

ESTUDO COMPARATIVO SOBRE CLASSIFICAÇÃO DE LESÕES DE PELE COM RNA-MLP E SVM

Gilson Saturnino dos Santos¹

Alex F. de Araujo²

¹*gilson.santos@ifms.edu.br*

²*alex.araujo@ifms.edu.br*

^{1,2}*Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul*

¹*Rua Salime Tanure, S/N - Santa Tereza, 79400-000, Mato Grosso do Sul / Coxim, Brazil*

²*Rua Angelo Melão, 790 - Jardim Das Paineiras, 79641-162, Mato Grosso do Sul / Três Lagoas, Brazil*

Abstract. Technological evolution, evidenced in recent years, has contributed significantly to both Artificial Intelligence (AI) and Digital Image Processing and Analysis (DIP) areas. Several computational methodologies can be found in the specialized literature, with different applications, such as the classification of skin lesions from dermoscopic images. Although the initial analysis of skin lesions is based on a set of visual rules known as the ABCD rule (Asymmetry, Borders, Color and Diameter), the performance of this visual analysis is influenced by factors such as variation of illumination during capture of the image, the presence of artifacts that cause noise, and the visual fatigue of the specialist during image analysis. A mistaken initial analysis may lead to delays in the elaboration of an adequate treatment plan, affecting the effectiveness of this treatment. The interdisciplinarity between AI and DIP has produced good results when the algorithms are intended for specific applications and use a priori knowledge based on the problem addressed. In this context, this work presents a comparative study between the classification methodologies based on Artificial Neural Networks (ANN) and Support Vector Machine (SVM), using the existing knowledge about the visual evaluation of skin lesions as knowledge base a priori to train these classifiers. In general terms, it is expected to better understand the performance of these two AI techniques when applied to classify skin lesions in malignant and non-malignant, evaluating the performance of each methodology. To perform this work we used the Weka tool, with the implementation of Artificial Neural Network (Multilayer Perceptron) and Support Vector Machine (Sequential Minimal Optimization). The dataset used has 200 examples and 14 attributes. The methodology used was Stratified Cross-Validation in 10 parts. The parameters of each algorithm were defined as standard in the tool. The results obtained were promising, showing the relevance of the AI algorithms to skin lesion classification and the relevant features of the Weka to improve the quality of the classification modeling solutions.

Keywords: Skin Lesion Classification, Artificial Neural Networks, Support Vector Machine, Image Processing and Analysis

1 Introdução

A evolução tecnológica, notada nos últimos anos, tem contribuído de forma significativa para que as áreas de Inteligência Artificial (IA) e Processamento e Análise de Imagens Digitais (PID) se destaquem dentro das comunidades acadêmica, científica e industrial. Esta interdisciplinaridade entre as duas áreas de conhecimento tem sido explorada por diferentes trabalhos apresentados na literatura, proporcionando o desenvolvimento de metodologias computacionais variadas, com diferentes aplicações, entre as quais destaca-se neste trabalho, a classificação de lesões de pele em imagens dermatoscópicas. Esta aplicação é relevante devido a alta incidência de câncer de pele registrada na população em geral, além da necessidade de diagnóstico precoce para aumentar a probabilidade de sucesso no tratamento destas lesões.

Para agilizar o diagnóstico, uma análise visual prévia é realizada pelo especialista médico, levando em consideração um conjunto de regras visuais conhecidos como ABCD (**A**ssimetria, **B**ordas, **C**or da lesão e **D**ímetro). Por envolver análise subjetiva, é uma tarefa complexa, cuja performance pode sofrer influência de fatores como a variação de iluminação durante a captura da imagem, a presença de artefatos que causam ruído, e o cansaço visual do especialista durante análise visual da imagem. Uma análise inicial equivocada pode gerar atraso na elaboração de um plano de tratamento adequado, afetando a efetividade desse tratamento, e colocando em risco a vida do paciente. Além disso, existem diferentes técnicas para obtenção da imagem a ser avaliada, tais como fotografia normal, e utilização de contraste junto com o equipamento chamado dermatoscópio, o qual pode ampliar a lesão, exibindo mais detalhes das suas características. Todos estes fatores somados, abrem espaço para a tecnologia computacional, mais especificamente algoritmos computacionais, possam ser desenvolvidos e usados para auxiliar o dermatologista na análise inicial das lesões de pele.

