo o caso ótimo onde todas as imagens retornadas são pertinentes. Além disso é comum utilizar técnicas de indexação a partir dos valores obtidos nas etapas de extração de características e avaliação de distância, já que índices métricos e multidimensionais e outros métodos de acesso como certos tipos de *hashing* podem ser utilizados para vetores de características.

Nas próximas seções serão apresentadas abordagens referentes a técnicas de extração de características e funções de distância, que podem ser consideradas técnicas mais "clássicas" por serem as abordagens utilizadas desde as primeiras propostas para recuperação de imagens. Além disso também serão apresentadas abordagens mais atuais baseadas em *deep learning* e como utilizar os dados obtidos como parte do processo de classificação de novas imagens.

2.2 Extratores de características

Uma imagem é representada digitalmente a partir de uma matriz de pontos (*pixels*). Cada *pixel* possui um conjunto de valores numéricos que representam a intensidade de cada um dos canais de cores associados. Estes canais de cores dependem do modelo adotado. Usualmente as imagens digitais são armazenadas utilizando o modelo de cores *RGB* (*Red*, *Green* e *Blue*) onde cada canal possui um valor de intensidade que normalmente varia entre 0 e 255. A composição dessas tonalidades determina a cor que será apresentada. Embora a cor final exibida em cada *pixel* seja independente do espaço de cores adotado, a escolha de cada modelo afeta como esses valores são manipulados e também o custo de memória associado.

Extratores de características são algoritmos que possuem como objetivo sintetizar as informações mais relevantes de uma imagem a partir de determinados critérios, critérios estes que dependem exclusivamente do tipo de característica que é relevante na análise em questão. Conceitualmente, este processo visa reproduzir matematicamente observações que descrevem determinado aspecto da imagem, identificando bordas, objetos, regiões relevantes, entre outros. Formalmente, a partir de imagens pertencentes ao domínio $\mathbb{I}$, um método de extração de características é uma função computacional não bijetiva $\mathbb{I} \rightarrow \mathbb{S}$, onde $\mathbb{I}$ é o domínio de imagens.

Em geral essas características podem ser englobadas em três grandes grupos: características de cor, características de textura e características de forma. Usualmente, um algoritmo de extração de características analisa exclusivamente um desses grupos, sendo necessário o uso simultâneo de diferentes extratores quando deseja-se observar diferentes grupos de características ao mesmo tempo. Esta composição possivelmente aumenta a qualidade do resultado porém aumenta significativamente o número de dimensões geradas.

Devido à natureza multidimensional de imagens digitais os algoritmos de extração de características podem ser interpretados como um caso especial de transformação e redução

de dimensões, já que os extratores consistem em uma função computacional que é capaz de transformar um conjunto de informações presentes na imagem (*pixels*) em um conjunto de valores reduzidos, seja ele multidimensional ou não, que recebe o nome de vetor de características (Figura 2). Esta é uma diferença entre extratores de características e algoritmos de seleção de dimensões, já que as técnicas de seleção constroem o vetor final a partir das características originais, o que faz com o que o resultado seja apenas um subconjunto dos valores já existentes.

Figura 2 – Processo de extração de características.

Assim como em algoritmos de seleção de características é esperado que os valores obtidos sejam representativos permitindo assim descrever a imagem original a partir de um conjunto menor de atributos. De acordo com Lippman (LIPPMANN, 1989), as características devem possuir informações suficientes para distinção entre as classes, não ser sensível a variações irrelevantes na entrada, e também possuir uma quantidade de dimensões que permita cálculos eficientes e que limite o total de dados necessários para treinamento. Percebe-se então que a escolha do algoritmo de extração deve ser coerente com as características desejadas para o contexto de aplicação. Nas próximas seções serão discutidas algumas técnicas de extração relevantes e suas particularidades.

2.2.1 Extratores globais

Descritores de imagens podem ser definidos como globais e locais (TIAN *et al.*, 2013) (Figura 3). Neste trabalho, associaremos extratores globais nos casos em que toda a imagem é analisada a fim de produzir um resultado capaz de descrever genericamente todo o conteúdo presente na mesma. De forma simplificada assumiremos que extratores globais são aqueles que analisam toda a imagem e geram um único vetor de características correspondente. No entanto tal tarefa é desafiadora já que existem diversos fatores que devem ser analisados e que possuem grande impacto neste processo, como iluminação, orientação dos objetos na cena, oclusões, entre outros. Os extratores globais são os mais utilizados na literatura mas também são conhecidos por uma série de limitações, principalmente quando a característica a ser analisada na imagem consiste apenas em um objeto ou região de interesse.

Uma das técnicas de extração global mais conhecida é o chamado histograma de cores (NOVAK; SHAFER, 1992). O histograma de cores faz uma contagem do total de *pixels* de cada uma das 256 intensidades (considerando uma imagem em escala de cinza) presentes na imagem,

Figura 3 – (a) Processo de extração global gerando um único vetor de características e (b) extração local onde se extrai informações de sub-regiões da imagem de entrada.

Fonte: Elaborada pelo autor.

sendo que cada um desses valores recebe o nome de *bin*. O conjunto desses *bins* formam o vetor de características da imagem, reduzindo assim a dimensão da sua representação original de $N \times M$, onde N é a altura e M é a largura da imagem, para um total de 256 valores. É importante que a dimensionalidade do vetor de características não seja demasiadamente grande pois irá afetar negativamente o processo de classificação e indexação, como será discutido mais adiante na seção de consultas por similaridade.

Os principais tipos de histogramas utilizados são os histogramas de cores considerando tons de cinza, o histograma total de cores e o histograma métrico (TRAINA *et al.*, 2002). Como já citado, o primeiro caso faz a contagem de apenas 256 valores possíveis e resulta em um vetor numérico onde cada posição representa o total de *pixels* com o respectivo valor. No caso do histograma total de cores são considerados todos os 256 valores presentes em cada um dos três canais de cores (assumindo o modelo de cores *RGB*), totalizando 768 *bins* no vetor de características. A figura 4 apresenta um exemplo de um histograma de cores e sua imagem de origem.

Figura 4 – Exemplo de um histograma de cores a partir de uma imagem em escala de cinza.

Fonte: Elaborada pelo autor.

Já o histograma métrico (TRAINA *et al.*, 2002) pode lidar da mesma forma que os casos anteriores alterando apenas a representação do vetor de características final. Porém, em vez de armazenar explicitamente o total de *pixels* em cada um dos *bins* associados é utilizada uma aproximação gráfica do histograma original utilizando pontos de inflexão. Com isso, o processamento da comparação entre imagens fica até 4 vezes mais rápido do que com histogramas tradicionais. A figura 5 apresenta um exemplo de um histograma métrico.

Figura 5 – (a) Exemplo de histograma original e (b) histograma métrico calculado sobre o original.

Fonte: Elaborada pelo autor.

2.2.2 Extratores locais

Imagens podem conter uma coleção de objetos ou regiões de interesse com capacidade para definirem semanticamente o conteúdo que está sendo apresentado. Dessa forma, faz-se necessária a criação de métodos capazes de analisar individualmente essas regiões eliminando assim informações que não são relevantes para a descrição da imagem. Para tal, é necessário que os descritores locais identifiquem previamente quais são essas regiões de interesse para então utilizar os respectivos métodos de extração de características. Por este motivo este tipo de descritores possuem duas principais etapas de execução: detecção e descrição.

Os extratores locais diferem dos globais principalmente pelo resultado gerado, já que ao invés de produzir um único vetor de características representativo para toda imagem, são produzidos vetores equivalentes ao total de regiões analisadas. De forma simplificada, neste trabalho consideraremos extratores locais como os extratores em que analisam sub-regiões da imagem original gerando vetores de características para cada uma delas, ou seja, o resultado final da análise da imagem será uma coleção de vetores de características. Esta característica impacta na análise de similaridade entre imagens, pois impede que comparações diretas sejam utilizadas. Dessa forma também é necessário utilizar abordagens que consigam avaliar a distância entre vetores com tamanhos diferentes.

Um algoritmo de extração que sintetiza bem a premissa de análise local é o *SIFT* (do inglês *Scale Invariant Feature Transform*) (LOWE, 1999) que consiste em uma abordagem

para detectar pontos de interesse em uma imagem. Esses pontos de interesse devem conter informações específicas o suficiente para permitirem uma identificação precisa do conteúdo presente na região a qual os mesmos se encontram. Dessa forma, uma imagem analisada pelo *SIFT* passa a ser descrita por uma grande coleção de vetores de características locais, sendo que o conteúdo de cada vetor é invariante a translação, rotação e escala, além de serem parcialmente invariantes a mudanças de iluminação.

A execução do *SIFT* pode ser dividida em quatro etapas: (i) detecção dos extremos; (ii) localização dos pontos chaves; (iii) definição da orientação; e (iv) descrição dos pontos chaves. Por serem obtidos um conjunto de pontos, cada imagem possui diversos descritores referentes a cada ponto chave identificado. Por este motivo a comparação entre diferentes imagens analisadas pelo *SIFT* não pode ser feita diretamente com as funções de distâncias tradicionais, o que exige o uso de uma técnica específica para essa extração de características. Usualmente essa comparação é feita a partir da redução do problema a uma busca de vizinho mais próximo.

Figura 6 – Exemplo de identificação dos pontos chaves via *SIFT*.

Fonte: Itseez (2014).

Devido à qualidade dos resultados obtidos pelo *SIFT* existem diversas propostas mais recentes para otimizar o tempo de execução do mesmo. A abordagem similar mais utilizada é o *SURF* (em inglês, *Speeded Up Robust Features*) (BAY; TUYTELAARS; GOOL, 2006) que apresenta resultados similares ao *SIFT* porém com menor tempo de execução. Uma das otimizações propostas no *SURF* está no cálculo aproximado do laplaciano da função gaussiana através de convoluções utilizando “caixas”, ou seja, agrupando regiões de *pixels* antes de aplicar o filtro de convolução.

2.2.3 Segmentação para extração de características locais

A segmentação de imagens consiste na divisão de uma imagem em sub-regiões por meio de determinado critério. Embora consista em uma das etapas mais importantes no processamento de imagens ainda é considerada uma das etapas mais difíceis de serem realizadas para imagens

complexas. Uma segmentação precisa favorece a execução correta dos algoritmos de extração de características pois identifica previamente quais segmentos da imagem efetivamente possuem informações relevantes para a análise. Da mesma forma, uma segmentação incorreta pode inutilizar todas as etapas posteriores de análise impedindo a obtenção de resultados relevantes ao final da execução de todas as etapas de um sistema *CBIR*.

Grande parte dos algoritmos de segmentação utilizam os conceitos de descontinuidade e similaridade para analisar os *pixels* da imagem e identificar regiões passíveis de segmentação (NIXON; AGUADO, 2012). No primeiro caso são localizadas mudanças bruscas no valor das intensidades dos *pixels* para localização dos limites da região segmentada, de forma que os valores médios dos *pixels* de um lado desta fronteira sejam muito maiores ou menores em relação aos valores presentes na fronteira adjacente. Este tipo de segmentação é chamada de detecção de bordas.

Dentre as diversas técnicas para segmentação, uma proposta com bastante destaque consiste na segmentação da imagem a partir do uso de *superpixels*. Esta técnica de segmentação tem como objetivo identificar regiões que apresentam *pixels* redundantes entre si, ou seja, que possuam valores similares entre os *pixels* adjacentes e os agrupam formando uma nova unidade de representação da imagem que recebe o nome de *superpixel*. Um algoritmo para geração de *superpixels* atua como uma função $G(I, K_p) = \{S_j \mid 0 \leq j < K_p\}$ que a partir de uma imagem I gera K_p *superpixels* S_j , $S_j \subset I$. É esperado que os *superpixels* identifiquem regiões com boa aderência às suas bordas de forma que atue como suporte a técnicas de segmentação.

Cada uma das técnicas propostas para geração de *superpixels* apresentam vantagens e desvantagens diretamente relacionadas ao tipo de aplicação desejada. Entretanto, é esperado que independentemente da técnica selecionada o processo de geração de *superpixels* seja eficiente o suficiente, já que o ganho obtido nas etapas posteriores devido a redução de tamanho da imagem pode não se justificar se o tempo de execução necessário para tal for elevado. Além disso os *superpixels* devem ter uma boa aderência de bordas na imagem para que regiões com informações visuais diferentes não sejam agrupadas em uma única unidade.

