



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Combinação da Regra ABCD e CNN's Pré-treinadas no Diagnóstico Automático de Melanoma

Nayara Holanda de Moura

Número de Ordem PPGCC: M001

Teresina-PI, Abril de 2018

Nayara Holanda de Moura

Combinação da Regra ABCD e CNN's Pré-treinadas no Diagnóstico Automático de Melanoma

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Rodrigo de Melo Souza Veras

Teresina-PI

Abril de 2018

Nayara Holanda de Moura

Combinação da Regra ABCD e CNN's Pré-treinadas no Diagnóstico Automático de Melanoma/ Nayara Holanda de Moura. – Teresina-PI, Abril de 2018-
61 p. : il. (algumas color.) ; 30 cm.

Orientador: Rodrigo de Melo Souza Veras

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação, Abril de 2018.

1. Melanoma. 2. Descritor Híbrido. 3. *Multilayer Perceptron*. I. Rodrigo de Melo Souza Veras. II. Universidade Federal do Piauí. III. Combinação da regra ABCD e CNN's Pré-treinadas no Diagnóstico Automático de Melanoma.

CDU 02:141:005.7

“Combinação da regra ABCD e CNN’s Pré-treinadas no Diagnóstico Automático de Melanoma”

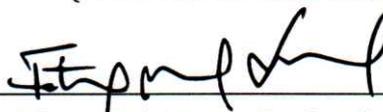
NAYARA HOLANDA DE MOURA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Aprovada por:



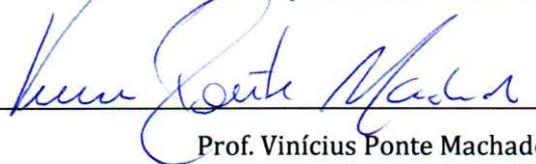
Prof. Rodrigo de Melo Souza Vêras
(Presidente da Banca Examinadora)



Profa. Fátima Nelsizeuma Sombra de Medeiros
(Examinadora Externa)



Prof. Kelson Rômulo Teixeira Aires
(Examinador Interna)



Prof. Vinícius Ponte Machado
(Examinador Interno)

Teresina, 04 de abril de 2018

Aos meus pais Maria de Fátima Holanda Moura e Gerson Raimundo de Moura, por todo amor, carinho, ensinamentos e por nunca deixarem de acreditar na minha capacidade.

Agradecimentos

A Deus por guiar meus passos e preencher-me de coragem.

Aos meus pais, Maria de Fátima Holanda Moura e Gerson Raimundo de Moura, pelo amor incondicional, apoio e inspiração para que eu alcance novas realizações pessoais e profissionais.

Aos meus irmãos, Ionara Holanda de Moura e Leonardo Holanda de Moura, por torcerem pelo meu sucesso, pela amizade e companheirismo, sempre me apoiando e me estimulando nos estudos.

À minha sobrinha, Heloísa Meneses Holanda Moura, por ser a alegria da nossa casa.

À minha avó (*In Memoriam*), Maria Francisca Holanda Silva, por seu amor, doçura, cuidados... Para sempre minha segunda mãe. E aos meus avôs, Francisco da Rocha Silva e Raimundo Borges, que muito ensinaram a todos os seus filhos e netos.

Ao meu companheiro, Hugo Santos Piauilino Neto, pelo amor e carinho, e por influenciar minha vida positivamente.

Ao meu orientador, Dr. Rodrigo de Melo Souza Veras, por sua paciência, pelos ensinamentos e pelo exemplo de determinação.

À Universidade Federal do Piauí – UFPI.

A todos os meus estimados professores, que transmitiram valiosos conhecimentos.

E aos meus amigos e demais familiares, por seguirem junto comigo nessa caminhada, dividindo alegrias e sofrimentos.

*“A cada nova manhã nasce junto
uma nova chance”
(Autor Desconhecido)*

Resumo

Melanoma é um tipo maligno de lesão de pele, e atualmente está entre os tipos existentes de câncer mais perigosos. Entretanto, o diagnóstico precoce dessa doença proporciona ao paciente uma maior chance de cura. Neste cenário, métodos computacionais para processamento e análise de imagens de lesão de pele têm sido estudados e desenvolvidos para auxiliar os profissionais da área médica. Esses métodos pretendem possibilitar uma facilidade e rapidez em relação ao diagnóstico da patologia através da interpretação de imagens médicas. Neste trabalho é proposto um método computacional visando auxiliar os médicos dermatologistas no diagnóstico de lesões de pele em melanoma ou não-melanoma, por meio de imagens dermatoscópicas. Com este método pretende-se classificar as lesões de pele utilizando um descritor híbrido. Esse descritor é obtido a partir da combinação da regra ABCD (Assimetria, Borda, Cor e Diâmetro) e Redes Neurais Convolucionais: CaffeNet, Vgg-m, Vgg-verydeep-19, Vgg-f e Vgg-s. Além disso, é realizado nesse descritor híbrido uma seleção de atributos com o algoritmo Razão de Ganho de Informação. As características extraídas das imagens são utilizadas como entradas para o classificador Perceptron de Múltiplas Camadas (MLP). Nos testes, foram utilizadas imagens de lesão de pele de 2 bases de imagens públicas: PH² e DermIS. O resultado obtido na classificação foi uma taxa de acerto igual a 94,9% e um índice Kappa de 0,8916, considerado “Excelente”.

Palavras-chaves: melanoma. descritor híbrido. *multilayer perceptron*.

Abstract

Melanoma is a malignant kind of skin lesion, and is currently among the most dangerous types of cancer present. However, the early diagnosis of this disease gives the patient a greater chance of cure. In this scenario, computational methods for processing and analyzing skin lesion images have been studied and developed to assist medical professionals. These methods aim to make it easier and quicker to diagnose pathology through the interpretation of medical images. This work proposes a computational method to assist dermatologists in the diagnosis of skin lesions in melanoma or non-melanoma by means of dermatoscopic images. With this method we intend to classify the skin lesions using a hybrid descriptor. This descriptor is obtained from the combination of the ABCD rule (Asymmetry, Border, Color and Diameter) and Convolutional Neural Networks: CaffeNet, Vgg-m, Vgg-verydeep-19, Vgg-f and Vgg-s. In addition, a selection of attributes with the Gain Ratio Information algorithm is performed in this hybrid descriptor. The characteristics extracted from the images are used as inputs for the MultiLayer Perceptron (MLP) classifier. In the tests, skin lesion images from 2 public image bases PH² and DermIS were used. The result obtained in the classification was an accuracy equal to 94.9% and a Kappa index of 0.8916, considered “Excellent”.

Keywords: melanoma. hybrid descriptor. multilayer perceptron.

Lista de ilustrações

Figura 1 – Metodologia geral de um sistema CAD.	2
Figura 2 – Melanoma. Lesão de pele de coloração variada, assimétrica e bordas irregulares. Extraído de: Dermatologia.net (2017).	2
Figura 3 – Exemplos de imagens da base PH ² , de acordo com a categoria.	11
Figura 4 – Exemplos de imagens da base DermIS, de acordo com a categoria.	12
Figura 5 – Borda da lesão de pele.	13
Figura 6 – Coloração artificial de uma imagem de lesão de pele.	14
Figura 7 – Orientações usadas para construir as matrizes de co-ocorrências.	15
Figura 8 – Exemplo de cálculo do descritor LBP. (a) Exemplo de janela; (b) Limiarização baseada no pixel central; (c) Padrão calculado a partir do resultado da limiarização. Fonte: Silva (2014).	18
Figura 9 – Arquitetura básica de uma MLP. Fonte: Silva (2014).	21
Figura 10 – Hiperplano de separação ótimo para classes linearmente separáveis.	22
Figura 11 – Passos do classificador <i>Random Forest</i> . Fonte: Silva (2014).	26
Figura 12 – Estrutura do método proposto.	29
Figura 13 – Componentes do modelo de cores HSV.	30
Figura 14 – Segmentação da lesão de pele a partir da máscara binária fornecida pelas bases.	30
Figura 15 – Etapa de extração de características.	31
Figura 16 – Resultados da variação da porcentagem de atributos selecionados para o classificador SVM.	40
Figura 17 – Resultados da variação da porcentagem de atributos selecionados para o classificador <i>Random Forest</i>	41
Figura 18 – Resultados da variação da porcentagem de atributos selecionados para o classificador MLP.	42
Figura 19 – Exemplo de erro de <i>under-segmentation</i> e <i>over-segmentation</i>	44

Lista de tabelas

Tabela 1 – Quadro-resumo dos trabalhos relacionados.	10
Tabela 2 – Nível de precisão da classificação, de acordo com o índice Kappa.	27
Tabela 3 – Composição do descritor híbrido proposto na união das bases de imagens PH ² + DermIS.	32
Tabela 4 – Resultados da classificação com as imagens sem segmentação da união das Bases PH ² e DermIS utilizando os classificadores MLP, SVM, Random Tree e KNN.	35
Tabela 5 – Resultados da classificação com as imagens sem segmentação da união das Bases PH ² e DermIS utilizando comitês.	36
Tabela 6 – Resultados da classificação com as imagens segmentadas da união das Bases PH ² e DermIS utilizando os classificadores MLP, SVM, Random Tree e KNN.	38
Tabela 7 – Resultados da classificação com as imagens segmentadas da união das Bases PH ² e DermIS utilizando comitês.	39
Tabela 8 – Resultados do descritor híbrido com o classificador MLP para as imagens da união das bases PH ² + DermIS com erro de segmentação.	44
Tabela 9 – Comparação do método proposto com o estado da arte.	48
Tabela 10 – Resultados da classificação com as imagens sem segmentação da Base PH ²	59
Tabela 11 – Resultados da classificação com as imagens sem segmentação da Base DermIS.	60
Tabela 12 – Resultados da classificação com as imagens segmentadas da Base PH ²	60
Tabela 13 – Resultados da classificação com as imagens segmentadas da Base DermIS.	61

Lista de abreviaturas e siglas

A	Acurácia
ABCD	Assimetria, Borda, Cor, Diâmetro
AUC	<i>Area Under Curve</i>
BoVW	<i>Bag of Visual Words</i>
CAD	<i>Computer Aided Diagnosis</i>
CNN	<i>Convolutional Neural Network</i>
Dermatlas	<i>Interactive Dermatology Atlas</i>
DermIS	<i>Dermatology Information System</i>
FN	Falso Negativo
FP	Falso Positivo
GLCM	<i>Gray-Level Co-occurrence Matrix</i>
GLN	<i>Gray Level Non-uniformity</i>
GLRLM	<i>Gray-Level Run Length Matrix</i>
HGLRE	<i>High Gray Level Runs Emphasis</i>
HOG	<i>Histograms of Oriented Gradients</i>
HSV	<i>Hue, Saturation, Value</i>
ILSVRC	<i>ImageNet Large Scale Visual Recognition Competition</i>
INCA	Instituto Nacional de Câncer
ISIC	<i>International Skin Imaging Collaboration</i>
K	índice Kappa
KNN	<i>K-Nearest Neighbor</i>
LBP	<i>Local Binary Pattern</i>
LGLRE	<i>Low Gray Level Runs Emphasis</i>

LIPU	<i>Logistic regression using Initial variables and Product Units</i>
LRE	<i>Long Runs Emphasis</i>
LRHGLE	<i>Long Run High Gray-Level Emphasis</i>
LRLGLE	<i>Long Run Low Gray-Level Emphasis</i>
MATLAB	<i>Matrix Laboratory</i>
MLP	<i>MultiLayer Perceptron</i>
MRF	<i>Markov Random Field</i>
P	Precisão
RGB	<i>Red, Green, Blue</i>
RLN	<i>Run Length Non-uniformity</i>
ROC	<i>Receiver Operating Characteristic</i>
RP	<i>Run Percentage</i>
S	Sensibilidade
SCF	<i>Skin Cancer Foundation</i>
SMC	Sistema de Múltiplos Classificadores
SRE	<i>Short Runs Emphasis</i>
SRHGLR	<i>Short Run High Gray-Level Emphasis</i>
SRLGLE	<i>Short Run Low Gray-Level Emphasis</i>
SVM	<i>Support Vector Machine</i>
TDS	<i>Total Dermoscopic Score</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
WHO	<i>World Health Organization</i>

Sumário

1	INTRODUÇÃO	1
1.1	Objetivos	3
1.1.1	Objetivo Geral	3
1.1.2	Objetivos Específicos	3
1.2	Contribuições	4
1.3	Produções Científicas	4
1.4	Organização da Dissertação	4
2	ESTADO DA ARTE	6
2.1	Sem Segmentação	6
2.2	Com Segmentação	7
2.3	Considerações Finais	9
3	MATERIAIS E MÉTODOS	11
3.1	Base de Imagens	11
3.2	Descritores	12
3.3	Classificadores	20
3.4	Seleção de Atributos	25
3.5	Métricas de Avaliação dos Resultados	26
3.6	Método de Avaliação dos Classificadores	28
3.7	Considerações Finais	28
4	MÉTODO PROPOSTO	29
4.1	Pré-processamento	29
4.2	Segmentação	30
4.3	Extração de Características	30
4.4	Classificação	32
5	RESULTADOS E DISCUSSÃO	33
5.1	Resultados com Imagens sem Segmentação	34
5.2	Resultados com Imagens Segmentadas pelos Especialistas	36
5.3	Resultados com Imagens com Erros de Segmentação	43
5.4	Considerações Finais	45
6	CONCLUSÃO E TRABALHOS FUTUROS	49
6.1	Conclusão	49
6.2	Trabalhos Futuros	50

REFERÊNCIAS	51
ANEXOS	58
ANEXO A – TESTES REALIZADOS COM AS BASES DE IMA- GENS PH² E DERMIS SEPARADAS	59

1 Introdução

A computação vem avançando em várias áreas distintas, tais como educação, comércio, segurança e saúde. Ela é um importante meio de auxílio na automatização de processos visando um objetivo desejado e promovendo a simplificação de diversos problemas. Na área da saúde, ela se destaca por conta dos inúmeros softwares criados para auxílio médico, que em sua maioria tem a finalidade de auxiliar na detecção e diagnóstico de doenças.

Os softwares médicos são de grande ajuda pois aliviam em muito os aspectos limitantes do olho e do cérebro humano, como fadiga e distração. Com os avanços da tecnologia da computação, diversos tipos de Diagnósticos Assistidos por Computador (*Computer Aided Diagnosis - CAD*) têm sido desenvolvidos nos últimos anos. Os sistemas CAD ajudam a equipe médica na interpretação de imagens e, conseqüentemente, proporcionam uma segunda opinião (VERAS et al., 2011). Eles são compostos, em geral, por três etapas: segmentação da estrutura de interesse, extração de características da região segmentada e classificação da imagem baseada nas características extraídas.

A Figura 1 mostra um exemplo de uma metodologia geral para a detecção automática do câncer de pele. Temos como entrada uma imagem de lesão de pele, podendo haver um pré-processamento para melhorar essa imagem. Depois disso é feita a segmentação da região de interesse, que nesse caso é somente a lesão de pele, e não a pele saudável. Em seguida, é realizada a extração de características que descreve a imagem em formato apropriado para a posterior classificação em melanoma ou não-melanoma.

Segundo o Instituto Nacional de Câncer, cerca de 180 mil novos casos de câncer de pele são registrados no Brasil. Dentre os tipos de câncer de pele, o melanoma representa menos de dois por cento dos casos, mas possui o pior prognóstico e o mais alto índice de mortalidade (INCA, 2017). De acordo com a *World Health Organization*, em torno de 132 mil novos casos de melanoma ocorrem globalmente a cada ano (WHO, 2017). Nos Estados Unidos, a estimativa feita pelo *Skin Cancer Foundation* é que cerca de 9.730 pessoas morreriam de melanoma em 2017, dos 87.110 novos casos que seriam diagnosticados (SCF, 2017).

Câncer é um grupo de doenças que ocorrem quando as células se tornam anormais dividindo-se e formando mais células, sem controle ou ordem. O câncer de pele é o mais comum de todos os tipos de câncer e o aumento da sua incidência, em parte, está relacionado ao comportamento das pessoas em relação à exposição ao sol. Contudo, é uma doença que tem cura se diagnosticada em seus estágios iniciais. Dentre os tipos de câncer de pele, temos o melanoma (A.C.CAMARGO, 2017).

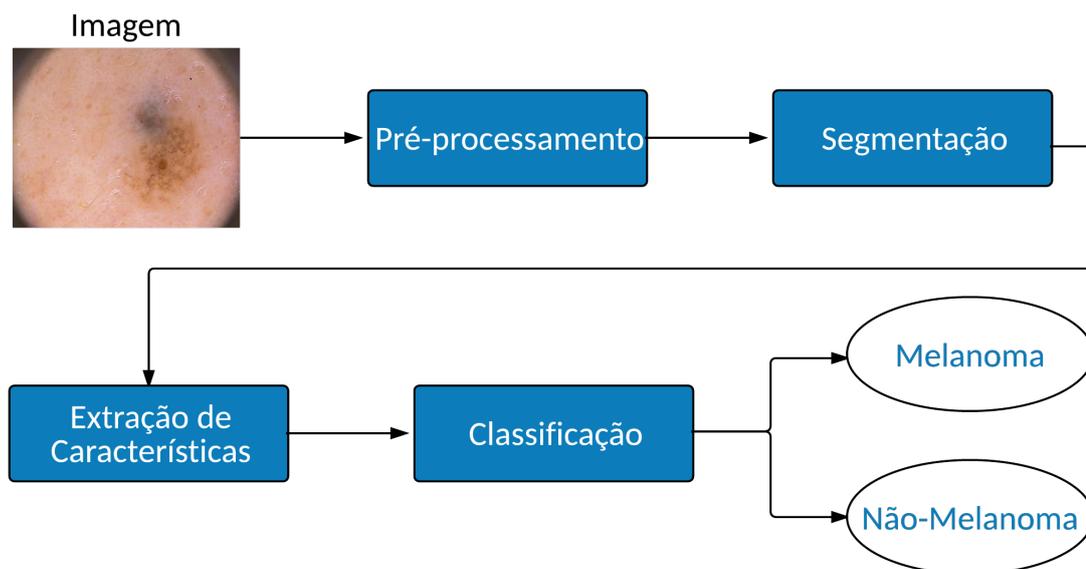


Figura 1 – Metodologia geral de um sistema CAD.

O melanoma é uma lesão maligna, que tem origem nos melanócitos na junção dermo-epidérmica (separa a derme da epiderme). É considerada a mais invasiva e agressiva entre os outros cânceres de pele, pois tende a se espalhar para outras partes do corpo em um processo chamado metástase¹ (A.C.CAMARGO, 2017). A Figura 2 apresenta o exemplo de um melanoma.



Figura 2 – Melanoma. Lesão de pele de coloração variada, assimétrica e bordas irregulares. Extraído de: Dermatologia.net (2017).

Para a detecção desta patologia, inicialmente os especialistas fazem uma inspeção visual das manchas ou feridas da pele, verificando se aparentam com feridas que não cicatrizam, se possuem crescimento assimétrico, se possuem bordas não delimitadas, entre

¹ Quando o câncer se espalha além do local onde começou (sítio primário) para outras partes do corpo. A metástase pode ocorrer quando as células cancerosas viajam através da corrente sanguínea ou dos vasos linfáticos para outras áreas do corpo.

outros aspectos. Entretanto, podem ser realizados alguns exames complementares para verificar seu estágio, como raio-X, biópsia e exames de sangue (DERMATOLOGIA, 2017).

A detecção automática (metodologia computacional) do câncer de pele é feita a partir da análise de imagens obtidas através do exame de dermatoscopia, com o aparelho dermatoscópico, ou de fotografias tiradas utilizando câmeras digitais convencionais. Para a obtenção de um diagnóstico das lesões de pele de forma mais precisa, algumas características visuais podem ser observadas, como assimetria, irregularidade da borda, variação da cor interna e o diâmetro, características essas conhecidas como regra ABCD (BARCELOS; BOAVENTURA; SILVA, 2003). Além dessa regra, outros descritores de cor, forma e textura podem ser utilizados nessas metodologias.

Apesar da alta taxa de mortalidade ligada ao melanoma, o diagnóstico precoce da doença desempenha um papel fundamental na diminuição desses números, já que se o mesmo for detectado durante as etapas iniciais do seu desenvolvimento, a chance de cura é elevada. Assim, visando diminuir a margem de erro dos diagnósticos não invasivos e auxiliar o dermatologista, metodologias para automatizar a análise de lesões de pele são cada vez mais estudadas e desenvolvidas. Isso possibilita a prevenção de lesões benignas se tornarem malignas e a elaboração de tratamentos adequados, com maior índice de cura para lesões malignas.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo principal deste trabalho é o desenvolvimento de um método computacional para a identificação do melanoma em imagens dermatoscópicas, com a finalidade de ser utilizado em um sistema CAD, auxiliando o dermatologista no diagnóstico.

1.1.2 Objetivos Específicos

Os seguintes objetivos específicos são propostos:

- estudo e experimentação das técnicas de visão computacional e aprendizagem de máquina mais recentes para reconhecimento do melanoma em imagens de lesão de pele;
- avaliar descritores e classificadores;
- analisar as informações extraídas com o propósito de mensurar a importância de cada informação para a classificação das imagens;
- analisar o método proposto em imagens de lesão de pele com erros de segmentação.

1.2 Contribuições

A principal contribuição desse trabalho é disponibilizar um método de identificação do melanoma em imagens dermatoscópicas para ser utilizado em um sistema CAD, auxiliando o dermatologista no diagnóstico das lesões de pele. Além de avaliar diferentes descritores e classificadores. O método, em resumo, é composto dos seguintes algoritmos:

- combinação de redes neurais convolucionais e de características de assimetria, borda, cor e diâmetro, conforme a regra ABCD das lesões de pele para a extração de características;
- seleção de atributos por meio do algoritmo Razão de Ganho de Informação (QUINLAN, 1986);
- classificação das lesões de pele por meio do MultiLayer Perceptron (MLP) (HAYKIN, 2009).

