

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ELIEZER SOARES FLORES

**Representações Esparsas Através de  
Dicionários para Processamento e  
Análise de Imagens Macroscópicas de  
Lesões Melanocíticas**

Tese apresentada como requisito parcial para  
a obtenção do grau de Doutor em Ciência da  
Computação

Orientador: Prof. Dr. Jacob Scharcanski

Porto Alegre  
2022

## CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Flores, Eliezer Soares

Representações Esparsas Através de Dicionários para Processamento e Análise de Imagens Macroscópicas de Lesões Melanocíticas / Eliezer Soares Flores. – Porto Alegre: PPGC da UFRGS, 2022.

145 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2022. Orientador: Jacob Scharcanski.

1. Representações esparsas. 2. Aprendizado de dicionários. 3. Processamento de imagens. 4. Análise de imagens. 5. Segmentação de imagens. 6. Classificação de imagens. I. Scharcanski, Jacob. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões Mendes

Vice-Reitora: Prof.<sup>a</sup> Patrícia Helena Lucas Pranke

Pró-Reitor de Pós-Graduação: Prof. Júlio Otávio Jardim Barcellos

Diretora do Instituto de Informática: Prof.<sup>a</sup> Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. Claudio Rosito Jung

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## AGRADECIMENTOS

A Deus, por ter me dado saúde e força para encarar as dificuldades da vida e por ter colocado ao meu lado pessoas muito especiais sem as quais certamente eu não teria conseguido atingir este objetivo.

Aos meus pais Mariza de Fátima Soares Flores e Ildefonso Flores, meus maiores incentivadores e exemplos de dignidade, humildade e resiliência, pela incansável dedicação durante toda a minha criação, pelo amor incondicional e pelos incontáveis conselhos valorosos que levarei para sempre comigo.

A Lorena Aparecida Nunes Viana, amor da minha vida, por ter me aguentado durante este período longo e árduo, em que tive que conciliar trabalho e doutorado, e por ter topado compartilhar a sua jornada comigo, ao infinito de além!

Ao meu irmão Filipe Soares Flores, meu melhor amigo e exemplo de perseverança, pelo companheirismo e pelos inúmeros momentos especiais vivenciados desde o dia do seu nascimento.

Ao meu irmão Nei Ribeiro Flores, que infelizmente acabou partindo de maneira precoce no ano passado, por ter me instigado a trabalhar com computação desde cedo e pelo exemplo de coragem e superação.

A todos os meus familiares (espero que entendam que não citarei nomes aqui afim de evitar cometer a injustiça de deixar alguém importante de fora), por sempre me incentivaram nos estudos e compartilharem comigo tantos momentos especiais.

Ao meu orientador Prof. Dr. Jacob Scharcanski, por ter confiado no meu potencial, pela oportunidade única de desenvolver esta pesquisa tão relevante que me abriu tantas portas (por exemplo, conquistei um cargo de professor efetivo na Universidade Federal do Pampa em concurso na área de processamento digital de imagens) e por ter aceitado continuar me orientando mesmo sem eu ter a possibilidade de me dedicar integralmente ao doutorado.

Aos professores Dr. André Carlos Ponce de Leon Ferreira de Carvalho, Dr. Claudio Rosito Jung e Dr. Eduardo Antônio Barros da Silva (em ordem alfabética), membros da banca examinadora desta tese, pelos tantos comentários e apontamentos que permitiram uma melhoria considerável na qualidade deste trabalho.

Aos meus amigos, colegas, ex-colegas, ex-professores, alunos, ex-alunos e pessoas que cruzaram a minha vida contribuindo direta ou indiretamente na minha trajetória (novamente, não citarei nomes aqui pelo mesmo motivo que não citei o nome dos meus familiares com exceção dos meus pais e irmãos).

## RESUMO

O melanoma é o tipo mais letal de câncer de pele, uma vez que é mais propenso à metástase. Especificamente, a taxa de pacientes que sobrevivem pelo menos cinco anos após o diagnóstico dessa doença no estágio inicial é superior a 99%. No entanto, essa taxa diminui para cerca de 25% se a detecção ocorre somente no último estágio. Nesse contexto, sistemas que auxiliem no diagnóstico precoce do melanoma podem desempenhar um papel de extrema importância, especialmente em regiões nas quais o acesso a dermatologistas é precário. Contudo, diferenciar um melanoma de lesões melanocíticas benignas pode ser uma tarefa desafiadora, mesmo para especialistas experientes. Para lidar com esse problema, nesta tese, propõe-se um sistema automático para detecção de melanoma a partir de uma simples fotografia digital, o qual baseia-se em modelos de representações esparsas. Os resultados apresentados pelo sistema proposto são promissores e sugerem que o sistema proposto pode potencialmente superar alternativas estado-da-arte e até mesmo médicos treinados

**Palavras-chave:** Representações esparsas. Aprendizado de dicionários. Processamento de imagens. Análise de imagens. Segmentação de imagens. Classificação de imagens.

# **Sparse Representation through Dictionaries for Processing and Analysis of Macroscopic Images of Melanocytic Lesions**

## **ABSTRACT**

Melanoma is the most lethal type of skin cancer, since it is most prone to metastasis. Specifically, the rate of patients who survive at least five years after early stage diagnosis of this disease is over 99%. However, this rate decreases to about 25% if detection occurs only at the last stage. In this context, systems that assist in the early diagnosis of melanoma can play an extremely important role, especially in regions where access to dermatologists is poor. However, differentiating melanoma from benign melanocytic lesions can be a challenging task, even for experienced specialists. To address this problem, in this thesis, an automatic system is proposed for melanoma detection from a simple digital photograph, which is based on sparse representation models. The results presented by the proposed system are promising and suggest that it can potentially outperform state-of-the-art alternatives and even trained dermatologists.

**Keywords:** sparse representation, dictionary learning, image processing, image analysis, image segmentation, image classification.

## LISTA DE ABREVIATURAS E SIGLAS

ABCD	<i>Asymmetry, Border irregularity, Color variation, Diameter</i>
ALS	<i>Alternating Least Squares</i>
ARL	<i>Attention Residual Learning</i>
AUC	<i>Area Under the Curve</i>
CIE	<i>Commission Internationale de l'Éclairage</i>
CAD	<i>Computer Aided Diagnosis</i>
CD	<i>Critical Difference</i>
DCNN	<i>Deep Convolutional Neural Network</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DSNet	<i>Dermoscopic Skin Network</i>
DW	<i>Dynamic Weights</i>
DWT	<i>Discrete Wavelet Transform</i>
EM	<i>Expectation-Maximization</i>
FCNet	<i>Fully Convolutional Network</i>
FrCNet	<i>Full-resolution Convolutional</i>
FISTA	<i>Fast Iterative Shrinkage-Thresholding Algorithm</i>
GPGPU	<i>General Purpose Graphic Processing Units</i>
GMM	<i>Gaussian Mixture Model</i>
GT	<i>Ground Truth</i>
HSV	<i>Hue, Saturation, Value</i>
ITDL	<i>Information-Theoretic Dictionary Learning</i>
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
ISIC	<i>International Skin Imaging Collaboration</i>

ISTA	<i>Iterative Shrinkage-Thresholding Algorithm</i>
IW	<i>ISIC Winner</i>
KNN	<i>K Nearest Neighbors</i>
LGPNet	<i>Local/Global Patch-based Network</i>
MLE	<i>Maximum Likelihood Estimation</i>
MOD	<i>Method of Optimal Directions</i>
NGC	<i>Normalized Graph Cut</i>
NMF	<i>Nonnegative Matrix Factorization</i>
OMP	<i>Orthogonal Matching Pursuit</i>
PSPNet	<i>Pyramid Scene Parsing Network</i>
RGB	<i>Red, Green, Blue</i>
ROC	<i>Receiver Operating Characteristic</i>
ROI	<i>Region of Interest</i>
SegNet	<i>Segmentation Network</i>
SLA	<i>Skin Lesion Analyzer</i>
SRC	<i>Sparse Representation-based Classification</i>
STA	<i>Shrinkage-Thresholding Algorithm</i>
SVD	<i>Singular Value Decomposition</i>
TFP	Taxa de Falsos Positivos
TVP	Taxa de Verdadeiros Positivos
UITDL	<i>Unsupervised Information-Theoretic Dictionary Learning</i>
U-Net	<i>U-shaped Network</i>

## LISTA DE SÍMBOLOS

### Arranjos Numéricos e Indexação

$a$ ou $\bar{a}$	Um escalar (e.g., um número real)
$\mathbf{a}$ ou $\mathbf{a}$	Um vetor <sup>1</sup>
$A$ ou $\mathbf{A}$	Uma matriz
$\mathcal{A}$	Um tensor tridimensional
$a_i$	Valor do $i$ -ésimo componente de $\mathbf{a}$ (i.e., o $i$ -ésimo elemento de $\mathbf{a}$ )
$A_{l,c}$	Valor do elemento localizado na $l$ -ésima linha e na $c$ -ésima coluna de $A$
$\mathcal{A}_{l,c}^{[k]}$	Valor do elemento localizado na $l$ -ésima linha, na $c$ -ésima coluna e na $k$ -ésima camada de $\mathcal{A}$
$\mathcal{A}^{[i]}$	Matriz correspondente à $i$ -ésima camada de $\mathcal{A}$
$\underline{a}_l$	Vetor (i.e., matriz coluna) cujos valores dos componentes correspondem àqueles dos elementos que estão na $l$ -ésima linha de $A$
$\mathbf{a}_c$	Vetor cujos valores dos componentes correspondem àqueles dos elementos que estão na $c$ -ésima coluna de $A$
$\mathbf{0}_n$	Vetor de zeros, no qual todos os seus $n$ componentes valem zero
$\mathbf{1}_n$	Vetor de uns, no qual todos os seus $n$ componentes valem um
$\mathbf{I}_n$	Matriz identidade com $n$ linhas (e $n$ colunas)
$\delta_k$	$k$ -ésima coluna de $\mathbf{I}_n$

---

<sup>1</sup>Nesta tese, os termos vetor e sequência podem ser considerados sinônimos e são empregados para referir-se a uma sequência ordenada e finita.

## Conjuntos

$\mathbb{A}$	Um conjunto $\mathbb{A}$
$\mathbb{A}^c$	Conjunto composto pelos elementos que não estão em $\mathbb{A}$ (i.e., o complemento de $\mathbb{A}$ )
$\mathbb{N}$	Conjunto dos números naturais
$\mathbb{R}$	Conjunto dos números reais
$\mathbb{C}$	Conjunto dos números complexos
$\emptyset$	Conjunto vazio
$\{a, \dots, b\}$	Conjunto com todos os inteiros entre $a$ e $b$ , incluindo $a$ e $b$
$[a, b]$	Conjunto com todos os reais entre $a$ e $b$ , incluindo $a$ e $b$
$[a, b)$	Conjunto com todos os reais entre $a$ e $b$ , incluindo $a$ e excluindo $b$
$\mathbb{A} \cap \mathbb{B}$	Conjunto composto pelos elementos que estão em $\mathbb{A}$ e $\mathbb{B}$ (i.e., a intersecção entre $\mathbb{A}$ e $\mathbb{B}$ )
$\mathbb{A} \cup \mathbb{B}$	Conjunto composto pelos elementos que estão em $\mathbb{A}$ ou $\mathbb{B}$ (i.e., a união entre $\mathbb{A}$ e $\mathbb{B}$ )
$\mathbb{A} - \mathbb{B}$	Conjunto composto pelos elementos que estão em $\mathbb{A}$ e não estão em $\mathbb{B}$ (i.e., a diferença entre $\mathbb{A}$ e $\mathbb{B}$ )
$\mathbb{A}_1 \times \dots \times \mathbb{A}_n$	Conjunto composto por todas as sequências $\mathbf{a}$ de $n$ componentes tais que $a_i$ está em $\mathbb{A}_i$ para todo $i$ em $\{1, \dots, n\}$ (i.e., o produto cartesiano entre os conjuntos $\mathbb{A}_1, \dots, \mathbb{A}_n$ )
$\mathbb{A}^n$	Conjunto composto por todas as sequências $\mathbf{a}$ de $n$ componentes tais que $a_i$ está em $\mathbb{A}$ para todo $i$ em $\{1, \dots, n\}$ (i.e., o produto cartesiano entre os conjuntos $\mathbb{A}_1, \dots, \mathbb{A}_n$ no caso em que $\mathbb{A}_i = \mathbb{A}$ para todo $i$ em $\{1, \dots, n\}$ )
$\mathbb{A}^{n_l \times n_c}$	Conjunto composto por todas as matrizes $\mathbf{A}$ de $n_l$ linhas e $n_c$ colunas tais que $A_{l,c}$ está em $\mathbb{A}$ para todo $(l,c) \in \{1, \dots, n_l\} \times \{1, \dots, n_c\}$

$\mathbb{A}^{n_l \times n_c \times n_k}$	Conjunto composto por todos os tensores tridimensionais $\mathcal{A}$ de $n_l$ linhas, $n_c$ colunas e $n_k$ camadas tais que $\mathcal{A}_{l,c}^{[k]}$ está em $\mathbb{A}$ para todo $(l,c,k) \in \{1, \dots, n_l\} \times \{1, \dots, n_c\} \times \{1, \dots, n_k\}$
$\text{sup}(\mathbf{a})$	Conjunto composto pelos índices $i$ em $\{1, \dots, n\}$ tais que os componentes $a_i$ de $\mathbf{a}$ são diferentes de zero
$\text{col}(\mathbf{A})$	Conjunto composto pelos vetores que podem ser gerados por uma combinação linear das colunas de $\mathbf{A}$ (i.e., o espaço coluna de $\mathbf{A}$ )

### Operações do Cálculo e da Álgebra Linear

$\mathcal{A} : \mathbb{A} \rightarrow \mathbb{B}$	Função <sup>2</sup> cujo domínio é $\mathbb{A}$ e o contradomínio é $\mathbb{B}$
$\mathcal{A} \circ \mathcal{B}$	Composição das funções $\mathcal{A}$ e $\mathcal{B}$ , de modo que $(\mathcal{A} \circ \mathcal{B})(\mathbf{a}) = \mathcal{A}(\mathcal{B}(\mathbf{a}))$
$\mathcal{A}(\mathbf{a}; \Theta)$	Função de $\mathbf{a}$ parametrizada por $\Theta$ ; alternativamente, com o intuito de simplificar a notação, pode-se escrever somente $\mathcal{A}(\mathbf{a})$ , omitindo o conjunto de parâmetros $\Theta$
$\frac{d\mathcal{A}}{da}$	Derivada de $\mathcal{A}$ em relação a $a$
$\frac{\partial \mathcal{A}}{\partial a_i}$	Derivada parcial de $\mathcal{A}$ em relação a $a_i$
$\nabla \mathcal{A}(\mathbf{a})$	Gradiente da função $\mathcal{A}$ em $\mathbf{a}$ : $\left( \frac{\partial \mathcal{A}(\mathbf{a})}{\partial a_1}, \dots, \frac{\partial \mathcal{A}(\mathbf{a})}{\partial a_n} \right)$ ; alternativamente, pode-se denotar $\frac{\partial \mathcal{A}(\mathbf{a})}{\partial \mathbf{a}}$ e empregar a expressão “a derivada de $\mathcal{A}$ em relação a $\mathbf{a}$ ” para referir-se ao gradiente em questão
$\int_a^b \mathcal{A}(x) dx$	Integral definida de $\mathcal{A}$ no intervalo $[a, b]$
$\int_{-\infty}^{\infty} \mathcal{A}(x) dx$	$\lim_{a \rightarrow -\infty} \int_a^b \mathcal{A}(x) dx + \lim_{c \rightarrow \infty} \int_b^c \mathcal{A}(x) dx$ para qualquer $b \in \mathbb{R}$ (i.e., a integral definida de $\mathcal{A}$ no intervalo $(-\infty, \infty)$ )

<sup>2</sup>Nesta tese, os termos função, operação e transformação são utilizados como sinônimos.

$\arg \min_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\}$	Vetor $\mathbf{a}$ em $\Omega$ tal que, para todo $\mathbf{a}'$ em $\Omega$ , $\mathcal{A}(\mathbf{a}) \leq \mathcal{A}(\mathbf{a}')$ (i.e., um ponto de mínimo global da função objetivo $\mathcal{A}$ sobre o conjunto de pontos viáveis $\Omega$ ); alternativamente, pode-se escrever $\arg \min_{\mathbf{a}} \{\mathcal{A}(\mathbf{a})\}$ sujeito a $\mathbf{a} \in \Omega$ , a fim de evidenciar o conjunto de pontos viáveis
$\arg \max_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\}$	Ponto de máximo global de $\mathcal{A}$ sobre $\Omega$ , ou seja, $\arg \min_{\mathbf{a} \in \Omega} \{-\mathcal{A}(\mathbf{a})\} = \arg \max_{\mathbf{a}} \{\mathcal{A}(\mathbf{a})\}$ sujeito a $\mathbf{a} \in \Omega$
$\min_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\}$	Valor mínimo global de $\mathcal{A}$ sobre $\Omega$ : $\mathcal{A} \left( \arg \min_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\} \right) = \min_{\mathbf{a}} \{\mathcal{A}(\mathbf{a})\}$ sujeito a $\mathbf{a} \in \Omega$
$\max_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\}$	Valor máximo global de $\mathcal{A}$ sobre $\Omega$ : $\mathcal{A} \left( \arg \max_{\mathbf{a} \in \Omega} \{\mathcal{A}(\mathbf{a})\} \right) = \max_{\mathbf{a}} \{\mathcal{A}(\mathbf{a})\}$ sujeito a $\mathbf{a} \in \Omega$
$\mathbf{a} \cdot \mathbf{b}$	Produto escalar entre os vetores $\mathbf{a}$ e $\mathbf{b}$ ; alternativamente, uma vez que os vetores $\mathbf{a}$ e $\mathbf{b}$ podem ser vistos como matrizes colunas, pode-se escrever $\mathbf{a}^T \mathbf{b}$
$ a $	Valor absoluto de $a$
$\ \mathbf{a}\ _p$	Norma $\ell_p$ de $\mathbf{a}$
$\ \mathbf{A}\ _F$	Norma matricial de Frobenius computada sobre $\mathbf{A}$
$\det(\mathbf{A})$	Determinante de $\mathbf{A}$
$\mathbf{A}^T$	Matriz transposta de $\mathbf{A}$
$\mathbf{A}^{-1}$	Matriz inversa de $\mathbf{A}$
$\mathbf{A}^\dagger$	Matriz pseudo-inversa de $\mathbf{A}$
$\text{card}(\mathbb{A})$	Quantidade de elementos de $\mathbb{A}$ (i.e., a cardinalidade de $\mathbb{A}$ )
$\mathbb{1}_{\{\text{cond}\}}$	Função que resulta em 1, se a condição $\text{cond}$ for verdadeira, ou resulta em 0, se $\text{cond}$ for falsa (i.e., uma função indicadora)

## Probabilidade e Teoria da Informação

$\mathbb{P}(\mathbb{A})$	Probabilidade de ocorrência do evento $\mathbb{A}$
--------------------------	--

$a \sim \mathbb{D}$	Uma variável aleatória $a$ que segue uma distribuição $\mathbb{D}$
$\mathbb{G}(\mathbf{a}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Distribuição Gaussiana, sobre $\mathbf{a}$ , parametrizada pelo vetor de médias $\boldsymbol{\mu}$ e pela matriz de covariâncias $\boldsymbol{\Sigma}$
$\mathbb{E}(\mathbf{a})$	Valor esperado de $\mathbf{a}$
$\mathbb{H}(\mathbf{a})$	Entropia de $\mathbf{a}$
$\mathbb{I}(\mathbf{a}, \mathbf{b})$	Informação mútua entre $\mathbf{a}$ e $\mathbf{b}$

### Símbolos Diversos e de Particular Interesse Nesta Tese

$\mathcal{F}$	Uma imagem na forma de uma função
$\mathbf{I}$	Uma imagem na forma de uma matriz
$\mathcal{I}$	Uma imagem na forma de um tensor tridimensional
$\mathbf{p} = (l, c)$	Uma posição (ou localização) na imagem
$\mathbf{v} = \mathcal{F}(\mathbf{p})$	O valor da imagem na posição $\mathbf{p}$
$n_l$	Número de linhas de uma imagem
$n_c$	Número de colunas de uma imagem
$n_b$	Número de canais (ou bandas) de uma imagem
$n_q$	Número de níveis de quantização em uma imagem
$\mathbf{R}, \mathbf{G}$ e $\mathbf{B}$	Canais de uma imagem em RGB
$\mathcal{C}$	Modelo de classificação
$\mathbb{F}$	Conjunto de amostras rotuladas
$\mathbf{f}^{(i)}$	$i$ -ésima amostra (i.e., $i$ -ésimo <i>feature vector</i> )
$r^{(i)}$	$i$ -ésimo rótulo de classe
$n_i$	Número de imagens
$n_d$	Número de dimensões
$n_r$	Número de rótulos

$\mathcal{F}$	Função de redução de dimensionalidade
$L = I \odot R$	Um canal com informação de luminância, fatorado em termos da iluminância $I$ e da refletância $R$ ( $\odot$ denota o produto matricial elemento a elemento – i.e., o produto de Hadamard)
$\mathcal{O}$	Conjunto de pixels que correspondem ao objeto (i.e., à lesão) de interesse
$M^*$	Imagem binária (i.e., máscara) com o resultado de segmentação desejado
$M$	Máscara produzida por algum método de segmentação
$\mathcal{Y}$	Conjunto de amostras não rotuladas
$y_i$	$i$ -ésima amostra não rotulada
$\mathcal{I}$	Conjunto de imagens rotuladas
$\mathcal{S}$	Modelo de segmentação
$y$	Amostra na forma de um vetor
$\mathcal{D}$	Dicionário na forma de um conjunto
$D$	Dicionário na forma matricial
$x$	Representação de uma amostra
$\mathcal{R}$	Função de regularização
$\epsilon$	Tolerância ao erro de reconstrução

## LISTA DE FIGURAS

Figura 1.1 Exemplos de imagens macroscópicas de lesões melanocíticas. ....	21
Figura 2.1 Canais RGB de uma imagem macroscópica de lesão melanocítica. ....	25
Figura 2.2 Fluxo de trabalho de um sistema tradicional de processamento e análise de imagens macroscópicas de lesões melanocíticas. ....	29
Figura 2.3 Topologia da DCNN VGGNet-16. ....	43
Figura 2.4 Exemplos de imagens dermatoscópicas de lesões melanocíticas em comparação com as imagens macroscópicas correspondentes. ....	46
Figura 4.1 Fluxo de trabalho do método proposto para a segmentação de lesões melanocíticas em imagens macroscópicas. ....	77
Figura 4.2 Exemplos de resultados obtidos ao longo do pré-processamento. ....	79
Figura 4.3 Exemplos de regiões retangulares avaliadas para produzir imagens de intensidades. ....	81
Figura 4.4 Exemplos de imagens de intensidades obtidas. ....	82
Figura 4.5 Exemplos de dicionários iniciais obtidos a partir de imagens de intensidades. ....	87
Figura 4.6 Exemplos de resultados obtidos ao longo do pós-processamento. ....	90
Figura 4.7 Fluxo de trabalho do método proposto para a classificação de lesões melanocíticas em imagens macroscópicas. ....	91
Figura 5.1 Ranqueamento médio de diferentes configurações do método de segmentação proposto. ....	99
Figura 5.2 Ranqueamento médio do método de segmentação proposto em comparação com versões alternativas do mesmo. ....	102
Figura 5.3 Ranqueamento médio do método de segmentação proposto em comparação com métodos representativos do estado-da-arte (resultados dos experimentos com a base DermNet). ....	103
Figura 5.4 Ranqueamento médio do método de segmentação proposto em comparação com métodos representativos do estado-da-arte (resultados dos experimentos com a base DermQuest). ....	104
Figura 5.5 Ranqueamento médio do método de classificação proposto em comparação com versões alternativas do mesmo. ....	109
Figura 5.6 Ranqueamento médio do método de classificação proposto, nos testes com as bases de <a href="#">Alcón et al. (2009)</a> e <a href="#">Brinker et al. (2019)</a> , em comparação com alternativas representativas do estado-da-arte e diagnósticos realizados por dermatologistas (humanos) especializados. ....	110
Figura 5.7 Matrizes de confusão correspondentes aos testes com o método ARL. ....	111
Figura 5.8 Matrizes de confusão correspondentes aos testes com o método DW. ....	112
Figura 5.9 Matrizes de confusão correspondentes aos testes com o método IW. ....	112
Figura 5.10 Matrizes de confusão correspondentes aos testes com o método SLA. ....	112
Figura 5.11 Matrizes de confusão correspondentes aos testes com o método proposto. ....	113

## LISTA DE TABELAS

Tabela 2.1	Camadas da DCNN VGGNet-16.....	43
Tabela 5.1	Detalhamento sobre as bases de imagens macroscópicas de lesões melanocíticas empregadas nos experimentos.....	96
Tabela 5.2	Ranqueamento médio de diferentes configurações do método de classificação proposto (experimento com as configurações selecionadas). .....	106
Tabela A.1	Ranqueamento médio de diferentes configurações do método de classificação proposto (experimento com todas as configurações). .....	133

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>19</b>
<b>2 PROCESSAMENTO E ANÁLISE DE IMAGENS DE LESÕES MELANOCÍTICAS: CONCEITOS, MÉTODOS E ESTADO-DA-ARTE</b> .....	<b>23</b>
<b>2.1 Conceitos Fundamentais</b> .....	<b>23</b>
<b>2.2 Sistemas de Processamento e Análise de Imagens Macroscópicas de Lesões Melanocíticas</b> .....	<b>26</b>
2.2.1 Métodos de Pré-Processamento.....	30
2.2.2 Métodos de Segmentação.....	31
2.2.3 Métodos de Classificação.....	38
<b>2.3 Estado-da-Arte</b> .....	<b>45</b>
<b>2.4 Considerações Finais</b> .....	<b>47</b>
<b>3 REPRESENTAÇÕES ESPARSAS ATRAVÉS DE DICIONÁRIOS: CONCEITOS E MÉTODOS</b> .....	<b>48</b>
<b>3.1 Conceitos Fundamentais</b> .....	<b>48</b>
3.1.1 Representação de Imagens em Bases Ortonormais.....	48
3.1.2 Representações Esparsas de Imagens via Dicionários.....	50
<b>3.2 Métodos para Obter Representações Esparsas e Dicionários</b> .....	<b>53</b>
3.2.1 <i>Orthogonal Matching Pursuit</i> (OMP).....	57
3.2.2 <i>Fast Iterative Shrinkage-Thresholding Algorithm</i> (FISTA).....	58
3.2.3 <i>Expectation-Maximization</i> e <i>Gaussian Mixture Models</i> (GMMs).....	62
3.2.4 <i>Method of Optimal Directions</i> (MOD) e <i>K-Singular Value Decomposition</i> (K-SVD).....	64
3.2.5 <i>Nonnegative Matrix Factorization</i> (NMF).....	68
3.2.6 <i>Information-Theoretic Dictionary Learning</i> (ITDL).....	70
3.2.7 <i>Sparse Representation-based Classification</i> (SRC).....	72
<b>3.3 Considerações Finais</b> .....	<b>74</b>
<b>4 REPRESENTAÇÕES ESPARSAS PARA PROCESSAMENTO E ANÁLISE DE IMAGENS MACROSCÓPICAS DE LESÕES MELANOCÍTICAS</b> .....	<b>75</b>
<b>4.1 Representações Esparsas para Segmentação de Imagens Macroscópicas de Lesões Melanocíticas</b> .....	<b>75</b>
4.1.1 Pré-processamento da Imagem de Entrada.....	77
4.1.2 Obtenção de uma Imagem de Intensidades.....	79
4.1.3 Aprendizado de um Dicionário usando UITDL.....	83
4.1.4 Agrupamento dos Blocos e Pós-Processamento.....	88
<b>4.2 Representações Esparsas para Classificação de Imagens Macroscópicas de Lesões Melanocíticas</b> .....	<b>90</b>
4.2.1 Aprendizado dos Dicionários de Características Profundas.....	92
4.2.2 Predição usando o SRC Modificado.....	93
<b>4.3 Considerações Finais</b> .....	<b>95</b>
<b>5 RESULTADOS EXPERIMENTAIS E DISCUSSÃO</b> .....	<b>96</b>
<b>5.1 Experimentos com o Método Proposto para Segmentação de Imagens Macroscópicas de Lesões Melanocíticas</b> .....	<b>98</b>
5.1.1 Seleção dos Hiperparâmetros de Configuração.....	98
5.1.2 Estudo de Ablação.....	100
5.1.3 Comparação com o Estado-da-Arte.....	102

<b>5.2 Experimentos com o Método Proposto para Classificação de Imagens</b>	
<b>Macroscópicas de Lesões Melanocíticas.....</b>	<b>104</b>
5.2.1 Seleção dos Hiperparâmetros de Configuração .....	104
5.2.2 Estudo de Ablação .....	108
5.2.3 Comparação com o Estado-da-Arte .....	109
<b>6 CONCLUSÃO.....</b>	<b>114</b>
<b>REFERÊNCIAS.....</b>	<b>115</b>
<b>APÊNDICE A — RESULTADOS SUPLEMENTARES .....</b>	<b>133</b>

## 1 INTRODUÇÃO

O melanoma, um tumor maligno que começa nos melanócitos<sup>1</sup> (i.e., uma lesão melanocítica maligna), embora possua um prognóstico favorável se detectado nos estágios iniciais, é o tipo mais letal de câncer de pele, uma vez que é mais propenso à metástase (KAUFMAN, 2016; RIKER, 2018). Especificamente, a taxa de pacientes que sobrevivem pelo menos cinco anos após o diagnóstico dessa doença no estágio inicial é superior a 99%. No entanto, essa taxa diminui para cerca de 30% se a detecção ocorre somente no último estágio (ACS, 2022).

O diagnóstico precoce do melanoma é essencial não só para reduzir o número de vítimas como também os custos envolvidos com o tratamento da doença (MORGAN et al., 2018; BRASIL, 2019). Por isso, inúmeras pesquisas promissoras têm sido apresentadas com o intuito de prover técnicas mais exatas de discriminação entre melanomas e nevos<sup>2</sup> benignos, por exemplo, adaptando alguma forma de exame genético (CLARKE et al., 2015; CLARKE et al., 2017), microscopia confocal de reflectância (CARRERA; MARGHOOB, 2016), tomografia de coerência óptica (TURANI et al., 2019), espectroscopia de massa (LAZOVA et al., 2020) ou expressão de microRNAs (TORRES et al., 2020).

A despeito dos avanços tecnológicos recentes, a triagem clínica de lesões melanocíticas continua desempenhando um papel de fundamental importância e é tipicamente respaldada por alguma lista de verificação médica tal como a “regra ABCD” (FRIEDMAN; RIGEL; KOPF, 1985; JOHR, 2002; GLAZER et al., 2017). Essencialmente, ABCD é um acrônimo para Assimetria, Borda, Cor e Diâmetro, os quais são aspectos morfológicos que ordinariamente viabilizam a distinção entre melanomas e nevos. Infelizmente, certos nevos, ditos atípicos, possuem características visuais de melanomas, o que dificulta consideravelmente a análise clínica (MITSUI et al., 2016).

Para lidar com esse tipo de situação, dermatologistas costumam utilizar uma técnica não invasiva, denominada dermatoscopia ou microscopia de epiluminescência, a qual consiste em empregar um equipamento especializado denominado dermatoscópio para controlar a iluminação da cena e obter uma visuali-

---

<sup>1</sup>Células localizadas na camada inferior da epiderme, as quais produzem a melanina pigmentada, responsável pela coloração da pele, dos cabelos e dos olhos.

<sup>2</sup>O termo médico “nevo” é empregado para denominar as lesões de pele informalmente conhecidas como “sinais” ou “pintas”.

zação amplificada da lesão, permitindo assim a verificação de determinadas estruturas características de melanomas e impossíveis de identificar a olho nu (WEBER et al., 2018). Entretanto, cabe salientar que em regiões remotas, bem como em países subdesenvolvidos ou em desenvolvimento (e.g., Brasil), raramente um dermatoscópio é utilizado na saúde primária e uma consulta com um dermatologista pode exigir um longo tempo de espera.

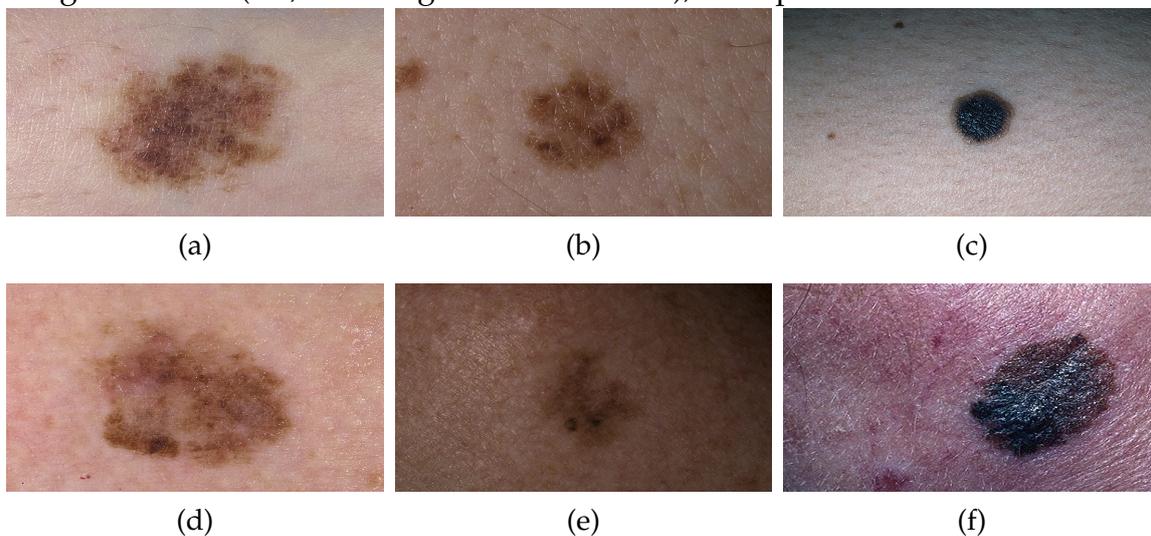
Se uma determinada lesão é considerada suspeita, pode ser recomendada a realização de uma biópsia excisional, seguida por um exame histopatológico (SWETTER et al., 2019). Embora tal exame não seja infalível, seu diagnóstico é amplamente reconhecido como uma espécie de verdade absoluta (CLARKE et al., 2017). Todavia, biópsias de lesões benignas implicam em despesas desnecessárias. Ademais, biópsias, de um modo geral, são desconfortáveis e, esporadicamente, podem resultar em complicações mais graves tais como hemorragia ou alergia aos anestésicos (NELSON et al., 2019).

Neste contexto, vários sistemas têm sido propostos para a pré-triagem de lesões melanocíticas partir de uma simples fotografia digital (i.e., uma imagem macroscópica ou clínica) (CAVALCANTI, 2013). Esses sistemas podem ser aplicados em diferentes cenários como, por exemplo,

- no acompanhamento teledermatológico de pacientes que já tiveram melanoma ou possuem a síndrome do nevo atípico (MELAMED et al., 2017);
- na triagem de pacientes em unidades básicas de saúde ou em consultas com um clínico geral;
- como ferramenta didática em cursos de dermatologia;
- para auxiliar o diagnóstico de um dermatologista.

Contudo, projetar um sistema que seja capaz de distinguir melanomas de determinadas lesões melanocíticas benignas, conhecidas como nevos atípicos, é uma tarefa complexa de processamento e análise de imagens. A Figura 1.1, a qual apresenta três amostras da classe “nevo atípico” (Subfiguras 1.1a, 1.1c e 1.1e) e três amostras da classe “melanoma” (Subfiguras 1.1b, 1.1d e 1.1f), está disposta de uma maneira que deve evidenciar ao leitor que resolver o problema em questão não é trivial, mesmo para um especialista, não só por causa da variabilidade entre

Figura 1.1: Exemplos de imagens macroscópicas de lesões melanocíticas. A primeira linha (i.e., as Subfiguras 1.1a até 1.1c) apresenta exemplos de nevos atípicos; a segunda linha (i.e., as Subfiguras 1.1d até 1.1f), exemplos de melanomas.



Fonte: Alcón et al. (2009).

imagens de uma mesma classe, mas também pelas similaridades visuais entre imagens das diferentes classes.

Cabe mencionar que sistemas análogos têm sido projetados para operar com base em diferentes técnicas de imageamento (MAGALHAES; MENDES; VARDASCA, 2019). Em particular, nos últimos anos, sistemas que operam a partir de uma imagem obtida por meio de um dermatoscópio digital (i.e., uma imagem dermatoscópica) vêm sendo bastante investigados (SCHARCANSKI; CELEBI, 2014; MARCHETTI et al., 2018; CELEBI; CODELLA; HALPERN, 2019). Porém, deve-se salientar que mesmo dermatoscópios profissionais de entrada podem custar mais de mil euros (MICALI; LACARRUBBA, 2018). Logo, sistemas baseados em imagens dermatoscópicas possuem uma gama de aplicações mais restrita do que aqueles baseados em imagens macroscópicas (PACHECO; KROHLING, 2019).

Neste cenário, a realização deste trabalho foi motivada pelo objetivo geral de contribuir com o desenvolvimento de soluções que possam auxiliar na triagem de lesões melanocíticas e, em especial, no diagnóstico precoce de casos de melanoma. Para isso, o objetivo específico é a proposição de um sistema completo de pré-triagem de lesões melanocíticas com base em imagens macroscópicas, o qual compreende:

- um método não supervisionado de segmentação de lesões melanocíticas,

baseado no particionamento das representações dos blocos quadrados da imagem de entrada em um dicionário obtido empregando uma versão adaptativa e não supervisionada do método ITDL (*Information-Theoretic Dictionary Learning*) (QIU; PATEL; CHELLAPPA, 2014; FLORES; SCHARCANSKI, 2016);

- um método supervisionado de classificação, o qual é baseado em uma representação esparsa da imagem segmentada de teste em um “dicionário de características profundas” e que pode ser visto como uma versão modificada, mais escalável e robusta para lidar com o problema do desbalanceamento entre classes, do método SRC (*Sparse Representation-based Classification*) (WRIGHT et al., 2009; FLORES; ZORTEA; SCHARCANSKI, 2019).

Outra contribuição relevante a ser considerada, além da exposição, da fundamentação, da implementação e da análise dos métodos desenvolvidos, é a revisão sobre a literatura de processamento e análise de imagens macroscópicas de lesões melanocíticas. Em linhas gerais, a principal hipótese desta pesquisa é que o sistema proposto, baseado em representações esparsas através de dicionários, pode alcançar um desempenho preditivo equiparável aos desempenhos observados em alternativas apresentadas recentemente na literatura, mesmo sendo treinado com uma quantidade relativamente limitada de exemplos.

Esta tese está organizada conforme delineado a seguir. Subsequentemente a esta introdução do Capítulo 1, no Capítulo 2, encontra-se o levantamento bibliográfico sobre os trabalhos relacionados ao tema desta pesquisa (i.e., processamento e análise de imagens de lesões melanocíticas). Em seguida, no Capítulo 2, está a fundamentação teórica, com um enfoque nos métodos de representações esparsas e aprendizado de dicionário empregados no desenvolvimento do sistema proposto. No Capítulo 4, por sua vez, está detalhado o sistema de processamento e análise de imagens macroscópicas de lesões melanocíticas proposto. Na sequência, no Capítulo 5, estão reportados os resultados dos experimentos realizados e, por fim, no Capítulo 6, as conclusões a respeito deste trabalho.

## 2 PROCESSAMENTO E ANÁLISE DE IMAGENS DE LESÕES MELANOCÍ- TICAS: CONCEITOS, MÉTODOS E ESTADO-DA-ARTE

Neste capítulo, é apresentada uma síntese da literatura relacionada aos sistemas de processamento e análise de lesões melanocíticas. Especificamente, na Seção 2.1, são estabelecidas algumas definições básicas, notações e terminologias consideradas no decorrer de todo este trabalho e, em especial, no restante deste capítulo. Na Seção 2.2, por sua vez, são tratados métodos que costumam compor os sistemas em questão, os quais estão organizados de acordo com o seu propósito no fluxo de operação do sistema: pré-processamento (Subseção 2.2.1); segmentação (Subseção 2.2.2); classificação (Subseção 2.2.3). Métodos representativos do estado-da-arte são tratados separadamente na Seção 2.3. Por fim, as considerações finais sobre este capítulo podem ser encontradas na Seção 2.4.

### 2.1 Conceitos Fundamentais

Ao longo tese, os termos “imagem” e “imagem digital” são utilizados indistintamente para designar dados ou sinais que podem ser descritos por meio de uma função na forma:

$$\begin{aligned} \mathcal{F} &: \{1, \dots, n_l\} \times \{1, \dots, n_c\} \rightarrow \{0, \dots, n_q - 1\}^{n_b} \\ &= \{(\mathbf{p}, \mathbf{v}) : \mathbf{p} \in \{1, \dots, n_l\} \times \{1, \dots, n_c\} \text{ e } \mathbf{v} \in \{0, \dots, n_q - 1\}^{n_b}\}, \end{aligned} \quad (2.1)$$

onde  $\mathbf{v}$  pode ser dito o valor de  $\mathcal{F}$  na posição  $\mathbf{p}$  (i.e.,  $\mathbf{v} = \mathcal{F}(\mathbf{p})$ ) e cada  $(\mathbf{p}, \mathbf{v}) \in \mathcal{F}$  pode ser dito um pixel (do inglês, *picture element*) (GONZALEZ; WOODS, 2010).

Do ponto de vista estrutural, uma imagem pode ser caracterizada pelas quantidades  $n_l, n_c, n_b, n_q \in \mathbb{N}$ , cujas denominações utilizadas nesta tese são apresentadas a seguir:

- $n_l$  – número de linhas da imagem;
- $n_c$  – número de colunas da imagem;
- $n_b$  – número de bandas (ou canais) da imagem;

- $n_q$  – número de níveis de quantização nos canais da imagem.<sup>1</sup>

Se  $n_b = 1$ , a imagem pode referida como monocanal. Por outro lado, se  $n_b > 1$ , a imagem pode ser referida como multicanal.