Neste contexto, um método muito utilizado recentemente com bons resultados é o chamado *Simple Linear Iterative Clustering (SLIC)* (ACHANTA *et al.*, 2012). O *SLIC* possui a vantagem de ser de simples utilização uma vez que o único parâmetro a ser definido é o valor de k , que corresponde ao número aproximado de *superpixels* de tamanhos similares que serão gerados sobre a imagem. A partir desse parâmetro são definidos k centros espaçados igualmente na imagem. Os *pixels* então são associados de forma iterativa a cada centro baseados na sua similaridade, de forma que ao final deste processo a região de cada *superpixel* está definida. A Figura 7 apresenta um exemplo do uso do *SLIC* para identificação de *superpixels* em diferentes imagens.

Como citado no artigo original (ACHANTA *et al.*, 2012), o *SLIC* possui a vantagem de ser mais eficiente e preciso em relação aos outros principais algoritmos da literatura como o

Figura 7 – Exemplo de identificação de *superpixels* utilizando o *SLIC*.

Fonte: Achanta *et al.* (2012).

GS04 (FELZENSZWALB; HUTTENLOCHER, 2004) e NC05 (REN; MALIK, 2003), além de sua versatilidade para diferentes domínios de aplicação. Por este motivo apresenta potencial para uso em imagens médicas, objeto de estudo deste trabalho.

2.3 Extração de características por meio de redes neurais

As redes convolucionais neurais (CNN) têm se destacado por ser uma abordagem extremamente eficaz na análise e classificação de imagens. Uma das características mais relevantes é que essas redes atuam como ótimos extratores de características, permitindo que se extraia atributos relevantes de uma imagem utilizando uma rede pré-treinada apenas fazendo ajustes em seus parâmetros (GOODFELLOW *et al.*, 2016)). Além disso devido aos filtros de diferentes tamanhos que usualmente compõem as arquiteturas dessas redes as mesmas acabam generalizando o processo de extração global e local, sendo invariantes a escala e translação dos objetos em cena. Outra vantagem das CNNs é o compartilhamento de pesos, o que reduz significativamente a quantidade de informações a serem armazenadas tornando-as escaláveis para imagens de alta resolução.

Uma CNN pode ser descrita como uma série de camadas convolucionais e de subamostragem comumente seguidas por uma camada totalmente conectada. Uma camada convolucional consiste na aplicação da operação de convolução na camada de entrada (basicamente cálculo do produto vetorial entre a matriz de pesos e região de sobreposição) e a propagação desse resultado para a camada posterior. Essa operação pode alterar a dimensão original dependendo do tamanho e quantidade de filtros utilizados. Já na operação de subamostragem (camada de *pooling*) o objetivo principal é reduzir a dimensão da camada anterior tentando propagar apenas a informação mais relevante. Existem uma série de abordagens para a realização dessa operação sendo o *max-pooling* o mais comum, que consiste em selecionar o maior elemento do grupo que está sendo observado.

Para determinar o resultado da rede são aplicadas as chamadas funções de ativação. Essas funções têm como objetivo tornar a transição entre as classificações da rede menos brusca, permitindo que pequenas alterações nos pesos reflitam também suavemente na saída final. A função de ativação é a transformação não linear aplicada ao longo do sinal de entrada. Caso a função de ativação seja omitida, os pesos simplesmente farão uma transformação linear, que embora seja uma operação simples de se calcular é extremamente limitada para a resolução de problemas mais complexos. Algumas das funções de ativação mais comuns são a sigmoide (Equação 2.1), a ReLU (Equação 2.2) e a softmax (Equação 2.3).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

$$f(x) = \max(0, x) \quad (2.2)$$

$$f(x)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \text{ para } j = 1, \dots, K \quad (2.3)$$

Enquanto na abordagem tradicional os extratores de características apresentem problemas na capacidade de generalização para diferentes contextos de aplicação (GOODFELLOW *et al.*, 2016), uma mesma CNN permite a obtenção de resultados satisfatórios sem a necessidade de grandes adaptações em sua arquitetura, muitas vezes exigindo apenas a variação de alguns de seus parâmetros. A seguir listamos algumas das CNNs mais conhecidas e utilizadas com maior frequência.

2.3.1 AlexNet

A CNN conhecida como AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) obteve grande destaque em 2012 quando venceu a competição ImageNet obtendo uma taxa de erro top-5 de apenas 15,3%, superando de forma expressiva o segundo colocado que obteve uma taxa de 26,2%. Com esse resultado a abordagem *Deep learning*, que não era alvo de estudos expressivos desde a Lenet (LECUN *et al.*, 1998) em 1998, passou a receber grande atenção e tornar-se padrão em estudos estado da arte em processamento de imagens.

Figura 8 – Arquitetura da AlexNet.

Fonte: Ramsundar e Zadeh (2018).

A AlexNet possui camada de entrada de 227×227 e se destaca por ser extremamente maior em comparação com as CNNs propostas até então. Ela possui cerca de 60 milhões de parâmetros e demorou entre cinco e seis dias para ser treinada. Sua arquitetura possui um total de 8 camadas. As duas primeiras são camadas convolucionais sucedidas por camadas de *max pooling*, para filtrar as informações mais relevantes de cada região. As três próximas camadas convolucionais são interligadas diretamente entre si enquanto a quinta é novamente sucedida por uma camada de *max pooling*. As camadas finais são compostas por duas camadas totalmente conectadas e finalmente por uma última de *softmax* responsável por gerar uma das 1000 classificações possíveis do ImageNet. A arquitetura simplificada pode ser vista na Figura 8.

Além de ser uma rede maior em relação às desenvolvidas até então, a AlexNet se destaca por ter usado a função de ativação ReLU (unidade linear retificada) diferenciando-se das funções Tahn e Sigmóide que eram padrões até então. Além disso pelos testes mostrou-se que o uso dessa função de ativação diminuiu o tempo de treinamento em aproximadamente 6 vezes em relação às outras funções. Além da melhora da função de ativação, este trabalho também contribuiu discutindo formas de redução do *overfitting*, como o uso de *data agumentation* através do espelhamento das imagens originais e a utilização de camadas de *Dropouts* como forma de regularização.

2.3.2 Inception

A motivação inicial ao se desenvolver a rede Inception (SZEGEDY *et al.*, 2015) focou-se em um dos principais problemas enfrentados ao se trabalhar com classificação de imagens. A região de interesse a ser classificada pode ocupar ou não toda imagem, sendo que nos casos em que não ocupa pode estar em qualquer posição e escala, o que dificulta determinar com precisão qual o tamanho do kernel mais adequado para não se perder essas informações.

Figura 9 – Módulo inception.

Fonte: Szegedy *et al.* (2015).

Para isso foi proposto o módulo inception (Figura 9) que consiste basicamente na ramificação da camada anterior para a aplicação de diferentes tamanhos de filtros ao mesmo tempo, seguindo um conceito de rede mais larga ao invés de mais profunda. Os filtros possuem tamanhos 1×1 , 3×3 e 5×5 além da aplicação de *max pooling*. A saída de cada uma dessas 4 operações são concatenadas e passadas para o próximo nível da rede. Ao todo, a rede *Inception* possui 9 módulos inceptions concatenados.

A primeira versão da Inception, que foi desenvolvida pela equipe da Google (SZEGEDY *et al.*, 2015), foi a vencedora do desafio Imagenet de 2014 obtendo um erro top-5 de apenas 6,67%, valor extremamente próximo ao do humano. Além disso a rede Inception original passou por diversas melhorias que contribuíram para que diversos outros estudos a utilizassem como base.

2.3.3 Resnet

Ao longo dos anos sempre existiu o interesse em desenvolver redes cada vez mais profundas, porém quanto maior a profundidade da rede maior a dificuldade envolvida no treinamento. O aumento dos parâmetros provocado pelo maior número de camadas torna o treinamento mais demorado, exige hardwares mais potentes e provoca ainda a saturação da acurácia da rede e degradação do resultado final. Para combater isso em 2015 foi proposta pela Microsoft a Resnet (HE *et al.*, 2016), rede vencedora do desafio Imagenet do mesmo ano obtendo um erro top-5 de 3,57%, valor incrivelmente baixo até para um humano. O grande diferencial proposto pela Resnet é o chamado “aprendizado residual” que é capaz de combater uma série de problemas existentes em redes mais profundas.

Figura 10 – Modulo residual.

Fonte: Ramsundar e Zadeh (2018).

A ideia central da Resnet é permitir a adição de mais camadas ao seu modelo sem aumentar o erro obtido pela arquitetura original. Para obter isso a Resnet compõe a saída obtida pela camada com os valores de entrada recebidos antes dessa última operação criando um "atalho" (Figura 10). Dessa forma a cada bloco está sendo aprendido um valor residual, ou seja, a subtração das características aprendidas em relação aos valores originais. Essa abordagem permitiu que a Resnet fosse a rede mais profunda até o momento de sua publicação possuindo 152 camadas. Além disso é citado no artigo original (HE *et al.*, 2016) que foram consideradas redes mais profundas com até 1200 camadas, porém a versão com 152 camadas foi o máximo que conseguiram chegar sem enfrentar o problema do *overfitting*.

2.4 Consultas por similaridade

Como já citado, para o armazenamento e recuperação de imagens digitais buscou-se novas maneiras de sintetizar os aspectos relevantes dessas "matrizes de *pixels*" para permitir a busca semântica dessas informações. Entretanto, usualmente a recuperação de dados é feita a partir da busca de valores exatos como números ou palavras, ou seja, por dados sujeitos a comparação por identidade dentro dos gerenciadores de dados tradicionais. Esse aspecto acaba por gerar uma limitação que afeta qualquer conjunto que não siga esse padrão ou onde buscas por valores exatos possuam pouco significado prático para o usuário, impedindo assim que diversas técnicas tradicionais para indexação e recuperação de dados possam ser utilizadas. Por esse motivo esse gerenciamento de dados tradicional torna-se ineficiente para o gerenciamento de dados multimídia, como imagens, vídeos e áudios.

Para tratar desses dados é necessário utilizar algoritmos e modelos mais elaborados capazes de realizar as buscas a partir de outros critérios. Embora existam diversos modelos, o principal modelo utilizado nesse contexto é o de consulta por similaridade (ZEZULA *et al.*, 2006), ou seja, a busca por elementos da base que sejam similares ou próximos do elemento de consulta a partir de um determinado critério. Uma modelagem consolidada para avaliação de

similaridade é chamada de consultas em espaços métricos. Um espaço métrico é definido por um par $\langle \mathbb{S}, \delta \rangle$, onde $\mathbb{S}$ é o domínio dos elementos e δ é uma função de distância métrica (ZEZULA *et al.*, 2006). Sejam s_i, s_j e s_l elementos de $\mathbb{S}$, uma função de distância δ é dita métrica caso apresente as seguintes propriedades:

- (i) Não-negatividade: $0 \leq \delta(s_i, s_j) < \infty$
- (ii) Simetria $\delta(s_i, s_j) = \delta(s_j, s_i)$
- (iii) Desigualdade triangular: $\delta(s_i, s_j) \leq \delta(s_i, s_l) + \delta(s_l, s_j)$

2.4.1 Funções de distância

Para que um sistema de recuperação por conteúdo tenha a capacidade de responder consultas por similaridade é necessário que ele possua meios de mensurar o quão similar são os objetos analisados. A representação numérica dessa similaridade usualmente é obtida a partir da utilização de uma função que calcula a distância (ou dissimilaridade) entre dois objetos e obtém um valor real (comumente positivo) que indica a similaridade entre eles. É esperado que quanto maior o valor obtido, menor a similaridade entre os objetos comparados, sendo que no caso particular da distância zero a comparação indique que ambos os objetos são idênticos.

Existem diversas funções de distâncias com uma série de aplicações e propriedades específicas. Essas funções podem ser classificadas em dois grupos: (i) discretas, caso os resultados obtidos pertençam a um pequeno conjunto de valores pré-definidos; e (ii) contínuas, se a cardinalidade dos valores obtidos seja infinita (ZEZULA *et al.*, 2006).

