

Divergências de Bregman aplicadas na Recuperação por Conteúdo de Imagens de Displasias Orais

Tiago R. M. Soares¹, Adriano B. Silva¹, Adriano M. Loyola²,
Sérgio V. Cardoso², Paulo R. de Faria², Leandro A. Neves³,
Marcelo Z. Nascimento¹, Humberto Razente¹

¹Faculdade de Computação – Universidade Federal de Uberlândia

²Escola de Odontologia – Universidade Federal de Uberlândia

³Departamento de Ciência da Computação e Estatística (DCCE) – UNESP

tiago.soares@ufu.br

Abstract. *In recent years, several works have employed non-metric functions to deal with the semantic gap between query results and users' perception of similarity in Content-Based Image Retrieval (CBIR) systems. In this study, we investigated the search of images of oral cavity dysplasias, obtained from histological slides that contained induced lesions of mice of the C57Bl/6 lineage. The images were segmented with a method based on the Mask R-CNN neural network to extract morphological and non-morphological descriptors. Bregman divergences (Kullback-Leibler and Mahalanobis) and metrics (Euclidean and Manhattan) were employed in searches, which were evaluated for precision and recall. Bregman divergences proved to be more effective in identifying the dysplasia levels.*

Resumo. *Nos últimos anos vários trabalhos empregaram funções não métricas para lidar com a descontinuidade semântica entre os resultados de consultas e a percepção de similaridade dos usuários em sistemas de Recuperação de Imagens Baseada em Conteúdo (CBIR). Neste estudo investigou-se a recuperação de imagens de displasias da cavidade oral, obtidas a partir de lâminas histológicas que continham lesões induzidas em camundongos da linhagem C57Bl/6. As imagens foram segmentadas com um método baseado na rede neural Mask R-CNN para extração de descritores morfológicos e não morfológicos. As divergências de Bregman (Kullback-Leibler e Mahalanobis) e métricas (Euclidiana e Manhattan) foram utilizadas em buscas, que foram avaliadas por precisão e revocação. As divergências de Bregman mostraram-se mais eficazes na identificação dos níveis de displasias.*

1. Introdução

Com o grande número de exames de imagens realizados diariamente, o simples armazenamento e manipulação individual dos exames por meio de uma chave de identificação é um desafio devido ao grande volume de dados. Para o desenvolvimento de sistemas de auxílio ao diagnóstico (*Computer Aided Diagnosis – CAD*) e sistemas de recuperação de imagem baseada em conteúdo (*Content-based Image Retrieval – CBIR*), surge a necessidade de encontrar similaridades ou dissimilaridades entre imagens de exames distintos. O

CBIR tem contribuído no processo de tomada de decisão clínica, recuperando conteúdos em bases de dados de caso similares que já possuem diagnósticos [Junior et al. 2020].

A medida de similaridade é um valor numérico que representa o quanto dois elementos são semelhantes, de modo que quanto maior o valor, maior a similaridade. Para a dissimilaridade, quanto menor o valor, maior a similaridade entre dois elementos. No modelo de dissimilaridade usualmente se empregam métricas, como nas distâncias Euclidiana e Manhattan. A maior parte dos trabalhos emprega distância Euclidiana para computação da dissimilaridade. No entanto, em várias aplicações do mundo real a medida de dissimilaridade não é uma métrica, por exemplo, quando se comparam distribuições de probabilidade, séries temporais, imagens, matrizes entre outros [Skopal 2006]. Para contornar esse problema, é possível a aplicação de funções não métricas como a divergência Kullback-Leibler (KL), usada para comparar duas funções de densidade de probabilidade.

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

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

Este trabalho aborda o desafio de realizar recuperação de imagem por conteúdo de lesões com potencial maligno obtidas de imagens histológicas da cavidade oral. Foram empregadas técnicas para segmentação das regiões de núcleos com uma rede Mask R-CNN. Os descritores obtidos de textura e baseados em modelos de redes neurais convolucionais foram extraídos das regiões segmentadas para composição dos vetores. Finalmente, foram aplicadas etapas de processamento que resultaram em espaços de alta dimensionalidade para avaliação da qualidade da recuperação com o emprego de diferentes medidas de dissimilaridade. Assim, as principais contribuições desse trabalho são:

- Investigação de métricas das divergências de Bregman no processo de recuperação

- de conteúdo para avaliação dos graus de displasia em imagens histológicas;
- Estudo comparativo entre as divergências de Bregman (KL e Mahalanobis) e as medidas tradicionais (Euclidiana e Manhattan) em imagens da cavidade oral;
 - Análise do impacto na eficiência do processo de recuperação de conteúdo na avaliação de descritores obtidos por meio de redes neurais convolucionais e características extraídas por meio de técnicas *handcraft*.