1.3 Produções Científicas

Durante o estudo e desenvolvimento do presente trabalho foram publicados os seguintes artigos:

- MOURA, N. H.; VERAS, R. M. S.; AIRES, K. R. T.; SANTOS, L. G. T.; MACHADO, V. P. *Proposta de um Descritor Híbrido para Aprimoramento da Identificação Automática de Melanoma*. Workshop de Informática Médica - WIM, 2017, São Paulo - SP. Anais do XVII Workshop de Informática Médica, 2017.
- MOURA, N. H.; VERAS, R. M. S.; AIRES, K. R. T.; MACHADO, V. P.; SILVA, R. R. V.; ARAÚJO, F.H. D.; CLARO, M. L. *Combining ABCD Rule, Texture Features and Transfer Learning in Automatic Diagnosis of Melanoma*. IEEE Symposium on Computers and Communications, Natal - RN, 2018.

1.4 Organização da Dissertação

O presente trabalho está organizado em 6 (seis) capítulos. Após a introdução (Capítulo 1), que apresentou o problema estudado, os objetivos e as produções científicas obtidas, os próximos capítulos estão organizados da seguinte forma:

Capítulo 2: são apresentados os trabalhos da literatura, fornecendo a sustentação teórica da proposição.

Capítulo 3: são apresentados e descritos os materiais e métodos utilizados nesse trabalho.

Capítulo 4: descreve o método proposto nesta dissertação, mostrando cada etapa de processamento.

Capítulo 5: são descritos os testes realizados e expostos os resultados do trabalho, bem como uma discussão sobre eles.

Capítulo 6: são apresentadas as considerações finais do trabalho realizado e sugestões de trabalhos futuros.

2 Estado da Arte

A fim de obter uma concepção clara do problema proposto pelo presente estudo, fez-se necessária a pesquisa, o estudo e a formulação de ideias sobre o tema. Para tanto, são apresentados a seguir os principais trabalhos que circundam a Combinação da Regra ABCD e CNN's (*Convolutional Neural Networks*) Pré-treinadas no Diagnóstico Automático de Melanoma. Eles servem como base teórica para a sustentação científica e definem o foco com o qual se formularão hipóteses para melhor entender a situação-problema.

Nos últimos anos, diversos trabalhos que propõem métodos automáticos para auxiliar o dermatologista foram desenvolvidos utilizando técnicas de processamento de imagens e visão computacional para detectar e classificar os tipos de lesões de pele. Esses métodos automáticos, em geral, estão divididos em quatro etapas: pré-processamento, segmentação, extração de características e classificação. Entretanto, existem trabalhos na literatura que não utilizam a segmentação das lesões de pele.

A seguir, são apresentados alguns trabalhos relevantes nessa área, divididos em duas seções: sem segmentação e com segmentação.

2.1 Sem Segmentação

No trabalho de [Chang et al. \(2013\)](#), um sistema CAD é proposto com o objetivo de diagnosticar lesões melanocíticas e não melanocíticas baseado em imagens fotográficas digitais convencionais. As características das imagens são extraídas através de descritores de cor, forma e textura em um total de 769 imagens de lesões de pele da base de dados do *Department of Dermatology, Kaohsiung Medical University*. Os algoritmos de textura utilizados foram: *Gray Level Run Length Matrix* (GLRLM) ([GALLOWAY, 1975](#)) e Tamura ([TAMURA; MORI; YAMAWAKI, 1978](#)). A classificação das imagens foi realizada com a utilização do classificador *Support Vector Machine* (SVM) ([CORTES; VAPNIK, 1995](#)). Os resultados mostram que o método proposto alcançou uma acurácia de 90,64% e uma sensibilidade de 85,63%.

A extração de características com base em uma nova abordagem para a detecção do melanoma é mostrada no trabalho de [Fornaciali et al. \(2014\)](#). Os descritores chamados BossaNova são utilizados nesse exame. Eles são uma extensão do modelo *Bag of Visual Words* (BoVW) ([SIVIC; ZISSERMAN, 2003](#)) e são baseados em uma distribuição da distância à palavra-chave. Para obtê-los, são utilizadas duas etapas: codificação e agrupamento. A primeira associa os descritores locais da imagem para o elemento mais próximo da tabela de codificação e a segunda calcula a média desses códigos da imagem e agrupa

toda a informação em um vetor de características. A classificação das imagens é feita com a utilização do classificador SVM. Este método atingiu uma área sob a curva (*Area Under Curve* - AUC) ROC (*Receiver Operating Characteristic*) (HANLEY; MCNEIL, 1982) de 0,937 com a utilização de 747 imagens dermatoscópicas, dentre elas, 187 melanomas e 560 lesões de pele benignas, coletadas do *Department of Medical Informatics, RWTH Aachen University*.

Em Codella et al. (2015), *Deep Learning* (Aprendizado Profundo) (JIA et al., 2014) e *Sparse Coding* (MAIRAL et al., 2014) são utilizados na extração de características de imagens de lesão de pele. Para a classificação, o algoritmo utilizado é o SVM. Os experimentos foram realizados com um total de 2624 imagens da base de dados *International Skin Imaging Collaboration* (ISIC) (HALPERN et al., 2015) (334 melanomas, 144 nevos atípicos² e 2146 lesões benignas) e divididos em duas tarefas: diferenciar o melanoma das lesões não-melanoma e diferenciá-lo somente dos nevos atípicos. A primeira tarefa atingiu uma acurácia de 93,1% e, a segunda, uma acurácia de 73,9%.

2.2 Com Segmentação

Araujo et al. (2012) propuseram um método de extrair as características da imagem baseado na regra ABCD. A definição da rugosidade é feita a partir da borda extraída, bem como a análise do tamanho da mesma e a quantidade de pontos de inflexão (onde as curvas mudam de côncavas para convexas) presentes. Além disso, o diâmetro, a assimetria e a variação de cores são calculados sobre a área lesionada. Na segmentação das lesões de pele, foram utilizadas duas técnicas: divisão e união (*split and merge*), com o intuito de encontrar a lesão e obter o contorno; e modelo de contorno ativo (KASS; WITKIN; TERZOPOULOS, 1988), para aperfeiçoar o contorno encontrado. Este método ainda possui algumas limitações, dentre elas, a sua ineficiência na presença excessiva de pelos e a utilização de apenas 40 imagens para a validação do método proposto. As imagens foram retiradas de 2 bases de imagens diferentes: *Dermatlas* (*Interactive Dermatology Atlas*) (COHEN; LEHMANN, 2012) e Goshtasby (GOSHTASBY, 2012). A acurácia atingida nos testes foi de 92%.

No trabalho de Cavalcanti, Scharcanski e Baranoski (2013) é apresentada uma nova abordagem para a distinção entre melanomas e nevos atípicos. Inicialmente, é realizada uma segmentação das lesões de pele proposta por Cavalcanti e Scharcanski (2011), que utiliza uma representação de imagem multicanal para maximizar a discriminação entre regiões de pele saudáveis e não-saudáveis. Na extração de características, foram utilizadas

² Nevos benignos e não usuais, que podem assemelhar-se ao melanoma.

doze características baseadas na concentração dos dois tipos de melanina³ (eumelanina⁴ e feomelanina⁵), em conjunto com a regra ABCD. Neste caso, a característica que pertence à letra D da regra é alterada para estruturas diferenciais: rede pigmentar, glóbulos e pontos, que caracterizam lesões malignas. Os experimentos foram feitos com um total de 152 imagens (45 nevos atípicos e 107 melanomas) da base de dados Dermnet (JUNIOR, 2009). Para a classificação das imagens, o algoritmo utilizado foi o *K-Nearest Neighbor* (KNN) (FIX; JR, 1951). Os resultados obtidos foram animadores atingindo 98,66% de acurácia, 96,18% de precisão e 99,70% de sensibilidade.

No trabalho proposto por Bhati e Singhal (2015), inicialmente, é realizado um pré-processamento para melhorar as imagens com o método de Otsu (OTSU, 1979), que converte a imagem original em imagem monocromática. O próximo passo é a detecção da borda da lesão, utilizando o detector de Harris Corner (HARRIS; STEPHENS, 1988). A classificação da lesão de pele como benigna ou maligna para detecção do melanoma é feita a partir de uma pontuação obtida pela *Total Dermoscopic Score* - TDS (ou Pontuação de Dermatoscopia Total) que leva em consideração uma pontuação obtida para cada característica A, B, C e D da regra ABCD (descriptor). A base de imagem é formada por 30 imagens (16 malignas e 14 benignas) de lesão de pele obtidas pelo *Cancer Research Center, Binaker* e algumas imagens da internet. Os resultados foram analisados de acordo com a sensibilidade e precisão, alcançando 92,30% e 84,61%, respectivamente.

No trabalho de Kasmi e Mokrani (2016), a segmentação das lesões de pele é feita a partir de duas etapas: a primeira é um pré-processamento que detecta automaticamente os pelos utilizando os filtros de Gabor (GABOR, 1946); a segunda é a detecção da borda da lesão utilizando contornos ativos geodésicos (CASELLES; KIMMEL; SAPIRO, 1997). A regra ABCD foi utilizada para a extração de características de lesões de pele na detecção do melanoma, sendo D, estruturas dermatoscópicas diferenciais. Essas estruturas avaliam a presença de rede pigmentar⁶, áreas de regressão (despigmentação ou *peppering*), estrias radiadas⁷, pontos⁸ e glóbulos⁸. A classificação foi realizada a partir da pontuação obtida pelo TDS. Nos testes, 200 imagens dermatoscópicas foram utilizadas, sendo 80 melanomas e 120 lesões benignas da base de dados *Interactive Atlas of Dermoscopy* (ARGENZIANO et al., 2000). Os testes mostraram que esse método alcançou uma taxa de sensibilidade de

³ É um tipo de proteína responsável por colorir a pele e pelos dos seres humanos, além de proteger o DNA das células contra a radiação ultravioleta emitida pelo sol.

⁴ Tipo de melanina que apresenta uma coloração que varia do negro ao marrom, além de possuir um alto peso molecular e capacidade de dispersar a luz ultravioleta.

⁵ Tipo de melanina que apresenta coloração que varia do vermelho ao amarelo.

⁶ É caracterizada por linhas marrons que formam uma rede sobre um fundo mais claro. É formada pela presença de pigmento na camada basal, a camada mais profunda da epiderme que faz contato direto com a derme.

⁷ São faixas radiadas e paralelas na periferia da lesão, frequentemente vistas no melanoma (distribuídas irregularmente).

⁸ São estruturas arredondadas, de tamanho variável e coloração preta, marrom ou cinza azulada, de acordo com a profundidade. Por definição, os pontos são $< 0,1\text{mm}$ e os glóbulos são $> 0,1\text{mm}$.

91,25%, precisão de 95,83% e acurácia de 94,0%.

Sánchez-Monedero et al. (2016) propuseram um sistema que combina regra ABCD, *Gray-Level Co-occurrence Matrix* (GLCM) (HARALICK; SHANMUGAM; DINSTEIN, 1973), *Local Binary Pattern* (LBP) (OJALA; PIETIKÄINEN; HARWOOD, 1996) e *Markov Random Field* (MRF) (LI, 2009) para a extração de características das lesões de pele. As imagens foram segmentadas utilizando um algoritmo de segmentação automática proposto por Saez, Serrano e Acha (2014), em que uma técnica de conjunto de níveis baseada em bordas é aplicada juntamente com um gradiente de cor adaptado (SÁEZ et al., 2013). Nesse trabalho, a base de dados é composta por 556 imagens obtidas pelo *Interactive Atlas of Dermoscopy* (ARGENZIANO et al., 2000). Vários métodos de classificação foram utilizados para os experimentos. O melhor resultado encontrado foi uma acurácia de 68,51% com o método de classificação *Logistic regression using Initial variables and Product Units* (LIPU) (GUTIÉRREZ; HERVÁS-MARTÍNEZ; MARTÍNEZ-ESTUDILLO, 2011).

Oliveira et al. (2016) apresentaram uma nova abordagem para extração de características de lesão de pele. A regra ABCD e o método *Box-Counting* (AL-AKAIDI, 2004) são utilizados como descritores das imagens. Entretanto, o diâmetro da regra ABCD não foi aplicado devido à sua grande dependência da resolução da imagem. Na segmentação das lesões de pele foi utilizado o modelo de Chan-Vese (CHAN; VESE, 2001), um modelo de contorno ativo sem bordas. As bases de imagens utilizadas possuem um total de 408 imagens coletadas de *Loyola University Chicago* (MELTON; SWANSON, 2012), *YSP Dermatology Image Database* (SUZUMURA, 2012), *DermAtlas* (COHEN; LEHMANN, 2012), *DermIS* (DIEPGEN; YIHUNE, 2012), *Saúde Total* (TOTAL, 2012), *Skin Cancer Guide* (GUIDE, 2012) e *Dermnet* (JUNIOR, 2009). Os experimentos foram divididos em três: classificação entre imagens de nevos e ceratose seborreica⁹, entre nevos e melanoma, e entre ceratose seborreica e melanoma. Neste último, o resultado da classificação não foi tão expressivo com uma acurácia de 74,33%, utilizando o SVM. O melhor resultado obtido foi uma acurácia de 79,01%, porém, para a diferenciação entre nevos e ceratose seborreica, o que não torna tão eficiente o método.

2.3 Considerações Finais

Neste capítulo são apresentados alguns trabalhos relacionados com a presente pesquisa, sendo estes divididos em: sem segmentação e com segmentação das imagens. A Tabela 1 apresenta um quadro-resumo onde é possível observar que diferentes descritores e classificadores são utilizados na detecção automática de câncer de pele. Dentre os descritores, a regra ABCD, comumente aplicada para a extração de características, permite fácil entendimento e fornece simplicidade de aplicação, ao mesmo tempo em que mostra

⁹ Lesão benigna da pele, geralmente arredondada ou irregular, de coloração acastanhada, amarronzada ou negra, e de aspecto verrucoso.

Tabela 1 – Quadro-resumo dos trabalhos relacionados.

Sem segmentação				
Trabalhos	Descritores	Classificadores	Número de Imagens	Métricas de Avaliação
Chang et al. (2013)	GLRLM e Tamura	SVM	729	A = 90,64%
Fornaciali et al. (2014)	BossaNova	SVM	747	AUC = 0,937
Codella et al. (2015)	<i>Deep Learning</i> e <i>Sparse Coding</i>	SVM	2624	A = 93,1%
Com segmentação				
Trabalhos	Descritores	Classificadores	Número de Imagens	Métricas de Avaliação
Araujo et al. (2012)	ABCD	-	40	A = 92,0%
Cavalcanti, Scharcanski e Baranoski (2013)	ABCD	KNN	152	A = 98,66%
Bhati e Singhal (2015)	ABCD	TDS	30	P = 84,61% S = 92,3%
Kasmi e Mokrani (2016)	ABCD	TDS	200	A = 94,0%
Sánchez-Monedero et al. (2016)	ABCD, GLCM, LBP e MRF	LIPU	556	A = 68,51%
Oliveira et al. (2016)	ABCD e <i>Box-Counting</i>	SVM	408	A = 74,33%

resultados confiáveis para detecção do melanoma. Essa regra pode ser calculada de diversas maneiras, como pode ser visto nos trabalhos apresentados.

Além das características ABCD, as redes neurais convolucionais e principalmente os descritores de textura também são muito utilizados para analisar computacionalmente lesões de pele a partir de imagens. Isso porque as características de textura das lesões de pele, como sua suavidade ou rugosidade e regularidade ou irregularidade podem ser utilizadas para identificar os padrões de uma determinada lesão.

3 Materiais e Métodos

Neste capítulo são apresentados os materiais e métodos necessários para a execução da pesquisa neste trabalho. Desse modo, são apresentadas as bases de imagens de lesões de pele utilizadas para os testes, o algoritmo utilizado na seleção de atributos, as métricas de avaliação dos resultados e o método de avaliação dos classificadores.

Além disso, são detalhados os descritores e os classificadores utilizados nos testes necessários para o desenvolvimento do método proposto. Os descritores são: regra ABCD, GLCM, GLRLM, HOG, LBP, Tamura, *Box-Counting* e CNN's. E os classificadores são: MLP, SVM, *Random Tree*, KNN, comitê (MLP + SVM + KNN) e *Random Forest*.

3.1 Base de Imagens

Nesse trabalho, foram utilizadas duas bases de imagens para os testes dos algoritmos: PH² (MENDONÇA et al., 2013) e DermIS (*Dermatology Information System*) (DIEPGEN; YIHUNE, 2012). A PH² é constituída por um total de 200 imagens dermatoscópicas: 80 nevos (manchas) comuns, 80 nevos atípicos e 40 melanomas (tumor maligno). As imagens estão no sistema de cores RGB (*Red, Green, Blue*) com uma resolução de 768 x 560 pixels. A Figura 3 apresenta exemplos de imagens de lesões de pele da base PH², distribuídas nas categorias benigna e melanoma.

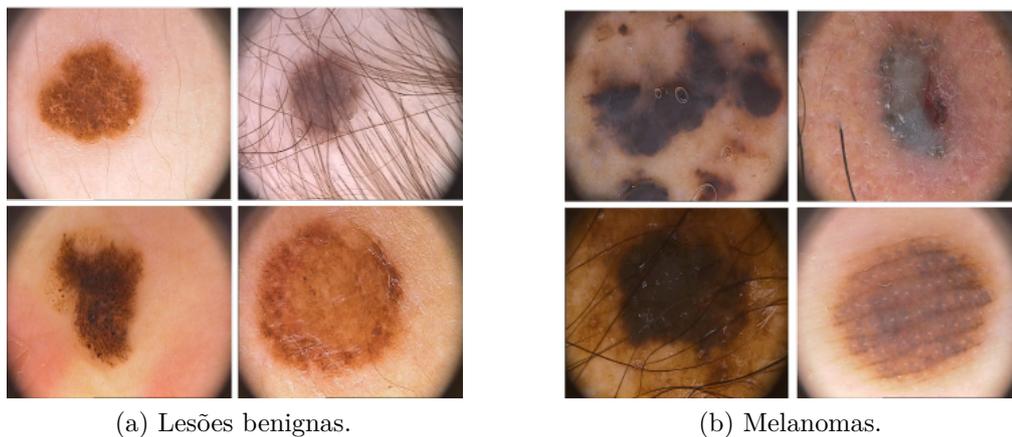


Figura 3 – Exemplos de imagens da base PH², de acordo com a categoria.

Já a base DermIS possui um total de 206 imagens dermatoscópicas, divididas em melanoma e benigna, 119 e 87, respectivamente. Nela, as imagens também estão no sistema de cores RGB. Essa base apresenta imagens de lesões de pele que são mais difíceis para a extração de características e classificação do que a base PH². Isso porque a semelhança entre as lesões benignas e os melanomas é maior, resultando numa maior complexidade

para a distinção entre as duas classes. A Figura 4 apresenta exemplos de imagens de lesões de pele das categorias benigna e melanoma da base DermIS.

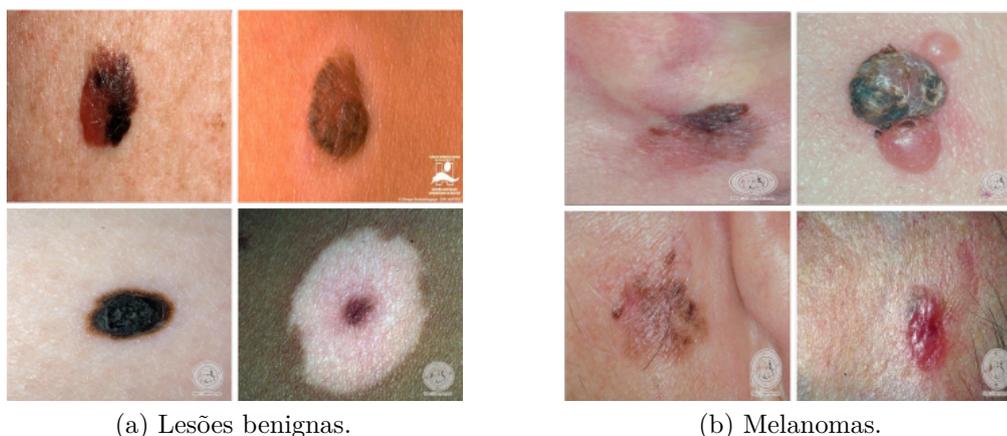


Figura 4 – Exemplos de imagens da base DermIS, de acordo com a categoria.

As imagens foram avaliadas e diagnosticadas por um dermatologista, apresentando a máscara binária da lesão segmentada, aspecto esse, fundamental pela escolha dessas bases.

3.2 Descritores

Atributos são propriedades que podem ser mensuradas a partir de uma imagem, como por exemplo forma, cor e textura. Esses atributos, agrupados em um vetor de características, são denominados de descritores de imagem. Neste trabalho foram utilizados alguns descritores para extrair atributos das imagens de lesão de pele, que são detalhados a seguir.

Regra ABCD

Essa regra é dividida em quatro partes para descrever a imagem e retorna 5 atributos: Assimetria (1 atributo), Irregularidade da Borda (2 atributos), Cor (1 atributo) e Diâmetro (1 atributo). A primeira característica calculada é o Diâmetro. Nessa etapa, a máscara binária é utilizada para obter a borda da lesão com a utilização do algoritmo de Sobel (1970). O resultado é mostrado na Figura 5. Assim, com a borda encontrada, o diâmetro é calculado. Ele consiste na maior distância entre dois pontos, $(p_1 = (x_1, y_1))$ e $p_2 = (x_2, y_2)$ pertencentes à borda de uma imagem, constituindo a diagonal maior. O cálculo dessa distância é feito com a utilização da equação da distância Euclidiana (Equação 3.1).

