

Aplicação de Redes de Aprendizado Profundo e Algoritmos de Aprendizado de Máquina para Classificar Imagens de Câncer de Pele

João P. C. A. de Sá¹, Leandro A. Ensina², Daniel C. Jeronymo¹

¹Universidade Tecnológica Federal do Paraná (UTFPR)
Toledo – PR – Brasil

²Faculdade Assis Gurgacz (FAG)
Toledo – PR – Brasil

joaosa@alunos.utfpr.edu.br, leandroensina@fag.edu.br, danielc@utfpr.edu.br

Abstract. *Skin cancer represents a significant challenge for public health, demanding technologies that aid in its diagnosis. In this work, we propose a method that combines deep learning algorithms, which we used for extracting features from skin cancer images, and traditional machine learning algorithms, which we used as classifiers. We used the HAM10000 dataset in our experiments, enabling us to perform multiclass classification. We utilized the VGG convolutional network as a feature extractor, which served as input for algorithms such as Perceptron and LR to perform the classification task. The results demonstrated the effectiveness of our method and supported the continuity of our research.*

Resumo. *O câncer de pele representa um grande desafio para a saúde pública, demandando de tecnologias que auxiliem em seu diagnóstico. Neste trabalho, propomos um método que combina algoritmos de aprendizagem profunda, utilizados para a extração de atributos de imagens de câncer de pele, e algoritmos de aprendizado de máquina tradicionais, utilizados como classificadores. Para isso, empregamos em nossos experimentos o conjunto de dados HAM10000 que nos permite realizar a classificação multiclasse. Utilizamos a rede convolucional VGG como extrator de atributos, os quais foram usados como entrada para algoritmos como Perceptron e LR realizarem a tarefa de classificação. Os resultados alcançados demonstram boa efetividade do nosso método, amparando a continuidade de nossa pesquisa.*

1. Introdução

O câncer representa um grande desafio para a saúde pública devido a sua complexidade, diversidade e impacto. Dentre os tipos mais comuns de câncer no mundo está o câncer de pele, cuja incidência vem aumentando gradualmente a cada ano [World Health Organization 2022]. O tipo de câncer de pele mais preocupante é o melanoma em razão aos seus altos índices de mortalidade e, em virtude da sua rápida evolução, seu diagnóstico faz com que a sua detecção precoce seja fundamental [O'Neill and Scoggins 2019]. Frente a isso, a dermatoscopia destaca-se como um método de diagnóstico não invasivo para a avaliação da pele por meio de imagens, permitindo a observação até mesmo das camadas mais profundas da pele [Giuffrida et al. 2020].

Com o avanço da tecnologia, algoritmos de inteligência artificial estão sendo aplicados diretamente na análise das imagens geradas pela dermatoscopia, acelerando o processo de identificação de câncer de pele e auxiliando os médicos dermatologistas no diagnóstico desta grave doença. Atualmente podemos encontrar diversos bancos de imagens de câncer de pele online, como o PH2 [Mendonça et al. 2013], o ISIC ISBI 2016 [Gutman et al. 2018], o HAM10000 [Tschandl 2018], entre outros.

Diferentes trabalhos na área já foram desenvolvidos a partir de algoritmos de aprendizado de máquina para a classificação de câncer de pele. Os autores de [Pathan et al. 2019] propuseram um método que extrai características definidas explicitamente pelos próprios pesquisadores, selecionando aqueles atributos mais relevantes para então fazerem a classificação entre melanoma e não melanoma. Por outro lado, [Layode et al. 2019] introduziram um método que extrai os atributos automaticamente com a utilização de redes de aprendizado profundo. Já os autores de [Auxilia et al. 2022] implementaram e avaliaram uma rede baseada em aprendizado profundo em conjunto com técnicas de *data augmentation* para a classificação das lesões de pele. Por fim, os autores de [Spolaôr et al. 2024] utilizaram a técnica de *fine-tuning* em uma rede pré-treinada VGG para avaliar seu método em conjuntos pequenos de dados.