De qualquer modo, explicitando  $\mathbf{p} = (l, c)$  e  $\mathbf{v} = (v_1, \dots, v_{n_b})$ , pode-se decompor cada pixel  $(\mathbf{p}, \mathbf{v}) = ((l, c), \mathbf{v}) \in \mathcal{F}$  em termos de um conjunto composto por  $n_b$  pixels básicos:  $\{((l, c), v_b)\}_{b=1}^{n_b}$ . Logo, a imagem  $\mathcal{F}$  pode ser decomposta em  $n_b$  canais (i.e., imagens monocanal),  $\mathcal{F}^{[b]} = \{((l, c), v_b) : (l, c) \in \{1, \dots, n_l\} \times \{1, \dots, n_c\} \text{ e } v_b \in \{0, \dots, n_q - 1\}\}$ ,  $b = 1, \dots, n_b$ , de modo que o  $b$ -ésimo canal de  $\mathcal{F}$  pode ser representado em um computador por meio da seguinte matriz (MARQUES FILHO; VIEIRA NETO, 1999):

$$\mathcal{I}^{[b]} = \begin{bmatrix} \mathcal{F}^{[b]}(1,1) & \mathcal{F}^{[b]}(1,2) & \dots & \mathcal{F}^{[b]}(1,c) & \dots & \mathcal{F}^{[b]}(1,n_c) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathcal{F}^{[b]}(l,1) & \mathcal{F}^{[b]}(l,2) & \dots & \mathcal{F}^{[b]}(l,c) & \dots & \mathcal{F}^{[b]}(l,n_c) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathcal{F}^{[b]}(n_l,1) & \mathcal{F}^{[b]}(n_l,2) & \dots & \mathcal{F}^{[b]}(n_l,c) & \dots & \mathcal{F}^{[b]}(n_l,n_c) \end{bmatrix} \in \mathbb{R}^{n_l \times n_c}. \quad (2.2)$$

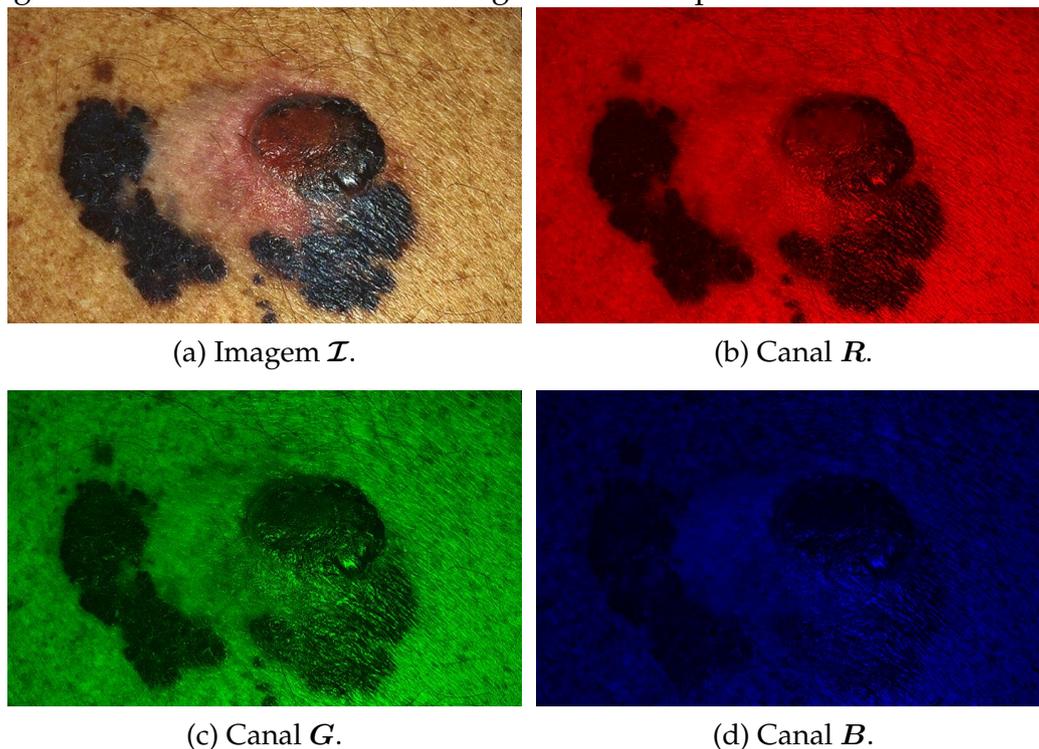
Note que, na Eq. (2.2),  $\mathcal{F}^{[b]}(l, c) \in \{0, \dots, n_q - 1\}$  para qualquer  $(l, c) \in \{1, \dots, n_l\} \times \{1, \dots, n_c\}$ . Consequentemente, poderia-se escrever  $\mathcal{I}^{[b]} \in \{0, \dots, n_q - 1\}^{n_l \times n_c} \subset \mathbb{N}^{n_l \times n_c} \subset \mathbb{R}^{n_l \times n_c}$ . Contudo, nessa equação, e no restante desta tese, optou-se por utilizar a notação simplificada  $\mathcal{I}^{[b]} \in \mathbb{R}^{n_l \times n_c}$ , a qual omite o detalhamento dos níveis de quantização, para indicar que  $\mathcal{I}^{[b]}$  é uma matriz com  $n_l$  linhas e  $n_c$  colunas (i.e., um tensor bidimensional). De maneira análoga, utiliza-se  $\mathcal{I} \in \mathbb{R}^{n_l \times n_c \times n_b}$  para indicar que a imagem  $\mathcal{I}$  é um tensor tridimensional com  $n_l$  linhas,  $n_c$  colunas e  $n_b$  bandas (VELHO; FRERY; GOMES, 2009).

Uma imagem multicanal cujos canais foram obtidos a partir de uma mesma cena, mas cada um deles com um conjunto de sensores sensíveis a uma diferente faixa do espectro eletromagnético, pode ser dita multiespectral. Em especial, se as faixas empregadas correspondem a bandas de cores, a imagem multiespectral pode ser dita colorida e cada uma dos seus canais, cujos valores podem ser referidos como intensidades, pode ser referido como uma imagem monocromá-

<sup>1</sup>Nesta tese, assume-se que a quantidade de níveis de quantização é sempre a mesma em todos os canais da imagem.

tica (PETROU; PETROU, 2010).<sup>2</sup> Por exemplo, considere a imagem colorida  $\mathcal{I}$  apresentada no seu monitor na Subfigura 2.1a (i.e., uma imagem macroscópica de lesão melanocítica), a qual está em RGB de modo que os seus canais, ilustrados nas Subfiguras 2.1b até 2.1d, contêm, respectivamente, as intensidades de vermelho (R – *Red*), verde (G – *Green*) e azul (B – *Blue*) capturadas em cada posição da imagem.

Figura 2.1: Canais RGB de uma imagem macroscópica de lesão melanocítica.



Fonte: Alcón et al. (2009).

Ao longo desta tese, o termo “processamento de imagem” refere-se a uma sequência de operações executadas por um computador a partir de uma dada imagem, a qual supõe-se que tenha sido obtida a partir de uma cena que contenha um único objeto como foco da atenção (i.e., um único objeto de interesse). Sistemas que processam uma imagem a fim de extrair informações úteis para suportar alguma tomada de decisão (e.g., a identificação de uma lesão melanocítica como “melanoma” ou “nevo atípico”), nesta tese, são denominados “sistemas de processamento e análise de imagens”.

<sup>2</sup>Em particular, uma imagem monocromática com somente duas intensidades pode ser dita binária.

## 2.2 Sistemas de Processamento e Análise de Imagens Macroscópicas de Lesões Melanocíticas

Sistemas de processamento e análise de imagens, em última instância, costumam desempenhar uma tarefa de reconhecimento (ou classificação) de padrões, sob a perspectiva do aprendizado automático supervisionado.<sup>3</sup> Nesse contexto, um método de aprendizado supervisionado é empregado para ajustar uma função  $\mathcal{C} : \mathbb{R}^{n_d} \rightarrow \{1, \dots, n_r\}$  (i.e., um modelo de classificação) com base em um conjunto de pares ordenados  $\mathbb{F} = \{(\mathbf{f}^{(i)}, r^{(i)})\}_{i=1}^{n_i}$ , os quais assume-se que sejam independentes e identicamente distribuídos (DUDA; HART; STORCK, 2001; MURPHY, 2012).

O conjunto de pares ordenados utilizado para ajustar  $\mathcal{C}$ ,  $\mathbb{F}$ , é dito um conjunto de amostras rotuladas. Nesse conjunto, cada amostra  $\mathbf{f}^{(i)} = \mathcal{F}(\mathcal{I}^{(i)}) \in \mathbb{R}^{n_d}$  é um vetor de características (do inglês, *feature vector*) obtido aplicando uma função  $\mathcal{F} : \mathbb{R}^{n_l \times n_c \times n_b} \rightarrow \mathbb{R}^{n_d}$  sobre uma imagem  $\mathcal{I}^{(i)} \in \mathbb{R}^{n_l \times n_c \times n_b}$  cuja classe do objeto de interesse está codificada pelo rótulo  $r^{(i)} \in \{1, \dots, n_r\}$ . O objetivo é que, após o ajuste de  $\mathcal{C}$ , a função composta  $(\mathcal{C} \circ \mathcal{F})(\mathcal{I}) = \mathcal{C}(\mathcal{F}(\mathcal{I}))$  possa ser usada para prever a classe do objeto de interesse em uma dada imagem  $\mathbb{R}^{n_l \times n_c \times n_b} \ni \mathcal{I} \notin \{\mathcal{I}^{(i)}\}_{i=1}^{n_i}$  (FUKUNAGA, 1990; BISHOP, 2006).

Com base em Hastie, Tibshirani e Friedman (2009), a densidade correspondente a  $n_1$  vetores uniformemente distribuídos em um intervalo de  $\mathbb{R}$ , com tamanho  $\ell$ , é  $\rho = n_1/\ell$ . Observe que para manter essa mesma densidade em um quadrado no  $\mathbb{R}^2$  ou em um cubo no  $\mathbb{R}^3$ , ambos com lado  $\ell$ , são necessários  $n_2 = n_1^2 = (\rho \cdot \ell)^2$  e  $n_3 = n_1^3 = (\rho \cdot \ell)^3$  vetores distribuídos nas respectivas área e volume. De uma forma geral, pode-se reproduzir tal densidade em um hipercubo no  $\mathbb{R}^{n_d}$ , com lado  $\ell$ , inserindo  $n_i = n_1^{n_d} = (\rho \cdot \ell)^{n_d}$  vetores (e.g.,  $\{\mathbf{f}^{(i)} \in \mathbb{R}^{n_d}\}_{i=1}^{n_i}$ ) no hipervolume em questão. Portanto,  $\rho = (1/\ell)n_i^{(1/n_d)}$ , de modo que quanto maior a dimensionalidade  $n_d$  dos  $n_i$  vetores dados, menos densa (i.e., mais esparsa) tende a ser a sua distribuição.

Adicionalmente, suponha que os vetores  $\{\mathbf{f}^{(i)}\}_{i=1}^{n_i}$  estejam uniformemente distribuídos em um hipercubo de lado unitário no  $\mathbb{R}^{n_d}$ . Assim, o hipercubo de lado  $\ell < 1$  com os  $n_v < n_i$  vetores mais próximos de  $\mathcal{F}(\mathcal{I})$  ocupa uma fração  $n_v/n_i$

<sup>3</sup>O aprendizado automático pode ser visto como uma abordagem da inteligência artificial, a qual também é conhecida como aprendizado de máquina (do inglês, *machine learning*) (MIT-CHILL, 1997; FACELI et al., 2021).

do hipervolume unitário, o que implica  $\ell = (n_v/n_i)^{(1/n_d)}$  e, conseqüentemente,  $\lim_{n_d \rightarrow \infty} \ell = \lim_{n_d \rightarrow \infty} (n_v/n_i)^{(1/n_d)} = 1$ . Logo, quando  $n_d$  tende ao infinito, todos os  $n_v$  vetores mais próximos de  $\mathcal{F}(\mathcal{I})$  estão nos lados do hipercubo unitário,<sup>4</sup> onde também estão os  $n_i - n_v$  vetores mais distantes de  $\mathbf{f}$ . Ou seja, a noção de proximidade vai perdendo o seu significado à medida que a dimensionalidade aumenta (BEYER et al., 1999).

As situações descritas nos dois últimos parágrafos são manifestações de um fenômeno mais amplo, conhecido na literatura como “a maldição da dimensionalidade” (BELLMAN, 1961; HUGHES, 1968), o qual é particularmente crítico nos casos em que o objetivo é processar e analisar imagens. Por exemplo, as imagens macroscópicas da base DermQuest (AMELARD et al., 2014), uma das bases usadas nos experimentos descritos nesta tese, possuem, na média, aproximadamente 1,5 milhão de pixels. Conseqüentemente, a função  $\mathcal{F}$  não só desempenha o papel de padronizar o tamanho das amostras, mas também o de reduzir a dimensionalidade das imagens (por isso, tal função pode ser referida como função de redução de dimensionalidade).

Uma vez que o conteúdo de uma imagem normalmente é redundante, torna-se viável descrevê-la de maneira eficiente (DA SILVA; MENDONÇA, 2005; SCHULZ; SILVA; VELHO, 2009). Contudo, o uso de diferentes funções de redução de dimensionalidade  $\mathcal{F}$  pode revelar ou ocultar as diferentes informações contidas na imagem. Por exemplo, se a imagem de interesse  $\mathcal{I}$  pode ser aproximada pela combinação linear de alguns poucos elementos de um conjunto gerador  $\mathbf{G}$ , é possível obter uma representação esparsa  $\mathbf{f} = \mathcal{F}(\mathcal{I}) \simeq \mathbf{G}^\dagger \text{vet}(\mathcal{I})$ ,<sup>5</sup> a qual evidencia a estrutura essencial subjacente à imagem (i.e., o princípio das representações esparsas (OLSHAUSEN; FIELD, 1996), o qual será tratado em detalhes no Capítulo 3).

Dentre os outros princípios, os quais assim como o princípio das representações esparsas podem ser considerados para orientar o desenvolvimento da função  $\mathcal{F}$ , estão (BENGIO; COURVILLE; VINCENT, 2013):

- o princípio da organização hierárquica que sugere uma função de redução de dimensionalidade composta por várias etapas de extração de características, as quais progressivamente incrementam a informação semântica por

<sup>4</sup>Caso contrário, estariam em um hipercubo com lado  $\ell < 1$ .

<sup>5</sup> $\text{vet}(\cdot)$  denota o operador de vetorização que rearranja a entrada na forma de um arranjo unidimensional.

meio da representação de características mais abstratas em termos de características menos abstratas (FUKUSHIMA, 1980; MARR, 1982);

- o princípio dos agrupamentos naturais que supõe que se duas imagens possuem o mesmo rótulo de classe, então tendem a concentrar-se próximas a uma mesma região (e.g., um *manifold*), cuja dimensionalidade efetiva – intrínseca – é consideravelmente menor do que a dimensionalidade original das imagens (SEUNG; LEE, 2000);
- o princípio dos fatores compartilhados, o qual parte da premissa de que as características relevantes para uma determinada tarefa podem ser transferidas e/ou adaptadas para a realização de outras tarefas similares (PAN; YANG, 2009; YOSINSKI et al., 2014).

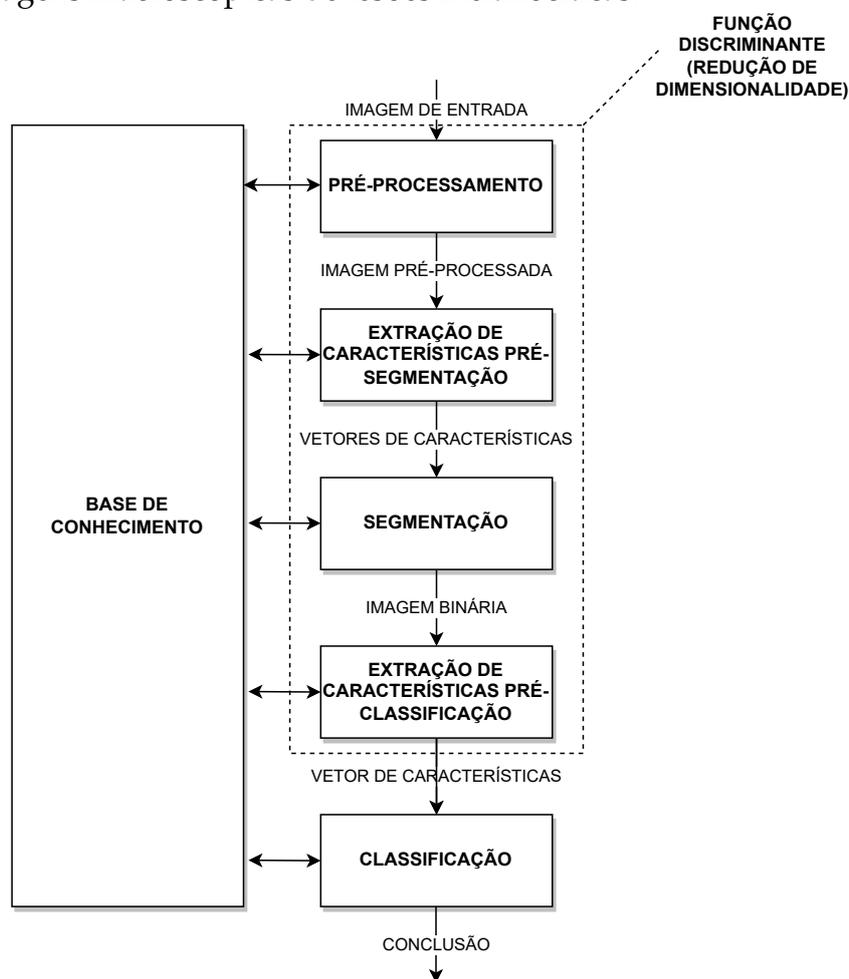
Embora os mesmos princípios normalmente norteiem o projeto da função de redução de dimensionalidade  $\mathcal{F}$ , são inúmeros os métodos que podem integrar tal função. Todavia, no caso dos sistemas tradicionais de processamento e análise de imagens macroscópicas de lesões melanocíticas, os quais têm sido projetados, desenvolvidos e investigados há pelo menos três décadas (PEREDNIA, 1991; STOECKER; MOSS, 1992; HALL; CLARIDGE; SMITH, 1995), a função  $\mathcal{F}$  é elaborada com base em diretrizes estabelecidas na dermatologia (e.g., regra ABCD) e engenhosidade humana<sup>6</sup> e, frequentemente, segue um fluxo de trabalho comum, o qual pode ser expresso pela forma geral  $\mathcal{F}(\cdot) = (\mathcal{F}_{ext} \circ \mathcal{F}_{seg} \circ \mathcal{F}_{pre})(\cdot) = \mathcal{F}_{ext}(\mathcal{F}_{seg}(\mathcal{F}_{pre}(\cdot)))$  que envolve:

- $\mathcal{F}_{pre} : \mathbb{R}^{n_l \times n_c \times n_b} \rightarrow \mathbb{R}^{n_l \times n_c \times n_b} \times \mathbb{R}^{n_l \times n_c \times n_b}$  (i.e., a etapa de pré-processamento), onde uma versão modificada da imagem de entrada é produzida para simplificar as etapas subsequentes;
- $\mathcal{F}_{seg} : \mathbb{R}^{n_l \times n_c \times n_b} \times \mathbb{R}^{n_l \times n_c \times n_b} \rightarrow \mathbb{R}^{n_l \times n_c \times n_b} \times \mathbb{R}^{n_l \times n_c \times n_b} \times \{0,1\}^{n_l \times n_c}$  (i.e., a etapa de segmentação), em que uma imagem binária é gerada com o objetivo de delimitar a lesão de interesse;
- $\mathcal{F}_{ext} : \mathbb{R}^{n_l \times n_c \times n_b} \times \mathbb{R}^{n_l \times n_c \times n_b} \times \{0,1\}^{n_l \times n_c} \rightarrow \mathbb{R}^{n_d}$  (i.e., a etapa de extração de características pré-classificação), cuja finalidade é representar a lesão segmentada por meio de um vetor de características conveniente.

<sup>6</sup>O comportamento dessas funções pode ser alterado por meio de hiperparâmetros de configuração.

O fluxo de trabalho completo de um sistema de processamento e análise de imagens macroscópicas de lesões melanocíticas típico (i.e., baseado em engenharia de características) é ilustrado no diagrama de blocos da Figura 2.2. Diversos métodos usualmente empregados na composição desses sistemas são apresentados a seguir, organizados de acordo com as etapas ilustradas na Figura 2.2. Concretamente, na Subseção 2.2.1, são discutidos métodos de pré-processamento. Em seguida, na Subseção 2.2.2, são apresentados métodos de segmentação (e extração de características pré-segmentação). Por fim, na Subseção 2.2.3, são tratados métodos de classificação (e extração de características pré-classificação).

Figura 2.2: Fluxo de trabalho de um sistema tradicional de processamento e análise de imagens macroscópicas de lesões melanocíticas.



Fonte: Os Autores.

Cabe ainda mencionar que métodos de extração de características, independentemente de precederem a segmentação ou a classificação, serão tratados em mais detalhes no Capítulo 3, no âmbito de representações esparsas através de dicionários. Tais representações, de particular interesse nesta tese, tendem

a produzir um resultado no qual somente alguns poucos coeficientes não nulos contribuem efetivamente na representação. Consequentemente, reduzem a dimensionalidade efetiva, a qual é determinada pelas posições dos coeficientes não nulos em questão.

### 2.2.1 Métodos de Pré-Processamento

Visto que imagens macroscópicas, normalmente, são obtidas sem nenhum tipo de controle sobre a fonte de iluminação, é necessário atenuar os eventuais efeitos de sombreamento e reflexo antes de efetivamente lidar com a segmentação (i.e., é necessário pré-processar a imagem). Em especial, os métodos de melhoramento de contraste e calibração de cores (e.g., métodos de constância de cor), os quais são ordinariamente empregados para pré-processar imagens dermatoscópicas, não estendem-se para imagens macroscópicas (BARATA; CELEBI; MARQUES, 2014; SCHARCANSKI; CELEBI, 2014). De fato, as estratégias convencionalmente adotadas, para mitigar os efeitos causados pela iluminação irregular, podem ser descritas em quatro passos gerais:

1. converter a imagem de entrada de RGB para um espaço de cores que centre a informação de luminância (i.e., a informação acromática) em um único canal;
2. estimar os efeitos da iluminação com base no canal de luminância;
3. remover os efeitos da iluminação estimados;
4. converter a imagem manipulada de volta para o espaço de cores RGB.

Primeiramente, considere  $L \in \mathbb{R}^{n_l \times n_c}$  um canal com informação de luminância, com as suas intensidades reescaladas de  $[0, n_q - 1]$  para  $[0, 1]$  (e.g., dividindo a intensidade original de cada pixel por  $n_q - 1$ ), o qual pode ser fatorado em termos da iluminância  $I \in \mathbb{R}^{n_l \times n_c}$  (i.e., a iluminação) e da refletância  $R \in \mathbb{R}^{n_l \times n_c}$ , conforme mostrado na equação a seguir:

$$L(l,c) = I(l,c)R(l,c), \quad (2.3)$$

em que  $L(l,c)$ ,  $I(l,c)$  e  $R(l,c)$  denotam os valores indexados por  $(l,c)$  em  $L$ ,  $I$  e  $R$ , respectivamente.

Note que se a iluminação  $I$  fosse conhecida, poderia-se remover os seus efeitos sobre  $L$  por meio de:

$$R(l,c) = \frac{L(l,c)}{I(l,c)}. \quad (2.4)$$

Contudo, a matriz  $I$  é desconhecida e, na prática, costuma ser estimada a partir de  $L$ .

Uma forma simples de realizar tal estimativa é assumindo que  $I$  corresponde à imagem  $L$  suavizada por morfologia matemática (e.g., aplicando um fechamento morfológico (MEYER, 1979)) ou filtragem com passa-baixas (veja, por exemplo, o trabalho de Alcón et al. (2009)). Todavia, é praticamente inviável selecionar um filtro (ou um elemento estruturante) que possa adaptar-se bem para diferentes imagens. Alternativamente, métodos mais robustos podem ser produzidos ajustando um modelo de  $I$  com base somente em alguns pixels de  $L$ . Exemplos representativos de estratégias que seguem esse tipo de abordagem podem ser encontrados nos trabalhos de Cavalcanti, Scharcanski e Lopes (2010), Glaister et al. (2013), Giotis et al. (2015) e Gautam et al. (2018).

## 2.2.2 Métodos de Segmentação

As características utilizadas para descrever as informações visuais (e.g., forma, cores e texturas) de uma lesão melanocítica, avaliadas pelo método de classificação para diferenciar um melanoma de um nevo atípico, são mensuradas com base no resultado da segmentação. No contexto desta tese, o objetivo da segmentação é identificar quais, dentre os pixels da imagem de entrada, correspondem ao objeto (i.e., à lesão) de interesse. O conjunto de tais pixels, ditos “de lesão”, pode ser denotado  $\mathbb{O}$  e referido como a ROI (do inglês, *Region of Interest* – região de interesse). O conjunto  $\mathbb{O}^c$  com os demais pixels, ditos “de não lesão”, os quais praticamente não possuem informações relevantes para distinguir um melanoma de um nevo atípico, pode ser referido como “o fundo” (do inglês, *background*) (HANCE et al., 1996; OLIVEIRA et al., 2016).

Equivalentemente, pode-se dizer que a finalidade da segmentação é produzir uma imagem binária  $M^* \in \{0,1\}^{n_l \times n_c}$  (i.e., uma máscara) na qual:

$$M^*(l,c) = \begin{cases} 1, & \text{se o pixel indexado por } (l,c), \text{ na imagem de entrada } \mathcal{I}, \text{ pertence} \\ & \text{ao conjunto } \mathcal{O}, \\ 0, & \text{se o pixel indexado por } (l,c), \text{ na imagem de entrada } \mathcal{I}, \text{ pertence} \\ & \text{ao conjunto } \mathcal{O}^c. \end{cases} \quad (2.5)$$

Contudo, na prática, um método de segmentação é empregado para estimar  $M^*$  e resulta na máscara  $M \in \{0,1\}^{n_l \times n_c}$  em que:

$$M(l,c) = \begin{cases} 1, & \text{se o pixel indexado por } (l,c), \text{ na imagem de entrada } \mathcal{I}, \text{ foi esti-} \\ & \text{mado como "de lesão"}, \\ 0, & \text{se o pixel indexado por } (l,c), \text{ na imagem de entrada } \mathcal{I}, \text{ foi esti-} \\ & \text{mado como "de não lesão"}. \end{cases} \quad (2.6)$$

Os métodos de segmentação, em geral, podem ser orientados à similaridade e/ou descontinuidade. Basicamente, os métodos orientados à similaridade modelam a ROI e o fundo explicitamente como sendo conjuntos de componentes (i.e., pixels, superpixels ou blocos) similares. Os métodos orientados à descontinuidade, por outro lado, modelam tais conjuntos implicitamente por meio da fronteira que os separa (DOUGHERTY, 2009; GONZALEZ; WOODS, 2018).

Os métodos orientados à descontinuidade costumam não lidar bem com imagens macroscópicas de lesões melanocíticas, porque tais imagens, frequentemente, apresentam um baixo contraste entre a ROI e o fundo e/ou contêm artefatos indesejáveis (e.g., pêlos) com bordas salientes (ALCÓN et al., 2009; WONG; SCHARCANSKI; FIEGUTH, 2011). Contudo, os ditos métodos de contornos ativos, os quais produzem contornos fechados e suaves, com frequência são empregados para refinar o resultado obtido com algum método orientado à similaridade (veja, por exemplo, os trabalhos de Tang (2009), Cavalcanti et al. (2011), Cavalcanti e Scharcanski (2013)).

Os métodos orientados à similaridade, sob a perspectiva do aprendizado automático não supervisionado (FACELI et al., 2021), lidam com uma tarefa de agrupamento (rígido) (JAIN; MURTY; FLYNN, 1999) de modo que ajustam uma função  $\mathcal{S} : \mathbb{R}^m \rightarrow \{0, 1\}$  a partir de um conjunto de amostras não rotuladas  $\mathbb{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^n$ , no qual cada amostra  $\mathbf{y}^{(i)} \in \mathbb{R}^m$  é um vetor de características que repre-

sentam um único componente de alguma imagem. Após esse ajuste, a função  $\mathcal{S}$  pode ser empregada para estimar a identidade de cada pixel de uma dada imagem  $\mathcal{I} \in \mathbb{R}^{n_l \times n_c \times n_b}$ , com o intuito de produzir a máscara  $M^*$  definida na Eq. (2.5), o que resulta na máscara da Eq. (2.6).

Os métodos orientados à similaridade podem ser subdivididos em “não contextuais” – os quais realizam a identificação dos componentes considerando somente o seu valor – e “contextuais” – que, além do valor, consideram também relações entre diferentes componentes. No caso dos métodos não contextuais, cada pixel de uma imagem de intensidades  $I \in \mathbb{R}^{n_l \times n_c}$  pode ser representado por um único valor  $I(l,c)$  e a função utilizada para estimar a identidade dos pixels possui a forma  $\mathcal{S}(I(l,c)) = \mathbb{1}_{\{I(l,c) \geq \tau\}}$ , onde  $\tau$  é um parâmetro configurável denominado limiar. Em especial, o clássico método de Otsu (1979), o qual tem sido utilizado como um componente em diversos métodos propostos para segmentar lesões melanocíticas em imagens macroscópicas (OLIVEIRA et al., 2016), permite ajustar o limiar  $\tau$  de maneira adaptativa.

Especificamente, uma abordagem simples e frequentemente empregada para segmentar lesões melanocíticas em imagens macroscópicas consiste em, primeiro, produzir um canal de intensidades que tendem a ser distintas para os pixels de lesão e não lesão e, depois, aplicar o método de Otsu (1979) nesse canal. Por exemplo, Cavalcanti e Scharcanski (2013) utilizaram um canal que visa representar a variabilidade local da luminância, a qual é usualmente maior nos pixels de lesão do que nos pixels de não lesão. Glaister, Wong e Clausi (2014), por sua vez, propuseram um canal cujas intensidades representam a distintividade textural, a qual também tende a ser diferente nos pixels de lesão e não lesão.

Outro tipos de estratégias envolvem a combinação de múltiplos canais e/ou múltiplos limiares. Por exemplo o método de Bernart, Scharcanski e Bampi (2016), primeiramente, aplica o método de Otsu (1979) independentemente nos três canais do espaço de cores RGB. Em seguida, estabelece a identidade de todos os pixels em que houve um consenso nas três limiarizações. Por fim, os demais pixels, ditos “suspeitos”, são identificados com base nos  $K$  vizinhos mais próximos (KNN – *K Nearest Neighbors*) dentre aqueles em que houve um consenso anteriormente.

No nosso trabalho em (ZORTEA; FLORES; SCHARCANSKI, 2017), por outro lado, propusemos ponderar dois limiares obtidos aplicando o método de

Otsu (1979) independentemente em duas diferentes regiões críticas de um mesmo canal cujas intensidades representam a distintividade em relação a cor mediana de pixels estimados como não lesão. O canal proposto em (ZORTEA; FLORES; SCHARCANSKI, 2017) também foi utilizado para compor o método de segmentação proposto nesta tese e, em função disso, será detalhado na Subseção 4.1.2.

Uma maneira tradicional de descrever o problema de agrupamento em questão é por meio do seguinte problema de otimização:

$$(\mathbf{D}^*, \mathbf{X}^*) = \arg \min_{\mathbf{D}, \mathbf{X}} \left\{ \sum_{i=1}^n \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \right\} \text{ sujeito a } \|\mathbf{x}_i\|_0 = 1 \text{ para } i = 1, \dots, n, \quad (2.7)$$

onde  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{K \times n}$  ( $K$  é um hiperparâmetro configurável) é uma matriz em que cada coluna  $\mathbf{x}_i$  é uma nova representação, baseada na matriz  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{m \times K}$ , para o vetor  $\mathbf{y}_i$ .

Note que se a restrição de que  $\|\mathbf{x}_i\|_0$  deve ser exatamente 1 para  $i = 1, \dots, n$  for satisfeita, então o produto  $\mathbf{D}\mathbf{x}_i$  deve ser igual a  $\mathbf{d}_k$  para algum  $k \in \{1, \dots, K\}$  (i.e., uma quantização vetorial). Por exemplo, se  $K = 2$ , a representação de  $\mathbf{y}_i$  por  $\mathbf{d}_1$  pode interpretada como  $\mathcal{S}(\mathbf{y}_i) = 0$  e a representação de  $\mathbf{y}_i$  por  $\mathbf{d}_2$  pode interpretada como  $\mathcal{S}(\mathbf{y}_i) = 1$  (ou vice-versa). Nesse contexto, o conjunto de todos os componentes  $\mathbf{y}_i$  que são representados pelo mesmo vetor  $\mathbf{d}_k$  é dito o  $k$ -ésimo agrupamento. Em todo o caso, tem-se que  $\|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 = \sum_{k=1}^K x_{i,k} \|\mathbf{y}_i - \mathbf{d}_k\|_2^2$ , com  $x_{i,k}$  denotando o  $k$ -ésimo componente de  $\mathbf{x}_i$ , de modo que o problema pode ser alternativamente descrito por (BISHOP, 2006):

$$(\mathbf{D}^*, \mathbf{X}^*) = \arg \min_{\mathbf{D}, \mathbf{X}} \left\{ \sum_{i=1}^n \sum_{k=1}^K x_{i,k} \|\mathbf{y}_i - \mathbf{d}_k\|_2^2 \right\} \text{ sujeito a } \|\mathbf{x}_i\|_0 = 1 \text{ para } i = 1, \dots, n. \quad (2.8)$$

Dentre os métodos que abordam o problema da Eq. (2.8), o mais popular é o método generalizado de Lloyd, referido nesta tese como o método de K-médias (KANUNGO et al., 2002; JAIN, 2010). Tal método atualiza os valores de  $\mathbf{D}$  e  $\mathbf{X}$  iterativamente, com base nas seguintes etapas:

1. Inicializar  $\mathbf{D}$ .
2. (a) Minimizar a função objetivo da Eq. (2.8) em relação a  $\mathbf{X}$  (com a restri-

ção de que a soma de cada coluna de  $\mathbf{X}$  deve resultar em 1), mantendo a matriz  $\mathbf{D}$  fixa.

(b) Minimizar a função objetivo da Eq. (2.8) em relação a  $\mathbf{D}$ , mantendo a matriz  $\mathbf{X}$  fixa.

3. Se pelo menos um componente mudar de agrupamento e o número máximo de iterações não for excedido, voltar para o início da etapa 2. Senão, encerrar.

Observe que, na etapa 2.(a), os valores de  $x_{i,k}$ , para  $i = 1, \dots, n$  e  $k = 1, \dots, K$ , devem ser atualizados simplesmente verificando o átomo  $\mathbf{d}_k$  mais próximo de cada componente  $\mathbf{y}_i$ :

$$x_{i,k} = \begin{cases} 1, & \text{se } k = \arg \min_{k'} \|\mathbf{y}_i - \mathbf{d}_{k'}\|_2^2, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.9)$$

Na etapa 2.(b), por sua vez, derivando a função objetivo em relação a  $\mathbf{d}_k$  e igualando o resultado ao vetor nulo, pode-se obter (BISHOP, 2006):

$$\mathbf{d}_k = \frac{\sum_{i=1}^n x_{i,k} \mathbf{y}_i}{\sum_{i=1}^n x_{i,k}}. \quad (2.10)$$

Em suma, na etapa 2.(a), cada componente  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ , é representado pelo vetor  $\mathbf{d}_k$  que está mais próximo (i.e., o componente é atribuído para o  $k$ -ésimo agrupamento). Na etapa 2.(b), cada átomo  $\mathbf{d}_k$ ,  $k = 1, \dots, K$ , é atualizado para ser a média (i.e., o centróide) dos componentes que foram atribuídas para o  $k$ -ésimo agrupamento na etapa anterior. Em função disso, pode-se convenientemente denotar  $\mathbf{d}_k$  como  $\boldsymbol{\mu}_k$ , para evidenciar o fato de que tal vetor é a média dos componentes do  $k$ -ésimo agrupamento – por isso o método é denominado K-médias.

A convergência desse método é garantida, pois cada iteração reduz (ou mantém) o valor da função objetivo da Eq. (2.8). No entanto, uma vez que tal função não é convexa, o mesmo pode convergir para um mínimo local ao invés de convergir para o mínimo global (BOYD; VANDENBERGHE, 2004). Assim, uma questão importante a ser considerada é a inicialização das médias na etapa 1 (CELEBI; KINGRAVI; VELA, 2013). Um método amplamente utilizado para inicia-

lizar as médias é o K-médias++ (ARTHUR; VASSILVITSKII, 2007), o qual tende a acelerar a convergência do K-médias, tornando-o também mais robusto, via a execução das seguintes etapas:

1. Selecionar aleatoriamente um componente  $\mathbf{y}_i$  para ser a média  $\boldsymbol{\mu}_1$ , considerando a probabilidade de selecionar o  $i$ -ésimo componente,  $\mathbf{y}_i$ , dada por:

$$\mathbb{P}(\mathbf{y}_i) = \frac{1}{n}. \quad (2.11)$$

2. Computar o quadrado da distância entre cada componente  $\mathbf{y}_i$  e  $\boldsymbol{\mu}_1$ :

$$d_i^2 = \|\mathbf{y}_i - \boldsymbol{\mu}_1\|_2^2. \quad (2.12)$$

3. Selecionar aleatoriamente um componente  $\mathbf{y}_i$  para ser a média  $\boldsymbol{\mu}_2$ , considerando a probabilidade de selecionar o  $i$ -ésimo componente,  $\mathbf{y}_i$ , dada por:

$$\mathbb{P}(\mathbf{y}_i) = \frac{d_i^2}{\sum_{i=1}^n d_i^2}. \quad (2.13)$$

4. Se  $K > 2$ , executar para  $t = 2, \dots, K$ :

- (a) Computar o quadrado da distância entre cada componente  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ , e a média do agrupamento mais próximo:

$$d_i^2 = \min_k \{\|\mathbf{y}_i - \boldsymbol{\mu}_k\|_2^2\}. \quad (2.14)$$

- (b) Selecionar aleatoriamente um componente  $\mathbf{y}_i$  para ser a média  $\boldsymbol{\mu}_{t+1}$ , considerando a probabilidade de selecionar o vetor  $\mathbf{y}_i$  dada pela Eq. (2.13).

Cabe mencionar que o método de K-médias, embora seja tradicionalmente empregado para a segmentação de imagens em geral (FORSYTH; PONCE, 2012), é incapaz de particionar adequadamente agrupamentos que não sejam linearmente separáveis no  $\mathbb{R}^m$ . Para contornar essa limitação, pode-se precedê-lo com a fatoração de  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$  em matrizes não negativas (DHILLON; GUAN; KULIS, 2004; DING; HE; SIMON, 2005).<sup>7</sup> Um exemplo representativo de método

<sup>7</sup>O problema da fatoração em matrizes não negativas será tratado em detalhes na Subseção 3.2.5.

que aplica essa ideia, para segmentar imagens macroscópicas de lesões melanocíticas, pode ser encontrado em (CAVALCANTI et al., 2014).

Outra alternativa bem conhecida para lidar com amostras que não são linearmente separáveis no espaço de entrada consiste em usar algum tipo de relaxação espectral em um modelo de agrupamento baseado em grafo (NASCIMENTO, 2010; NASCIMENTO; CARVALHO, 2011). Nesse contexto, o conjunto  $\{\mathbf{y}^{(i)}\}_{i=1}^n$  pode ser representado por um grafo  $G = \{\mathbb{V}, \mathbb{E}\}$ . Nesse grafo  $G$ ,  $\mathbb{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  é o conjunto de vértices, em que  $\mathbf{v}_i$  representa  $\mathbf{y}_i$ , e  $\mathbb{E}$  é o conjunto de arestas, no qual uma aresta  $(i, j) \in \mathbb{E}$  se e somente se  $\mathbf{v}_i \in \mathbb{V}$  é adjacente a  $\mathbf{v}_j \in \mathbb{V}$ .

Estritamente falando, tal grafo é ponderado e não direcionado de modo que pode ser caracterizado por meio de uma matriz simétrica de similaridade  $\mathbf{W} \in \mathbb{R}^{n \times n}$ , a qual atribui um peso para cada aresta e pode ser inicializada conforme mostrado a seguir (LUXBURG, 2007):

$$W_{i,j} = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{v}_i - \mathbf{v}_j\|_2^2\right). \quad (2.15)$$

Após a inicialização, todos os valores de  $\mathbf{W}$  menores que um determinado valor  $\epsilon$  podem ser setados para zero de modo que dois vértices  $\mathbf{v}_i$  e  $\mathbf{v}_j$  são considerados adjacentes se e somente se  $W_{i,j} > 0$ .

O grafo  $G$  pode ser particionado em dois subgrafos, cujos conjuntos de vértices são  $\mathbb{A}$  e  $\mathbb{B} = \mathbb{V} - \mathbb{A}$ , removendo as arestas que conectam  $\mathbb{A}$  e  $\mathbb{B}$ . Dessa maneira, o peso total das arestas removidas (i.e., o corte) é dado por:

$$\text{Corte}(\mathbb{A}, \mathbb{B}) = \sum_{\mathbf{v}_i \in \mathbb{A}, \mathbf{v}_j \in \mathbb{B}} W_{i,j}. \quad (2.16)$$

Infelizmente, como observado por Wu e Leahy (1993), a minimização do corte definido na Eq. (2.16) pode resultar no agrupamento de pequenos conjuntos de vértices isolados, pois não considera a informação intragrupo.

Para lidar com esse problema, Shi e Malik (1997, 2000) propuseram a minimização do corte normalizado, o qual é definido por:

$$\text{CorteN}(\mathbb{A}, \mathbb{B}) = \frac{\text{Corte}(\mathbb{A}, \mathbb{B})}{\text{Corte}(\mathbb{A}, \mathbb{V})} + \frac{\text{Corte}(\mathbb{A}, \mathbb{B})}{\text{Corte}(\mathbb{B}, \mathbb{V})}. \quad (2.17)$$

Além disso, Shi e Malik (1997, 2000) mostraram que:

$$\arg \min_{\mathbf{A}, \mathbf{B}} \{\text{CorteN}(\mathbf{A}, \mathbf{B})\} = \arg \min_{\mathbf{Q}} \left\{ \frac{\mathbf{Q}^T (\mathbf{D} - \mathbf{W}) \mathbf{Q}}{\mathbf{Q}^T \mathbf{D} \mathbf{Q}} \right\} \text{ sujeito a } \mathbf{Q}^T \mathbf{D} \mathbf{1}_n = 0, \quad (2.18)$$

onde  $\mathbf{D} \in \mathbb{R}^{n \times n}$  é a matriz diagonal na qual:

$$\mathbf{D}_{i,i} = \sum_{j=1}^n W_{i,j} \quad (2.19)$$

e  $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_n]^T \in \{0,1\}^{n \times 2}$  é uma matriz em que:

$$\mathbf{q}_i = \begin{cases} (1,0), & \text{se } \mathbf{v}_i \in \mathbf{A}, \\ (0,1), & \text{se } \mathbf{v}_i \in \mathbf{B}. \end{cases} \quad (2.20)$$

De fato, uma versão relaxada do problema de otimização na segunda linha da Eq. (2.18) pode ser resolvida com base em autovalores e autovetores relacionados com a matriz Laplaciana  $\mathbf{D} - \mathbf{W}$ , o que é usualmente referido na literatura como corte normalizado de grafo ou NGC (do inglês, *Normalized Graph Cut*) (SHI; MALIK, 1997; SHI; MALIK, 2000). Contudo, o elevado custo computacional do NGC, especialmente no que diz respeito ao consumo de memória, é uma possível justificativa para o mesmo ter sido tão pouco explorado para segmentar lesões melanocíticas em imagens macroscópicas (OLIVEIRA et al., 2016). Para mitigar esse problema, nesta tese, aplica-se o NGC sobre representações de blocos em um dicionário construído combinando NMF com uma adaptação do método ITDL (QIU; PATEL; CHELLAPPA, 2014), conforme será detalhado na Seção 4.1.

### 2.2.3 Métodos de Classificação

Conforme pormenorizado anteriormente, sistemas de processamento e análise de imagens macroscópicas de lesões melanocíticas, em última instância, tradicionalmente ajustam uma função  $\mathcal{C} : \mathbb{R}^{nd} \rightarrow \{1, \dots, n_r\}$  para, dado um vetor com as características extraídas a partir de uma lesão de interesse, predizer um rótulo que codifique a classe da lesão em questão.

Note que o rótulo predito  $r$  é, efetivamente, uma estimativa do rótulo correto  $r^*$ , de modo que o erro quadrático médio do método de aprendizado (i.e.,

do estimador), ao realizar uma predição, pode ser decomposto como segue (GE-MAN; BIENENSTOCK; DOURSAT, 1992):

$$\begin{aligned}
\mathbb{E}((r^* - \mathbf{r})^2) &= \mathbb{E}((r^* - \mathbb{E}(\mathbf{r}) + \mathbb{E}(\mathbf{r}) - \mathbf{r})^2) \\
&= \mathbb{E}((r^* - \mathbb{E}(\mathbf{r}))^2 + (\mathbb{E}(\mathbf{r}) - \mathbf{r})^2 + 2(r^* - \mathbb{E}(\mathbf{r}))(\mathbb{E}(\mathbf{r}) - \mathbf{r})) \\
&= \mathbb{E}((r^* - \mathbb{E}(\mathbf{r}))^2) + \mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})^2) + \mathbb{E}(2(r^* - \mathbb{E}(\mathbf{r}))(\mathbb{E}(\mathbf{r}) - \mathbf{r})) \\
&= (r^* - \mathbb{E}(\mathbf{r}))^2 + \mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})^2) + 2(r^* - \mathbb{E}(\mathbf{r}))\mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})) \\
&= (r^* - \mathbb{E}(\mathbf{r}))^2 + \mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})^2) + 2(r^* - \mathbb{E}(\mathbf{r}))(\mathbb{E}(\mathbb{E}(\mathbf{r})) - \mathbb{E}(\mathbf{r})) \\
&= (r^* - \mathbb{E}(\mathbf{r}))^2 + \mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})^2) + 2(r^* - \mathbb{E}(\mathbf{r}))\overset{0}{\cancel{\mathbb{E}(\mathbf{r}) - \mathbb{E}(\mathbf{r})}} \\
&= \text{viés}(\mathbf{r})^2 + \text{variância}(\mathbf{r}), \tag{2.21}
\end{aligned}$$

com

$$\mathbb{E}(\mathbf{r}) = \int_{-\infty}^{\infty} \mathbf{r} \cdot \mathbb{P}(\mathbf{r}) \cdot d\mathbf{r} \tag{2.22}$$

denotando o valor esperado da variável aleatória  $\mathbf{r} \sim \mathbb{P}(\mathbf{r})$  da qual  $r$  é um dos possíveis valores.

