

Otimização de CNN aplicada ao Diagnóstico de Melanoma

**Carlos Vinícios Martins Rocha, Lucas Bezerra Maia, Geraldo Braz Junior,
João Dallyson Sousa de Almeida, Anselmo Cardoso de Paiva**

¹Núcleo de Computação Aplicada, Universidade Federal do Maranhão (UFMA)
Av. dos Portugueses, 1966 - Bacanga, São Luís - MA, 65080-805, Brasil

{carlos.martins, lucas.maia, geraldo, jdallyson, paiva}@nca.ufma.br

Abstract. *Melanoma is the most lethal skin cancer compared to others, however, patients have a high cure rate when diagnosed in their early stages. Exist several approaches to automatic diagnosis and detection proposed by different authors. However, the training of models in small and unbalanced databases presents several obstacles. Thus, this work in progress aims to apply the techniques of transfer of learning to training models capable of assisting in the diagnosis and screening of melanoma. Preliminary results showed that the use of synthetic data generation in conjunction with the fine-tuning of VGG16, showed crucial improvements, approximately 100% sensitivity and 93.75% specificity.*

Resumo. *O melanoma é o câncer de pele mais letal em comparação aos outros, entretanto, os pacientes apresentam um elevado índice de cura quando diagnosticado em seu estágio inicial. Por este motivo, tem sido propostas abordagens para o diagnóstico e detecção automática propostas por diferentes autores. Entretanto, o treinamento de modelos em bases de dados pequenas e não balanceadas apresentam diversos obstáculos. Dessa forma, esse trabalho em andamento tem o objetivo aplicar as técnicas de transferência de aprendizado para treinamento de modelos capazes de auxiliar no diagnóstico e triagem do melanoma. Resultado preliminares mostraram que uso da geração de dados sintéticos em conjunto com o fine-tuning da VGG16, apresentaram melhorias cruciais, aproximadamente 100% de sensibilidade e 93,75% de especificidade.*

1. Introdução

Dentre os tipos de cânceres, o câncer de pele é o mais comum e sua incidência aumenta cada vez mais, por parte, em um reflexo do comportamento humano à exposição excessiva de sol. O melanoma, mesmo sendo considerado o tipo de câncer de pele mais letal, apresenta índices de cura elevados quando diagnosticado em seu estágio inicial. A doença se manifesta em pele normal a partir do surgimento de uma pinta com tonalidade escura e bordas irregulares, ou de uma lesão pigmentada pré-existente, onde o tumor irá evoluir apresentando mudanças de cores e aumento da área da lesão [Soares 2008].

A última estimativa mundial apresentou que 232.130 novos casos de melanoma seriam diagnosticados em todo o globo. Enquanto que, em termos de mortalidade, foram estimados 55.488 óbitos, sendo os países com população caucasiana, os lugares com maiores taxas de incidência. Atualmente, estima-se, para homens e mulheres, respectivamente, 85.170 e 80.410 novos casos de neoplasia não melanoma no Brasil, cujos valores são correspondentes a um risco estimado de 82,53 casos novos a cada 100 mil homens e 75,84 para cada 100 mil mulheres.

As maiores chances de cura do melanoma são quando diagnosticado precocemente, ou seja, quando as células cancerosas ainda estão apenas na superfície da pele. Além da biópsia, existem alguns exames que servem para detecção e prevenção do câncer de pele. Um deles é a dermatoscopia, técnica não-invasiva, no qual o dermatologista usa um aparelho chamado dermatoscópio para fotografar a região da lesão, aumentando-a de 10 a 70 vezes.

Embora exista método por imagem que possa ser analisado pelo médico, ainda consiste num problema de saúde pública conseguir aproximar com agilidade o paciente que necessita do tratamento do médico. O quantitativo de análises pode tornar o acesso proibitivo, assim como cansativo e sujeito a erros para o profissional sobrecarregado. A ideia por trás de sistemas de auxílio ao diagnóstico é inicialmente realizar a triagem de maneira a concentrar os esforços do profissional no casos corretos.

Este trabalho, em andamento, propõe a utilização de técnicas de transfer learning na área de deep learning para descrever e criar modelos preditivos com a finalidade de diagnosticar melanoma através de imagens de dermatoscopia. Nossa contribuição consiste em demonstrar um método para otimização automática de redes neurais, com objetivo de estimar uma rede que acoplada a uma rede base, pré-treinada, tenha a capacidade de generalização mesmo em bases de dados variadas e desbalanceadas.

2. Metodologia

A metodologia proposta se utiliza do conceito de *Transfer Learning* em *Deep Learning* para propôr uma arquitetura que utiliza a rede do *Visual Geometry Group* (VGG16) [Simonyan and Zisserman 2014], pré-treinada como rede base e realiza o *fine tuning* de seus pesos e o acoplamento de novas camadas para resolver o problema específico de diagnóstico de melanoma.

A utilização da mesma se justifica para extrair características básicas, tomando-se como base o modelo baseado na ImageNet. Isso faz com que o treinamento seja mais rápido e ainda garante que características primitivas tenham sido corretamente treinadas. A partir desse modelo, é construída uma nova rede para entender e diagnosticar melanoma a partir de exames dermatoscópicos. O diagrama da metodologia proposta está ilustrado na Figura 1, onde o mesmo demonstra que as bases de dados passaram pelo pré-processamento e *augmentation*, seguidos pela fase de *fine-tuning* e treino da rede e no final a análise dos resultados obtidos. As próximas seções apresentam detalhes da metodologia proposta.