A classificação do câncer de pele como multiclasse ainda é um grande desafio, justamente pela pequena quantidade de exemplos disponíveis para treinamentos e avaliação dos algoritmos. Além disso, o uso de algoritmos de aprendizado profundo, como as redes neurais convolucionais, tem demonstrado boa efetividade para a extração de características de imagens, especialmente do domínio médico. Logo, podemos utilizar estes atributos como entrada para algoritmos de classificação convencionais, como *K-Nearest Neighbors* (KNN) e *Support Vector Machine* (SVM).

Este trabalho tem como objetivo classificar o tipo de câncer de pele a partir da combinação de algoritmo de aprendizado profundo, utilizado como extrator de atributos, e algoritmos tradicionais de aprendizado de máquina, utilizados como classificadores. Os resultados alcançados até o momento demonstram boa efetividade do nosso método e nos permite amparar e fomentar nossos próximos passos na pesquisa.

A organização do restante do documento está estruturada da seguinte maneira. Na Seção 2 descrevemos o método proposto para a resolução do problema de classificação multiclasse de câncer de pele, enquanto na Seção 3 descrevemos o conjunto de experimentos realizados. Já na Seção 4 apresentamos e discutimos os resultados parciais alcançados, juntamente com direcionamentos para a continuidade desta pesquisa. Por fim, na Seção 5 realizamos a conclusão do nosso trabalho.

2. Método Proposto

Nosso método divide-se em duas etapas: (1) extração de características por um algoritmo de aprendizado profundo e (2) classificação. O emprego de redes profundas como extratores de características está relacionado a sua capacidade de reconhecimento de características relevantes em imagens.

Para a primeira etapa, utilizamos a rede neural convolucional VGG19 [Simonyan and Zisserman 2015] unicamente como extratora de atributos. Em síntese, a VGG 19 possui 16 camadas de convolução e 3 camadas totalmente conectadas. A escolha inicial desta rede se deve ao seu bom desempenho para o problema de câncer de pele

conforme amparado pela literatura [Spolaôr et al. 2024]. Cabe ressaltar que esta rede foi treinada para um problema multiclasse, vide base de dados descrita na Seção 3. A rede foi usada com os pesos de seu pré-treinamento na ImageNet. Adicionalmente, neste estudo não realizamos *fine-tuning*.

Após a extração dos atributos pela VGG, estas características servirão como entrada para diferentes algoritmos de classificação, como *Linear Discriminant Analysis* (LDA), *K-Nearest Neighbors* (KNN) com 5 vizinhos, *Support Vector Machine* (SVM), *Perceptron* e *Logistic Regression* (LR), além dos algoritmos *ensemble Random Forest* e *Extra Trees* com 100 estimadores cada [Alpaydin 2020].

3. Protocolo Experimental

As imagens utilizadas são provenientes do banco de imagens HAM10000 [Tschandl 2018], o qual é composto de mais de 10.000 imagens de lesões de pele, que podem ser classificadas entre 327 imagens de queratoses actínicas, 514 de carcinoma basocelular, 1.099 de lesões benignas do tipo queratose, 115 dermatofibroma, 1.113 de melanoma, 6.705 de nevos melanocíticos e 142 de lesões vasculares. Diferente dos outros dois bancos de imagens citados na Seção 1 (PH2 e ISIC ISBI 2016), a base HAM10000 possui sete diferentes classes, enquanto as bases PH2 e ISIC ISBI 2016 possuem apenas três classes, sendo comumente transformadas e usadas pelos trabalhos da literatura para a resolução de classificação binária.