Lesões malignas tendem a possuir um diâmetro maior ou igual a 6mm com a possibilidade de variação de tamanho e forma. Entretanto, as não malignas normalmente



Figura 5 – Borda da lesão de pele.

possuem um diâmetro menor que 6mm e não mudam de tamanho. Assim, o diâmetro é um importante atributo a ser calculado. Para a classificação, a distância é utilizada em *pixels*.

$$D_{(p_1,p_2)} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (3.1)$$

A Assimetria é encontrada a partir do Diâmetro, onde a maior distância, em *pixels*, entre a reta (diagonal maior) e um ponto pertencente à borda é obtida pela [Equação 3.2](#), onde a , b e c são os coeficientes da reta e (x_0, y_0) os pontos da coordenada da imagem. Lesões que tendem a ser mais arredondadas são geralmente não malignas. Assim, a assimetria de uma lesão diz respeito à sua forma.

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}, \quad (3.2)$$

Lesões não malignas geralmente possuem bordas lisas e suaves, e já as malignas, bordas irregulares. A irregularidade da borda é calculada a partir de duas técnicas. Na primeira, pesos são atribuídos para os quatro vizinhos da direita e da esquerda de todos os pontos da borda da lesão. O vizinho que estiver acima do ponto analisado em relação à coordenada y , recebe peso -1, caso contrário, recebe peso 1. Se a soma desses pesos for maior ou igual a 2 ou menor ou igual a -2, para ambos os lados, considera-se uma inflexão. Além disso, se a soma dos pesos for positiva essa inflexão é considerada um pico, mas se negativa, é considerada um vale. Assim, cada ponto de inflexão está associado a uma oscilação da borda. O objetivo é descobrir a quantidade total de picos e vales menores na lesão ([ARAUJO et al., 2012](#)).

Já na segunda técnica, encontra-se os picos e vales maiores da lesão. O produto vetorial ([Equação 3.3](#)) é empregado para todos os pontos da borda da lesão, considerando os seus vizinhos da direita e esquerda $(p - 15; p; p + 15)$, sendo p o ponto, ou seja, variação de 15 *pixels* à esquerda e 15 *pixels* à direita do ponto p . Assim, caso o resultado do produto

vetorial seja maior que zero, $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ e $p_3 = (x_3, y_3)$ constituem um pico e, se menor que zero, constituem um vale (ARAÚJO et al., 2012).

$$p = (x_2 - x_1) \cdot (y_3 - y_1) - (y_2 - y_1) \cdot (x_3 - x_1). \quad (3.3)$$

Na característica Cor, a quantidade de cores encontradas na lesão é utilizada na classificação. Inicialmente, é realizada uma divisão da imagem em dez intervalos (cores pertencentes ao mesmo foram consideradas semelhantes). Para cada intervalo, é feita a contagem do número de *pixels*. Entretanto, as regiões muito pequenas (menos de 100 *pixels*) foram desconsideradas. Geralmente, lesões com cor interna homogênea são não malignas. Já os melanomas, possuem grande variação das intensidades marrom e preta, e podem conter regiões isoladas em branco. (ARAÚJO et al., 2012). A Figura 6 mostra um exemplo do resultado do cálculo da quantidade de cores presentes na lesão de pele, onde foram encontradas sete cores diferentes na lesão.

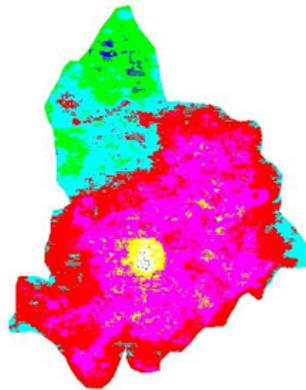


Figura 6 – Coloração artificial de uma imagem de lesão de pele.

Gray-Level Co-occurrence Matrix (GLCM)

A GLCM ou Matrizes de Co-ocorrências de Níveis de Cinza, é uma técnica que tem como base a análise de textura em imagens. Nela são analisadas as co-ocorrências existentes entre pares de *pixels* através de algum padrão. Uma GLCM é sempre uma matriz quadrada, onde as informações das intensidades relativas dos *pixels* são armazenadas.

As probabilidades de co-ocorrências são calculadas entre dois níveis de cinza i e j , utilizando uma orientação θ (theta) e uma distância conhecida como *pixel pair spacing*. Essa orientação pode assumir os valores de 0° , 45° , 90° ou 135° , conforme a Figura 7 (BARALDI; PARMIGGIANI, 1995). Para cada relacionamento espacial possível (distância e orientação) existe uma matriz de co-ocorrência. Desse modo, todas as informações sobre a textura de uma imagem estarão contidas nessa matriz.

Haralick, Shanmugam e Dinstein (1973) definiram 14 características significativas para a GLCM, porém foram utilizadas as mais importantes: contraste (Equação 3.4), correlação (Equação 3.5), energia (Equação 3.6), homogeneidade (Equação 3.7) e entropia (Equação 3.8). Uma vez que a utilização de algumas dessas características pode gerar um melhor resultado do que a utilização de todas (SILVA, 2013). Assim, são retornados 17 atributos. Onde G a quantidade de níveis de cinza na imagem. C_{ij} é a entrada da matriz GLCM na linha i e coluna j . Note também que (μ_x, μ_y) e (σ_x, σ_y) representam média e desvio padrão, respectivamente, para a linha i e coluna j da matriz GLCM.

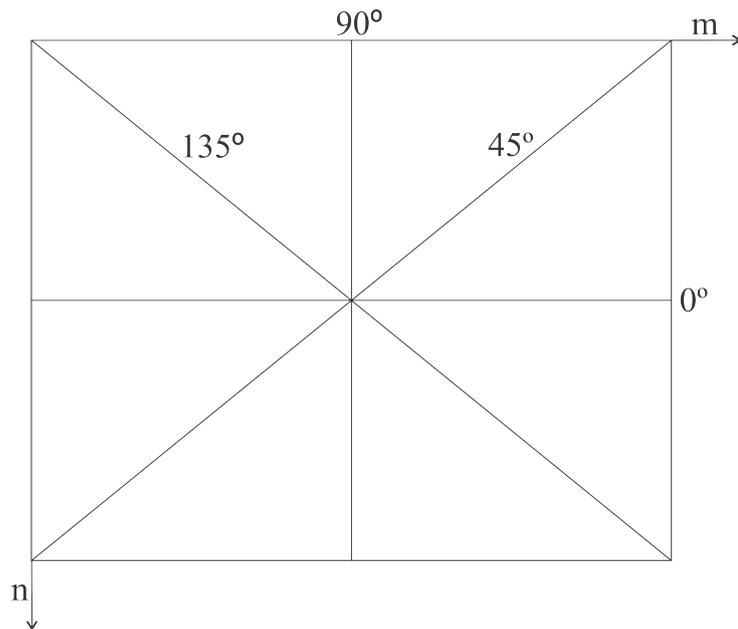


Figura 7 – Orientações usadas para construir as matrizes de co-ocorrências.

$$\text{Contraste} = \sum_{i,j=1}^G C_{ij}(i-j)^2. \quad (3.4)$$

$$\text{Correlação} = \sum_{i,j=1}^G \frac{(i-\mu_x)(j-\mu_y)C_{ij}}{(\sigma_x)(\sigma_y)}. \quad (3.5)$$

$$\text{Energia} = \sum_{i,j=1}^G (C_{ij})^2. \quad (3.6)$$

$$\text{Homogeneidade} = \sum_{i,j=1}^G \frac{C_{ij}}{1+|i-j|}. \quad (3.7)$$

$$\text{Entropia} = - \sum_{i,j=1}^G C_{ij} \log C_{ij}. \quad (3.8)$$

O contraste representa uma medida da intensidade de contraste de um *pixel* e seu vizinho sobre toda a imagem. A correlação representa o quão correlacionado um *pixel* é

com o seu vizinho sobre toda a imagem. A energia apresenta a quantidade de variação na intensidade em uma dada região da imagem. A homogeneidade representa a proximidade da distribuição dos elementos na GLCM com a sua diagonal. E por fim, a entropia indica o quão normal é a distribuição dos níveis de cinza na imagem.

Gray Level Run Length Matrix (GLRLM)

A GLRLM ou Método de Comprimento de Primitiva a Nível de Cinza, também analisa a textura em imagens. Ela utiliza matrizes auxiliares e verifica as repetições dos *pixels* (primitivas). Uma matriz GLRLM é criada para cada corte e possui o número de primitivas dos níveis de cinza na direção θ (0° , 45° , 90° ou 135°) e a uma distância d . Assim, esse descritor retorna 44 características. A soma de todas as matrizes GLRLM de cada corte é feita, e a partir do número de repetições encontradas, a matriz é dividida e normalizada.

Segundo Galloway (1975), a partir da matriz do GLRLM cinco medidas estatísticas podem ser extraídas das imagens. São elas, a ênfase em primitivas curtas (*Short Runs Emphasis* - SRE) (Equação 3.9), a ênfase em primitivas longas (*Long Runs Emphasis* - LRE) (Equação 3.10), a não-uniformidade do nível de cinza (*Gray Level Non-uniformity* - GLN) (Equação 3.11), a não-uniformidade do comprimento da primitiva (*Run Length Non-uniformity* - RLN) (Equação 3.12) e a percentagem de primitivas (*Run Percentage* - RP) (Equação 3.13). Entretanto, mais duas medidas estatísticas foram determinadas Chu, Sehgal e Greenleaf (1990), a ênfase em primitivas de baixo nível de cinza (*Low Gray Level Runs Emphasis* - LGLRE) (Equação 3.14) e a ênfase em primitivas de alto nível de cinza (*High Gray Level Runs Emphasis* - HGLRE) (Equação 3.15).

Outras quatro funções de extração de características seguindo a ideia das medidas estatísticas foram descritas por Dasarathyand e Holder (1991). São elas: ênfase em primitivas curtas de baixo nível de cinza (*Short Run Low Gray-Level Emphasis* - SRLGLE) (Equação 3.16), ênfase em primitivas curtas de alto nível de cinza (*Short Run High Gray-Level Emphasis* - SRHGLE) (Equação 3.17), ênfase em primitivas longas de baixo nível de cinza (*Long Run Low Gray-Level Emphasis* - LRLGLE) (Equação 3.18) e ênfase em primitivas longas de alto nível de cinza (*Long Run High Gray-Level Emphasis* - LRHGLE) (Equação 3.19).

$$SRE = \frac{\sum_{i=1}^G \sum_{j=1}^R \frac{p(i, j|\theta)}{j^2}}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)}. \quad (3.9)$$

$$LRE = \frac{\sum_{i=1}^G \sum_{j=1}^R j^2 p(i, j|\theta)}{\sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta)}. \quad (3.10)$$

$$GLN = \sum_{i=1}^G \left(\sum_{j=1}^R p(i, j|\theta) \right)^2 / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.11)$$

$$RLN = \sum_{j=1}^R \left(\sum_{i=1}^G p(i, j|\theta) \right)^2 / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.12)$$

$$RP = \frac{1}{N} \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.13)$$

$$LGLRE = \sum_{i=1}^G \sum_{j=1}^R \frac{p(i, j|\theta)}{i^2} / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.14)$$

$$HGLRE = \sum_{i=1}^G \sum_{j=1}^R i^2 p(i, j|\theta) / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.15)$$

$$SRLGLE = \sum_{i=1}^G \sum_{j=1}^R \frac{p(i, j|\theta)}{i^2 \cdot j^2} / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.16)$$

$$SRHGLE = \sum_{i=1}^G \sum_{j=1}^R \frac{i^2 \cdot p(i, j|\theta)}{j^2} / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.17)$$

$$LRLGLE = \sum_{i=1}^G \sum_{j=1}^R \frac{j^2 \cdot p(i, j|\theta)}{i^2} / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.18)$$

$$LRHGLE = \sum_{i=1}^G \sum_{j=1}^R i^2 \cdot j^2 \cdot p(i, j|\theta) / \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta). \quad (3.19)$$

Onde G é o número de níveis de cinza. R é o número de diferentes comprimentos de execução que ocorre e N é o número de *pixels* na imagem. A GLRLM é uma matriz dimensional de $(G \times R)$ elementos em que cada elemento $p(i, j|\theta)$ retorna o número total de ocorrências de execuções de comprimento j do nível de cinza i , em uma dada direção θ .

Histograms of Oriented Gradients (HOG)

O Histograma de Gradientes Orientados (DALAL; TRIGGS, 2005), calcula o histograma da orientação dos gradientes na imagem. Ele se baseia na ideia que a forma e a aparência de um objeto podem ser descritas muitas vezes pela intensidade dos gradientes ou pela direção das bordas, mesmo sem um conhecimento prévio da posição de tais bordas.

O HOG consiste em converter a imagem para escala de cinzas e normalizá-la de acordo com a iluminação. Além disso, é feita a computação dos gradientes e uma divisão da imagem em pequenas regiões espaciais chamadas células. Para cada célula é calculado

um histograma local *1-D* das orientações sobre os *pixels* da célula. As células podem ter forma retangular ou radial. Após os histogramas serem computados, eles são normalizados através de um acúmulo de histogramas locais em regiões espaciais um pouco maiores chamadas de blocos.

Os histogramas acumulados são então utilizados para normalizar todas as células naquele bloco. Após a normalização, uma janela de detecção é coletada sobre os histogramas gerados, que consiste na saída do descritor HOG com 81 atributos.

Local Binary Pattern (LBP)

O descritor *Local Binary Pattern* (LBP), ou Padrão Binário Local, apresenta um bom desempenho em várias aplicações como por exemplo, na classificação e na segmentação de textura, e inspeção de superfície (SURULIANDI; MEENA; ROSE, 2012) (SILVA; AIRES; VERAS, 2014).

No descritor LBP original (OJALA; PIETIKÄINEN; HARWOOD, 1996), os *pixels* de uma imagem em níveis de cinza por limiarização são rotulados. A Figura 8 mostra o funcionamento do descritor. O *pixel* central da janela de vizinhança 3 x 3 constitui o valor do limiar, onde cada *pixel* é comparado com o central. O resultado final, considerado um número binário, é feito para todas as janelas escolhidas da imagem. Além disso, quando todas as janelas forem computadas, o vetor de atributos é montado com a junção dos valores obtidos. Assim, com a limiarização feita, um histograma da célula é calculado, ou seja, as quantidades de “0” e “1” são computadas.

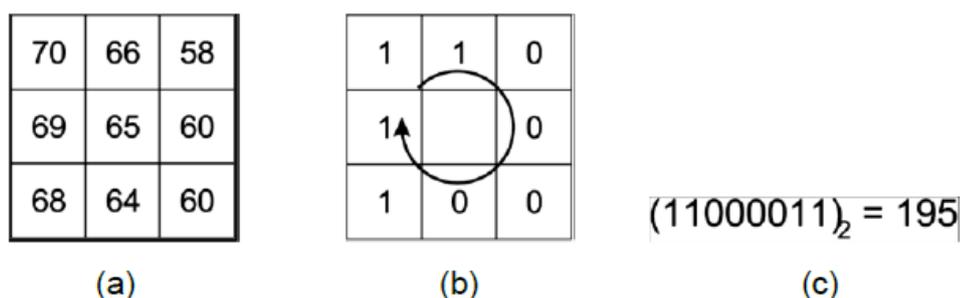


Figura 8 – Exemplo de cálculo do descritor LBP. (a) Exemplo de janela; (b) Limiarização baseada no *pixel* central; (c) Padrão calculado a partir do resultado da limiarização. Fonte: Silva (2014).

Tamura

Esse descritor de textura calcula as seguintes três características: granularidade (*coarseness*), contraste (*contrast*) e direção (*orientation*) (TAMURA; MORI; YAMAWAKI, 1978). Essas características estão relacionadas à percepção visual humana (ABBADENI; ZIOU; WANG, 2000).

A característica *coarseness* é a propriedade mais fundamental. Ela indica a densidade da textura. Quanto menores e mais juntos estão os elementos da textura, mais densa ela é. Quanto maiores ou mais espaçados os elementos da textura, menos densa, ou mais esparsa ela é.

O contraste mede o grau de clareza com o qual se pode distinguir entre diferentes áreas texturizadas em uma imagem. Uma imagem é bem contrastada se suas diferentes áreas são claramente visíveis.

A direção é uma propriedade global em uma área de imagem que busca medir o grau de direcionalidade e não a direção específica. Ela indica se uma textura é direcional ou não. Esse conceito está relacionado tanto à forma do elemento quanto à distribuição dos elementos na textura. Duas texturas iguais que diferem por uma rotação devem apresentar o mesmo grau de direcionalidade.

Box-Counting

O método *Box-Counting* também é um descritor de textura, que calcula as dimensões fractais sob uma imagem, retornando 11 atributos. O fractal, abordado primeiramente por Mandelbrot, representa um objeto geométrico que pode ser dividido em várias partes, onde cada parte é similar ao objeto maior (AL-AKAIDI, 2004). A dimensão fractal (D) é uma medida que quantifica o nível de irregularidade ou auto-similaridade dos fractais (OLIVEIRA, 2012).

A dimensão fractal é dada pela Equação 3.20. Onde N representa a quantidade de elementos das partes auto-similares que reconstrói a imagem original e T representa a quantidade de quadrantes correspondente a uma fração do tamanho anterior.

$$D = \frac{\log(N)}{\log(1/T)}. \quad (3.20)$$

Existem diversas técnicas para estimar a dimensão fractal. O método *Box-Counting*, desenvolvido por Al-Akaidi (2004), foi utilizado nesse trabalho. Ele é um dos algoritmos mais populares para calcular a dimensão fractal de sinais e imagens.

Esse método caracteriza-se pela definição de uma grade com n caixas sobre a imagem, para manter o sinal em um quadrado. E então é feita uma contagem da quantidade de caixas que cobrem o fractal para representar a dimensão. Esse processo é iterativo, ao qual o tamanho das caixas vai se reduzindo, conforme cada iteração.

Convolutional Neural Network (CNN)

A CNN ou Rede Neural Convolutiva possui a técnica de transferência de aprendizado proposta por Shin, Roth et al. (2016). Ela é inicialmente treinada em uma grande

base natural de imagens fazendo com que assimile características genéricas, facilitando a sua aplicabilidade em pequenas bases de dados.

Duas formas de transferências são apresentadas no trabalho de [Castelluccio et al. \(2015\)](#), a primeira consiste no *fine-tune* da rede, onde sua estrutura é modificada, congelando-se as camadas de alto nível. A segunda, empregada nesse trabalho, é utilizando a penúltima camada da rede como matriz de características, contendo 4096 atributos por instância.

As CNN's utilizadas são pré-treinadas na base natural de imagens *ImageNet* e estão disponíveis no *website MatConvNet*. Sete arquiteturas propostas recentemente foram utilizadas na construção da abordagem:

- *AlexNet*: é uma arquitetura desenvolvida por [Krizhevsky, Sutskever e Hinton \(2012\)](#) e foi utilizada na *ImageNet Large Scale Visual Recognition Competition (ILSVRC)* de 2010 para treinar e classificar o banco de dados *ImageNet*. Compreende oito camadas que precisam ser treinadas, cinco camadas convolucionais com filtros de tamanhos 5×5 e 7×7 , seguidas de três camadas totalmente conectadas, bem como camadas de agrupamento máximo.
- *CaffeNet*: esta arquitetura foi desenvolvida no Centro de Visão e Aprendizado de Bekeley e é considerada uma das CNNs mais populares em aprendizado profundo ([JIA et al., 2014](#)). Compreende cinco camadas convolucionais, cada uma seguida por uma camada de *pooling* e três camadas totalmente conectadas. A principal diferença entre essa arquitetura e a AlexNet é a ordem da camada de agrupamento e a camada de normalização.
- *Vgg-f*, *Vgg-m* e *Vgg-s*: propostas por [Chatfield et al. \(2014\)](#), essas arquiteturas foram projetadas para aumentar a profundidade das redes; contém oito camadas de aprendizado, sendo cinco convolucionais e três totalmente conectadas. A *Vgg-m* é mais larga em relação a *Vgg-s* (que possui menos filtros em uma das camadas de convolução).
- *Vgg-verydeep-16* e *Vgg-verydeep-19*: propostas por [Simonyan e Zisserman \(2014\)](#), foram descobertas no *ImageNet Challenge 2014*. Elas são redes de maior profundidade que utilizam uma arquitetura com filtros de convolução muito pequenos (3×3).

3.3 Classificadores

Um classificador tem por objetivo dividir o espaço de características em regiões de decisão. A entrada dos classificadores é o vetor de atributos resultante dos descritores,

e a saída é a classe que a região pertence. Os classificadores que foram utilizados nesse trabalho são apresentados a seguir.

MultiLayer Perceptron (MLP)

O *MultiLayer Perceptron* (Perceptron de Múltiplas Camadas) é uma rede neural artificial utilizada para resolver problemas mais complexos, que não poderiam ser resolvidos pelo modelo de neurônio básico. Ele possui uma rede de perceptrons dispostos em camadas. Na camada oculta, os neurônios são denominados ocultos porque eles não têm acesso direto à saída da rede, onde os erros de aproximação são calculados (HAYKIN, 2009).

Na MLP, o método mais popular para aprendizado é chamado de retro-propagação (*backpropagation*). A aprendizagem por retro-propagação do erro consiste em dois passos através das diferentes camadas da rede: propagação e retro-propagação. No primeiro, um padrão de atividades é aplicado aos nós sensoriais e a resposta real da rede é produzida. No segundo, os pesos sinápticos são ajustados de acordo com um sinal de erro para ser propagado na direção contrária das conexões sinápticas. Isso porque o objetivo é minimizar o erro médio global, ou seja, fazer com que a resposta real da rede se aproxime da resposta desejada.

A Figura 9 mostra a arquitetura básica de uma MLP, onde um sistema de neurônios ligados por conexões sinápticas e dividido em neurônios de entrada, que recebem estímulos do meio externo, neurônios internos (ocultos) e neurônios de saída, que se comunicam com o exterior.

