

SEGMENTAÇÃO, CARACTERIZAÇÃO E CLASSIFICAÇÃO DE IMAGENS DERMOSCÓPICAS USANDO SELEÇÃO DE ATRIBUTOS

Ana Paula Benetti Grezzana (PIBIC/CNPq)¹

Newton Spolaôr¹, Feng Chung Wu^{1 2}

Huei Diana Lee (Orientadora)¹

Laboratório de Bioinformática (LABI)/
Universidade Estadual do Oeste do Paraná (Unioeste)¹
Faculdade de Ciências Médicas (FCM)/
Universidade Estadual de Campinas (Unicamp)²

{ana.grezzana, huei.lee}@unioeste.br

Objetivos

Pesquisar e aplicar técnicas de segmentação customizada, extração de características e Seleção de Atributos (SA) para auxiliar na classificação de imagens dermoscópicas benignas e malignas.

Métodos e Procedimentos

Abordagens baseadas em segmentação e caracterização de imagens, bem como SA têm obtido sucesso na classificação de imagens dermoscópicas, exibindo potencial para apoiar a detecção precoce de doenças desafiadoras, como câncer de pele [1]. Essas abordagens constituem a base para o método deste trabalho, composto pelas seguintes etapas: (1) investigação de técnicas de apoio e coleta do conjunto de imagens; (2) pré-processamento de imagens; (3) extração de características; (4) seleção de características (atributos) e construção de classificador; e (5) realização de experimentos e análise dos resultados. Algoritmos, estratégias e soluções úteis à construção do modelo foram estudadas na Etapa (1). Ainda, nesta fase, foi coletado o conjunto de imagens dermoscópicas com lesões anotadas utilizado neste trabalho, proveniente do repositório *International Skin*

Imaging Collaboration (ISIC) [2]. Essa seleção contou com 226 imagens de lesões do tipo nevos e 48 de lesões do tipo melanoma. Na Etapa (2), foi desenvolvida uma segmentação justa simples das imagens a partir de suas máscaras (anotações) de lesão que já haviam sido adquiridas. Essa segmentação de imagens consiste no recorte da imagem na região determinada por sua máscara, que cobre a localidade importante para a análise. Na Etapa (3) foram aplicadas as técnicas de caracterização de Haralick *et al.* [3] (14 atributos), e *Neighborhood Grey Tone Difference Matrix* (NGTDM) [4] (5 atributos). Na Etapa (4), foi realizada a divisão das imagens em *folds* e a validação cruzada, que permite que todos os exemplos (imagens) e atributos passem pela fase de teste, e, em seguida, foi aplicada a seleção de atributos pelo método *Correlation-based Feature Selection* (CFS) [5] em cada *fold* de treinamento. Ao final, os atributos escolhidos foram aqueles que apresentaram ocorrência em todos os *folds* (intersecção) [6]. A construção do modelo, ainda nesta etapa, foi desenvolvida pelos algoritmos J48 [5] e *Support Vector Machine* (SVM) [6]. Cada um deles recebeu como entrada (1) todos os atributos (*Baseline* (B)) e (2) os atributos selecionados previamente. Na Etapa (5), foram analisados os valores de

acurácia para verificar o desempenho dos algoritmos de classificação com e sem SA. Esta análise, foi realizada usando o teste estatístico Friedman, comparando os resultados obtidos por J48 e SVM com e sem a aplicação de SA.

Resultados

O estudo de imagens dermoscópicas é essencial para o diagnóstico precoce de um possível câncer de pele e para que, a partir dele, possa ser realizado o tratamento precoce da lesão. Visando este objetivo e possuindo como base trabalhos bem-sucedidos na literatura, foi possível a construção dos modelos que apresentaram resultados satisfatórios. Esses resultados se devem, especialmente, pela adição de mecanismos de segmentação justa, como mostra na Figura 1, e de intersecção entre conjuntos de atributos.

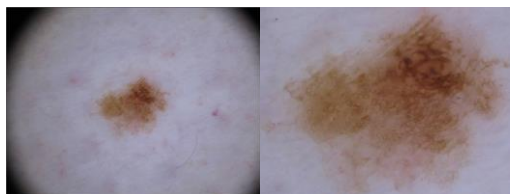


Figura 1: À esquerda, uma imagem original do repositório ISIC; e à direita, a mesma imagem com a aplicação da segmentação justa

Os resultados encontrados pela aplicação dos algoritmos J48 e SVM apresentaram acurácias acima de 80%. Em relação ao SVM, os resultados se mostraram iguais para o seu uso com todos os atributos e com os atributos selecionados por CFS e pela intersecção de *folds*, indicando que possivelmente não há presença de atributos redundantes e ou irrelevantes para a classe determinada. Já para o algoritmo J48, houve uma melhora no desempenho pelo uso de SA e de intersecção, o que demonstra o benefício do uso desses métodos. Ainda, esse último modelo obteve o melhor resultado encontrado, superando SVM com e sem o uso de SA. A aplicação do teste de Friedman, não revelou diferença estatisticamente significativa na comparação dos resultados. Na Tabela 1 são mostrados valores médios de acurácia e respectivos desvios-padrão entre parênteses.

Tabela 1: Resultados da aplicação dos algoritmos de classificação

Acurácia Média	J48	SVM
Todos os atributos	80,28 (3,98)	82,49 (1,46)
Com SA e intersecção	83,17 (7,52)	82,49 (1,46)

Conclusões

O método implementado neste trabalho cumpriu o objetivo, sendo que sua aplicação possibilitou a classificação de imagens com lesões de pele. Resultados promissores em um conjunto de imagens de um repositório importante em dermoscopia foram obtidos. Trabalhos futuros envolvem a aplicação do método em mais bases de imagens médicas e o uso de mais medidas de avaliação para averiguar o potencial de generalização do melhor modelo. Algoritmos de classificação alternativos, como redes neurais rasas e profundas, também serão explorados.

Referências Bibliográficas

- [1] Lee, H.D. et al. (2018). Dermoscopic assisted diagnosis in melanoma: Reviewing results, optimizing methodologies and quantifying empirical guidelines. *Knowledge-Based Systems*, **158**, 9-24.
- [2] ISIC Archive. (2021). The International Skin Imaging Collaboration. Acessado em: 06/09/2021. Disponível em: <https://www.isic-archive.com/>.
- [3] Haralick, R.M. et al. (1973). Textural features for image classification. *IEEE Transaction on System, Man and Cybernetics*, **3**, 610-621.
- [4] Amadasun, M. & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on systems, man, and Cybernetics*, **19**, 1264-1274.
- [5] Witten, I.H. et al. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Cambridge: Morgan Kaufmann.
- [6] Gheiratmand, M. et al. (2017). Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. *npj Schizophr*, **3**, 22, 1-12.
- [7] Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*, Cambridge: Cambridge University Press.