2. Materiais e Métodos

O método proposto envolveu um processo de segmentação, extração de características e avaliação das métricas, como apresentado na Figura 1. O sistema foi desenvolvido usando a linguagem Python. Para a extração de características morfológicas específicas das regiões de interesse nas imagens histológicas foram empregados segmentação e pós-processamento das imagens com o Python “regionprops_table” antes da extração.

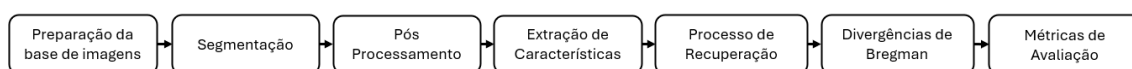


Figura 1. Fluxograma dos principais estágios empregados para recuperação de imagem por conteúdo em imagens histológicas.

O câncer é uma condição genética desencadeada por mutações no DNA, que podem surgir espontaneamente ou por fatores ambientais. Algumas dessas mutações podem ser transmitidas hereditariamente. As células com mutações passam por um processo de seleção natural, onde aquelas que promovem vantagens em termos de crescimento e sobrevivência podem proliferar e se tornar dominantes na população celular [Teixeira 2007].

Lesões pré-cancerosas são anormalidades benignas que têm potencial para evoluir para diagnósticos malignos, com as displasias sendo um exemplo importante. Displasias são caracterizadas por alterações na morfologia celular, como o formato geométrico, tamanho e características internas da pigmentação, especialmente na camada epitelial. No entanto, o diagnóstico preciso de displasias é desafiador devido à subjetividade na classificação das lesões por especialistas, levando à necessidade de abordagens quantitativas para uma caracterização objetiva. Embora a análise histológica de tecidos seja uma das melhores maneiras de caracterizar os níveis de displasia, desafios como carga de trabalho elevada e subjetividade podem impactar nessa análise [van der Waal 2009].

O diagnóstico por imagens e o emprego de sistemas de recuperação de imagens desempenha um papel crucial no diagnóstico das displasias. Essas imagens auxiliam especialistas na localização precisa de tumores e na orientação durante procedimentos cirúrgicos, sendo relevantes em intervenções minimamente invasivas, contribuindo para diagnósticos mais precisos e tratamentos eficazes. As displasias podem apresentar graus de evolução, sendo leves, moderadas ou severas, dependendo das alterações e intensidades encontradas na camada epitelial [Silva et al. 2022].

2.1. Preparação da Base de Imagens

A avaliação dos tecidos histológicos é amplamente considerada como uma abordagem confiável para a detecção de doenças. Especificamente, no campo da saúde bucal, são conduzidas pesquisas sobre lesões, chamadas de pré-câncer, com o objetivo de esclarecer as características e anomalias que desempenham um papel significativo no surgimento de

patologias. A histologia compreende a investigação das estruturas, funções e organização celular para a formação de órgãos. Além disso, envolve a análise das dimensões celulares por meio do uso de microscópio, sendo necessário a preparação do material fatiado em cortes histológicos, resultando em lâminas obtidas por etapas que compõem o processo de preparação dos tecidos. Essas etapas podem ser denominadas por: fixação, desidratação e diafanização, inclusão, microtomia, montagem e coloração [Gonzalez and Woods 2009].

Para este trabalho as imagens histológicas foram coletadas, tratadas e fornecidas por [Silva et al. 2022]. Foi criado um repositório de imagens a partir de lâminas que continham lesões induzidas experimentalmente na língua de camundongos da linhagem C57Bl/6, utilizando o carcinógeno 4NQO. O experimento *in vivo* foi previamente autorizado pelo Comitê de Ética na Utilização de Animais, sob o protocolo número 038/09, e conduzido no período compreendido entre 2009 e 2010 [Faria et al. 2011]. Essas imagens foram empregadas em um estudo apresentado por [Silva et al. 2022] para criação de um sistema de apoio a tomada de decisão constituído das etapas: segmentação, pós-processamento, extração de características e classificação.

2.1.1. Segmentação

Durante a fase de segmentação de núcleos displásicos foi implementado um método baseado na rede neural Mask R-CNN com a arquitetura ResNet-50. A Mask R-CNN é uma rede neural convolucional orientada por regiões, projetada para analisar agrupamentos de pixels semelhantes organizados em regiões para fins de classificação [He et al. 2016].

O processo de segmentação ocorreu em três etapas, integrando uma rede denominada *Region Proposal Network* (RPN). A ResNet-50 foi empregada para identificar objetos na imagem, utilizando suas camadas convolucionais para gerar uma pirâmide de características conhecida como *Feature Proposal Network* (FPN). As camadas da FPN, contendo informações de alto e baixo nível da imagem, foram utilizadas para identificar objetos por meio do cálculo de similaridade entre regiões [He et al. 2016, Silva et al. 2022].