Se fosse factível fixar o termo  $\text{viés}(\mathbf{r})^2 = (r^* - \mathbb{E}(\mathbf{r}))^2$ , a minimização do erro na Eq. (2.21) poderia ser efetuada simplesmente diminuindo ao máximo o valor do termo não negativo  $\text{variância}(\mathbf{r}) = \mathbb{E}((\mathbb{E}(\mathbf{r}) - \mathbf{r})^2)$ .<sup>8</sup> Infelizmente, reduzir a variância (e.g., forçando a obtenção de uma função  $\mathcal{C}$  menos flexível) tende a elevar o viés, ao passo que reduzir o viés tende a elevar a variância. Esse efeito indesejável é referido na literatura de aprendizado de máquina como “o dilema viés-variância” (do inglês, *the bias-variance tradeoff*) (MITCHELL, 1997; FACELI et al., 2021). De uma forma geral, funções mais flexíveis apresentam um baixo viés e uma alta variância (i.e., sobreajustam-se aos exemplos de treinamento) em comparação com funções menos flexíveis, as quais, comparativamente, possuem um alto viés e uma baixa variância (i.e., subajustam-se aos exemplos de treinamento) (BURKOV, 2019; IZBICKI; SANTOS, 2020).

Na prática, selecionar uma função que exiba um bom compromisso entre o viés e a variância costuma ser um dos maiores desafios ao lidar com um pro-

<sup>8</sup>De maneira análoga, poderia-se manter fixa a variância e diminuir, tanto quanto possível, o quadrado do viés.

blema de classificação tal como este que está sendo abordado nesta tese (WOLPERT, 1996; DOMINGOS, 2012). Por isso, vários métodos de aprendizado automático supervisionado têm sido empregados para discriminar lesões melanocíticas em imagens macroscópicas, incluindo métodos baseados em distâncias, métodos baseados em árvores de decisão, métodos baseados em aprendizado Bayesiano, métodos baseados em maximização de margens e métodos baseados em modelos múltiplos preditivos (OLIVEIRA et al., 2018; PATHAN; PRABHU; SIDDALINGASWAMY, 2018; OKUBOYEJO; OLUGBARA, 2018).

Todavia, sistemas que utilizam métodos de aprendizado mais tradicionais (e.g., KNN), normalmente, requerem projetos mais elaborados de redução de dimensionalidade a fim de descrever as lesões de uma maneira discriminativa. Nesse sentido, características globais da lesão inspiradas na regra ABCD, ou em algum outro protocolo dermatológico, estão entre as abordagens mais convencionais, assim como o uso de características locais baseadas nas cores e/ou nas texturas dos componentes (e.g., superpixels) da imagem. Exemplos representativos desse tipo de sistema incluem aqueles propostos por Alcón et al. (2009), Cavalcanti e Scharcanski (2011), Cavalcanti, Scharcanski e Baranoski (2013), Amelard et al. (2014) e Bernart, Scharcanski e Bampi (2016).

Nos últimos anos, depois da publicação do influente trabalho de Krizhevsky, Sutskever e Hinton (2012), sistemas baseados em redes neurais convolucionais profundas (do inglês, *Deep Convolutional Neural Networks* – DCNNs) e aprendizado profundo<sup>9</sup> passaram a ser considerados “o padrão-ouro” para tarefas de processamento e análise de exames médicos que recaem em classificação de imagens (SHIN et al., 2016; LITJENS et al., 2017; MENEGOLA et al., 2017).

Uma DCNN, em termos gerais, é um modelo na forma  $(f \circ \mathcal{F})(\mathcal{I}; \Theta)$ , composto por um classificador  $\mathcal{C} = \mathcal{F}_l$  e por uma função de redução de dimensionalidade  $\mathcal{F} = \mathcal{F}_{l-1} \circ \mathcal{F}_{l-2} \circ \dots \circ \mathcal{F}_1$ , o qual é convencionalmente treinado por um método de aprendizado profundo (regularmente, alguma variante do método retropropagação de erro), a fim de minimizar uma determinada função de perda  $\mathcal{P}(\mathbb{I}, \Theta)$ , de um modo que todos os seus parâmetros  $\Theta = \{\theta_i\}_{i=1}^{n_p}$  são ajustados diretamente a partir dos dados brutos contidos em  $\mathbb{I} = \{(\mathbf{I}^{(i)}, r^{(i)})\}_{i=1}^n$  (e.g., um conjunto de imagens de lesões melanocíticas rotuladas) (GOODFELLOW; BEN-

---

<sup>9</sup>Uma forma de aprendizado automático com algumas adaptações necessárias para treinar apropriadamente algum modelo profundo (e.g., uma DCNN) (LECUN; BENGIO; HINTON, 2015).

GIO; COURVILLE, 2016).

Os componentes primordiais das DCNNs são as ditas camadas convolucionais ou convolutivas. Para  $k \in \{1, \dots, \ell - 2\}$ , uma camada convolucional é uma função  $\mathcal{F}_k : \mathbb{R}^{n_i^{[k-1]} \times n_c^{[k-1]} \times n_b^{[k-1]}} \rightarrow \mathbb{R}^{n_i^{[k]} \times n_c^{[k]} \times n_b^{[k]}}$  que produz um arranjo tridimensional  $\mathcal{Z}^{[k]} \in \mathbb{R}^{n_i^{[k]} \times n_c^{[k]} \times n_b^{[k]}}$ , denominado o volume de saída da  $k$ -ésima (i.e., o volume de entrada da  $(k + 1)$ -ésima) camada, conforme apresentado a seguir:

$$\mathcal{Z}_{l,c,h}^{[k]} = g^{[k]} \left( \left( \mathcal{W}_h^{[k]} * \mathcal{Z}^{[k-1]} \right)_{l,c} + \beta_h^{[k]} \right), \quad (2.23)$$

onde  $g^{[k]} : \mathbb{R} \rightarrow \mathbb{R}$  e  $*$  denotam, respectivamente, a função de transferência (ou ativação) da  $k$ -ésima camada e a operação de convolução com múltiplos canais;  $\Theta^{[k]} = \{(\mathcal{W}_h^{[k]}, \beta_h^{[k]}) : \mathcal{W}_h^{[k]} \in \mathbb{R}^{w^{[k]} \times w^{[k]} \times n_b^{[k-1]}}$  e  $\beta_h^{[k]} \in \mathbb{R}\}_{h=1}^{n_i^{[k]}} \subset \Theta$  é denominado o banco de filtros da  $k$ -ésima camada e contém os pesos (ou seja, os parâmetros) da camada em questão.

Normalmente, DCNNs empregam também camadas de agregação (do inglês, *pooling*). Para  $k \in \{1, \dots, \ell - 2\}$ , uma camada de agregação é uma função  $\mathcal{F}_k : \mathbb{R}^{n_i^{[k-1]} \times n_c^{[k-1]} \times n_b^{[k-1]}} \rightarrow \mathbb{R}^{n_i^{[k-1]}/w^{[k]} \times n_c^{[k-1]}/w^{[k]} \times n_b^{[k-1]}}$ , na qual  $w^{[k]} \in \mathbb{N} - \{0,1\}$ , que reduz as quantidades de linhas e colunas de  $\mathcal{Z}^{[k-1]} \in \mathbb{R}^{n_i^{[k-1]} \times n_c^{[k-1]} \times n_b^{[k-1]}}$  (i.e., reduz o volume de saída da  $(k - 1)$ -ésima camada ou, em outras palavras, o volume de entrada da  $k$ -ésima camada), de acordo com a seguinte equação:

$$\mathcal{Z}_{l',c',h}^{[k]} = g^{[k]} \left( \left\{ \mathcal{Z}_{l',c',h}^{[k-1]} : s(l-1) + 1 \leq l' \leq s.l, s(c-1) + 1 \leq c' \leq s.c \right\} \right), \quad (2.24)$$

em que  $g^{[k]} : \mathbb{S} \rightarrow \mathbb{R}$  é a função de agregação da  $k$ -ésima camada.

As últimas camadas de uma DCNN, usualmente, são as intituladas densas ou totalmente conectadas. Para  $k \in \{2, \dots, \ell - 1\}$ , uma camada totalmente conectada é uma função  $\mathcal{F}_k : \mathbb{R}^{n^{[k-1]}} \rightarrow \mathbb{R}^{n^{[k]}}$  que gera um vetor  $\mathbf{z} \in \mathbb{R}^{n^{[k]}}$  no qual os componentes são dados por:

$$z_j = g^{[k]} \left( \left( \mathcal{W}^{[k]} \mathbf{z}^{[k-1]} \right)_j + \beta_j^{[k]} \right), \quad (2.25)$$

onde  $g^{[k]} : \mathbb{R} \rightarrow \mathbb{R}$  é uma função de ativação;  $\Theta^{[k]} = \{\mathcal{W}^{[k]} \in \mathbb{R}^{n^{[k]} \times n^{[k-1]}}\}$ ,  $\{\beta_j^{[k]} \in \mathbb{R}\}_{j=1}^{n^{[k]}} \subset \Theta$  são os pesos associados à  $k$ -ésima camada.

Basicamente, uma DCNN efetua uma redução gradativa na dimensionalidade de uma dada imagem de entrada por meio de uma série de camadas con-

volucionais e de agregação até a obtenção de um determinado volume  $\mathcal{Z}^{[k]} \in \mathbb{R}^{n_i^{[k]} \times n_c^{[k]} \times n_b^{[k]}}$ . Em seguida, esse volume é rearranjado na forma de um vetor unidimensional  $\mathbf{z}^{[k]} \in \mathbb{R}^{n^{[k]}}$  ( $n^{[k]} = n_i^{[k]} \cdot n_c^{[k]} \cdot n_b^{[k]}$ ) que, subsequentemente, é fornecido para a primeira dentre uma sequência de camadas totalmente conectadas, as quais resultam noutro vetor  $\mathbf{z}^{[k]} \in \mathbb{R}^{n_r}$  cujos componentes podem ser interpretados como os escores de predição da DCNN para as possíveis classes da imagem avaliada.

Um exemplo canônico de DCNN é a VGGNet-16, proposta por [Simonyan e Zisserman \(2015\)](#), a qual contém 16 camadas parametrizáveis (13 camadas convolutivas e três camadas totalmente conectadas), além de cinco camadas não parametrizáveis (camadas de agregação). A Tabela 2.1, onde

$$\text{ReLU}(z) = \max\{0, z\} \quad (2.26)$$

e

$$\sigma(z) = (1 + \exp(-z))^{-1}, \quad (2.27)$$

fornece um detalhamento sobre cada uma dessas camadas. A Figura 2.3, por sua vez, ilustra os volumes que representam a imagem de entrada ao longo de todo o seu processamento pela VGGNet-16.

Para treinar a VGGNET-16, [Simonyan e Zisserman \(2015\)](#) empregaram mais de um milhão e duzentas mil imagens rotuladas, as quais foram disponibilizadas para a competição do ImageNet (i.e., ILSVRC – *ImageNet Large Scale Visual Recognition Challenge*) ([RUSSAKOVSKY et al., 2015](#)). Mesmo assim, técnicas de regularização precisaram ser utilizadas para de induzir uma flexibilidade reduzida no modelo. Por exemplo, a função de perda considerada,

$$\mathcal{P}(\mathbb{I}, \Theta) = \left( - \sum_{i=1}^n \sum_{j=1}^{n_r} \mathbb{1}_{\{r^{(i)}=j\}} \cdot \log \left( [(f \circ \mathcal{F})(\mathcal{I}^{(i)}; \Theta)]_j \right) \right) + \lambda \sum_{i=1}^{n_p} \theta_i^2, \quad (2.28)$$

é uma versão regularizada da clássica perda em ([NELDER; WEDDERBURN, 1972](#)). Além disso, a técnica de regularização por *dropout* ([SRIVASTAVA et al., 2014](#)) também foi usada no treinamento, o qual apesar de tudo levou algo entre duas e três semanas em um computador equipado com quatro GPUs NVIDIA Titan Black.



classificação (KHAN et al., 2020). A saída desses modelos é um vetor no  $\mathbb{R}^{n_r}$  em que cada componente indicará o escore de predição da imagem de entrada para algum dos  $n_r$  rótulos possíveis. Visto que segmentar uma imagem de entrada equivale a classificar cada um dos seus pixels, adaptações arquiteturais cuja saída está em  $\mathbb{R}^{n_l \times n_c \times n_r}$ , e não em  $\mathbb{R}^{n_r}$ , podem ser realizadas para tarefas de segmentação (SHELHAMER; LONG; DARRELL, 2017). Exemplos representativos de arquiteturas baseadas nessa ideia incluem a FCNet (*Fully Convolutional Network*) (LONG; SHELHAMER; DARRELL, 2015), a U-Net (RONNEBERGER; FISCHER; BROX, 2015), a PSPNet (*Pyramid Scene Parsing Network*) (ZHAO et al., 2017) e a SegNet (BADRINARAYANAN; KENDALL; CIPOLLA, 2017).

Um resultado clássico das teorias do aprendizado estatístico e computacional estabelece que quanto mais flexível o modelo, maior é a quantidade necessária de exemplos (e.g., imagens macroscópicas de lesões melanocíticas rotuladas como “maligna” ou “nevo atípico”) para garantir que o mesmo será ajustado de um modo “provavelmente aproximadamente correto” (VALIANT, 1984; EHRENFUCHT et al., 1989). Em particular, a flexibilidade de uma DCNN, medida conforme proposto por Vapnik e Chervonenkis (1971), depende do seu número de camadas e, principalmente, do total de parâmetros contidos nas mesmas (BARTLETT et al., 2019).

Na prática, é conhecido o fato de que treinar DCNNs demanda uma quantidade consideravelmente maior de exemplos do que para treinar modelos mais tradicionais de aprendizado automático (POUYANFAR et al., 2018). Em parte por isso que, somente na última década, tornou-se mais viável treinar DCNNs, as quais já vinham sendo investigadas na literatura há bastante tempo (veja, por exemplo, Fukushima (1988) e/ou LeCun e Bengio (1995)).<sup>11</sup>

De acordo com Menegola et al. (2017), são necessárias pelo menos algumas dezenas de milhares de exemplos para treinar adequadamente uma DCNN da maneira mais tradicional (ou seja, inicializando os pesos pseudoaleatoriamente – i.e., “do zero” – e atualizando-os iterativamente com uma variante do *backpropagation*, do modo que foi efetuado por Simonyan e Zisserman (2015)). Porém, para desenvolver aplicações na área médica, normalmente, trabalha-se em um cenário no qual a quantidade de exemplos disponibilizados é escassa (e.g., menos de mil

<sup>11</sup> Além do fenômeno do *big data* (CHEN; ZHANG, 2014), outro fator determinante para a popularização das DCNNs foi o advento das GPGPUs (*General Purpose Graphic Processing Units* – unidades de processamento gráfico de propósito geral) (TAN; LIM, 2018).

exemplos) (SHEN; WU; SUK, 2017). Uma estratégia convencional para contornar essa limitação é partir de uma DCNN pré-treinada (e.g., com as imagens do ILSVRC), o que pode ou não envolver um ajuste fino nos pesos pré-ajustados e é conhecido na literatura como aprendizado por transferência.<sup>12</sup>

### 2.3 Estado-da-Arte

Em seu trabalho seminal, Esteva et al. (2017) propuseram empregar mais de 120 mil imagens, macroscópicas e dermatoscópicas, para realizar um ajuste fino nos pesos da DCNN Inception-v3 (SZEGEDY et al., 2016), previamente ajustados, com a base do ILSVRC, conforme detalhado no artigo de Szegedy et al. (2016). Tal DCNN foi testada com 130 imagens macroscópicas para distinguir melanomas de nevos benignos (não necessariamente atípicos). A curva ROC (*Receiver Operating Characteristic*) obtida a partir dos resultados desses testes ficou acima de 12 dos 22 pontos ROC correspondentes aos desempenhos preditivos dos dermatologistas comparados. Embora os resultados em (ESTEVA et al., 2017) sejam promissores, os dados utilizados (e.g., as imagens clínicas e as categorias taxonômicas fornecidas aos autores pelo centro médico da universidade de Stanford) não foram disponibilizados, inviabilizando assim a reprodutibilidade desse trabalho (PACHECO; KROHLING, 2019; FORNACIALI, 2019).

Mais recentemente, diversos outros sistemas de classificação de lesões melanocíticas baseados em aprendizado profundo têm sido propostos na literatura, geralmente, combinando aprendizado por transferência com outras abordagens tais como o aumento de dados (do inglês, *data augmentation*), a incorporação de metadados, o uso de mecanismos de atenção e a composição de modelos múltiplos preditivos com DCNNs. Exemplos representativos de tais sistemas incluem aqueles em (WOZA, 2018; ZHANG et al., 2019; HA; LIU; LIU, 2020; PACHECO; TRAPPENBERG; KROHLING, 2020). Contudo, a maioria desses sistemas foram projetados para operar com base em imagens dermatoscópicas, principalmente, pelos motivos a seguir.

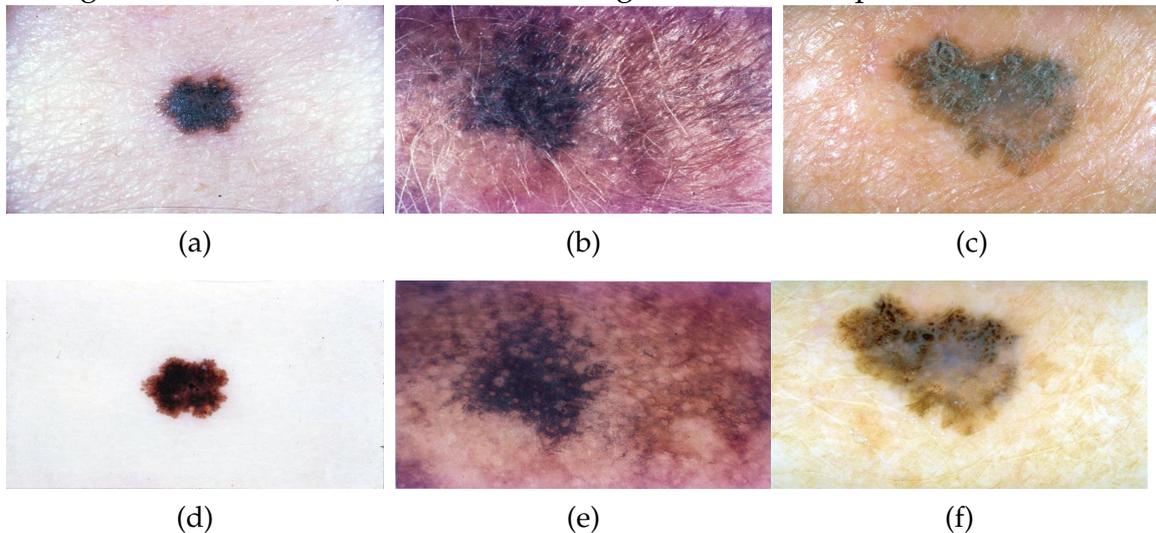
1. A tarefa de processar e analisar uma imagem macroscópica de lesão melano-

---

<sup>12</sup>A premissa básica desta estratégia é que, após o pré-treinamento, as primeiras camadas da DCNN detectem características genéricas em imagens (e.g., bordas), as quais são úteis para discriminar imagens em diferentes domínios de aplicação.

nocítica, em vez de uma imagem dermatoscópica, é mais desafiadora não só porque a classificação da lesão deve ser realizada com base em uma imagem que contém menos informações sugestivas, mas também porque é necessário elaborar etapas de processamento de imagens mais robustas para lidar com complicações adicionais (e.g., iluminação irregular e maior quantidade de artefatos indesejáveis), conforme exemplificado na Figura 2.4, onde cada linha apresenta uma determinada lesão melanocítica capturada pelas duas modalidades de imageamento. Concretamente, na primeira linha (i.e., Subfiguras 2.4a, 2.4c e 2.4e), essas lesões são apresentadas em imagens macroscópicas. Na segunda linha (i.e., nas Subfiguras 2.4b, 2.4d e 2.4f), por sua vez, essas mesmas lesões são apresentadas em imagens dermatoscópicas.

Figura 2.4: Exemplos de imagens dermatoscópicas de lesões melanocíticas em comparação com as imagens macroscópicas correspondentes. A primeira linha (i.e., as Subfiguras 2.4a até 2.4c) apresenta imagens macroscópicas de lesões melanocíticas, as mesmas que também são apresentadas na segunda linha (i.e., nas Subfiguras 2.4d até 2.4f), desta vez, em imagens dermatoscópicas.



Fonte: Kawahara et al. (2018).

2. A existência, desde 2016, de uma competição anual, organizada pelo ISIC (*International Skin Imaging Collaboration*) (CODELLA et al., 2018), a qual atualmente disponibiliza uma base pública com mais de 70 mil imagens dermatoscópicas, ou seja, uma ordem de magnitude maior que a base de imagens macroscópicas comercializada em (UNIVERSITY OF EDINBURGH, 2022) e duas ordens de magnitude maior que as bases de imagens macroscópicas públicas propostas por Amelard et al. (2014), Giotis et al. (2015), Brinker et

al. (2019).

Cabe ainda mencionar que a maior parte dos sistemas baseados em aprendizado profundo propostos para classificar imagens de lesões melanocíticas não incorpora uma etapa própria de segmentação das lesões, embora métodos baseados em aprendizado profundo especificamente projetados para a segmentação possam ser encontrados na literatura, por exemplo, em (JAFARI et al., 2016; ALMASNI et al., 2018; HASAN et al., 2020).

## 2.4 Considerações Finais

Neste capítulo, foi apresentada uma síntese da literatura relacionada aos sistemas de processamento e análise de imagens de lesões melanocíticas. Conforme apresentado, a maioria dos sistemas propostos recentemente foram desenvolvidos para operar com base em imagens dermatoscópicas e, consequentemente, podem não desempenhar a sua função da maneira esperada quando a entrada é uma imagem macroscópica. Além disso, tais sistemas foram treinados com uma quantidade de imagens maior do que as quantidades de imagens disponíveis nas bases públicas de imagens macroscópicas.

Evidencia-se também o fato que os sistemas existentes, normalmente, não incluem uma etapa específica de segmentação, o que eleva os riscos de problemas como aqueles reportados em (BISSOTO et al., 2019). Ademais, a avaliação desses sistemas costuma basear-se em validações cruzadas sobre uma mesma base de imagens (ao invés de basear-se em testes com bases cruzadas) e, por conseguinte, pode resultar em previsões de desempenho pouco realistas.

O sistema proposto nesta tese, basicamente, foi projetado para lidar bem com todos os desafios delineados nos dois últimos parágrafos. Antes da apresentação detalhada desse sistema no Capítulo 4, no próximo capítulo (i.e., no Capítulo 3) serão tratados os métodos para obter representações esparsas e dicionários que fundamentaram a elaboração do sistema proposto.

### 3 REPRESENTAÇÕES ESPARSAS ATRAVÉS DE DICIONÁRIOS: CONCEITOS E MÉTODOS

Neste capítulo, na Seção 3.2, são tratados em detalhes métodos para a obtenção de representações esparsas e dicionários, os quais viabilizaram o desenvolvimento do sistema proposto nesta tese. Anteriormente, na Seção 3.1, são apresentadas definições e conceitos que servem de embasamento teórico para os métodos da Seção 3.2. Além disso, na Seção 3.3, encontram-se as considerações finais deste capítulo.

#### 3.1 Conceitos Fundamentais

Nesta seção, são introduzidas algumas definições e conceitos fundamentais. O intuito principal desta seção é esclarecer detalhes referentes à terminologia empregada não só na próxima seção (i.e., na Seção 3.2), mas também no restante desta tese e, em especial, no capítulo subsequente (i.e., no Capítulo 4).

##### 3.1.1 Representação de Imagens em Bases Ortonormais

Sabe-se que, qualquer arranjo unidimensional  $\mathbf{y} \in \mathbb{R}^m$  (e.g., uma imagem rearranjada como um vetor (PRATT, 1975)) pode ser decomposto da seguinte forma:

$$\mathbf{y} = \mathbf{I}_m \mathbf{y} = \sum_{k=1}^m y_k \delta_k, \quad (3.1)$$

onde  $\delta_k \in \mathbb{R}^m$  denota a  $k$ -ésima coluna da matriz identidade  $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ .<sup>1</sup>

O conjunto  $\mathbb{D} = \{\delta_k\}_{k=1}^m$ , dito “a base canônica”, é somente um caso particular de base ortonormal para o  $\mathbb{R}^m$  e os coeficientes da combinação linear da Eq. (3.1),  $\{y_k\}_{k=1}^m$ , podem ser referidos como “a representação de  $\mathbf{y}$  em  $\mathbb{D}$ ”. Todavia, é importante mencionar que a representação de um sinal na base canônica raramente é uma opção adequada para tarefas de análise (RUBINSTEIN; BRUCKSTEIN; ELAD, 2010).

<sup>1</sup>O vetor  $\delta_k$ , na área de processamento digital de sinais, é dito um “impulso unitário deslocado” (DINIZ; DA SILVA; NETTO, 2014).

De uma forma mais geral, se  $\mathbb{D} = \{\mathbf{d}_k\}_{k=1}^m$  é uma base ortonormal para o  $\mathbb{R}^m$ , então  $\mathbf{d}_k \cdot \mathbf{d}_j = \delta_{k,j}$ , para quaisquer  $j, k \in \{1, \dots, m\}$ , e qualquer vetor  $\mathbf{y} \in \mathbb{R}^m$  pode ser escrito de acordo com (BURRUS; GOPINATH; GUO, 1998):

$$\begin{aligned}\mathbf{y} &= \sum_{k=1}^m x_k \mathbf{d}_k \\ &= \mathbf{D}\mathbf{x},\end{aligned}\tag{3.2}$$

em que  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_m] \in \mathbb{R}^{m \times m}$  e  $\mathbf{x} = (x_1, \dots, x_m) = [x_1, \dots, x_m]^T \in \mathbb{R}^m$ .

Fazendo o produto escalar dos dois lados da primeira linha da Eq. (3.2) com algum  $\mathbf{d}_j$ ,  $j \in \{1, \dots, m\}$ , tem-se que:

$$\begin{aligned}\mathbf{y} \cdot \mathbf{d}_j &= \left( \sum_{k=1}^m x_k \mathbf{d}_k \right) \cdot \mathbf{d}_j \\ &= x_j \mathbf{d}_j \cdot \mathbf{d}_j + \sum_{k \neq j} x_k \mathbf{d}_k \cdot \mathbf{d}_j \\ &= x_j \overset{1}{\mathbf{d}_j \cdot \mathbf{d}_j} + \sum_{k \neq j} x_k \overset{0}{\mathbf{d}_k \cdot \mathbf{d}_j} \\ &= x_j.\end{aligned}\tag{3.3}$$

Assim, pode-se concluir que se  $\mathbb{D} = \{\mathbf{d}_k\}_{k=1}^m$  é uma base ortonormal para o  $\mathbb{R}^m$ , então qualquer vetor  $\mathbf{y} \in \mathbb{R}^m$  pode ser representado em  $\mathbb{D}$  através dos coeficientes  $\{x_k\}_{k=1}^m$ , os quais podem ser determinados de maneira única pela Eq. (3.3) (com  $j = k$ ). A mesma conclusão pode ser obtida observando que se as colunas de  $\mathbf{D}$  são ortonormais, então  $\mathbf{D}\mathbf{D}^T = \mathbf{D}^T\mathbf{D} = \mathbf{I}_m$ , ou seja, a inversa de  $\mathbf{D}$  existe e é igual a sua transposta, de modo que

$$\mathbf{x} = \mathbf{D}^{-1}\mathbf{y} = \mathbf{D}^T\mathbf{y}\tag{3.4}$$

é a única solução para o sistema de equações lineares da Eq. (3.2).

Essa ideia pode ser estendida para representar  $\mathbf{y} \in \mathbb{R}^m$  em uma base unitária  $\mathbf{D} \in \mathbb{C}^{m \times m}$  (i.e., a inversa de  $\mathbf{D}$  é igual a sua Hermitiana), de maneira que a representação  $\mathbf{x} \in \mathbb{C}^m$  é um vetor com  $m$  componentes complexas. Um exemplo clássico desse tipo de situação ocorre quando o valor do elemento localizado na

$l$ -ésima linha e na  $c$ -ésima coluna de  $\mathbf{D}$  é dado por:

$$D_{l,c} = \frac{1}{\sqrt{m}} \exp\left(-\frac{2\pi\sqrt{-1}(l-1)(c-1)}{m}\right). \quad (3.5)$$

Nesse exemplo, as Eqs. (3.2) e (3.4) correspondem, respectivamente, às equações de síntese e análise da DFT (*Discrete Fourier Transform* – Transformada Discreta de Fourier) (RAO; YIP, 2001).

### 3.1.2 Representações Esparsas de Imagens via Dicionários

Uma base ortonormal é, de fato, um caso especial de dicionário que, por sua vez, é um conjunto de vetores  $\mathbb{D} = \{\mathbf{d}_k\}_{k=1}^K$ , tipicamente referidos como átomos ou itens do dicionário.<sup>2</sup> Note que o problema mais geral, no qual é dado um dicionário  $\mathbb{D}$  qualquer e um vetor  $\mathbf{y} \in \mathbb{R}^m$  que deseja-se escrever como uma combinação linear dos átomos de  $\mathbb{D}$ , equivale algebricamente a determinar um vetor  $\mathbf{x} = [x_1, \dots, x_K]^T \in \mathbb{R}^K$  tal que:

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^K x_k \mathbf{d}_k \\ &= \mathbf{D}\mathbf{x}, \end{aligned} \quad (3.6)$$

onde  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{m \times K}$  é o dicionário  $\mathbb{D}$  disposto em um formato matricial no qual as colunas de  $\mathbf{D}$  correspondem aos átomos do dicionário  $\mathbb{D}$ .

Uma vez que  $\mathbb{D}$  não é necessariamente uma base, o sistema de equações lineares da Eq. (3.6) pode não ter solução ou possuir infinitas soluções (i.e., a matriz  $\mathbf{D}$  pode ser singular) (HEATH, 2018). Efetivamente, um dicionário pode ser categorizado como (LOVISOLO, 2006; LOVISOLO; DA SILVA, 2014; MALLAT, 2009):

- completo e não redundante (i.e., uma base para o  $\mathbb{R}^m$ ), se o dicionário  $\mathbb{D}$  é composto por  $K = m$  átomos linearmente independentes e  $\text{col}(\mathbf{D}) = \mathbb{R}^m$ ;
- completo e redundante (i.e., um dicionário sobrecompleto), se o dicionário  $\mathbb{D}$  é composto por  $K > m$  átomos linearmente dependentes e  $\text{col}(\mathbf{D}) = \mathbb{R}^m$ ;

<sup>2</sup>Ao longo desta tese, assume-se que os átomos do dicionário são  $\ell_2$ -normalizados, de forma que  $\|\mathbf{d}_k\|_2 = 1$  para  $k = 1, \dots, K$ .

- subcompleto, se  $\text{col}(\mathbf{D}) \subset \mathbb{R}^m$  (i.e., se  $\text{col}(\mathbf{D}) \subseteq \mathbb{R}^m$  e  $\text{col}(\mathbf{D}) \neq \mathbb{R}^m$ ).

A forma mais tradicional de contornar as eventuais singularidades consiste em relaxar o problema de maneira que o objetivo passe a ser determinar:

$$\begin{aligned} \mathbf{x}_2 &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \right\} \\ &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{D}\mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{D}^T \mathbf{D}\mathbf{x} \right\}. \end{aligned} \quad (3.7)$$

Assim, derivando a função objetivo – convexa – da Eq. (3.7) em relação ao vetor  $\mathbf{x}$ , e igualando o resultado ao vetor nulo, segue que (BOYD; VANDENBERGHE, 2004):

$$\mathbf{x}_2 = \mathbf{D}^\dagger \mathbf{y}, \quad (3.8)$$

onde

$$\mathbf{D}^\dagger = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \quad (3.9)$$

denota a matriz pseudo-inversa de  $\mathbf{D}$ .

Contudo, é mais conveniente empregar uma relaxação alternativa que tolere um erro de reconstrução menor ou igual ao hiperparâmetro<sup>3</sup>  $\epsilon$  e, em contrapartida, busque pela minimização de alguma função de regularização  $\mathcal{R} : \mathbb{R}^m \rightarrow \mathbb{R}$  que, de certa forma, meça o quão “indesejável” é uma dada representação  $\mathbf{x}$ , conforme mostrado a seguir (ELAD, 2010):

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \{ \mathcal{R}(\mathbf{x}) \} \text{ sujeito a } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon. \quad (3.10)$$

Note que se o hiperparâmetro que quantifica o nível de tolerância ao erro de reconstrução,  $\epsilon$ , é menor que  $\|\mathbf{y} - \mathbf{D}\mathbf{x}_2\|_2$ , nenhum vetor  $\mathbf{x}$  satisfaz a restrição imposta na Eq. (3.10), o que inviabiliza a obtenção de  $\mathbf{x}^*$ .

Visto que a maioria dos sinais de interesse (e.g., imagens macroscópicas de lesões melanocíticas) possuem redundâncias consideráveis (DA SILVA; MENDONÇA, 2005), funções de regularização que promovem representações espar-

<sup>3</sup>Ao longo desta tese, o termo hiperparâmetro é usado para denominar uma variável de configuração cujo valor deve ser especificado antes da execução do método em questão. O termo parâmetro, por outro lado, indica uma variável cujo valor é ajustado durante a execução do método (GOODFELLOW; BENGIO; COURVILLE, 2016).

sas, as quais concentram a energia em alguns poucos coeficientes, podem ser empregadas para evidenciar a “essência subjacente” do sinal. Cabe mencionar que tais representações são adequadas não só para o reconhecimento de padrões (WRIGHT et al., 2009), mas também para diversas tarefas de processamento e análise de imagens (veja, por exemplo, (RABBANI; JOSHI, 2002), (ELAD; AHARON, 2006) e (SCHULZ; SILVA; VELHO, 2009)).

Concretamente, a esparsidade de um vetor  $\mathbf{x}$  pode ser mensurada pela sua pseudo-norma  $\ell_0$ , a qual é denotada  $\|\mathbf{x}\|_0$  e definida por (PATEL; CHELLAPA, 2013):

$$\|\mathbf{x}\|_0 = \text{card}(\text{sup}(\mathbf{x})) = \lim_{p \rightarrow 0} \|\mathbf{x}\|_p^p, \quad (3.11)$$

onde

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \quad (3.12)$$

é a norma  $\ell_p$  de  $\mathbf{x}$ ,

$$\text{sup}(\mathbf{x}) = \{i : x_i \neq 0\} \quad (3.13)$$

é o suporte de  $\mathbf{x}$ , ou seja, o conjunto composto pelos índices dos componentes não-nulos de  $\mathbf{x}$ , e  $\text{card}(S)$  denota a cardinalidade do conjunto  $S$ , isso é, a quantidade de elementos do conjunto em questão. Em outras palavras, a “norma”  $\ell_0$  resulta na quantidade de componentes não-nulos do vetor.

Fazendo  $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_0$  na Eq. (3.10), pode-se escrever:

$$\begin{aligned} \hat{\mathbf{x}}_0 &= \arg \min_{\mathbf{x}} \{\|\mathbf{x}\|_0\} \text{ sujeito a } \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \leq \epsilon \\ &= \arg \min_{\mathbf{x}} \{\|\mathbf{x}\|_0\} \text{ sujeito a } \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \leq \frac{1}{2} \epsilon^2. \end{aligned} \quad (3.14)$$

Em seguida, reescrevendo na forma Lagrangiana generalizada, segue que:

$$\begin{aligned}\hat{\boldsymbol{x}}_0 &= \arg \min_{\boldsymbol{x}} \left\{ \|\boldsymbol{x}\|_0 + \frac{1}{\lambda} \left[ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 - \frac{1}{2} \epsilon^2 \right] \right\} \\ &= \arg \min_{\boldsymbol{x}} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_0 \right\}.\end{aligned}\quad (3.15)$$

Na prática, embora o valor de  $\lambda$  seja unicamente definido a partir do valor de  $\epsilon$ , o mesmo costuma ser tratado diretamente como um hiperparâmetro, pois a sua relação com  $\epsilon$  depende também dos valores de  $\boldsymbol{D}$  e  $\boldsymbol{y}$  (RISH; GRABARNIK, 2015).

### 3.2 Métodos para Obter Representações Esparsas e Dicionários

Nesta seção, são descritos métodos para obter uma representação esparsa através de um dado dicionário e também métodos para construir dicionários. Os métodos para obter uma representação esparsa em um dado dicionário, dependendo do contexto no qual são aplicados, são denominados métodos de aproximação esparsa (TROP; WRIGHT, 2010), métodos de codificação esparsa (BAO et al., 2015), métodos de recuperação esparsa (MARQUES et al., 2018), métodos de decomposição atômica (CHEN; DONOHO; SAUNDERS, 2001) ou métodos de perseguição (MALLAT; ZHANG, 1993) (em inglês, *sparse approximation*, *sparse coding*, *sparse recovery*, *atomic decomposition* e *pursuit methods*, respectivamente).

Idealmente, o propósito dos métodos de codificação esparsa é determinar a solução  $\hat{\boldsymbol{x}}_0 \in \mathbb{R}^K$  das Eqs. (3.14) e (3.15). Efetivamente, é obtida uma estimativa de  $\hat{\boldsymbol{x}}_0$ , a qual deve ser produzida a partir da amostra  $\boldsymbol{y} \in \mathbb{R}^m$ , do dicionário  $\boldsymbol{D} \in \mathbb{R}^{m \times K}$  e do hiperparâmetro (i.e., do escalar  $\epsilon$  ou do escalar  $\lambda$ ). Note que se o valor de  $\|\hat{\boldsymbol{x}}_0\|_0$  fosse conhecido *a priori*, poderia-se reescrever o problema da Eq. (3.15) na forma:

$$\hat{\boldsymbol{x}}_0 = \arg \min_{\boldsymbol{x}_S} \left\{ \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{D}_S \boldsymbol{x}_S\|_2^2 \right\}, \quad (3.16)$$

em que  $S$  denota algum suporte de cardinalidade  $\|\hat{\boldsymbol{x}}_0\|_0$ ,  $\boldsymbol{D}_S$  é a matriz composta pelas colunas de  $\boldsymbol{D}$  cujos índices estão em  $S$  e  $\boldsymbol{x}_S$  é o vetor formado pelos componentes de  $\boldsymbol{x}$  cujos índices estão em  $S$ .

Assim, poderia-se, em tese, calcular o valor de  $\boldsymbol{x}_S$  para cada suporte de cardinalidade  $\|\hat{\boldsymbol{x}}_0\|_0$ , computar a função objetivo para cada um desses valores e,

em seguida, selecionar aquele que produziu o menor resultado. No entanto, deve ser evidenciado que a quantidade total de suportes de cardinalidade  $\|\hat{\mathbf{x}}_0\|_0$  é:

$$\binom{K}{\|\hat{\mathbf{x}}_0\|_0} = \frac{K!}{\|\hat{\mathbf{x}}_0\|_0!(K - \|\hat{\mathbf{x}}_0\|_0)!}. \quad (3.17)$$

Além disso, o valor de  $\|\hat{\mathbf{x}}_0\|_0$  não é conhecido *a priori*, o que requer a avaliação, primeiramente, de todos os suportes de cardinalidade 1, depois de todos os suportes de cardinalidade 2 e assim por diante, até que uma aproximação suficientemente robusta de  $\hat{\mathbf{x}}_0$  seja obtida.

O procedimento de busca exaustiva (i.e., força bruta) mencionado anteriormente pode ser sintetizado da seguinte forma:

1. Fazer  $s \leftarrow 1$ .
2. Para cada suporte  $\mathcal{S}$  de cardinalidade  $s$ :
  - 2.1. Obter  $\mathbf{x}_{\mathcal{S}}$  através da Eq. (3.8).
  - 2.2. Se  $\|\mathbf{y} - \mathbf{D}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|_2 \leq \epsilon$ , encerrar. Senão, fazer  $s \leftarrow s + 1$  e retornar ao início do passo 2.

Repare que a quantidade total de suportes avaliados, para os quais os passos 2.1 e 2.2 devem ser executados, é:

$$T(K, \|\hat{\mathbf{x}}_0\|_0) = 1 + \sum_{s=1}^{\|\hat{\mathbf{x}}_0\|_0-1} \binom{K}{s} \quad (3.18)$$

no melhor caso e

$$T(K, \|\hat{\mathbf{x}}_0\|_0) = \sum_{s=1}^{\|\hat{\mathbf{x}}_0\|_0} \binom{K}{s} \quad (3.19)$$

no pior caso.

Para exemplificar a ineficiência desse método, suponha que o hiperparâmetro  $\epsilon$  tenha sido escolhido permitindo obter a solução  $\hat{\mathbf{x}}_0$  com  $K = 1000$  componentes, dentre os quais somente  $\|\hat{\mathbf{x}}_0\|_0 = 10$  são não-nulos. Assumindo que cada avaliação demore 1 nanosegundo, a busca exaustiva levaria mais de 85 mil anos, no melhor caso, e aproximadamente 8,5 milhões de anos, no pior caso, para encontrar a solução desse problema. De fato, o problema das Eqs. (3.14) e (3.15)

é, em geral, NP-difícil (NATARAJAN, 1995; DAVIS; MALLAT; AVELLANEDA, 1997).

Apesar disso, diversos métodos numéricos vêm sendo propostos na literatura a fim de propiciar formas práticas de aproximar a solução desse problema. Muitos desses baseiam-se no fato de que a solução  $\hat{x}_0$  é composta de duas partes: 1) o suporte da solução, isso é, as posições onde estão localizados os componentes não-nulos de  $\hat{x}_0$  e 2) os valores desses componentes nas posições indicadas pelo suporte. Deve-se evidenciar que, se o suporte da solução for conhecido, os valores dos componentes não-nulos podem ser determinados através da Eq. (3.8). Baseado nisso, os métodos em questão abordam o problema estimando iterativamente, de modo heurístico, as posições dos componentes não-nulos, o que em essência nada mais é do que a concepção de um método guloso. O primeiro método a empregar esse tipo de abordagem nesse contexto foi denominado *Matching Pursuit* (MP) e introduzido por Mallat e Zhang (1993). Uma versão aperfeiçoada desse método, o *Orthogonal Matching Pursuit* (OMP) (DAVIS; MALLAT; ZHANG, 1994; PATI; REZAIIFAR; KRISHNAPRASAD, 1993), é apresentado em detalhes na Subseção 3.2.1.

Outra estratégia amplamente utilizada consiste em relaxar o problema original da Eq. (3.15), substituindo a norma  $\ell_0$  pela norma  $\ell_1$ , o que resulta no seguinte problema de otimização convexa:

$$\hat{x}_1 = \arg \min_x \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}, \quad (3.20)$$

o qual é frequentemente referido, na literatura de processamento de sinais, como BPDN (*Basis Pursuit DeNoising*) (CHEN; DONOHO; SAUNDERS, 1998) e é essencialmente equivalente ao problema de regressão linear via LASSO (*Least Absolute Shrinkage and Selection Operator*) (TIBSHIRANI, 1996). Surpreendentemente, pode-se mostrar que, na maioria dos casos, a solução  $\hat{x}_0$  da Eq. (3.15) é satisfatoriamente aproximada pela solução  $\hat{x}_1$  da Eq. (3.20) (DONOHO, 2006; CANDES; ROMBERG; TAO, 2006).

O método do gradiente<sup>4</sup> descendente pode parecer, à primeira vista, uma abordagem natural para o problema, pois é globalmente convergente para problemas de otimização irrestrita nos quais a função objetivo é convexa. Especifica-

<sup>4</sup>Formalmente, a função objetivo da Eq. (3.20) não possui gradiente em  $\mathbf{x} = \mathbf{0}_K$ . Porém, como essa função é convexa, um subgradiente pode ser usado para desempenhar a função do gradiente (SHOR, 1985).

mente, seja  $\mathbf{x}^{(t)}$  a  $t$ -ésima estimativa produzida pelo método em questão, tem-se que a diferença entre o valor da função objetivo em  $\mathbf{x}^{(t)}$  e o valor da função objetivo no ponto de mínimo global é, no pior caso, proporcional a  $\ln(t)/\sqrt{t}$ , ou seja, a taxa de convergência do método é  $\mathcal{O}(\ln(t)/\sqrt{t})$  (NESTEROV, 2018). Por outro lado, métodos baseados em operadores proximais costumam ser mais adequados para esse tipo de situação, pois levam em consideração a estrutura do problema (PARIKH; BOYD et al., 2014).

O método ISTA (*Iterative STA*), baseado no operador proximal *shrink* (conhecido como STA (*Shrinkage–Thresholding Algorithm*)), é bem difundido na literatura de otimização e possui uma taxa de convergência  $\mathcal{O}(1/t)$ . O método FISTA (*Fast ISTA*), por sua vez, é uma versão mais eficiente do método ISTA, com taxa de convergência  $\mathcal{O}(1/t^2)$ , e vem sendo considerado por diversos autores da área como “o método estado-da-arte” para esse tipo de problema (BECK; TEBOULLE, 2009; BACH et al., 2012; FLOREA; VOROBYOV, 2017). Na Subseção 3.2.2, o método FISTA é explicado detalhadamente.