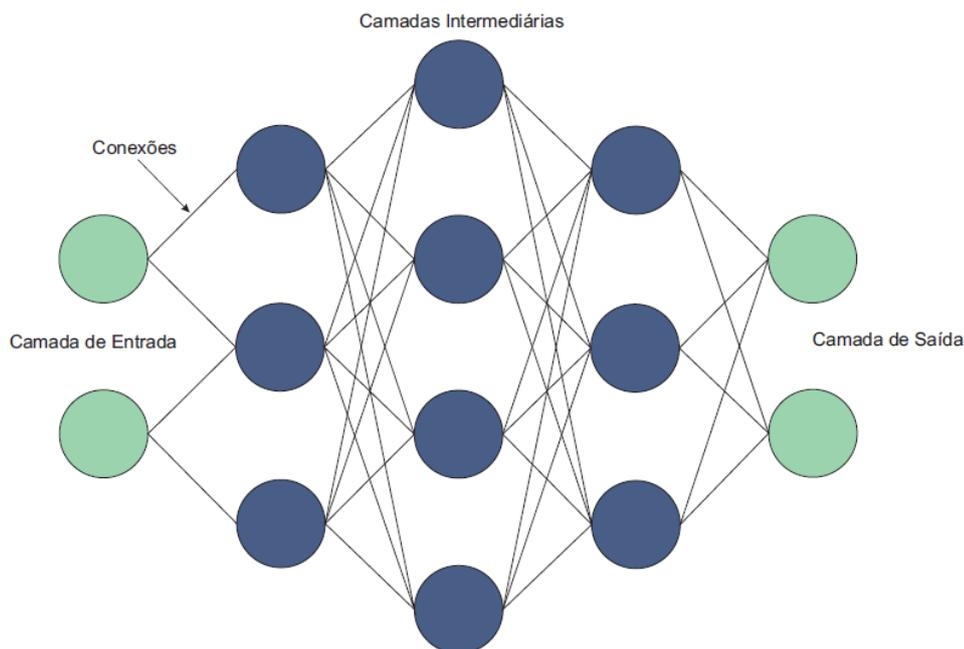


Figura 9 – Arquitetura básica de uma MLP. Fonte: Silva (2014).

Nesse trabalho, a MLP é utilizada com três camadas (uma camada de entrada, uma camada oculta e uma camada de saída). A camada oculta é composta por duas unidades de processamento simples, às quais correspondem os neurônios. A função de ativação utilizada é a sigmóide aproximada e a função de perda (critério para otimização das redes) é o erro quadrático. Foram testadas outras quantidades de camadas ocultas e também um modelo sem essa camada. Os resultados obtidos nas outras configurações não foram melhores que com uma camada. Outros valores para a quantidade de neurônios foram testados também.

Support Vector Machine (SVM)

A *Support Vector Machine* (Máquina de Vetor de Suporte), assim como outros classificadores, é uma técnica de aprendizagem supervisionada, que contém as fases de treinamento e teste; no treinamento, atributos e saídas desejadas são dadas a fim de projetar os vetores de suporte utilizados para obter um classificador; no teste, o classificador é utilizado para encontrar uma saída a partir de um vetor de entrada. Em geral, é feito um mapeamento do espaço de entrada para um espaço de dimensionalidade maior. Em seguida, é calculado um hiperplano de separação ótimo a fim de maximizar a distância de separação entre as classes.

A [Figura 10](#) mostra duas classes (azuis e vermelhos) linearmente separáveis e seu hiperplano de separação ótimo. É possível observar que existem três hiperplanos possíveis: o verde, o amarelo e o preto. Contudo, tanto o hiperplano verde, quanto o amarelo passa próximo de algum ponto. Neste caso, o hiperplano ótimo é o preto, pois maximiza a distância entre os pontos. A margem é representada pelas linhas pontilhadas.

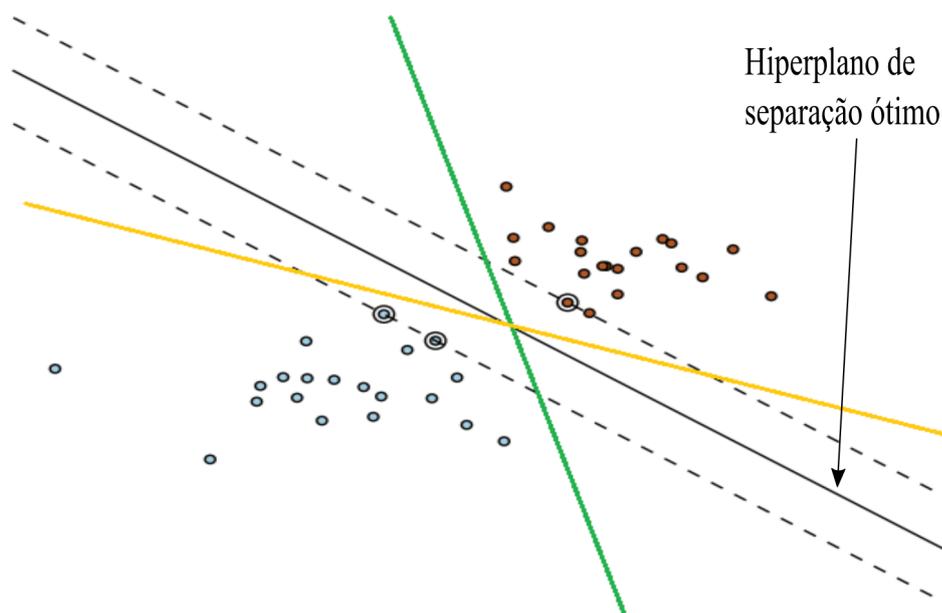


Figura 10 – Hiperplano de separação ótimo para classes linearmente separáveis.

É necessário que a dimensionalidade do problema seja aumentada ao se utilizar SVM para reconhecimento de padrões, tornando uma função não-linearmente separável em uma função linearmente separável (COVER, 1965). Para isso, são aplicadas as funções de Kernel nos vetores de entrada. Em um vetor de entrada de dimensão N , por exemplo, ao aplicar a função de Kernel, é obtido um novo vetor de dimensão X maior que N . Após isso, são obtidos os vetores de suporte, tornando possível definir o hiperplano de separação ótimo. Nesse trabalho, a função de Kernel utilizada foi a linear.

Random Tree

A *Random Tree* é uma árvore induzida aleatoriamente a partir de um conjunto de árvores possíveis, usando m atributos aleatórios em cada nodo. Em outras palavras, cada árvore tem uma chance igual de ser amostrada. Uma das formas mais utilizadas de se construir uma *Random Tree*, é através do processo de ramificação Galton-Watson (RODRIGUEZ, 2012).

Esse processo de ramificação pode ser explicado da seguinte maneira: suponha que temos árvores que dão nascimento a novas árvores segundo a variável aleatória X com valores no conjunto $\{0, 1, 2, \dots\}$ e função de distribuição de probabilidade expressa por $P(X = k) = p_k$, para $\{k = 0, 1, 2, \dots, n\}$. Inicialmente suponha que, temos uma única árvore na iteração $n = 0$. Essa árvore dá nascimento, na iteração $n = 1$ a novas árvores segundo a variável aleatória X . Dizemos que essas novas árvores são descendentes diretas da anterior. Isto é, ela dá nascimento a k descendentes diretos com probabilidade p_k .

Se temos um certo número de árvores na iteração n , cada uma delas dá nascimento a novas árvores na iteração $n + 1$. Em cada caso, isto é feito de acordo com uma variável aleatória independente e identicamente distribuída a variável aleatória X . Dizemos que a árvore da iteração $n = 0$ constitui a geração 0 e as árvores que nascem na iteração n constituem a n -ésima geração do processo, $n > 1$. Denotamos por Z_n a variável aleatória que conta o número de árvores da n -ésima geração.

Nesse trabalho, o classificador *Random Tree* foi utilizado com uma profundidade máxima das árvores ilimitada, com o número aleatório de semente (*seed*) (usado para selecionar atributos) igual a 1 e o número preferido de instâncias (*batchSize*) para processar, se a predição do lote está sendo executada, igual a 100.

K-Nearest Neighbors (KNN)

Proposto por Fix e Jr (1951), o K-Vizinhos Mais Próximos (KNN) é um classificador baseado em memória, que frequentemente é utilizado em aplicações do mundo real devido à sua simplicidade. Mesmo sendo uma técnica simples, obteve considerável exatidão para classificação em diversas aplicações e é comumente utilizado na comparação com outros

classificadores (SILVA, 2007).

Com relação ao custo de memória, o KNN armazena todos os padrões de treinamento e quanto ao desempenho, para cada padrão a ser classificado o KNN deve calcular sua distância para todos os padrões de treinamento. É usual escolher o “k” ímpar, para reduzir as ocorrências de impossibilidade de classificação devido a empate na quantidade de vizinhos mais próximos (SILVA, 2007).

A fase de treinamento do KNN consiste simplesmente em armazenar todos os padrões de treinamento. O KNN tem um parâmetro k que é o número de vizinhos a ser considerado para a classificação. Para $k = 1$, KNN é também chamado de método do vizinho mais próximo (*Nearest Neighbor* - NN). Ele classifica um dado padrão como pertencente a mesma classe do padrão mais próximo a ele no conjunto de treinamento. Há um número de medidas de distâncias usadas neste processo, entretanto a distância euclidiana é a mais utilizada.

Quando $k > 1$, primeiramente são calculadas as distâncias do novo padrão a ser classificado com todos os padrões do conjunto de treinamento. Subsequentemente, o algoritmo considera os k padrões do conjunto treinamento com as menores distâncias ao padrão a ser classificado. Finalmente, o novo padrão é classificado como pertencente a classe da maioria dos k padrões mais próximos do conjunto de treinamento.

Nesse trabalho, o classificador KNN foi utilizado com o $k = 1$. Uma vez que, foram realizados testes para o KNN com o $k = 1$, $k = 3$, $k = 5$ e $k = 7$ e os melhores resultados foram alcançados com o $k = 1$.

Comitê (MLP + SVM + KNN)

Um comitê é um conjunto de classificadores cujas decisões individuais são combinadas de alguma forma para classificar um novo caso. Um resultado interessante é que comitês de classificadores podem ser mais precisos que os classificadores individuais que compõem o comitê (BERNARDINI, 2006).

Sistema de Múltiplos Classificadores (SMC) são utilizados para combinar classificadores. O estudo de SMCs é necessário, visto que não há um único classificador que seja considerado ideal para todos os problemas. Também, não há orientação clara de como escolher um conjunto de métodos de aprendizagem. É raro quando se tem um conhecimento completo sobre a distribuição de dados. Além disso, os detalhes de como o algoritmo de classificação se comporta não são compreendidos. Portanto, em tarefas de classificação de padrões, é difícil encontrar um bom classificador individual (JUNIOR, 2011).

O resultado da combinação dos C classificadores é realizado usando um SMC. Nesse trabalho, foi realizada a combinação dos classificadores MLP, SVM e KNN. A regra de combinação utilizada nesse trabalho foi a média de probabilidades (média das saídas de

cada componente). Sendo considerado o combinador mais simples e muito utilizado de acordo com [Barbosa \(2009\)](#).

Random Forest

O classificador *Random Forest* ([BREIMAN, 2001](#)) é um tipo de SMC. Ele é uma combinação de predições de diversas árvores onde cada árvore depende dos valores de um vetor independente, amostrados aleatoriamente e com a mesma distribuição para todas as árvores da floresta (se denomina para uma série de árvores de decisão). Por fim, quando um grande número de árvores são geradas, as classes com maior número de votos são eleitas.

Em um *Random Forest* cada nó é dividido usando o melhor dentre um subconjunto de indicadores escolhidos aleatoriamente naquele nó. Esta estratégia um tanto contraditória acaba por funcionar muito bem em comparação com muitos outros classificadores, além de ser robusto a superajuste nos parâmetros. Além disso, é de fácil utilização, no sentido que ele tem apenas dois parâmetros (o número de variáveis no subconjunto aleatório em cada nó e o número de árvores da floresta) e, normalmente, não é muito sensível aos seus valores.

A [Figura 11](#) mostra os passos do classificador *Random Forest*. A partir de um vetor de atributos D , são gerados outros vetores de atributos, que são embaralhados em relação ao vetor original. É gerado um vetor para cada árvore da *Random Forest*. Em seguida, eles são passados como parâmetro para as árvores de decisão. Cada árvore irá gerar um resultado para a classificação e, após isso, os resultados são combinados obtendo uma saída unificada.

Nesse trabalho, o classificador *Random Forest* foi utilizado com uma profundidade máxima das árvores ilimitada e o número de árvores foi 100.

3.4 Seleção de Atributos

O processo de seleção de atributos é necessário para identificar e eliminar inconsistências presentes nos dados. Ainda que também possam ser utilizados todos os atributos de uma base de dados, a seleção de atributos relevantes e a eliminação dos que são totalmente desnecessários podem diminuir a complexidade do problema e alcançar um desempenho ótimo de aprendizagem ([NEVES, 2003](#)).

A seleção de atributos desempenha um papel importante na preparação de dados para a mineração, pois quanto mais se remove dados redundantes e irrelevantes, mais se reduz a dimensionalidade do espaço de atributos ([BATISTA, 2003](#))

O algoritmo Razão de Ganho de Informação ([QUINLAN, 1986](#)) (*Gain Ratio*

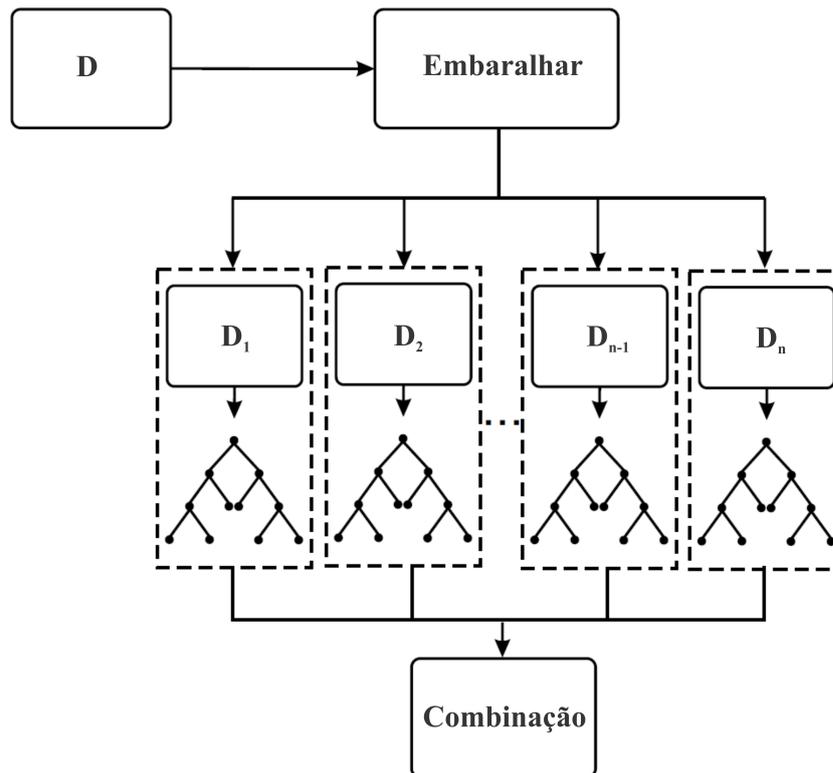


Figura 11 – Passos do classificador *Random Forest*. Fonte: [Silva \(2014\)](#).

Information) foi utilizado nesse trabalho para a seleção de atributos. A técnica *Gain Ratio* é uma versão ponderada da técnica *Information Gain*, uma generalização da métrica ganho de informação. Ele é um filtro de atributos supervisionado muito flexível e permite combinar vários métodos de pesquisa e avaliação. O método utilizado foi o *Ranker*, que classifica os atributos por suas avaliações individuais.

A técnica *Gain Ratio* visa solucionar uma limitação da técnica de *Information Gain*, a qual tende a selecionar atributos com maior número de valores distintos. Ele procura selecionar atributos que maximizam o ganho de informação, enquanto minimizam o número de valores de um atributo. Após o cálculo do valor da razão de ganho de cada atributo, é gerado um *ranking*, sendo selecionados os n atributos desse *ranking* com os maiores valores ([LIMA, 2016](#)).

3.5 Métricas de Avaliação dos Resultados

Para avaliar o desempenho dos algoritmos de classificação foram utilizadas algumas métricas conhecidas na literatura: Acurácia (A), índice Kappa (K), Precisão (P) e Sensibilidade (S). Esses métodos de avaliação dos resultados foram calculados a partir da matriz de confusão, composta por 4 valores: O Verdadeiro Positivo (VP) e o Verdadeiro Negativo (VN), que apresentam a quantidade de lesões não-melanocíticas e melanocíticas, respectivamente, classificadas corretamente; e o Falso Positivo (FP) e o Falso Negativo (FN), que

apresentam a quantidade de lesões não-melanocíticas e melanocíticas, respectivamente, classificadas incorretamente.

Acurácia: é a porcentagem de casos corretamente classificados em um conjunto de teste (POWERS, 2007). Ela é calculada a partir da Equação 3.21.

$$A = \frac{VP + VN}{VP + FP + FN + VN}. \quad (3.21)$$

Índice Kappa: calculado a partir da Equação 3.22, o Kappa é utilizado como uma medida apropriada da exatidão por representar inteiramente a matriz de confusão. Ele é um coeficiente de concordância para escalas nominais que mede o relacionamento entre a concordância, além da casualidade, e a discordância esperada (ROSENFELD; FITZPATRICK-LINS, 1986). O nível de precisão do índice Kappa para a avaliação dos resultados foi classificado conforme a Tabela 2, de acordo com o estabelecido por Landis e Kock (LANDIS; KOCH, 1977).

$$K = \frac{\theta_1 - \theta_2}{1 - \theta_2}, \quad (3.22)$$

$$\theta_1 = \frac{VP + VN}{VP + VN + FP + FN}, \quad (3.23)$$

$$\theta_2 = \frac{[(VP + FN)(VP + FP)] + [(VN + FN)(VN + FP)]}{(VP + VN + FP + FN)^2}. \quad (3.24)$$

Tabela 2 – Nível de precisão da classificação, de acordo com o índice Kappa.

Índice Kappa (K)	Qualidade
$K \leq 0.2$	Ruim
$0.2 < K \leq 0.4$	Razoável
$0.4 < K \leq 0.6$	Bom
$0.6 < K < 0.8$	Muito Bom
$K \geq 0.8$	Excelente

Precisão: calculada pela Equação 3.25, a precisão indica a proporção de verdadeiros positivos em relação a todas as predições positivas, ou seja, mostra a quantidade de objetos da classe X classificados corretamente em relação a todos os objetos classificados como sendo da classe X (POWERS, 2007).

$$P = \frac{VP}{VP + FP} \quad (3.25)$$

Sensibilidade: calculada pela Equação 3.26, a sensibilidade indica a proporção de verdadeiros positivos em relação a suas predições positivas e as suas incorretas predições

negativas. Ou seja, de todos os objetos da classe X, quantos foram classificados como X (POWERS, 2007).

$$S = \frac{VP}{VP + FN} \quad (3.26)$$

3.6 Método de Avaliação dos Classificadores

O método para avaliação dos classificadores utilizado nesse trabalho foi o *k-fold cross-validation* (ou validação cruzada) ($k = 10$). O conjunto de dados originais é aleatoriamente particionado em 10 subconjuntos de dados com o mesmo tamanho. A partir dos 10 subconjuntos, um único subconjunto é escolhido como conjunto de validação para testar o modelo, e os 9 subconjuntos restantes são utilizados como dados de treinamento. Esse processo é repetido 10 vezes. Para a validação dos dados, cada um dos 10 subconjuntos é utilizado uma vez. Assim, a média dos 10 resultados gerados é obtida para produzir uma estimativa única.

3.7 Considerações Finais

Neste capítulo foram apresentadas as bases de imagens e suas características, os descritores e classificadores utilizados para os testes, o algoritmo de seleção de atributos, as métricas utilizadas para a avaliação dos resultados obtidos e o método de avaliação dos classificadores.

Considerando a importância de diferenciar lesões benigna de malignas, foram realizados diversos testes com diferentes combinações de descritor e classificador. Descritores de imagens são uma classe de algoritmos de visão computacional. Eles são responsáveis por extrair atributos importantes das imagens. Algoritmos de classificação são utilizados em várias áreas da computação, pode-se observar na literatura uma grande variedade de algoritmos com esse propósito. Nesse trabalho foram testados os algoritmos mais utilizados na literatura. Utilizou-se também SMCs com o objetivo de melhorar os obtidos. A principal análise feita em relação à classificação foi a taxa de acerto e o índice Kappa.

4 Método Proposto

Esse trabalho propõe um método de visão computacional para a identificação do melanoma em imagens dermatoscópicas, com intuito de auxiliar o dermatologista no seu diagnóstico. Na [Figura 12](#) é apresentada a estrutura do método proposto. Este método é baseado nas etapas de processamento de imagens digitais de acordo com [Gonzalez e Woods \(2002\)](#). As seguintes etapas de processamento fazem parte do seu escopo: pré-processamento, extração de características e classificação das lesões de pele conforme as suas classes estabelecidas.

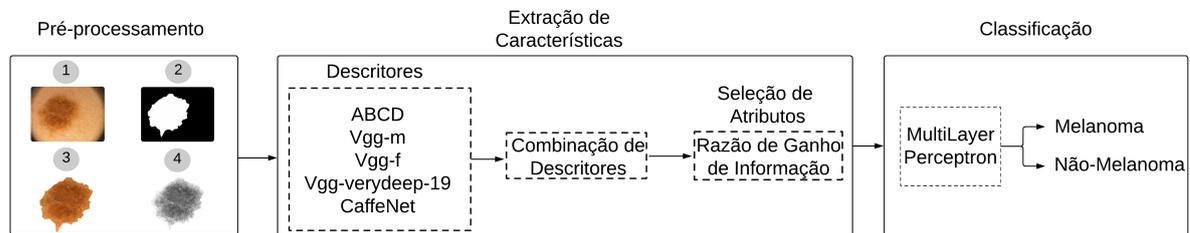


Figura 12 – Estrutura do método proposto.