Na segunda etapa, utilizando uma janela deslizante de 3x3 pixels nas regiões propostas pela RPN, foi gerado um vetor de características. Esse vetor foi então processado por duas camadas totalmente conectadas: uma para regressão de caixas delimitadoras e outra para classificação de regiões. Durante cada passo da janela deslizante, k regiões candidatas foram calculadas simultaneamente. A camada de regressão teve tamanho $4k$, representando coordenadas e dimensões de cada região, enquanto a camada de classificação teve tamanho $2k$, indicando a probabilidade de pertencer às classes de núcleo ou fundo. Na terceira etapa, a rede gerou máscaras binárias para cada objeto na imagem, por meio de uma rede convolucional em conjunto com mapas de características, estimando uma máscara de tamanho $m \times n$ para cada núcleo presente na imagem [He et al. 2016].

Durante o treinamento, dez imagens por grau de displasia foram usadas, totalizando 40 imagens. Com a colaboração de um especialista, foram geradas 1.220 máscaras binárias para os núcleos presentes nessas imagens. A rede foi pré-treinada com a base ImageNet e passou por um segundo treinamento usando a arquitetura ResNet-50, com 40 *epochs* e 150 iterações por *epoch*. A taxa de aprendizado foi 0,001, e o otimizador empregado foi o *Stochastic Gradient Descent* (SGD), com momento 0,9 [Silva et al. 2022].

2.1.2. Pós-Processamento

Após gerar as imagens binárias para cada núcleo, é possível que ocorram falhas nos contornos, áreas não totalmente preenchidas e pequenos artefatos, prejudicando a segmentação do núcleo. Para eliminar essas regiões de falsos positivos e aprimorar os núcleos segmentados, foram aplicadas operações morfológicas na etapa de pós-processamento conforme proposto por [Silva et al. 2022].

O processo inicia com o preenchimento dos contornos dos núcleos e eliminação de regiões de falsos negativos, ou seja, preenchendo os núcleos, por meio da operação morfológica de dilatação. Essa operação amplia os objetos em uma imagem binária, preenchendo espaços vazios e destacando linhas e contornos de bordas [Silva et al. 2022].

Após a dilatação, regiões de falsos negativos, como áreas não preenchidas, podem ainda estar presentes nas imagens. Nesse caso, uma função de preenchimento de regiões foi utilizada, seguida por um filtro de erosão para eliminar ruídos e restaurar o tamanho inicial dos núcleos. Após esses procedimentos, qualquer objeto na imagem com uma área calculada inferior a 30 pixels foi designado como região de fundo. Na Figura 2 são apresentadas as etapas do processo de segmentação de uma imagem [Silva et al. 2022].

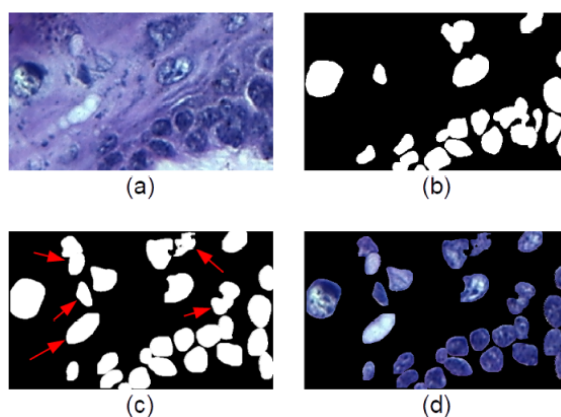


Figura 2. Segmentação em uma imagem saudável com uso da Mask R-CNN: (a) imagem original, (b) padrão-ouro, (c) máscara resultante e (d) aplicação da máscara sobre a imagem original. As setas em vermelho indicam regiões Falsos Positivos (FP). [Silva et al. 2022]

2.1.3. Extração de Características nas Imagens Segmentadas

Após as etapas de segmentação e pós-processamento, as características morfológicas e não morfológicas foram extraídas das regiões de interesse. Essas características incluíram informações sobre a forma, tamanho e outras propriedades geométricas das regiões identificadas, fornecendo *insights* valiosos para análises. Essa metodologia abrangente de extração de características, desde a segmentação inicial até a análise detalhada das propriedades morfológicas, será denominada *Handcraft*, enfatizando o processo manual e cuidadoso envolvido em cada etapa do processo [Silva et al. 2022].

As análises dos tecidos para definir os níveis de displasia envolveram a obtenção de características morfológicas e não morfológicas das imagens. Para as características

morfológicas foram obtidos as métricas: Área (A), Excentricidade (E), Orientação (O_R), Perímetro (P_E) e Solidez (S). Além disso, foram consideradas características não morfológicas, como Entropia (E_N) e Índice de Moran (I_M), que exploram informações internas das regiões em imagens histológicas [Silva et al. 2022].

