

# Análise e Classificação de Lesões Dermatológicas com Aprendizado de Máquinas

Hugo R. Gomes, Francisco J. A. de Aquino

**Resumo**—Este artigo foca em aplicar os conceitos de aprendizado de máquinas e de processamento de imagens ao reconhecimento de lesões dermatológicas, comparando os resultados obtidos a partir da utilização de árvores de decisão e redes neurais artificiais. São descritas as etapas do desenvolvimento dos classificadores e de um aplicativo de classificação interativo. Os resultados obtidos foram de 81.6% de exatidão para árvores de decisão e 85% para redes neurais. Ao todo foram utilizadas 200 imagens para o treinamento e classificação, as imagens foram disponibilizadas pelo Hospital Pedro Hispano, Matosinhos.

**Index Terms**—Aprendizado de máquinas, processamento de imagens, melanoma, árvores de decisão, redes neurais.

## I. INTRODUÇÃO

Melanomas são a contraparte maligna de tumores melanocíticos, sendo um dos tumores malignos a se espalhar mais rapidamente, podendo ocorrer em qualquer idade [1]. Em [2] é descrito um estudo sobre o relacionamento do estágio de desenvolvimento de lesões melanoma com a taxa de recorrência ou metástase, conclui-se que casos nos quais a doença se encontra em estágios iniciais de desenvolvimento as chances de ocorrência de metástase e de recorrência são muito menores. São necessários, portanto, meios de facilitar o diagnóstico.

Uma lesão dermatológica pode ser classificada a partir da regra ABCD descrita por Stolz [3]. A regra propõe um sistema de pontuação que avalia lesões com base na (A) assimetria, (B) padrão de borda (C) variação de cor e (D) diferentes componentes estruturais. A Tabela I mostra as respectivas pontuações e pesos dos critérios componentes da regra.

Ao fim do processo obtém-se a pontuação final a partir da soma das pontuações dos critérios levando em consideração seus respectivos pesos.

Pontuações com valores menores ou iguais a 4.75 são consideradas benignas, valores maiores que 4.75 representam casos suspeitos, ou atípicos, valores maiores do que 5.45 podem ser considerados como melanoma.

## II. CLASSIFICAÇÃO

O aprendizado de máquinas propõe vários métodos que possibilitam a classificação através da extração de características e da generalização dos dados referentes a uma determinada classe.

Tabela I.

| Cálculo da pontuação ABCD da dermatologia |                      |      |
|---|----------------------|------|
| Critério                                  | Possíveis pontuações | Peso |
| Assimetria                                | 0-2                  | 1.3  |
| Borda                                     | 0-8                  | 0.1  |
| Cor                                       | 1-6                  | 0.5  |
| Diferentes Estruturas                     | 1-5                  | 0.5  |

Algoritmos de classificação podem utilizar quatro diferentes métodos, são eles simbólico, conexionista, genético e estocástico. Neste artigo foram utilizados métodos simbólicos e conexionistas. Representados por árvores de decisão e redes neurais respectivamente [4].

Como descrito em [5] árvores de decisão são exemplos de métodos simbólicos heurísticos que abordam uma estratégia de aprendizado não-incremental por exemplos.

Uma rede neural é uma máquina projetada para modelar a maneira como o cérebro realiza uma tarefa em particular ou função de interesse. Para alcançar um bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples, denominadas “neurônios” ou “unidades de processamento” [6].

Seguindo a regra ABCD, um classificador de lesões dermatológicas deve levar em consideração características de simetria, cor, borda e textura. Essas características são extraídas a partir de descritores, que produzirão dados que representam a lesão.

## III. METODOLOGIA

O processo de classificação de lesões dermatológicas desenvolvido pode ser observado na Fig. 1, representando uma visão simplificada do método proposto. As imagens utilizadas para o treinamento e classificação foram retiradas do dataset PH<sup>2</sup> disponibilizado pelo Hospital Pedro Hispano[7].

Cada imagem foi pré-processada para a remoção de possíveis elementos como pelos, suor e sangue que pudessem interferir na extração de características. Posteriormente, as imagens foram segmentadas a partir de uma máscara de limiar.

O processo removerá todos os elementos que não representem a lesão, isso pode ser observado na Fig. 2, que demonstra o funcionamento de um aplicativo desenvolvido no Matlab que recebe uma imagem, faz seu pré-processamento, extrai suas características e por fim atribui uma das classes a ela, fazendo o seu diagnóstico com os classificadores.

Posteriormente é feita a extração dos descritores de características, apresentadas a seguir.

A cor representa a quantidade de cores presentes na lesão, ela será representada por um histograma de cores. A Fig. 3 mostra a plotagem da sobreposição dos histogramas das

Hugo R. Gomes é aluno do curso de Tecnólogo em Telemática, IFCE. E-mail: [hugo.ifce@gmail.com](mailto:hugo.ifce@gmail.com).

Francisco José A. de Aquino é professor do IFCE – Campus Fortaleza, Departamento de Telemática. E-mail: [fcovalves\\_Aq@ifce.edu.br](mailto:fcovalves_Aq@ifce.edu.br).

classes comum, atípico e melanoma respectivamente.

É possível perceber que lesões melanoma apresentam maior quantidade de cores do que lesões comuns, a diferença entre as classes comuns e atípicas, porém, não pode ser distinguida facilmente apenas com a análise da cor.