Além de métodos para obter uma representação esparsa em um dicionário dado, nesta subseção, considera-se opções de métodos para gerar um dicionário. Para um dado conjunto de amostras  $\{\mathbf{y}_i\}_{i=1}^n$ , esses métodos têm por objetivo produzir um dicionário  $\mathbb{D}$  tal que cada amostra  $\mathbf{y}_i \in \mathbb{R}^m$ ,  $i \in \{1, \dots, n\}$ , possa ser aproximada como uma combinação linear de alguns poucos átomos de  $\mathbb{D}$ . Esse dicionário, em geral, pode ser pré-definido ou aprendido a partir do conjunto de amostras (RUBINSTEIN; BRUCKSTEIN; ELAD, 2010).

Um dicionário pré-definido é obtido através de uma fórmula fechada, a qual não depende diretamente das amostras, mas sim de um modelo matemático pré-estabelecido para descrevê-las. A base ortonormal da Eq. (3.5) (i.e., a matriz da DFT) é um exemplo de dicionário pré-definido, bem como as matrizes das DWTs (*Discrete Wavelet Transforms*) e das DCTs (*Discrete Cosine Transforms*), as quais são amplamente difundidas na literatura de processamento e análise de sinais (WANG, 2012).

Um dicionário aprendido, por outro lado, é treinado/adaptado a partir do conjunto de amostras (TOSIC; FROSSARD, 2011). Embora o esforço computacional para obter um dicionário pré-definido seja consideravelmente menor do que para aprender um dicionário, ao longo da última década, vem sendo evidenciado o fato que dicionários aprendidos fornecem representações mais adequadas não

só para as aplicações tratadas nesta proposta de tese, mas também para diversas outras tarefas práticas tais como processamento de áudio, compressão de dados, codificação de vídeo, restauração de imagens e reconhecimento facial (YANG et al., 2009; PATEL; CHELLAPA, 2013; GANGEH et al., 2015).

### 3.2.1 Orthogonal Matching Pursuit (OMP)

Basicamente, o objetivo do OMP é estimar a solução  $\hat{x}_0$  da Eq. (3.14) iterativamente. Para isso, a cada iteração, um único átomo é ativado, ou seja, o coeficiente da estimativa que corresponde ao átomo em questão deixa de valer zero. Especificamente, partindo de  $\mathbf{x}^{(0)} = \mathbf{0}_K$ , na  $t$ -ésima iteração, uma estimativa  $\mathbf{x}^{(t)}$  é obtida ativando o átomo que permite causar uma maior diminuição na energia do resíduo, de acordo com os passos enumerados a seguir (ELAD, 2010; TCHEOU, 2011):

1. Inicializar  $t = 0$ ,  $\mathbf{x}^{(t)} = \mathbf{0}_K$ ,  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{D}\mathbf{x}^{(t)} = \mathbf{y}$  e  $S^{(t)} = \emptyset$ ;  $\mathbf{r}^{(t)}$  denota o resíduo na  $t$ -ésima iteração;  $S^{(t)}$  denota o conjunto com os índices dos átomos ativos na  $t$ -ésima iteração.
2. Fazer  $t = t + 1$  e computar a diminuição máxima na energia do resíduo que pode ser provocada pela ativação do  $k$ -ésimo átomo,

$$\mathcal{E}(k) = \min_a \|a \mathbf{d}_k - \mathbf{r}^{(t-1)}\|_2^2, \quad (3.21)$$

para  $k = 1, \dots, K$ .

3. Ativar o átomo que permite causar uma maior diminuição na energia do resíduo, isso é, seja  $k^* \in \{1, \dots, K\}$  um índice tal que  $\mathcal{E}(k^*) \leq \mathcal{E}(k)$  para todo  $k \in \{1, \dots, K\}$ , fazer  $S^{(t)} = S^{(t-1)} \cup \{k^*\}$ .
4. Computar

$$\mathbf{x}^{(t)} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \text{ sujeito a } \text{sup}(\mathbf{x}) = S^{(t)}. \quad (3.22)$$

5. Computar  $\mathbf{r}^{(t)} = \mathbf{y} - \mathbf{D}\mathbf{x}^{(t)}$ . Caso o critério de convergência tenha sido satisfeito (i.e.,  $\|\mathbf{r}^{(t)}\|_2^2 \leq \epsilon$  ou  $t > t_{\max}$ , com  $\epsilon$  e  $t_{\max}$  sendo hiperparâmetros do método), encerrar. Caso contrário, retornar ao passo 2.

Note que o valor ótimo de  $a$  no passo 2 pode ser determinado derivando a função objetivo da Eq. (3.21), igualando o resultado a zero e manipulando para obter:

$$a^* = \frac{\mathbf{d}_k^T \mathbf{r}^{(t-1)}}{\mathbf{d}_k^T \mathbf{d}_k}. \quad (3.23)$$

Além disso, como os átomos do dicionário são  $\ell_2$ -normalizados, tem-se:

$$\bar{a}^* = \mathbf{d}_k^T \mathbf{r}^{(t-1)}. \quad (3.24)$$

Substituindo o resultado da Eq. (3.24) na Eq. (3.21), segue que:

$$\begin{aligned} \mathcal{E}(k) &= \|\bar{a}^* \mathbf{d}_k - \mathbf{r}^{(t-1)}\|_2^2 = \|\mathbf{d}_k^T \mathbf{r}^{(t-1)} \mathbf{d}_k - \mathbf{r}^{(t-1)}\|_2^2 \\ &= [\mathbf{d}_k^T \mathbf{r}^{(t-1)} \mathbf{d}_k - \mathbf{r}^{(t-1)}]^T [\mathbf{d}_k^T \mathbf{r}^{(t-1)} \mathbf{d}_k - \mathbf{r}^{(t-1)}] \\ &= [\mathbf{d}_k^T \mathbf{r}^{(t-1)} \mathbf{d}_k^T - (\mathbf{r}^{(t-1)})^T] [\mathbf{d}_k^T \mathbf{r}^{(t-1)} \mathbf{d}_k - \mathbf{r}^{(t-1)}] \\ &= (\mathbf{d}_k^T \mathbf{r}^{(t-1)})^2 \underbrace{\mathbf{d}_k^T \mathbf{d}_k}_1 - (\mathbf{d}_k^T \mathbf{r}^{(t-1)})^2 - (\mathbf{d}_k^T \mathbf{r}^{(t-1)})^2 + (\mathbf{r}^{(t-1)})^T \mathbf{r}^{(t-1)} \\ &= \|\mathbf{r}^{(t-1)}\|_2^2 - (\mathbf{d}_k^T \mathbf{r}^{(t-1)})^2. \end{aligned} \quad (3.25)$$

Assim, uma vez que o termo  $\|\mathbf{r}^{(t-1)}\|_2^2$  independe de  $k$ , conclui-se que o problema de encontrar um índice  $k^* \in \{1, \dots, K\}$  tal que  $\mathcal{E}(k^*) \leq \mathcal{E}(k)$  para todo  $k \in \{1, \dots, K\}$  (passo 3) pode ser resolvido de maneira equivalente e mais prática selecionando-se um índice  $k^* \in \{1, \dots, K\}$  tal que  $|\mathbf{d}_{k^*}^T \mathbf{r}^{(t-1)}| \geq |\mathbf{d}_k^T \mathbf{r}^{(t-1)}|$  para todo  $k \in \{1, \dots, K\}$ .

### 3.2.2 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

Primeiramente, considere o caso particular em que o dicionário  $\mathbf{D}$  é ortogonal<sup>5</sup>, no qual pode-se reescrever a Eq. (3.15) na forma:

<sup>5</sup>Um dicionário  $\mathbf{D} \in \mathbb{R}^{m \times K}$  é dito ortogonal se e somente se é quadrado (i.e.,  $m = K$ ) e  $\mathbf{D}^T \mathbf{D} = \mathbf{D} \mathbf{D}^T = \mathbf{I}_m = \mathbf{I}_K$ .

$$\begin{aligned}
\hat{\mathbf{x}}_1 &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{D}\mathbf{D}^T \mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{D}(\mathbf{z} - \mathbf{x})\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} (\mathbf{z} - \mathbf{x})^T \mathbf{D}^T \mathbf{D} (\mathbf{z} - \mathbf{x}) + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} (\mathbf{z} - \mathbf{x})^T (\mathbf{z} - \mathbf{x}) + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\
&= \arg \min_{\mathbf{x}} \left\{ \sum_{k=1}^K \left[ \frac{1}{2} (x_k - z_k)^2 + \lambda |x_k| \right] \right\}, \tag{3.26}
\end{aligned}$$

em que  $\mathbf{z} = \mathbf{D}^T \mathbf{y}$ .

Derivando a função objetivo da Eq. (3.26), a qual é convexa, em relação a  $x_k$  (assumindo  $x_k \neq 0$ ) e igualando essa derivada a zero, tem-se que:

$$x_k - z_k + \lambda \frac{\partial |x_k|}{\partial x_k} = 0, \tag{3.27}$$

ou, equivalentemente,

$$x_k = z_k - \lambda \frac{\partial |x_k|}{\partial x_k}, \tag{3.28}$$

com:

$$\frac{\partial |x_k|}{\partial x_k} = \frac{\partial \sqrt{x_k^2}}{\partial x_k} = \frac{1}{2\sqrt{x_k^2}} 2x_k = \frac{x_k}{|x_k|} = \begin{cases} 1, & \text{se } x_k > 0, \\ -1, & \text{se } x_k < 0. \end{cases} \tag{3.29}$$

Se  $x_k > 0$ , então substituir o resultado da Eq. (3.29) na Eq. (3.28) resulta em  $x_k = z_k - \lambda$  e, conseqüentemente,  $z_k > \lambda$ , pois  $x_k = z_k - \lambda > 0$ . Por outro lado, se  $x_k < 0$ , então substituir o resultado da Eq. (3.29) na Eq. (3.28) resulta em  $x_k = z_k + \lambda$  e, conseqüentemente,  $z_k < -\lambda$ , pois  $x_k = z_k + \lambda < 0$ . Logo, se  $-\lambda \leq z_k \leq \lambda$ , então  $x_k = 0$ . Assim, os componentes de  $\hat{\mathbf{x}}_1$  podem ser obtidos pelo

seguinte operador (DONOHO; JOHNSTONE, 1994):

$$\text{shrink}(z_k, \lambda) = \begin{cases} z_k - \lambda, & \text{se } z_k > \lambda, \\ z_k + \lambda, & \text{se } z_k < -\lambda, \\ 0, & \text{se } |z_k| \leq \lambda. \end{cases} \quad (3.30)$$

Para o caso geral (i.e., para um dicionário  $\mathbf{D}$  não necessariamente ortogonal), o problema da Eq. (3.20) pode ser reescrito como:

$$\begin{aligned} \hat{\mathbf{x}}_1 &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\ &= \arg \min_{\mathbf{x}} \{f(\mathbf{x}) + \lambda \|\mathbf{x}\|_1\}, \end{aligned} \quad (3.31)$$

onde  $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{D}\mathbf{x} - \mathbf{y}\|_2^2$ .

Fazendo a expansão em série de Taylor de segunda ordem de  $f$ , em torno de um ponto  $\mathbf{x}^{(t)} \in \mathbb{R}^K$ , segue que:

$$\begin{aligned} f(\mathbf{x}) &\simeq f(\mathbf{x}^{(t)}) + (\mathbf{x} - \mathbf{x}^{(t)})^T \nabla f(\mathbf{x}^{(t)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(t)})^T \nabla^2 f(\mathbf{x}^{(t)}) (\mathbf{x} - \mathbf{x}^{(t)}) \\ &= \frac{1}{2} \|\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}\|_2^2 + (\mathbf{x} - \mathbf{x}^{(t)})^T \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(t)})^T \mathbf{D}^T \mathbf{D} (\mathbf{x} - \mathbf{x}^{(t)}). \end{aligned} \quad (3.32)$$

Em seguida, aproximando  $\mathbf{D}^T \mathbf{D}$  por  $L\mathbf{I}_K$ , com  $L$  denotando o dobro do maior autovalor de  $\mathbf{D}^T \mathbf{D}$ , tem-se que (ZHANG et al., 2015)<sup>6</sup>:

$$\begin{aligned} f(\mathbf{x}) &\simeq \frac{1}{2} \|\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}\|_2^2 + (\mathbf{x} - \mathbf{x}^{(t)})^T \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) + \frac{L}{2} (\mathbf{x} - \mathbf{x}^{(t)})^T (\mathbf{x} - \mathbf{x}^{(t)}) \\ &= \frac{1}{2} \|\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}\|_2^2 + \frac{L}{2} \left[ 2(\mathbf{x} - \mathbf{x}^{(t)})^T \frac{1}{L} \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) + (\mathbf{x} - \mathbf{x}^{(t)})^T (\mathbf{x} - \mathbf{x}^{(t)}) \right] \\ &= \frac{1}{2} \|\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}\|_2^2 + \frac{L}{2} \left[ \left\| (\mathbf{x} - \mathbf{x}^{(t)}) + \frac{1}{L} \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) \right\|_2^2 - \frac{1}{L^2} \left\| \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) \right\|_2^2 \right] \\ &= Q(\mathbf{x}, \mathbf{x}^{(t)}). \end{aligned} \quad (3.33)$$

<sup>6</sup>Pode-se mostrar que o gradiente  $\nabla f(\mathbf{x}^{(t)}) = \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y})$  é uma função Lipschitz contínua para a qual a menor constante de Lipschitz é  $L$ . Assim, pode-se aproximar a Hessiana  $\nabla^2 f(\mathbf{x}^{(t)}) = \mathbf{D}^T \mathbf{D}$  conforme indicado no texto.

A Eq. (3.33) viabiliza abordar o problema da Eq. (3.31) iterativamente, de acordo com a seguinte equação:

$$\mathbf{x}^{(t+1)} = \arg \min_{\mathbf{x}} \{ Q(\mathbf{x}, \mathbf{x}^{(t)}) + \lambda \|\mathbf{x}\|_1 \}. \quad (3.34)$$

No entanto, observe que os termos  $\frac{1}{2} \|\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}\|_2^2$  e  $-\frac{1}{L^2} \|\mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y})\|_2^2$ , os quais compõem  $Q(\mathbf{x}, \mathbf{x}^{(t)})$ , independem de  $\mathbf{x}$  e permitem reescrever a Eq. (3.34) na forma (FIGUEIREDO; NOWAK, 2003; DAUBECHIES; DEFRISE; MOL, 2004):

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \arg \min_{\mathbf{x}} \left\{ \frac{L}{2} \left\| \left( \mathbf{x} - \mathbf{x}^{(t)} \right) + \frac{1}{L} \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y}) \right\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\ &= \arg \min_{\mathbf{x}} \left\{ \frac{L}{2} \|\mathbf{x} - \theta(\mathbf{x}^{(t)})\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\} \\ &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \theta(\mathbf{x}^{(t)})\|_2^2 + \frac{\lambda}{L} \|\mathbf{x}\|_1 \right\}, \end{aligned} \quad (3.35)$$

em que  $\theta(\mathbf{x}^{(t)}) = \mathbf{x}^{(t)} - \frac{1}{L} \mathbf{D}^T (\mathbf{D}\mathbf{x}^{(t)} - \mathbf{y})$ .

O problema da Eq. (3.35) pode ser convertido no problema da Eq. (3.26) (penúltima linha) de modo que os componentes de  $\mathbf{x}^{(t+1)}$ ,  $x_k^{(t+1)}$ ,  $k = 1, \dots, K$ , podem ser obtidos por:

$$x_k^{(t+1)} = \text{shrink}(z_k^{(t)}, \lambda/L) = \begin{cases} z_k^{(t)} - \frac{\lambda}{L}, & \text{se } z_k^{(t)} > \frac{\lambda}{L}, \\ z_k^{(t)} + \frac{\lambda}{L}, & \text{se } z_k^{(t)} < -\frac{\lambda}{L}, \\ 0, & \text{se } |z_k^{(t)}| \leq \frac{\lambda}{L}, \end{cases} \quad (3.36)$$

onde  $z_k^{(t)}$  denota o  $k$ -ésimo componente de  $\mathbf{z}^{(t)} = \theta(\mathbf{x}^{(t)})$  e  $\mathbf{z}^{(t)}$  pode ser visto como o resultado de uma iteração do método do gradiente descendente (CHENG, 2015).

Dados um número máximo de iterações,  $t_{\max}$ , e um chute inicial (e.g.,  $\mathbf{x}^{(0)} = \mathbf{0}_K$ , com  $\mathbf{0}_K$  denotando o vetor do  $\mathbb{R}^K$  cujos componentes são todos iguais a zero), o procedimento iterativo descrito pela Eq. (3.35) permite gerar uma sequência de pontos  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(t_{\max})}$ . Para um valor de  $t_{\max}$  suficientemente grande, tem-se que  $\mathbf{x}^{(t_{\max})} = \hat{\mathbf{x}}_1$ . Beck e Teboulle (2009) propuseram utilizar o esquema do gradiente descendente acelerado (NESTEROV, 1983) para produzir uma sequência de pontos  $\tilde{\mathbf{x}}^{(1)}, \tilde{\mathbf{x}}^{(2)}, \dots, \tilde{\mathbf{x}}^{(t_{\max})}$  que requer uma quantidade menor de iterações para convergir para  $\hat{\mathbf{x}}_1$  (e.g.,  $t_{\max} = 100$ ). O método proposto em (BECK; TE-

**BOULLE, 2009**) foi intitulado FISTA (*Fast Iterative Thresholding Algorithm*) e pode ser sumarizado pelos seguintes passos:

1.  $L \leftarrow 2\Lambda_{\max}(\mathbf{D}^T \mathbf{D})$ , com  $\Lambda_{\max}(\mathbf{D}^T \mathbf{D})$  denotando o maior autovalor de  $\mathbf{D}^T \mathbf{D}$ .
2.  $\tilde{\mathbf{x}}^{(1)} \leftarrow \mathbf{0}_K$ .
3.  $\mu^{(1)} \leftarrow 1$ .
4.  $t \leftarrow 1$ .
5.  $\mathbf{z}^{(t)} \leftarrow \theta(\mathbf{x}^{(t)})$
6. Obter  $\mathbf{x}^{(t)}$  computando cada um dos seus componentes,  $x_k^{(t)}$ ,  $k = 1, \dots, K$ , através da Eq. (3.36).
7.  $\mu^{(t+1)} \leftarrow \frac{1}{2} \left( 1 + \sqrt{1 + 4[\mu^{(t)}]^2} \right)$ .
8.  $\tilde{\mathbf{x}}^{(t+1)} \leftarrow \mathbf{x}^{(t)} + \left( \frac{\mu^{(t)} - 1}{\mu^{(t+1)}} \right) (\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})$ .
9. Se  $t > t_{\max}$ , encerrar. Senão, fazer  $t \leftarrow t + 1$  e retornar ao passo 5.

### 3.2.3 Expectation-Maximization e Gaussian Mixture Models (GMMs)

De fato, o K-médias é um caso particular de uma ideia mais geral que consiste em construir o dicionário com base em algum método paramétrico de estimativa de densidade, o qual é empregado para determinar os parâmetros de uma distribuição de probabilidades (i.e., de um modelo estatístico) a partir do conjunto de amostras (**LEWICKI; SEJNOWSKI, 2000; KREUTZ-DELGADO et al., 2003**). Concretamente, o K-médias emprega o clássico EM (*Expectation Maximization*) (**DEMPSTER; LAIRD; RUBIN, 1977**) para ajustar as  $K$  médias, as quais podem ser vistas como uma versão simplificada do modelo de mistura de Gaussianas (GMM – *Gaussian Mixture Model*), de modo que cada amostra é representada através de uma única dentre as  $K$  médias.

O GMM, o qual pode ser visto como um método de agrupamento suave (do inglês, *soft clustering*), efetivamente, serve como um aproximador univer-

sal para distribuições de probabilidades e pode ser definido por (CARREIRA-PERPINAN, 2000):

$$\mathbb{P}(\mathbf{y}_i; \Theta) = \sum_{k=1}^K \alpha_k \mathbb{G}(\mathbf{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (3.37)$$

onde  $\Theta = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\Sigma}_K, \alpha_1, \alpha_2, \dots, \alpha_K\}$  é o conjunto de parâmetros do modelo e  $\mathbb{G}(\mathbf{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  denota a distribuição de probabilidades Gaussiana multivariada com média  $\boldsymbol{\mu}_k$  e matriz de covariância  $\boldsymbol{\Sigma}_k$ :

$$\mathbb{G}(\mathbf{y}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi)^{m/2} \det^{1/2}(\boldsymbol{\Sigma}_k)} \exp\left(-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k)\right). \quad (3.38)$$

Nesse contexto, usualmente, supõe-se que as amostras  $\mathbf{y}_i$ ,  $i = 1, \dots, n$ , são independentes e identicamente distribuídas. Logo, o problema de estimar os parâmetros a partir das amostras pode ser formulado como (i.e., MLE – *Maximum Likelihood Estimation*) (DUDA; HART; STORCK, 2001):

$$\begin{aligned} \Theta^* &= \arg \max_{\Theta} \{\mathbb{P}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n; \Theta)\} \\ &= \arg \max_{\Theta} \left\{ \prod_{i=1}^n \mathbb{P}(\mathbf{y}_i; \Theta) \right\} \\ &= \arg \max_{\Theta} \left\{ \ln \left[ \prod_{i=1}^n \mathbb{P}(\mathbf{y}_i; \Theta) \right] \right\} \\ &= \arg \max_{\Theta} \left\{ \sum_{i=1}^n \ln [\mathbb{P}(\mathbf{y}_i; \Theta)] \right\}. \end{aligned} \quad (3.39)$$

Derivando a função objetivo da Eq. (3.39), a qual é conhecida como *log-likelihood*, em relação a  $\boldsymbol{\mu}_k$ ,  $\boldsymbol{\Sigma}_k$  e  $\alpha_k$ , igualando as derivadas parciais aos respectivos vetores nulos e manipulando as equações resultantes, pode-se obter (BILMES, 1998):

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^n x_{i,k} \mathbf{y}_i}{\sum_{i=1}^n x_{i,k}}, \quad (3.40)$$

$$\Sigma_k = \frac{\sum_{i=1}^n x_{i,k} (\mathbf{y}_i - \boldsymbol{\mu}_k)(\mathbf{y}_i - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^n x_{i,k}} \quad (3.41)$$

e

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n x_{i,k} \quad (3.42)$$

de modo que:

$$x_{i,k} = \frac{\alpha_k \mathbb{G}(\mathbf{y}_i; \boldsymbol{\mu}_k, \Sigma_k)}{\sum_{k'=1}^K \alpha_{k'} \mathbb{G}(\mathbf{y}_i; \boldsymbol{\mu}_{k'}, \Sigma_{k'})}. \quad (3.43)$$

Conforme pode ser observado nas Eqs. (3.40) até (3.43), os parâmetros do GMM não são dados por fórmulas fechadas. Tipicamente, esses parâmetros são obtidos pelo EM, um método numérico e iterativo, o qual efetua as seguintes etapas (BISHOP, 2006):

1. Inicializar os parâmetros  $\{\boldsymbol{\mu}_k\}_{k=1}^K$ ,  $\{\Sigma_k\}_{k=1}^K$  e  $\{\alpha_k\}_{k=1}^K$ .
2. Calcular/recalcular  $x_{i,k}$ , para  $i = 1, 2, \dots, n$  e  $k = 1, 2, \dots, K$ , utilizando os parâmetros correntes, de acordo com a Eq. (3.43). Essa etapa é denominada *Expectation-step* ou, de forma abreviada, *E-step*.
3. Re-estimar os parâmetros  $\{\boldsymbol{\mu}_k\}_{k=1}^K$ ,  $\{\Sigma_k\}_{k=1}^K$  e  $\{\alpha_k\}_{k=1}^K$ , utilizando os valores correntes de  $x_{i,k}$ . Essa etapa é denominada *Maximization-step* ou, de forma abreviada, *M-step*.
4. Avaliar a função objetivo e checar a convergência. Caso o critério de convergência tenha sido satisfeito, encerrar. Caso contrário, retornar para *E-step*.

### 3.2.4 Method of Optimal Directions (MOD) e K-Singular Value Decomposition (K-SVD)

Diversos métodos generalizam o K-médias (a rigor, generalizam o problema de otimização da Eq. (2.7)) no sentido de permitirem uma quantidade

maior de componentes não-nulos nas representações esparsas, de modo que recaem no seguinte problema:

$$(\mathbf{D}^*, \mathbf{X}^*) = \arg \min_{\mathbf{D}, \mathbf{X}} \left\{ \sum_{i=1}^n \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \right\} \text{ sujeito a } \|\mathbf{x}_i\|_0 \leq s_0 \text{ para } i = 1, \dots, n, \quad (3.44)$$

onde  $s_0$  é um hiperparâmetro que controla a esparsidade das representações.

O MOD (*Method of Optimal Directions*) (ENGAN; AASE; HUSOY, 1999) e o K-SVD (AHARON; ELAD; BRUCKSTEIN, 2006) são dois dos métodos mais bem conhecidos para abordar o problema da Eq. (3.44) (PATEL; CHELLAPA, 2013). Tais métodos atualizam os valores de  $\mathbf{D}$  e  $\mathbf{X}$  iterativamente e podem ser descritos de uma maneira geral por meio dos seguintes passos:

1. Inicializar a matriz  $\mathbf{D}$  aleatoriamente, com a condição de que cada átomo seja  $\ell_2$ -normalizado, ou seja, com a condição de que  $\|\mathbf{d}_k\|_2 = 1$  para  $k = 1, \dots, K$ .
2. Atualizar os valores de  $\mathbf{X}$  mantendo a matriz  $\mathbf{D}$  fixa, o que reduz o problema da Eq. (3.44) à forma:

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \left\{ \sum_{i=1}^n \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \right\} \text{ sujeito a } \|\mathbf{x}_i\|_0 \leq s_0 \text{ para } i = 1, \dots, n. \quad (3.45)$$

Observe que a função objetivo da Eq. (3.45) pode ser reescrita como:

$$\sum_{i=1}^n \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 = \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \sum_{i' \neq i} \|\mathbf{y}_{i'} - \mathbf{D}\mathbf{x}_{i'}\|_2^2. \quad (3.46)$$

Além disso, fixando todas, exceto a  $i$ -ésima coluna de  $\mathbf{X}$ , recai-se no seguinte problema de otimização:

$$\mathbf{x}_i^* = \arg \min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 \text{ sujeito a } \|\mathbf{x}_i\|_0 \leq s_0. \quad (3.47)$$

Assim, nesse passo, as colunas de  $\mathbf{X}$  são atualizadas uma a uma, empregando-se algum método de codificação esparsa (e.g., OMP) afim de obter a solução do problema da Eq. (3.47) para atualizar a  $i$ -ésima coluna de  $\mathbf{X}$ .

3. Atualizar os valores de  $D$  mantendo a matriz  $X$  fixa, o que reduz o problema da Eq. (3.44) à forma:

$$\begin{aligned} D^* &= \arg \min_D \left\{ \sum_{i=1}^n \|y_i - D x_i\|_2^2 \right\} \\ &= \arg \min_D \{ \|Y - DX\|_F^2 \}, \end{aligned} \quad (3.48)$$

em que

$$\|A\|_F = \sqrt{\sum_{l=1}^{n_l} \sum_{c=1}^{n_c} A_{l,c}^2}, \quad (3.49)$$

denota a norma de Frobenious da matriz  $A \in \mathbb{R}^{n_l \times n_c}$ , com elementos  $A_{l,c}$ ,  $l = 1, \dots, n_l$ ,  $c = 1, \dots, n_c$ . Nesse passo é que diferenciam-se o MOD e o K-SVD, conforme detalhado a seguir.

4. Verificar a convergência. Caso o critério de convergência tenha sido satisfeito, encerrar. Caso contrário, retornar ao passo 2.

Conforme mencionado acima, o MOD e o K-SVD diferem-se somente no passo 3 (i.e., na etapa de atualização do dicionário). Especificamente, a regra de atualização de dicionário do MOD pode ser deduzida derivando a segunda linha da Eq. (3.48) em relação a  $D$ , igualando o resultado à matriz de zeros correspondente e manipulando a expressão resultante para obter:

$$D^* = Y X^T (X X^T)^{-1}. \quad (3.50)$$

No K-SVD, por sua vez, o dicionário é atualizado átomo por átomo, de uma maneira mais eficiente, a qual não exige a inversão de matrizes. Para isso, primeiramente, [Aharon, Elad e Bruckstein \(2006\)](#) observaram que a Eq. (3.48) pode ser escrita conforme segue:

$$\begin{aligned}
\mathbf{D}^* &= \arg \min_{\mathbf{D}} \left\{ \left\| \mathbf{Y} - \sum_{k=1}^K \mathbf{d}_k \underline{\mathbf{x}}_k^T \right\|_F^2 \right\} \\
&= \arg \min_{\mathbf{D}} \left\{ \left\| \mathbf{Y} - \mathbf{d}_k \underline{\mathbf{x}}_k^T - \sum_{k' \neq k} \mathbf{d}_{k'} \underline{\mathbf{x}}_{k'}^T \right\|_F^2 \right\} \\
&= \arg \min_{\mathbf{D}} \left\{ \left\| \left( \mathbf{Y} - \sum_{k' \neq k} \mathbf{d}_{k'} \underline{\mathbf{x}}_{k'}^T \right) - \mathbf{d}_k \underline{\mathbf{x}}_k^T \right\|_F^2 \right\}, \tag{3.51}
\end{aligned}$$

com  $\mathbf{d}_k$  denotando a  $k$ -ésima coluna (i.e., o  $k$ -ésimo átomo) de  $\mathbf{D}$  e  $\underline{\mathbf{x}}_k^T$  denotando a  $k$ -ésima linha de  $\mathbf{X}$ .

Os autores do K-SVD também observaram que fixando todos, exceto o  $k$ -ésimo átomo, o problema da Eq. (3.51) reduz-se à forma:

$$\mathbf{d}_k^* = \arg \min_{\mathbf{d}_k} \left\{ \left\| \mathbf{E}_k - \mathbf{d}_k \underline{\mathbf{x}}_k^T \right\|_F^2 \right\}, \tag{3.52}$$

de modo que

$$\mathbf{E}_k = \mathbf{Y} - \sum_{k' \neq k} \mathbf{d}_{k'} \underline{\mathbf{x}}_{k'}^T \tag{3.53}$$

denota a matriz de erros obtida quando o  $k$ -ésimo átomo é desconsiderado na representação em questão.

Considerando  $i_1, \dots, i_s$  os elementos de  $\text{sup}(\underline{\mathbf{x}}_k)$  ( $s = \text{card}(\text{sup}(\underline{\mathbf{x}}_k)) = \|\underline{\mathbf{x}}_k\|_0$ ), os mesmos definiram a matriz  $\mathbf{\Omega}_k \in \mathbb{R}^{n \times s}$ , de acordo com a equação abaixo:

$$\mathbf{\Omega}_k = [\delta_{i_1}, \dots, \delta_{i_s}], \tag{3.54}$$

em que  $\delta_i, i \in \text{sup}(\underline{\mathbf{x}}_k)$ , denota a  $i$ -ésima coluna da matriz identidade  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ .

Em seguida, atentando para o fato que o produto  $\tilde{\underline{\mathbf{x}}}_k^T = \underline{\mathbf{x}}_k^T \mathbf{\Omega}_k \in \mathbb{R}^{1 \times s}$  descarta os componentes nulos de  $\underline{\mathbf{x}}_k^T$  e que o produto  $\tilde{\mathbf{E}}_k = \mathbf{E}_k \mathbf{\Omega}_k \in \mathbb{R}^{m \times s}$  seleciona somente as colunas de  $\mathbf{E}_k$  correspondentes aos erros de representação das amostras (i.e., das colunas de  $\mathbf{Y}$ ) que efetivamente utilizam  $\mathbf{d}_k$ , **Aharon, Elad e Bruckstein (2006)** propuseram atualizar o dicionário resolvendo, para  $k = 1, \dots, K$ , o seguinte problema de otimização:

$$\begin{aligned}
(\mathbf{d}_k^*, \mathbf{x}_k^*) &= \arg \min_{(\mathbf{d}_k, \mathbf{x}_k)} \left\{ \left\| \mathbf{E}_k \boldsymbol{\Omega}_k - \mathbf{d}_k \mathbf{x}_k^T \boldsymbol{\Omega}_k \right\|_F^2 \right\} \\
&= \arg \min_{(\mathbf{d}_k, \mathbf{x}_k)} \left\{ \left\| \tilde{\mathbf{E}}_k - \mathbf{d}_k \tilde{\mathbf{x}}_k^T \right\|_F^2 \right\}.
\end{aligned} \tag{3.55}$$

Uma vez que o valor mínimo global da função objetivo da Eq. (3.55) ocorre quando a matriz  $\tilde{\mathbf{E}}_k$  é igual a  $\mathbf{d}_k \tilde{\mathbf{x}}_k^T$ , que é o produto externo (do inglês, *outer product*) entre  $\mathbf{d}_k$  e  $\tilde{\mathbf{x}}_k$ , tal matriz possui posto 1. Consequentemente, a decomposição em valores singulares (SVD – *Singular Value Decomposition*) de  $\tilde{\mathbf{E}}_k$  permite otimizar  $\mathbf{d}_k$  e  $\tilde{\mathbf{x}}_k^T$  (GOLUB; VAN LOAN, 2013). Especificamente, se  $\tilde{\mathbf{E}}_k = \sum_{i=1}^s \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ , com  $\sigma_1 > \dots > \sigma_s$  denotando os valores singulares de  $\tilde{\mathbf{E}}_k$  e  $\mathbf{u}_i$  e  $\mathbf{v}_i$  denotando os vetores singulares associados a  $\sigma_i$ , o ponto de ótimo global ocorre quando  $\mathbf{d}_k = \mathbf{u}_1$  e  $\tilde{\mathbf{x}}_k^T = \sigma_1 \mathbf{v}_1^T$ .

Note que esse método, após cada etapa de atualização do dicionário, preserva as colunas de  $\mathbf{D}$  normalizadas e mantém (ou eventualmente aumenta) a esparsidade das representações. Note-se que o nome K-SVD é uma alusão ao fato de que enquanto o K-médias atualiza o dicionário computando  $K$  médias, o K-SVD atualiza o dicionário realizando  $K$  decomposições em valores singulares (SVDs) (AHARON; ELAD; BRUCKSTEIN, 2006). Resultados experimentais sugerem que o K-SVD não só converge mais rápido que o MOD como também tende a produzir dicionários/representações mais robusto(a)s (ELAD, 2010).

### 3.2.5 Nonnegative Matrix Factorization (NMF)

Nos casos em que o conjunto de amostras fornecido pode ser organizado na forma de uma matriz não negativa<sup>7</sup>  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{m \times n}$ , para um dado  $K < \min\{m, n\}$ , pode-se desejar determinar duas matrizes também não negativas  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbb{R}^{m \times K}$  (i.e., um dicionário) e  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{K \times n}$  (i.e., um conjunto de representações) tais que o produto  $\mathbf{DX}$  aproxime  $\mathbf{Y}$  (i.e., fatoração de matrizes não negativas ou fatoração em matrizes não negativas – NMF) (LEE; SEUNG, 1999). Especificamente, pode-se estipular a solução ideal

<sup>7</sup>Uma matriz é dita não negativa se e somente se todos os seus elementos são maiores ou iguais a zero.

desse problema como sendo:

$$(\mathbf{D}^*, \mathbf{X}^*) = \arg \min_{\mathbf{D}, \mathbf{X}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \right\}. \quad (3.56)$$

De um modo geral, tal problema tende a produzir representações esparsas, uma vez que permite somente coeficientes não negativos nas combinações lineares e que existe uma íntima relação entre não negatividade e esparsidade (LI; NGOM, 2013). No entanto, uma desafio importante a ser considerado na otimização da Eq. (3.56) é a existência de mínimos locais devido à não convexidade da função objetivo. Na prática, diversos métodos de otimização podem ser empregados para abordar esse problema, dentre os quais pode-se destacar o método de iteração multiplicativa, o método do gradiente descendente e o método ALS (*Alternating Least Squares* – mínimos quadrados alternados). Em geral, o ALS converge mais rápido, de forma mais consistente e é menos sensível à inicialização (BERRY et al., 2007).

Basicamente, o ALS explora o fato de que embora a função objetivo da Eq. (3.56) não seja convexa em relação a  $\mathbf{D}$  e  $\mathbf{X}$ , a mesma é convexa em relação a  $\mathbf{D}$ , caso a matriz  $\mathbf{X}$  seja considerada fixa, e é convexa em relação a  $\mathbf{X}$ , caso a matriz  $\mathbf{D}$  seja considerada fixa. Especificamente, derivando tal função em relação a  $\mathbf{X}$  e mantendo a matriz  $\mathbf{D}$  fixa, segue que:

$$\mathbf{X} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{Y}. \quad (3.57)$$

Por sua vez, derivando a função em questão em relação a  $\mathbf{D}$  e mantendo a matriz  $\mathbf{X}$  fixa, tem-se que:

$$\mathbf{D} = \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1}. \quad (3.58)$$

Assim, o método ALS para NMF pode ser descrito através dos seguintes passos:

1. Inicializar as matrizes  $\mathbf{D}$  e  $\mathbf{X}$  aleatoriamente.
2. Atualizar os valores de  $\mathbf{X}$  mantendo a matriz  $\mathbf{D}$  fixa, de acordo com a Eq. (3.57).
3. Substituir todos os valores negativos de  $\mathbf{X}$  por zero.

4. Atualizar os valores de  $D$  mantendo a matriz  $X$  fixa, de acordo com a Eq. (3.58).
5. Substituir todos os valores negativos de  $D$  por zero.
6. Verificar a convergência. Caso o critério de convergência não tenha sido satisfeito, retornar ao passo 2.
7. Dividir todos elementos em uma mesma linha de  $X$  pela norma  $\ell_2$  da linha.
8. Multiplicar todos os elementos em uma mesma coluna de  $D$  pela norma  $\ell_2$  computada no passo 7 para a linha de  $X$  correspondente.
9. Encerrar.

### 3.2.6 Information-Theoretic Dictionary Learning (ITDL)

Considere o problema de construir um dicionário otimizado  $D^*$  através da seleção de um subconjunto dos átomos de um dado dicionário inicial  $D^{(0)}$ . Esse problema foi abordado por [Qiu, Patel e Chellappa \(2014\)](#) de maneira supervisionada considerando, efetivamente, as entradas listadas a seguir:

- um conjunto de amostras de treinamento  $\{\mathbf{y}_i\}_{i=1}^n$ , expresso no formato matricial  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{m \times n}$ ;
- um conjunto com os rótulos das amostras de treinamento, descrito pelo vetor  $\mathbf{r} = [r_1, \dots, r_n]^T$  em que  $r_i \in \{1, \dots, n_r\}$ ,  $i = 1, \dots, n$ , permite identificar a classe de  $\mathbf{y}_i$ ;
- um conjunto  $\{\mathbf{d}_k^{(0)}\}_{k=1}^K$  de átomos de um dicionário inicial, na forma  $\mathbf{D}^{(0)} = [\mathbf{d}_1^{(0)}, \dots, \mathbf{d}_K^{(0)}] \in \mathbb{R}^{m \times K}$ , bem como um conjunto  $\{\mathbf{x}_k^{(0)}\}_{k=1}^K$  no qual  $\mathbf{x}_k^{(0)}$  é o vetor com os coeficientes de representação associados ao átomo  $\mathbf{d}_k^{(0)}$ , na forma  $\mathbf{X}^{(0)} = [\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_K^{(0)}]^T \in \mathbb{R}^{K \times n}$ ; <sup>8</sup>
- um valor para o hiperparâmetro  $K' < K$ , o qual especifica a quantidade de átomos a serem selecionados em  $\mathbf{D}^{(0)}$  para compor  $D$ .

---

<sup>8</sup> $\mathbf{D}^{(0)}$  e  $\mathbf{X}^{(0)}$  podem ser previamente construídos a partir de  $\mathbf{Y}$  utilizando algum método de obtenção de dicionário (e.g., K-SVD).

A partir dessas entradas, e com base na teoria da informação (SHANNON, 1948), Qiu, Patel e Chellappa (2014) sugeriram estipular o dicionário ótimo por:

$$D^* = \arg \max_D \left\{ \lambda_1 \mathbb{I}(D, Y) + \lambda_2 \mathbb{I}(D, D^{(0)} - D) + \lambda_3 \mathbb{I}(X_D, r) \right\}, \quad (3.59)$$

onde  $\lambda_1$ ,  $\lambda_2$  e  $\lambda_3$  são os parâmetros que balanceiam as importâncias dadas a aspectos desejáveis em um dicionário: a representatividade  $\mathbb{I}(D, Y)$ , a compactividade  $\mathbb{I}(D, D^{(0)} - D)$  e a discriminabilidade  $\mathbb{I}(X_D, r)$ ;  $\mathbb{I}(A, B)$  denota a informação mútua entre os conjuntos  $A$  e  $B$ , a qual é dada por (COVER; THOMAS, 2006):

$$\mathbb{I}(A, B) = \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \mathbb{P}(\mathbf{a}_i \cap \mathbf{b}_j) \log \left( \frac{\mathbb{P}(\mathbf{a}_i \cap \mathbf{b}_j)}{\mathbb{P}(\mathbf{a}_i) \mathbb{P}(\mathbf{b}_j)} \right), \quad (3.60)$$

com  $\{\mathbf{a}_i\}_{i=1}^{n_A}$  e  $\{\mathbf{b}_j\}_{j=1}^{n_B}$  denotando, respectivamente, os possíveis valores de  $A$  e  $B$ .

Basicamente, a representatividade mede a quantidade de informação que os átomos em  $D$  contêm sobre as amostras em  $Y$ . De forma análoga, a compactividade do dicionário mensura a quantidade de informação que ele carrega sobre os átomos que estavam no dicionário inicial e foram deixados de fora da sua composição, assim como a discriminabilidade relaciona as representações em  $X_D$  com os rótulos em  $r$ .<sup>9</sup>

Na prática, além do dicionário ser coerente com as amostras que serão representadas nele (i.e., ser representativo), deseja-se que o mesmo seja compacto, ou seja, composto por átomos incoerentes de modo a encorajar amostras próximas, as quais provavelmente pertencem à mesma classe, a serem descritas por coeficientes similares, em termos de um conjunto semelhante de átomos. Por outro lado, também deseja-se que o dicionário seja discriminativo e incentive amostras distantes, as quais provavelmente não pertencem à mesma classe, a serem descritas por conjuntos diferentes de átomos, ou, eventualmente, por um conjunto semelhante de átomos, porém, com coeficientes evidentemente distintos (QIU; JIANG; CHELLAPPA, 2011; QIU; PATEL; CHELLAPPA, 2014).

De fato, o problema da Eq. (3.59) é NP-difícil e possui uma função objetivo submodular (KRAUSE; SINGH; GUESTRIN, 2008). Por isso, Qiu, Patel e Chellappa (2014) propuseram um método voraz para a obtenção de um dicionário subótimo, referido nesta tese como método de aprendizado de dicionário base-

<sup>9</sup> $X_D$  denota a matriz com os coeficientes de  $X^{(0)}$  associados aos átomos contidos em  $D$ .

ado em teoria da informação (ITDL – *Information-Theoretic Dictionary Learning*), o qual pode ser sumarizado por meio das etapas abaixo:

1. Inicializar o conjunto de átomos selecionados como sendo um conjunto vazio:

$$\hat{\mathbf{D}} = \emptyset. \quad (3.61)$$

2. Para  $t = 1, \dots, K'$ , adicionar, na  $t$ -ésima coluna de  $\hat{\mathbf{D}}$ , o átomo:

$$\begin{aligned} \mathbf{d}_k^* = \arg \max_{\mathbf{d}_k \in \mathbf{D}^{(0)} - \hat{\mathbf{D}}} \left\{ \lambda_1 \left[ \mathbb{I}(\hat{\mathbf{D}} \cup \mathbf{d}_k, \mathbf{Y}) - \mathbb{I}(\hat{\mathbf{D}}, \mathbf{Y}) \right] \right. \\ \left. + \lambda_2 \left[ \mathbb{I}(\hat{\mathbf{D}} \cup \mathbf{d}_k, \mathbf{D}^{(0)} - (\hat{\mathbf{D}} \cup \mathbf{d}_k)) - \mathbb{I}(\hat{\mathbf{D}}, \mathbf{D}^{(0)} - \hat{\mathbf{D}}) \right] \right. \\ \left. + \lambda_3 \left[ \mathbb{I}(\mathbf{X}_{\hat{\mathbf{D}} \cup \mathbf{d}_k}, \mathbf{r}) - \mathbb{I}(\mathbf{X}_{\hat{\mathbf{D}}}, \mathbf{r}) \right] \right\}. \end{aligned} \quad (3.62)$$

Para estimar as informações mútuas envolvidas na Eq. (3.62), os conjuntos fornecidos como entrada para o problema,  $\mathbf{Y}$ ,  $\mathbf{r}$ ,  $\mathbf{D}^{(0)}$  e  $\mathbf{X}^{(0)}$ , e consequentemente os seus subconjuntos, são modelados por processos estocásticos Gaussianos (ou seja, os elementos desses conjuntos são associados a variáveis aleatórias com distribuições Gaussianas) (RASMUSSEN; WILLIAMS, 2006).