Dentre a infinidade de funções de distâncias existentes a família de funções Minkowski, também chamadas de normas L_p , se destacam pela versatilidade de aplicações. Elas consistem em um conjunto de funções de distâncias, entre elas a distância Euclidiana, onde cada variação é definida a partir da escolha do parâmetro $p \in \mathbb{R}$ utilizado. A função Minkowski generalizada é apresentada na Equação 2.4, onde $\mathbb{S} = \mathbb{R}$.

$$L_p[s_i, s_j] = \sqrt[p]{\sum_{l=1}^d |s_{il} - s_{jl}|^p} \quad (2.4)$$

Quando p na Equação 2.4 assume o valor 1 a função recebe o nome de *City-Block* ou *Manhattan*. Já no caso de $p = 2$ obtém-se a distância Euclidiana. Quando o valor de p tende a infinito, a equação obtida recebe o nome de *Chebyshev* e é aproximada pela Equação 2.5. A Figura 11 apresenta a distância resultante dependendo do parâmetro p utilizado, considerando um espaço bi-dimensional.

$$L_\infty = \max_{l=1}^d |s_{il} - s_{jl}| \quad (2.5)$$

Figura 11 – Exemplos de distâncias (a) *City-block*, (b) *Euclidiana*, (c) *Chebyshev* e (d) *Norma fracionada* com $p = 0,5$. O ponto preto indica o centro a partir do qual todos os elementos que estão na borda possuem mesma distância.

Fonte: Elaborada pelo autor.

Além dos valores inteiros para p tradicionalmente utilizados também é possível utilizar valores fracionados, tal que $0 < p < 1$ para calcular essa distância, inclusive sendo mais ou menos adequada dependendo do conjunto de dados. Em geral, o uso da norma fracionada resulta em uma distribuição de pontos com melhor contraste que os obtidos por valores inteiros de p , o que aumenta a qualidade das consultas inclusive para conjuntos com alta dimensionalidade (AGGARWAL; HINNEBURG; KEIM, 2001). Em processos que envolvam agrupamento, a norma fracionada tende a selecionar os elementos de forma mais adequada no espaço de busca e aumentar a distância entre máximo e mínimo, auxiliando a construção de *clusters* melhor definidos. Infelizmente, ao contrário das outras funções da família Minkowski, a norma fracionada não respeita todas as propriedades necessárias para ser uma métrica (propriedade *iii*).

As distâncias Minkowski são bastante utilizadas na literatura, particularmente devido a sua simplicidade e qualidade dos resultados obtidos, diversos algoritmos e estruturas de indexação fazem uso amplo da distância Euclidiana como base para consulta de dados. O uso dessas funções é indicado para conjuntos onde o vetor numérico analisado possua coordenadas independentes entre si. Além disso as distâncias da família Minkowski apresentam grande sensibilidade mesmo para pequenas variações em uma única dimensão, o que pode ser um problema dependendo do contexto em que estão sendo aplicadas. Por exemplo, no caso de análise de imagens é usual extrair histogramas de cores para em seguida utilizar funções de distância no processo de recuperação por conteúdo, entretanto a distância Euclidiana pode beneficiar uma ou outra tonalidade de cor caso os histogramas não estejam normalizados (WANG; ZHANG; FENG, 2005). Além disso as funções da família Minkowski estão restritas à manipulação de valores reais, impedindo por exemplo, que seja utilizada para avaliar diferenças entre conjuntos de elementos.

2.4.2 Consultas em espaços métricos

Como diversas estratégias de busca e estruturas de indexação são baseadas no formalismo de espaço métrico, as distâncias métricas permitem que algoritmos sejam otimizados em relação ao uso de memória e *CPU* possam ser utilizados independentemente do contexto de aplicação

(RAFF; NICHOLAS, 2018).

Figura 12 – Exemplo de (a) consulta por abrangência, (b) aos k -vizinhos mais próximos ($k = 4$) e (c) diversidade aos k vizinhos mais próximos de acordo com (SANTOS *et al.*, 2013). A estrela representa o elemento de consulta s_q e os quadrados em verde consistem nos elementos retornados em cada uma das consultas enquanto ε representa o raio da consulta por abrangência.

Fonte: Elaborada pelo autor.

As consultas por similaridades podem ser divididas em dois tipos principais: as consultas por abrangência e as consultas aos k -vizinhos mais próximos (ZEZULA *et al.*, 2006). No primeiro caso a região de interesse está centrada em todos os objetos da base de dados que sejam dissimilares do objeto de consulta até um limiar máximo. É necessário destacar que por considerar apenas um limiar como limitador não é possível prever o total de elementos que será retornado, o que pode ser um problema dependendo da aplicação a qual essa consulta está sendo empregada. Além disso a distribuição dos elementos no espaço métrico e o valor do limiar utilizado pode resultar em consultas pouco relevantes como por exemplo, uma consulta em que a resposta obtida contenha quase todos os elementos da base. Uma definição mais formal desta consulta é apresentada na Definição 1.

Definição 1 (Consulta por abrangência). Seja um subconjunto de dados $S \in \mathbb{S}$ e uma função de distância δ . Uma consulta por abrangência com o elemento de consulta $s_q \in \mathbb{S}$ e com uma dissimilaridade máxima ξ pode ser explicitada como:

$$R_q(s_q, \xi_q) = \{s_i \mid s_i \in S, \delta(s_i, s_q) \leq \xi_q\}$$

Já no caso dos k -vizinhos mais próximos, o interesse está na recuperação dos k objetos mais próximos do objeto de consulta. Esta consulta é apresentada na Definição 2.

Definição 2 (Consulta aos vizinhos mais próximos). Seja um subconjunto de dados $S \in \mathbb{S}$ e uma função de distância δ . Uma consulta aos k -vizinhos mais próximos com o elemento de consulta $s_q \in \mathbb{S}$ e um número de objetos a ser retornados igual a k é dada por:

$$kNN(s_q, k) = A = \{s_j \mid s_j \in \mathbb{S}, \forall s_i \in S - A : \delta(s_j, s_q) \leq \delta(s_i, s_q)\}$$

A consulta por k -vizinhos mais próximos é muito utilizada devido a sua simplicidade e por possuir uma ótima performance para diferentes conjuntos de dados (ZHANG *et al.*, 2018).

Além disso, esta consulta resolve o problema de definir um ξ já que a quantidade k de elementos é mais intuitiva para o usuário. A consulta kNN lida, muitas vezes, com dados binários e de classes compostas de forma automática, o que faz com que essa abordagem não exija uma fase prévia de treinamento. Usualmente a classificação de um novo dado a partir deste tipo de consulta é realizada recuperando os k casos com menor distância em relação ao elemento de consulta e é utilizada uma estratégia de votação para que a classe majoritária retornada seja atribuída para o dado consultado. Semanticamente, dizemos que os objetos retornados são os k mais “similares” ao objeto consultado em nossa base de dados, considerando a medida de distância utilizada.

Embora as consultas citadas anteriormente sejam as mais utilizadas as mesmas possuem limitações que motivam o desenvolvimento de novas abordagens. Tanto a consulta por abrangência quanto a consulta aos k vizinhos mais próximos verificam apenas a relevância individual de cada elemento considerando o valor de sua distância em relação ao elemento central de consulta. Dessa forma é possível que diversos elementos bastante similares entre si estejam presentes na resposta obtida, fazendo com que o usuário tenda a perder interesse na consulta devido aos elementos retornados repetidamente (ZIEGLER; LAUSEN; SCHMIDT-THIEME, 2004). Neste contexto, outras consultas, como a consulta por diversidade, podem ser empregadas com o objetivo de aumentar a relevância das respostas e diminuir o tempo de análise do usuário. Este tipo de consulta é muito utilizado em sistemas de recomendação (GOGNA; MAJUMDAR, 2017) e recuperação de informação (XU *et al.*, 2018). Infelizmente essas abordagens possuem um custo computacional elevado já que a princípio exigem uma comparação entre todos os pares da base de dados. A Figura 12 mostra exemplos gráficos de três das consultas citadas.

Também existem outras consultas que podem considerar o relacionamento entre os elementos da base, tais como a consulta aos k vizinhos mais próximos reverso (*1reverse-kNN*) (TAO; PAPADIAS; LIAN, 2004). Nessa consulta, dado um conjunto de dados S e um ponto s_q , o kNN reverso retorna todos os pontos $s \in S$ que possuem s_q como um dos seus respectivos vizinhos mais próximos. Este tipo de consulta possui diversas aplicações para problemas geográficos (WU *et al.*, 2008) (TRAN; TANJAR; SAFAR, 2009) entretanto exige um custo computacional maior do que um kNN tradicional.

2.5 Considerações finais

Neste capítulo foram brevemente apresentados os principais conceitos tradicionais utilizados nesta pesquisa. Tais conceitos são a base do que se convencionou a chamar de descrição de imagens e *CBIR* canônico, que utiliza engenharia de características.

Os extratores de características constituem uma etapa essencial em sistemas de recuperação de imagens a partir de seu conteúdo por ser a ponte entre o entendimento e compreensão do usuário e o sistema computacional. Consequentemente, é necessário que os algoritmos empregados possuam a capacidade para identificar os aspectos visuais esperados como cores, texturas e

formas, resultando em técnicas otimizadas para cada domínio de aplicação.

O uso de redes convolucionais (CNNs) tem se apresentado uma alternativa extremamente poderosa para análise e classificação de imagens, permitindo uma maior generalização de suas aplicações e lidando com os diversos problemas descritos no processo de extração de características tradicional. Uma única CNN pode extrair características locais e globais simultaneamente e gerar a classificação esperada em segundos, poupando a necessidade de construir um extrator especializado e que na maioria das vezes extrai apenas características de baixo nível que não refletem a percepção de um humano sobre a imagem. Essa simplificação em termos de desenvolvimento é acompanhada de um processo de treinamento que tende a ser custoso em termos computacionais, exigindo hardwares mais potentes e um tempo de processamento muito maior.

As consultas por similaridade atuam como uma alternativa às abordagens tradicionais permitindo que dados até então inviáveis de serem analisados sejam manipulados seguindo o desejo / intenção dos usuários. A similaridade entre os objetos é verificada não por *matching* exato mas pelo uso de uma função de distância, preferencialmente métrica. Neste contexto foram descritos os principais conceitos necessários para a realização dessas consultas, os impactos do uso de cada abordagem e os desafios inerentes de consultas por similaridade que devem ser compreendidos para o desenvolvimento correto de aplicações. Considerando os conceitos apresentados é crucial identificar qual função de distância representa as imagens médicas analisadas neste trabalho em conjunto com os valores obtidos a partir dos extratores de características para que o resultado esperado pelo usuário de um *CBMIR* seja plenamente atendido.

ASPECTOS DA RECUPERAÇÃO DE IMAGENS MÉDICAS POR CONTEÚDO

3.1 Visão geral

Para análise de imagens médicas são utilizados os mesmos conceitos de um sistema *CBIR* tradicional, obviamente escolhendo os parâmetros perceptuais coerentemente (BUGATTI *et al.*, 2011). Os parâmetros perceptuais podem ser vistos como parâmetros que são perceptualmente consistentes com as expectativas de um dado usuário buscando caracterizar e quantificar essa similaridade perceptual. No caso médico podemos ilustrar essa similaridade como a visão que o especialista possui e as características que ele busca sobre o domínio de imagens que estão sendo avaliadas. Além disso, sistemas especializados em imagens médicas vêm sendo objeto de estudos não somente para auxílio de diagnósticos mas também para permitir o compartilhamento dos dados armazenados entre diferentes bases. Este conceito recebe o nome de *Radiomics* (GILLIES; KINAHAN; HRICAK, 2016). Ao contrário de sistemas *CAD* que resultam em uma única resposta a uma consulta realizada, em *Radiomics* é esperado que uma grande quantidade de dados sejam extraídas das imagens em questão e armazenados em bases de dados compartilhadas permitindo que outros sistemas possam utilizar essas informações para geração de hipóteses, testes ou ambos. Esse compartilhamento tem potencial para aumentar a acurácia de diagnósticos e identificação de padrões, além de criar um ambiente adequado para o desenvolvimento de futuras pesquisas.