Neste contexto, este trabalho apresenta um estudo comparativo entre a aplicação das metodologias de classificação baseadas em Redes Neurais Artificiais (RNA) e Máquina de Vetores de Suporte (*Support Vector Machines* - SVM), usando o conhecimento existente sobre a avaliação visual de lesões de pele como base de conhecimento a priori para treinar, de forma supervisionada, estes classificadores. A existência de trabalhos relevantes na literatura especializada tem demonstrado que a interdisciplinaridade presente entre IA e PID pode produzir bons resultados quando os algoritmos desenvolvidos destinam-se a aplicações específicas e usam uma base de conhecimento a priori sobre o problema abordado.

Como objetivo geral deste trabalho, espera-se entender melhor o desempenho dessas duas técnicas de IA quando aplicadas para classificar lesões de pele nas classes *Common-Nevus*, *Atypical-Nevus* e Melanoma, avaliando a taxa de acerto e falhas de cada metodologia. Para alcançar este objetivo, foi utilizado a ferramenta Weka, a qual está disponível para uso livre, com as implementações de RNA e SVM, chamadas na ferramenta de *Multilayer Perceptron* (MLP) *Sequential Minimal Optimization* (SMO), respectivamente, bem como os respectivos algoritmos de treinamento. O conjunto de dados utilizado possui 200 exemplos, sendo 14 atributos para descrever cada elemento. A metodologia de treinamento utilizada foi a *Stratified Cross-Validation* em 10 partes, sendo os parâmetros de cada algoritmo definidos como padrão na ferramenta. Os resultados obtidos foram promissores, apresentando acurácias de 89% e 92,5%, respectivamente para os classificadores RNA e SVM. Também foi realizado um experimento com o pacote AutoWeka e foi encontrado uma configuração de RNA que obteve uma acurácia de 92,5%. Para avaliação estatística dos resultados obtidos adotou-se as taxas de verdadeiros positivos (VP), estatística Kappa e Área ROC, para cada classe, sendo que o modelo SVM também foi superior ao modelo RNA com configurações padrões do algoritmo para as três classes testadas. A RNA com as configurações encontradas pelo AutoWeka obteve resultados iguais ao modelo SVM, como descrito detalhadamente na seção de resultados.

O restante deste artigo está organizado da seguinte forma: a fundamentação teórica é apresentada na Seção 2. Na seção 3 descreve-se a metodologia adotada, enquanto na seção 4 os testes e resultados obtidos são apresentados, seguidos pelas conclusões e considerações finais.

2 Fundamentação teórica

O processamento e análise de imagens possui um fluxo de etapas, como a etapa de classificação, na qual pixels ou regiões semelhantes de uma imagem podem ser agrupados para separar e representar os diferentes grupos de pixels da imagem em classes de interesse. A classificação de imagens é uma técnica muito utilizada para identificar objetos e regiões de interesse em imagens de várias áreas e aplicações, como em imagens de satélite (Fucci et. al. [1]), imagens médicas (Dos Santos e Araujo [2]), e na indústria (Nacereddine et. al. [3]). Na área médica, as tarefas de interpretação visual de imagens de exames médico vêm sendo cada vez mais auxiliadas por algoritmos computacionais, otimizando a elaboração dos planos de tratamentos.