3. Computar as representações no dicionário estimado:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \|\mathbf{Y} - \hat{\mathbf{D}}\mathbf{X}\|_F^2 \right\} = \hat{\mathbf{D}}^\dagger \mathbf{Y}. \quad (3.63)$$

### 3.2.7 Sparse Representation-based Classification (SRC)

Ao contrário do ITDL, a maioria dos métodos supervisionados de aprendizado de dicionário partem da premissa de que cada átomo serve somente para representar amostras de uma determinada classe. Com base nisso, pode-se estruturar o dicionário geral como uma composição de subdicionários (i.e., dicionários de classe), os quais, por sua vez, são constituídos pelos átomos que servem a uma mesma classe. Em especial, o método SRC (*Sparse Representation-based Classification*) (WRIGHT et al., 2009), proposto originalmente para reconhecimento facial

e de particular interesse nesta tese, baseia as suas predições em um dicionário estruturado desse modo.

Especificamente, as amostras de treinamento rotuladas de acordo com a  $r$ -ésima classe,  $r \in \{1, \dots, n_r\}$ , podem ser organizadas na forma de um dicionário de classe:

$$\mathbf{D}^{(r)} = [\mathbf{d}_1^{(r)}, \dots, \mathbf{d}_{K_r}^{(r)}] \in \mathbb{R}^{m \times K_r}, \quad (3.64)$$

no qual cada átomo  $\mathbf{d}_k^{(r)} \in \mathbb{R}^m$ ,  $k = 1, \dots, K_r$ , é uma amostra de treinamento da  $r$ -ésima classe, rearranjada na forma de um vetor unidimensional. Assim, pode-se construir um único dicionário geral:

$$\mathbf{D} = [\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(n_r)}] \in \mathbb{R}^{m \times K}, \quad (3.65)$$

o qual é a concatenação dos dicionários  $\mathbf{D}^{(r)}$ ,  $r = 1, \dots, n_r$ , de modo que  $K = \sum_{r=1}^{n_r} K_r$ .

Para classificar uma amostra de teste, rearranjada na forma de um vetor unidimensional,  $\mathbf{y} \in \mathbb{R}^m$ , o SRC primeiramente aplica algum método tal como o OMP ou o FISTA a fim de obter uma representação esparsa de  $\mathbf{y}$  em  $\mathbf{D}$ , resultando em:

$$\mathbf{x} = [\mathbf{x}^{(1)T}, \dots, \mathbf{x}^{(n_r)T}]^T \in \mathbb{R}^K, \quad (3.66)$$

onde  $\mathbf{x}^{(r)} \in \mathbb{R}^{K_r}$ ,  $r = 1, \dots, n_r$ , é o bloco do vetor  $\mathbf{x}$  com os coeficientes correspondentes ao bloco de átomos referentes ao dicionário  $\mathbf{D}^{(r)}$ . Subsequentemente, realiza a classificação de  $\mathbf{y}$  como segue:

$$\hat{r} = f(\mathbf{y}; \mathbf{D}) = \arg \min_r \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{D}^{(r)} \mathbf{x}^{(r)}\|_2^2 \right\}. \quad (3.67)$$

Em um cenário ideal, tem-se que  $\mathbf{y} \in \text{col}(\mathbf{D}^{(r^*)})$  e  $\text{col}(\mathbf{D}^{(r_i)}) \cap \text{col}(\mathbf{D}^{(r_j)}) = \{\mathbf{0}_m\}$  para  $r_i \neq r_j$ , de maneira todas as entradas de  $\mathbf{x}$  devem ser nulas, exceto algumas daquelas associadas com a classe  $r^*$  (i.e.,  $\mathbf{x}^{(r)} = \mathbf{0}_{K_r}$  para todo  $r \neq r^*$ ). Em um cenário mais realista, espera-se que o vetor  $\mathbf{x}$  seja esparsa e que a maioria dos coeficientes não nulos de  $\mathbf{x}$  ativem átomos referentes ao dicionário da classe correta  $r^*$ , permitindo assim identificar a classe de  $\mathbf{y}$  por meio da Eq. (3.67).

### 3.3 Considerações Finais

Neste capítulo, foi abordada principalmente a questão da obtenção de representações esparsas e dicionários, com ênfase nos métodos que serviram de base para o desenvolvimento do sistema proposto nesta tese. Antes disso, foram apresentadas algumas definições e conceitos fundamentais para um melhor entendimento acerca desses métodos e dos métodos propostos. Estes últimos compõem o sistema desenvolvido e serão detalhados no capítulo a seguir (i.e., no Capítulo 4), o qual também descreve a integração entre os métodos propostos e o funcionamento do sistema como um todo.

## 4 REPRESENTAÇÕES ESPARSAS PARA PROCESSAMENTO E ANÁLISE DE IMAGENS MACROSCÓPICAS DE LESÕES MELANOCÍTICAS

Neste capítulo, é apresentado o sistema de processamento e análise de imagens macroscópicas de lesões melanocíticas proposto nesta tese, cujo processo de operação passa pelas etapas introduzidas no Capítulo 2, a saber, pré-processamento, extração de características pré-segmentação, segmentação, extração de características pré-classificação e classificação, de acordo com o fluxo de execução apresentado na Figura 2.2.

Para fins de organização, as etapas de pré-processamento, extração de características pré-segmentação e segmentação são abordadas conjuntamente na Seção 4.1 e são etapas do método de segmentação proposto. As etapas de extração de características pré-classificação e classificação, por sua vez, são tratadas na Seção 4.2 e, simultaneamente, denominadas “o método de classificação proposto”. Cabe mencionar que tanto o método de segmentação quanto o método de classificação propostos, os quais estão detalhados nas Seções 4.1 e 4.2, baseiam-se nas técnicas discutidas anteriormente no Capítulo 3. As considerações finais a respeito deste capítulo, por fim, podem ser encontradas na Seção 4.3.

### 4.1 Representações Esparsas para Segmentação de Imagens Macroscópicas de Lesões Melanocíticas

A primeira etapa do método de segmentação proposto consiste em converter a imagem de entrada, pré-processada conforme detalhado na Subseção 4.1.1, em uma imagem de intensidades, com o intuito de evidenciar as dissimilaridades entre as cores dos pixels de lesão e não lesão, bem como para simplificar as etapas subsequentes do método. Esse processo de obtenção da imagem de intensidades, descrito na Subseção 4.1.2, é uma contribuição original do nosso trabalho, a qual foi publicada primeiramente em (ZORTEA; FLORES; SCHARCANSKI, 2017).

A aplicação direta de métodos robustos de agrupamento – e.g., NGC (SHI; MALIK, 2000) – para segmentar a imagem de intensidades obtida, em nível de pixels, demandaria recursos computacionais consideráveis. Para lidar com essa questão, a imagem de intensidades é caracterizada de modo que os blocos (*pa-*

*tches*) quadrados da imagem são representados por meio de combinações lineares de alguns poucos blocos geradores básicos (i.e., átomos de um dicionário), conforme tratado na Subseção 4.1.3.

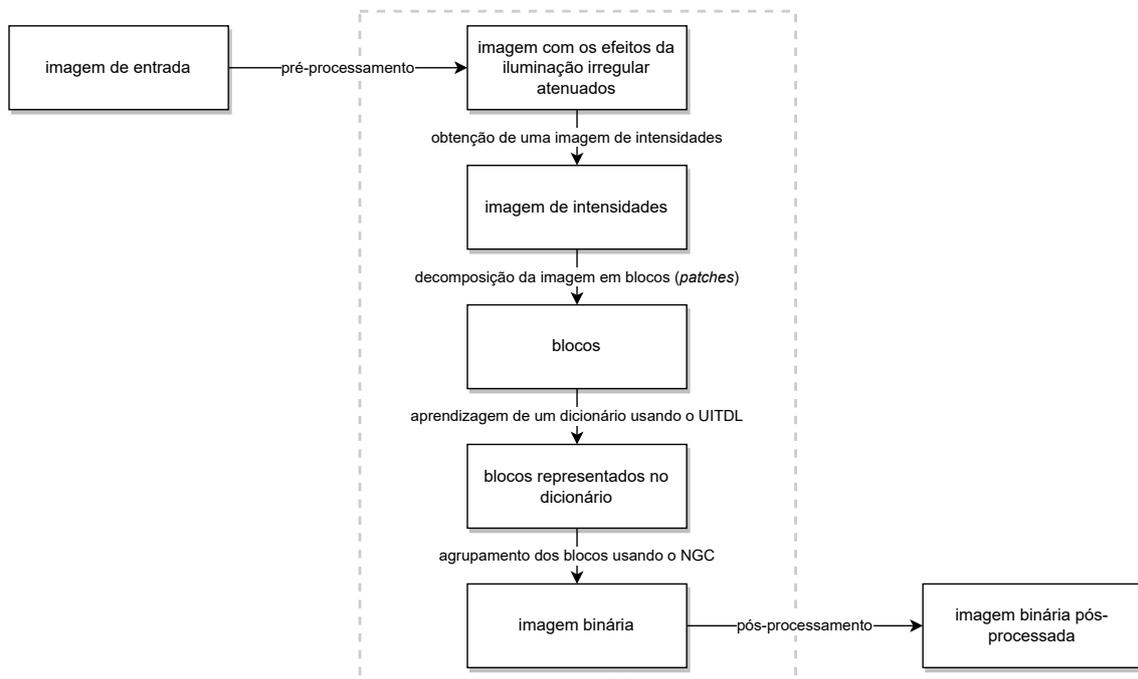
Estritamente falando, o dicionário utilizado é constituído por um subconjunto dos átomos de um dicionário inicial, o qual é aprendido via NMF a partir dos blocos quadrados da imagem, com o objetivo de produzir um dicionário otimizado em termos de compactividade e representatividade. Cabe mencionar que esse esquema de otimização de dicionário, intitulado UITDL (*Unsupervised Information-Theoretic Dictionary Learning*), é uma versão não supervisionada e adaptativa do método ITDL proposto em (QIU; PATEL; CHELLAPPA, 2014).<sup>1</sup> Tal adaptação é uma contribuição original do nosso trabalho, a qual foi publicada em (FLORES; SCHARCANSKI, 2016).

Por fim, o método NGC é usado para particionar as representações dos blocos no dicionário otimizado em dois grupos, resultando em uma imagem binária, a qual rotula os pixels da imagem como “lesão” ou “pele normal”, e é pós-processada para padronizar a identificação e remover eventuais artefatos indesejáveis, conforme tratado na Subseção 4.1.4. A Figura 4.1 ilustra todo o fluxo de trabalho do método de segmentação proposto.

---

<sup>1</sup>Esse método foi apresentado em detalhes na Subseção 3.2.6

Figura 4.1: Fluxo de trabalho do método proposto para a segmentação de lesões melanocíticas em imagens macroscópicas.



Fonte: Os Autores.

#### 4.1.1 Pré-processamento da Imagem de Entrada

Imagens macroscópicas usualmente são adquiridas em condições de iluminação irregular e, portanto, tendem a conter áreas sombreadas, as quais podem ser confundidas com partes de lesões de pele. A fim de evitar que tais artefatos induzam o método proposto a cometer esse tipo de erro, a imagem de entrada, antes de ser efetivamente processada pelo método de segmentação, é pré-processada pelo método de atenuação de sombra proposto em (CAVALCANTI; SCHARCANSKI; LOPES, 2010), o qual consiste em executar as seguintes etapas:

1. converter a imagem de entrada do espaço de cores RGB para o espaço de cores HSV;
2. computar o canal  $V^*$  através da seguinte equação:

$$V^*(l,c) = \frac{V(l,c)}{\theta_1 l^2 + \theta_2 c^2 + \theta_3 lc + \theta_4 l + \theta_5 c + \theta_6'} \quad (4.1)$$

onde  $V(l,c)$  denota a intensidade do pixel indexado pelo par ordenado  $(l,c)$  no canal  $V$  (i.e., no terceiro canal do espaço de cores HSV) e  $\{\theta_i\}_{i=1}^6$  é o

conjunto de parâmetros, ajustados pelo método de regressão por mínimos quadrados ordinários a partir dos pixels localizados nos quatro cantos  $\kappa \times \kappa$  do canal  $V$ , do modelo quadrático de iluminação;<sup>2</sup>

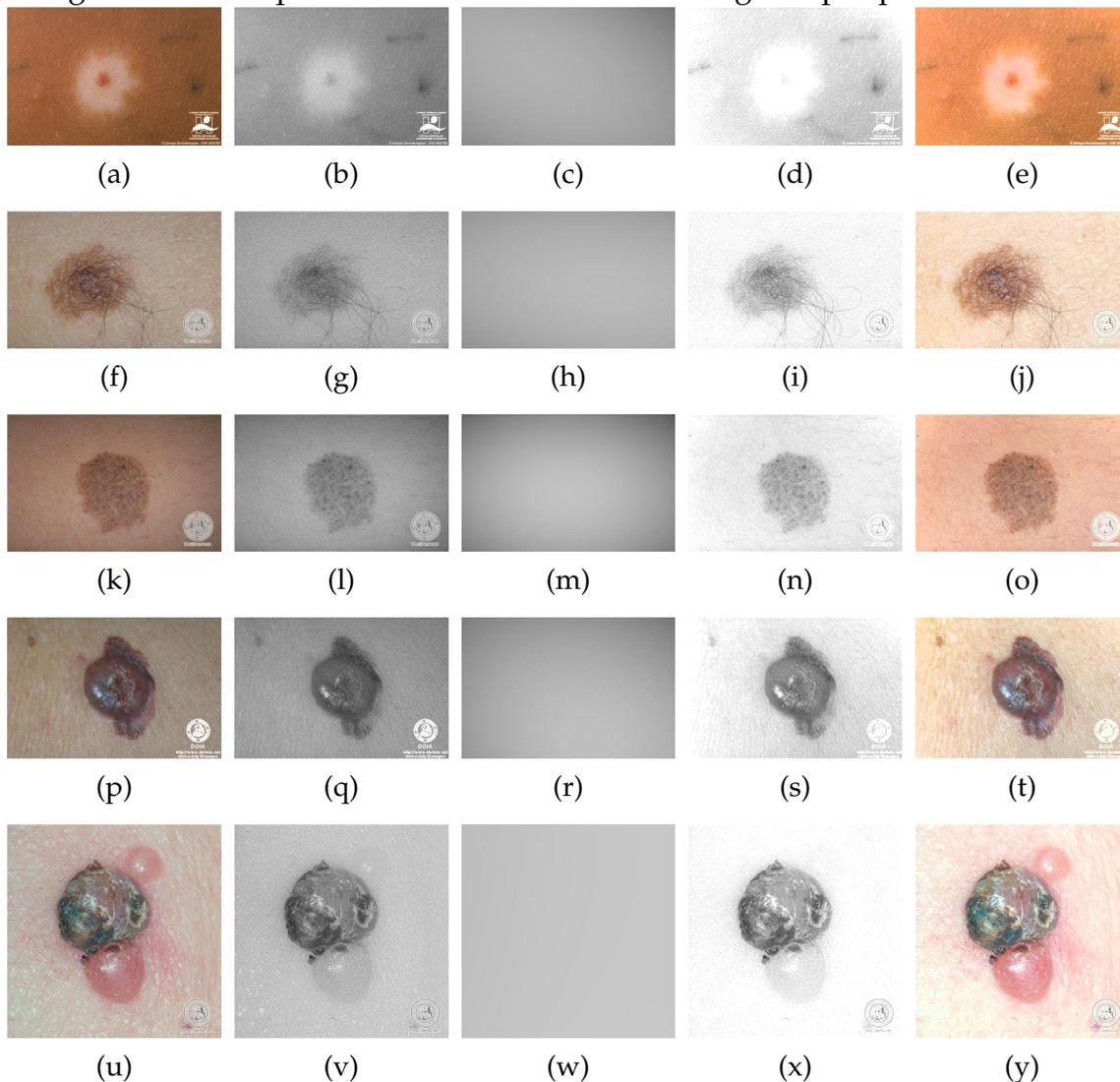
3. substituir o canal  $V$  original pelo canal  $V^*$ ;
4. converter a imagem de volta do espaço de cores HSV para o espaço de cores RGB.

Na Figura 4.2, em cada linha, é apresentado um exemplo ao longo das etapas do pré-processamento. Especificamente, na primeira coluna, as Subfiguras 4.2a, 4.2f, 4.2k, 4.2p e 4.2u mostram algumas imagens macroscópicas de lesões melanocíticas. Na segunda coluna, as Subfiguras 4.2b, 4.2g, 4.2l, 4.2q e 4.2v apresentam os canais  $V$  correspondentes às imagens da primeira coluna. Na terceira coluna, as Subfiguras 4.2c, 4.2h, 4.2m, 4.2r e 4.2w mostram os modelos  $Z$  obtidos a partir dos canais  $V$  da segunda coluna. Na quarta coluna, as Subfiguras 4.2d, 4.2i, 4.2n, 4.2s e 4.2x apresentam os canais  $V^*$  computados e, por fim, na quinta coluna, as Subfiguras 4.2e, 4.2j, 4.2o, 4.2t e 4.2y mostram os resultados após a conclusão do pré-processamento (i.e., as imagens com os efeitos da iluminação irregular atenuados).

---

<sup>2</sup>Em todos os experimentos realizados nesta tese, utilizou-se  $\kappa = \frac{1}{5} \min\{n_r, n_c\}$ .

Figura 4.2: Exemplos de resultados obtidos ao longo do pré-processamento.



Fonte: Os Autores.

#### 4.1.2 Obtenção de uma Imagem de Intensidades

A partir da imagem de entrada pré-processada  $\mathcal{I} \in \mathbb{R}^{n_l \times n_c \times n_b}$ , a primeira etapa do método de segmentação proposto produz uma imagem de intensidades  $I \in \mathbb{R}^{n_l \times n_c}$ . Em termos gerais, a obtenção da imagem de intensidades começa com a realização de uma busca por uma região retangular, a qual espera-se que seja composta predominantemente por pixels de não lesão. Em seguida, é determinada a cor média da região selecionada. Por fim, uma distância é calculada entre essa cor e a cor de cada pixel da imagem de entrada, resultando em uma imagem de intensidades.

Durante a busca, são avaliadas diversas regiões retangulares não sobre-

postas e com lados paralelos às bordas externas da imagem (tais como as regiões delimitadas por linhas pretas ou azuis na Figura 4.3), começando com a região mais externa e movendo-se em direção ao centro da imagem até percorrer uma determinada quantidade de pixels. Cada linha da Figura 4.3 apresenta um exemplo de busca realizada. Tais buscas começam avaliando as regiões mostradas nas colunas mais à esquerda, isso é, as regiões das Subfiguras 4.3a, 4.3e, 4.3i, 4.3m e 4.3r, e terminam avaliando as regiões mostradas nas colunas mais à direita, isso é, as regiões das Subfiguras 4.3d, 4.3h, 4.3l, 4.3q e 4.3x. Em todos os experimentos realizados nesta tese, estabeleceu-se que a busca deve ser por regiões disjuntas com espessuras fixas de  $\frac{1}{50} \max\{n_r, n_c\}$  e encerrar após percorrer  $\frac{1}{4} \min\{n_r, n_c\}$  pixels.

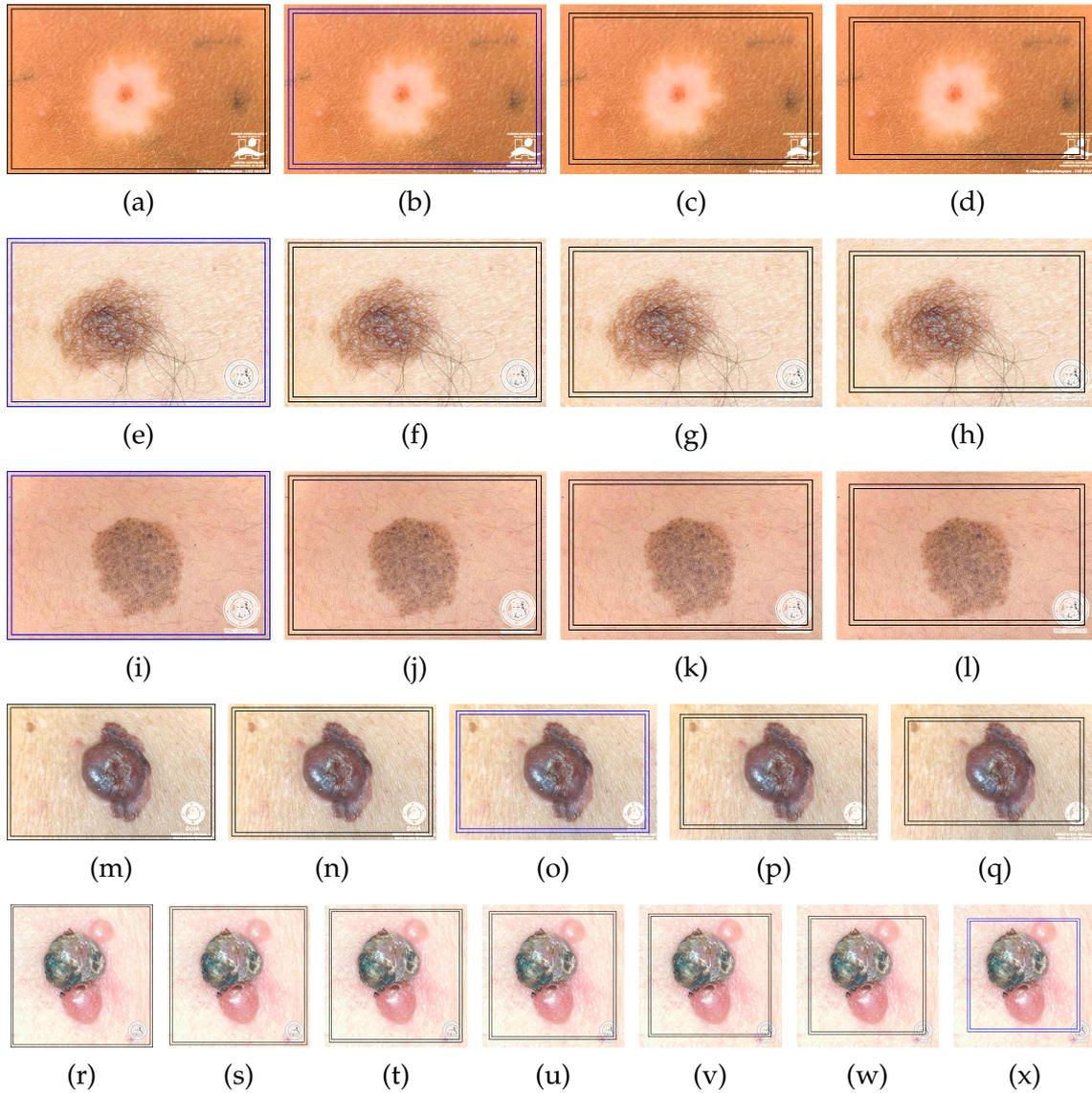
Especificamente, estipula-se que a região ideal é, dentre aquelas avaliadas, a que minimiza o coeficiente total de variação  $\sum_{b=1}^3 \sigma_b / \mu_b$ , no qual  $\sigma_b$  e  $\mu_b$ ,  $b = 1, 2, 3$ , denotam, respectivamente, o desvio padrão e a média das intensidades da imagem de entrada no  $b$ -ésimo canal do espaço de cores CIELAB, considerando somente as intensidades das posições correspondentes à região em questão. Em especial, as regiões ideais são delimitadas por linhas azuis na Figura 4.3. Chama-se a atenção do leitor para o fato que o coeficiente total de variação diminui à medida que a região possui tons mais homogêneos, claros e/ou avermelhados, como é esperado no caso de regiões que contêm prevalentemente pixels de não lesão.

Em suma, a região ideal é selecionada com base no coeficiente total de variação e a sua cor média, no espaço de cores CIELAB,  $(L_0, A_0, B_0)$ , é determinada. Subsequentemente, a imagem de intensidades é computada por meio de:

$$I(l, c) = \sqrt{(L(l, c) - L_0)^2 + (A(l, c) - A_0)^2 + (B(l, c) - B_0)^2}, \quad (4.2)$$

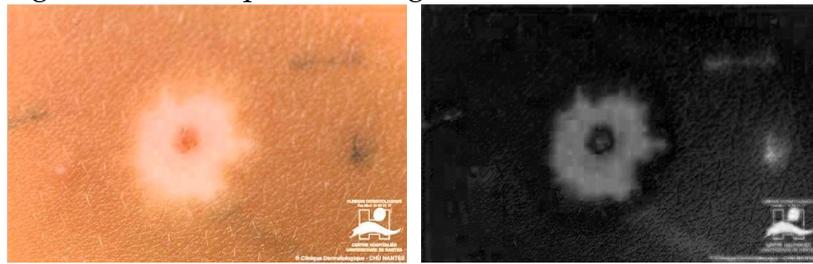
onde  $(l, c) \in \{1, 2, \dots, n_l\} \times \{1, 2, \dots, n_c\}$  e  $(L(l, c), A(l, c), B(l, c))$  denota a cor do pixel localizado na  $l$ -ésima linha e na  $c$ -ésima coluna da imagem de entrada, representada no espaço de cores CIELAB. A Figura 4.4 apresenta alguns exemplos de imagens de intensidades obtidas a partir de imagens de entrada pré-processadas.

Figura 4.3: Exemplos de regiões retangulares avaliadas para produzir imagens de intensidades.



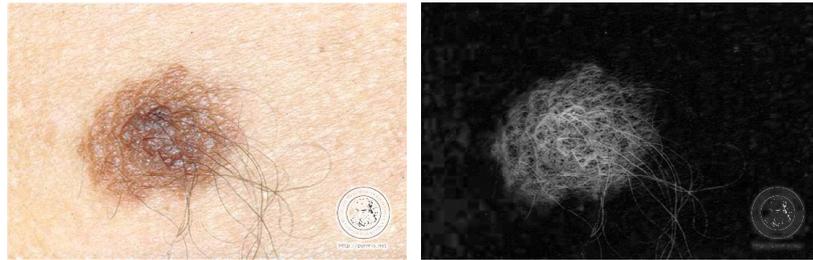
Fonte: Os Autores.

Figura 4.4: Exemplos de imagens de intensidades obtidas.



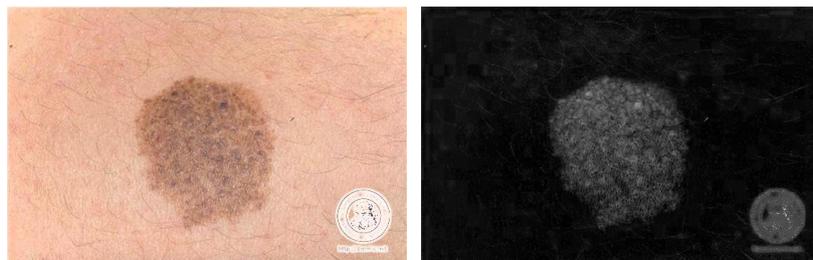
(a)

(b)



(c)

(d)



(e)

(f)



(g)

(h)



(i)

(j)

Fonte: Os Autores.

### 4.1.3 Aprendizado de um Dicionário usando UITDL

Antes de efetivamente iniciar o processo de aprendizado de dicionário, a imagem de intensidades,  $\mathbf{I} \in \mathbb{R}^{n_l \times n_c}$ , é decomposta em  $n = \lfloor \frac{n_l}{w} \rfloor \lfloor \frac{n_c}{w} \rfloor$  blocos não sobrepostos de dimensões  $w \times w$ . Em seguida, é organizado um conjunto de sinais na forma de uma matriz não negativa  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$ , na qual  $\mathbf{y}_i \in \mathbb{R}^m$ ,  $i = 1, \dots, n$ , é um bloco rearranjado na forma de um vetor unidimensional (i.e.,  $m = w^2$ ). Em todos os experimentos efetuados nesta tese, utilizou-se  $w$  como sendo o inteiro ímpar mais próximo de  $\frac{1}{100} \max\{n_l, n_c\}$ .

Subsequentemente, o processo de aprendizado de dicionário começa com a aplicação do método ALS para NMF<sup>3</sup>, o qual, a partir de  $\mathbf{Y}$  e para um dado  $K < \min\{m, n\}$ ,<sup>4</sup> resulta em duas matrizes não negativas:

1.  $\mathbf{D}^{(0)} = [\mathbf{d}_1^{(0)}, \dots, \mathbf{d}_K^{(0)}] \in \mathbb{R}^{m \times K}$  – o dicionário inicial, em que cada coluna  $\mathbf{d}_k^{(0)}$ ,  $k = 1, \dots, K$ , é um átomo do dicionário; e
2.  $\mathbf{X}^{(0)} = [\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_n^{(0)}] \in \mathbb{R}^{K \times n}$  – a representação inicial, onde cada coluna  $\mathbf{x}_i^{(0)}$ ,  $i = 1, \dots, n$ , representa  $\mathbf{y}_i$  em  $\mathbf{D}^{(0)}$ .

Cada linha da Figura 4.5 apresenta, na coluna à direita, isso é, nas Subfiguras 4.5b, 4.5d, 4.5f, 4.5h e 4.5j, um exemplo de dicionário inicial (cada átomo do dicionário foi rearranjado na forma de um bloco quadrado) obtido com  $K = 9$  a partir dos blocos  $5 \times 5$  da imagem correspondente na coluna à esquerda, isso é, nas Subfiguras 4.5a, 4.5c, 4.5e, 4.5g e 4.5i.

Basicamente, o UITDL é um método voraz que produz um dicionário otimizado a partir de um dicionário inicial, de modo que a seleção dos átomos ocorre iterativamente com base nos ganhos de compactividade e representatividade fornecidos pelos mesmos. Especificamente, dada a matriz dos blocos  $\mathbf{Y}$ , bem como o dicionário e a representação inicial (i.e., as matrizes  $\mathbf{D}^{(0)}$  e  $\mathbf{X}^{(0)}$  obtidas anteriormente), a etapa de aprendizado de dicionário do método proposto prossegue aplicando o UITDL, o qual efetua os passos enumerados a seguir:

1. Inicializar um conjunto de átomos (i.e., um dicionário) como sendo um conjunto vazio:

$$\tilde{\mathbf{D}} = \emptyset. \quad (4.3)$$

<sup>3</sup>O leitor pode consultar a Subseção 3.2.5 para mais detalhes.

<sup>4</sup> $K$  é um hiperparâmetro do método de segmentação proposto.

2. Para  $t = 1, \dots, K$ , adicionar, na  $t$ -ésima coluna de  $\tilde{\mathbf{D}}$ , o átomo

$$\mathbf{d}_k^* = \arg \max_{\mathbf{d}_k \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}} \left\{ \lambda_1 [\mathbb{I}(\tilde{\mathbf{D}} \cup \mathbf{d}_k, \mathbf{D}^{(0)} - (\tilde{\mathbf{D}} \cup \mathbf{d}_k)) - \mathbb{I}(\tilde{\mathbf{D}}, \mathbf{D}^{(0)} - \tilde{\mathbf{D}})] + \lambda_2 [\mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}} \cup \mathbf{d}_k) - \mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}})] \right\} \quad (4.4)$$

e registrar o par ordenado  $(t, f(t))$ , onde:

$$f(t) = \max_{\mathbf{d}_k \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}} \left\{ \lambda_1 [\mathbb{I}(\tilde{\mathbf{D}} \cup \mathbf{d}_k, \mathbf{D}^{(0)} - (\tilde{\mathbf{D}} \cup \mathbf{d}_k)) - \mathbb{I}(\tilde{\mathbf{D}}, \mathbf{D}^{(0)} - \tilde{\mathbf{D}})] + \lambda_2 [\mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}} \cup \mathbf{d}_k) - \mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}})] \right\}. \quad (4.5)$$

3. Determinar:

$$t^* = \arg \max_t \{t\} \text{ sujeito a } f(t) > 0. \quad (4.6)$$

4. Construir o dicionário otimizado,  $\mathbf{D}_{t^*}$ , o qual é composto pelas  $t^*$  primeiras colunas de  $\tilde{\mathbf{D}}$ .

5. Obter a representação  $\mathbf{X}$ , de  $\mathbf{Y}$  em  $\mathbf{D}_{t^*}$ , por meio de:

$$\mathbf{X} = \mathbf{D}_{t^*}^\dagger \mathbf{Y}. \quad (4.7)$$

Os dois termos que são multiplicados por  $\lambda_1$  e  $\lambda_2$  nas Eqs. (4.4) e (4.5) quantificam, respectivamente, os ganhos de compactividade e representatividade. Esses termos podem ser computados de forma conveniente levando em consideração os seguintes fatos:

1. Determinar um átomo  $\mathbf{d}_k^* \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}$  que maximize  $\mathbb{I}(\tilde{\mathbf{D}} \cup \mathbf{d}_k, \mathbf{D}^{(0)} - (\tilde{\mathbf{D}} \cup \mathbf{d}_k)) - \mathbb{I}(\tilde{\mathbf{D}}, \mathbf{D}^{(0)} - \tilde{\mathbf{D}})$ , através da modelagem de um processo Gaussiano, equivale a resolver o seguinte problema de otimização (QIU; JIANG; CHELLAPPA, 2011):

$$\mathbf{d}_k^* = \arg \max_{\mathbf{d}_k \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}} \left\{ \frac{\sigma^2(\mathbf{d}_k) - \boldsymbol{\sigma}_{\tilde{\mathbf{D}}_k}^T \boldsymbol{\Sigma}_{\tilde{\mathbf{D}}}^{-1} \boldsymbol{\sigma}_{\tilde{\mathbf{D}}_k}}{\sigma^2(\mathbf{d}_k) - \boldsymbol{\sigma}_{\tilde{\mathbf{D}}'_k}^T \boldsymbol{\Sigma}_{\tilde{\mathbf{D}}'}^{-1} \boldsymbol{\sigma}_{\tilde{\mathbf{D}}'_k}} \right\}, \quad (4.8)$$

no qual  $\sigma^2(\mathbf{d}_k)$  representa a variância do átomo  $\mathbf{d}_k$ ,  $\boldsymbol{\sigma}_{\tilde{\mathbf{D}}_k} \in \mathbb{R}^{t-1}$  denota o vetor com as covariâncias entre o átomo  $\mathbf{d}_k$  e os átomos em  $\tilde{\mathbf{D}}$  e  $\boldsymbol{\Sigma}_{\tilde{\mathbf{D}}} \in \mathbb{R}^{(t-1) \times (t-1)}$  é a matriz com as covariâncias entre todos os pares de átomos em  $\tilde{\mathbf{D}}$ . Além disso,  $\tilde{\mathbf{D}}' = \mathbf{D}^{(0)} - (\tilde{\mathbf{D}} \cup \mathbf{d}_k)$ ,  $\boldsymbol{\sigma}_{\tilde{\mathbf{D}}'_k} \in \mathbb{R}^{K-t+1}$  denota o vetor com as covariâncias entre o átomo  $\mathbf{d}_k$  e os átomos em  $\tilde{\mathbf{D}}'$  e  $\boldsymbol{\Sigma}_{\tilde{\mathbf{D}}'} \in \mathbb{R}^{(K-t+1) \times (K-t+1)}$  é a matriz com as covariâncias entre todos os pares de átomos em  $\tilde{\mathbf{D}}'$ .

2. Determinar um átomo  $\mathbf{d}_k^* \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}$  que maximize  $\mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}} \cup \mathbf{d}_k) - \mathbb{I}(\mathbf{Y}, \tilde{\mathbf{D}})$ , assumindo que os blocos em  $\mathbf{Y}$  sejam independentes, equivale a resolver o seguinte problema de otimização (QIU; PATEL; CHELLAPPA, 2014):

$$\mathbf{d}_k^* = \arg \max_{\mathbf{d}_k \in \mathbf{D}^{(0)} - \tilde{\mathbf{D}}} \left\{ \sum_{i=1}^n \mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}}) - \sum_{i=1}^n \mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k) \right\}, \quad (4.9)$$

em que  $\mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}})$  denota a entropia de  $\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}})$ , a qual é definida como:

$$\mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}}) = \sum_{i=1}^n \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}}) \log \left( \frac{1}{\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}})} \right). \quad (4.10)$$

A probabilidade condicional em questão pode ser estimada por:

$$\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}}) = \exp \left( -\frac{\|\mathbf{r}_i\|_2^2}{2\sigma^2(\mathbf{r}_i)} \right), \quad (4.11)$$

onde  $\mathbf{r}_i = \mathbf{y}_i - \tilde{\mathbf{D}}\mathbf{x}_i$  e  $\sigma^2(\mathbf{r}_i)$  representa a variância de  $\mathbf{r}_i$ ,  $\mathbf{x}_i$  denota a  $i$ -ésima coluna de  $\mathbf{X}_{\tilde{\mathbf{D}}}^{(0)}$  e  $\mathbf{X}_{\tilde{\mathbf{D}}}^{(0)}$  é a matriz formada pelas linhas de  $\mathbf{X}^{(0)}$  correspondentes às colunas selecionadas em  $\tilde{\mathbf{D}}$  (e.g., se  $\tilde{\mathbf{D}}$  é composta pelas 1ª, 3ª e 5ª colunas de  $\mathbf{D}^{(0)}$ , então  $\mathbf{X}_{\tilde{\mathbf{D}}}^{(0)}$  é formada pelas 1ª, 3ª e 5ª linhas de  $\mathbf{X}^{(0)}$ ).

Deve ser observado que a entropia  $\mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k)$  é definida de maneira análoga como:

$$\mathbb{H}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k) = \sum_{i=1}^n \mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k) \log \left( \frac{1}{\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k)} \right), \quad (4.12)$$

de modo que a probabilidade condicional na equação acima pode ser esti-

mada por:

$$\mathbb{P}(\mathbf{y}_i | \tilde{\mathbf{D}} \cup \mathbf{d}_k) = \exp\left(-\frac{\|\mathbf{r}'_i\|_2^2}{2\sigma^2(\mathbf{r}'_i)}\right), \quad (4.13)$$

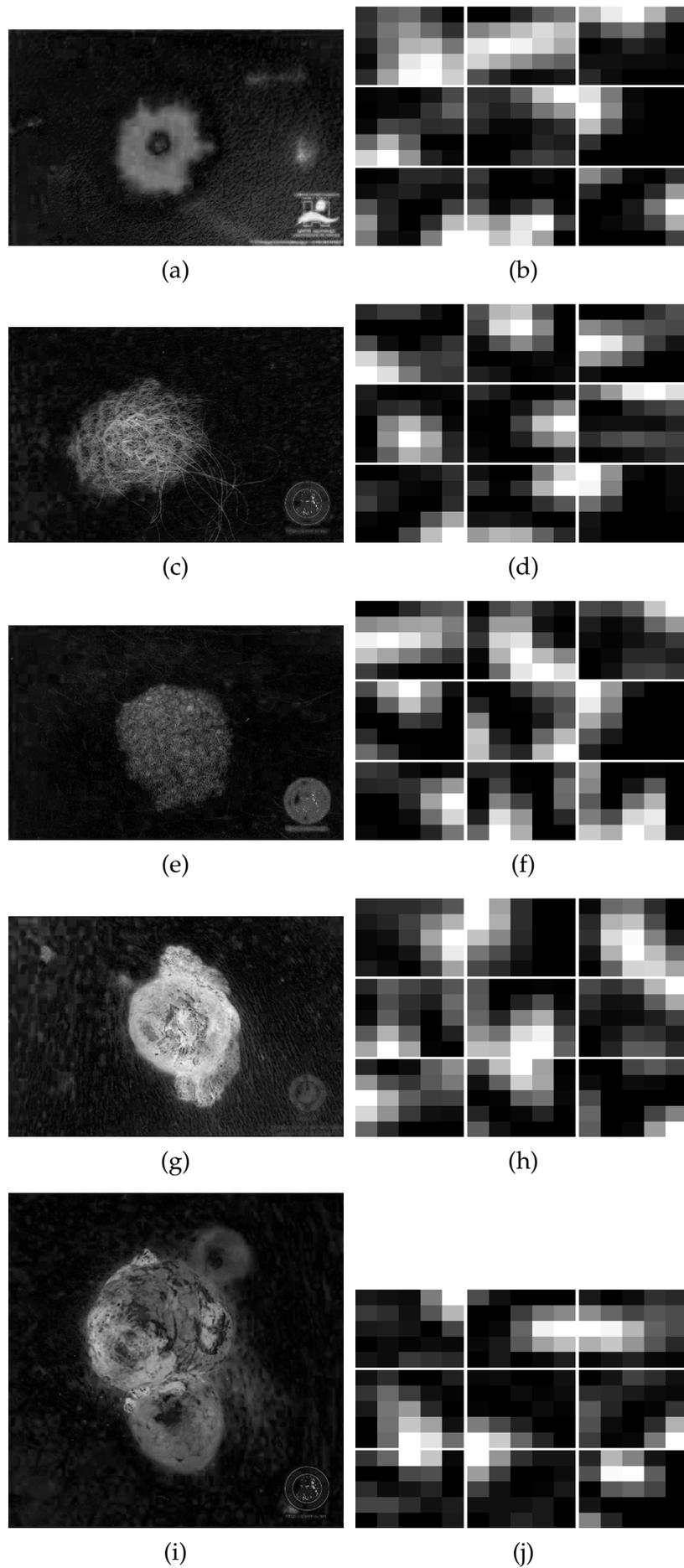
onde  $\mathbf{r}'_i = \mathbf{y}_i - (\tilde{\mathbf{D}} \cup \mathbf{d}_k)\mathbf{x}'_i$  e  $\sigma^2(\mathbf{r}'_i)$  representa a variância de  $\mathbf{r}'_i$ ,  $\mathbf{x}'_i$  denota a  $i$ -ésima coluna de  $\mathbf{X}_{\tilde{\mathbf{D}} \cup \mathbf{d}_k}^{(0)}$  e  $\mathbf{X}_{\tilde{\mathbf{D}} \cup \mathbf{d}_k}^{(0)}$  é a matriz formada pelas linhas de  $\mathbf{X}^{(0)}$  correspondentes às colunas em  $\mathbf{D}^{(0)} \cup \mathbf{d}_k$ .

Por fim, cabe mencionar que  $\lambda_1$  e  $\lambda_2$  não são hiperparâmetros do UITDL, mas sim parâmetros, extraídos diretamente a partir dos dados, que normalizam as contribuições dos ganhos de compactividade e representatividade, de acordo com:

$$\lambda_1 = 1, \quad (4.14)$$

$$\lambda_2 = \max_{k \in \{1, \dots, K\}} \left\{ \frac{\mathbb{I}(\mathbf{Y}, \mathbf{d}_k)}{\mathbb{I}(\mathbf{d}_k, \mathbf{D}^{(0)} - \mathbf{d}_k)} \right\}. \quad (4.15)$$

Figura 4.5: Exemplos de dicionários iniciais obtidos a partir de imagens de intensidades.



#### 4.1.4 Agrupamento dos Blocos e Pós-Processamento

Em particular, nesta etapa do método proposto, o método NGC (SHI; MALIK, 2000; COUR; YU; SHI, 2010) é primeiramente aplicado para minimizar a Eq. (2.17), resultando em duas partições:  $A$  e  $B$ . Então, essas partições são empregadas para gerar uma segmentação (i.e., uma máscara binária) inicial,  $M^{(0)} \in \{0,1\}^{n_l \times n_c}$ , de tal forma que  $M^{(0)}(l,c) = 0$  se o pixel indexado por  $(l,c)$  foi agrupado na partição  $A$  ou  $M^{(0)}(l,c) = 1$  se o pixel indexado por  $(l,c)$  foi agrupado na partição  $B$ .

Imagens macroscópicas, tipicamente, possuem outros tipos de artefatos indesejáveis, além dos efeitos de sombra, tais como lesões menores que não eram o foco da atenção no momento da aquisição da imagem, pêlos e reflexos, os quais também podem ser confundidos com a lesão de interesse e/ou produzir falhas na identificação da lesão. Para lidar com essas questões, o método proposto é equipado com algumas etapas simples de pós-processamento, as quais são aplicadas imediatamente após a segmentação, conforme descrito a seguir.