A primeira etapa do método proposto é o pré-processamento. As lesões de pele segmentadas são obtidas a partir da máscara binária fornecida pelas bases. Em seguida, as imagens são transformadas no sistema de cores HSV. Para a extração de características é realizada uma combinação de descritores e seleção de atributos, dando origem ao descritor híbrido para a extração de características das imagens. As características obtidas são utilizadas como dados de entrada pelo MLP para classificar as lesões em melanoma ou não-melanoma.

4.1 Pré-processamento

O modelo de cores HSV é definido por três propriedades: matiz (*Hue* – H), saturação (*Saturation* – S) e valor (*Value* – V). A matiz identifica o tipo de cor desde o vermelho até o violeta e mais o magenta, podendo atingir valores de 0 a 100%. A saturação pode atingir valores de 0 a 100% e quanto menor esse valor, mais cinzenta é a cor. Por fim, o valor, que define o brilho da cor, também atinge valores de 0 a 100%. A [Figura 13](#) mostra os componentes de uma imagem da base de dados PH² no modelo de cores HSV.

Desse modelo de cores, as características foram extraídas pelos descritores a partir da componente gerada pela soma das bandas S e V. Em outras palavras, os valores dos *pixels* dessas duas bandas de cores são somados. Isso se deve ao fato que após diversos

testes com diferentes combinações de modelos de cores e bandas no trabalho de [Moura et al. \(2015\)](#), essa nova componente conseguiu diferenciar melhor os tipos de lesão de pele.

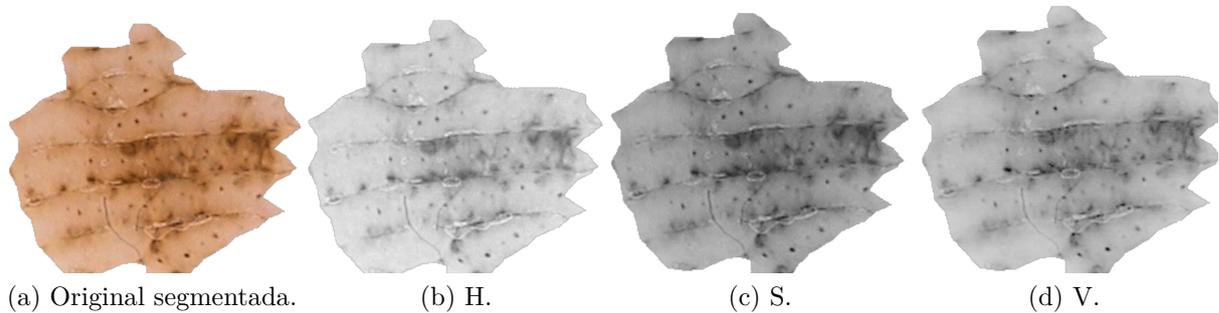


Figura 13 – Componentes do modelo de cores HSV.

4.2 Segmentação

A segmentação das imagens de lesão de pele consiste em separar a região doente da região saudável, para que as lesões possam ser detectadas. Entretanto, nesse trabalho, não utilizamos nenhum método automático para segmentação das imagens e sim as máscaras binárias das lesões segmentadas por especialistas. A [Figura 14](#) mostra um exemplo de como foi feito o processo para a obtenção das lesões segmentadas a partir da imagem original e da máscara binária da lesão.



Figura 14 – Segmentação da lesão de pele a partir da máscara binária fornecida pelas bases.

4.3 Extração de Características

Nessa etapa de extração de características, o vetor é formado a partir de duas etapas, apresentadas na [Figura 15](#): combinação de descritores e seleção de atributos.

1. Combinação de Descritores: é proposto um descritor híbrido para extração de atributos após uma pesquisa na literatura sobre quais descritores estavam sendo mais utilizados na detecção de câncer de pele em imagens médicas, assim como, quais estavam presentes em trabalhos mais recentes e qual seu custo computacional.

Os descritores que foram escolhidos para compor o descritor híbrido proposto foram: regra ABCD e as redes neurais convolucionais (Vgg- m, Vgg-s, Vgg-verydeep-19 e CaffeNet). Os mesmos foram apresentados e detalhados anteriormente no [Capítulo 3](#).

A regra ABCD calcula na imagem 5 atributos: Assimetria (1 atributo), Irregularidade da Borda (2 atributos), Cor (1 atributo) e Diâmetro (1 atributo). Já as CNN's pré-treinadas: Vgg-m, Vgg-s, Vgg-verydeep-19 e CaffeNet possui, cada uma, 4096 atributos. Desse modo, totalizando 16.389 atributos.

Após essa combinação (ou concatenação) de descritores é realizada uma seleção de atributos. Isso porque a adição de atributos irrelevantes às instâncias de uma base de dados pode dificultar o processo de aprendizado do algoritmo (classificador), causando erros aleatórios durante os testes ([KUMAR et al., 2016](#)).

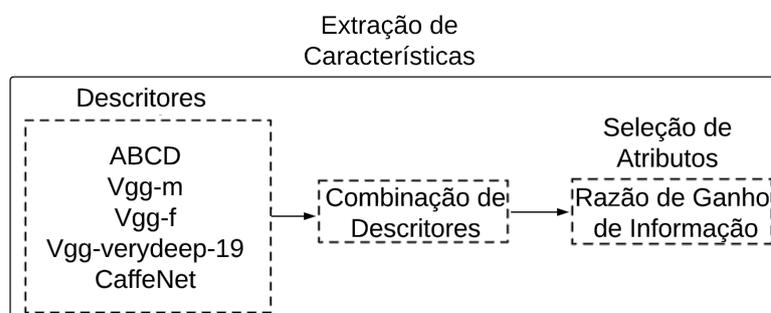


Figura 15 – Etapa de extração de características.

2. Seleção de Atributos: o processo de seleção de atributos pode produzir uma representação mais compacta do conceito a ser aprendido, ou seja, focar nos atributos que realmente são importantes para a classificação.

O algoritmo Razão de Ganho de Informação (*Gain Ratio Information*) foi utilizado nesse trabalho. Ele é um filtro de atributos supervisionado que pode ser usado para selecionar atributos. É muito flexível e permite combinar vários métodos de pesquisa e avaliação. O método utilizado foi o *Ranker*, que classifica os atributos por suas avaliações individuais. Depois dos atributos classificados, somente 5% (820 atributos) mais relevantes foram utilizados. Visto que os descritores utilizados retornam uma grande quantidade de atributos (16.389 no total), ou seja, uma grande possibilidade de existir atributos que não sejam significativos ou que sejam duplicados, dificultando a classificação das lesões de pele ([KUMAR et al., 2016](#)).

É importante ressaltar que a escolha desses descritores, assim como a porcentagem de atributos mais relevantes utilizados foi realizada a partir dos testes feitos nesse trabalho, que serão apresentados no próximo capítulo, na [seção 5.2](#). Em suma, essa porcentagem foi variada em 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% e 100%. Diante disso, a utilização de 5% dos atributos mais relevantes apresentou o melhor

resultado.

Com os descritores escolhidos e a seleção de atributos realizada através do algoritmo *Gain Ratio Information*, compomos assim o descritor híbrido para a extração de características em imagens de lesão de pele.

A [Tabela 3](#) apresenta a porcentagem de cada descritor utilizado na composição do descritor híbrido para as bases de imagens a partir da seleção de 5% dos atributos mais relevantes feita com o algoritmo *Gain Ratio Information*. Pode-se observar que os atributos mais utilizados pertencem ao algoritmo Vgg-verydeep-19 (48,04%), em seguida a rede Vgg-m (33,65%), CaffeNet (9,63%), Vgg-f (8,29%) e por fim a regra ABCD com 0,39%.

Entretanto, o algoritmo ABCD possui apenas 5 atributos em que 3 deles foram selecionados pelo *Gain Ratio Information*, enquanto as CNN's pré-treinadas possuem cada uma 4096 atributos. Isso mostra a importância dos atributos da regra ABCD para o classificador diferenciar as duas classes do problema.

Tabela 3 – Composição do descritor híbrido proposto na união das bases de imagens PH² + DermIS.

Descritor	Quantidade Total de Atributos (16389)	Quantidade de Atributos Selecionados (820)	Porcentagem (%)
CaffeNet	4096	79	9,63
Vgg-m	4096	276	33,65
Vgg-verydeep-19	4096	394	48,04
ABCD	5	3	0,39
Vgg-f	4096	68	8,29

4.4 Classificação

Na identificação das lesões de pele, de acordo com as características extraídas a partir do descritor híbrido, é utilizado o classificador MLP (que alcançou o melhor resultado nos testes realizados), que permite reconhecer os padrões desejados por meio da aprendizagem. Em poucas palavras, o MLP é um sistema de neurônios ligados por conexões sinápticas e dividido em neurônios de entrada, que recebem estímulos do meio externo, neurônios internos (ocultos) e neurônios de saída, que se comunicam com o exterior. Ele foi concebido para resolver problemas mais complexos, os quais não poderiam ser resolvidos pelo modelo de neurônio básico. Para isto são necessárias mais conexões, os quais só existem em uma rede de *perceptrons* dispostos em camadas.

5 Resultados e Discussão

Neste capítulo, são apresentados e analisados os resultados encontrados nos testes realizados. A realização dos testes tem por objetivo avaliar o método proposto neste trabalho para a extração de características e classificação de lesões de pele, visando o auxílio ao dermatologista em seu diagnóstico.

Para isso, nos testes dos algoritmos, foram utilizadas duas bases de imagens: PH² e DermIS. Os testes foram realizados com a união dessas duas bases, proporcionando uma maior diversidade de imagens, 406 no total.

A base PH² possui um total de 200 imagens dermatoscópicas, dentre elas, 40 melanomas e 160 não-melanomas. Já a base DermIS é formada por um total de 206 imagens, dentre elas, 119 melanomas e 87 não-melanomas.

Para a classificação, foram utilizados nos testes os algoritmos *MultiLayer Perceptron* (MLP), *Support Vector Machine* (SVM), *Random Tree* e *K-Nearest Neighbors* (KNN) e dois comitês de classificadores. O primeiro comitê é formado pelos algoritmos MLP, SVM e KNN e o segundo é o *Random Forest*.

Os testes foram divididos em três grupos ou cenários, a saber:

- imagens sem segmentação;
- imagens segmentadas pelos especialistas;
- imagens com erros de segmentação.

No primeiro grupo foram utilizadas todas as imagens originais das bases, ou seja, sem nenhum pré-processamento ou segmentação. No segundo grupo foram utilizadas todas as imagens segmentadas pelos especialistas, sendo elas convertidas ao modelo de cores HSV, utilizando somente a soma das bandas S e V. Já no terceiro grupo, todas as imagens foram processadas automaticamente com dois erros de segmentação (detalhados na [seção 5.3](#)): *under-segmentation* e *over-segmentation*.

Todos os algoritmos foram implementados utilizando a ferramenta computacional MATLAB[®] (*Matrix Laboratory*) (THOMPSON; SHURE, 1995). Já para a classificação, a ferramenta WEKA[®] (*Waikato Environment for Knowledge Analysis*) (HALL et al., 2009) foi utilizada.

Nos testes foi utilizado o método estatístico *k-fold cross-validation*, com o *k* igual a 10. Para avaliar os resultados obtidos foram utilizadas as métricas mostradas na [Seção 3.5](#): Acurácia (A), Precisão (P), Sensibilidade (S) e o índice Kappa (K).

5.1 Resultados com Imagens sem Segmentação

Neste experimento, os testes são realizados com as imagens originais das bases de dados sem segmentação e nenhum pré-processamento. Vale ressaltar que as características do descritor ABCD não foram utilizadas, pois para serem calculadas é necessário ter informações da borda da lesão.

Todos os descritores detalhados na Seção 3.2 e os classificadores da Seção 3.3 foram utilizados para os testes com as imagens sem segmentação. O objetivo é avaliar a classificação feita a partir das informações extraídas por esses descritores.

A Tabela 4 apresenta os resultados obtidos na classificação das imagens sem segmentação da união das bases PH² e DermIS. Os classificadores utilizados foram: MLP, SVM, Random Tree e KNN.

Para o classificador MLP, pode-se observar que a maioria dos resultados obtiveram classificação considerada “Boa” de acordo com a Tabela 2 que mostra os níveis de exatidão em relação ao índice Kappa. O melhor resultado alcançado foi com a utilização do descritor HOG com uma acurácia e sensibilidade de 77,8%, precisão de 77,9% e índice Kappa igual a 0,5421. Já com o classificador SVM somente quatro dos treze resultados obtiveram classificação considerada “Boa”, o restante obteve classificação “Razoável”. Com a utilização da CNN AlexNet como descritor, esse classificador obteve o melhor resultado com Kappa igual a 0,5863 e acurácia de 79,8%.

O classificador *Random Tree* apresentou melhores resultados se comparados ao SVM. Dentre eles, ocorreram seis resultados considerados “Bons”, onde a CNN Vgg-verydeep-19 alcançou a maior taxa de acerto com 76,3% e Kappa de 0,5038. Entretanto, o pior resultado obtido em todos os testes com as imagens sem segmentação foi com a utilização do classificador *Random Tree* e o descritor *Box-Counting*, que atingiu um índice Kappa considerado “Razoável” de 0,2075. Já com o classificador KNN o melhor resultado também atingiu uma acurácia de 76,3%, contudo com o descritor Vgg-m e um Kappa menor igual a 0,4924. Ao todo, o KNN apresentou onze resultados considerados “Bons”.

A Tabela 5 mostra os resultados obtidos com a utilização de comitês para a classificação das imagens de lesão de pele sem segmentação. O primeiro comitê é formado pelos algoritmos MLP, SVM e KNN e o segundo é o *Random Forest*. No primeiro, o melhor resultado é alcançado com o descritor Vgg-f com uma taxa de acerto igual a 80,3%, sendo esta, a maior acurácia encontrada para os testes das imagens sem segmentação. Próximo a esse resultado, o *Random Forest* obteve uma acurácia de 79,5% e índice Kappa de 0,5705 com o descritor Vgg-verydeep-16.

Em relação aos descritores é possível observar que houve uma variedade dentre os melhores resultados com cada classificador utilizado. Ou seja, os classificadores obtiveram uma maior taxa de acerto com diferentes descritores.

Tabela 4 – Resultados da classificação com as imagens sem segmentação da união das Bases PH² e DermIS utilizando os classificadores MLP, SVM, Random Tree e KNN.

Descritor	Base de Imagens PH ² + DermIS															
	MLP				SVM				Random Tree				KNN			
	P	S	K	A	P	S	K	A	P	S	K	A	P	S	K	A
GLCM	77,4	76,8	0,5227	76,8	70,4	70,9	0,3659	70,9	74,3	73,6	0,4573	73,6	71,9	71,7	0,4102	71,6
GLRLM	77,9	77,8	0,5369	77,8	70,6	69,7	0,3945	69,7	72,9	72,9	0,4314	72,9	74,7	74,6	0,4694	74,6
HOG	78,2	78,3	0,5421	78,3	71,3	71,7	0,3940	71,6	66,0	66,3	0,2863	66,2	73,8	74,1	0,4442	74,1
LBP	69,1	69,2	0,3517	69,2	75,2	70,75	0,4218	70,2	69,3	69,7	0,3548	69,7	72,8	72,7	0,4282	72,6
Tamura	74,7	74,9	0,4668	74,8	71,3	68,7	0,3764	68,7	74,2	74,4	0,4576	74,3	69,3	69,5	0,3562	69,4
Box-Counting	73,8	70,4	0,4163	70,4	70,5	68,5	0,3668	68,4	62,2	62,1	0,2075	62,1	63,6	63,1	0,2349	63,1
AlexNet	75,9	76,1	0,4935	76,1	80,6	79,8	0,5863	79,8	73,6	73,6	0,4463	73,6	75,1	75,4	0,4664	75,3
CaffeNet	75,8	75,9	0,4923	75,8	77,0	76,6	0,5160	76,6	70,0	70,4	0,3684	70,4	75,8	76,1	0,4830	76,1
Vgg-f	74,8	75,1	0,4666	75,1	77,6	77,3	0,5287	77,3	70,7	70,7	0,3856	70,6	74,6	74,9	0,4631	74,8
Vgg-m	70,7	70,7	0,3842	70,6	76,2	70,2	0,2857	70,1	67,4	67,5	0,3162	67,4	76,1	76,4	0,4924	76,3
Vgg-s	75,4	75,6	0,4819	75,6	74,5	70,9	0,3148	70,9	69,8	70,0	0,3665	69,9	73,5	73,9	0,4382	73,8
Vgg-verydeep-16	73,9	73,9	0,4521	73,8	72,5	72,2	0,3716	72,1	74,2	73,6	0,4561	73,6	75,8	75,9	0,4912	75,8
Vgg-verydeep-19	74,3	74,6	0,4585	74,6	72,2	71,9	0,3668	71,9	76,4	76,4	0,5038	76,3	75,1	75,4	0,4676	75,3

Tabela 5 – Resultados da classificação com as imagens sem segmentação da união das Bases PH² e DermIS utilizando comitês.

Base de Imagens PH ² + DermIS								
Descritor	MLP + SVM + KNN				Random Forest			
	P	S	K	A	P	S	K	A
GLCM	73,7	73,9	0,4484	73,8	78,3	78,1	0,5435	78,0
GLRLM	74,6	74,6	0,4393	74,6	79,3	78,8	0,5623	78,8
HOG	74,6	73,9	0,4112	73,8	77,2	77,3	0,5191	77,3
LBP	73,7	73,4	0,4044	73,3	75,1	75,1	0,4774	75,1
Tamura	72,8	72,4	0,4275	72,4	76,7	76,6	0,5106	76,6
Box-Counting	69,5	67,7	0,3484	67,7	66,8	64,8	0,2918	64,7
AlexNet	77,0	77,1	0,5177	77,1	79,3	79,1	0,5641	79,0
CaffeNet	75,9	75,9	0,4946	75,8	78,3	78,3	0,5451	78,3
Vgg-f	80,3	80,3	0,5856	80,3	78,5	78,1	0,5466	78,0
Vgg-m	72,7	72,4	0,3794	72,4	76,6	76,8	0,5075	76,8
Vgg-s	74,2	73,6	0,4064	73,6	76,6	76,8	0,5007	76,8
Vgg-verydeep-16	72,5	72,7	0,3944	72,6	79,5	79,6	0,5705	79,5
Vgg-verydeep-19	73,3	72,4	0,3718	72,4	78,7	78,8	0,5525	78,8

5.2 Resultados com Imagens Segmentadas pelos Especialistas

Nesta seção, são apresentados os resultados obtidos com as imagens segmentadas a partir das máscaras binárias fornecidas pelas próprias bases (segmentação feita manualmente por especialistas).

Todos os descritores detalhados na Seção 3.2 e os classificadores da Seção 3.3 foram utilizados para os testes com as imagens segmentadas. O objetivo é avaliar a classificação feita a partir das informações extraídas por esses descritores.

A Tabela 6 mostra os resultados da classificação para cada um dos descritores utilizados na união das bases PH² + DermIS com as imagens segmentadas. O descritor CaffeNet obteve o melhor resultado com o uso do classificador MLP, atingindo 83,0% de acurácia e 0,5987 de índice Kappa. Ao todo foram obtidos doze resultados considerados “Bons” e somente dois “Razoáveis”.

Em relação ao classificador SVM essa taxa de acerto aumentou para 84,4% em conjunto com a CNN Vgg-f. O índice Kappa de 0,6098 classificou o resultado como “Muito Bom”. Dentre os treze resultados restantes, nove foram considerados “Bons”, três “Razoáveis” e apenas um “Ruim” obtido com o uso do descritor GLRLM.

Com o classificador *Random Tree* a maior acurácia obtida foi de 76,6% com o descritor ABCD, onde obteve um índice Kappa igual a 0,5095 considerado “Bom”. A precisão e sensibilidade também obtiveram uma taxa de 76,6%. Nesse classificador, a maioria dos resultados foram considerados “Razoáveis” de acordo com Kappa. Além disso,

dois foram considerados “Ruins”.

Por fim, o KNN se mostrou melhor que o *Random Tree* pois apenas um dos resultados obteve classificação “Razoável” e apenas um “Ruim”. O restante foi considerado “Bom”, onde o melhor resultado alcançou uma acurácia de 76,6% e índice Kappa de 0,5335 com o descritor LBP. Mesmo que com o descritor AlexNet foi atingida uma taxa de acerto igual a 78,5%, o índice Kappa obtido foi menor com apenas 0,4733.

Os resultados com as imagens segmentadas utilizando comitês são apresentados na [Tabela 7](#). Dois comitês foram utilizados, o primeiro composto pelo MLP, SVM e KNN e o segundo pelo *Random Forest*.

No comitê 1, onze resultados obtiveram uma classificação considerada “Boa” e dois considerados “Razoáveis”. O melhor resultado foi com a utilização do descritor Vgg-s, onde obteve uma acurácia igual a 81,2% e índice Kappa de 0,6027 considerado “Muito Bom”.

Em relação ao classificador *Random Forest*, o melhor resultado obtido teve uma acurácia de 81,5% e um índice Kappa de 0,5267 com a utilização do descritor CaffeNet. Entretanto, é possível observar que com o descritor Vgg-f também foi obtida uma acurácia de 81,5%, mas o índice Kappa foi menor, igual a 0,5199. De todos os resultados, treze foram considerados “Bons” e apenas um considerado “Ruim” com o descritor *Box-Counting*.

Dentre todos os descritores, os piores resultados ocorreram com a utilização do descritor *Box-Counting*, onde em todos os classificadores só foram obtidos resultados considerados “Razoáveis” e “Ruins” de acordo com o Kappa. com o *Random Tree* foi alcançada uma acurácia de apenas 57,6% e um índice Kappa de 0,0988.