Foram extraídas 13 características para cada núcleo n em uma imagem I_k , onde k representa o número da imagem. Cada imagem I_k é representada por uma matriz de tamanho $13 \times n$. Em seguida, medidas estatísticas de média (M) e desvio padrão (D) foram calculadas para cada característica em I , resultando em dois vetores, M_k e D_k , contendo 13 descritores cada grupo. Esses vetores foram concatenados para formar uma representação final, $I_k = [M_k, D_k]$, utilizado na etapa de classificação das imagens [Silva et al. 2022].

2.2. Extração de Características com Redes Neurais Convolucionais

As Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNN) são empregadas na extração de características de imagens histológicas podendo gerar descritores relevantes, até mesmo de imagens que não foram segmentadas. Ao contrário de métodos tradicionais que demandam a identificação prévia de regiões de interesse, as CNNs são capazes de aprender automaticamente padrões e características diretamente a partir dos dados brutos da imagem [Silva et al. 2022].

A obtenção das características das imagens foi realizada por meio da aplicação de dois modelos CNNs. Nos experimentos, foram empregados as camadas de convolução, de agrupamento máximo e precedendo a camada de achatamento. As saídas resultantes do processo de achatamento foram armazenadas em vetores de características [Silva et al. 2022]. Para treinamento dos modelos foi empregado a técnica de transferência de aprendizagem com os pesos extraídos da base ImageNet.

O modelo AlexNet é composto por cinco camadas convolucionais intercaladas com três operações de pooling, seguidas por duas camadas totalmente conectadas e uma função softmax. Por outro lado, o modelo ResNet-50 consiste em uma camada de entrada seguida por 48 camadas convolucionais organizadas em blocos repetidos, apresentando 16 conexões de salto e finalizando com a camada de classificação [Krizhevsky et al. 2017, Silva et al. 2022].

As camadas iniciais dessas arquiteturas são responsáveis por extrair características que quantificam bordas, formas e cores dos objetos nas imagens. As camadas finais são empregadas na identificação de padrões de textura globais. No caso do modelo AlexNet, foram extraídos 616.032 atributos provenientes das cinco camadas convolucionais, excluindo as camadas totalmente conectadas e a softmax. No que diz respeito à arquitetura ResNet-50, um total de 1.095.488 atributos foi obtido, considerando as duas primeiras camadas e as três últimas do modelo [Silva et al. 2022].

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

2.3. Processo de Recuperação

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

Uma função de similaridade é aquela na qual duas imagens são mais similares quanto maior o valor retornado pela função de comparação. Em muitos casos, essas funções são normalizadas para resultarem em valores entre 0 e 1, facilitando sua interpretação. Sendo X um conjunto qualquer e considerando que $s(x, y)$ é a similaridade entre os elementos x e y , então, s deve satisfazer as propriedades:

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

A dissimilaridade é obtida por meio de funções de distância (métricas), ou por meio de funções em que uma ou mais propriedades das métricas não são garantidas, com o objetivo de melhorar a qualidade semântica [Santini and Jain 1999]. Uma métrica é função $d : X \cdot X \rightarrow R$ que satisfaz as propriedades:

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

2.4. Métricas de Avaliação

Quando um usuário realiza uma consulta por meio do processo de recuperação de imagens, ele espera que o sistema retorne apenas imagens que sejam pertinentes aos seus interesses (ou seja, relevantes). É possível avaliar a satisfação gerada por um processo de recuperação medindo a quantidade de elementos relevantes recuperados e a precisão da recuperação. A precisão refere-se à capacidade do sistema de não apresentar um grande número de elementos irrelevantes para obter uma quantidade satisfatória de relevantes.

As principais medidas utilizadas para se avaliar o desempenho em um processo de recuperação de imagens são precisão e revocação. Dada uma consulta x , a precisão é definida como a fração entre o número de imagens relevantes recuperadas, representado por $R_a(x)$, e o número total de imagens recuperadas, $P_r(x) = R_a(x)/A$. A revocação é definida com a fração entre o número de imagens relevantes recuperadas, $R_a(x)$, e o número total de imagens relevantes na coleção, $R_e(x) = R_a(x)/R$.

Essas duas medidas estão correlacionadas, pois dependem do número de elementos relevantes recuperados, R_a . Em geral, aumentar a revocação tende a diminuir a precisão. Isso acontece porque, à medida que mais elementos são incluídos na resposta a uma determinada consulta, a quantidade de elementos relevantes recuperados aumenta. No entanto, juntamente com os relevantes, elementos não relevantes são incluídos em uma proporção maior que de relevantes. Isso resulta na redução da precisão, uma vez que a qualidade da resposta é comprometida.