Primeiramente, com base na suposição de que a imagem de entrada foi adquirida de um modo que a maioria dos pixels sobre os quatro cantos  $\kappa \times \kappa$  corresponde à pele normal, é conveniente consultar as intensidades desses cantos em  $M^{(0)}$ . Concretamente, uma segmentação  $M \in \{0,1\}^{n_l \times n_c}$  é produzida a partir de  $M^{(0)}$ , de acordo com:

$$M(l,c) = \begin{cases} 1 - M^{(0)}(l,c), & \text{se a soma das intensidades nos quatro} \\ & \text{cantos } \kappa \times \kappa \text{ de } M^{(0)} \text{ é maior que } 2\kappa^2, \\ M^{(0)}(l,c), & \text{caso contrário,} \end{cases} \quad (4.16)$$

com o intuito de assegurar que os pixels identificados como lesão possuam intensidade “1” e os pixels identificados como pele normal possuam intensidade “0”.

Depois, é mantida somente uma das regiões previamente identificadas como lesão. Especificamente, aquela cujo conjunto de pares ordenados correspondente é:

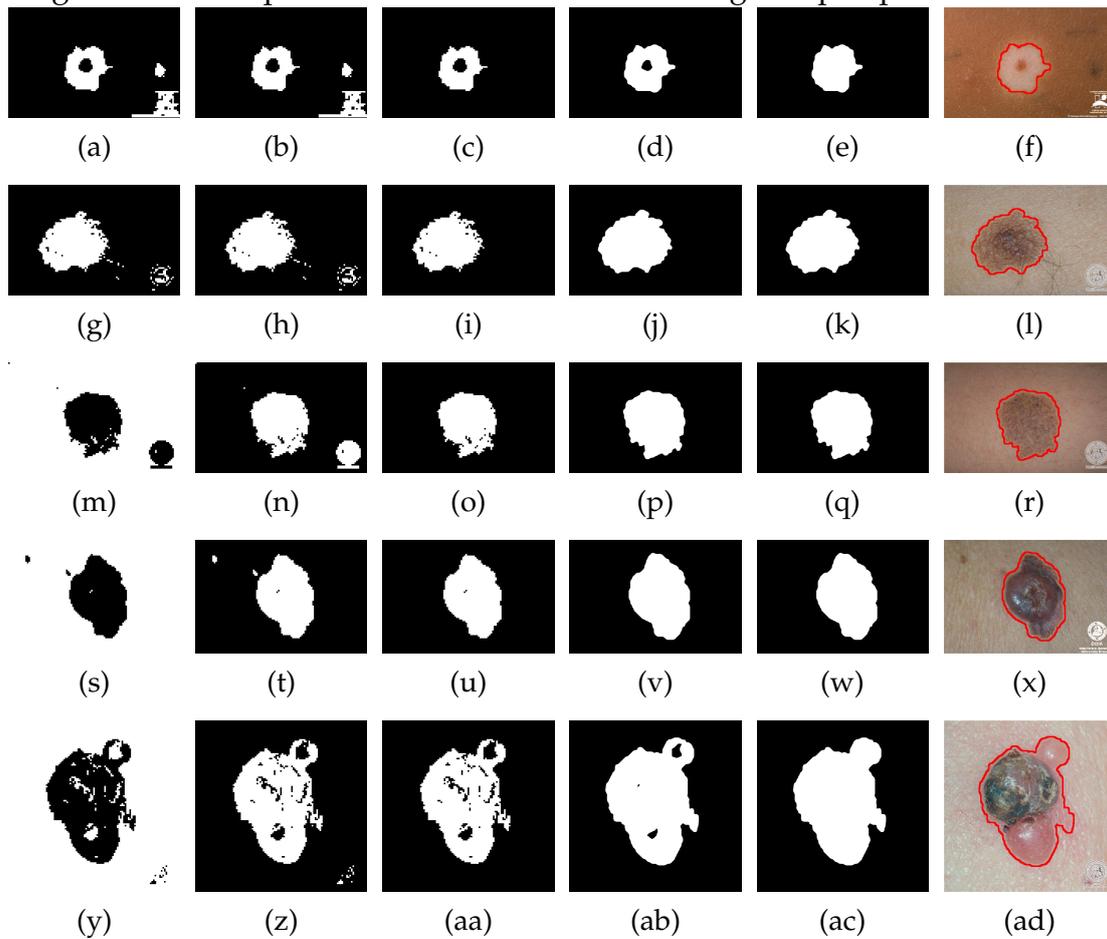
$$\psi_h^* = \arg \max_{\psi_h} \left\{ \sum_{(l,c) \in \psi_h} \exp \left( -\frac{(l-l_0)^2 + (c-c_0)^2}{2\sigma^2} \right) \right\}, \quad (4.17)$$

onde  $\psi_h \in \{\psi_h\}_{h=1}^{h=H}$  é um conjunto de pares ordenados, o qual contém as coordenadas de todos os pixels de uma mesma região, de acordo com  $M$ , identificada como sendo uma região de lesão;  $(l_0, c_0) = (\lfloor n_l/2 \rfloor, \lfloor n_c/2 \rfloor)$  denota o par ordenado com as coordenadas do pixel central de  $M$  e  $\sigma$  é um hiperparâmetro configurável (em todos os experimentos realizados nesta tese, foi utilizado  $\sigma = 0,1 \max\{n_r, n_c\}$ ).

Em termos gerais, a função objetivo da Eq. (4.17) aumenta à medida que uma determinada região cresce e/ou aproxima-se do centro da imagem. Consequentemente, essa etapa do pós-processamento retém apenas a região “mais destacada”, a qual espera-se que preponderadamente corresponda à lesão de interesse. Subsequentemente, as bordas dessa região são suavizadas aplicando um filtro Gaussiano e, em seguida, uma limiarização (em todos os experimentos, foi utilizado um filtro Gaussiano com dimensões  $3w \times 3w$  e desvio padrão  $w$ , combinado com um limiar fixo em 0,1). Por fim, em caso de haver necessidade, tal região ainda é preenchida por um método de preenchimento de regiões.

Na Figura 4.6, em cada linha, é apresentado um exemplo ao longo das etapas de pós-processamento. Concretamente, na primeira coluna, as Subfiguras 4.6a, 4.6g, 4.6m, 4.6s e 4.6y mostram alguns resultados obtidos imediatamente após a segmentação usando NGC. Na segunda coluna, as Subfiguras 4.6b, 4.6h, 4.6n, 4.6t e 4.6z apresentam esses resultados logo depois da verificação dos quatro cantos  $\kappa \times \kappa$  e do eventual complemento das intensidades. Na terceira coluna, as Subfiguras 4.6c, 4.6i, 4.6o, 4.6u e 4.6aa mostram os resultados em questão imediatamente após a seleção da região de interesse. Na quarta coluna, as Subfiguras 4.6d, 4.6j, 4.6p, 4.6v e 4.6ab apresentam tais resultados logo depois da suavização das bordas. Na quinta coluna, as Subfiguras 4.6e, 4.6k, 4.6q, 4.6w e 4.6ac mostram os resultados após o eventual preenchimento de regiões. Por fim, na sexta coluna, as Subfiguras 4.6f, 4.6l, 4.6r, 4.6x e 4.6ad apresentam as mesmas informações disponíveis nas imagens da quinta coluna, porém, alternativamente representadas através de um contorno vermelho sobposto à imagem de entrada.

Figura 4.6: Exemplos de resultados obtidos ao longo do pós-processamento.



Fonte: Os Autores.

## 4.2 Representações Esparsas para Classificação de Imagens Macroscópicas de Lesões Melanocíticas

Nesta seção, é apresentado um método para a classificação de lesões melanocíticas em imagens macroscópicas. Esse método, o qual pode ser visto como uma versão modificada – mais escalável e robusta para lidar com o problema do desbalanceamento entre classes – do método SRC (WRIGHT et al., 2009), é uma contribuição original do nosso trabalho, a qual foi publicada primeiramente, no contexto de análise de imagens de satélite, em (FLORES; ZORTEA; SCHARCANSKI, 2019).

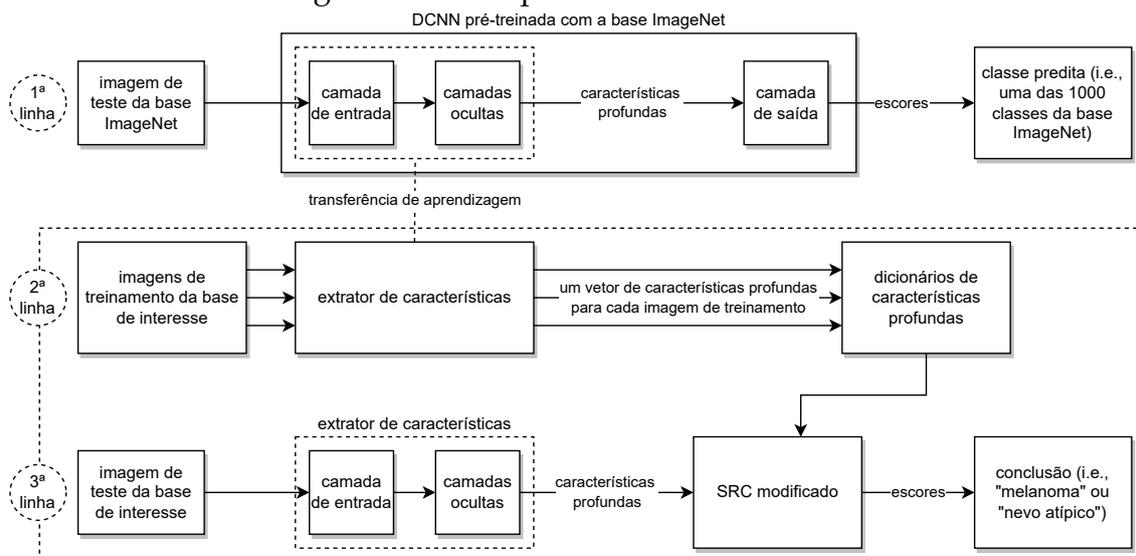
O diagrama da Figura 4.7 ilustra o fluxo de trabalho do método de classificação proposto que consiste em duas etapas principais: 1) aprendizado dos dicionários de características profundas; 2) predição usando o SRC modificado. Na primeira “linha horizontal”, no topo do diagrama, é apresentada a arquitetura

tura de uma DCNN composta por camadas cujos parâmetros foram previamente ajustados com as imagens disponibilizadas pro ImageNet. Ordinariamente, dada uma imagem natural de teste, essa DCNN produziria um escore para cada uma das mil classes consideradas no ImageNet, permitindo assim a classificação da mesma.

Todavia, no método proposto, os parâmetros pré-ajustados das camadas de entrada e ocultas da DCNN são usados como um extrator de características, o qual produz um vetor “características profundas” (do inglês, *deep features*) para cada imagem macroscópica de lesão melanocítica de treinamento. Esses vetores de características profundas são usados como entrada para o aprendizado dos dicionários de características profundas, conforme detalhado na Subseção 4.2.1.

O SRC modificado, por sua vez, combina o vetor de características profundas extraído a partir de uma imagem macroscópica de lesão melanocítica de teste com os dicionários de características profundas para produzir o escore da classe “melanoma” e o escore da classe “nevo atípico”, viabilizando assim a tomada de alguma decisão, conforme discutido na Subseção 4.2.2.

Figura 4.7: Fluxo de trabalho do método proposto para a classificação de lesões melanocíticas em imagens macroscópicas.



Fonte: Os Autores.

### 4.2.1 Aprendizado dos Dicionários de Características Profundas

Seja  $q_i$  o número de imagens de treinamento rotuladas de acordo com a  $i$ -ésima classe,  $i \in \{1, \dots, n_r\}$ , pode-se arbitrariamente atribuir um índice distinto  $j \in \{1, \dots, q_i\}$  para cada imagem de treinamento da  $i$ -ésima classe. Desse modo, cada imagem no conjunto de treinamento pode ser unicamente identificada pelo par  $(i, j)$ .

Considere  $\mathbf{f}_{i,j} \in \mathbb{R}^p$  o vetor contendo as características profundas extraídas a partir da imagem identificada pelo par  $(i, j)$  usando uma DCNN pré-treinada qualquer. Assim, os vetores de características associados com a  $i$ -ésima classe podem ser arranjados na forma:

$$\mathbf{F}_i = [\mathbf{f}_{i,1}, \dots, \mathbf{f}_{i,q_i}], \quad (4.18)$$

onde  $\mathbf{F}_i \in \mathbb{R}^{p \times q_i}$ .

Embora a matriz  $\mathbf{F}_i$  pudesse ser usada diretamente como um dicionário para a  $i$ -ésima classe, no método proposto, um modelo de mistura de Gaussianas (GMM – *Gaussian Mixture Model*) (CARREIRA-PERPINAN, 2000) é adicionalmente empregado para obter uma representação mais compacta para cada classe. Para ajustar esse modelo, os parâmetros  $\hat{\Theta}_i$  são estimados para maximizar a função *log-likelihood* de acordo com a seguinte equação:

$$\hat{\Theta}_i = \arg \max_{\Theta_i} \sum_{j=1}^{q_i} \sum_{k=1}^{C_i} \log(\alpha_{i,k} \mathbb{P}(\mathbf{f}_{i,j}; \boldsymbol{\mu}_{i,k}, \Sigma_{i,k})), \quad (4.19)$$

em que

$$\Theta_i = \{\boldsymbol{\mu}_{i,1}, \dots, \boldsymbol{\mu}_{i,C_i}, \Sigma_{i,1}, \dots, \Sigma_{i,C_i}, \alpha_{i,1}, \dots, \alpha_{i,C_i}\} \quad (4.20)$$

deve satisfazer a restrição

$$\sum_{k=1}^{C_i} \alpha_{i,k} = 1. \quad (4.21)$$

Para isso, foi utilizado o clássico método iterativo EM (*Expectation-Maximization*) (DEMPSTER; LAIRD; RUBIN, 1977; BISHOP, 2006).

No método proposto, probabilidades *a priori* idênticas são atribuídas na inicialização do EM, ou seja,  $\alpha_{i,k} = 1/n_r$ ,  $k = 1, \dots, C_i$ . As médias do GMM,  $\boldsymbol{\mu}_{i,k}$ ,

$k = 1, \dots, C_i$ , por sua vez, são inicializadas com  $C_i$  vetores de características aleatoriamente selecionados a partir de  $F_i$ . Além disso, as matrizes de covariância  $\Sigma_{i,k}$ ,  $k = 1, \dots, C_i$ , são assumidas como sendo diagonais. Por fim, após a execução do EM, o dicionário de características profundas associado à  $i$ -ésima classe é dado por:

$$D_i = [\boldsymbol{\mu}_{i,1}, \dots, \boldsymbol{\mu}_{i,C_i}] \in \mathbb{R}^{p \times C_i}, \quad (4.22)$$

com  $C_i$ , a quantidade de átomos no  $i$ -ésimo dicionário de características profundas, definindo um hiperparâmetro do método proposto.

#### 4.2.2 Predição usando o SRC Modificado

Suponha que  $\mathbf{y}_i$ , o vetor contendo as características profundas extraídas a partir de uma imagem de teste da  $i$ -ésima classe, esteja no subespaço gerado pelos átomos do dicionário associado com a  $i$ -ésima classe (i.e.,  $\mathbf{y}_i \in \text{col}(D_i)$ ). Dessa maneira,  $\mathbf{y}_i$  pode ser escrito como uma combinação linear das colunas de  $D_i$ , ou seja,

$$\mathbf{y}_i = D_i \mathbf{x}_i, \quad (4.23)$$

de forma que  $\mathbf{x}_i \in \mathbb{R}^{C_i}$  é o vetor com os coeficientes dessa combinação linear.

Uma vez que a classe da imagem de teste é desconhecida *a priori*, é conveniente definir a matriz  $D \in \mathbb{R}^{p \times C}$ ,  $C = \sum_{i=1}^{n_r} C_i$ , como a concatenação dos dicionários de características profundas:

$$D = [D_1, \dots, D_{n_r}]. \quad (4.24)$$

Logo, a Eq. (4.23) pode ser reescrita como:

$$\mathbf{y}_i = D \mathbf{x}, \quad (4.25)$$

onde  $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_{n_r}^T]^T \in \mathbb{R}^C$ .

Se  $\text{col}(D_\alpha) \cap \text{col}(D_\beta) = \{\mathbf{0}_p\}$ , para  $\alpha \neq \beta$ , e  $C = p$ , então as entradas de  $\mathbf{x}$  são todas nulas, com exceção daquelas que correspondem à  $i$ -ésima classe (i.e.,  $\mathbf{x}_j = \mathbf{0}_{C_j}$  para todo  $j \neq i$ , com  $\mathbf{0}_{C_j}$  denotando o vetor cujas  $C_j$  componentes são

iguais a zero). Nesse caso,  $\mathbf{x}$  é a única solução do sistema de equações lineares da Eq. (4.25), para  $\mathbf{y}_i$  e  $\mathbf{D}$  dados, e a classificação pode ser efetuada determinando:

$$\arg \min_i \|\mathbf{y}_i - \mathbf{D}_i \mathbf{x}_i\|_2 = \arg \min_i \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}_i \mathbf{x}_i\|_2^2, \quad (4.26)$$

de modo que  $\|\cdot\|_2$  denota a norma  $\ell_2$  do vetor e  $i \in \{1, \dots, n_r\}$ .

Infelizmente, a solução da Eq. (4.25) tipicamente não é única. Consequentemente, uma restrição poderia ser incorporada à Eq. (4.25) para a obtenção de uma solução única em termos de  $\mathbf{x}$  e que fosse a mais esparsa dentre todas as soluções da Eq. (4.25):

$$\mathbf{x}_0 = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{sujeito a} \quad \mathbf{y}_i = \mathbf{D}\mathbf{x}, \quad (4.27)$$

com  $\|\mathbf{x}\|_0$  denotando a norma  $\ell_0$  de  $\mathbf{x}$ , a qual corresponde ao número de componentes não-nulos do vetor  $\mathbf{x}$ .

Entretanto, o problema em questão pode ser resolvido de maneira mais apropriada utilizando a norma  $\ell_1$  para a obtenção de uma estimativa da solução  $\hat{\mathbf{x}}_1$  do seguinte problema de otimização (consulte a Subseção 3.1.2 para detalhes):

$$\hat{\mathbf{x}}_1 = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (4.28)$$

em que  $\|\mathbf{x}\|_1 = \sum_{k=1}^C |x_k|$  denota a norma  $\ell_1$  de  $\mathbf{x}$ . No método proposto, o método FISTA (BECK; TEBoulLE, 2009), explicado em detalhes na Subseção 3.2.2, é empregado para obter essa estimativa.

Aqui, cabe salientar que o escalar  $\lambda$ , o qual fornece o *tradeoff* entre o termo  $\frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}\|_2^2$  e o termo  $\|\mathbf{x}\|_1$ , é um hiperparâmetro do método proposto. O outro hiperparâmetro do método proposto, por conveniência, é definido como segue:

$$\varphi = \frac{C_i}{q_i}, \quad (4.29)$$

onde  $q_i$  é o número de imagens no conjunto de treinamento rotuladas de acordo com a  $i$ -ésima classe e  $C_i$  é o número de átomos no  $i$ -ésimo dicionário de características profundas. Note que os valores de  $n_i$  e  $C_i$  definem o problema de otimização da Eq. (4.19).

Seja  $\mathbf{x}^* = [\mathbf{x}_1^{*T}, \dots, \mathbf{x}_c^{*T}]^T \in \mathbb{R}^C$  a aproximação da solução do problema da Eq. (4.28) obtida pelo método FISTA; o erro de representação da  $i$ -ésima classe

é:

$$e_i = \frac{1}{2} \|\mathbf{y}_i - \mathbf{D}_i \mathbf{x}_i^*\|_2^2, \quad (4.30)$$

onde  $\mathbf{x}_i^* \in \mathbb{R}^{C_i}$  é a “parte” de  $\mathbf{x}^*$  correspondente à  $i$ -ésima classe. Assim sendo, o escore  $s_i$  associado com a  $i$ -ésima classe pode ser computado por:

$$s_i = \begin{cases} \frac{e_i^{-1}}{q_i \sum_{j=1}^{n_r} e_j^{-1}}, & \text{se } e_i \neq 0, \\ \frac{1}{q_i}, & \text{se } e_i = 0, \end{cases} \quad (4.31)$$

de modo que o erro de representação  $e_i$  é assumido como sendo não-nulo para duas ou mais classes.

### 4.3 Considerações Finais

Neste capítulo, foi apresentado em detalhes o sistema proposto nesta tese para processar e analisar imagens macroscópicas de lesões melanocíticas. Por questões de organização, tal sistema foi visto como uma integração entre o método de segmentação de imagens macroscópicas de lesões melanocíticas, detalhado na Seção 4.1, e o método de classificação de imagens macroscópicas de lesões melanocíticas, detalhado na Seção 4.2.

O próximo capítulo apresentará os resultados obtidos nos experimentos realizados para avaliar o sistema proposto e está organizado de maneira análoga, de modo que a Seção 5.1 discute os experimentos conduzidos e resultados alcançados na avaliação do método de segmentação proposto, separadamente dos experimentos e resultados com o método de classificação, os quais estão na Seção 5.2.

## 5 RESULTADOS EXPERIMENTAIS E DISCUSSÃO

Este capítulo reporta os experimentos conduzidos, bem como os resultados obtidos, para a avaliação dos métodos que compõem o sistema proposto nesta tese e detalhado no Capítulo 4. Especificamente, a Seção 5.1 refere-se ao método de segmentação apresentado na Seção 4.1. A Seção 5.2, por sua vez, refere-se ao método de classificação na Seção 4.2 e pode ser vista como a avaliação do sistema como um todo.

Quatro bases de imagens de referência – *benchmarks* –, cujas quantidades de imagens macroscópicas de lesões melanocíticas (melanomas; nevos) e outros detalhes são apresentados na tabela abaixo, foram empregadas ao longo dos experimentos realizados.

Tabela 5.1: Detalhamento sobre as bases de imagens macroscópicas de lesões melanocíticas empregadas nos experimentos. Para as imagens das bases DermIS, DermNet e DermQuest, foram disponibilizados os GTs (*ground-truths*) de segmentação correspondentes; as bases DermIS, DermQuest e MClassND estavam disponíveis publicamente na data da submissão desta tese.

Base	Quantidade de imagens	GTs	Pub.
DermIS (AMELARD et al., 2014)	69 (43 mel.; 26 nev.)	✓	✓
DermNet (ALCÓN et al., 2009)	152 (107 mel.; 45 nev.)	✓	
DermQuest (AMELARD et al., 2014)	137 (76 mel.; 61 nev.)	✓	✓
MClassND (BRINKER et al., 2019)	100 (20 mel.; 80 nev.)		✓

Fonte: Os autores.

Como as imagens da base MClassND não possuem GTs (*ground-truths*) de segmentação, as mesmas foram utilizadas somente nos experimentos da Seção 5.2, ao contrário das imagens das demais bases, as quais foram usadas tanto nos experimentos da Seção 5.2 quanto nos experimentos da Seção 5.1.

Uma vez que o método de segmentação proposto não é do tipo supervisionado, os experimentos da Seção 5.1 não fizeram uso de um conjunto de treinamento. Ao invés disso, apenas a menor dentre as bases de imagens consideradas (i.e., a base DermIS) foi empregada para selecionar os hiperparâmetros de configuração do método, seguindo o mesmo procedimento experimental proposto em (AMELARD et al., 2014) e adotado, por exemplo, no nosso trabalho em (ZORTEA; FLORES; SCHARCANSKI, 2017). Todas as imagens da base DermNet, por

sua vez, foram utilizadas para a avaliação do desempenho do método de segmentação, assim como todas as imagens da base DermQuest.

O método de classificação proposto, por outro lado, é do tipo supervisionado. Por isso, o procedimento experimental empregado na Seção 5.2 consiste em, primeiramente, aplicar um *cross-dataset holdout*, designando as imagens das bases DermNet e MClassND como conjuntos de teste e, em seguida, reamostrar por validação cruzada  $5 \times 2$  (DIETTERICH, 1998) o conjunto composto pelas bases DermIS e DermQuest (ou seja, pelas bases que estão disponíveis publicamente<sup>1</sup> e que contêm os GTs de segmentação) de modo que o mesmo seja particionado dez vezes em subconjuntos estratificados de treinamento e validação cujas cardinalidades diferem-se por no máximo uma unidade.

Como métrica de desempenho de segmentação, considerou-se o escore  $BF_1$  (CSURKA et al., 2013), o qual é computado comparando somente o contorno da segmentação obtida com o contorno do *ground-truth*. Essa métrica é mais conveniente do que aquelas baseadas em regiões (HANCE et al., 1996) e, mais frequentemente utilizadas, no sentido de não penalizar codificações invertidas na identificação dos pixels (i.e., essa medida produz resultados idênticos para duas segmentações que diferem-se apenas por terem utilizado zero ao invés de um – e vice-versa – para rotular um pixel de lesão).

Como métrica de desempenho de classificação, considerou-se a área abaixo da curva ROC, a qual pode ser referida simplesmente como AUC (do inglês, *Area Under the Curve*). Essa medida é mais conveniente do que a acurácia (i.e., a taxa de acerto), pois independe do limiar aplicado sobre os escores para produzir os rótulos, sendo assim mais conveniente para avaliar cenários de treinamento com classes desbalanceadas (BRADLEY, 1997).

As comparações estatísticas entre métodos avaliados em diferentes conjuntos pareados de dados (i.e., conjuntos de pixels, no caso da segmentação, ou conjuntos de imagens, no caso da classificação), basearam-se no teste de Friedman (1937), seguido pelo pós-teste de Nemenyi (1963), para comparar vários métodos entre si, ou pelo pós-teste de Dunn (1961), com a correção de Bonferroni (1936), para comparar vários métodos com um único método de controle (e.g., com o método proposto) (DEMŠAR, 2006; FACELI et al., 2021). Em todos os testes e pós-testes, considerou-se que um nível de confiança de 95% é suficiente para

---

<sup>1</sup>Viabilizando assim a reprodutibilidade dos experimentos.

rejeitar a hipótese nula de que a(s) diferença(s) nos desempenhos observados não é estatisticamente significativa.

## 5.1 Experimentos com o Método Proposto para Segmentação de Imagens Macroscópicas de Lesões Melanocíticas

O número de átomos no dicionário inicial,  $K$ , é o único hiperparâmetro configurável do método de segmentação proposto. Assim, na seção Subseção 5.1.1, o método de segmentação proposto é avaliado variando esse hiperparâmetro.

O método de segmentação proposto pode ser visto como a integração das etapas apresentadas nas Subseções 4.1.1 até 4.1.4, de acordo com a visão geral mostrada na Figura 4.1 e em concordância com o que foi detalhado ao longo do Capítulo 4. Com o objetivo de avaliar a contribuição individual de cada uma dessas etapas, na Subseção 5.1.2, o desempenho do método de segmentação proposto é comparado com o desempenho de alguns métodos alternativos que correspondem ao próprio método proposto, porém, com uma, e somente uma, das suas etapas alterada ou removida (i.e., um estudo de ablação).

Por fim, na Subseção 5.1.3, o método proposto é comparado com métodos estado da arte em segmentação (e.g., métodos de segmentação propostos recentemente para lidar com o problema da segmentação de lesões melanocíticas em imagens macroscópicas).

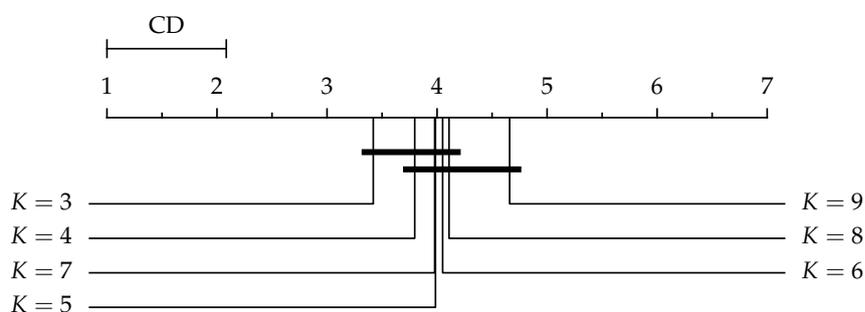
### 5.1.1 Seleção dos Hiperparâmetros de Configuração

A fim de determinar um valor apropriado para o hiperparâmetro de configuração, o método de segmentação proposto foi aplicado em todas as imagens em questão para cada valor  $K \in \{3, \dots, 9\}$ . Em seguida, com base nas segmentações obtidas a partir de uma mesma imagem, as variações do método proposto foram ranqueadas de 1 (melhor desempenho) até 7 (pior desempenho). Como medida de desempenho, foram utilizados os escores  $BF_1$  das segmentações produzidas, de acordo com o *ground-truth* correspondente à imagem em questão.

Baseado nos ranqueamentos, o teste de hipótese de [Friedman \(1937\)](#) foi

empregado e resultou em um valor-p de aproximadamente 4,78%, o que sugere que existem diferenças estatisticamente significativas nos desempenhos das variantes do método proposto. Subsequentemente, foi efetuado o pós-teste de **Nemenyi (1963)**, cujos resultados estão ilustrados no diagrama da Figura 5.1.

Figura 5.1: Ranqueamento médio de diferentes configurações do método de segmentação proposto (i.e., do método de segmentação proposto com  $K = 3, \dots, 9$ ). As segmentações produzidas pelas diferentes configurações, para cada imagem da base DermIS, foram ranqueadas considerando o escore  $BF_1$  como métrica de desempenho.



Fonte: Os Autores.

Conforme ilustrado na Figura 5.1, o melhor (i.e., o menor) ranqueamento médio ( $\approx 3,42$ ) foi alcançado pelo método proposto com  $K = 3$  e o maior ( $\approx 4,66$ ) com  $K = 9$ . Além disso,  $CD \approx 1,08$  denota a diferença crítica (do inglês, *Critical Difference*) do pós-teste realizado, de modo que a diferença de desempenho entre duas variantes do método proposto é considerada significativa se os respectivos ranqueamentos médios diferem-se por pelo menos  $CD$ . Por exemplo, a diferença de desempenho entre o método proposto com  $K = 3$  e  $K = 9$  é considerada estatisticamente significativa, pois  $4,66 - 3,42 > 1,08$ . Por isso, os seus ranqueamentos médios não estão “conectados” por uma das duas barras horizontais mais espessas, as quais possuem comprimento  $CD$ , mostradas na Figura 5.1.

Cabe ainda mencionar que o pós-teste de **Nemenyi (1963)** não detectou diferenças significativas entre o método proposto com  $K = 3$  e as suas variações com  $K = 4, 5, 6, 7$  e  $8$ , cujos ranqueamentos médios (aproximados) são, respectivamente,  $3,80, 3,99, 4,05, 3,98$  e  $4,11$ . Em função disso, uma das barras horizontais mais espessas conecta os ranqueamentos médios correspondentes ao método proposto com  $K = 3, 4, 5, 6, 7$  e  $8$ . De maneira análoga, note que não foram iden-

tificadas diferenças significativas entre as variantes do método proposto com  $K = 4, 5, 6, 7, 8, 9$ . Contudo, visto que  $K = 3$  resultou no melhor ranqueamento médio, daqui em diante, assume-se que o método proposto utiliza  $K = 3$  como hiperparâmetro de configuração.

### 5.1.2 Estudo de Ablação

O desempenho do método proposto foi comparado com o desempenho dos “esquemas alternativos” listados a seguir.

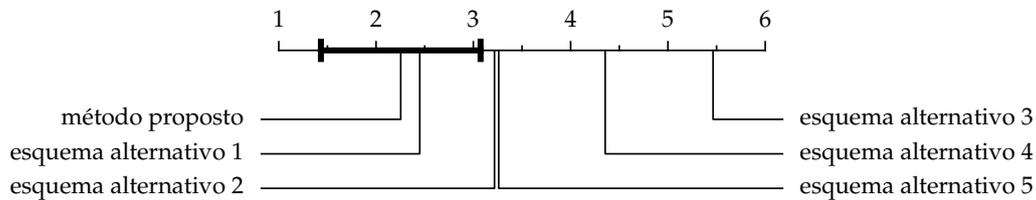
- Método 1: consiste em todas as etapas do método proposto, com exceção do pré-processamento detalhado na Subseção 4.1.1. Esse esquema foi testado com o objetivo de avaliar a influência do pré-processamento no desempenho do método proposto.
- Método 2: consiste no método proposto, porém, usando o canal I do espaço de cores HSI ao invés da imagem de intensidades obtida conforme descrito na Subseção 4.1.2. Esse esquema foi testado para avaliar se existem benefícios em utilizar o processo de obtenção da imagem de intensidades empregado pelo método proposto em vez de usar uma imagem de intensidades mais convencional.
- Método 3: consiste no método proposto, entretanto, utilizando o método K-SVD (ver Subseção 3.2.4), com  $K = 2w^2$  e  $s_0 = 3$ , ao invés do método ALS, conforme detalhado na Subseção 4.1.3, para efetuar a inicialização do dicionário. Esse esquema foi testado a fim de avaliar se existem vantagens em inicializar o dicionário via NMF em vez de construir um dicionário inicial supercompleto.
- Método 4: consiste no método proposto modificado para realizar o agrupamento com base no método de K-médias (ver Subseção 2.2.2), com  $K = 2$  e inicialização por meio do método K-médias++, ao invés de usar o método NGC conforme descrito na Subseção 4.1.4. Esse esquema foi testado com o intuito de avaliar se existem benefícios em utilizar o método NGC em vez de um método mais simples de agrupamento.

- Método 5: consiste em todas as etapas do método proposto, com exceção do pós-processamento descrito na Subseção 4.1.4. Esse esquema foi testado com o propósito de avaliar a importância do pós-processamento no desempenho do método proposto.

Especificamente, o método de segmentação proposto e os cinco esquemas alternativos descritos acima foram empregados para segmentar todas as imagens da base Amelard et al. (2014) oriundas do DermIS. A partir das segmentações produzidas, para cada uma dessas imagens, os métodos foram ranqueados de 1 a 6. Com base nesses ranqueamentos, o valor-p obtido pelo teste de Friedman (1937) foi menor do que  $10^{-30}$ , indicando uma significância estatística nas diferenças de desempenho observadas.

Adicionalmente, foi aplicado o pós-teste de Dunn (1961). Os resultados desse pós-teste estão compilados na Figura 5.2. Conforme pode ser observado, o método proposto apresentou o melhor ranqueamento médio ( $\approx 2,25$ ), seguido pelos esquemas alternativos 1, 2, 5, 4 e 3, cujos ranqueamentos médios (aproximados) são, respectivamente, 2,45, 3,22, 3,26, 4,36 e 5,46. Ademais, o CD  $\approx 0,82$  obtido, neste pós-teste, sugere que os métodos alternativos cujos ranqueamentos médios são maiores que  $2,45 + \text{CD} \approx 3,07$  possuem um desempenho significativamente inferior ao método proposto. Logo, de acordo com o pós-teste de Dunn (1961), o método proposto é significativamente superior a todas as alternativas com exceção à alternativa 5. Todavia, uma vez que o custo computacional do pós-processamento é relativamente baixo e os seus benefícios são evidentes, conforme exemplificado na Figura 4.6, optou-se por mantê-lo incorporado ao método proposto.

Figura 5.2: Ranqueamento médio do método de segmentação proposto em comparação com versões alternativas do mesmo. As segmentações produzidas pelas diferentes variantes, para cada imagem da base DermIS, foram ranqueadas considerando o escore  $BF_1$  como métrica de desempenho.



Fonte: Os autores.

### 5.1.3 Comparação com o Estado-da-Arte

Nesta subseção, o desempenho do método proposto foi comparado com o desempenho de métodos representativos do estado-da-arte, os quais foram contextualizados na Seção 2.2 e estão listados abaixo:<sup>2</sup>

- DSNet (*Dermoscopic Skin Network*) (HASAN et al., 2020);
- FCNet (*Fully Convolutional Network*) (LONG; SHELHAMER; DARRELL, 2015);
- FrCNet (*Full-resolution Convolutional Network*) (AL-MASNI et al., 2018);
- LGPNet (*Local/Global Patch-based Network*) (JAFARI et al., 2016);
- PSPNet (*Pyramid Scene Parsing Network*) (ZHAO et al., 2017);
- SegNet (*Segmentation Network*) (BADRINARAYANAN; KENDALL; CIPOLLA, 2017);
- U-Net (*U-shaped Network*) (RONNEBERGER; FISCHER; BROX, 2015).

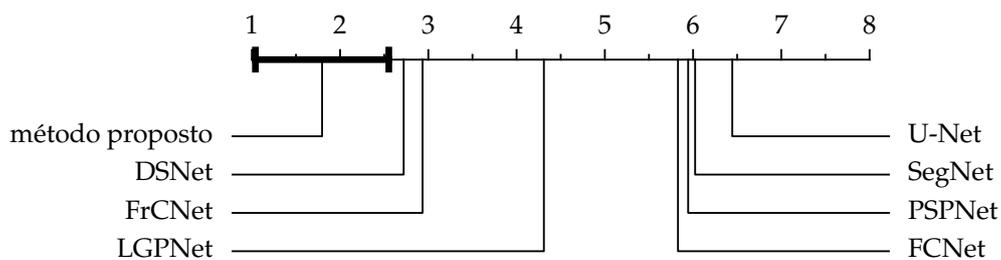
Na primeira avaliação descrita nesta subseção, os métodos em questão foram aplicados para segmentar todas as imagens da base DermNet. Para cada uma dessas imagens, tais métodos foram ranqueados de 1 a 8, com base no desempenho da segmentação produzida. O valor-p obtido pelo teste de Friedman (1937), a partir desses ranqueamentos, foi menor do que  $10^{-123}$ , o que indica que

<sup>2</sup>Os códigos e demais detalhes de implementação necessários para reproduzir todos os experimentos desta tese estarão disponíveis em: <<https://github.com/eliezersflores/thesis>>.

as diferenças de desempenho entre os métodos, observadas nas segmentações das imagens da base DermNet, são estatisticamente significantes.

O pós-teste de [Dunn \(1961\)](#) foi empregado para complementar o teste de [Friedman \(1937\)](#) por meio de uma comparação dos desempenhos médios dos métodos representativos do estado da arte em relação ao desempenho médio do método proposto. Os resultados desse pós-teste são apresentados na [Figura 5.3](#) e sugerem que o desempenho do método proposto, na base DermNet, é significativamente superior aos desempenhos dos métodos comparados.

Figura 5.3: Ranqueamento médio do método de segmentação proposto em comparação com métodos representativos do estado-da-arte. As segmentações produzidas pelos diferentes métodos, para cada imagem da base DermNet, foram ranqueadas considerando o escore  $BF_1$  como métrica de desempenho.

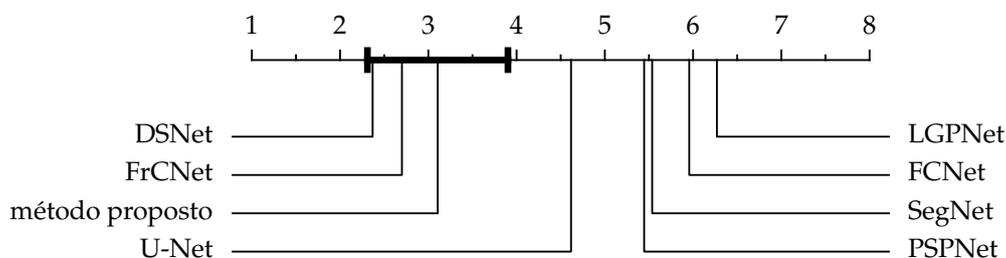


Fonte: Os autores.

Em uma segunda avaliação, o teste de [Friedman \(1937\)](#) foi utilizado do mesmo modo que na primeira avaliação, porém, considerando as imagens da base DermQuest, ao invés daquelas da base DermNet. O valor-p resultante desse teste foi menor do que  $10^{-80}$ , o que aponta que as diferenças de desempenho entre os métodos, verificadas nas segmentações das imagens em questão, também possuem significância estatística.

O pós-teste de [Dunn \(1961\)](#) foi usado como complemento ao teste de [Friedman \(1937\)](#) do parágrafo anterior, de maneira análoga ao pós-teste da primeira avaliação. Os resultados desse pós-teste estão compilados na [Figura 5.4](#) e sugerem que, nessas imagens, não existem diferenças de desempenho relevantes entre o método proposto e os métodos DSNet e FrCNet. Por outro lado, tais resultados indicam que o desempenho do método proposto é estatisticamente superior àqueles dos demais métodos analisados.

Figura 5.4: Ranqueamento médio do método de segmentação proposto em comparação com métodos representativos do estado-da-arte. As segmentações produzidas pelos diferentes métodos, para cada imagem da base DermQuest, foram ranqueadas considerando o escore  $BF_1$  como métrica de desempenho.



Fonte: Os autores.

## 5.2 Experimentos com o Método Proposto para Classificação de Imagens Macroscópicas de Lesões Melanocíticas

Nesta seção, são avaliadas as etapas de extração de características e classificação empregadas pelo sistema proposto, as quais foram apresentadas em detalhes na Seção 4.2. Para isso, primeiramente, na Subseção 5.2.1, são avaliados os efeitos de variar hiperparâmetros configuráveis utilizados nessas etapas e que influenciam no desempenho do sistema proposto. Depois, na Subseção 5.2.2, o desempenho do método proposto é comparado com alternativas, também treinadas com os vetores de características profundas extraídos pela DCNN ResNet-50, porém, usando um classificador mais tradicional (e.g., KNN ou SVM) ao invés do SRC modificado. Por fim, na Subseção 5.2.3, o método de classificação proposto é comparado com algumas alternativas estado-da-arte.

### 5.2.1 Seleção dos Hiperparâmetros de Configuração

Diferentes variantes do método de classificação proposto podem ser produzidas alterando a DCNN, pré-treinada com a base ImageNet, usada como extrator de características.<sup>3</sup> Com base nisso, foram investigadas as variações geradas com as seguintes DCNNs:

<sup>3</sup>Pode ser útil ao leitor rever a Figura 4.7.

- DenseNet-121, DenseNet-169 e DenseNet-201 (HUANG et al., 2017);
- EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, EfficientNet-B3, EfficientNet-B4, EfficientNet-B5, EfficientNet-B6 e EfficientNet-B7 (TAN; LE, 2019);
- InceptionNet-v3 (SZEGEDY et al., 2016);
- InceptionResNet-v2 (SZEGEDY et al., 2017);
- MobileNet (HOWARD et al., 2017);
- MobileNet-v2 (SANDLER et al., 2018);
- MobileNet-v3-small e MobileNet-v3-large (HOWARD et al., 2019);
- NASNet-large e NASNet-mobile (ZOPH et al., 2018);
- ResNet50, ResNet101 e ResNet152 (HE et al., 2016a);
- ResNet50-v2, ResNet101-v2 e ResNet152-v2 (HE et al., 2016b);
- VGGNet-16 e VGGNet-19 (SIMONYAN; ZISSERMAN, 2015);
- XceptionNet (CHOLLET, 2017).

A fim de avaliar o desempenho dessas variantes, foi aplicada uma validação cruzada  $5 \times 2$  sobre a base DermQuest, resultando em dez medidas de desempenho preditivo por variação avaliada. Para cada uma dessas medidas, as variantes foram ranqueadas de 1 (melhor desempenho) até 28 (pior desempenho), usando a área abaixo da curva (AUC – *Area Under the Curve*) ROC (*Receiver Operating Characteristic*) como métrica de desempenho (quanto maior a AUC, melhor o desempenho preditivo). Na Tabela 5.2, é apresentado o ranqueamento médio ( $\overline{\text{ranq}}$ ) com cada uma das DCNNs listadas anteriormente, assim como os volumes de entrada e saída dos respectivos extratores de características.

Note que, em cada linha da Tabela 5.2, também são apresentados os hiperparâmetros  $\lambda$  e  $\varphi$  empregados em combinação com a DCNN em questão. Cabe mencionar que todas as DCNNs, efetivamente, foram avaliadas em combinação com todos os pares de hiperparâmetros  $(\lambda, \varphi) \in \{1.10^{-2}, 2.10^{-2}, 5.10^{-2}, 1.10^{-1}, 2.10^{-1}, 5.10^{-1}, 1, 2, 5\} \times \{2.10^{-1}, 4.10^{-1}, 6.10^{-1}, 8.10^{-1}, 1\}$  (os resultados com todas as  $1260 = 28.9.5$  configurações estão em anexo na Tabela A.1) de modo que

Tabela 5.2: Ranqueamento médio de diferentes configurações do método de classificação proposto (i.e., do método de classificação proposto variando a DCNN empregada para a extração de características, bem como os hiperparâmetros  $\lambda$  e  $\varphi$ ). As predições realizadas pelas diferentes configurações, para cada conjunto de validação, foram ranqueadas considerando a AUC como métrica de desempenho.