O desenvolvimento de *Radiomics* apresenta diversos desafios sendo o principal deles o compartilhamento de dados. É necessário que os dados obtidos sejam padronizados, garantindo que outros sistemas possam manipular essas informações coerentemente. Isso exigiria um comprometimento de todas as instituições envolvidas para definir esses padrões e garantia de seu cumprimento, afinal um grande volume de dados desorganizados impede o desenvolvimento de qualquer aplicação. Além da padronização, outro grande desafio é manipular esse volume de

dados de forma eficiente, garantindo que a informação de milhares de exames anteriores sejam chegadas a cada nova consulta realizada. Embora sistemas de gerenciamento de bases de dados apresentem boas soluções, a comunicação entre sistemas descentralizados apresentam novas dificuldades que exigem o desenvolvimento de novas soluções para tornar seu uso eficiente.

3.2 Sistemas CBMIR

Existem diversos *CBMIRs* com foco em imagens médicas (LI *et al.*, 2018) e, apesar do avanço neste tipo de sistema, ainda existem também diversos desafios para unir esta tecnologia à prática clínica. Do ponto de vista médico, diversos estudos, como (DEPEURSINGE *et al.*, 2011), sugerem que a principal limitação é a falta de ferramentas interativas que permitam ao especialista testar e validar o que é realmente similar considerando a sua percepção. Dessa forma é essencial unir os conhecimentos de ciência de computação e da medicina / radiologia para que o resultado atingido seja realmente o esperado pelo especialista que fará uso deste sistema.

Grande parte dos desafios a serem superados estão sumarizados no trabalho clássico de Deserno (DESERNO; ANTANI; LONG, 2009), que sugere uma visão sistemática dos *gaps* na pesquisa dos sistemas de recuperação de imagens médicas a partir de seu conteúdo. Ele organiza os desafios em quatro grandes categorias: *gaps* de conteúdo, *gaps* de características, *gaps* de performance e *gaps* de usabilidade. Os *gaps* de conteúdo consistem na compreensão do usuário na utilização das imagens médicas e no contexto clínico para uso deste sistema. Os *gaps* de características incluem a escolha dos métodos de extração e representação do conteúdo das imagens além da compreensão de suas limitações. Os *gaps* de performance abordam o tempo de resposta e integração com os diversos sistemas de informações dos pacientes. Finalmente, os *gaps* de usabilidade incluem os tipos de consultas disponíveis e a capacidade do sistema em aprender a partir do *feedback* do usuário.

Considerando o quesito usabilidade, o sistema *Kundaha* (SANTOS *et al.*, 2018) traz uma proposta mais flexível ao propor uma arquitetura modular permitindo a variação dos parâmetros do descritor a serem utilizados no conjunto de imagens de mamografias, oferecendo mais de 70 combinações de extrator de características e funções de distâncias possíveis. As imagens utilizadas foram obtidas a partir da identificação de regiões de interesse dos exames originais destacando locais com massas malignas e benignas totalizando 2892 instâncias, um valor considerável devido à dificuldade existente em se obter e armazenar tais imagens. Outra característica interessante neste *CBMIR* é a presença de ferramentas de processamento de imagens, como o janelamento dos canais de cores, auxiliando os especialistas na visualização das imagens obtidas, como visto na Figura 13.

Embora essas características demonstrem o esforço em também superar os *gaps* de usabilidade do sistema, infelizmente o *Kundaha* peca no compartilhamento destes dados, já que todo o seu processamento é *offline*, fator não desejado para integrá-lo no conceito de *Radiomics*.

Figura 13 – Interface do Kundaha para realização de consultas por similaridade e diversidade.

Fonte: Santos *et al.* (2018).

Devido à grande complexidade existente nas diversas etapas dos sistemas *CBMIR* sempre acaba existindo um *trade-off* nos diversos *gaps* propostos por Deserno (DESERNO; ANTANI; LONG, 2009), pela necessidade de se focar em alguns aspectos para otimizar o resultado obtido. Espera-se que *Radiomics* proporcione cada vez mais um ambiente propício para compartilhamento de dados e consequentemente sistemas cada vez mais robustos.

Outro sistema que faz uso de imagens de tomografia, porém para recuperação de imagens pulmonares, foi proposto por Ashis em (DHARA *et al.*, 2017). Nesta abordagem são deixados os *gaps* de usabilidade para um segundo momento enquanto foca-se na otimização e acurácia das imagens recuperadas. Para tal baseia-se na sequência de imagens obtidas de cada paciente para extração de características tridimensionais dos nódulos pulmonares, empregando extratores de características globais de identificação de formas. Para as consultas de similaridade foram empregadas as funções da família Minkowski aplicadas em 891 nódulos presentes na base de imagens rotuladas em 5 diferentes classes. Dentre as diversas configurações avaliadas obteve-se uma acurácia de aproximadamente 80% como melhor resultado.

Considerando abordagens *deep learning* o trabalho de Esteva (ESTEVA *et al.*, 2017) discute a qualidade de classificação de imagens dermatológicas utilizando uma CNN que foi originalmente proposta para classificação genérica de imagens do desafio Imagenet. Para isso utilizou-se uma rede InceptionV3 e propôs uma taxonomia para agrupar lesões em grandes classes em uma espécie de "árvore" auxiliando no processo de classificação (Figura 14). Para o treinamento da rede foram utilizadas 129450 imagens distribuídas em mais de 2032 tipos de doenças. Uma das características mais interessantes deste estudo é que embora as redes treinadas não obtiveram 100% de precisão na classificação dessas imagens, as mesmas obtiveram acurácia

Figura 14 – Taxonomia empregada para classificação de imagens dermatológicas em (ESTEVA *et al.*, 2017).

Fonte: Esteva *et al.* (2017).

superior em relação a classificação realizada por dermatologistas experientes no mesmo conjunto de imagens.

Considerando os trabalhos aqui apresentados e analisando outras publicações relevantes sobre sistemas médicos existentes, nota-se que esses sistemas geralmente focam na qualidade do resultado e na performance das consultas, como ocorre, por exemplo, nos sistemas *MammoSVD* (OLIVEIRA *et al.*, 2009), *FIRE* (Flexible Image Retrieval Engine) (DESELAERS; KEYSERS; NEY, 2005), para classificação de Alzheimer via recuperação de ressonâncias magnéticas (KRUTHIKA *et al.*, 2018), entre outros. Outros sistemas, como o *MedFMI-SiR* (KASTER *et al.*, 2011) apresentam capacidade para lidar com grandes bases de dados, aproximando dos conceitos do *Radiomics*.

3.3 Análise de imagens de úlceras venosas

As imagens foco deste estudo consistem em imagens dermatológicas de úlceras venosas de membros inferiores. O tratamento deste tipo de lesão é fortemente baseado no estágio em que a mesma se encontra sendo o diagnóstico pela observação da lesão fator predominante para a classificação do especialista. Esses estágios podem ser agrupados em 4 principais etapas: (i) fibrina; (ii) granulação; (iii) necrose; e (iv) saudável. Como cada estágio possui uma predominância de cor específica que faz com que a classificação dessas etapas possa ser feita utilizando um "modelo de 4 cores", sendo a fibrina caracterizada pela cor amarela, a granulação pela cor vermelha, necrose pela cor preta e saudável pela própria tom de pele da pessoa tratada (Figura 15).

Figura 15 – Exemplos de imagens de úlceras venosas.

Outra especificidade deste tipo de lesão é o fato da mesma não ser necessariamente uniforme, mas apresentar regiões internas em diferentes estágios de evolução. Por este motivo muitas vezes os especialistas optam por fazer uma segmentação manual colocando um plástico sobre a lesão e marcando com caneta específica, estimando assim a evolução da lesão desde a última consulta. Porém esse processo é extremamente moroso e por ser necessário se apoiar sobre a lesão do paciente torna-se doloroso para o mesmo. Devido ao eventual impacto clínico que uma abordagem automática traria existe um esforço global na solução desse problema que está sintetizado em diversos estudos que de segmentação de imagens dermatológicas, sendo os principais resultados até então apresentados na Tabela 1.

Tabela 1 – Resumo de estudos focados na segmentação de lesões dermatológicas.

Abordagem	Acurácia	Extrator	Classificador
Veredas <i>et al.</i> (2010)	84.84%	Texture Descriptors	MLP
Pereyra <i>et al.</i> (2014)	61.70%	Color Descriptors	IbL
Mukherjee <i>et al.</i> (2014)	87.61%	Color-based Features	SVM
Blanco <i>et al.</i> (2016)	86.81%	Color Structure	RandomForest
Kavitha <i>et al.</i> (2017)	87.05%	Spectral Histograms	MLP
Chino <i>et al.</i> (2018)	89.87%	Bag-of-Signatures	RandomForest

3.4 Considerações finais

Percebe-se que embora CBMIR inicialmente sejam apenas sistemas de recuperação de imagens aplicados em um contexto médico, existem uma série de considerações extras que devem ser analisadas durante o desenvolvimento e avaliação de novas abordagens. Os gaps apresentados por Deserno (DESERNO; ANTANI; LONG, 2009) atuam como guia das principais dificuldades enfrentadas e quais cuidados devem ser considerados desde o início do desenvolvimento de tais sistemas médicos. Os problemas por ele relatados, principalmente em questão de usabilidade, seguem relevantes e destacam a necessidade de focar não somente na avaliação matemática dos resultados obtidos mas também em como tornar novas propostas desenvolvidas simples e viáveis de serem portadas para um contexto clínico.

O MÉTODO CL-MEASURE (COUNTING-LABELS SIMILARITY MEASURE)

4.1 Visão geral

Durante o primeiro ano de mestrado foi desenvolvida uma abordagem inicial para análise de imagens dermatológicas que resultou na publicação (BLANCO *et al.*, 2016). A motivação é que diversos extratores e funções de distância que foram apresentados nas seções anteriores são genéricos o suficiente para serem utilizados em uma grande variedade de aplicações. No entanto, para uso em contextos mais específicos como no caso de ambientes clínicos, existem diversas particularidades que devem ser consideradas durante a análise das imagens em questão. Por este motivo foi desenvolvida uma técnica para comparação de imagens médicas, mais especificamente imagens de úlceras venosas. A abordagem chamada de *Counting-Labels Similarity Measure* (*CL-Measure*) possui a vantagem de lidar com imagens segmentadas de forma automática e ser versátil o suficiente para permitir adaptação para outros domínios. Os detalhes desta abordagem assim como os resultados obtidos serão apresentados nas próximas seções.

4.2 A abordagem CL-Measure

A ideia principal do *CL-Measure* consiste na criação de sub-imagens baseadas em uma classificação prévia, calcular a distância entre elas e agregar as distâncias parciais obtidas a partir de uma função compatível. Para isso é necessário definir os parâmetros dessa função: Seja I_i^l uma região da imagem I_i rotulada por um $l \in \mathcal{L}$. Da mesma forma, seja I_j^l uma região pertencente a mesma classe l em uma segunda imagem I_j . Ambas as imagens I_i e I_j podem ser representadas a partir de um extrator ε como vetores de características x_l e y_l , respectivamente. Seja δ uma

métrica que compara qualquer par de vetores de características gerado a partir de ε . Então, a distância composta Δ entre as imagens I_i e I_j pode ser descrita como a composição ponderada das distâncias individuais $\delta(x_l, y_l)$ como apresentado na Equação 4.1, onde w_l é o peso atribuído para a respectiva classe do objeto.

$$\Delta(I_i, I_j) = \sum_l^{\mathcal{L}} w_l \cdot \delta(x_l, y_l) \quad (4.1)$$

Note que esta abordagem exige apenas que a função de distância interna seja capaz de comparar os vetores de características resultantes. Entretanto, dificilmente grupos de objetos pertencentes a mesma classe são representados como um único vetor de características já que embora similares cada um possui características e propriedades distintas. Dessa forma, é esperado que uma representação detalhada utilize um extrator de características que gere um conjunto de vetores de características, fato que impede o uso de funções de distância tradicionais como a distância euclidiana. Para superar este problema, o *CL-Measure* utiliza o conjunto de sub-imagens representadas a partir de *superpixels* e realiza o processo de comparação a partir de dois estágios: (i) Criação, representação e rotulação dos *superpixels*, e (ii) Avaliação de similaridade dos *superpixels*.