Vale ressaltar novamente que os melhores resultados ocorreram com diferentes descritores, com exceção da CNN CaffeNet em que apresentou a maior taxa de acerto tanto para o classificador MLP como para o classificador *Random Forest*, 83,0% e 81,5%, respectivamente.

Os testes com as imagens segmentadas apresentaram melhores resultados na classificação do que com as imagens sem segmentação. Isso se deve ao fato de que a maior acurácia (84,5%) e índice Kappa (0,6098) foram obtidos com a utilização das imagens segmentadas manualmente pelos especialistas. Além disso, enquanto no primeiro grupo de teste (imagens sem segmentação) foram obtidas 53 (cinquenta e três) classificações consideradas “Boas” e “Muito Boas”, no segundo grupo de teste (imagens segmentadas) foram obtidas 64. Desse modo, os testes a seguir foram realizados somente nas imagens segmentadas, pois a segmentação da lesão de pele em imagens dermatoscópicas aumenta as taxas de acerto na classificação, já que as características são extraídas apenas da lesões.

Tabela 6 – Resultados da classificação com as imagens segmentadas da união das Bases PH² e DermIS utilizando os classificadores MLP, SVM, Random Tree e KNN.

Descritor	Base de Imagens PH ² + DermIS															
	MLP				SVM				Random Tree				KNN			
	P	S	K	A	P	S	K	A	P	S	K	A	P	S	K	A
ABCD	78,2	78,3	0,5400	78,3	77,5	77,3	0,5276	77,3	76,6	76,6	0,5095	76,6	76,4	76,6	0,5051	76,6
GLCM	72,3	71,9	0,4173	71,9	76,7	75,6	0,5022	75,6	69,8	69,7	0,3663	69,7	75,4	74,9	0,4809	74,8
GLRLM	77,3	76,8	0,5216	76,8	61,2	58,6	0,1759	58,6	74,9	74,6	0,4717	74,6	72,9	72,7	0,4307	72,6
HOG	70,8	70,7	0,3870	70,6	75,2	75,1	0,4797	75,1	74,6	74,4	0,4672	74,3	76,5	76,4	0,5060	76,3
LBP	73,5	73,4	0,4430	73,3	64,0	63,5	0,2435	63,5	65,8	65,5	0,2828	65,5	79,2	76,6	0,5335	76,6
Tamura	77,5	77,3	0,5276	77,3	73,3	67,7	0,2187	67,7	72,5	72,9	0,4184	72,9	69,7	68,5	0,3571	68,4
Box-Counting	63,8	65,0	0,2112	65,0	67,5	68,0	0,2744	67,9	57,1	57,6	0,0988	57,6	61,4	61,1	0,1906	61,0
AlexNet	78,8	79,1	0,5056	79,0	81,2	81,0	0,5070	81,0	67,8	68,2	0,2497	68,2	77,8	78,6	0,4733	78,5
CaffeNet	82,8	83,0	0,5987	83,0	83,9	83,7	0,5864	83,7	66,0	66,0	0,2094	66,0	76,9	77,6	0,4565	77,5
Vgg-f	79,9	80,3	0,5316	80,2	84,5	84,5	0,6098	84,4	63,4	64,0	0,1489	64,0	76,6	77,1	0,4543	77,0
Vgg-m	79,6	79,6	0,5724	79,5	80,4	80,5	0,5884	80,5	71,3	71,4	0,3977	71,4	76,3	76,6	0,4971	76,6
Vgg-s	77,2	77,3	0,5223	77,3	78,7	78,8	0,5525	78,8	69,4	69,5	0,3576	69,4	76,5	76,6	0,4901	76,6
Vgg-verydeep-16	76,4	76,4	0,5049	76,3	78,0	78,1	0,5384	78,0	71,5	71,7	0,4022	71,6	76,4	76,6	0,4936	76,6
Vgg-verydeep-19	78,6	78,8	0,5494	78,8	79,2	79,3	0,5619	79,3	69,2	69,0	0,3531	68,9	76,1	76,4	0,4901	76,3

Tabela 7 – Resultados da classificação com as imagens segmentadas da união das Bases PH² e DermIS utilizando comitês.

Base de Imagens PH ² + DermIS								
Descritor	MLP + SVM + KNN				Random Forest			
	P	S	K	A	P	S	K	A
ABCD	77,9	77,6	0,5047	77,5	76,3	76,4	0,5015	76,3
GLCM	75,9	75,6	0,4934	75,6	74,2	73,4	0,4540	73,3
GLRLM	74,1	74,4	0,4398	74,3	80,4	79,8	0,5845	79,8
HOG	74,8	74,9	0,4467	74,8	76,4	76,4	0,5038	76,3
LBP	76,5	76,6	0,4889	76,6	76,9	76,4	0,5125	76,3
Tamura	77,1	77,3	0,5102	77,3	74,0	74,4	0,4501	74,3
Box-Counting	68,7	68,7	0,2851	68,7	62,0	63,5	0,1838	63,5
AlexNet	74,4	75,6	0,3752	75,6	80,3	80,3	0,4866	80,2
CaffeNet	79,0	79,6	0,4811	79,5	81,5	81,5	0,5267	81,5
Vgg-f	80,1	80,5	0,5061	80,5	81,8	81,5	0,5199	81,5
Vgg-m	79,7	79,8	0,5723	79,8	80,3	80,0	0,5855	80,0
Vgg-s	81,1	81,3	0,6027	81,2	79,5	79,3	0,5697	79,3
Vgg-verydeep-16	78,9	79,1	0,5572	79,0	79,6	79,1	0,5679	79,0
Vgg-verydeep-19	80,6	80,8	0,5885	80,7	79,1	78,6	0,5577	78,5

A partir dos resultados apresentados anteriormente nas [Tabela 4](#), [Tabela 5](#), [Tabela 6](#) e [Tabela 7](#), foi realizada a escolha dos classificadores. Foram somados os resultados de cada classificador que obtiveram índice Kappa maior que 0,5. Tanto o MLP como o *Random Forest* alcançaram 10 (dez) resultados do Kappa maiores que 0,5. Próximo a eles, o SVM alcançou 9 (nove). Já o comitê 1, KNN e *Random Tree* obtiveram 7 (sete), 5 (cinco) e 1 (um), respectivamente.

Depois dessa análise, por atingirem os melhores resultados, foram escolhidos os classificadores MLP, SVM e *Random Forest*. Em seguida, todos os resultados maiores que 0,5 foram ordenados em ordem decrescente e apenas os 5 (cinco) primeiros foram utilizados para a criação do descritor híbrido, que consiste na concatenação de todos os atributos dos cinco descritores escolhidos.

Outro ponto importante que deve ser observado é que após essa formação do descritor híbrido, foi realizada uma seleção de atributos com o algoritmo *Gain Ratio Information* (QUINLAN, 1986), ou Razão de Ganho de Informação, utilizando o método *Ranker*, que classifica os atributos por suas avaliações individuais.

Depois dos atributos classificados, foram realizados testes para identificar o melhor subconjunto de características (atributos). Utilizamos vetores de características com comprimentos variando de 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% e 100% dos atributos mais significativos de acordo com a seleção de atributos. Um total de 90 testes foram realizados (15 testes de porcentagens de seleção de atributos \times 6 descritores — os 5 escolhidos e o descritor híbrido) para cada um dos três classificadores.

Com o classificador SVM, o descritor híbrido foi formado a partir das CNNs Vgg-f ($K = 0,6098$), Vgg-m ($K = 0,5884$), CaffeNet ($K = 0,5864$), Vgg-verydeep-19 ($K = 0,5619$) e Vgg-s ($K = 0,5525$), pois apresentaram as cinco melhores taxas. Seus resultados podem ser visualizados no gráfico da Figura 16. Os resultados foram bastante variados para os descritores em relação às porcentagens de atributos selecionados.

Pode-se observar que o melhor resultado foi obtido com o descritor Vgg-f utilizando 100% dos seus atributos, onde atingiu uma acurácia de 83,2%. Já com o descritor CaffeNet utilizando somente 3% dos seus atributos foi obtida uma acurácia de 81,7%.

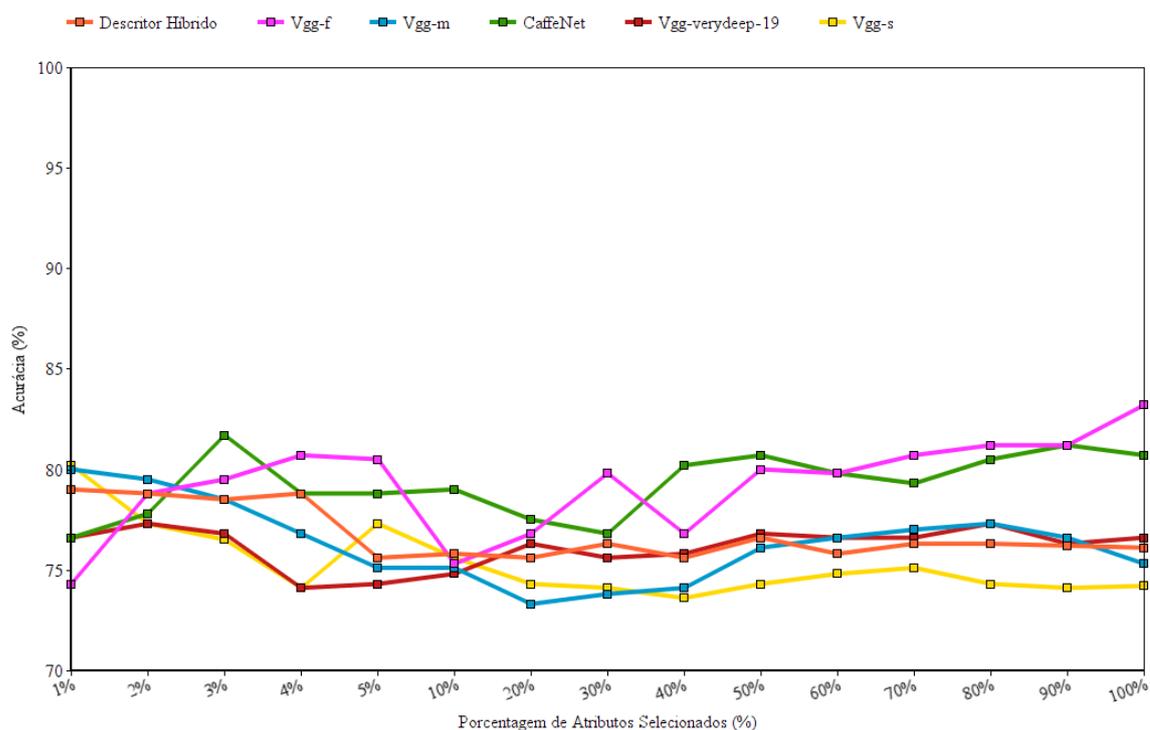


Figura 16 – Resultados da variação da porcentagem de atributos selecionados para o classificador SVM.

Somente os descritores Vgg-s e Vgg-m obtiveram os melhores resultados com a utilização de somente 1% de seus atributos mais relevantes, obtendo uma acurácia de 80,2% e 80,0%, respectivamente. E o descritor Vgg-verydeep-19 teve sua melhor taxa de acerto com uma acurácia de 77,3% utilizando 2% dos atributos mais relevantes.

Em relação ao descritor híbrido, não foram obtidos valores de acurácia promissores. O maior valor obteve somente 78,8% casos de sucesso utilizando 4% dos atributos mais relevantes. Isso indica que a combinação desses descritores para a formação do descritor híbrido não obteve relevância para que o classificador SVM pudesse distinguir melhor as duas classes: melanoma e não-melanoma.

Com o classificador *Random Forest*, o descritor híbrido foi formado a partir dos descritores Vgg-m ($K = 0,5855$), GLRLM ($K = 0,5845$), Vgg-s ($K = 0,5697$), Vgg-verydeep-16 ($K = 0,5679$) e Vgg-verydeep-19 ($K = 0,5577$), pois apresentaram as cinco melhores taxas. A Figura 17 mostra o gráfico da variação da porcentagem de atributos, apresentando a acurácia para cada um dos descritores utilizados. Já o *Random Forest* não apresentou tanta variação, os resultados foram bastantes próximos uns dos outros, com exceção de alguns do descritor híbrido e da GLRLM. Além disso, o *Random Forest* apresentou melhores resultados.

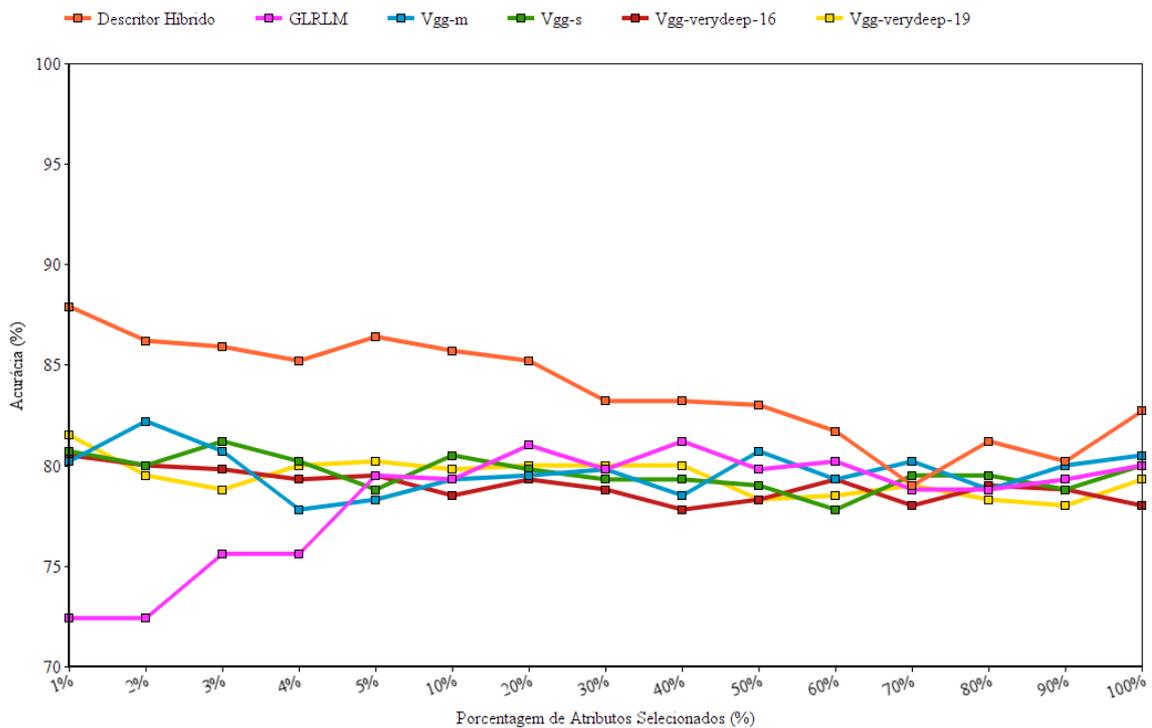


Figura 17 – Resultados da variação da porcentagem de atributos selecionados para o classificador *Random Forest*.

O descritor híbrido se mostrou o melhor descritor se comparado ao restante (GLRLM, Vgg-m, Vgg-s, Vgg-verydeep-16 e Vgg-verydeep-19), onde atingiu uma acurácia igual a 87,9% com a utilização de somente 1% dos atributos mais revelantes. Com exceção da utilização de 70% dos atributos onde o descritor Vgg-m obteve uma acurácia maior, o descritor híbrido alcançou todos os melhores resultados para todo o restante da variação de porcentagens de atributos.

O GLRLM obteve o pior resultado com uma acurácia de 72,4% utilizando 1% dos atributos. Somente com 40% dos atributos foi que esse descritor obteve uma maior taxa de acerto igual à 81,2%. O descritor Vgg-s também atingiu uma acurácia de 81,2%, contudo

a porcentagem de atributos selecionados foi igual a 3%.

O Vgg-m com apenas 2% dos atributos selecionados alcançou uma acurácia (82,2%) maior do que o GLRLM e Vgg-s, e do que os descritores Vgg-verydeep-16 (80,5%) e Vgg-verydeep-19 (81,5%), ambos com 1% dos atributos mais relevantes selecionados.

Finalmente, com o classificador MLP, o descritor híbrido foi formado a partir dos descritores CaffeNet ($K = 0,5987$), Vgg-m ($K = 0,5724$), Vgg-verydeep-19 ($K = 0,5494$), ABCD ($K = 0,5400$) e Vgg-f ($K = 0,5316$). Seus resultados em relação à variação de porcentagens de atributos selecionados podem ser visualizados no gráfico da Figura 18.

É importante ressaltar que foi com o classificador MLP em que obtivemos os melhores resultados para a diferenciação entre os dois tipos de lesão: melanoma e benigna. Assim como no classificador *Random Forest*, o descritor híbrido apresentou novamente as maiores taxas. Com 5% dos atributos mais relevantes selecionados, foi possível atingir uma acurácia de 94,9%, precisão e sensibilidade de 94,8%, e um índice Kappa considerado “Excelente” de 0,8916.

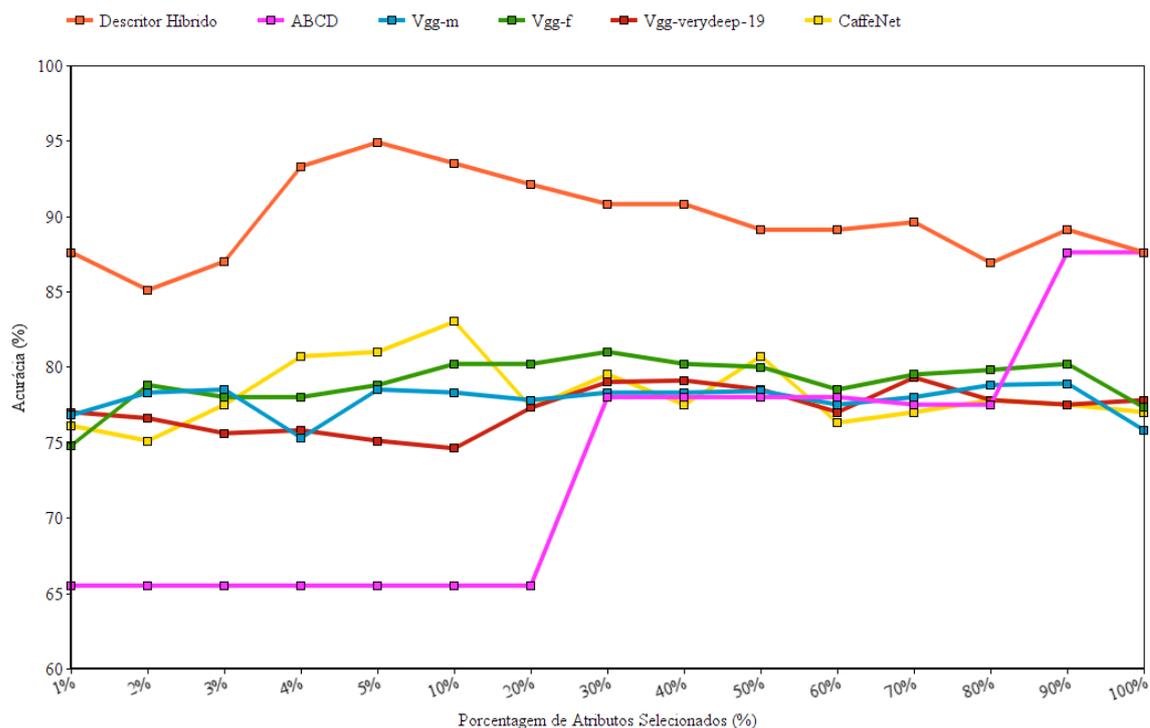


Figura 18 – Resultados da variação da porcentagem de atributos selecionados para o classificador MLP.

Os descritores Vgg-m, Vgg-f, Vgg-verydeep-19 e CaffeNet apresentaram resultados muito próximos entre si. O melhor resultado para o Vgg-m foi uma acurácia de 78,9%

com a utilização de 90% de seus atributos mais relevantes. O melhor para o descritor Vgg-f foi uma acurácia de 81,0% com a utilização de 30% de seus atributos mais relevantes. Já os descritores Vgg-verydeep-19 e CaffeNet obtiveram, respectivamente, uma acurácia de 79,3% (com 70% dos atributos selecionados) e 83,0% (com 10% dos atributos selecionados).

Com o descritor regra ABCD os valores quase não obtiveram diferença na variação da porcentagem de atributos selecionados, já que possui somente 5 atributos. O melhor resultado foi obtido com o uso de todos os seus atributos, onde atingiu uma acurácia igual 87,6%.

Apesar dos classificadores SVM e *Random Forest* também retornaram bons resultados, ao analisar todos os resultados que foram apresentados para as imagens segmentadas por especialistas, é possível constatar que o descritor híbrido atingiu a maior taxa de acerto com a utilização do classificador MLP, sendo assim o método proposto.

Uma segmentação ótima, no entanto, ainda é um desafio no desenvolvimento de sistemas CAD. Assim, com o objetivo de analisar a robustez do método proposto mesmo quando há erros de segmentação, foram realizados testes com dois tipos de erros de segmentação: *under-segmentation* e *over-segmentation*. Na subseção seguinte, são apresentados os resultados obtidos nesses testes.

5.3 Resultados com Imagens com Erros de Segmentação

A segmentação desempenha um papel importante em uma ampla gama de aplicações de visão por computador. Nos sistemas CAD, por exemplo, a busca por uma segmentação que seja a mais próxima da verdade terrestre (feita por especialistas) é um problema que ainda é muito pesquisado.

Como nesse trabalho o foco é a etapa de extração de características e classificação, dois erros de segmentação foram simulados nas imagens de lesão de pele. Dado que o objetivo é verificar a robustez do método proposto na ausência de uma boa segmentação.