DCNN	vol. (entrada $\rightarrow$ saída)	$\lambda$	$\varphi$	$\overline{\text{ranq}}$
DenseNet-121	$224 \times 224 \times 3 \rightarrow 1024$	5,00	1,00	14,9
DenseNet-169	$224 \times 224 \times 3 \rightarrow 1664$	5,00	1,00	14,9
DenseNet-201	$224 \times 224 \times 3 \rightarrow 1920$	5,00	0,60	14,9
EfficientNet-B0	$224 \times 224 \times 3 \rightarrow 1280$	0,50	0,60	10,3
EfficientNet-B1	$240 \times 240 \times 3 \rightarrow 1280$	5,00	0,80	6,5
EfficientNet-B2	$260 \times 260 \times 3 \rightarrow 1408$	2,00	1,00	13,9
EfficientNet-B3	$300 \times 300 \times 3 \rightarrow 1536$	5,00	1,00	16,9
EfficientNet-B4	$380 \times 380 \times 3 \rightarrow 1792$	2,00	1,00	15,3
EfficientNet-B5	$456 \times 456 \times 3 \rightarrow 2048$	0,50	0,40	15,3
EfficientNet-B6	$528 \times 528 \times 3 \rightarrow 2304$	0,50	1,00	12,8
EfficientNet-B7	$600 \times 600 \times 3 \rightarrow 2560$	0,50	0,60	8,8
InceptionResNet-v2	$299 \times 299 \times 3 \rightarrow 1536$	5,00	1,00	19,2
InceptionNet-v3	$299 \times 299 \times 3 \rightarrow 2048$	5,00	0,60	18,6
MobileNet	$224 \times 224 \times 3 \rightarrow 1024$	5,00	0,80	17,9
MobileNet-v2	$224 \times 224 \times 3 \rightarrow 1280$	5,00	1,00	9,5
MobileNet-v3-large	$224 \times 224 \times 3 \rightarrow 1280$	5,00	1,00	9,6
MobileNet-v3-small	$224 \times 224 \times 3 \rightarrow 1024$	2,00	0,60	11,2
NASNet-large	$331 \times 331 \times 3 \rightarrow 4032$	0,20	0,20	19,2
NASNet-mobile	$224 \times 224 \times 3 \rightarrow 1056$	0,50	0,40	17,4
ResNet-101	$224 \times 224 \times 3 \rightarrow 2048$	0,10	0,60	14,8
ResNet-101-v2	$224 \times 224 \times 3 \rightarrow 2048$	0,02	0,20	26,1
ResNet-152	$224 \times 224 \times 3 \rightarrow 2048$	0,20	0,20	9,6
ResNet-152-v2	$224 \times 224 \times 3 \rightarrow 2048$	1,00	0,40	22,0
ResNet-50	$224 \times 224 \times 3 \rightarrow 2048$	5,00	0,80	5,2
ResNet-50-v2	$224 \times 224 \times 3 \rightarrow 2048$	5,00	0,40	17,9
VGGNet-16	$224 \times 224 \times 3 \rightarrow 512$	5,00	1,00	9,2
VGGNet-19	$224 \times 224 \times 3 \rightarrow 512$	5,00	1,00	14,8
XceptionNet	$299 \times 299 \times 3 \rightarrow 2048$	0,50	0,40	19,1

Fonte: Os autores.

a Tabela 5.2 mostra os resultados levando em consideração somente as melhores configurações com cada DCNN.

A aplicação do teste de hipótese de [Friedman \(1937\)](#), a partir dos ranqueamentos sintetizados na Tabela 5.2, resultou num valor-p menor que  $10^{-8}$ , detectando assim uma significância estatística nas diferenças de desempenho observadas (considerando um nível de significância de 5%). A aplicação do pós-teste de [Nemenyi \(1963\)](#), por sua vez, identificou uma diferença crítica de aproximadamente 13,67, o que implica a existência de quatro grupos cujos desempenhos não diferem-se significativamente:

- Grupo 1, composto pelas variantes produzidas com as DCNNs ResNet-50, EfficientNet-B1, EfficientNet-B7, VGGNet-16, MobileNet-v2, MobileNet-v3-large, ResNet-152, EfficientNet-B0, MobileNet-v3-small, EfficientNet-B6, EfficientNet-B2, VGGNet-19, ResNet-101, DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B4, EfficientNet-B5, EfficientNet-B3, NASNet-mobile, ResNet-50-v2, MobileNet e InceptionNet-v3, as quais alcançaram ranqueamentos médios entre 5,2 e 18,6.
- Grupo 2, compreendendo as variações com as DCNNs EfficientNet-B1, EfficientNet-B7, VGGNet-16, MobileNet-v2, MobileNet-v3-large, ResNet-152, EfficientNet-B0, MobileNet-v3-small, EfficientNet-B6, EfficientNet-B2, VGGNet-19, ResNet-101, DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B4, EfficientNet-B5, EfficientNet-B3, NASNet-mobile, ResNet-50-v2, MobileNet, InceptionNet-v3, XceptionNet, NASNet-large e InceptionResNet-v2, cujos ranqueamentos médios estão entre 6,5 e 19,2.
- Grupo 3, formado pelas variantes geradas com as DCNNs EfficientNet-B7, VGGNet-16, MobileNet-v2, MobileNet-v3-large, ResNet-152, EfficientNet-B0, MobileNet-v3-small, EfficientNet-B6, EfficientNet-B2, VGGNet-19, ResNet-101, DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B4, EfficientNet-B5, EfficientNet-B3, NASNet-mobile, ResNet-50-v2, MobileNet, InceptionNet-v3, XceptionNet, NASNet-large, InceptionResNet-v2 e ResNet-152-v2, as quais obtiveram ranqueamentos médios entre 8,8 e 22,0.
- Grupo 4, consistindo nas variações com as DCNNs EfficientNet-B6, EfficientNet-B2, VGGNet-19, ResNet-101, DenseNet-169, DenseNet-121, DenseNet-201, EfficientNet-B4, EfficientNet-B5, EfficientNet-B3, NASNet-mobile,

ResNet-50-v2, MobileNet, InceptionNet-v3, XceptionNet, NASNet-large, InceptionResNet-v2, ResNet-152-v2 e ResNet-101-v2, cujos ranqueamentos médios estão entre 12,8 e 26,1.

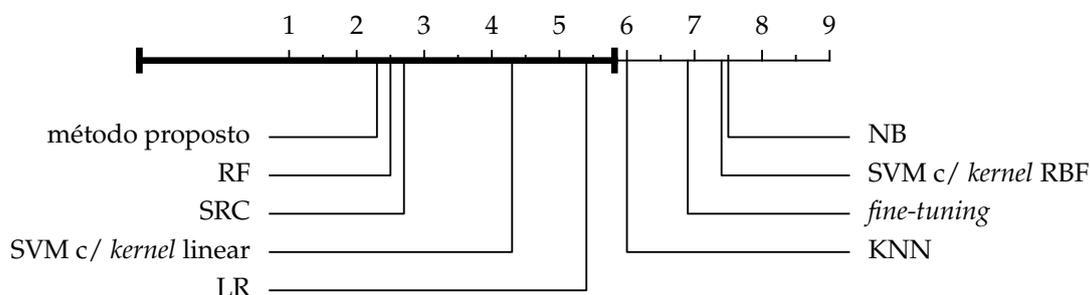
Uma vez que o melhor ranqueamento médio foi alcançado com a DCNN ResNet-50, daqui em diante, assume-se que o método proposto utiliza essa DCNN como extrator de características. Cabe enfatizar que não foram observadas diferenças estatisticamente significativas entre a DCNN ResNet e as demais alternativas do grupo 1. Esse pode ser considerado um aspecto positivo pois flexibiliza o sistema proposto a empregar diferentes DCNNs sem prejuízos no desempenho.

### 5.2.2 Estudo de Ablação

Alternativas ao método proposto podem ser produzidas substituindo o esquema de classificação baseado em dicionários por algum outro classificador treinado a partir dos mesmos vetores de características profundas extraídos pela DCNN ResNet-50 (veja a Figura 4.7). Com base nisso, foram investigadas as alternativas geradas com os seguintes classificadores: *K Nearest Neighbors* (KNN); *Logistic Regression* (LR); *Naive Bayes* (NB); *Random Forest* (RF); *Sparse Representation Classifier* (SRC); *Support Vector Machine* (SVM) com *kernel* linear e SVM com *kernel* do tipo *Radial Basis Function* (RBF). Ademais, foi analisado o *fine-tuning* (FT) dos pesos da DCNN ResNet-50, antes de empregá-la para a extração dos vetores de características profundas usados na construção dos dicionários, ao invés de deixar fixos os pesos resultantes do pré-treinamento com a base ImageNet.

Cada uma das alternativas avaliadas, de acordo com a validação cruzada  $5 \times 2$  sobre a base DermQuest descrita na Subseção 5.2.1, resultou em dez medidas (AUC) de desempenho preditivo, as quais foram ranqueadas de 1 até 9. O valor-p obtido pelo teste de Friedman (1937), a partir desses ranqueamentos, foi menor do que  $10^{-7}$ , o que indica que as diferenças de desempenho são estatisticamente significativas. Os resultados do pós-teste de Dunn (1961) estão condensados na Figura 5.5.

Figura 5.5: Ranqueamento médio do método de classificação proposto em comparação com versões alternativas do mesmo. As predições realizadas pelas diferentes variantes, para cada conjunto de validação, foram ranqueadas considerando a AUC como métrica de desempenho.



Fonte: Os autores.

### 5.2.3 Comparação com o Estado-da-Arte

Nesta subseção, o desempenho do método proposto foi comparado com o desempenho de métodos representativos do estado-da-arte, os quais foram contextualizados na Seção 2.3 e estão listados abaixo:

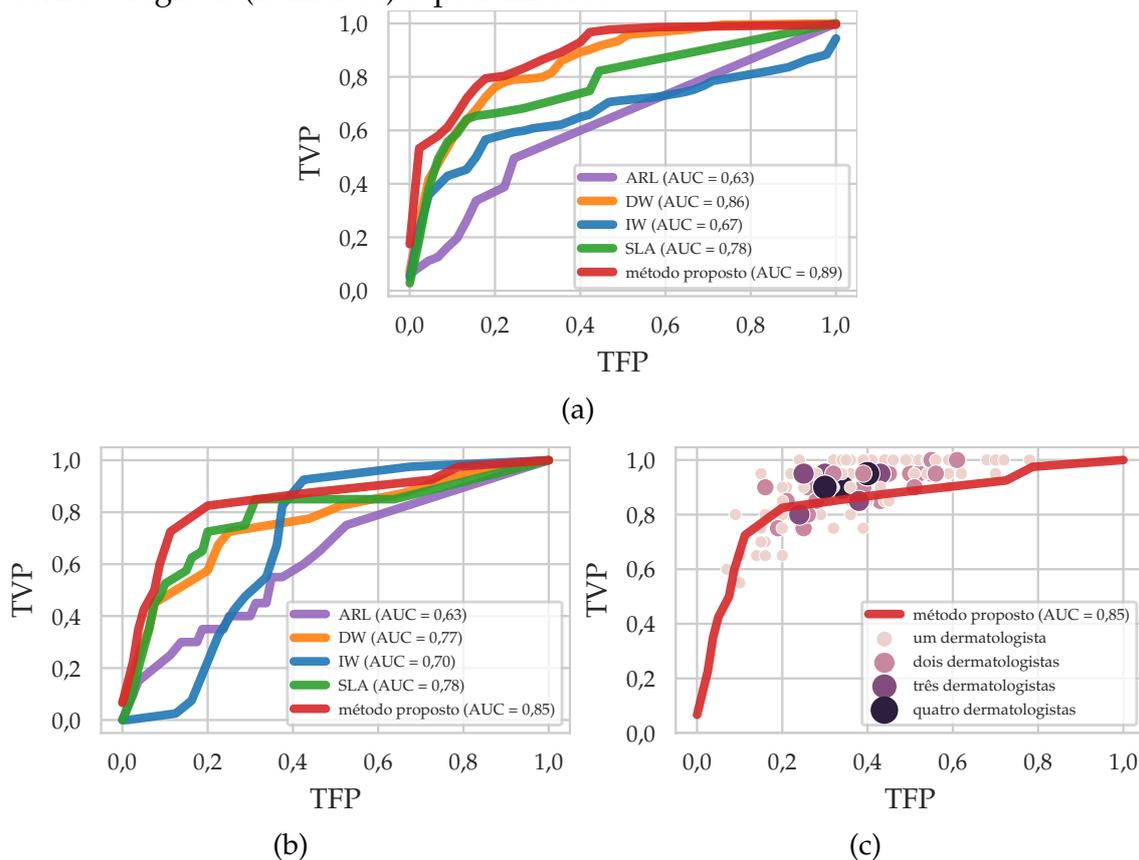
- *Attention Residual Learning* (ARL) (ZHANG et al., 2019);
- *Dynamic Weights* (DW) (PACHECO; TRAPPENBERG; KROHLING, 2020);
- *ISIC Winner* (IW) (HA; LIU; LIU, 2020);
- *Skin Lesion Analyzer* (SLA) (WOZA, 2018).

A Subfigura 5.6a exibe um comparativo das curvas ROC, obtidas pelo método proposto e pelos métodos representativos do estado da arte em questão, nos testes realizados com as imagens da base DermQuest. A Subfigura 5.6b também compara as curvas ROC produzidas a partir desses mesmos métodos, porém, nos testes efetuados com as imagens da base MClassND. Conforme pode ser observado nas Subfiguras 5.6a e 5.6b, o desempenho do método proposto, em termos da métrica AUC, supera o desempenho dos demais métodos investigados, tanto nos testes com as imagens da base DermQuest quanto nos testes com as imagens da base MClassND.

A Subfigura 5.6c, por sua vez, compara o método proposto com os desempenhos de 157 dermatologistas especializados Brinker et al. (2019). Cada círculo no plano ROC indica o desempenho de um ou mais desses dermatologistas (se

mais que um dermatologista obteve o mesmo desempenho, um círculo maior e mais escuro é apresentado). Como pode ser notado, o método proposto, potencialmente, pode superar o desempenho de alguns desses dermatologistas.

Figura 5.6: Ranqueamento médio do método de classificação proposto, nos testes com as bases de [Alcón et al. \(2009\)](#) e [Brinker et al. \(2019\)](#), em comparação com alternativas representativas do estado-da-arte e diagnósticos realizados por dermatologistas (humanos) especializados.



Fonte: Os autores.

As Figuras 5.7 até 5.11, a seguir, apresentam as matrizes de confusão correspondentes aos testes realizados com o método proposto e com cada um dos métodos estado-da-arte listados acima. A partir dos dados que estão nessas figuras, é possível verificar que o método proposto é mais preciso do que todas as alternativas em questão, tanto na base DermNet quanto na base MClassND.

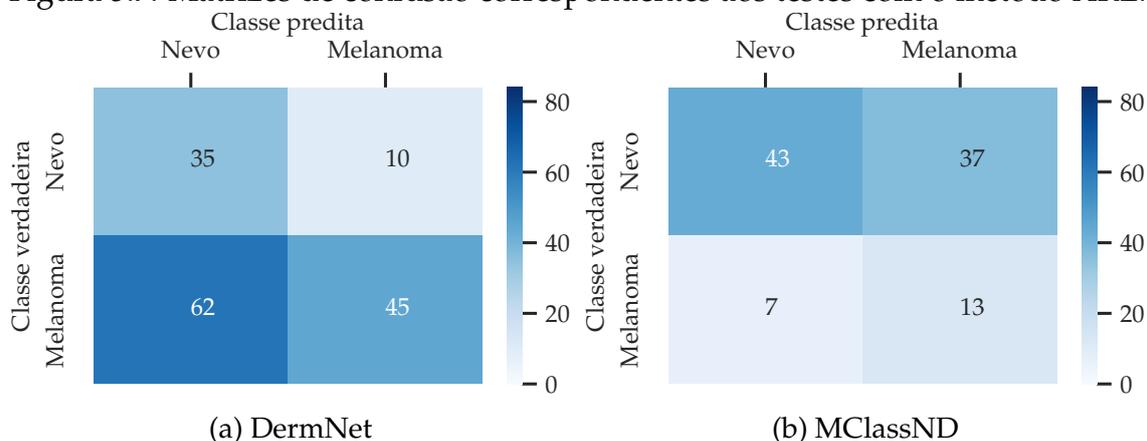
Com relação à taxa de detecção de melanoma (i.e., a taxa de verdadeiros positivos – TVP –, já considerada anteriormente na análise ROC, a qual pode ser alternativamente referida como sensibilidade ou revocação), o método proposto também supera os demais métodos na base DermNet, porém, é superado pelo método IW na base MClassND. Contudo, levando em conta a média (aritmética,

geométrica ou harmônica) entre a precisão e a revocação, o método proposto supera todas as demais alternativas na base DermNet e também na base MClassND.

Cabe mencionar que, embora a AUC tenha sido considerada como métrica de desempenho de classificação, os resultados dos dois últimos parágrafos não só corroboram com a tese de que o método proposto apresenta uma taxa satisfatória de detecção de melanoma, mas também tende a ser preciso na realização dos seus diagnósticos.

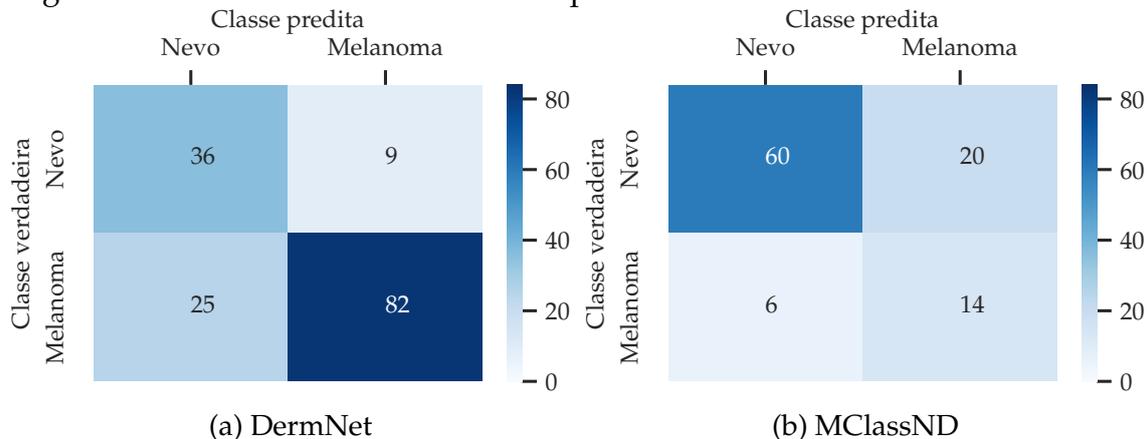
Por fim, deve-se salientar que o sistema proposto é um sistema de processamento e análise de imagens baseado em aprendizado automático e, por tanto, teve que lidar com muitos dos problemas recorrentemente encontrados no desenvolvimento deste tipo de sistema (e.g., quantidade limitada de dados de treinamento, número de amostras desbalanceadas, etc). Consequentemente, algumas questões importantes que não foram tratadas nesta tese, por exemplo, a integração dos métodos de segmentação e classificação em uma única rede que possa ser treinada de maneira unificada, bem como a adaptação do sistema para fornecer uma predição mais interpretável, são aspectos relevantes que podem vir a ser investigadas em extensões deste trabalho.

Figura 5.7: Matrizes de confusão correspondentes aos testes com o método ARL.



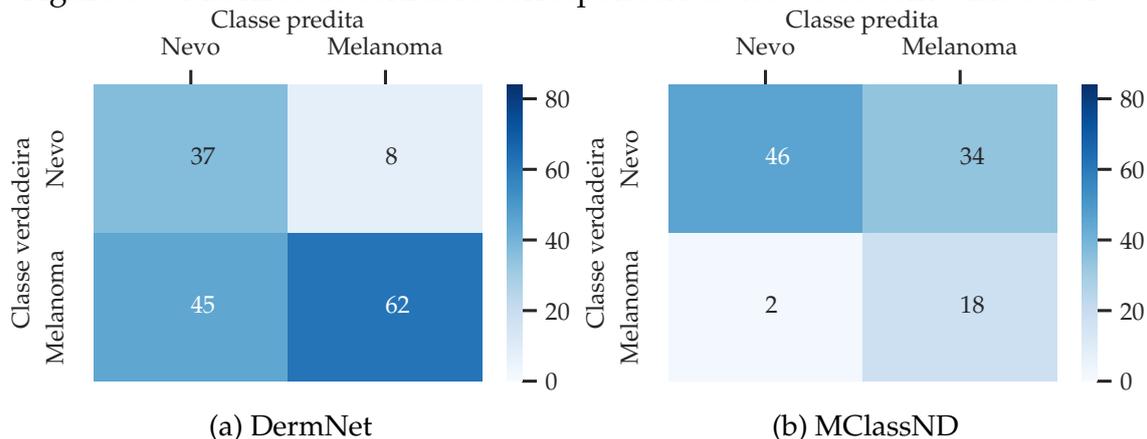
Fonte: Os autores.

Figura 5.8: Matrizes de confusão correspondentes aos testes com o método DW.



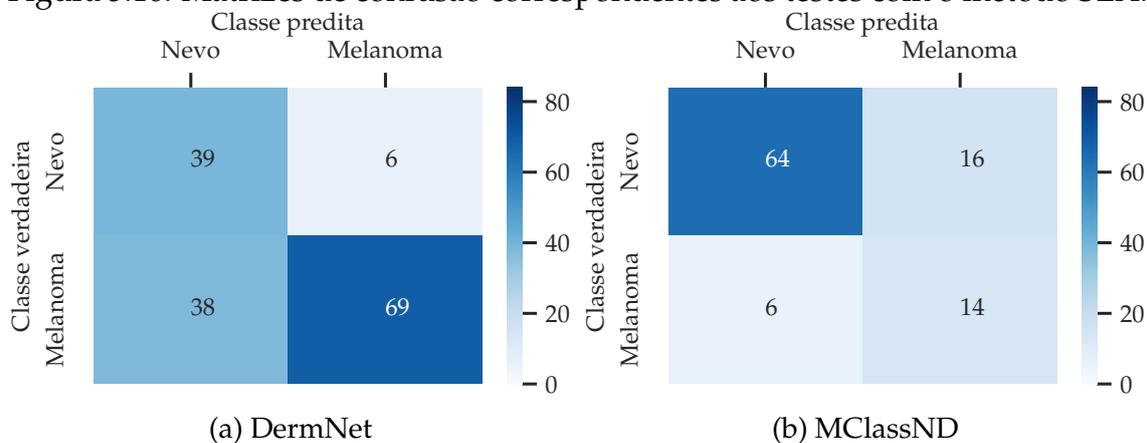
Fonte: Os autores.

Figura 5.9: Matrizes de confusão correspondentes aos testes com o método IW.



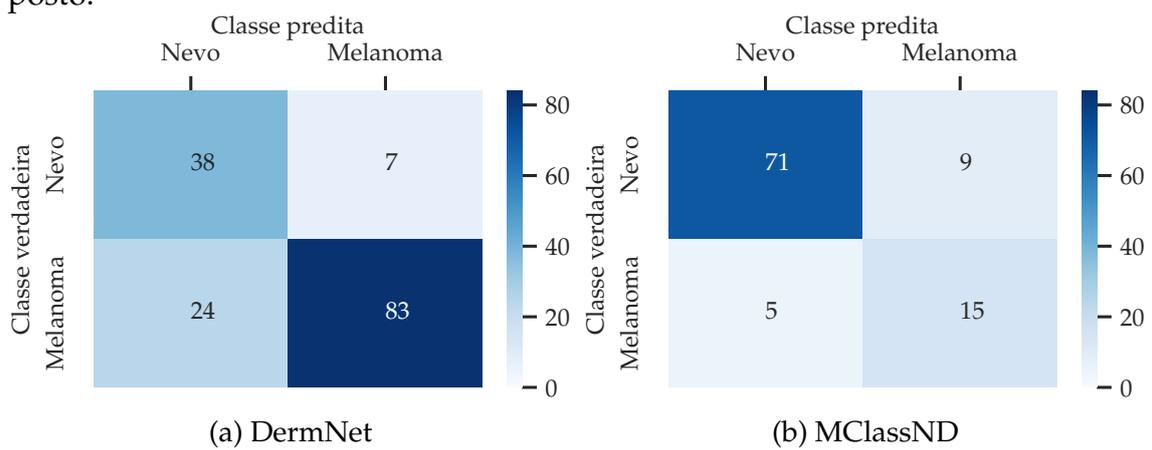
Fonte: Os autores.

Figura 5.10: Matrizes de confusão correspondentes aos testes com o método SLA.



Fonte: Os autores.

Figura 5.11: Matrizes de confusão correspondentes aos testes com o método proposto.



Fonte: Os autores.

## 6 CONCLUSÃO

Nesta tese de doutorado, foi proposto um sistema completo de pré-triagem de lesões melanocíticas em imagens macroscópicas, o qual é composto por um método de segmentação (não supervisionado) e um método de classificação (supervisionado). O método de segmentação baseou-se na proposta de uma versão adaptativa e não supervisionada do método de aprendizado de dicionário ITDL (QIU; PATEL; CHELLAPPA, 2014). O método de classificação, por sua vez, pode ser visto como uma extensão do método SRC (WRIGHT et al., 2009), a qual torna-o mais escalável e flexível para lidar com os eventuais desbalanceamentos nas quantidades de exemplos de cada classe.

Os resultados obtidos sugerem que os métodos propostos são promissores em tarefas desafiadoras de processamento e análise de imagens. Especificamente, foi observado que tanto o método de segmentação quanto de classificação podem superar significativamente alternativas representativas do estado da arte. Ademais, o desempenho do sistema como um todo, potencialmente, é superior até mesmo ao desempenho reportado por alguns dermatologistas especializados.

Além das publicações em eventos, ao longo do desenvolvimento desta tese, foram publicados três artigos em periódicos internacionais, bem como um capítulo de livro, com processos de revisão rigorosos, os quais estão listados a seguir.

- FLORES, E.; SCHARCANSKI, J. Segmentation of melanocytic skin lesions using feature learning and dictionaries. **Expert Systems with Applications**, Amsterdam, v. 56, p. 300–309, 2016.
- ZORTEA, M.; FLORES, E.; SCHARCANSKI, J. A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images. **Pattern Recognition**, Amsterdam, v. 64, p. 92–104, 2017.
- FLORES, E.; ZORTEA, M.; SCHARCANSKI, J. Dictionaries of deep features for landuse scene classification of very high spatial resolution images. **Pattern Recognition**, Amsterdam, v. 89, p. 32–44, 2019.
- BERNART, M.; FLORES, E.; SCHARCANSKI, J. Macroscopic Pigmented Skin Lesion Prescreening. In: NARAYIAN, R (Ed.). **Encyclopedia of Biomedical Engineering**. Amsterdam: Elsevier, 2019. v. 2, p. 561–573.

## REFERÊNCIAS

- ACS – American Cancer Society. **Survival Rates for Melanoma Skin Cancer**. 2022. Disponível em: <https://www.cancer.org/cancer/melanoma-skin-cancer/detection-diagnosis-staging/survival-rates-for-melanoma-skin-cancer-by-stage.html>. Último acesso em: 09 set. 2022.
- AHARON, M.; ELAD, M.; BRUCKSTEIN, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. **IEEE Transactions on Signal Processing**, Piscataway, v. 54, n. 11, p. 4311–4322, 2006.
- AL-MASNI, M. A.; AL-ANTARI, M. A.; CHOI, M.-T.; HAN, S.-M.; KIM, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. **Computer Methods and Programs in Biomedicine**, Amsterdam, v. 162, p. 221–231, 2018.
- ALCÓN, J. F.; CIUHU, C.; KATE, W. ten; HEINRICH, A.; UZUNBAJAKAVA, N.; KREKELS, G.; SIEM, D.; HAAN, G. de. Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. **IEEE Journal of Selected Topics in Signal Processing**, Piscataway, v. 3, n. 1, p. 14–25, 2009.
- AMELARD, R.; GLAISTER, J.; WONG, A.; CLAUSI, D. A. High-level intuitive features (hlifs) for intuitive skin lesion description. **IEEE Transactions on Biomedical Engineering**, Piscataway, v. 62, n. 3, p. 820–831, 2014.
- ARTHUR, D.; VASSILVITSKII, S. K-means++: the advantages of careful seeding. In: ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS (SODA), 18., 2007, New Orleans. **Proceedings [...]**. Philadelphia: SIAM, 2007. p. 1027–1035.
- BACH, F.; JENATTON, R.; MAIRAL, J.; OBOZINSKI, G. et al. Optimization with sparsity-inducing penalties. **Foundations and Trends® in Machine Learning**, Boston, v. 4, n. 1, p. 1–106, 2012.
- BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 39, n. 12, p. 2481–2495, 2017.
- BAO, C.; JI, H.; QUAN, Y.; SHEN, Z. Dictionary learning for sparse coding: algorithms and convergence analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 38, n. 7, p. 1356–1369, 2015.
- BARATA, C.; CELEBI, M. E.; MARQUES, J. S. Improving dermoscopy image classification using color constancy. **IEEE Journal of Biomedical and Health Informatics**, Piscataway, v. 19, n. 3, p. 1146–1152, 2014.
- BARTLETT, P. L.; HARVEY, N.; LIAW, C.; MEHRABIAN, A. Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks.

**Journal of Machine Learning Research**, v. 20, n. 63, p. 1–17, 2019. Disponível em: <http://jmlr.org/papers/v20/17-612.html>. Último acesso em: 09 set. 2022.

BECK, A.; TBOULLE, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. **SIAM Journal on Imaging Sciences**, Philadelphia, v. 2, n. 1, p. 183–202, 2009.

BELLMAN, R. E. **Adaptive Control Processes: a guided tour**. 1. ed. Princeton: Princeton University Press, 1961.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 35, n. 8, p. 1798–1828, 2013.

BERNART, E.; SCHARCANSKI, J.; BAMPI, S. Segmentation and classification of melanocytic skin lesions using local and contextual features. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP), 23., 2016, Phoenix. **Proceedings [...]**. Piscataway: IEEE, 2016. p. 2633–2637.

BERRY, M. W.; BROWNE, M.; LANGVILLE, A. N.; PAUCA, V. P.; PLEMMONS, R. J. Algorithms and applications for approximate nonnegative matrix factorization. **Computational Statistics and Data Analysis**, v. 52, n. 1, p. 155–173, 2007.

BEYER, K.; GOLDSTEIN, J.; RAMAKRISHNAN, R.; SHAFT, U. When is “nearest neighbor” meaningful? In: INTERNATIONAL CONFERENCE ON DATABASE THEORY (ICDT), 7., 1999, Jerusalem. **Proceedings [...]**. Berlin, Heidelberg: Springer, 1999. (Lecture Notes in Computer Science 1540), p. 217–235.

BILMES, J. A. **A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models (Technical Report)**. Berkeley, 1998.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. 1. ed. Singapore: Springer, 2006.

BISSOTO, A.; FORNACIALI, M.; VALLE, E.; AVILA, S. (de) constructing bias on skin lesion datasets. In: IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW), 32., 2019, Long Beach. **Proceedings [...]**. Piscataway: IEEE, 2019. p. 2766–2774.

BONFERRONI, C. E. Teoria statistica delle classi e calcolo delle probabilita. **Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze**, Venezia, v. 8, p. 3–62, 1936.

BOYD, S.; VANDENBERGHE, L. **Convex Optimization**. 1. ed. Cambridge: Cambridge University Press, 2004.

BRADLEY, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. **Pattern Recognition**, Amsterdam, v. 30, n. 7, p. 1145–1159, 1997.

BRASIL. Ministério da Saúde. Comissão Nacional de Incorporação de Tecnologias no SUS. **Terapia-alvo (vemurafenibe, dabrafenibe, cobimetinibe, trametinibe) e imunoterapia (ipilimumabe, nivolumabe, pembrolizumabe) para o tratamento de primeira linha do melanoma avançado não-cirúrgico e metastático**. 1. ed. Brasília: CONITEC, 2019.

BRINKER, T. J.; HEKLER, A.; HAUSCHILD, A.; BERKING, C.; SCHILLING, B.; ENK, A. H.; HAFERKAMP, S.; KAROGLAN, A.; KALLE, C. von; WEICHENTHAL, M. et al. Comparing artificial intelligence algorithms to 157 german dermatologists: the melanoma classification benchmark. **European Journal of Cancer**, Amsterdam, v. 111, p. 30–37, 2019.

BURKOV, A. **The hundred-page machine learning book the hundred-page machine learning book**. 1. ed. Quebec: Andriy Burkov, 2019.

BURRUS, C. S.; GOPINATH, R. A.; GUO, H. **Introduction to Wavelets and Wavelet Transforms: a primer**. 1. ed. Upper Saddle River: Prentice Hall, 1998.

CANDES, E. J.; ROMBERG, J. K.; TAO, T. Stable signal recovery from incomplete and inaccurate measurements. **Communications on Pure and Applied Mathematics**, Hoboken, v. 59, n. 8, p. 1207–1223, 2006.

CARREIRA-PERPINAN, M. A. Mode-finding for mixtures of gaussian distributions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 22, n. 11, p. 1318–1323, 2000.

CARRERA, C.; MARGHOOB, A. A. Discriminating nevi from melanomas: clues and pitfalls. **Dermatologic Clinics**, Amsterdam, v. 34, n. 4, p. 395–409, 2016.

CAVALCANTI, P. G. **Automated prescreening of melanocytic skin lesions using standard camera images**. 2013. 108 p. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

CAVALCANTI, P. G.; SCHARCANSKI, J. Automated prescreening of pigmented skin lesions using standard cameras. **Computerized Medical Imaging and Graphics**, Amsterdam, v. 35, n. 6, p. 481–491, 2011.

CAVALCANTI, P. G.; SCHARCANSKI, J. A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images. **Computer Methods and Programs in Biomedicine**, Amsterdam, v. 112, n. 3, p. 684–693, 2013.

CAVALCANTI, P. G.; SCHARCANSKI, J.; BARANOSKI, G. V. A two-stage approach for discriminating melanocytic skin lesions using standard cameras. **Expert Systems with Applications**, Amsterdam, v. 40, n. 10, p. 4054–4064, 2013.

CAVALCANTI, P. G.; SCHARCANSKI, J.; LOPES, C. B. O. Shading attenuation in human skin color images. In: INTERNATIONAL SYMPOSIUM ON VISUAL COMPUTING (ISVC), 6., 2010, Las Vegas. **Proceedings** [...]. Berlin, Heidelberg: Springer, 2010. (Lecture Notes in Computer Science 6453), p. 190–198.

CAVALCANTI, P. G.; SCHARCANSKI, J.; MARTINEZ, C. E.; PERSIA, L. E. D. Segmentation of pigmented skin lesions using non-negative matrix factorization. In: INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE (I2MTC), 31., 2014, Montivideo. **Proceedings** [...]. Piscataway: IEEE, 2014. p. 72–75.

CAVALCANTI, P. G.; SCHARCANSKI, J.; PERSIA, L. E. D.; MILONE, D. H. An ICA-based method for the segmentation of pigmented skin lesions in macroscopic images. In: INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC), 33., 2011, Boston. **Proceedings** [...]. Piscataway: IEEE, 2011. p. 5993–5996.

CELEBI, M. E.; CODELLA, N.; HALPERN, A. Dermoscopy image analysis: overview and future directions. **IEEE Journal of Biomedical and Health Informatics**, Piscataway, v. 23, n. 2, p. 474–478, 2019.

CELEBI, M. E.; KINGRAVI, H. A.; VELA, P. A. A comparative study of efficient initialization methods for the k-means clustering algorithm. **Expert Systems with Applications**, Amsterdam, v. 40, n. 1, p. 200–210, 2013.

CHEN, C. P.; ZHANG, C.-Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. **Information Sciences**, Amsterdam, v. 275, p. 314–347, 2014.

CHEN, S. S.; DONOHO, D. L.; SAUNDERS, M. A. Atomic decomposition by basis pursuit. **SIAM Journal on Scientific Computing**, Philadelphia, v. 43, n. 1, p. 33–61, 1998.

CHEN, S. S.; DONOHO, D. L.; SAUNDERS, M. A. Atomic decomposition by basis pursuit. **SIAM Review**, Philadelphia, v. 43, n. 1, p. 129–159, 2001.

CHENG, H. **Sparse Representation, Modeling and Learning in Visual Recognition: theory, algorithms and applications**. 1. ed. London: Springer, 2015.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 30., 2017, Honolulu. **Proceedings** [...]. Piscataway: IEEE, 2017. p. 1251–1258.

CLARKE, L. E.; FLAKE, D. D.; BUSAM, K.; COCKERELL, C.; HELM, K.; MCNIFF, J.; REED, J.; TSCHEN, J.; KIM, J.; BARNHILL, R. et al. An independent validation of a gene expression signature to differentiate malignant melanoma from benign melanocytic nevi. **Cancer**, Hoboken, v. 123, n. 4, p. 617–628, 2017.

CLARKE, L. E.; WARF, M. B.; FLAKE, D. D.; HARTMAN, A.-R.; TAHAN, S.; SHEA, C. R.; GERAMI, P.; MESSINA, J.; FLORELL, S. R.; WENSTRUP, R. J. et al. Clinical validation of a gene expression signature that differentiates benign nevi from malignant melanoma. **Journal of Cutaneous Pathology**, Hoboken, v. 42, n. 4, p. 244–252, 2015.

CODELLA, N. C. F.; GUTMAN, D.; CELEBI, M. E.; HELBA, B.; MARCHETTI, M. A.; DUSZA, S. W.; KALLOO, A.; LIOPYRIS, K.; MISHRA, N.; KITTLER, H.; HALPERN, A. Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC). In: IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI), 15., 2018, Washington. **Proceedings** [...]. Piscataway: IEEE, 2018. p. 168–172.

COUR, T.; YU, S.; SHI, J. **MATLAB Normalized Cuts Segmentation Code**. 2010. Disponível em: <<https://www.cis.upenn.edu/~jshi/software/>>. Último acesso em: 09 set. 2022.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. 2. ed. Nashville: John Wiley & Sons, 2006.

CSURKA, G.; LARLUS, D.; PERRONNIN, F.; MEYLAN, F. What is a good evaluation measure for semantic segmentation? In: BRITISH MACHINE VISION CONFERENCE (BMVC), 24., 2013, Bristol. **Proceedings** [...]. Durham: BMVA Press, 2013. p. 32.1–32.11.

DA SILVA, E. A.; MENDONÇA, G. V. Digital image processing. In: CHEN, W.-K. (Ed.). **The Electrical Engineering Handbook**. Boston: Elsevier, 2005. p. 561–573.

DAUBECHIES, I.; DEFRISE, M.; MOL, C. D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. **Communications on Pure and Applied Mathematics**, Hoboken, v. 57, n. 11, p. 1413–1457, 2004.

DAVIS, G.; MALLAT, S.; AVELLANEDA, M. Adaptive greedy approximations. **Constructive Approximation**, Berlin, v. 13, n. 1, p. 57–98, 1997.

DAVIS, G. M.; MALLAT, S. G.; ZHANG, Z. Adaptive time-frequency decompositions. **Optical Engineering**, Bellingham, v. 33, n. 7, p. 2183–2192, 1994.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society: Series B (Methodological)**, Hoboken, v. 39, n. 1, p. 1–22, 1977.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine Learning Research**, v. 7, n. 1, p. 1–30, 2006. Disponível em: <<https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf>>. Último acesso em: 09 set. 2022.

DHILLON, I. S.; GUAN, Y.; KULIS, B. Kernel k-means: Spectral clustering and normalized cuts. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (KDD), 10., 2004, Seattle. **Proceedings** [...]. New York: ACM, 2004. p. 551–556.

DIETTERICH, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. **Neural Computation**, Cambridge, v. 10, n. 7, p. 1895–1923, 1998.

DING, C.; HE, X.; SIMON, H. D. On the equivalence of nonnegative matrix factorization and spectral clustering. In: SIAM INTERNATIONAL CONFERENCE ON DATA MINING (SDM), 5., 2005, Newport Beach. **Proceedings** [...]. Philadelphia: SIAM, 2005. p. 606–610.

DINIZ, P. S. R.; DA SILVA, E. A. B.; NETTO, S. L. **Processamento Digital de Sinais**. 2. ed. Porto Alegre: Bookman, 2014.

DOMINGOS, P. A few useful things to know about machine learning. **Communications of the ACM**, New York, v. 55, n. 10, p. 78–87, 2012.

DONOHO, D. L. For most large underdetermined systems of equations, the minimal  $l_1$ -norm near-solution approximates the sparsest near-solution. **Communications on Pure and Applied Mathematics**, Hoboken, v. 59, n. 7, p. 907–934, 2006.

DONOHO, D. L.; JOHNSTONE, J. M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, Oxford, v. 81, n. 3, p. 425–455, 1994.

DOUGHERTY, G. **Digital Image Processing for Medical Applications**. 1. ed. Cambridge: Cambridge University Press, 2009.

DUDA, R. O.; HART, P. E.; STORCK, D. G. **Pattern Classification**. 2. ed. New York: Wiley, 2001.

DUNN, O. J. Multiple comparisons among means. **Journal of the American Statistical Association**, Boston, v. 56, n. 293, p. 52–64, 1961.

EHRENFEUCHT, A.; HAUSSLER, D.; KEARNS, M.; VALIANT, L. A general lower bound on the number of examples needed for learning. **Information and Computation**, Amsterdam, v. 82, n. 3, p. 247–261, 1989.

ELAD, M. **Sparse and Redundant Representations: from theory to applications in signal and image processing**. 1. ed. New York: Springer, 2010.

ELAD, M.; AHARON, M. Image denoising via sparse and redundant representations over learned dictionaries. **IEEE Transactions on Image Processing**, Piscataway, v. 15, n. 12, p. 3736–3745, 2006.

ENGAN, K.; AASE, S. O.; HUSOY, J. H. Method of optimal directions for frame design. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 24., 1999, Phoenix. **Proceedings** [...]. Piscataway: IEEE, 1999. p. 2443–2446.

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**, Berlin, v. 542, n. 7639, p. 115–118, 2017.

FACELI, K.; LORENA, A. C.; GAMA, J.; ALMEIDA, T. A. de; CARVALHO, A. C. P. L. F. de. **Inteligência Artificial: uma abordagem de aprendizado de máquina**. 2. ed. Rio de Janeiro: LTC, 2021.

FIGUEIREDO, M. A.; NOWAK, R. D. An em algorithm for wavelet-based image restoration. **IEEE Transactions on Image Processing**, Piscataway, v. 12, n. 8, p. 906–916, 2003.

FLOREA, M. I.; VOROBYOV, S. A. A robust FISTA-like algorithm. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 42., 2017, New Orleans. **Proceedings [...]**. Piscataway: IEEE, 2017. p. 4521–4525.

FLORES, E.; SCHARCANSKI, J. Segmentation of melanocytic skin lesions using feature learning and dictionaries. **Expert Systems with Applications**, Amsterdam, v. 56, p. 300–309, 2016.

FLORES, E.; ZORTEA, M.; SCHARCANSKI, J. Dictionaries of deep features for land-use scene classification of very high spatial resolution images. **Pattern Recognition**, Amsterdam, v. 89, p. 32–44, 2019.

FORNACIALI, M. S. **Reliable Automated Melanoma Screening for the Real World**. 2019. 102 p. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2019.

FORSYTH, D. A.; PONCE, J. **Computer Vision: a modern approach**. 2. ed. Boston: Pearson, 2012.

FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. **Journal of the American Statistical Association**, Boston, v. 32, n. 200, p. 675–701, 1937.

FRIEDMAN, R. J.; RIGEL, D. S.; KOPF, A. W. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. **CA: a Cancer Journal for Clinicians**, Hoboken, v. 35, n. 3, p. 130–151, 1985.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. 2. ed. San Diego: Academic Press, 1990.

FUKUSHIMA, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. **Biological Cybernetics**, Berlin, v. 36, n. 4, p. 193–202, 1980.

FUKUSHIMA, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. **Neural networks**, Amsterdam, v. 1, n. 2, p. 119–130, 1988.

GANGEH, M. J.; FARAHAT, A. K.; GHODSI, A.; KAMEL, M. S. **Supervised Dictionary Learning and Sparse Representation – a review**. 2015. Disponível em: <<http://arxiv.org/abs/1502.05928>>. Último acesso em: 09 set. 2022.

GAUTAM, D.; AHMED, M.; MEENA, Y. K.; HAQ, A. U. Machine learning-based diagnosis of melanoma using macro images. **International Journal for Numerical Methods in Biomedical Engineering**, v. 34, n. 5, p. e2953, 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/cnm.2953>>. Último acesso em: 09 set. 2022.