2.5. Divergência de Bregman

Dada uma função convexa $\varphi : R^d \rightarrow R$, a divergência de Bregman entre x e $y \in \text{dom}(\varphi)$ é dada pela Equação 1 [Song et al. 2022], onde $\nabla\varphi(y)$ é o vetor gradiente de φ em y .

$$d_\varphi(x, y) = \varphi(x) - \varphi(y) - (\nabla\varphi(y), x - y) \quad (1)$$

Quando diferentes funções convexas são empregadas, as distâncias de Bregman definem várias funções de distância conhecidas, entre elas:

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

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

Os vetores de características podem conter valores iguais a zero. Esta questão exige um tratamento, uma vez que a divergência KL é definida utilizando a função logarítmica cujo domínio é $x > 0$. A divergência KL entre dois vetores x e y , $x = (x_1, \dots, x_d)$ e $y = (y_1, \dots, y_d)$ é definida pela Equação 2, onde x e y pertencem ao conjunto d -Simplex definido na Equação 3.

$$d_\varphi(x, y) = \sum_{j=1}^d x_j \log_2\left(\frac{x_j}{y_j}\right) \quad (2)$$

$$\Delta^d = \{x = (x_1, x_2, \dots, x_d); 0 \leq x_j \leq 1, \text{ tal que } \sum_{j=1}^d x_j = 1\} \quad (3)$$

Assim, se para algum j , $x_j = 0$ e $y_j \neq 0$, a parcela $0 \log_2(0/y_j)$ tende a nula, uma vez que o limite $\lim_{x_j \rightarrow 0} x_j \log_2 x_j = 0$. Entretanto se $x_j \neq 0$ e $y_j = 0$, pode-se assumir que $x_j \log_2(x_j/0) = \infty$, como apresentado na Equação 4. É possível ter duas opções de tratamento para os dados quando os vetores x e y apresentarem coordenadas onde $x_j \neq 0$ e $y_j = 0$. A primeira opção, denominada Kullback-Leibler baseado na Teoria da Informação (KL/TI), é assumir $x_j \log_2(x_j/0) = \gamma$, onde γ é um valor suficientemente grande. A segunda, denominada denominado Kullback-Leibler com o tratamento do ϵ (KL/ ϵ), considera que se $y_j \rightarrow 0$, então $y_j = \epsilon$, onde ϵ é um valor próximo de zero [Rocha et al. 2021]. Neste trabalho foi utilizada a opção Kullback-Leibler com o tratamento do ϵ (KL/ ϵ).

$$x_j \log_2\left(\frac{x_j}{y_j}\right) = \begin{cases} 0, & \text{se } x_j = 0 \\ \infty, & \text{se } x_j \neq 0 \text{ e } y_j = 0 \end{cases} \quad (4)$$

3. Experimentos e Análise dos Resultados

Neste estudo foram empregados os descritores baseados em modelos de redes neurais convolucionais obtidos das arquiteturas AlexNet e ResNet-50 para a extração de características das imagens de cavidade bucal em relação aos graus de displasia. Para obtenção

das informações *handcrafts*, as imagens foram segmentadas por uma rede Mask R-CNN. Foram incluídas informações de textura obtidas por meio de entropia e índice de Moran. Também foram investigados as composição desses descritores não morfológicos com outras informações morfológicas obtidos por métricas como área, excentricidade, orientação, perímetro e solidez. A associação da extração de características por meio de redes neurais (AlexNet e ResNet-50) e informações *handcraft* permitiu obter um conjunto abrangente de informações descritivas das imagens, que foram fundamentais para o estudo e análise dos graus de displasia das imagens histológicas da cavidade oral.

Para que fosse possível mensurar as diferenças entre as divergências de Bregman e as distâncias métricas, a metodologia foi aplicada em uma série de associações investigadas na composição dos vetores obtidos pelas técnicas de extração das características: AlexNet, ResNet-50, AlexNet + ResNet-50, Handcraft, Handcraft + AlexNet, Handcraft + ResNet-50, e Handcraft + AlexNet + ResNet-50. Para avaliar a precisão e revocação foi realizado a comparação entre os graus de displasia: Saudável x Leve, Saudável x Moderado, Saudável x Severo, Leve x Moderado, Leve x Severo e Moderado x Severo.

Para a análise de recuperação de imagens utilizando as divergências de Bregman, foram selecionadas as duas abordagens diferentes de distâncias métricas, Euclidiana e Manhattan, e as duas distâncias não-métricas, KL e Mahalanobis. A partir destas distâncias, foram calculadas as médias de precisão e revocação para as 10 primeiras imagens recuperadas do banco de dados de displasias. Para as combinações feitas entre as técnicas de extração das características, nota-se que a utilização de somente AlexNet ou ResNet-50 proporcionou resultados de descritores iguais a zero, tendo impacto direto na precisão da divergência de KL com o tratamento do ϵ . Para contornar essa dificuldade, foi utilizada a combinação de técnicas de extração apresentada na Tabela 1.