No primeiro estágio, as imagens comparadas são divididas em grupos de *superpixels*, os quais são representados a partir de um extrator ε e agrupados a partir da classificação supervisionada. Para qualquer par de imagens I_i e I_j , o estágio de criação de *superpixels* gera dois conjuntos de sub-imagens I_i^l e I_j^l onde um classificador C rotula qualquer *superpixel* S_k como $l, l \in \mathcal{L}$. No segundo estágio, o *CL-Measure* calcula a similaridade entre I_i e I_j estendendo a Equação 4.1 para suportar as sub-imagens I_q^l . Esta medida utiliza tipos específicos de funções de distância para comparar estes conjuntos de *superpixels* e para considerar apenas o conjunto de rótulos $\mathcal{L}'$ definidos pelo usuário, $\mathcal{L}' \subset \mathcal{L}$. Formalmente, seja D uma variação métrica (que atenda todos os três critérios de espaços métricos) que compare dois conjuntos de imagens (a partir do processo de extração de características). O *CL-Measure* pode ser definido pela Equação 4.2, onde w_l é calculado a partir da influência de cada rótulo sobre a imagem.

$$CL - Measure(I_i, I_j) = \sum_l^{\mathcal{L}'} w_l \cdot D(I_i^l, I_j^l) \quad (4.2)$$

A influência de cada classe pode ser definida como a razão entre o número de *pixels* de cada classe e o total de *pixels* das regiões segmentadas como expresso pela Equação 4.3.

$$w_l = \frac{|I_i^l|}{|\bigcup_{l \in \mathcal{L}'} I_i^l|} \quad (4.3)$$

Entretanto, ao adotar a Equação 4.3, o *CL-Measure* não cumpre todos os axiomas de metricidade, particularmente a simetria e a desigualdade triangular. Para tornar o *CL-Measure*

Figura 16 – Pipeline de execução do *CL-Measure* para comparação de duas imagens dermatológicas.

Fonte: Elaborada pelo autor.

métrico, w_l pode ser definido como a razão entre os *pixels* classificados de cada classe e o total de *pixels* rotulados, como definido na Equação 4.4.

$$w_l = \frac{|I_i^l| + |I_j^l|}{|\bigcup_{l \in \mathcal{L}'} I_i^l| + |\bigcup_{l \in \mathcal{L}'} I_j^l|} \quad (4.4)$$

A Figura 16 apresenta o fluxo de execução do *CL-Measure*. Primeiramente, as imagens comparadas são divididas em *superpixels* representados por um extrator ε . Cada vetor de características é então rotulado a partir de um classificador supervisionado a partir do conjunto de classes definidas previamente pelo usuário, onde vetores de classes não relevantes são descartados e os restantes são agrupados em sub-imagens. A partir das sub-imagens obtidas é utilizada uma função de distância compatível para avaliar as distâncias entre regiões pertencentes à mesma classe. Finalmente, o *CL-Measure* calcula a influência de cada classe de forma que a similaridade entre as duas imagens originais seja a distância ponderada acumulada entre todos os pares de sub-imagens.

4.3 Experimentos

A performance do *CL-Measure* foi avaliada a partir da realização das seguintes etapas:

1. Calcular o *F-Measure* para a classificação de *superpixels* considerando três extratores;
2. Calcular o tempo de execução para duas funções de distância adotando o extrator com o melhor *F-Measure*;

3. Realizar uma análise de qualidade do *CL-Measure* considerando o caso de estudo de úlceras dermatológicas, utilizando o extrator mais adequado e função de distância mais rápida;
4. Realizar um estudo de precisão e revocação ($P \times R$), comparando a abordagem tradicional *CBIR* baseada na extração global e a distância Euclidiana em relação à abordagem métrica e não métrica do *CL-Measure*.

Para a realização dos testes foi utilizada a base de imagens de úlceras *BASE_DERMA* obtidas em parceria com o Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto - Universidade de São Paulo (HC-FMRP). Ela é composta por 186 imagens obtidas a partir de pacientes com insuficiência venosas ou arteriais. Essas imagens podem ser segmentadas em quatro classes distintas como $\mathcal{L} = \{\text{granulação, fibrina, necrose, saudável}\}$. Entretanto, no contexto em questão as classes de interesse incluem apenas regiões lesionadas, ou seja $\mathcal{L}' = \mathcal{L} - \{\text{saudável}\}$. A partir deste conjunto original, foram selecionadas 15 imagens representativas as quais foram avaliadas e segmentadas por uma especialista. Ao final do processo, foram obtidos 26813 *superpixels*, manualmente rotulados como 20465 instâncias da classe saudável, 2629 instâncias de granulação, 3433 instâncias de fibrina e 286 instâncias de necrose. Portanto, as 15 imagens em questão foram utilizadas para treinamento do classificador supervisionado.

4.3.1 Configurações para o *CL-Measure*

Para determinar qual extrator é mais adequado para representar as características presentes nos *superpixels* de úlceras venosas, a *BASE_DERMA* foi testada considerando três diferentes extratores de cores: *Color Layout*, *Scalable Color* e *Color Structure*. A escolha destes extratores deve-se ao fato da otimização em termos de tempo de execução e uso de memória presentes nos extratores do padrão *MPEG-7*. A qualidade de cada extrator foi avaliada utilizando o classificador *Random Forest* (BREIMAN, 2001) utilizando o procedimento *10-fold cross validation* a partir das regiões manualmente rotuladas por especialistas. O *Random Forest* foi escolhido devido a sua acurácia e robustez considerando outros classificadores da literatura (BREIMAN, 2001). Para avaliar a performance da classificação foi calculado o *F-Measure* médio para cada uma das classes presentes. Os resultados obtidos são apresentados na Tabela 2.

Tabela 2 – Comparação de extratores para *superpixels* de úlceras venosas.

Extrator	F-Measure por classe			
	Média	Granulação	Fibrina	Necrose
Color Layout	0.840	0.759	0.118	0.040
Scalable Color	0.893	0.760	0.538	0.335
Color Structure	0.903	0.786	0.555	0.403

Fonte: Elaborada pelo autor.

O extrator *Color Structure* apresentou melhor performance em relação aos outros dois competidores considerando todas as classes e, conseqüentemente, para todo o conjunto testado. Por este motivo, o *Color Structure* foi definido como um dos parâmetros do descritor para os testes utilizando o *CL-Measure*. A seguir é necessário definir qual função de distância é a mais indicada para avaliar conjuntos de *superpixels* das imagens em questão. Para este teste foram utilizadas duas funções de distância capazes de comparar conjuntos de vetores de características: a distância *Earth-Mover's* e a distância de *Jaccard*. Cada distância foi avaliada em relação ao seu tempo de execução considerando 100 pares de imagens distintas comparadas utilizando o *CL-Measure*. Como em sistemas *CBIR* reais o número de elementos da base de dados tende a ser muito maior que 100, o cálculo dessas funções impacta diretamente na usabilidade do sistema e deve ser observada cuidadosamente. A Tabela 3 compara os resultados do uso da distância *Earth-Mover's* e *Jaccard* na Equação 4.2 considerando o tempo médio e desvio padrão. Pelo experimento nota-se que a distância de *Jaccard* é aproximadamente 1000 vezes mais rápida em relação ao seu concorrente. Por este motivo *Jaccard* foi utilizada como distância padrão para o cálculo do *CL-Measure*.

Tabela 3 – Comparação do tempo de execução das funções de distância.

Distância	Tempo de execução	
	Média	Desvio padrão
Earth-Mover	159.751 ms	± 21.403 ms
Jaccard	000.053 ms	$\pm$ 00.731 ms

Fonte: Elaborada pelo autor.

4.3.2 Avaliação do *CL-Measure*

Foi utilizado o extrator *Color Structure* e a distância de *Jaccard* como parâmetros do descritor do *CL-Measure* para comparação das imagens baseadas em *superpixels*. Para avaliar a performance desta proposta foram considerados três diferentes cenários. Primeiramente, cada imagem da *BASE_DERMA* foi representada a partir de um único vetor de características resultante do processo de extração global utilizando o *Color Structure* e comparados a partir da distância Euclidiana. Em outras palavras, o primeiro cenário consiste na abordagem tradicional de recuperação de imagens, onde é feita a extração sobre toda a imagem e é empregada uma função de distância amplamente utilizada. Este cenário será chamado de *MEDIDA_PADRAO*. Os outros dois cenários considerados consistem nas versões métrica e não métrica do *CL-Measure* para recuperação das imagens.

As diferentes abordagens foram avaliadas a partir da recuperação das 5 imagens mais próximas a partir de um elemento fixo, ou seja, uma consulta aos 5 vizinhos mais próximos. A Figura 17 mostra o resultado de uma consulta, onde o elemento de consulta é o primeiro à esquerda de cada linha. As imagens marcadas com um X vermelho são falsos positivos, em

Figura 17 – Resultado da consulta aos 5 vizinhos mais próximos considerando os cenários (a) *CL-Measure* não métrico, (b) *CL-Measure* métrico, e (c) *MEDIDA_PADRAO*.

Fonte: Elaborada pelo autor.

outras palavras, imagens que não são similares ao elemento de consulta. Este primeiro resultado exemplifica a capacidade de recuperação do *CL-Measure* em relação à abordagem tradicional, apresentando um bom resultado para o contexto de úlceras venosas.

Finalmente, avaliou-se as curvas de precisão e revocação para verificar a qualidade das consultas realizadas. Para obter essas curvas, foi executada uma consulta para cada um dos elementos da *BASE_DERMA* recuperando os k vizinhos mais próximos para $k = 1, \dots, 186$ e armazenando as médias dos valores de precisão e revocação. É esperado que as curvas permaneçam o mais próximo possível do topo do gráfico, ou seja, a precisão permaneça alta para maiores valores de revocação. As curvas de cada uma das 3 abordagens de similaridade avaliadas são apresentadas na Figura 18.

Figura 18 – Curvas de precisão e revocação para as medidas de similaridade *MEDIDA_PADRAO*, *CL-Measure* não métrico e *CL-Measure* métrico utilizando a *BASE_DERMA*.

Fonte: Elaborada pelo autor.

Pela Figura 18, podemos avaliar a qualidade de recuperação calculando a área abaixo de

cada uma das curvas. Considerando a *BASE_DERMA*, o *CL-Measure* não métrico apresentou melhor performance em relação aos adversários, sendo em média 5% mais preciso do que a sua versão métrica e 8% melhor do que *MEDIDA_PADRAO*. O *CL-Measure* não métrico também foi o mais preciso considerando qualquer valor de revocação. Além disso, o *CL-Measure* não métrico apresentou a maior precisão para pequenos valores de revocação, o que consiste no cenário mais importante em sistemas *CBIR*, já que é esperado que essas aplicações retornem poucas imagens porém que sejam relevantes. Dessa forma o *CL-Measure* não métrico utilizando como parâmetros perceptuais o extrator *Color Structure* e a distância de *Jaccard* apresentaram desempenho superior do que a abordagem tradicional.

4.4 Considerações finais

A proposta inicial *CL-Measure* calcula a distância entre *superpixels* rotulados que compõem a imagem original ponderando sobre a influência de cada rótulo a partir da proporção em que os mesmos aparecem na cena. Os experimentos realizados demonstraram a vantagem em utilizar o *CL-Measure* para análise e recuperação de imagens de úlceras venosas, indicando que a configuração testada consiste em uma melhor abordagem neste contexto em relação às comumente utilizadas. Mais especificamente, os experimentos determinaram: (i) a melhor configuração para o *CL-Measure* neste domínio, e (ii) o *CL-Measure* apresentou melhor acurácia para todos os valores de revocação.