A textura será descrita com momentos estatísticos, sendo representado por um vetor de características contendo a média, variância e desvio padrão dos canais RGB e da imagem em escala de cinza.

A simetria será descrita a partir do cálculo da assinatura da lesão, sendo representada por uma função que descreve a distância euclidiana da borda da lesão em relação a centroide. A variação da assinatura será utilizada para descrever a borda.

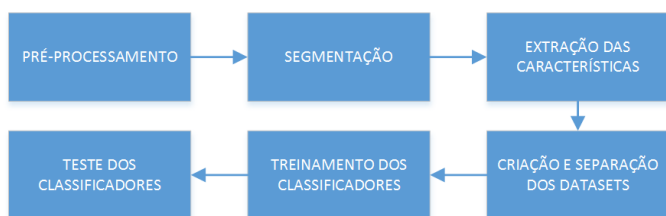


Fig. 1. Representação do processo desenvolvido



Fig. 2. Software de classificação interativa de lesões.

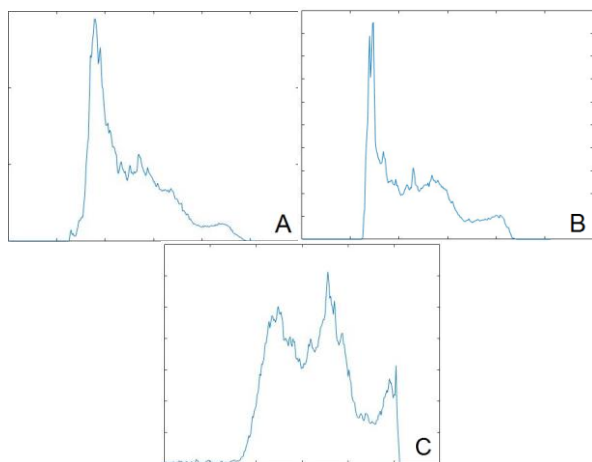


Fig. 3. Histogramas sobrepostos de todas as lesões: (a) comum; (b) atípico e (c) melanoma.

Após a formação do vetor de características foram criados 3 datasets contendo os descritores das respectivas classes. Posteriormente os *datasets* foram juntados e distribuídos entre treino e teste, atribuindo a eles 70% e 30% das imagens respectivamente.

Para evitar estouro de memória o processo será dividido em cinco diferentes scripts, descritos a seguir.

- Primeiro: descreve as imagens quanto a cor e textura e salva seus resultados em arquivos “.mat”.
- Segundo: divide os dados de cor e textura em treino e teste.
- Terceiro: descreve as imagens quanto a simetria e borda.
- Quarto: divide os dados de simetria e borda.
- Quinto: Treina e testa os classificadores utilizando os dados produzidos, exibindo a porcentagem de acerto.

#### IV. RESULTADOS

O desempenho dos classificadores pode ser observado na tabela II, onde são apresentados os resultados em árvores de decisão e rede neural de reconhecimento de padrões.

TABELA II  
EXATIDÃO DOS CLASSIFICADORES

| Classificador     | Comum | Atípico | Melanoma | Total    |
|-------------------|-------|---------|----------|----------|
| Árvore de Decisão | 62.5% | 100%    | 83.3%    | 81.667 % |
| Rede Neural       | 75%   | 100%    | 75%      | 85 %     |

A rede neural obteve melhores resultados do que o a árvore de decisão, respondendo positivamente a descritores estatísticos e globais, porém a utilização de assinaturas não representou resultados satisfatórios.

Apesar de ter obtido uma melhor precisão do que a árvore de decisão, a rede neural obteve menor pontuação quanto a precisão da classe melanoma, se mostrando menos apta a fazer predições com o vetor de características utilizado.

#### V. CONCLUSÕES

Ao longo do período de desenvolvimento foi possível concluir que o aprendizado de máquinas é capaz de prover soluções simples para problemas complexos, porém a aprendizagem nem sempre ocorre de maneira esperada.

Ficou claro também a importância da qualidade dos dados utilizados, e, por fim, conclui-se que não se deve confiar cegamente em tais algoritmos, eles devem ser utilizados como sistemas de apoio a decisão, neste caso para um dermatologista.

#### VI. REFERÊNCIAS

- [1] Mohan, H. (2010). Textbook of Pathology Sixth Edition. Nova Deli: Jaypee.
- [2] BRESLOW, A. (1970). Thickness, Cross-Sectional Areas and Depth of Cutaneous Melanoma. *Annals of Surgery* 172.5, 902–908.
- [3] Stolz, W. (1994). The ABCD Rule of Dermatoscopy. High Prospective Value in The Diagnosis of Doubtful Melanocytic Skin Lesions. *Journal of the American Academy of Dermatology*, 551-559.
- [4] Luger, G. F. (2013). Inteligência Artificial 6ª Edição. São Paulo: Pearson.
- [5] Quinlan, R. (1986). Induction of Decision Trees. *Machine Learning*, pp. 81–106.
- [6] Haykin, S. (2001). Redes Neurais Princípios e Prática. Porto Alegre: Bookman.
- [7] Teresa Mendonça, Pedro M. Ferreira, Jorge Marques, Andre R. S. Marcal, Jorge Rozeira. PH<sup>2</sup> - A dermoscopic image database for research and benchmarking, 35th International Conference of the IEEE Engineering in Medicine and Biology Society, July 3-7, 2013, Osaka, Japan.