Os algoritmos de classificação podem ser supervisionados ou não supervisionados, sendo o seu tipo definido a partir da forma usada para realizar o treinamento do classificador, podendo ter ou não a intervenção de um usuário externo ao processo de treinamento. A classificação supervisionada acontece com intervenção desse usuário externo, exigindo usar conhecimento prévio das regiões de interesse para criar amostras relevantes para as classes desejadas, construindo-se assim um modelo de aprendizagem a ser usado nas classificações seguintes. Por outro lado, na classificação não supervisionada ocorre menos interação de um usuário na fase de treinamento. O algoritmo usa a semelhança de características dos pixels para gerar conhecimento suficiente para definir as classes e presentes no conjunto de imagens. Para isso, são usados métodos estatísticos e matemáticos para calcular os limiares de cada classe (Meneses e Almeida [4]).

Para classificação de imagens médicas, pode-se usar algoritmos apresentados na literatura, tais como: SVM (*Support Vector Machine*), MLP (*MultiLayer Perceptron*), e KNN (*K-Nearest-Neighbor*) e *Randon Forest* (Hosni et.al. [5]). Para que os algoritmos possam ser treinados e, posteriormente possam realizar a classificação das estruturas presentes nas imagens, faz-se necessário rotular os atributos relevantes para a definição da classe para a qual uma estrutura pertence. Neste contexto, os descritores, também conhecidos como atributos ou características (Remeseiro e Bolon-Canedo [6]), de uma imagem tornam-se fundamentais para a etapa de classificação. Os descritores são dados que descrevem um objeto e devem ser calculados de forma que objetos de classes distintas possuam combinações de valores diferentes. Por isso, é comum os algoritmos de classificação usarem combinações de vários descritores. Para a análise de imagens, os descritores podem representar informação de textura, intensidade, forma entre outros.

A partir de um estudo da literatura, foi possível verificar que alguns atributos mostram-se relevantes para a classificação de lesões de pele, como apresentado em Mendonça et. al. [7], onde os autores detalham e recomendam usar os atributos apresentados na Tabela 1. Segundo os autores, estes atributos foram definidos com auxílio de especialistas médicos e são relevantes para descrever as classes de lesões de pele *Common Nevus*, *Atypical Nevus* e *Melanoma*. As imagens na Figura 1 representam as 3 referidas classes. É importante destacar que *Common-Nevus* são agrupamentos de células de pigmento da pele, e é uma lesão de pele não maligna presente na maioria dos adultos. Geralmente aparecem nas áreas frequentemente expostas ao sol, tendo aparência mais regular e cor homogênea. Por outro lado *Atypical Nevus* são maiores do que os *Common Nevus*, apresentam mistura de várias cores e possuem bordas irregulares, podendo às vezes, evoluir para um melanoma. O melanoma é uma lesão perigosa por possuir capacidade de se espalhar para outros órgãos do corpo. Este tipo de lesão pode mudar de forma com o passar do tempo, possui bordas irregulares e variação grande cores.



Figura 1. Imagens referentes às classes *Common Nevus*, *Atypical Nevus* e *Melanoma* (Mendonça et. al. [7]).

Quando um método de classificação é proposto ou uma nova aplicação é realizada, pode-se avaliar a precisão (acurácia) dos resultados obtidos, comparando-os com as amostras *Gold Standard*. Na literatura especializada, encontra-se estudos relevantes e reconhecidos, nos quais técnicas para esta avaliação são apresentadas. A matriz confusão é uma dessas técnicas, sendo constantemente adotada nos trabalhos de classificação (Ponzoni e Almeida [8]). Ela permite quantificar a acurácia de uma porção dos dados obtidos de acordo as taxas de erro e acerto em relação à amostra real. Geralmente, é representada na forma matricial, na qual as colunas representam os dados reais ou previstos nas amostras *Gold Standard*, e as linhas representam as classes geradas pelo classificador. Os dados classificados de acordo com o esperado são exibidos na diagonal da matriz e constituem os conjuntos conhecidos como verdadeiro positivo e verdadeiro negativo. Enquanto isso, os demais dados da matriz são interpretados como classificações incorretas ou falso positivo e falso negativo. Outras técnicas derivadas da matriz de confusão também são apresentadas na literatura e são bastante usadas, como o índice Kappa, que busca avaliar a assertividade do algoritmo de classificação. É importante ressaltar que não existe uma regra que estabelece os valores máximo e mínimo para o índice Kappa que permitem avaliar se uma classificação foi eficiente ou não, mas é consenso de que valores mais próximos de 1 (ou 100%) representam melhores resultados. Estes limiares são definidos de acordo com cada aplicação e com os resultados mínimos que se espera obter.