A Tabela 1 apresenta os resultados obtidos utilizando as comparações propostas em seus níveis de precisão e revocação. O efeito de visualização foi realizado por uma média dos resultados pelos grupos de (50/40/30/20/10) características mais relevantes das imagens. Entretanto, optou-se por focar exclusivamente nas distâncias Euclidiana e KL, que demonstraram resultados mais significativos ao examinar as médias dos conjuntos de dados. No entanto, vale ressaltar que as distâncias de Manhattan e Mahalanobis apresentaram resultados satisfatórios quando consideradas individualmente em outros casos.

É possível notar que a média de precisão nas 10 primeiras imagens recuperadas de todas as classificações dos graus de displasia, utilizando somente “AlexNet”, resultaram em uma precisão de 80,1 e 80,7, com desvio padrão de 1,9 e 5,7 para Euclidiana e KL respectivamente. Já com os descritores da “ResNet-50”, os valores foram de 79,0 e 78,0, com desvio padrão de 1,9 e 6,2 respectivamente, para Euclidiana e KL. Com os descritores *Handcraft*, os resultados mostram valores exatos, sem a inclusão de médias, a precisão foi de 61,3 e 72,0, respectivamente, com as métricas Euclidiana e KL. Para as combinações “AlexNet + ResNet-50”, a precisão foi de 81,0 e 86,9, com desvio padrão de 0,9 e 3,8 respectivamente, enquanto com a combinação Handcraft + AlexNet”, a precisão média foi de 82,1 e 90,6 com desvio padrão de 1,2 e 2,8. Com a associação *Handcraft* + ResNet-50, os valores de precisão média foi de 79,8 e 83,7, com desvio padrão de 1,0 e 4,0 respectivamente, para Euclidiana e KL, e para os descritores *Handcraft* + ResNet-50 + AlexNet, a precisão encontrada foi de 81,6 e 88,3 com desvio padrão de 1,0 e 3,5. O fato do desvio ser maior para divergência KL está relacionado a baixa performance utili-

Tabela 1. Valores da métricas Precisão (Prec.) e Revocação (Rev.) dos descritores e distâncias para as displasias nos graus leve (leve), moderada (mod.), severa (sev.) e saudável (saud.).

CNN Descritores	Distância Métrica	saud. x leve		saud. x mod.		saud. x sev.		leve x mod.		leve x sev.		mod. x sev.	
		Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.	Prec.	Rev.
AlexNet	Euclideana	92,0	12,4	91,1	12,3	82,3	11,1	70,6	9,5	72,4	9,8	72,5	9,8
	KL	81,8	9,6	62,9	8,6	85,8	7,9	82,1	8,4	89,3	9,1	82,2	8,7
ResNet-50	Euclideana	92,7	12,5	87,5	11,8	87,0	11,8	66,2	8,9	72,6	9,8	68,0	9,2
	KL	91,7	9,4	96,0	12,0	89,9	8,6	61,9	7,7	50,3	8,6	77,4	8,9
AlexNet + ResNet-50	Euclideana	91,7	12,4	90,9	12,3	88,9	12,0	70,8	9,6	74,3	10,0	69,2	9,3
	KL	98,2	11,3	96,4	11,2	99,0	11,6	73,7	8,2	74,5	9,8	79,5	9,1
<i>Handcraft</i>	Euclideana	57,8	8,1	63,4	8,9	60,9	8,5	56,5	7,6	64,2	8,7	65,3	8,8
	KL	76,5	8,6	77,5	9,5	74,6	9,2	60,6	7,8	69,7	9,2	72,8	8,9
<i>Handcraft</i> + AlexNet	Euclidiana	94,4	13,2	93,1	13,5	85,6	11,9	70,7	9,8	75,2	10,2	73,5	10,2
	KL	96,1	11,6	94,6	11,4	96,3	9,3	84,7	8,7	89,7	9,1	82,3	9,0
<i>Handcraft</i> + ResNet-50	Euclidiana	94,6	13,2	89,2	12,9	87,6	12,3	66,7	9,4	72,9	9,9	68,0	9,5
	KL	98,5	10,8	96,1	12,9	96,0	10,1	72,6	8,3	61,9	8,7	77,3	9,2
<i>Handcraft</i> + Resnet-50 + AlexNet	Euclideana	93,1	13,0	91,9	13,3	89,4	12,5	71,4	10,0	74,8	10,1	69,0	9,6
	KL	98,3	11,5	96,9	12,2	99,1	12,0	79,0	8,5	76,6	9,9	80,0	9,3