É importante também destacar que o *CL-Measure* não está restrito ao contexto analisado, mas pode ser utilizado em qualquer domínio onde a segmentação através de *superpixels* e a rotulação dessas regiões sejam uma boa opção. Por este motivo, pretende-se futuramente analisar a utilização do *CL-Measure* em outros contextos genéricos e também em outros domínios médicos como no caso de vértebras lombares. Este resultado preliminar indica o potencial desta nova abordagem e permite utilizá-la como um guia para o desenvolvimento de técnicas futuras e consequentemente para que os resultados alcançados neste trabalho sejam utilizados em outros domínios de imagens.

O MÉTODO QTDU (QUALITY OF TISSUES FROM DERMATOLOGICAL ULCERS)

5.1 Visão geral

A partir dos resultados positivos obtidos pelo uso do *CL-Measure* discutiu-se melhorias que poderiam ser propostas para otimizar ainda mais a classificação deste tipo de imagens. Visto o crescimento do uso de técnicas de redes neurais para a classificação de imagens, como nos estudos (ESTEVA *et al.*, 2017) e (HOO-CHANG *et al.*, 2016), faz-se necessário refletir sobre o uso dessas novas técnicas em função das abordagens tradicionais empregadas no *CL-Measure*.

Neste contexto foi proposto o *Quality of Tissues from Dermatological Ulcers (QTDU)* que faz uso de redes convolucionais (*CNNs*) para rotulação dos superpixels com a vantagem de compor todo o processo de identificação de características e classificação, dispensando a necessidade de identificar qual o extrator de características mais adequado para o contexto em questão. Além disso é possível fazer uso de redes bem estabelecidas como a *Inception* (SZEGEDY *et al.*, 2015) e *Resnet* (HE *et al.*, 2016) utilizando técnicas de transferência de aprendizado e ajustes de parâmetros de cada arquitetura.

5.2 A abordagem QTDU

Estudos de imagens dermatológicas baseiam-se em conjuntos de dados reduzidos comumente com algumas centenas de imagens, exatamente o oposto esperado para o uso de *deep learning* que usualmente possuem dezenas de milhares de imagens de exemplo. Para contornar este problema, empregamos uma abordagem de "dividir para conquistar" utilizando os superpixels gerados ao invés das 186 imagens existentes na *BASE_DERMA*. Para tal pedimos que dermatologistas rotulassem os superpixels de 40 imagens do conjunto original utilizando o modelo de quatro cores $\mathcal{L} = \{\text{granulação, fibrina, necrose, saudável}\}$. Como resultado foram ob-

tidos 44893 superpixels distribuídos em: (i) 37187 regiões classificadas como saudável, (ii) 3974 superpixels classificados como fibrina, (iii) 3284 com granulação como classe predominante e (iv) 448 regiões com necrose.

5.2.1 Experimentos

A performance do *QTDU* foi avaliada a partir da realização das seguintes etapas:

1. Definir quais extratores da abordagem *machine learning* tradicional possuem melhor capacidade de representação para o conjunto de imagens em questão;
2. Definir melhor maneira de lidar com o conjunto de imagens reduzidas para o uso de CNNs no contexto de úlceras venosas;
3. Realizar uma análise de qualidade do *QTDU* considerando o tempo necessário para sua execução e a acurácia obtida;
4. Visualizar a capacidade de segmentação e representação das classificações obtidas em relação a análise realizada por um humano.

5.2.2 Configurações para o QTDU

Características foram extraídas dos superpixels obtidos a partir dos extratores *Color layout*, *Color Structure* e *Scalable Color*, que geram representações multidimensionais de 12, 128 e 256 dimensões, respectivamente. Para avaliar o processo de extração, aplicou-se a técnica de visualização *T-Distributed Neighbor Embedding (T-SNE)* (MAATEN; HINTON, 2008) permitindo avaliar o relacionamento entre cada classe e suas respectivas representações. A Figura 19 (a) mostra a visualização *T-SNE* dos vetores do *Color-Structure* após amostragem do conjunto original de 44893 superpixels em 285 instâncias por classe e utilizando a parametrização de 1000 iterações e paciência de 200 ciclos.

Figura 19 – Análise visual da *BASE_DERMA*. (a) Visualização *T-SNE* das características do *Color-structure*. (b) *PCA Scree plot*. (c) Variância individual do *PCA*.

Fonte: Elaborada pelo autor.

Devido à alta dimensionalidade das representações do *Color Structure* e *Scalable-color* e dos problemas (já conhecidos) que isso pode acarretar optou-se por aplicar uma redução de dimensionalidade antes do processo de classificação. Para tal utilizou-se o *Principal Components Analysis (PCA)* e avaliou-se a contribuição de cada dimensão considerando suas variâncias. A Figura 19 (b) e (c) mostram o comportamento das variâncias cumulativas e individual para o *Color-structure*, respectivamente. Foram empregados dois critérios distintos para o processo de redução: (i) *Kaiser-Guttman (KG)*, o qual seleciona componentes os quais possuem variância individual maior que 1, e (ii) *Scree-Plot*, o qual seleciona componentes que juntos possuem variância acumulada de 80% em relação a dimensionalidade total. O critério de KG indicou 31 componentes do total de 128 para o *Color Structure*, enquanto o *Scree-Plot* selecionou 12 componentes que representam 80.2% do total de variância. De forma análoga, o critério KG selecionou 18 dimensões do *Scalable Color* relevantes para a análise, enquanto o *Scree-Plot* definiu 6 componentes que representam 80.4% do total de variância.

A partir das reduções descritas foram obtidas 7 diferentes representações para o mesmo conjunto que foram nomeados como *Color layout*, *Color Structure*, *Color Structure - KG*, *Color Structure - SP*, *Scalable Color*, *Scalable Color - KG*, *Scalable Color - SP*. Para o processo de classificação empregou-se os classificadores *Random Tree*, *Naive-Bayes*, *Bayes-Net*, *Random Forest*, *SVM*, *MLP* e *IBL* considerando tanto a norma L_1 quanto a L_2 .

Figura 20 – Análise dos extratores do *MPEG – 7* e classificadores considerando *ROC – AUC* e coeficiente *KAPPA*.

	RandomForest	Random Tree	SVM	MLP	Naïve-Bayes	Bayes-Net	IbL- L_2	IbL- L_1
Color Layout	.902-.448	.698-.374	.646-.331	.879-.427	.822-.336	.853-.376	.696-.376	.397-.376
Color Structure	.952-.665	.791-.548	.675-.410	.880-.563	.815-.149	.875-.438	.815-.592	.812-.583
Color Structure - KG	.947-.641	.790-.542	.630-.331	.921-.568	.704-.140	.872-.351	.806-.577	.809-.582
Color Structure - SP	.940-.604	.777-.516	.604-.262	.900-.531	.706-.133	.850-.333	.794-.551	.794-.549
Scalable Color	.947-.606	.760-.485	.601-.207	.815-.381	.502-.000	.619-.058	.827-.613	.796-.551
Scalable Color - KG	.955-.645	.782-.534	.600-.200	.806-.330	.497-.000	.650-.000	.807-.569	.813-.584
Scalable Color - SP	.906-.507	.715-.414	.600-.200	.807-.310	.495-.000	.774-.243	.740-.439	.746-.452

Fonte: Elaborada pelo autor.

O conjunto de classificadores e representações (56 combinações) foram avaliados utilizando uma validação cruzada de 10 – *fold*. A Figura 20 apresenta os resultados considerando a área sob a curva *ROC-AUC* e o coeficiente *KAPPA*. Os valores em negrito representam o melhor resultado de cada coluna, ou seja, qual classificador foi mais preciso para cada representação. Pelos resultados, nota-se que as configurações utilizando *Random Forest* apresentaram médias superiores em relação às outras abordagens. Combinando este classificador com os extratores *Color Structure* e *Scalable Color* foram obtidos valores *ROC-AUC* de 8% e 5% e coeficientes *KAPPA* de 6% e 3%, respectivamente, superiores aos outros concorrentes. Optamos então por eleger essas duas configurações como representantes da abordagem tradicional visto que obtiveram valores superiores a qualquer outra configuração.

Figura 21 – Arquitetura geral do QTDU. Na fase de aprendizado, a imagem da úlcera é dividida em um conjunto de superpixels para ajustar a CNN e suas seis novas camadas finais.

Fonte: Elaborada pelo autor.

Nesta nova abordagem "ignoramos" o processo de extração de características confiando a tarefa de identificar qual representação multidimensional é mais apropriada a uma CNN. Para tal, o QTDU baseia-se nos resultados obtidos dos modelos de *deep learning* propostos no *ImageNet* e nos resultados obtidos em trabalhos anteriores (ESTEVA *et al.*, 2017) para definir as CNNs candidatas, no caso as redes *ResNet* (HE *et al.*, 2016) e *Inception V3* (SZEGEDY *et al.*, 2015). Além disso, utilizou-se como base os mesmos parâmetros aplicados nos trabalhos em questão, ou seja, *learning rate* de 0,001, *momentum* de 0,88, *loss function* como *categorical cross entropy* e *batch size* de 24.

Conhecendo as limitações do conjunto também preparou-se para evitar o *overfitting* das redes, sabendo que tanto a *Resnet* quanto a *Inception V3* possuem milhões de parâmetros. Alternativas viáveis para tal incluem o uso de regularizações e a inclusão de diversas camadas de *dropout*, o que potencialmente alteraria significativamente a topologia das redes originais. Entretanto, optou-se por manter as arquiteturas mais próximas possíveis das originais para uma comparação mais justa em relação aos extratores já descritos que também não passaram por qualquer alteração. Dessa forma o QTDU faz uso de duas técnicas: (i) *data augmentation* para aumentar a quantidade de superpixels utilizados durante o treinamento da rede, e (ii) a adição de 6 novas camadas na saída de cada uma das redes mantendo os pesos da rede previamente treinada (*transfer learning*).

Para o *data augmentation* foram aplicadas uma série de transformações nos superpixels

originais: (i) rotação de 90° , (ii) fator de escala de 0,2 para mais e para menos, e (iii) espelhamento horizontal e vertical. Essas escolhas permitem a generalização a partir da simetria das lesões (rotação) e da região atingida (escala). Como resultado, foram obtidas 179572 instâncias dos 4 rótulos previamente descritos. Para o transfer learning, manteve-se as arquiteturas originais tanto da Resnet quanto inception e foram adicionadas novas camadas compostas por 3 camadas totalmente conectadas (*dense*) separadas por duas camadas de *dropout* e uma camada final *softmax* para classificação final. Como as redes originais tiveram a arquitetura inalteradas utilizamos os pesos dessas redes previamente treinadas para o Imagenet e as camadas finais inicializadas com pesos aleatórios.

A Figura 21 ilustra o *pipeline* baseado em três estágios da nossa proposta. A fase de treinamento do QTDU consiste no treinamento da CNN a partir dos parâmetros originais do ImageNet e variáveis aleatórias (novas camadas) que são treinadas de acordo com o conteúdo dos superpixels apresentados.

5.2.3 Avaliação do QTDU

QTDU foi testado em uma estação com sistema operacional UBUNTU 16.04.4 LTS executando um cluster local com 2 GPUs com 2560 núcleos de 1607 MHz cada, 64 GB de memória RAM. Estes testes foram construídos para identificar (i) qual melhor configuração para o QTDU, e (ii) quantificar os ganhos ou perda do QTDU em relação às abordagens *Machine learning* empregadas anteriormente. Com o objetivo rotular essas abordagens anteriores a serem comparadas com o QTDU *deep learning* passaremos a chamá-las de abordagens *machine learning*, embora essencialmente o uso de *deep learning* também consiste na aplicação de técnicas de aprendizado de máquina. Dessa forma, selecionou-se as duas melhores abordagens anteriores e comparou-se com as duas CNNs testadas no QTDU, Inception V3 e ResNet. Todas as comparações foram baseadas nos resultados obtidos via validação cruzada com 10 – *fold*.

5.2.3.1 Métricas de tempo

Mediu-se o tempo gasto para treinar a *Inception V3* e *ResNet* como CNNs para o QTDU. De forma análoga, mediu-se também o tempo necessário para treinamento da *Random Forest* tanto com o extrator *Color Structure* quanto ao *Scalable Color - KG*. A Figura 22 apresenta a média e o desvio padrão dos tempos obtidos. Nota-se que o *Random Forest* em conjunto com o extrator *Scalable Color - KG* é 1,78 ordem de magnitude mais rápido do que a CNN mais rápida (*Inception V3*).