Para garantir uma avaliação robusta do desempenho dos modelos de classificação, foi adotado o método de validação cruzada estratificada 5×2 (5×2 *stratified cross-validation*). A escolha por esta abordagem baseia-se pela natureza desbalanceada da base de dados HAM10000, conforme protocolo já utilizado pela literatura [Roy et al. 2018]. Este método é útil quando lidamos com conjuntos de dados desbalanceados, como é o caso do conjunto de imagens dermatológicas HAM10000, onde as classes têm números variados de amostras.

Os desempenhos neste trabalho são avaliados de acordo com as métricas a seguir: acurácia (Equação 1), precisão (Equação 2), *F1-Score* (Equação 3) e especificidade (Equação 4). As médias dessas medidas em todos os experimentos 5×2 estão relatados nos resultados dispostos na Seção 4.

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

$$Precisão = \frac{VP}{VP + FP} \quad (2)$$

$$F1 - Score = \frac{VP}{VP + \frac{FN+FP}{2}} \quad (3)$$

$$Especificidade = \frac{VN}{VN + FN} \quad (4)$$

sendo VP, VN, FP e FN os valores de verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, respectivamente.

4. Resultados Parciais

Na Figura 1 estão apresentados os desempenhos alcançados para cada algoritmo mencionado na Seção 2, incluindo a própria VGG19 como classificador. Para justificar o motivo de usarmos a rede VGG19 apenas para extrair atributos e os algoritmos tradicionais de aprendizado de máquina como classificadores, também a testamos para realizar a classificação.



Figura 1. Desempenho dos classificadores utilizados.

Os resultados alcançados demonstram a boa efetividade de nosso método frente ao simples uso da VGG19 para realizar todo o processo, incluindo a própria classificação. O F1-Score obtido para a VGG19 foi de $43\% \pm 0,38\%$, enquanto os algoritmos Perceptron, SVM e LR alcançaram $72\% \pm 0,24\%$, $72\% \pm 0,58\%$ e $71\% \pm 0,68\%$, respectivamente. A acurácia de $74\% \pm 0,51\%$ da VGG19 demonstra a tendência da classificação deste algoritmo para a classe majoritária, o que justifica em parte o valor inferior para o F1-Score comparado aos demais algoritmos.

Apesar dos desempenhos dos algoritmos *Random Forest* e *Extra Trees* (F1-Score de $62\% \pm 0,49\%$ e $61\% \pm 0,66\%$, respectivamente) serem inferiores ao de algoritmos como Perceptron e LR, pretendemos investigar mais a fundo o uso de algoritmos *ensemble* em busca de resultados mais acurados e precisos. Neste aspecto, podemos utilizar os algoritmos que demonstraram os melhores resultados (*e.g.*, Perceptron e LR) como preditores base para gerarmos novos conjuntos de modelos *ensemble*. Para isso, técnicas como *Bagging* e *Random Subspace Method* serão úteis para a geração de conjuntos de modelos múltiplos preditivos diversificados [Breiman 1996, Ho 1998]. Tal geração poderá ser

homogênea, quando o mesmo algoritmo base é usado, ou heterogênea, quando diferentes algoritmos base são treinados e combinados no conjunto gerado.

Neste cenário, iremos explorar o emprego da seleção dinâmica de classificadores a partir dos modelos provenientes do conjunto de classificadores gerado [Britto et al. 2014, Cruz et al. 2018]. Em suma, na seleção dinâmica um único modelo ou um grupo de modelos é selecionado para classificar individualmente cada caso de teste desconhecido. Esta técnica é uma das abordagens mais promissoras do *ensemble*, uma vez que permite selecionar, em tempo real, os modelos base mais adequados (competentes) para classificar individualmente cada novo caso de teste. Para isso, essa técnica estima a competência de cada classificador do conjunto, de modo que apenas o(s) mais competente(s) seja(m) selecionado(s) para predizer o rótulo de um dado exemplo.