GEMAN, S.; BIENENSTOCK, E.; DOURSAT, R. Neural networks and the bias/variance dilemma. **Neural Computation**, Cambridge, v. 4, n. 1, p. 1–58, 1992.

GIOTIS, I.; MOLDERS, N.; LAND, S.; BIEHL, M.; JONKMAN, M. F.; PETKOV, N. MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. **Expert Systems with Applications**, Amsterdam, v. 42, n. 19, p. 6578–6585, 2015.

GLAISTER, J.; AMELARD, R.; WONG, A.; CLAUSI, D. A. MSIM: multistage illumination modeling of dermatological photographs for illumination-corrected skin lesion analysis. **IEEE Transactions on Biomedical Engineering**, Piscataway, v. 60, n. 7, p. 1873–1883, 2013.

GLAISTER, J.; WONG, A.; CLAUSI, D. A. Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. **IEEE Transactions on Biomedical Engineering**, Piscataway, v. 61, n. 4, p. 1220–1230, 2014.

GLAZER, A. M.; RIGEL, D. S.; WINKELMANN, R. R.; FARBERG, A. S. Clinical diagnosis of skin cancer: enhancing inspection and early recognition. **Dermatologic Clinics**, Amsterdam, v. 35, n. 4, p. 409–416, 2017.

GOLUB, G. H.; VAN LOAN, C. F. **Matrix Computations**. 4. ed. Baltimore: JHU, 2013.

GONZALEZ, R. C.; WOODS, R. E. **Processamento Digital de Imagens**. 3. ed. São Paulo: Pearson, 2010.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. 4. ed. New York: Pearson, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT, 2016.

HA, Q.; LIU, B.; LIU, F. **Identifying melanoma images using EfficientNet ensemble: Winning solution to the SIIM-ISIC melanoma classification challenge**. 2020. Disponível em: <<https://arxiv.org/pdf/2010.05351.pdf>>. Último acesso em: 09 set. 2022.

HALL, P.; CLARIDGE, E.; SMITH, J. M. Computer screening for early detection of melanoma—is there a future? **British Journal of Dermatology**, Hoboken, v. 132, n. 3, p. 325–338, 1995.

HANCE, G. A.; UMBAUGH, S. E.; MOSS, R. H.; STOECKER, W. V. Unsupervised color image segmentation: with application to skin tumor borders. **IEEE Engineering in Medicine and Biology Magazine**, Piscataway, v. 15, n. 1, p. 104–111, 1996.

HASAN, M. K.; DAHAL, L.; SAMARAKOON, P. N.; TUSHAR, F. I.; MARTÍ, R. DSNet: Automatic dermoscopic skin lesion segmentation. **Computers in Biology and Medicine**, Amsterdam, v. 120, p. 103738, 2020.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. **The Elements of Statistical Learning: data mining, inference, and prediction**. 2. ed. New York: Springer, 2009.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 29., 2016, Las Vegas. **Proceedings** [...]. Piscataway: IEEE, 2016. p. 770–778.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Identity mappings in deep residual networks. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 14., 2016, Amsterdam. **Proceedings** [...]. Berlin, Heidelberg: Springer, 2016. p. 630–645.

HEATH, M. **Scientific Computing: an introductory survey**. 2. ed. Philadelphia: SIAM, 2018.

HOWARD, A.; SANDLER, M.; CHEN, B.; WANG, W.; CHEN, L.-C.; TAN, M.; CHU, G.; VASUDEVAN, V.; ZHU, Y.; PANG, R.; ADAM, H.; LE, Q. Searching for mobilenetv3. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 17., 2019, Seoul. **Proceedings** [...]. Piscataway: IEEE, 2019. p. 707–714.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017. Disponível em: <<https://arxiv.org/abs/1704.04861>>. Último acesso em: 09 set. 2022.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 30., 2017, Honolulu. **Proceedings** [...]. Piscataway: IEEE, 2017. p. 4700–4708.

HUGHES, G. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, Piscataway, v. 14, n. 1, p. 55–63, 1968.

IZBICKI, R.; SANTOS, T. M. dos. **Inteligência Artificial: uma abordagem estatística**. 1. ed. São Carlos: Rafael Izbicki, 2020.

JAFARI, M. H.; KARIMI, N.; NASR-ESFAHANI, E.; SAMAVI, S.; SOROUSH-MEHR, S. M. R.; WARD, K.; NAJARIAN, K. Skin lesion segmentation in clinical images using deep learning. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR), 23 2016, Cancun. **Proceedings** [...]. Piscataway: IEEE, 2016. p. 337–342.

JAIN, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, Amsterdam, v. 31, n. 8, p. 651–666, 2010.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, New York, v. 31, n. 3, p. 264–323, 1999.

JOHR, R. H. Dermoscopy: alternative melanocytic algorithms—the abcd rule of dermatoscopy, menzies scoring method, and 7-point checklist. **Clinics in Dermatology**, Amsterdam, v. 20, n. 3, p. 240–247, 2002.

KANUNGO, T.; MOUNT, D. M.; NETANYAHU, N. S.; PIATKO, C. D.; SILVERMAN, R.; WU, A. Y. An efficient  $k$ -means clustering algorithm: Analysis and implementation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 24, n. 7, p. 881–892, 2002.

KAUFMAN, H. **Melanoma**. 1. ed. Cham: Springer, 2016.

KAWAHARA, J.; DANESHVAR, S.; ARGENZIANO, G.; HAMARNEH, G. Seven-point checklist and skin lesion classification using multitask multimodal neural nets. **IEEE Journal of Biomedical and Health Informatics**, Piscataway, v. 23, n. 2, p. 538–546, 2018.

KHAN, A.; SOHAIL, A.; ZAHOORA, U.; QURESHI, A. S. A survey of the recent architectures of deep convolutional neural networks. **Artificial intelligence review**, Berlin, v. 53, n. 8, p. 5455–5516, 2020.

KRAUSE, A.; SINGH, A.; GUESTRIN, C. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. **Journal of Machine Learning Research**, v. 9, n. 2, 2008. Disponível em: <<https://dl.acm.org/doi/10.5555/1390681.1390689>>. Último acesso em: 09 set. 2022.

KREUTZ-DELGADO, K.; MURRAY, J. F.; RAO, B. D.; ENGAN, K.; LEE, T.-W.; SEJNOWSKI, T. J. Dictionary learning algorithms for sparse representation. **Neural Computation**, Cambridge, v. 15, n. 2, p. 349–396, 2003.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 26., 2012, Stateline. **Proceedings** [...]. Online: Curran Associates, Inc., 2012. (Advances in Neural Information Processing Systems, v. 25), p. 3320–3328. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>. Último acesso em: 09 set. 2022.

LAZOVA, R.; SMOOT, K.; ANDERSON, H.; POWELL, M. J.; ROSENBERG, A. S.; RONGIOLETTI, F.; PILLONI, L.; D’HALLEWIN, S.; GUEORGUIEVA, R.; TANTCHEVA-POÓR, I. et al. Histopathology-guided mass spectrometry differentiates benign nevi from malignant melanoma. **Journal of Cutaneous Pathology**, Hoboken, v. 47, n. 3, p. 226–240, 2020.

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. In: ARBIB, M. A. (Ed.). **The Handbook of Brain Theory and Neural Networks**. London: MIT, 1995. v. 2, p. 255–258.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Berlin, v. 521, n. 7553, p. 436–444, 2015.

LEE, D. D.; SEUNG, H. S. Learning the parts of objects by non-negative matrix factorization. **Nature**, Berlin, v. 401, n. 6755, p. 788–791, 1999.

LENAIL, A. NN-SVG: Publication-ready neural network architecture schematics. **Journal of Open Source Software**, v. 4, n. 33, p. 747, 2019. Disponível em: <<https://joss.theoj.org/papers/10.21105/joss.00747>>. Último acesso em: 09 set. 2022.

LEWICKI, M. S.; SEJNOWSKI, T. J. Learning overcomplete representations. **Neural Computation**, Cambridge, v. 12, n. 2, p. 337–365, 2000.

LI, Y.; NGOM, A. Sparse representation approaches for the classification of high-dimensional biological data. **BMC Systems Biology**, Berlin, v. 7, n. S4, p. S6, 2013.

LITJENS, G.; KOOL, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. **Medical Image Analysis**, Amsterdam, v. 42, p. 60–88, 2017.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 28., 2015, Boston. **Proceedings [...]**. Piscataway: IEEE, 2015. p. 3431–3440.

LOVISOLO, L. **Representações de Sinais Usando Dicionários Redundantes**. 2006. 237 p. Tese (Doutorado em Engenharia Elétrica) – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

LOVISOLO, L.; DA SILVA, E. A. B. Frames in signal processing. In: CHELLAPPA, R.; THEODORIDIS, S. (Ed.). **Signal Processing Theory and Machine Learning**. Amsterdam: Elsevier, 2014. v. 1, p. 561–590.

LUXBURG, U. von. A tutorial on spectral clustering. **Statistics and Computing**, Berlin, v. 17, n. 4, p. 395–416, 2007.

MAGALHAES, C.; MENDES, J.; VARDASCA, R. The role of AI classifiers in skin cancer images. **Skin Research and Technology**, Hoboken, v. 25, n. 5, p. 750–757, 2019.

MALLAT, S. G. **A Wavelet Tour of Signal Processing: the sparse way**. 3. ed. Burlington: Elsevier, 2009.

MALLAT, S. G.; ZHANG, Z. Matching pursuits with time-frequency dictionaries. **IEEE Transactions on Signal Processing**, Philadelphia, v. 41, n. 12, p. 3397–3415, 1993.

MARCHETTI, M. A.; CODELLA, N. C.; DUSZA, S. W.; GUTMAN, D. A.; HELBA, B.; KALLOO, A.; MISHRA, N.; CARRERA, C.; CELEBI, M. E.; DEFAZIO, J. L. et al. Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. **Journal of the American Academy of Dermatology**, Amsterdam, v. 78, n. 2, p. 270–277, 2018.

MARQUES, E. C.; MACIEL, N.; NAVINER, L.; CAI, H.; YANG, J. A review of sparse recovery algorithms. **IEEE Access**, Piscataway, v. 7, p. 1300–1322, 2018.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**. 1. ed. Rio de Janeiro: Brasport, 1999.

MARR, D. **Vision: A computational investigation into the human representation and processing of visual information**. 1. ed. New York: Henry Holt and Company, 1982.

MELAMED, R. D.; AYDIN, I. T.; RAJAN, G. S.; PHELPS, R.; SILVERS, D. N.; EMMETT, K. J.; BRUNNER, G.; RABADAN, R.; CELEBI, J. T. Genomic characterization of dysplastic nevi unveils implications for diagnosis of melanoma. **Journal of Investigative Dermatology**, Amsterdam, v. 137, n. 4, p. 905–909, 2017.

MENEGOLA, A.; FORNACIALI, M.; PIRES, R.; BITTENCOURT, F. V.; AVILA, S.; VALLE, E. Knowledge transfer for melanoma screening with deep learning. In: IEEE INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI), 14., 2016, Melbourne. **Proceedings [...]**. Piscataway: IEEE, 2017. p. 297–300.

MEYER, F. Iterative image transformations for an automatic screening of cervical smears. **Journal of Histochemistry & Cytochemistry**, Los Angeles, v. 27, n. 1, p. 128–135, 1979.

MICALI, G.; LACARRUBBA, F. Dermatoscopy: instrumental update. **Dermatologic Clinics**, Amsterdam, v. 36, n. 4, p. 345–348, 2018.

MITCHELL, T. M. **Machine Learning**. 1. ed. New York: McGraw-Hill, 1997.

MITSUI, H.; KIECKER, F.; SHEMER, A.; CANNIZZARO, M. V.; WANG, C. Q.; GULATI, N.; OHMATSU, H.; SHAH, K. R.; GILLEAUDEAU, P.; SULLIVAN-WHALEN, M. et al. Discrimination of dysplastic nevi from common melanocytic nevi by cellular and molecular criteria. **Journal of Investigative Dermatology**, Amsterdam, v. 136, n. 10, p. 2030–2040, 2016.

MORGAN, F. C.; DURAN, J.; FRAILE, B.; KARIA, P. S.; LIN, J. Y.; OTT, P. A.; RUIZ, E. S.; WANG, D. M.; ZHANG, Y.; SCHMULTS, C. D. A comparison of skin cancer screening and treatment costs at a massachusetts cancer center, 2008 versus 2013. **Journal of the American Academy of Dermatology**, Amsterdam, v. 79, n. 5, p. 921–928, 2018.

MURPHY, K. **Machine Learning: a probabilistic perspective**. 1. ed. Cambridge: MIT, 2012.

NASCIMENTO, M. C. V. **Metaheurísticas para o problema de agrupamento de dados em grafo**. 2010. 105 p. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – ICMC, Universidade de São Paulo, São Carlos, 2010.

NASCIMENTO, M. C. V.; CARVALHO, A. C. P. de Leon Ferreira de. Spectral methods for graph clustering—a survey. **European Journal of Operational Research**, Amsterdam, v. 211, n. 2, p. 221–231, 2011.

NATARAJAN, B. K. Sparse approximate solutions to linear systems. **SIAM Journal on Computing**, Philadelphia, v. 24, n. 2, p. 227–234, 1995.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. **Journal of the Royal Statistical Society: Series A (General)**, Hoboken, v. 135, n. 3, p. 370–384, 1972.

NELSON, K. C.; SWETTER, S. M.; SABODA, K.; CHEN, S. C.; CUIEL-LEWANDROWSKI, C. Evaluation of the number-needed-to-biopsy metric for the diagnosis of cutaneous melanoma: A systematic review and meta-analysis. **JAMA Dermatology**, Chicago, v. 155, n. 10, p. 1167–1174, 2019.

NEMENYI, P. B. **Distribution-free multiple comparisons**. 1963. 254 p. Tese (Doutorado em Matemática) – University Microfilms International, Princeton University, Ann Arbor, 1963.

NESTEROV, Y. A method of solving a convex programming problem with convergence rate  $o(\frac{1}{k^2})$ . **Doklady Akademii Nauk SSSR**, Moscow, v. 269, n. 3, p. 543–547, 1983.

NESTEROV, Y. **Lectures on Convex Optimization**. 2. ed. Cham: Springer, 2018.

OKUBOYEJO, D. A.; OLUGBARA, O. O. A review of prevalent methods for automatic skin lesion diagnosis. **The Open Dermatology Journal**, v. 12, n. 1, 2018. Disponível em: <<https://opendermatologyjournal.com/VOLUME/12/PAGE/14/FULLTEXT/>>. Último acesso em: 09 set. 2022.

OLIVEIRA, R. B.; FILHO, E. M.; MA, Z.; PAPA, J. P.; PEREIRA, A. S.; TAVARES, J. M. R. Computational methods for the image segmentation of pigmented skin lesions: a review. **Computer Methods and Programs in Biomedicine**, Amsterdam, v. 131, p. 127–141, 2016.

OLIVEIRA, R. B.; PAPA, J. P.; PEREIRA, A. S.; TAVARES, J. M. R. Computational methods for pigmented skin lesion classification in images: review and future trends. **Neural Computing and Applications**, Berlin, v. 29, n. 3, p. 613–636, 2018.

OLSHAUSEN, B. A.; FIELD, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. **Nature**, Berlin, v. 381, n. 6583, p. 607, 1996.

OTSU, N. A threshold selection method from gray-level histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, Piscataway, v. 9, n. 1, p. 62–66, 1979.

PACHECO, A. G.; KROHLING, R. A. Recent advances in deep learning applied to skin cancer detection. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS) – RETROSPECTIVES WORKSHOP, 33., 2019, Vancouver. **Proceedings** [...]. Online: Curran Associates, Inc., 2019. (Advances in Neural Information Processing Systems, v. 32), p. 1–9. Disponível em: <<https://arxiv.org/abs/1912.03280>>. Último acesso em: 09 set. 2022.

PACHECO, A. G.; TRAPPENBERG, T.; KROHLING, R. A. Learning dynamic weights for an ensemble of deep models applied to medical imaging classification. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 30., 2020, Glasgow. **Proceedings** [...]. Piscataway: IEEE, 2020. p. 1–8.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, Piscataway, v. 22, n. 10, p. 1345–1359, 2009.

PARIKH, N.; BOYD, S. et al. Proximal algorithms. **Foundations and Trends® in Optimization**, Boston, v. 1, n. 3, p. 127–239, 2014.

PATEL, V. M.; CHELLAPA, R. **Sparse Representations and Compressive Sensing for Imaging and Vision**. 1. ed. New York: Springer, 2013.

PATHAN, S.; PRABHU, K. G.; SIDDALINGASWAMY, P. Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. **Biomedical Signal Processing and Control**, Amsterdam, v. 39, p. 237–262, 2018.

PATI, Y. C.; REZAIIFAR, R.; KRISHNAPRASAD, P. S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In: ASILOMAR CONFERENCE ON SIGNALS, SYSTEMS AND COMPUTERS (ACSSC), 27., 1993, Pacific Grove. **Proceedings [...]**. Piscataway: IEEE, 1993. p. 40–44.

PEREDNIA, D. A. What dermatologists should know about digital imaging. **Journal of the American Academy of Dermatology**, Amsterdam, v. 25, n. 1, p. 89–108, 1991.

PETROU, M.; PETROU, C. **Image Processing: the fundamentals**. 2. ed. Chichester: Wiley, 2010.

POUYANFAR, S.; SADIQ, S.; YAN, Y.; TIAN, H.; TAO, Y.; REYES, M. P.; SHYU, M.-L.; CHEN, S.-C.; IYENGAR, S. A survey on deep learning: Algorithms, techniques, and applications. **ACM Computing Surveys**, New York, v. 51, n. 5, p. 1–36, 2018.

PRATT, W. K. Vector space formulation of two-dimensional signal processing operations. **Computer Graphics and Image Processing**, Amsterdam, v. 4, n. 1, p. 1–24, 1975.

QIU, Q.; JIANG, Z.; CHELLAPPA, R. Sparse dictionary-based representation and recognition of action attributes. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 13., 2011, Barcelona. **Proceedings [...]**. Piscataway: IEEE, 2011. p. 707–714.

QIU, Q.; PATEL, V. M.; CHELLAPPA, R. Information-theoretic dictionary learning for image classification. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 36, n. 11, p. 2173–2184, 2014.

RABBANI, M.; JOSHI, R. An overview of the JPEG 2000 still image compression standard. **Signal Processing: Image communication**, Amsterdam, v. 17, n. 1, p. 3–48, 2002.

RAO, K. R.; YIP, P. C. **The Transform and Data Compression Handbook**. 1. ed. Boca Raton: CRC, 2001.

RASMUSSEN, C. E.; WILLIAMS, C. K. I. **Gaussian processes for machine learning**. 1. ed. London: MIT Press, 2006.

RIKER, A. **Melanoma: a modern multidisciplinary approach**. 1. ed. Cham: Springer, 2018.

RISH, I.; GRABARNIK, G. Y. **Sparse Modeling: theory, algorithms, and applications**. 1. ed. Boca Raton: CRC, 2015.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional networks for biomedical image segmentation. In: **MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION (MICCAI)**, 18., 2015, Munich. **Proceedings** [...]. Cham: Springer, 2015. p. 234–241.

RUBINSTEIN, R.; BRUCKSTEIN, A. M.; ELAD, M. Dictionaries for sparse representation modeling. **Proceedings of the IEEE**, Piscataway, v. 98, n. 6, p. 1045–1057, 2010.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, Berlin, v. 115, n. 3, p. 211–252, 2015.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)**, 31., 2018, Salt Lake City. **Proceedings** [...]. Piscataway: IEEE, 2018. p. 4510–4520.

SCHARCANSKI, J.; CELEBI, M. E. **Computer Vision Techniques for the Diagnosis of Skin Cancer**. 1. ed. Berlin, Heidelberg: Springer, 2014. (Series in BioEngineering).

SCHULZ, A.; SILVA, E. A. B. D.; VELHO, L. **Compressive Sensing**. 1. ed. Rio de Janeiro: IMPA, 2009. (Publicações Matemáticas).

SEUNG, H. S.; LEE, D. D. The manifold ways of perception. **Science**, Washington, v. 290, n. 5500, p. 2268–2269, 2000.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, Piscataway, v. 27, n. 3, p. 379–423, 1948.

SHELHAMER, E.; LONG, J.; DARRELL, T. Fully convolutional networks for semantic segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 39, n. 4, p. 640–651, 2017.

SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. **Annual Review of Biomedical Engineering**, San Mateo, v. 19, p. 221–248, 2017.

SHI, J.; MALIK, J. Normalized cuts and image segmentation. In: **IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)**, 11., 1997, San Juan. **Proceedings** [...]. Piscataway: IEEE, 1997. p. 731–737.

SHI, J.; MALIK, J. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 22, n. 8, p. 888–905, 2000.

SHIN, H. C.; 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**, Piscataway, v. 35, n. 5, p. 1285–1298, 2016.

SHOR, N. **Minimization Methods for Non-Differentiable Functions**. 1. ed. Berlin: Springer, 1985.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR), 14., 2015, San Diego. **Proceedings** [...]. Online: ICLR, 2015. Disponível em: <<https://arxiv.org/abs/1409.1556>>. Último acesso em: 09 set. 2022.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1929–1958, 2014. Disponível em: <<http://jmlr.org/papers/v15/srivastava14a.html>>. Último acesso em: 09 set. 2022.

STOECKER, W. V.; MOSS, R. H. Editorial: digital imaging in dermatology. **Computerized Medical Imaging and Graphics**, Amsterdam, v. 16, n. 3, p. 145–150, 1992.

SWETTER, S. M.; TSAO, H.; BICHAKJIAN, C. K.; CURIEL-LEWANDROWSKI, C.; ELDER, D. E.; GERSHENWALD, J. E.; GUILD, V.; GRANT-KELS, J. M.; HALPERN, A. C.; JOHNSON, T. M. et al. Guidelines of care for the management of primary cutaneous melanoma. **Journal of the American Academy of Dermatology**, Amsterdam, v. 80, n. 1, p. 208–250, 2019.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 31., 2017, San Francisco. **Proceedings** [...]. Palo Alto: AAAI, 2017. p. 4278–4284.

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 29., 2016, Las Vegas. **Proceedings** [...]. Piscataway: IEEE, 2016. p. 2818–2826.

TAN, K.-H.; LIM, B. P. The artificial intelligence renaissance: Deep learning and the road to human-level machine intelligence. **APSIPA Transactions on Signal and Information Processing**, Cambridge, v. 7, 2018.

TAN, M.; LE, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML), 36, 2019, Long Beach. **Proceedings** [...]. Online: PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 6105–6114. Disponível em: <<https://proceedings.mlr.press/v97/tan19a.html>>. Último acesso em: 09 set. 2022.

TANG, J. A multi-direction GVF snake for the segmentation of skin cancer images. **Pattern Recognition**, Amsterdam, v. 42, n. 6, p. 1172–1179, 2009.

TCHEOU, M. P. **Compressão de Sinais Usando Decomposições Atômicas com Base em Dicionários Redundantes**. 2011. 167 p. Tese (Doutorado em Engenharia Elétrica) – COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

- TIBSHIRANI, R. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, Hoboken, v. 58, n. 1, p. 267–288, 1996.
- TORRES, R.; LANG, U. E.; HEJNA, M.; SHELTON, S. J.; JOSEPH, N. M.; SHAIN, A. H.; YEH, I.; WEI, M. L.; OLDHAM, M. C.; BASTIAN, B. C. et al. MicroRNA ratios distinguish melanomas from nevi. **Journal of Investigative Dermatology**, Amsterdam, v. 140, n. 1, p. 164–173, 2020.
- TOSIC, I.; FROSSARD, P. Dictionary learning: what is the right representation for my signal? **IEEE Signal Processing Magazine**, Piscataway, v. 28, n. 2, p. 27–38, 2011.
- TROPP, J. A.; WRIGHT, S. J. Computational methods for sparse solution of linear inverse problems. **Proceedings of the IEEE**, Piscataway, v. 98, n. 6, p. 948–958, 2010.
- TURANI, Z.; FATEMIZADEH, E.; BLUMETTI, T.; DAVELUY, S.; MORAES, A. F.; CHEN, W.; MEHREGAN, D.; ANDERSEN, P. E.; NASIRIAVANAKI, M. Optical radiomic signatures derived from optical coherence tomography images improve identification of melanoma. **Cancer Research**, Philadelphia, v. 79, n. 8, p. 2021–2030, 2019.
- UNIVERSITY OF EDINBURGH. **Dermofit Image Library**. 2022. Disponível em: <<https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library>>. Último acesso em: 09 set. 2022.
- VALIANT, L. G. A theory of the learnable. **Communications of the ACM**, New York, v. 27, n. 11, p. 1134–1142, 1984.
- VAPNIK, V. N.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. **Theory of Probability and its Applications**, Philadelphia, v. 16, n. 2, p. 264, 1971.
- VELHO, L.; FRERY, A. C.; GOMES, J. **Image Processing for Computer Graphics and Vision**. 2. ed. London: Springer, 2009.
- WANG, R. **Introduction to Orthogonal Transforms: with applications in data processing and analysis**. 1. ed. Cambridge: Cambridge University Press, 2012.
- WEBER, P.; TSCHANDL, P.; SINZ, C.; KITTLER, H. Dermatoscopy of neoplastic skin lesions: recent advances, updates, and revisions. **Current Treatment Options in Oncology**, Berlin, v. 19, n. 11, p. 56, 2018.
- WOLPERT, D. H. The lack of a priori distinctions between learning algorithms. **Neural Computation**, Cambridge, v. 8, n. 7, p. 1341–1390, 1996.
- WONG, A.; SCHARCANSKI, J.; FIEGUTH, P. Automatic skin lesion segmentation via iterative stochastic region merging. **IEEE Transactions on Information Technology in BioMedicine**, Piscataway, v. 15, n. 6, p. 929–936, 2011.
- WOZA, M. **Skin Lesion Analyzer**. 2018. Disponível em: <<http://skin.test.woza.work/>>. Último acesso em: 09 set. 2022.

WRIGHT, J.; YANG, A. Y.; GANESH, A.; SASTRY, S. S.; MA, Y. Robust face recognition via sparse representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 31, n. 2, p. 210–227, 2009.

WU, Z.; LEAHY, R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Piscataway, v. 15, n. 11, p. 1101–1113, 1993.

YANG, J.; PENG, Y.; XU, W.; DAI, Q. Ways to sparse representation: an overview. **Science in China Series F: Information Sciences**, Berlin, v. 52, n. 4, p. 695–703, 2009.

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks? In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 28., 2014, Montréal. **Proceedings** [...]. Online: Curran Associates, Inc., 2014. (Advances in Neural Information Processing Systems, v. 27), p. 3320–3328. Disponível em: <<http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf>>. Último acesso em: 09 set. 2022.

ZHANG, J.; XIE, Y.; XIA, Y.; SHEN, C. Attention residual learning for skin lesion classification. **IEEE Transactions on Medical Imaging**, Piscataway, v. 38, n. 9, p. 2092–2103, 2019.

ZHANG, Z.; XU, Y.; YANG, J.; LI, X.; ZHANG, D. A survey of sparse representation: algorithms and applications. **IEEE Access**, Piscataway, v. 3, p. 490–530, 2015.

ZHAO, H.; SHI, J.; QI, X.; WANG, X.; JIA, J. Pyramid scene parsing network. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 30., 2017, Honolulu. **Proceedings** [...]. Piscataway: IEEE, 2017. p. 6230–6239.

ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable architectures for scalable image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 31., 2018, Salt Lake City. **Proceedings** [...]. Piscataway: IEEE, 2018. p. 8697–8710.

ZORTEA, M.; FLORES, E.; SCHARCANSKI, J. A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images. **Pattern Recognition**, Amsterdam, v. 64, p. 92–104, 2017.

## APÊNDICE A — RESULTADOS SUPLEMENTARES

Tabela A.1: Ranqueamento médio de diferentes configurações do método de classificação proposto (versão completa).

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	857,75	834,75	664,05	673,45	626,40
0,02	869,95	737,35	677,50	688,90	627,15
0,05	792,15	713,50	669,30	623,50	624,65
0,1	821,05	806,80	860,55	644,35	623,50
0,2	824,50	787,60	711,70	640,65	616,60
0,5	884,15	746,60	638,00	604,15	599,95
1	906,55	702,35	581,20	566,75	589,95
2	809,65	736,90	681,00	508,20	551,35
5	865,15	798,55	548,40	517,45	<b>497,00</b>

(a) DenseNet-121.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	749,15	782,80	563,35	693,35	593,25
0,02	815,55	606,75	591,65	570,65	593,25
0,05	744,30	667,60	653,70	648,90	588,30
0,1	867,85	823,75	616,10	577,70	586,90
0,2	782,90	525,20	632,40	651,65	580,45
0,5	795,35	819,05	500,55	549,35	560,20
1	888,50	723,30	584,10	530,85	543,60
2	817,60	716,15	547,75	578,10	508,95
5	909,15	648,75	634,90	481,75	<b>463,50</b>

(b) DenseNet-169.

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
0,01	460,80	618,65	539,40	701,00	738,70
0,02	612,00	592,75	573,35	702,45	738,10
0,05	524,70	696,55	559,85	659,25	734,55
0,1	570,35	505,25	776,55	780,40	729,05
0,2	648,30	577,45	602,90	711,55	712,90
0,5	582,35	649,75	668,15	652,50	685,45
1	480,45	722,60	549,00	656,60	656,25
2	632,65	541,45	565,70	631,80	596,55
<b>5</b>	682,45	553,90	<b>436,35</b>	525,75	495,30

(c) DenseNet-201.

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
0,01	485,50	486,75	312,45	423,10	385,60
0,02	477,40	399,95	311,65	405,45	385,45
0,05	611,55	457,00	387,45	416,25	374,55
0,1	470,95	373,55	403,25	470,50	366,95
0,2	604,15	423,00	358,40	339,75	359,00
<b>0,5</b>	474,00	452,70	<b>292,45</b>	353,55	335,30
1	529,40	395,30	364,65	329,50	329,70
2	344,90	466,20	430,90	393,95	346,10
5	678,90	657,55	441,10	476,40	441,65

(d) EfficientNet-B0.

$\lambda \backslash \varphi$	0,2	0,4	0,6	<b>0,8</b>	1
0,01	336,95	282,85	208,45	336,95	326,65
0,02	245,10	270,15	250,10	350,25	326,05

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,05	222,80	301,75	348,80	320,80	325,00
0,1	254,05	156,85	362,10	370,85	315,35
0,2	382,05	204,55	272,45	294,50	292,15
0,5	355,50	228,75	241,90	270,00	262,95
1	276,60	323,85	170,80	230,05	220,10
2	233,10	153,60	227,30	227,35	196,40
5	375,90	406,90	251,00	<b>152,15</b>	179,80

(e) EfficientNet-B1.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	866,45	567,80	601,90	479,30	509,10
0,02	603,60	606,35	590,30	555,70	508,30
0,05	891,10	671,20	613,50	514,70	507,20
0,1	758,10	736,00	597,60	533,70	506,25
0,2	720,10	700,40	585,85	486,75	502,55
0,5	788,25	575,60	538,90	572,30	496,25
1	905,65	567,50	599,60	568,15	482,95
2	759,35	733,25	577,05	495,85	<b>466,20</b>
5	1032,95	804,90	706,10	687,75	497,20

(f) EfficientNet-B2.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	879,20	827,70	838,55	871,35	823,10
0,02	835,95	731,30	715,85	706,95	821,50
0,05	855,00	736,10	859,30	698,50	815,40
0,1	903,55	887,00	668,25	820,65	803,15

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,2	868,75	804,05	772,55	791,25	780,40
0,5	813,00	688,25	725,90	796,75	731,25
1	825,70	801,75	794,95	711,95	679,95
2	850,10	851,65	613,20	711,00	645,35
5	955,50	802,85	747,90	733,90	<b>544,95</b>

(g) EfficientNet-B3.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	573,20	811,35	803,75	883,45	878,95
0,02	844,95	731,85	857,85	817,75	876,95
0,05	770,30	763,20	946,05	861,35	868,70
0,1	734,70	621,90	892,70	817,50	846,90
0,2	680,10	607,15	732,95	736,95	813,50
0,5	535,75	734,55	558,45	788,95	718,25
1	814,60	588,00	670,95	531,00	588,30
2	680,95	810,95	608,35	470,10	<b>423,75</b>
5	828,90	803,65	653,95	584,45	487,70

(h) EfficientNet-B4.

$\lambda \backslash \varphi$	0,2	<b>0,4</b>	0,6	0,8	1
0,01	848,90	656,25	706,15	709,60	651,05
0,02	719,75	783,30	702,50	664,55	644,75
0,05	595,95	681,90	707,10	699,05	635,60
0,1	637,60	808,75	681,00	617,15	619,40
0,2	596,55	737,35	601,00	576,35	585,55
<b>0,5</b>	607,50	<b>481,60</b>	615,15	502,90	502,65

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	<b>0,4</b>	0,6	0,8	1
1	753,10	731,65	587,65	500,15	484,35
2	766,55	723,60	488,45	563,65	524,45
5	1022,45	909,70	803,50	734,50	643,30

(i) EfficientNet-B5.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	<b>1</b>
0,01	571,40	630,45	588,55	439,00	451,30
0,02	478,15	392,00	502,25	456,15	448,25
0,05	562,90	626,90	555,50	444,20	432,85
0,1	594,45	547,95	455,40	487,15	413,50
0,2	539,00	529,60	486,00	419,50	396,75
<b>0,5</b>	624,05	529,75	417,45	435,10	<b>371,85</b>
1	518,45	486,55	470,30	400,60	388,45
2	770,30	714,45	562,40	450,40	443,50
5	675,15	775,30	580,15	635,90	509,65

(j) EfficientNet-B6.

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
0,01	411,95	475,20	359,05	394,25	451,95
0,02	406,40	365,15	369,65	508,00	450,60
0,05	460,00	292,50	457,15	448,40	438,55
0,1	285,65	447,80	482,70	385,00	429,55
0,2	410,75	447,80	322,60	370,85	410,20
<b>0,5</b>	455,10	293,60	<b>250,95</b>	340,85	402,55
1	594,90	499,30	397,10	311,35	378,60
2	689,50	493,05	416,90	433,80	426,30

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
5	784,80	633,60	748,20	600,50	549,05

(k) EfficientNet-B7.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	<b>1</b>
0,01	1017,40	847,95	842,35	749,85	794,00
0,02	1095,70	947,15	938,80	844,00	790,80
0,05	934,45	974,70	726,50	764,50	787,95
0,1	986,65	808,50	819,90	837,65	777,90
0,2	990,40	1022,70	719,80	857,05	766,45
0,5	947,55	999,15	823,30	831,25	737,15
1	1093,70	882,10	785,75	817,40	702,70
2	1055,85	873,70	830,20	646,30	672,60
<b>5</b>	1058,40	825,70	722,15	752,95	<b>616,40</b>

(l) InceptionResNet-v2.

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
0,01	721,05	833,80	921,85	986,00	956,80
0,02	782,70	819,45	858,55	937,95	954,65
0,05	846,00	790,65	800,15	941,10	951,40
0,1	706,20	816,15	899,85	929,40	942,70
0,2	781,75	815,80	933,90	871,40	938,25
0,5	725,70	757,95	868,55	956,10	904,30
1	796,10	731,70	847,75	931,05	858,90
2	858,30	719,15	780,10	755,15	780,90
<b>5</b>	744,50	737,85	<b>642,20</b>	734,50	702,55

(m) InceptionNet-v3.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	731,30	721,30	756,05	713,65	826,20
0,02	788,55	676,10	799,10	843,85	824,75
0,05	865,05	683,35	648,55	753,80	823,30
0,1	714,10	717,70	739,45	839,30	817,75
0,2	742,75	825,45	806,35	709,00	818,05
0,5	935,30	717,90	670,15	767,50	789,80
1	656,10	779,15	738,30	662,35	775,20
2	581,25	704,30	647,65	720,50	721,75
5	672,90	676,15	598,70	<b>554,90</b>	612,60

(n) MobileNet.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	444,30	499,65	432,75	410,25	420,60
0,02	450,45	478,05	504,60	387,55	420,40
0,05	594,20	426,30	409,40	396,10	420,15
0,1	482,35	404,90	408,85	450,40	415,40
0,2	465,90	367,30	343,00	403,15	410,15
0,5	442,35	449,70	342,15	401,95	394,50
1	477,70	306,90	308,40	285,60	367,55
2	482,30	327,80	308,90	347,15	322,45
5	337,00	256,15	277,55	275,65	<b>239,40</b>

(o) MobileNet-v2.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	859,05	771,45	561,75	461,00	425,05
0,02	764,40	614,30	527,45	515,80	421,95

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,05	759,75	591,45	560,35	436,90	418,05
0,1	782,25	858,90	589,75	407,40	407,30
0,2	578,75	562,50	596,40	494,10	395,00
0,5	625,00	532,85	583,55	496,25	366,85
1	659,00	676,00	512,55	409,70	347,80
2	658,20	623,80	544,40	405,70	296,10
5	723,15	623,50	535,20	437,10	<b>245,45</b>

(p) MobileNet-v3-large.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	632,35	542,75	561,30	494,15	475,75
0,02	612,80	386,55	439,85	476,80	476,15
0,05	664,55	457,70	532,30	507,80	470,60
0,1	547,70	521,45	486,95	401,70	463,15
0,2	706,05	497,55	466,85	447,80	455,85
0,5	600,85	478,50	540,95	472,80	441,75
1	646,90	350,80	479,25	392,30	425,25
2	588,35	434,90	<b>325,85</b>	370,95	377,50
5	611,65	521,05	411,20	444,85	381,25

(q) MobileNet-v3-small.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	697,60	756,30	877,00	822,10	871,90
0,02	883,25	882,55	788,00	890,90	873,70
0,05	767,55	883,70	880,85	893,10	868,30
0,1	848,70	790,35	827,15	821,95	860,10

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
<b>0,2</b>	<b>645,25</b>	910,35	849,95	826,20	847,90
0,5	772,45	670,65	737,80	755,85	824,15
1	725,50	747,00	765,70	758,00	792,70
2	840,25	731,90	819,00	745,10	770,90
5	836,90	891,15	879,95	727,65	788,50

(r) NASNet-large.

$\lambda \backslash \varphi$	0,2	<b>0,4</b>	0,6	0,8	1
0,01	730,50	652,70	744,05	632,60	674,55
0,02	852,15	700,50	682,15	753,65	676,05
0,05	671,95	630,50	670,10	689,45	675,50
0,1	677,75	788,75	677,80	673,20	676,95
0,2	860,00	733,35	703,70	598,00	683,30
<b>0,5</b>	688,45	<b>568,95</b>	699,10	627,70	699,85
1	890,25	654,85	728,30	797,55	728,05
2	644,95	609,20	715,25	743,80	805,35
5	1048,70	777,75	840,55	790,75	894,55

(s) NASNet-mobile.

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
0,01	796,00	598,75	455,80	587,85	600,45
0,02	716,55	487,35	530,60	590,45	599,70
0,05	696,40	584,35	580,15	571,70	599,15
<b>0,1</b>	661,15	482,45	<b>433,85</b>	599,25	595,75
0,2	736,15	495,20	499,15	645,10	591,70
0,5	808,70	626,00	505,80	607,55	576,45

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	<b>0,6</b>	0,8	1
1	636,45	738,95	508,85	567,05	563,65
2	801,80	621,40	463,85	617,55	544,40
5	664,75	682,45	558,70	554,25	477,20

(t) ResNet-101.

$\lambda \backslash \varphi$	<b>0,2</b>	0,4	0,6	0,8	1
0,01	1045,10	1137,05	1077,50	1085,65	1090,90
<b>0,02</b>	<b>995,90</b>	1156,05	1140,75	1041,00	1090,90
0,05	1123,20	1109,85	1097,95	1106,75	1090,35
0,1	1022,45	1054,45	1104,80	1068,60	1092,30
0,2	1009,50	1170,15	1134,65	1063,55	1092,60
0,5	1167,15	1176,70	1154,10	1081,85	1091,45
1	1159,55	1152,85	1149,65	1109,10	1092,70
2	1020,10	1081,95	1094,30	1064,35	1086,15
5	1071,85	1069,65	1116,70	1046,10	1076,80

(u) ResNet-101-v2.

$\lambda \backslash \varphi$	<b>0,2</b>	0,4	0,6	0,8	1
0,01	518,85	458,55	463,15	525,85	571,80
0,02	507,45	380,40	526,95	553,10	570,40
0,05	761,25	510,20	442,50	524,40	571,60
0,1	434,05	368,20	480,10	546,65	568,40
<b>0,2</b>	<b>270,80</b>	491,70	556,15	530,60	561,85
0,5	348,65	568,00	439,45	567,95	556,85
1	281,85	461,50	455,50	457,45	545,30
2	480,50	481,95	304,00	464,00	523,75

(continua na próxima página)

$\lambda \backslash \varphi$	<b>0,2</b>	0,4	0,6	0,8	1
5	497,00	350,80	418,00	458,10	433,45

(v) ResNet-152.

$\lambda \backslash \varphi$	0,2	<b>0,4</b>	0,6	0,8	1
0,01	789,45	923,10	911,30	930,95	875,50
0,02	939,25	845,65	824,55	872,60	875,70
0,05	840,30	979,95	927,15	991,95	878,80
0,1	881,15	958,25	894,80	946,00	877,35
0,2	832,50	968,65	954,00	918,90	873,35
0,5	863,30	895,30	939,05	925,45	867,75
<b>1</b>	855,40	<b>782,20</b>	892,70	979,50	860,70
2	1043,20	885,45	881,95	900,60	855,10
5	890,80	830,50	890,90	867,05	826,10

(w) ResNet-152-v2.

$\lambda \backslash \varphi$	0,2	0,4	0,6	<b>0,8</b>	1
0,01	434,75	340,00	236,80	210,05	193,65
0,02	284,45	203,10	268,15	231,85	192,60
0,05	472,90	281,50	137,40	169,55	191,90
0,1	417,20	279,05	160,75	227,80	190,85
0,2	326,50	131,25	237,60	190,35	192,85
0,5	249,55	242,30	278,15	215,20	186,75
1	353,80	301,65	123,25	159,30	184,70
2	465,90	214,15	207,75	139,30	170,25
<b>5</b>	353,90	227,65	139,00	<b>109,35</b>	131,90

(x) ResNet-50.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	746,25	787,25	901,20	807,50	763,45
0,02	789,55	879,35	849,95	673,70	763,90
0,05	783,50	786,50	781,05	784,40	762,25
0,1	691,80	660,15	765,05	839,45	763,95
0,2	651,25	952,00	889,70	755,20	763,20
0,5	668,35	780,80	853,70	893,00	756,55
1	742,60	802,30	733,15	849,00	745,30
2	697,45	764,00	801,35	750,15	736,70
5	829,45	<b>586,60</b>	712,00	789,00	688,90

(y) ResNet-50-v2.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	334,40	363,75	381,05	333,25	281,20
0,02	259,75	408,50	354,40	348,65	280,75
0,05	244,65	429,60	316,85	272,55	283,20
0,1	322,65	390,00	390,75	444,45	282,45
0,2	234,95	471,20	321,45	374,65	279,80
0,5	374,00	454,85	426,70	364,00	265,80
1	235,35	331,35	358,95	292,30	264,60
2	341,90	574,95	328,40	268,60	259,40
5	361,30	360,90	438,20	272,15	<b>222,70</b>

(z) VGGNet-16.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	656,70	699,05	598,65	658,95	568,80
0,02	652,05	726,15	620,30	605,15	568,05

(continua na próxima página)

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,05	549,50	634,90	526,25	600,60	564,40
0,1	643,95	668,60	616,35	717,00	563,70
0,2	568,25	510,35	579,05	623,80	561,10
0,5	620,65	597,55	664,95	540,35	551,60
1	718,00	719,35	675,80	609,15	530,25
2	606,50	682,00	600,00	623,85	486,30
5	677,15	636,45	454,05	520,70	<b>392,70</b>

(aa) VGGNet-19.

$\lambda \backslash \varphi$	0,2	0,4	0,6	0,8	1
0,01	957,65	899,30	858,15	761,05	734,35
0,02	826,30	806,10	699,20	808,65	734,10
0,05	930,10	709,50	807,75	726,90	729,90
0,1	924,80	730,60	693,50	709,50	724,15
0,2	926,80	775,65	941,15	787,60	701,15
0,5	916,45	<b>672,20</b>	805,10	797,25	692,05
1	1064,50	931,25	896,65	707,35	688,30
2	1002,85	1019,30	888,25	784,30	706,40
5	1023,25	1065,15	873,45	858,15	780,05

(ab) XceptionNet.

Fonte: Os autores.