Segundo Witten & Frank [9] o Weka (*Waikato Environment for Knowledge Analysis*) é uma coleção do estado da arte de implementações de algoritmos de Aprendizado de Máquina (AM), subárea da Inteligência Artificial, permitido utilizar recursos de pré-processamento e extração de conhecimento. A ferramenta foi desenvolvida pela Universidade de Waikato da Nova Zelândia, utilizando a linguagem Java, sob a licença *General Public License*. Portanto, é um software livremente distribuído e de código fonte aberto.

O Weka permite utilizar implementações de algoritmos de Aprendizagem de Máquina para tarefas de regressão, classificação, agrupamento, extração de regras de associação e seleção de atributos relevantes. Além disso, a ferramenta permite verificar várias métricas de avaliação de classificadores, por exemplo: Acurácia, Verdadeiros Positivos, Verdadeiros Negativos, Área ROC e índice Kappa. Verdadeiros Positivos e Negativos são o número de exemplos positivos ou negativos classificados corretamente, sendo Positivo e Negativo exemplos de classes. Área ROC (*Receiver Operating Characteristic*), pode ser utilizada para designar a qualidade de um classificador. Quanto mais próximo de 1 for o valor da Área ROC, melhor é o classificador, ou seja, significa que nas avaliações dos resultados de verdadeiros positivos foram próximos a 100% e de falsos positivos foram próximos a 0%. Índice Kappa é uma importante métrica de confiabilidade que indica o quanto dois anotadores, juizes, humanos ou não, concordam sobre determinada classificação. No AM pode ser utilizada para analisar a predição e a realidade dos rótulos de classes do conjunto de dados. Quanto mais próximo de 1 maior a qualidade do classificador (Witten e Frank [9]; Landis e Koch [10]).

Para realizar o treinamento pode-se optar entre as técnicas *Holdout* e *Stratified Cross-Validation*. A técnica *Holdout* divide os dados disponíveis em uma porcentagem fixa para treinamento e teste, enquanto na técnica *Stratified Cross-Validation* o conjunto de dados é dividido em r partes (*fold*s). Os $r-1$ *fold*s são usados para treinamento e a parte restante é utilizada para teste. O processo é repetido r vezes utilizando sempre um *fold* diferente para teste. A proporção de erro é a média dos erros de cada execução. Cada *fold* possui a mesma distribuição de classes que o conjunto total de exemplos disponíveis (Witten e Frank [9]).

Tabela 1. Descritores recomendados pela literatura para classificação de lesões de pele (MENDONÇA et. al. [7]).

Atributo	Valores
Assimetria	0 – Completamente simétrico 1 – Assimétrico em um eixo 2 – Completamente assimétrico
Pigmentos	AT – Atípico T – Típico
Pontos ou glóbulos	A – Ausente AT – Atípico T – Típico
Listras	A – Ausente P – Presente
Área de regressão	A – Ausente P – Presente
<i>Blue withish veil</i>	A – Ausente P – Presente
Cor	1 – Branco 2 – Vermelho 3 – Marrom claro 4 – Marrom escuro 5 – Azul acinzentado 6 – Preto

Para utilizar as funções do Weka pode-se fazer uso de interface gráfica, do console de linha de comando ou incorporá-lo no código fonte de um software por meio da utilização da linguagem de programação Java (Witten e Frank [9]).