5.2.3.2 Métricas de acurácia

A Tabela 4 mostra a comparação geral da classificação média de todas as classes entre os métodos de classificação baseados em *machine learning* e QTDU com diferentes parametrizações considerando o coeficiente KAPPA (Kap.), F-Measure (F1), Sensividade (Sen.), Especificidade

Figura 22 – Média e desvio padrão do tempo necessário para treinamento da (i) RandomForest com Color Structure and Scalable Color – KG e (ii) QTDU com Inception V3 and ResNet.

Fonte: Elaborada pelo autor.

(Spe.) e área sob a curva (ROC-AUC). QTDU obteve melhores resultados em relação à abordagem anterior considerando qualquer cenário e métrica. Em particular, QTDU com *Inception V3* superou o *Color Structure* com *Random Forest* em aproximadamente 6,4% e 7,6% considerando sensibilidade e especificidade, respectivamente.

Tabela 4 – Comparação das abordagens machine learning e QTDU.

	Kap.	F1	Sen.	Esp.	AUC
RF / Color Structure	.665	.901	.910	.902	.952
RF / Scalable Color – KG	.645	.897	.904	.896	.955
QTDU – Inception V3	.716	.969	.968	.971	.986
QTDU – ResNet	.721	.971	.970	.974	.986

Fonte: Elaborada pelo autor.

De forma análoga, QTDU com ResNet superou o *Scalable Color - KG* com *Random Forest* em 7,3% e 8,7% considerando sensibilidade e especificidade, respectivamente. QTDU também superou ambas as abordagens baseadas em *machine learning* em 3,6% considerando AUC e obteve elevados valores do coeficiente KAPPA e *F-Measure* em todas as comparações.

Em um primeiro momento, *Inception V3* e *ResNet* obtiveram resultados similares para as métricas avaliadas na Tabela 4. Inclusive, ambas as CNNs atingiram virtualmente os mesmos valores de AUC, sendo que a ResNet obteve resultados sensivelmente melhores considerando o coeficiente KAPPA e *F-Measure*. Porém como a média entre as classes pode levar a um valor de acurácia ruim para a classe minoritária, principalmente em conjuntos altamente desbalanceados (nosso contexto), para fins de validação foram geradas as matrizes de confusão para as duas CNNs competidoras do QTDU detalhando assim suas performances para cada uma das classes de lesão. Para fins de comparação, também foram geradas as matrizes para ambos os modelos baseados em *machine learning*.

A Figura 23 mostra uma análise mais detalhada entre as abordagens discutidas destacando as diferenças entre as mesmas. Em particular, a Figura 23 (a-b) mostra o *trade-off* entre o *Color*

Figura 23 – F-Measures para todas as classes de lesões (maior o valor, melhor). (a) Color Structure and RandomForest. (b) Scalable Color – KG and RandomForest. (c) ResNet QTDU. (d) Inception V3 QTDU.

Fonte: Elaborada pelo autor.

Structure (mais efetivo para as classes saudável e fibrina) e *Scalable Color - KG* (mais efetivo para as classes granulação e necrose). De forma similar, a Figura 23(c-d) mostra que o QTDU é mais efetivo do que as abordagens *machine learning*, considerando qualquer uma das classes. Os mapas de calor (quanto mais quente - próximo ao vermelho - melhor) também revelaram que a ResNet é mais efetiva que a Inception V3 para a classificação de regiões sem lesões, fibrina e necrose, enquanto a Inception V3 atua melhor na classificação de regiões com granulação. Tais resultados indicam que a Resnet não é apenas sensivelmente melhor nas métricas da Tabela 4, mas também proporciona uma melhor separação entre as quatro classes de superpixels. Por este motivo, a ResNet foi escolhida para a realização do próximo conjunto de experimentos.

5.2.3.3 Quantificação de áreas lesionadas a partir do QTDU

Cinco imagens não utilizadas para as análises anteriores foram separadas da *BASE_DERMA* e segmentadas manualmente a partir de dois especialistas a partir da regra das 4 classes. Foram selecionadas imagens com todas as diferentes combinações de classes possíveis, isto é, granulação e fibrina, fibrina e necrose e granulação, fibrina e necrose. A Figura 24(a-b) apresenta um exemplo de uma imagem escolhida e sua segmentação manual realizada por um especialista.

QTDU foi aplicado para as imagens selecionadas realizando automaticamente o processo de segmentação (Figura 24(c)). Os pixels segmentados foram então comparados à classificação realizada pelos especialistas. A acurácia média do QTDU foi de 93,56%, que é 4,2% maior do que a melhor acurácia obtida em estudos anteriores baseados em *machine learning*. O erro médio absoluto (MAE) também foi calculado para as imagens segmentados via contagem do total de pixels classificados errados dividido pelo total de pixels das regiões lesionadas. A média de MAE para as imagens segmentadas foi de .089, e sua variância de 0.015. Estes resultados mostram a capacidade do QTDU em segmentar úlceras dermatológicas por uma abordagem baseada em superpixels e *deep learning* de forma correta e eficaz.

Figura 24 – Exemplo de (a) imagem original, (b) segmentação manual, e (c) quantificação da área via QTDU.

Fonte: Elaborada pelo autor.

5.3 Considerações finais

No caso do QTDU, foi obtido um avanço expressivo na classificações das áreas lesionadas mostrando mais uma vez o impacto positivo no uso de abordagens baseadas em *deep learning* em relação às técnicas tradicionais. Esta abordagem faz uso de *transfer learning* e da adição de algumas camadas finais nas arquiteturas de redes amplamente conhecidas (Inception V3 e ResNet) para otimizar o processo de classificação dos superpixels obtidos no conjunto de imagens de úlceras venosas.

A análise empírica sobre os 179572 superpixels divididos em 4 classes indicam que o QTDU é a abordagem mais eficaz neste contexto de classificação (médias de $AUC=0.986$, $sensibilidade=0.97$, $especificidade=0.974$) superando todas as abordagens anteriores baseadas em *machine learning* em 11,7% e 8,2% considerando o coeficiente KAPPA e *F-Measure*, respectivamente. Os experimentos também mostraram que a ResNet apresentou melhor acurácia do que a Inception V3 para ser utilizada junto com o QTDU, sendo que tal combinação foi responsável por uma acurácia média de 93,56%, que é 4,2% maior do que a apresentada que a abordagem atual com maior acurácia (89,87%) (CHINO *et al.*, 2018). Finalmente, uma análise com 5 imagens retiradas de um contexto clínico real indicam que o QTDU classificou corretamente as áreas de cada lesão com um erro absoluto médio de 0.089.

CONCLUSÕES

6.1 Considerações finais

Essa dissertação de mestrado visou contribuir com algoritmos e técnicas para classificação de imagens médicas considerando as restrições existentes em conjuntos reduzidos, fato recorrente na análise de imagens dermatológicas. As novas abordagens apresentadas avançaram o estado da arte na classificação de imagens de úlceras venosas obtendo resultados superiores a todos os trabalhos publicados que fizeram uso deste mesmo dataset.

Devido a estrutura modular da técnica pode-se alterar os parâmetros perceptuais empregados de forma simples e direta. Dessa forma é possível aplicar as abordagens apresentadas em outros sistemas *CBIR* que realizam consultas por similaridade permitindo assim sua avaliação em novos contextos. Os códigos desenvolvidos também serão disponibilizados facilitando a comparação com novas abordagens e sugestão de futuras melhorias.

6.2 Contribuições

Este projeto apresentou duas contribuições principais. A primeira delas consistiu na abordagem *Counting-label Similarity Measure* (CL-Measure) para classificação de imagens de úlceras venosas. Para tal o CL-Measure faz uso de uma arquitetura modular que organiza o processo de classificação em 4 etapas e propõe uma função de distância para a ponderação entre as diferentes classes presentes nas imagens e classificação final. Os experimentos demonstraram a melhora na acurácia em relação a todos os trabalhos anteriores até o momento de sua publicação.

A segunda contribuição deste trabalho foi a abordagem *Quality of Tissues from Dermatological ulcers* (QTDU) que expande os testes inicialmente realizados para o CL-Measure observando a qualidade de cada extrator via algoritmos de visualização, a redução de dimensionalidade por meio de diferentes conceitos e o uso de abordagens *deep learning* que são estado

da arte no momento. Os experimentos realizados demonstraram uma taxa de acurácia estado da arte para este conjunto de imagens provando ser uma forma eficaz para identificar regiões lesionadas nas imagens com potencial para ser utilizada como parte da rotina clínica neste tipo de tratamento.

Durante o desenvolvimento desta pesquisa foram produzidos os seguintes artigos:

- Blanco, Gustavo; Bedo, Marcos; Cazzolato, Mirela; Santos, Lucio; Jorge, Ana Elisa; Traina, Caetano; Azevedo-Marques, Paulo; Traina, Agma JM - "A label-scaled similarity measure for content-based image retrieval."(ISM 2016).
- Santos, Lucio; Blanco, Gustavo; Oliveira, Daniel; Traina, Agma JM; Traina Jr, Caetano; Bedo, Marcos VN - "Exploring Diversified Similarity with Kundaha."(CIKM 2018).
- Blanco, Gustavo; Jorge, Ana Elisa; Azevedo-Marques, Paulo; Oliveira, Daniel; Bedo, Marcos VN; Traina Jr, Caetano; Traina, Agma JM - "A superpixel-driven deep learning approach for the analysis of dermatological wounds.". (em processo de revisão)

6.3 Trabalhos futuros

Há diversas formas de explorar futuramente as propostas aqui apresentadas. Entre elas pode-se citar:

- Adaptação mais profunda das CNNs empregadas otimizando sua arquitetura para identificar qual seria o máximo de acurácia que poderia ser obtida para imagens de úlceras venosas.
- Analisar o impacto de usar uma rede neural para o processo de segmentação ao invés da abordagem SLIC.
- Construção de uma ferramenta para dermatologistas que apresente o resultado da classificação juntamente com a imagem original segmentada por regiões.
- Uso do CL-Measure e do QTDU em conjuntos genéricos de imagem para avaliar a capacidade de generalização em relação a outras abordagens já existentes.
- Propor formas de *data augmentation* que poderiam ser aplicadas especificamente para imagens dermatológicas viabilizando outras abordagens que exigem um conjunto maior de imagens.

REFERÊNCIAS

- ACHANTA, R.; SHAJI, A.; SMITH, K.; LUCCHI, A.; FUA, P.; SÜSSTRUNK, S. Slic superpixels compared to state-of-the-art superpixel methods. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 34, n. 11, p. 2274–2282, 2012. Citado nas páginas 32 e 33.
- AGGARWAL, C. C.; HINNEBURG, A.; KEIM, D. A. On the surprising behavior of distance metrics in high dimensional space. In: _____. **Database Theory — ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 420–434. ISBN 978-3-540-44503-6. Disponível em: <http://dx.doi.org/10.1007/3-540-44503-X_27>. Citado na página 39.
- BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. **Computer vision—ECCV 2006**, Springer, p. 404–417, 2006. Citado na página 31.
- BLANCO, G.; BEDO, M. V. N.; CAZZOLATO, M. T.; SANTOS, L. F. D.; JORGE, A. E. S.; TRAINA, C.; AZEVEDO-MARQUES, P. M.; TRAINA, A. J. M. A label-scaled similarity measure for content-based image retrieval. In: **2016 IEEE International Symposium on Multimedia (ISM)**. [S.l.: s.n.], 2016. p. 20–25. Citado na página 49.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. ISSN 1573-0565. Disponível em: <<http://dx.doi.org/10.1023/A:1010933404324>>. Citado na página 52.
- BUGATTI, P. H.; KASTER, D. S.; PONCIANO-SILVA, M.; TRAINA, A. J.; TRAINA, C. Integrating user profile in medical cbir systems to answer perceptual similarity queries. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging 2011: Computer-Aided Diagnosis**. [S.l.], 2011. v. 7963, p. 79630Q. Citado nas páginas 26 e 43.
- CHANG, S.-K.; KUNIL, T. Pictorial data-base systems. **Computer**, IEEE, v. 14, n. 11, p. 13–21, 1981. Citado na página 25.
- CHINO, D. Y. T.; SCABORA, L. C.; CAZZOLATO, M. T.; JORGE, A. E. S.; TRAINA, C.; TRAINA, A. J. M. Icarus: Retrieving skin ulcer images through bag-of-signatures. In: IEEE. **2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2018. p. 82–87. Citado nas páginas 21, 23 e 64.
- DEPEURSINGE, A.; FISCHER, B.; MÜLLER, H.; DESERNO, T. M. Prototypes for content-based image retrieval in clinical practice. **The open medical informatics journal**, v. 5, n. 1, 2011. Citado na página 44.
- DESELAERS, T.; KEYSERS, D.; NEY, H. Fire – flexible image retrieval engine: Imageclef 2004 evaluation. In: _____. **Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004, Bath, UK, September 15-17, 2004, Revised Selected Papers**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 688–698. ISBN 978-3-540-32051-7. Disponível em: <http://dx.doi.org/10.1007/11519645_67>. Citado na página 46.