Os erros de segmentação utilizados foram a *under-segmentation* e a *over-segmentation*. A Figura 19 exemplifica como eles foram obtidos. A partir da verdade terrestre da Figura 19(b) ilustrada pelo contorno de cor preta, para se obter o erro de *under-segmentation*, a área foi diminuída em 10%, sem modificar a borda. O resultado pode ser visto na Figura 19(c), ilustrado pelo contorno de cor verde.

No erro de *over-segmentation*, a partir da verdade terrestre da Figura 19(b) ilustrada pelo contorno de cor preta, a área foi aumentada em 10%, sem modificar a borda. O resultado pode ser visto na Figura 19(d), ilustrado pelo contorno de cor azul.

Aplicou-se o valor de 10% como erro absoluto aceito nesse estudo, tendo como base uma margem em que esses tipos de erros (aumentar ou diminuir a área segmentada)

fossem de ocorrência mais comum. Além disso, [Sousa et al. \(2017\)](#), em seu estudo sobre a análise de textura da região da imagem do disco óptico para diagnosticar o glaucoma, utilizou este mesmo valor.

A [Tabela 8](#) apresenta os resultados do descritor híbrido e o classificador MLP com os dois erros de segmentação nas imagens da união das bases PH² e DermIS.

O erro de *under-segmentation* atingiu uma acurácia igual a 87,0%, com precisão e sensibilidade de 87,3% e 86,9%, respectivamente. O índice Kappa classificou o resultado como “Muito Bom” de acordo com a tabela que mostra os níveis de exatidão.

O erro de *over-segmentation* apresentou uma maior taxa de acerto com acurácia igual a 90,8%, com precisão e sensibilidade de mesmo valor (90,9%). O índice Kappa classificou o resultado como “Excelente” de acordo com a [Tabela 2](#) que mostra os níveis de exatidão.

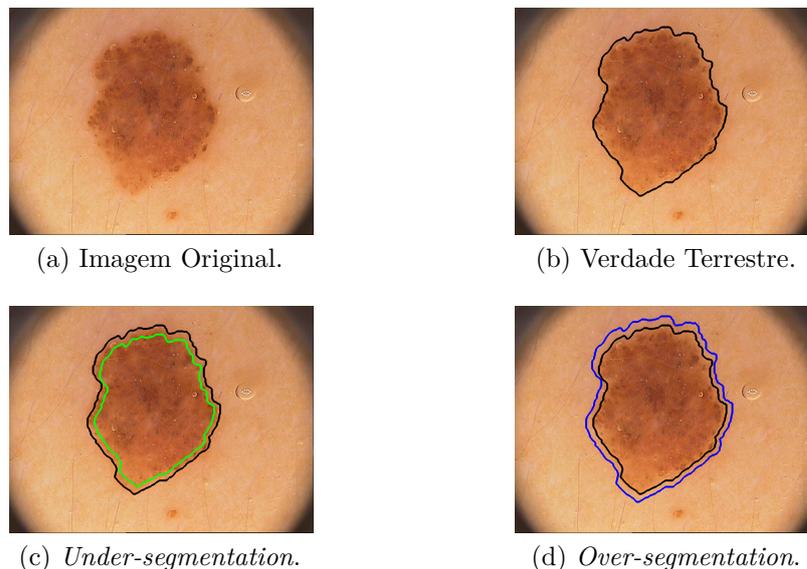


Figura 19 – Exemplo de erro de *under-segmentation* e *over-segmentation*.

Tabela 8 – Resultados do descritor híbrido com o classificador MLP para as imagens da união das bases PH² + DermIS com erro de segmentação.

Base de Imagens PH ² + DermIS				
Erro de Segmentação	P	S	K	A
<i>Under</i>	87,3	86,9	0,7249	87,0
<i>Over</i>	90,9	90,9	0,8094	90,8

Apesar dos resultados obtidos com o erro de *under-segmentation* não apresentarem muita diferença em relação aos da *over-segmentation*; eles obtiveram diferentes classificações em relação ao índice Kappa. A *under-segmentation* ocasionou ao descritor e classificador uma maior dificuldade para a diferenciação entre as duas classes do problema, uma vez que com este erro parte da lesão é desconsiderada.

5.4 Considerações Finais

Os resultados mostrados exemplificam o quão importante é a análise de diferentes cenários para a extração de características de lesões de pele. A união das bases PH² e DermIS proporcionou uma maior diversidade de imagens. Essa diversidade muitas vezes dificulta a distinção entre o melanoma e não-melanoma para o descritor e o classificador.

Foi feita uma análise comparativa utilizando diferentes descritores e classificadores. Tal análise foi realizada com o objetivo de justificar a escolha dos algoritmos utilizados no método proposto. Foram testados além dos descritores ABCD, GLCM, GLRLM, HOG, LBP, Tamura, *Box-Counting*, AlexNet, CaffeNet, Vgg-f, Vgg-m, Vgg-s, Vgg-verydeep-16 e Vgg-verydeep-19, os classificadores MLP, SVM, *Random Tree*, KNN, Comitê (MLP + SVM + KNN) e *Random Forest*. Obteve-se, então, a melhor combinação do classificador com o descritor.

No primeiro cenário, onde as imagens das bases não possuíam nenhuma segmentação e nenhum pré-processamento, foi encontrada uma acurácia, precisão e sensibilidade de 80,3%, obtendo um desempenho “Bom” com um índice Kappa igual a 0,5856. O descritor utilizado foi o Vgg-f com o Comitê formado pelos classificadores MLP, SVM e KNN. Entretanto, com o classificador SVM, a precisão (80,6%) e o índice Kappa (0,5863) obtiveram maiores valores com o descritor AlexNet.

Em relação aos demais classificadores MLP, *Random Tree*, KNN e *Random Forest*, os resultados apresentados foram bastante próximos entre si. O MLP teve como melhor resultado uma acurácia de 78,3% com o descritor HOG. O *Random Tree* teve como melhor resultado uma acurácia de 76,3% com o descritor Vgg-verydeep-19, assim como o KNN com o descritor Vgg-m.

Nesse mesmo cenário, é importante lembrar que as características ABCD não foram calculadas, já que não puderam ser calculadas, devido as imagens não serem segmentadas.

No segundo cenário, as imagens foram segmentadas através das máscaras binárias feitas por especialistas e fornecidas pelas bases. Inicialmente, foram realizados testes com todos os descritores e classificadores analisados. A partir disso, pode-se perceber que os resultados com as imagens segmentadas por especialistas alcançou melhores resultados na classificação do que com as imagens do primeiro cenário (imagens sem segmentação).

Os melhores resultados foram obtidos com os classificadores SVM, *Random Forest* e MLP. Isso porque foi realizada uma soma dos resultados de cada classificador que obtiveram o índice Kappa maior que 0,5 (classificação considerada “Boa”). De forma geral, após essa escolha dos melhores classificadores, foram realizados testes com diferentes variações de porcentagem de atributos dos descritores.

O melhor resultado encontrado para o classificador SVM foi bastante promissor,

com uma acurácia de 83,2% utilizando o descritor Vgg-f para a extração de características com 100% dos seus atributos. Já o descritor híbrido obteve apenas uma acurácia de 78,8% utilizando 4% dos seus atributos mais relevantes.

Com o classificador *Random Forest*, obtivemos melhores resultados do que com o classificador SVM. A melhor acurácia alcançada foi de 87,9% com a utilização de somente 1% dos atributos mais relevantes do descritor híbrido. Em relação aos outros descritores, os resultados de acordo com a variação de porcentagem dos atributos obtiveram valores de acurácia bem próximos.

Por fim, o classificador MLP e o descritor híbrido, que compõem o método proposto, obtiveram o melhor resultado de classificação, atingindo uma acurácia igual a 94,9% utilizando 5% de seus atributos mais relevantes selecionados.

A partir da escolha da melhor combinação descritor e classificador e a quantidade de atributos mais relevantes para serem utilizados, surge o terceiro cenário, as imagens com erros de *under-segmentation* e *over-segmentation*. O objetivo desse teste, é verificar quão robusto é o método proposto, mesmo quando há erros de segmentação.

Os resultados obtidos mostram que os dois erros de segmentação apresentam pequenas diferenças em relação as taxas de acerto obtidas no método proposto. Entretanto, os erros de *under-segmentation* obtiveram piores resultados que o de *over-segmentation*.

Em relação aos resultados obtidos com a segmentação das lesões que atingiram um índice Kappa “Excelente” para a união das bases PH² e DermIS, os erros de *under-segmentation* e *over-segmentation* alcançaram um índice Kappa com desempenho “Muito Bom” e “Excelente”, respectivamente. Isso mostra que mesmo quando ocorrer erros de segmentação, o método proposto ainda obtém resultados bastante promissores.

A [Tabela 9](#) apresenta um resumo comparativo com os resultados obtidos de nossa abordagem e os principais métodos do estado da arte. Os resultados mostram que nossa abordagem obteve maiores taxas de sucesso do que métodos de trabalhos anteriores: [Araujo et al. \(2012\)](#) (A = 92,0%), [Chang et al. \(2013\)](#) (A = 90,64%), [Codella et al. \(2015\)](#) (A = 93,1%), [Bhati e Singhal \(2015\)](#) (P = 84,61% e S = 92,3%), [Kasmi e Mokrani \(2016\)](#) (A = 94,0%), [Sánchez-Monedero et al. \(2016\)](#) (A = 68,51%) e [Oliveira et al. \(2016\)](#) (A = 74,33%).

O trabalho de [Cavalcanti, Scharcanski e Baranoski \(2013\)](#) (A = 98,66%) apresentou melhor desempenho do que o nosso método proposto (A = 94,9%). Uma razão para a diferença no desempenho é a quantidade e a diversidade das imagens usadas no teste. Nosso método foi avaliado usando 406 imagens de dois bancos de dados diferentes e o método de [Cavalcanti, Scharcanski e Baranoski \(2013\)](#) foi avaliado em 152 imagens de um único banco de dados. Além disso, analisando a [Tabela 9](#) identificamos que entre os vários métodos de diagnóstico de lesões cutâneas, a regra ABCD tem sido comumente aplicada

para extrair características. No entanto, esse descritor pode ser calculado de diversas maneiras. Assim, uma diretriz a ser seguida é o teste de diferentes tipos de descritores ABCD para a extração de características.

Outras observações precisam ser apontadas. No que diz respeito ao trabalho de [Codella et al. \(2015\)](#), [Chang et al. \(2013\)](#) e [Sánchez-Monedero et al. \(2016\)](#) os métodos foram avaliados usando mais imagens do que as nossas. No entanto, todas as imagens pertenciam ao mesmo banco de dados, enquanto avaliamos imagens de dois bancos de dados. Isso levou a um maior grau de diversidade entre as imagens e sugere que a abordagem proposta seja mais confiável.

Já no trabalho de [Oliveira et al. \(2016\)](#) foram utilizadas imagens de 7 bases de imagens diferentes, atingindo 74,33% de acurácia. Enquanto o método proposto foi avaliado em 2 bases de imagens diferentes, obtendo uma acurácia de 94,9%. Embora o número de bases seja diferente (7×2), é importante ressaltar que o número total de imagens é semelhante (408 contra 406) e que existe um cruzamento de mais da metade dessas imagens, desde que [Oliveira et al. \(2016\)](#) também utilizou a base DermIS (206 imagens).

No próximo capítulo tem-se as conclusões obtidas bem como os trabalhos futuros que podem ser realizados a partir do método proposto.

Tabela 9 – Comparação do método proposto com o estado da arte.

Trabalhos	Descritores	Classificadores	Número de Imagens	Métricas de Avaliação
Araujo et al. (2012)	ABCD	-	40 (2 bases de imagens)	A = 92,0%
Cavalcanti, Scharcanski e Baranoski (2013)	ABCD	KNN	152 (1 base de imagens)	A = 98,66%
Chang et al. (2013)	GLRLM e Tamura	SVM	729 (1 base de imagens)	A = 90,64%
Codella et al. (2015)	<i>Deep Learning</i> e <i>Sparse Coding</i>	SVM	2624 (1 base de imagens)	A = 93,1%
Bhati e Singhal (2015)	ABCD	TDS	30 (1 base de imagens*)	P = 84,61% S = 92,3%
Kasmi e Mokrani (2016)	ABCD	TDS	200 (1 base de imagens)	A = 94,0%
Sánchez-Monedero et al. (2016)	ABCD, GLCM, LBP e MRF	LIPU	556 (1 base de imagens)	A = 68,51%
Oliveira et al. (2016)	ABCD e <i>Box-Counting</i>	SVM	408 (7 bases de imagens)	A = 74,33%
Método Proposto	ABCD, Vgg-m Vgg-f, Vgg-verydeep-19 e CaffeNet	MLP	406 (2 bases de imagens)	A = 94,9% P = 94,8% S = 94,8%

*Algumas imagens são coletadas da internet.

6 Conclusão e Trabalhos Futuros

Este trabalho trata da identificação de melanoma em imagens dermatoscópicas com o objetivo de auxiliar o dermatologista no diagnóstico das lesões de pele. Para isso, foi proposto um método de visão computacional, que utiliza um descritor híbrido (formado a partir da combinação de vários descritores e da seleção de atributos) para a extração de características de lesões de pele e o algoritmo MLP para a classificação.

6.1 Conclusão

Este estudo mostra a possibilidade de utilizar técnicas de inteligência computacional para a diferenciação de lesões de pele. A principal contribuição está no fato de considerar que as informações obtidas pelo método podem ser disponibilizadas ao dermatologista com o objetivo de auxiliá-lo no diagnóstico das lesões de pele. Além disso, apresentou-se uma avaliação de diferentes descritores e classificadores.

Em relação à seleção de atributos foram realizados vários testes variando a porcentagem de atributos selecionados. Desse modo, obteve-se o melhor resultado utilizando apenas 5% dos atributos mais relevantes do descritor híbrido. Além disso, foram testados diferentes classificadores e combinações de algoritmos para a formação do descritor híbrido.

A precisão, a sensibilidade, o índice Kappa e a acurácia foram as métricas utilizadas nesse trabalho. Elas são apropriadas para a avaliação dos resultados de cada experimento e teste que foi realizado. Isto porque avaliam o desempenho dos algoritmos, retornando bons resultados somente quando a classificação teve boas taxas de acerto.

O descritor híbrido obteve resultados promissores para o problema em questão com o classificador MLP, obtendo uma acurácia de 94,9% e índice Kappa considerado “Excelente” em 406 imagens de duas bases de dados diferentes: PH² e DermIS. Isso significa que os vetores de atributos computados por ele constitui um bom descritor para o classificador conseguir separar as duas classes.

Nos cenários das imagens com erros de *under-segmentation* e *over-segmentation*, o método proposto proporcionou resultados animadores na classificação dos testes que foram realizados. Desse modo, com o aprimoramento das técnicas utilizadas para solucionar os problemas encontrados, como a complexidade da base de imagens DermIS, o método desenvolvido pode obter resultados ainda mais expressivos dos que foram apresentados.

6.2 Trabalhos Futuros

Os resultados apresentados são promissores, mas podem ser melhorados. Para a continuidade, as seguintes tarefas podem ser analisadas para o desenvolvimento de métodos relacionados à detecção e classificação de imagens médicas de lesão de pele: utilizar técnicas de pré-processamento para melhorar as imagens das bases (como por exemplo a remoção de pelos); desenvolver uma técnica para a segmentação de lesões de pele; utilizar outros métodos para extrair as características referentes à regra ABCD; realizar testes em outras bases de imagens de lesão de pele; e realizar testes com outros métodos para a seleção de atributos.

Referências

- ABBADENI, N.; ZIOU, D.; WANG, S. Perceptual textural features corresponding to human visual perception. In: *Proc. of the Thirteenth Vision Interface Conference, Montreal, Quebec (Canada)*. [S.l.: s.n.], 2000. p. 365–372. Citado na página 18.
- A.C.CAMARGO, H. *Tudo sobre o Câncer*. 2017. <<http://www.accamargo.org.br/>>. Acessado em: 2017. Citado 2 vezes nas páginas 1 e 2.
- AL-AKAIDI, M. *Fractal speech processing*. [S.l.]: Cambridge university press, 2004. Citado 2 vezes nas páginas 9 e 19.
- ARAUJO, A. F. et al. Análise e caracterização de lesões de pele para auxílio ao diagnóstico médico. *Avanços em Visão Computacional*, 2012. Citado 6 vezes nas páginas 7, 10, 13, 14, 46 e 48.
- ARGENZIANO, G. et al. Interactive atlas of dermoscopy (book and cd-rom). EDRA Medical Publishing & New Media, 2000. Disponível em: <<http://www.dermoscopy.org/atlas/default.asp>>. Citado 2 vezes nas páginas 8 e 9.
- BARALDI, A.; PARMIGGIANI, F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, v. 33, n. 2, p. 293–304, 1995. Citado na página 14.
- BARBOSA, B. H. G. *Computação Evolucionária e Máquinas de Comitê na Identificação de Sistemas Não-Lineares*. [S.l.], 2009. 197 f. Tese (Doutorado em Engenharia Elétrica) — Programa de Pós-Graduação em Engenharia Elétrica, UFMG, Belo Horizonte. Citado na página 25.
- BARCELOS, C. A. Z.; BOAVENTURA, M.; SILVA, E. A well-balanced flow equation for noise removal and edge detection. *IEEE Transactions on Image Processing*, IEEE, v. 12, n. 7, p. 751–763, 2003. Citado na página 3.
- BATISTA, G. E. de A. P. A. *Pré-processamento de Dados em Aprendizado de Máquina Supervisionado*. [S.l.], 2003. 232 f. Tese (Doutorado em Ciências da Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, USP, São Carlos. Citado na página 25.
- BERNARDINI, F. C. *Combinação de classificadores simbólicos utilizando medidas de regras de conhecimento e algoritmos genéticos*. [S.l.], 2006. 210 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, USP, São Carlos. Citado na página 24.
- BHATI, P.; SINGHAL, M. Early stage detection and classification of melanoma. In: *IEEE. Communication, Control and Intelligent Systems (CCIS), 2015*. [S.l.], 2015. p. 181–185. Citado 4 vezes nas páginas 8, 10, 46 e 48.
- BREIMAN, L. Random forests. *Machine learning*, v. 45, n. 1, p. 5–32, 2001. Citado na página 25.

- CASELLES, V.; KIMMEL, R.; SAPIRO, G. Geodesic active contours. *International journal of computer vision*, Springer, v. 22, n. 1, p. 61–79, 1997. Citado na página 8.
- CASTELLUCCIO, M. et al. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*, 2015. Citado na página 20.
- CAVALCANTI, P. G.; SCHARCANSKI, J. Automated prescreening of pigmented skin lesions using standard cameras. *Computerized Medical Imaging and Graphics*, Elsevier, v. 35, n. 6, p. 481–491, 2011. Citado na página 7.
- CAVALCANTI, P. G.; SCHARCANSKI, J.; BARANOSKI, G. V. A two-stage approach for discriminating melanocytic skin lesions using standard cameras. *Expert Systems with Applications*, Elsevier, v. 40, n. 10, p. 4054–4064, 2013. Citado 4 vezes nas páginas 7, 10, 46 e 48.
- CHAN, T. F.; VESE, L. A. Active contours without edges. *IEEE Transactions on Image Processing*, IEEE, v. 10, n. 2, p. 266–277, 2001. Citado na página 9.
- CHANG, W.-Y. et al. Computer-aided diagnosis of skin lesions using conventional digital photography: a reliability and feasibility study. *PloS One*, Public Library of Science, v. 8, n. 11, p. e76212, 2013. Citado 5 vezes nas páginas 6, 10, 46, 47 e 48.
- CHATFIELD, K. et al. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014. Citado na página 20.
- CHU, A.; SEHGAL, C. M.; GREENLEAF, J. F. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, Elsevier, v. 11, n. 6, p. 415–419, 1990. Citado na página 16.
- CODELLA, N. et al. Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In: SPRINGER. *International Workshop on Machine Learning in Medical Imaging*. [S.l.], 2015. p. 118–126. Citado 5 vezes nas páginas 7, 10, 46, 47 e 48.
- COHEN, B. A.; LEHMANN, C. U. *Johns Hopkins University - DermAtlas, Dermatology Image Atlas*. 2012. <<http://dermatlas.med.jhmi.edu/derm>>. Citado 2 vezes nas páginas 7 e 9.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine learning*, Springer, v. 20, n. 3, p. 273–297, 1995. Citado na página 6.
- COVER, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, IEEE, n. 3, p. 326–334, 1965. Citado na página 23.
- DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE. *Conference on Computer Vision and Pattern Recognition*. [S.l.], 2005. v. 1, p. 886–893. Citado na página 17.
- DASARATHYAND, B. R.; HOLDER, E. B. Image characterizations based on joint gray-level run-length distributions. *Pattern Recognition Letters*, v. 12, n. 8, p. 497–502, 1991. Citado na página 16.