zando apenas 10 descritores ou em algumas classes específicas de displasias como "leve x severa". No entanto, ao analisar as imagens para o grupo de características *Handcraft* + ResNet-50 + AlexNet, percebeu-se que a aplicação da distância de Mahalanobis não ofereceu resultados potencialmente positivos. Essa discrepância foi atribuída à sensibilidade dessa métrica à matriz de covariância das características extraídas. Uma vez que as características obtidas pelas arquiteturas AlexNet e ResNet50 apresentam uma correlação significativa, a distância de Mahalanobis pode não ser capaz de capturar adequadamente a variação dos dados, resultando em desempenho inferior. Isso contrasta com os resultados obtidos ao calcular as distâncias para os grupos AlexNet e ResNet50 separadamente.

Acredita-se que os resultados obtidos com a abordagem Kullback-Leibler estão de acordo com a observação de [Banerjee et al. 2004] onde a divergência deve capturar as propriedades de similaridade desejáveis na aplicação, e em conjuntos de dados de imagens que são esparsos ou de alta dimensão, a divergência pode ser mais robusta do que a distância Euclidiana. Também foi feita a análise observando o desempenho ao recuperar todas as imagens presentes no banco coletado. Na Figura 3 a divergência de Bregman com a abordagem KL é visualmente melhor com o aumento na quantidade de imagens recuperadas utilizando a associação *Handcraft* + ResNet-50 + AlexNet.

4. Conclusão

Os sistemas CBIR têm tido ampla aplicação na análise e recuperação de imagens. No entanto, as aplicações atuais apresentam limitações evidentes nos processos devido a utilização de métricas. Para mitigar essas limitações, alguns estudos têm explorado a aplicação desses sistemas em conjunto com diversas medidas de similaridade para o tratamento da descontinuidade semântica. Neste trabalho foram investigadas as divergências de Bregman (Kullback-Leibler e Mahalanobis) para computar a dissimilaridade entre vetores de características que representam dados de imagens histológicas de displasia bucal.

Para avaliação da metodologia proposta foram apresentados os resultados de pre-

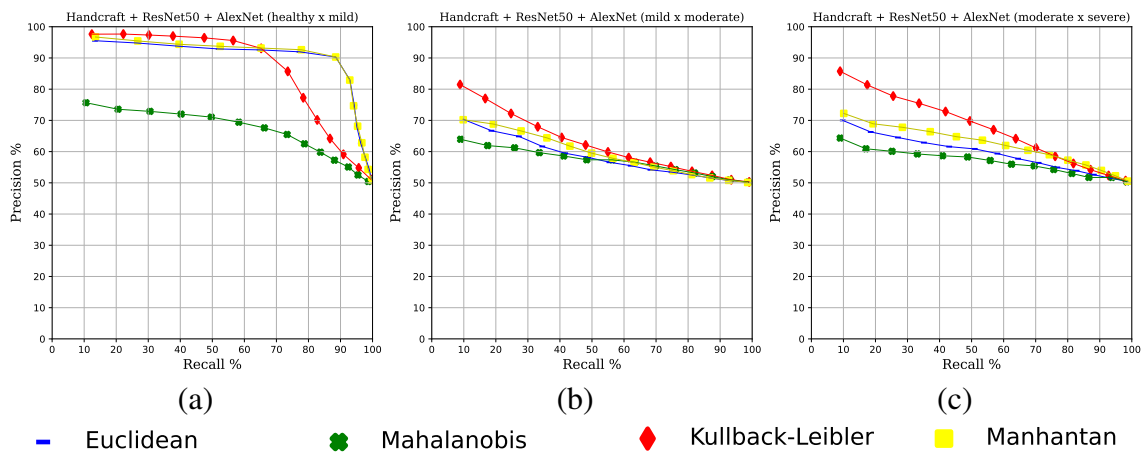


Figura 3. Gráficos de precisão x revocação. (a) *Handcraft* + ResNet-50 + AlexNet (saúdável x leve) (b) *Handcraft* + ResNet-50 + AlexNet (leve x moderada) (c) *Handcraft* + ResNet-50 + AlexNet (moderada x severa)

cisão e revocação entre os diversos níveis de displasia bucal (saúdável, leve, moderado e severa) e diferentes combinações de técnicas para extração das características. Os resultados permitem observar que na maioria dos casos o uso das divergências de Bregman (KL) foi superior às métricas Euclidiana e Manhattan, mostrando serem mais eficazes na identificação de características que melhor identificam os níveis de displasias.

A divergência de Bregman, com a abordagem Kullback-Leibler, torna os sistemas CBIR mais eficientes do que o uso das distâncias tradicionais (Euclidiana e Manhattan). Observa-se que a metodologia abordada poderá auxiliar especialistas da área da saúde a estudar e analisar lesões displásicas da cavidade oral. Em trabalhos futuros, outros descritores obtidos de camadas ou modelos de redes neurais convolucionais serão investigados em relação a separação dos graus de displasia em imagens histológicas.