O Weka oferece implementações de vários algoritmos de aprendizado de máquina, inclusive algoritmos clássicos de aprendizado supervisionado. Os algoritmos utilizados neste trabalho foram o de Redes Neurais Artificiais, chamado de *Multilayer Perceptron* (MLP) e de Máquina de Vetores de Suporte (SVM), chamado no Weka de SMO (*Sequential Minimal Optimization*). O Weka permite utilizar o pacote AutoWeka (Kotthoff [11]) para avaliar de maneira automática todos os possíveis algoritmos de aprendizado de máquina e suas diversas configurações informando o melhor algoritmo e seu respectivo conjunto de configuração (Witten e Frank [9]).

3 Metodologia

Foi utilizada a ferramenta Weka 3.8.1 para construção e avaliação dos modelos de classificadores RNAs MLP e SVM. O conjunto de dados utilizado para treinamento, possui 200 exemplos, 14 atributos (incluindo o atributo classe) e 3 classes (Melanoma com 40 instâncias, Common-Nevus e Atypical Nevus com 80 exemplos cada).

Foi realizado um experimento usando o RNA MLP e um segundo experimento com o algoritmo SMO, ambos com parâmetros padrões da ferramenta Weka. No primeiro experimento a RNA utilizou 9 camadas, 0.3 de taxa de aprendizado e 0.2 de momentum. No segundo experimento o SMO foi configurado com *Kernel* Polinomial. Para avaliação dos modelos foi utilizada a metodologia *Stratified Cross-Validation* dividindo o conjunto de dados em 10 partes. Após a execução dos dois primeiros experimentos foi realizado um terceiro experimento utilizando o pacote AutoWeka 2.6.1 com suas configurações padrões para identificar o melhor modelo dentre os possíveis do Weka e seus respectivos parâmetros.

4 Testes e Resultados

Na Tabela 2 pode ser verificado os resultados de acurácia, índice Kappa, Falsos Positivos e Falsos Negativos obtidos pelos modelos MLP e SVM. O modelo obtido com o algoritmo SVM é superior em todas as métricas ao algoritmo RNA.

O algoritmo encontrado com o pacote AutoWeka foi o *Multilayer Perceptron* com taxa de aprendizagem de 0.8348336141050203, *momentum* 0.5680343213959079 e 17 camadas. O classificador obtido por este experimento obteve resultados iguais ao modelo obtido com SVM: 92,5% de acurácia (Tabela 2). A matriz confusão também é igual ao modelo SVM (Tabela 3). Pode-se constatar que a utilização do pacote AutoWeka se mostrou interessante ao encontrar os parâmetros ideais para o modelo MLP igualar os resultados de avaliação do modelo SVM. Em uma única métrica apresentada pelo Weka o modelo MLP com AutoWeka foi um pouco superior ao modelo SVM. Esta métrica foi a Área ROC, como pode ser verificado na Tabela 4. Nas demais métricas apresentadas pelo Weka os modelos apresentaram resultados iguais.

Tabela 2. Resultados das métricas de Acurácia, Índice Kappa, Taxa de Verdadeiros Positivos e Taxa de Falsos Positivos dos modelos RNA (MLP), SVM (SMO) e RNA (MLP AutoWeka).

Algoritmo	Acurácia	Índice Kappa	Taxa de Verdadeiros Positivos	Taxa de Falsos Positivos
RNA (MLP)	89,0%	0,82	0,89	0,06
SVM (SMO)	92,5%	0,88	0,92	0,04
RNA (MLP AutoWeka)	95,5%	0,88	0,92	0,04

Na Tabela 3 é apresentado a matriz confusão dos dois modelos gerados. A classe com melhor resultado foi a *Common-Nevus* com 100% de acertos com o modelo SVM. As classes *Atypical-Nevus* e *Melanoma* foram incorretas 7 e 8 vezes, respectivamente, utilizando o classificador obtido com SVM.

Tabela 3. Matriz Confusão dos modelos MLP e SVM.

RNA (MLP) Classes			SVM (SMO) Classes		
Common-Nevus	Atypical-Nevus	Melanoma	Common-Nevus	Atypical-Nevus	Melanoma
76	4	0	80	0	0
3	71	6	3	73	4
1	8	31	1	7	32

Tabela 4. Área ROC das classes e média ponderada dos modelos obtidos com RNA (MLP) com AutoWeka e SVM (SMO).