- DERMATOLOGIA, S. B. de. *Câncer da pele*. 2017. <<http://www.sbd.org.br/doenca/cancer-da-pele/>>. Acessado em: 2017. Citado na página 3.
- DERMATOLOGIA.NET. *Melanoma – O câncer de pele mais perigoso*. 2017. <<http://www.dermatologia.net/cat-doencas-da-pele/melanoma-maligno/>>. Acessado em: 2017. Citado 2 vezes nas páginas 9 e 2.
- DIEPGEN, T. L.; YIHUNE, G. *Dermatology Information System – DermIS*. 2012. <<http://dermis.net/>>. Acessado em: 2017. Citado 2 vezes nas páginas 9 e 11.
- FIX, E.; JR, J. L. H. *Discriminatory analysis-nonparametric discrimination: consistency properties*. [S.l.], 1951. Citado 2 vezes nas páginas 8 e 23.
- FORNACIALI, M. et al. Statistical learning approach for robust melanoma screening. In: IEEE. *Conference on Graphics, Patterns and Images (27th SIBGRAPI)*. [S.l.], 2014. p. 319–326. Citado 2 vezes nas páginas 6 e 10.
- GABOR, D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, IET, v. 93, n. 26, p. 429–441, 1946. Citado na página 8.
- GALLOWAY, M. M. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, Elsevier, v. 4, n. 2, p. 172–179, 1975. Citado 2 vezes nas páginas 6 e 16.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. [S.l.]: Prentice hall New Jersey, 2002. 793 p. p. Citado na página 29.
- GOSHTASBY, A. *Segmentation of skin cancer images*. 2012. <<https://engineering-computer-science.wright.edu/>>. Citado na página 7.
- GUIDE, S. C. *Melanoma*. 2012. <http://www.skincancerguide.ca/melanoma/images/melanoma_images.html>. Citado na página 9.
- GUTIÉRREZ, P. A.; HERVÁS-MARTÍNEZ, C.; MARTÍNEZ-ESTUDILLO, F. J. Logistic regression by means of evolutionary radial basis function neural networks. *IEEE Transactions on Neural Networks*, v. 22, n. 2, p. 246–263, 2011. Citado na página 9.
- HALL, M. et al. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, ACM, v. 11, n. 1, p. 10–18, 2009. Citado na página 33.
- HALPERN, A. et al. *International Skin Imaging Collaboration*. 2015. <<https://isic-archive.com/>>. Citado na página 7.
- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, v. 143, n. 1, p. 29–36, 1982. Citado na página 7.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Ieee, v. 3, n. 6, p. 610–621, 1973. Citado na página 9.
- HARRIS, C.; STEPHENS, M. A combined corner and edge detector. In: CITESEER. *Alvey vision conference*. [S.l.], 1988. v. 15, n. 50, p. 10–5244. Citado na página 8.

- HAYKIN, S. S. *Neural networks and learning machines*. [S.l.]: Pearson Upper Saddle River, NJ, USA.; 2009. v. 3. Citado 2 vezes nas páginas 4 e 21.
- INCA. *Estatísticas do Câncer*. 2017. <<http://www1.inca.gov.br/vigilancia/>>. Acessado em: 2017. Citado na página 1.
- JIA, Y. et al. Caffe: Convolutional architecture for fast feature embedding. In: ACM. *Proceedings of the 22nd ACM international conference on Multimedia*. [S.l.], 2014. p. 675–678. Citado 2 vezes nas páginas 7 e 20.
- JUNIOR, J. L. C. *Dermnet skin disease image atlas*. 2009. <<http://www.dermnet.com/>>. Citado 2 vezes nas páginas 8 e 9.
- JUNIOR, M. P. P. Combining classifiers: from the creation of ensembles to the decision fusion. In: IEEE. *Conference on Graphics, Patterns and Images Tutoriais (24th SIBGRAPI)*. [S.l.], 2011. p. 1–10. Citado na página 24.
- KASMI, R.; MOKRANI, K. Classification of malignant melanoma and benign skin lesions: implementation of automatic abcd rule. *IET Image Processing*, IET, v. 10, n. 6, p. 448–455, 2016. Citado 4 vezes nas páginas 8, 10, 46 e 48.
- KASS, M.; WITKIN, A.; TERZOPOULOS, D. Snakes: Active contour models. *International journal of computer vision*, Springer, v. 1, n. 4, p. 321–331, 1988. Citado na página 7.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F. et al. (Ed.). *Advances in Neural Information Processing Systems 25*. [S.l.]: Curran Associates, Inc., 2012. p. 1097–1105. Citado na página 20.
- KUMAR, A. et al. An ensemble of fine-tuned convolutional neural networks for medical image classification. *IEEE Journal of Biomedical and Health Informatics*, IEEE, 2016. Citado na página 31.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, JSTOR, p. 159–174, 1977. Citado na página 27.
- LI, S. Z. *Markov random field modeling in image analysis*. [S.l.]: Springer Science & Business Media, 2009. Citado na página 9.
- LIMA, R. A. F. *Estratégias de Seleção de Atributos para Detecção de Anomalias em Transações Eletrônicas*. [S.l.], 2016. 94 f. Dissertação (Mestrado em Ciência da Computação) — Programa de Pós-Graduação em Ciência da Computação, UFMG, Belo Horizonte. Citado na página 26.
- MAIRAL, J. et al. Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, Now Publishers, Inc., v. 8, n. 2-3, p. 85–283, 2014. Citado na página 7.
- MELTON, J. L.; SWANSON, J. R. *Loyola University Dermatology Medical Education, Skin Cancer and Benign Tumor Image Atlas*. 2012. <<http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/content1.html>>. Citado na página 9.

- MENDONÇA, T. et al. Ph 2-a dermoscopic image database for research and benchmarking. In: IEEE. *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. [S.l.], 2013. p. 5437–5440. Citado na página 11.
- MOURA, N. H. et al. Classificação de lesões de pele utilizando atributos de cor e textura. *Anais do XV Workshop de Informática Médica – WIM*, 2015. Citado na página 30.
- NEVES, R. de Cássia David das. *Pré-processamento no processo de descoberta de conhecimento em banco de dados*. [S.l.], 2003. 137 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre. Citado na página 25.
- OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, Elsevier, v. 29, n. 1, p. 51–59, 1996. Citado 2 vezes nas páginas 9 e 18.
- OLIVEIRA, R. B. *Método de detecção e classificação de lesões de pele em imagens digitais a partir do modelo Chan-Vese e máquina de vetor de suporte*. [S.l.], 2012. 134 f. Dissertação (Mestrado em Ciência da Computação) — Programa de Pós-Graduação em Ciência da Computação, Área de Concentração – Computação Aplicada, do Instituto de Biociências, Letras e Ciências Exatas, UNESP, São José do Rio Preto. Citado na página 19.
- OLIVEIRA, R. B. et al. A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications*, Elsevier, v. 61, p. 53–63, 2016. Citado 5 vezes nas páginas 9, 10, 46, 47 e 48.
- OTSU, N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, IEEE, v. 9, n. 1, p. 62–66, 1979. Citado na página 8.
- POWERS, D. Evaluation: from precision, recall and f-factor to roc, informedness, markedness & correlation (technical report). *Adelaide, Australia*, 2007. Citado 2 vezes nas páginas 27 e 28.
- QUINLAN, J. R. Induction of decision trees. *Machine Learning*, Springer, v. 1, n. 1, p. 81–106, 1986. Citado 3 vezes nas páginas 4, 25 e 39.
- RODRIGUEZ, P. M. Processos de ramificação: teoria e aplicações. *Technical Report*, ICMC/USP, São Carlos, 2012. Citado na página 23.
- ROSENFELD, G. H.; FITZPATRICK-LINS, K. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing*, v. 52, n. 2, p. 223–227, 1986. Citado na página 27.
- SÁEZ, A. et al. Development and evaluation of perceptually adapted colour gradients. *IET Image Processing*, IET, v. 7, n. 4, p. 355–363, 2013. Citado na página 9.
- SAEZ, A.; SERRANO, C.; ACHA, B. Model-based classification methods of global patterns in dermoscopic images. *IEEE Transactions on Medical Imaging*, IEEE, v. 33, n. 5, p. 1137–1147, 2014. Citado na página 9.

- SÁNCHEZ-MONEDERO, J. et al. Classification of melanoma presence and thickness based on computational image analysis. In: SPRINGER. *International Conference on Hybrid Artificial Intelligence Systems*. [S.l.], 2016. p. 427–438. Citado 5 vezes nas páginas 9, 10, 46, 47 e 48.
- SCF. *Skin Cancer & Facts Statistics*. 2017. <<http://www.skincancer.org/skin-cancer-information/skin-cancer-facts>>. Acessado em: 2017. Citado na página 1.
- SHIN, H.-C.; ROTH, H. R. et al. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1285–1298, 2016. Citado na página 19.
- SILVA, E. R. J. *Investigação de técnicas de Extração e Seleção de Características e Classificadores Aplicados ao Problema de Classificação de Dígitos Manuscritos de Imagens de Documentos Históricos*. [S.l.], 2007. 51 f. Monografia (Bacharelado em Engenharia da Computação) — Escola Politécnica de Pernambuco – Universidade de Pernambuco, Pernambuco. Citado na página 24.
- SILVA, M. A. T. *Uso de matrizes de co-ocorrências para classificação automática de imagens de cana-de-açúcar*. [S.l.], 2013. 75 f. Monografia (Bacharelado em Ciência da Computação) — Universidade Federal de Lavras, Pernambuco. Citado na página 15.
- SILVA, R. R. V. *Detecção do Uso de Capacete por Motociclistas em Vias Públicas*. [S.l.], 2014. 81 f. Dissertação (Mestrado em Ciência da Computação) — Programa de Pós-Graduação em Ciência da Computação, UFPI, Teresina. Citado 4 vezes nas páginas 9, 18, 21 e 26.
- SILVA, R. R. V. e; AIRES, K. R. T.; VERAS, R. d. M. S. Helmet detection on motorcyclists using image descriptors and classifiers. In: IEEE. *Conference on Graphics, Patterns and Images (27th SIBGRAPI)*. [S.l.], 2014. p. 141–148. Citado na página 18.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 20.
- SIVIC, J.; ZISSERMAN, A. Video google: A text retrieval approach to object matching in videos. In: IEEE. *null*. [S.l.], 2003. p. 1470. Citado na página 6.
- SOBEL, I. *Camera models and machine perception*. [S.l.], 1970. Citado na página 12.
- SOUSA, J. A. de et al. Texture based on geostatistic for glaucoma diagnosis from fundus eye image. *Multimedia Tools and Applications*, Springer, v. 76, n. 18, p. 19173–19190, 2017. Citado na página 44.
- SURULIANDI, A.; MEENA, K.; ROSE, R. R. Local binary pattern and its derivatives for face recognition. *IET Computer Vision*, IET, v. 6, n. 5, p. 480–488, 2012. Citado na página 18.
- SUZUMURA, Y. *YSP Dermatology Image Database*. 2012. <http://homepage1.nifty.com/ysh/soft_e_ysp.htm>. Citado na página 9.

TAMURA, H.; MORI, S.; YAMAWAKI, T. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, IEEE, v. 8, n. 6, p. 460–473, 1978. Citado 2 vezes nas páginas 6 e 18.

THOMPSON, C. M.; SHURE, L. Image processing toolbox [for use with matlab]. The Math Works, 1995. Citado na página 33.

TOTAL, S. *Câncer da Pele: fotoproteção, Vida saudável com o sol*. 2012. <<http://www.saudetotal.com.br/prevencao/topicos/default.asp>>. Citado na página 9.

VERAS, R. M. S. et al. Aplicação de aprendizagem de máquina para classificação automática de retinas patológicas. *Anais do X Congresso Brasileiro de Inteligência Computacional*, 2011. Citado na página 1.

WHO. *How common is skin cancer?* 2017. <<http://www.who.int/uv/faq/skincancer/en/index1.html>>. Acessado em: 2017. Citado na página 1.

Anexos

ANEXO A – Testes realizados com as bases de imagens PH² e DermIS separadas

A [Tabela 10](#) apresenta os resultados obtidos pela classificação com as imagens sem segmentação da base PH². É possível perceber que o descritor CaffeNet alcançou a melhor taxa de acerto com o classificador SVM, atingindo uma acurácia de 92,5% e um índice Kappa de 0,7588 considerado “Muito Bom”. Com o classificador MLP, o melhor resultado também foi com a utilização do descritor CaffeNet, alcançando uma acurácia de 92,0% e Kappa de 0,7500. E com o classificador *Random Forest*, o descritor Tamura obteve melhor resultado com uma taxa de acerto igual a 90,0% e um Kappa de 0,6815.

Tabela 10 – Resultados da classificação com as imagens sem segmentação da Base PH².

Descritor	Base de Imagens PH ²											
	MLP				SVM				Random Forest			
	P	S	K	A	P	S	K	A	P	S	K	A
GLCM	86,2	87,0	0,5423	87,0	87,6	88,5	0,5194	88,5	85,2	86,0	0,5270	86,0
GLRLM	86,3	86,0	0,5706	86,0	83,6	76,0	0,3593	76,0	88,2	88,0	0,6319	88,0
HOG	84,1	85,0	0,4932	85,0	83,9	85,0	0,4828	85,0	85,1	86,0	0,4737	86,0
LBP	89,1	89,5	0,6557	89,5	83,5	84,5	0,3825	84,5	86,2	87,0	0,5423	87,0
Tamura	88,8	89,0	0,6497	89,0	86,2	87,0	0,5324	87,0	89,8	90,0	0,6815	90,0
Box-Counting	82,3	83,5	0,4369	83,5	83,8	84,5	0,1689	84,5	81,7	82,0	0,4268	82,0
AlexNet	85,8	86,0	0,5541	86,0	83,0	82,5	0,4681	82,5	88,1	88,5	0,5818	88,5
CaffeNet	92,0	92,0	0,7500	92,0	92,3	92,5	0,7588	92,5	88,8	88,0	0,5276	88,0
Vgg-f	89,6	89,5	0,6749	89,5	88,2	88,5	0,6302	88,5	89,1	89,5	0,6341	89,5
Vgg-m	87,2	87,5	0,5981	87,5	83,9	85,0	0,4828	85,0	83,2	84,5	0,4484	84,5
Vgg-s	87,0	87,5	0,5902	87,5	84,1	85,0	0,4932	85,1	84,4	85,5	0,4948	85,5
Vgg-verydeep-16	87,8	88,0	0,6178	88,0	87,5	88,0	0,6026	88,0	87,5	88,0	0,6026	88,0
Vgg-verydeep-19	88,5	89,0	0,6284	89,0	86,6	87,0	0,5779	87,0	89,6	90,0	0,6552	90,0

Os resultados da classificação com as imagens sem segmentação da base DermIS podem ser vistos na [Tabela 11](#). Pode-se observar que os resultados obtidos foram menores do que com a base PH². A melhor acurácia alcançada foi com a utilização da CNN Vgg-verydeep-16 e o classificador SVM, onde a taxa de acerto é igual a 83,9% e o índice Kappa igual a 0,6665. Já com o *Random Forest* e MLP essa taxa diminuiu para 78,1% e 75,7%, respectivamente, utilizando o descritor CaffeNet.

Tabela 11 – Resultados da classificação com as imagens sem segmentação da Base DermIS.

Base de Imagens DermIS												
Descritor	MLP				SVM				Random Forest			
	P	S	K	A	P	S	K	A	P	S	K	A
GLCM	65,7	66,0	0,2949	66,0	71,1	69,9	0,3399	69,9	72,6	72,8	0,4323	72,8
GLRLM	65,5	65,5	0,2925	65,5	77,6	74,8	0,4773	74,7	66,6	67,0	0,3063	66,9
HOG	64,2	63,6	0,2641	63,5	61,5	61,7	0,2116	61,6	64,1	64,6	0,2612	64,5
LBP	67,6	68,0	0,3310	67,9	83,3	80,6	0,4929	80,5	58,1	59,2	0,1321	59,2
Tamura	68,1	68,4	0,3421	68,4	65,4	65,0	0,2739	65,0	67,8	68,0	0,3393	67,9
Box-Counting	67,6	68,0	0,3331	67,9	65,4	65,0	0,2739	65,0	66,7	67,0	0,2974	66,9
AlexNet	72,3	72,3	0,4320	72,3	73,4	73,3	0,4553	73,3	73,1	73,3	0,4434	73,3
CaffeNet	76,0	75,7	0,5071	75,7	73,3	73,3	0,4520	73,3	78,1	78,2	0,5502	78,1
Vgg-f	72,9	72,8	0,4446	72,8	74,8	74,8	0,4827	74,7	71,6	71,8	0,4139	71,8
Vgg-m	69,3	69,4	0,3703	69,4	69,0	68,9	0,3652	68,9	73,2	73,3	0,4398	73,3
Vgg-s	71,2	71,4	0,4103	71,3	69,6	69,4	0,3761	69,4	71,3	71,4	0,4121	71,3
Vgg-verydeep-16	71,5	71,4	0,4157	71,3	84,1	84,0	0,6665	83,9	73,2	73,3	0,4381	73,3
Vgg-verydeep-19	70,4	69,9	0,3907	70,0	69,9	69,9	0,3832	69,9	70,7	70,9	0,3860	70,8

A Tabela 12 mostra os resultados da classificação para cada um dos descritores utilizados na base PH² com as imagens segmentadas. Pode-se observar que o melhor resultado foi obtido com o descritor ABCD e o classificador SVM, onde atingiu uma acurácia de 92,5% e um índice Kappa igual a 0,7492, considerado “Muito Bom”. Com os classificadores MLP e *Random Forest*, os melhores resultados foram obtidos com o descritor ABCD e a CNN Vgg-verydeep-16, respectivamente. Ambos atingiram uma acurácia de 91,5% e índice Kappa de 0,7213.

Tabela 12 – Resultados da classificação com as imagens segmentadas da Base PH².

Base de Imagens PH ²												
Descritor	MLP				SVM				Random Forest			
	P	S	K	A	P	S	K	A	P	S	K	A
ABCD	91,2	91,5	0,7213	91,5	92,3	92,5	0,7492	92,5	90,7	91,0	0,7078	91,0
GLCM	81,4	82,0	0,4156	82,0	79,3	80,5	0,3478	80,5	78,6	80,5	0,3206	80,5
GLRLM	89,0	89,0	0,6563	89,0	77,2	66,0	0,2273	66,0	89,7	90,0	0,6753	90,0
HOG	85,1	85,5	0,5338	85,5	88,0	88,5	0,6154	88,5	90,1	90,5	0,6758	90,5
LBP	86,2	87,0	0,5423	87,0	85,0	81,5	0,1148	81,5	84,4	85,5	0,4610	85,5
Tamura	89,2	89,5	0,6624	89,5	89,0	89,5	0,6048	89,5	88,5	89,0	0,6358	89,0
Box-Counting	84,7	85,0	0,5223	85,0	87,3	85,5	0,5892	85,5	89,0	89,5	0,6416	89,5
AlexNet	85,7	86,5	0,5392	86,5	87,5	88,0	0,6026	88,0	90,2	90,5	0,6690	90,5
CaffeNet	87,3	88,0	0,5862	88,0	87,2	87,5	0,5981	87,5	89,9	90,0	0,6403	90,0
Vgg-f	87,3	88,0	0,5862	88,0	86,2	87,0	0,5517	87,0	90,7	91,0	0,6897	91,0
Vgg-m	89,1	89,5	0,6488	89,5	90,1	90,5	0,6758	90,5	88,5	89,0	0,6284	89,0
Vgg-s	87,3	88,0	0,5862	88,0	90,7	91,0	0,6897	91,0	90,7	91,0	0,6959	91,0
Vgg-verydeep-16	91,2	91,5	0,7157	91,5	91,8	92,0	0,7297	92,0	91,2	91,5	0,7213	91,5
Vgg-verydeep-19	88,7	89,0	0,6429	89,0	91,3	91,5	0,7099	91,5	89,7	90,0	0,6753	90,0

Por fim, os resultados para a base de imagens DermIS com segmentação podem ser vistos na Tabela 13. Em comparação com os resultados da base de dados PH², as taxas de acerto foram novamente menores. Ambos os descritores Vgg-f e AlexNet atingiram bons resultados. O melhor resultado foi obtido com o descritor Vgg-f e o classificador MLP, onde alcançou uma acurácia igual a 84,9% e índice Kappa igual a 0,6892. Com

os classificadores SVM e *Random Forest*, o descritor AlexNet obteve melhor resultado atingindo uma acurácia de 82,0% e 83,4%, respectivamente e índice Kappa considerado “Muito Bom”.

Tabela 13 – Resultados da classificação com as imagens segmentadas da Base DermIS.

Base de Imagens DermIS												
Descritor	MLP				SVM				Random Forest			
	P	S	K	A	P	S	K	A	P	S	K	A
ABCD	69,7	69,9	0,3793	69,9	68,6	68,9	0,3492	68,9	66,0	66,5	0,2950	66,5
GLCM	61,9	62,6	0,2133	62,6	70,1	67,0	0,2644	66,9	69,2	69,4	0,3522	69,4
GLRLM	64,2	64,6	0,2658	64,5	54,2	54,4	0,0619	54,3	66,2	66,5	0,3038	66,5
HOG	67,1	67,5	0,3198	67,4	65,4	65,5	0,2557	65,5	70,1	70,4	0,3846	70,3
LBP	58,4	59,7	0,1242	59,7	80,2	70,4	0,2930	70,3	65,0	65,5	0,2746	65,5
Tamura	67,1	67,5	0,3176	67,4	68,4	68,4	0,3479	68,4	66,1	66,5	0,2995	66,5
Box-Counting	56,2	56,8	0,1021	56,7	67,6	68,0	0,3297	67,9	51,4	52,4	0,0035	52,4
AlexNet	81,1	81,1	0,6126	81,0	82,0	82,0	0,6290	82,0	83,5	83,5	0,6575	83,4
CaffeNet	83,6	83,5	0,6628	83,4	78,6	78,6	0,5623	78,6	80,5	80,6	0,5983	80,5
Vgg-f	84,9	85,0	0,6892	84,9	81,6	81,6	0,6231	81,5	83,1	83,0	0,6458	83,0
Vgg-m	71,8	71,8	0,4212	71,8	64,5	64,1	0,2706	64,0	70,7	0,709	0,3860	70,8
Vgg-s	66,9	67,0	0,3214	66,9	67,2	67,0	0,3276	66,9	68,7	68,9	0,3430	68,9
Vgg-verydeep-16	70,7	70,9	0,3994	70,8	63,8	63,1	0,2554	63,1	64,1	64,6	0,2588	64,5
Vgg-verydeep-19	67,3	67,5	0,3303	67,4	70,0	69,9	0,3851	69,9	69,6	69,9	0,3715	69,9

Após a realização dos testes podemos perceber que na base de imagens DermIS houve uma maior dificuldade na diferenciação das duas classes de lesões: melanoma e não-melanoma; uma vez que as imagens dessa base possuem mais semelhanças entre essas classes do que as imagens da base PH².