- DESERNO, T. M.; ANTANI, S.; LONG, R. Ontology of gaps in content-based image retrieval. **Journal of Digital Imaging**, v. 22, n. 2, p. 202–215, 2009. ISSN 1618-727X. Disponível em: <http://dx.doi.org/10.1007/s10278-007-9092-x>. Citado nas páginas 44, 45 e 47.
- DHARA, A. K.; MUKHOPADHYAY, S.; DUTTA, A.; GARG, M.; KHANDELWAL, N. Content-based image retrieval system for pulmonary nodules: Assisting radiologists in self-learning and diagnosis of lung cancer. **Journal of Digital Imaging**, v. 30, n. 1, p. 63–77, Feb 2017. ISSN 1618-727X. Disponível em: <https://doi.org/10.1007/s10278-016-9904-y>. Citado na página 45.
- ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, Nature Publishing Group, v. 542, n. 7639, p. 115, 2017. Citado nas páginas 15, 26, 45, 46, 57 e 60.
- FELZENSZVALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. **International journal of computer vision**, Springer, v. 59, n. 2, p. 167–181, 2004. Citado na página 33.
- GILLIES, R. J.; KINAHAN, P. E.; HRICAK, H. Radiomics. **Radiology**, Radiological Society of North America Inc., v. 278, n. 2, p. 563–577, 2016. Citado na página 43.
- GOGNA, A.; MAJUMDAR, A. Diablo: Optimization based design for improving diversity in recommender system. **Information Sciences**, Elsevier, v. 378, p. 59–74, 2017. Citado na página 41.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado nas páginas 33 e 34.
- HARE, J. S.; LEWIS, P. H.; ENSER, P. G.; SANDOM, C. J. Mind the gap: Another look at the problem of the semantic gap in image retrieval. SPIE and IS&T, 2006. Citado na página 26.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778. Citado nas páginas 36, 37, 57 e 60.
- HOO-CHANG, S.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. **IEEE transactions on medical imaging**, NIH Public Access, v. 35, n. 5, p. 1285, 2016. Citado na página 57.
- ITSEEZ. **The OpenCV Reference Manual**. 2.4.9.0. ed. [S.l.], 2014. Citado na página 31.
- JENITTA, A.; RAVINDRAN, R. S. Image retrieval based on local mesh vector co-occurrence pattern for medical diagnosis from mri brain images. **Journal of medical systems**, Springer, v. 41, n. 10, p. 157, 2017. Citado na página 22.
- KASTER, D. S.; BUGATTI, P. H.; PONCIANO-SILVA, M.; TRAINA, A. J. M.; MARQUES, P. M. A.; SANTOS, A. C.; TRAINA, C. Medfmi-sir: A powerful dbms solution for large-scale medical image retrieval. In: _____. **Information Technology in Bio- and Medical Informatics: Second International Conference, ITBAM 2011, Toulouse, France, August 31 - September 1, 2011. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 16–30. ISBN 978-3-642-23208-4. Disponível em: http://dx.doi.org/10.1007/978-3-642-23208-4_2. Citado na página 46.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105. Citado na página 34.

KRUTHIKA, K.; MAHESHAPPA, H.; INITIATIVE, A. D. N. *et al.* Cbir system using capsule networks and 3d cnn for alzheimer's disease diagnosis. **Informatics in Medicine Unlocked**, Elsevier, 2018. Citado na página 46.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 34.

LI, Z.; ZHANG, X.; MÜLLER, H.; ZHANG, S. Large-scale retrieval for medical image analytics: A comprehensive review. **Medical image analysis**, Elsevier, v. 43, p. 66–84, 2018. Citado na página 44.

LING, C. S.; ZAKI, W. M. D. W.; HUSSAIN, A.; AHMAD, W. S. H. M. W.; HING, E. Y. Shape based image retrieval system for mri spine. In: **2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)**. [S.l.: s.n.], 2017. p. 1–6. ISSN 2155-6830. Citado na página 21.

LIPPMANN, R. P. Pattern classification using neural networks. **IEEE communications magazine**, IEEE, v. 27, n. 11, p. 47–50, 1989. Citado na página 28.

LIU, Y.; PENG, Y.; LI, D.; FAN, J.; LI, Y. Crime scene investigation image retrieval with fusion cnn features based on transfer learning. In: **Proceedings of the 3rd International Conference on Multimedia and Image Processing**. New York, NY, USA: ACM, 2018. (ICMIP 2018), p. 68–72. ISBN 978-1-4503-6468-3. Disponível em: <<http://doi.acm.org/10.1145/3195588.3195605>>. Citado na página 26.

LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. **Computer vision, 1999. The proceedings of the seventh IEEE international conference on**. [S.l.], 1999. v. 2, p. 1150–1157. Citado na página 30.

MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. **Journal of machine learning research**, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 58.

MISHRA, N. K.; CELEBI, M. E. An overview of melanoma detection in dermoscopy images using image processing and machine learning. **arXiv preprint arXiv:1601.07843**, 2016. Citado na página 22.

MURAMATSU SHUNICHI HIGUCHI, T. M. M. O. H. F. C. **Similarity estimation for reference image retrieval in mammograms using convolutional neural network**. 2018. 10575 - 10575 - 6 p. Disponível em: <<https://doi.org/10.1117/12.2293979>>. Citado na página 21.

NIXON, M. S.; AGUADO, A. S. **Feature extraction & image processing for computer vision**. [S.l.]: Academic Press, 2012. Citado na página 32.

NOVAK, C. L.; SHAFER, S. A. Anatomy of a color histogram. In: IEEE. **Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on**. [S.l.], 1992. p. 599–605. Citado na página 28.

OLIVEIRA, J. E. E. de; LOPES, A. P. B.; CAMARA-CHAVEZ, G.; ARAUJO, A. de A.; DESERNO, T. M. Mammosvd: A content-based image retrieval system using a reference database of mammographies. In: **2009 22nd IEEE International Symposium on Computer-Based Medical Systems**. [S.l.: s.n.], 2009. p. 1–4. ISSN 1063-7125. Citado na página 46.

RAFF, E.; NICHOLAS, C. Toward metric indexes for incremental insertion and querying. **arXiv preprint arXiv:1801.05055**, 2018. Citado na página 40.

RAMSUNDAR, B.; ZADEH, R. B. **TensorFlow for Deep Learning: From Linear Regression to Reinforcement Learning**. 1st. ed. [S.l.]: O'Reilly Media, Inc., 2018. ISBN 1491980451, 9781491980453. Citado nas páginas 35 e 37.

REN, X.; MALIK, J. Learning a classification model for segmentation. In: IEEE. **null**. [S.l.], 2003. p. 10. Citado na página 33.

SANTOS, L. F.; BLANCO, G.; OLIVEIRA, D. d.; TRAINA, A. J.; JR, C. T.; BEDO, M. V. Exploring diversified similarity with kundaha. In: ACM. **Proceedings of the 27th ACM International Conference on Information and Knowledge Management**. [S.l.], 2018. p. 1903–1906. Citado nas páginas 22, 44 e 45.

SANTOS, L. F. D.; OLIVEIRA, W. D.; FERREIRA, M. R. P.; TRAINA, A. J. M.; TRAINA JR., C. Parameter-free and domain-independent similarity search with diversity. In: **Proceedings of the 25th International Conference on Scientific and Statistical Database Management**. New York, NY, USA: ACM, 2013. (SSDBM), p. 5:1–5:12. ISBN 978-1-4503-1921-8. Disponível em: <<http://doi.acm.org.ez67.periodicos.capes.gov.br/10.1145/2484838.2484854>>. Citado nas páginas 15 e 40.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2015. p. 1–9. Citado nas páginas 35, 36, 57 e 60.

TAO, Y.; PAPADIAS, D.; LIAN, X. Reverse knn search in arbitrary dimensionality. In: **Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30**. VLDB Endowment, 2004. (VLDB '04), p. 744–755. ISBN 0-12-088469-0. Disponível em: <<http://dl.acm.org.ez67.periodicos.capes.gov.br/citation.cfm?id=1316689.1316754>>. Citado na página 41.

TEKESTE, I.; DEMIR, B. Advanced local binary patterns for remote sensing image retrieval. In: **IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium**. [S.l.: s.n.], 2018. p. 6855–6858. ISSN 2153-7003. Citado na página 26.

TIAN, D. ping *et al.* A review on image feature extraction and representation techniques. **International Journal of Multimedia and Ubiquitous Engineering**, Citeseer, v. 8, n. 4, p. 385–396, 2013. Citado na página 28.

TORRES, R. da S.; FALCAO, A. X. Content-based image retrieval: theory and applications. 2006. Citado na página 26.

TRAINA, A. J.; JR, C. T.; BUENO, J. M.; AZEVEDO-MARQUES, P. M. The metric histogram: A new and efficient approach for content-based image retrieval. In: **Visual and Multimedia Information Management**. [S.l.]: Springer, 2002. p. 297–311. Citado nas páginas 29 e 30.

TRAN, Q. T.; TANIAR, D.; SAFAR, M. Reverse k nearest neighbor and reverse farthest neighbor search on spatial networks. In: _____. **Transactions on Large-Scale Data- and Knowledge-Centered Systems I**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 353–372. ISBN 978-3-642-03722-1. Disponível em: <http://dx.doi.org/10.1007/978-3-642-03722-1_14>. Citado na página 41.

WANG, L.; ZHANG, Y.; FENG, J. On the euclidean distance of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, n. 8, p. 1334–1339, 2005. Citado na página 39.

WU, W.; YANG, F.; CHAN, C.-Y.; TAN, K.-L. Finch: Evaluating reverse k-nearest-neighbor queries on location data. **Proc. VLDB Endow.**, VLDB Endowment, v. 1, n. 1, p. 1056–1067, ago. 2008. ISSN 2150-8097. Disponível em: <<http://dx.doi.org.ez67.periodicos.capes.gov.br/10.14778/1453856.1453970>>. Citado na página 41.

XU, B.; LIN, H.; YANG, L.; XU, K.; ZHANG, Y.; ZHANG, D.; YANG, Z.; WANG, J.; LIN, Y.; YIN, F. Improve diversity-oriented biomedical information retrieval using supervised query expansion. In: IEEE. **2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. [S.l.], 2018. p. 556–559. Citado na página 41.

ZEZULA, P.; AMATO, G.; DOHNAL, V.; BATKO, M. **Similarity search: the metric space approach**. [S.l.]: Springer Science & Business Media, 2006. v. 32. Citado nas páginas 37, 38 e 40.

ZHANG, S.; LI, X.; ZONG, M.; ZHU, X.; WANG, R. Efficient knn classification with different numbers of nearest neighbors. **IEEE transactions on neural networks and learning systems**, IEEE, v. 29, n. 5, p. 1774–1785, 2018. Citado na página 40.

ZIEGLER, C.-N.; LAUSEN, G.; SCHMIDT-THIEME, L. Taxonomy-driven computation of product recommendations. In: **Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management**. New York, NY, USA: ACM, 2004. (CIKM '04), p. 406–415. ISBN 1-58113-874-1. Disponível em: <<http://doi.acm.org.ez67.periodicos.capes.gov.br/10.1145/1031171.1031252>>. Citado na página 41.