Algoritmo	Área ROC	Área ROC
	RNA (MLP) com AutoWeka	SVM (SMO)
Classe Common-Nevus	0,981	0,983
Classe Atypical-Nevus	0,955	0,927
Classe Melanoma	0,959	0,935
Média Ponderada	0,966	0,951

5 Conclusões

Os estudos e testes realizados durante este trabalho demonstraram que a classificação lesões de pele usando algoritmos de Processamento e Análise de Imagens Digitais é difícil, mas é possível obter resultados relevantes a partir do uso de metodologias e ferramentas disponíveis na literatura. Notou-se também, que os classificadores RNA e SVM são muito usadas na classificação de imagens médicas, e produziram resultados satisfatórios para classificar as lesões de pele usadas como teste neste trabalho. Durante o desenvolvimento deste trabalho, percebeu-se que a escolha e cálculo dos atributos, bem como a configuração dos parâmetros e estrutura dos classificadores, influenciam significativamente na acurácia dos resultados obtidos. Para superar o obstáculo da configuração dos classificadores, a ferramenta Weka mostrou-se bastante útil, principalmente por oferecer a funcionalidade do AutoWeka, que ajuda a otimizar muito o classificador. Como trabalhos futuros, sugere-se fazer um estudo mais detalhado sobre a escolha e cálculo dos atributos usados como entrada para o classificador para buscar aumentar os índices de acerto da classificação.

Agradecimentos

Os autores agradecem ao Instituto Federal de Ciência, Educação e Tecnologia de Mato Grosso do Sul pelo apoio e suporte durante o desenvolvimento deste trabalho, disponibilizando também acesso aos laboratórios de Informática nos Campi Coxim e Três Lagoas para uso da ferramenta Weka e realização dos testes.

Referências

[1] Fucci, R. R.; Padovani, C. R.; Araujo, A. F. de; Nunes, G. M. "Classificação De Imagens Do Pantanal Usando Redes Neurais Artificiais Mlp E Perceptron Linear". 10.20906/CPS/CILAMCE2015-0613. 2015.

- [2] Dos Santos; G. S, Araujo, A. F. Estudo sobre classificação de lesões de pele com RNA-MLP. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics (CNMAC 2018). V. 06. N. 02.
- [3] Nacereddine, N.; Goumeidane A. B.; Ziou, D.. Unsupervised weld defect classification in radiographic images using multivariate generalized Gaussian mixture model with exact computation of mean and shape parameters, Computers in Industry, Volume 108, 2019, Pages 132-149, ISSN 0166-3615
- [4] Meneses, P. R.; Almeida, T. D. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Universidade de Brasília, Brasília, 2012, 266 p.
- [5] Hosni, M.; Abnane, I.; Idri, A.; De Gea, J. M. C.; Alemán, J. L. F., Reviewing ensemble classification methods in breast cancer, Computer Methods and Programs in Biomedicine, Volume 177, 2019, Pages 89-112, ISSN 0169-2607.
- [6] Remeseiro, B., Bolon-Canedo, V. A review of feature selection methods in medical applications, Computers in Biology and Medicine, Volume 112, 2019, 103375,ISSN 0010-4825
- [7] Mendonça, Teresa; Ferreira, P. M.; Marques, J., Marcal, A. R. S.; Rozeira, J.. PH² - 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.
- [8] Ponzoni, Flávio Jorge; Almeida, E. S. A estimativa do parâmetro kappa (K) da análise multivariada discreta no contexto de um SIG. Simpósio Brasileiro de Sensoriamento Remoto, v. 8, p. 52-58, 1996.
- [9] Witten, I. H. e Frank, E. (2011). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- [10] Landis, J. R. e Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, (pp. 159–174).
- [11] Kotthoff, Lars et al. Auto-WEKA 2.0: Automatic model selection and hyperparameter optimization in WEKA. The Journal of Machine Learning Research, v. 18, n. 1, p. 826-830, 2017.