A motivação para o emprego da seleção dinâmica também é justificada pelo fato de que seu uso pode acarretar no ganho de desempenho para bases de dados desbalanceadas [Roy et al. 2018], cenário que a base de dados utilizada em nosso trabalho (HAM10000) está inserida (vide Secção 3). Entre as técnicas de seleção dinâmica que pretendemos avaliar estão o META-DES e o KNORA [Cruz et al. 2018].

5. Conclusão

Propomos neste estudo um método que combina algoritmos de aprendizado profundo com algoritmos tradicionais para a classificação de imagens de câncer de pele. O algoritmo de aprendizado profundo inicialmente considerado foi o VGG19 devido ao seu bom desempenho já embasado pela literatura para o problema em questão (*i.e.*, câncer de pele) [Spolaôr et al. 2024].

Os resultados alcançados demonstram que o emprego de algoritmos tradicionais, especialmente Perceptron, SVM e LR, atingiram desempenhos superiores ao utilizarem os atributos extraídos pela VGG19 frente ao uso da própria VGG para realizar todo o processo de classificação. Tais resultados demonstram que nosso projeto é promissor e eficaz, atendendo ao objetivo proposto.

Trabalhos futuros incluem (1) avaliar outros algoritmos de aprendizado profundo como o Xception para a extração de atributos, (2) analisar de modo abrangente o emprego de algoritmos *ensemble*, em especial (3) explorar a técnica de seleção dinâmica de classificadores, como o META-DES e o KNORA, visto que estes métodos tendem a obter melhores resultados ao classificar *datasets* desbalanceados [Roy et al. 2018], como é o caso do banco de imagens HAM10000.

Referências

- Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press, Cambridge, 4 edition.
- Auxilia, V., Meenakshi, A., Prabahavathy, P., and Shamreen, A. B. (2022). Skin lesion classification and prediction by data augmentation in ham10000 and isic 2019 dataset. <https://doi.org/10.21203/rs.3.rs-2242686/v2>. Acessado em 05/05/2024.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- Britto, A. S., Sabourin, R., and Oliveira, L. E. S. (2014). Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, 47(11):3665–3680.

- Cruz, R. M., Sabourin, R., and Cavalcanti, G. D. C. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, 41:195–216.
- Giuffrida, R., Conforti, C., Meo, N. D., Deinlein, T., Guida, S., and Zalaudek, I. (2020). Use of noninvasive imaging in the management of skin cancer. *Current Opinion in Oncology*, 32(2):98–105.
- Gutman, D., Codella, N., Celebi, E., Helba, B., Marchetti, M., Mishra, N., and Halpern, A. (2018). Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 168–172.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.
- Layode, O., Alam, T., and Rahman, M. M. (2019). Deep Learning Based Integrated Classification and Image Retrieval System for Early Skin Cancer Detection. *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pages 1–7.
- Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S., and Rozeira, J. (2013). Ph2 - a dermoscopic image database for research and benchmarking. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5437–5440.
- O’Neill, C. H. and Scoggins, C. R. (2019). Melanoma. *J Surg Oncol*.
- Pathan, S., Prabhu, K. G., and Siddalingaswamy, P. (2019). Automated detection of melanocytes related pigmented skin lesions: A clinical framework. *Biomedical Signal Processing and Control*, 51(1):59–72.
- Roy, A., Cruz, R. M., Sabourin, R., and Cavalcanti, G. D. (2018). A study on combining dynamic selection and data preprocessing for imbalance learning. *Neurocomputing*, 286:179–192.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 6.
- Spolaôr, N., Lee, H. D., Mendes, A. I., Nogueira, C. V., Parmezan, A. R. S., Takaki, W. S. R., Coy, C. S. R., Wu, F. C., and Fonseca-Pinto, R. (2024). Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets. *Multimedia Tools and Applications*, 83:27305–27329.
- Tschandl, P. (2018). The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Harvard Dataverse*, 4.
- World Health Organization (2022). Skin cancer. Disponível em: <https://www.iarc.who.int/cancer-type/skin-cancer>. Acessado em: 28 fev. 2024.