5. Agradecimentos

Os autores agradecem o apoio financeiro do Conselho Nac. Desenvolvimento Científico e Tecnológico (CNPq): Projetos #313643/2021-0, #311404/2021-9 e #307318/2022-2.

Referências

- Banerjee, A., Merugu, S., Dhillon, I. S., and Ghosh, J. (2004). Clustering with bregman divergences. In *SIAM Int'l Conf. Data Mining*, pages 234–245. SIAM.
- Faria, P. R., Chammas, R., de Melo, T. L., Hsu, D. K., Liu, F.-T., Nonogaki, S., Cardoso, S. V., and Loyola, A. M. (2011). Absence of galectin-3 does not affect the development of experimental tongue carcinomas in mice. *Exp. Mol. Pathol.*, 90(2):189–193.
- Gonzalez, R. C. and Woods, R. C. (2009). *Processamento digital de imagens*. Pearson.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conf. Comp. Vision and Pat. Recog. (CVPR)*, pages 770–778.
- Junior, J. R. F., Santos, M. K., and Azevedo-Marques, P. M. (2020). Arcabouço para classificação, recuperação por conteúdo e radiômica de imagens médicas: uma investigação de biomarcadores quantitativos para o câncer de pulmão. In *Anais Estendidos do XX Simpósio Bras. Computação Aplicada à Saúde*, pages 13–18. SBC.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.
- Liu, M., Vemuri, B. C., Amari, S.-I., and Nielsen, F. (2012). Shape retrieval using hierarchical total bregman soft clustering. *IEEE TPAMI*, 34(12):2407–2419.
- Rocha, B. M., Ferreira, D. P. L., Barcelos, C. A. Z., and Soares, F. (2021). I-generalized and kullback-leibler divergences for content-based image retrieval. In *34th IEEE Canadian Conf. Elect. and Comp. Eng. (CCECE)*, pages 1–4. IEEE.
- Santini, S. and Jain, R. C. (1999). Similarity measures. *IEEE TPAMI*, 21(9):871–883.
- Santos, M. d. O., Lima, F. C. d. S. d., Martins, L. F. L., Oliveira, J. F. P., Almeida, L. M., and Cancela, M. d. C. (2023). Estimativa de incidência de câncer no brasil, 2023-2025. *Rev. Bras. Cancerol.*, 69(1):1–12.
- Silva, A., Oliveira, C., Pereira, D., Tosta, T., Martins, A., Loyola, A., Cardoso, S., Faria, P., Neves, L., and Nascimento, M. (2022). Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. In *SIBGRAPI*, pages 264–269. IEEE.
- Silva, A. B., Martins, A. S., Tosta, T. A. A., Loyola, A. M., Cardoso, S. V., Neves, L. A., de Faria, P. R., and do Nascimento, M. Z. (2024). OralEpitheliumDB: A dataset for oral epithelial dysplasia image segmentation and classification. *Journal of Imaging Informatics in Medicine*. pages 1-20.
- Skopal, T. (2006). On fast non-metric similarity search by metric access methods. In *EDBT*, volume 3896 of *LNCS*, pages 718–736. Springer.
- Song, Y., Gu, Y., Zhang, R., and Yu, G. (2022). BrePartition: Optimized high-dimensional ik/iNN search with bregman distances. *IEEE TKDE*, 34(3):1053–1065.
- Sung, H., Ferlay, J., Siegel, R., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 71(3):209–249.
- Teixeira, M. (2007). Explicação diversa para a origem do câncer, com foco nos cromossomos, e não nos genes, ganha corpo no establishment científico. *Revista Latinoamericana de Psicopatologia Fundamental*, 10(4):664–676.
- Tommasino, C., Merolla, F., Russo, C., Staibano, S., and Rinaldi, A. M. (2023). Histopathological image deep feature representation for CBIR in smart PACS. *Journal of Digital Imaging*, 36(5):2194–2209.
- Tosta, T., Faria, P., Servato, J. P., Neves, L., Roberto, G., Martins, A., and Nascimento, M. Z. (2019). Unsupervised method for normalization of hematoxylin-eosin stain in histological images. *Comput. Med. Imaging Graph.*, 77:101646.
- van der Waal, I. (2009). Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. *Oral Oncology*, 45(4–5):317–323.
- Zheng, Y., Jiang, Z., Zhang, H., Xie, F., Ma, Y., Shi, H., and Zhao, Y. (2018). Histopathological whole slide image analysis using context-based cbir. *IEEE Transactions on Medical Imaging*, 37(7):1641–1652.