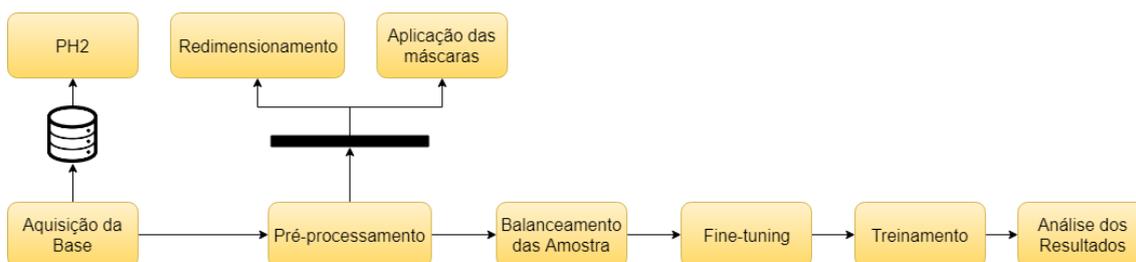


Figura 1. Diagrama da metodologia proposta.

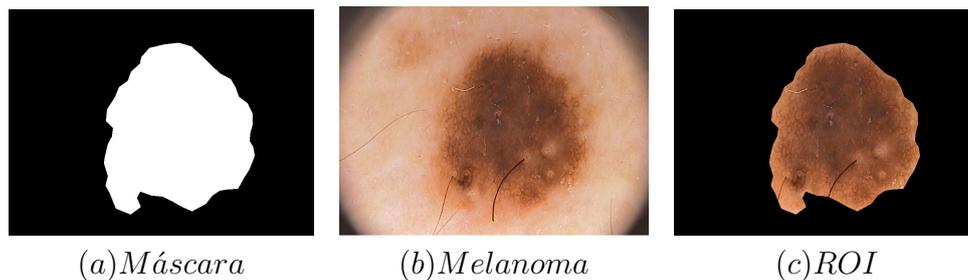


Figura 2. Extração da ROI.

2.1. Aquisição e Pré-Processamento

A base utilizada para estudo foi a PH2 [Mendonça et al. 2013]. A mesma contém 200 imagens dermatoscópicas divididas entre 3 classes: Normal, contendo 80 imagens, Atípica que também apresenta 80 e Melanoma, que apresenta 40 amostras. Para o estudo, foi realizada a união entre as classes Normal e Atípica, resultando em apenas 2 classes principais, Melanoma e Não Melanoma. Assim, resultou em uma base com 160 exemplos de não melanoma e 40 exemplos de melanoma.

A partir das máscaras binárias criadas por especialistas, contendo apenas a região da lesão, as imagens foram pré-processadas removendo-se as regiões da imagem que não são de interesse para o nosso estudo (Figura 2). Dessa forma, o aprendizado da rede é focado apenas na região de interesse (ROI), evitando que características da pele interfiram no processo de aprendizagem. Portanto, as máscaras são aplicadas às imagens originais, mantendo apenas o ROI.

Após o processo de extração do ROI, devido ao fato de as imagens apresentarem tamanhos diferentes, foi realizado uma fase de redimensionamento, em que as mesmas foram redimensionadas para 768 largura por 578 de altura.

2.2. Balanceamento da Amostra

A base PH2 é severamente desbalanceada. Percebe-se uma grande diferença entre as classes, em termos quantitativos, onde a quantidade de amostras da classe normal é sempre superior às de melanoma. Essa diferença interfere no processo de treinamento de nossa rede neural de convolução, fazendo com que o aprendizado acerca da menor classe seja afetado negativamente.

Visando contornar tal situação, foi utilizado um algoritmo que tem como objetivo, realizar um aumento no número de exemplos, contidos na pasta de treino, da menor classe. Esse algoritmo faz uso da classe ImageDataGenerator [Ima], encontrado na biblioteca do Keras [Chollet et al. 2015], para gerar as imagens. E além disso, sempre gera o número necessário de amostras para que ambas as classes estejam o mais próximo possível da igualdade.

A função apresenta certos parâmetros que permite a criação de uma variedade de imagens sintéticas, entretanto, foram utilizados apenas uma quantidade limitada. Dentre eles:

1. Intervalo de deslocamento vertical e horizontal, variando até 0,05;
2. Faixa de cisalhamento, com 0,02 graus de inclinação;

3. Rotação com 180 graus de limite;
4. Preenchimento constantes nos pontos fora dos limites das imagens, preenchendo os pontos fora dos limites com 0;
5. Limite de zoom em 0,3;
6. Permitindo o giro horizontal.

A utilização desses parâmetros tem como objetivo a criação de imagens do melanoma, que apresentem as deformidades das bordas em posições variadas e sempre garantindo que as características principais das mesmas não sejam perdidas durante o processo. Alguns exemplos das imagens geradas são apresentados na Figura 3.



Figura 3. Balanceamento da base via geração sintética de imagens

2.3. Fine-Tuning e Treinamento

O treinamento de uma *Convolutional Neural Network* (CNN) com bases pequenas é um problema devido a grande quantidade de parâmetros e pesos a serem ajustados na rede. Para solucionar tal problema de inicialização da rede, a literatura atual opta para o uso do método de transferência de aprendizado (*Fine-tuning*), que consiste em usar uma rede que já foi pré-treinada e dar continuidade ao processo de treinamento usando a nossa base específica, garantindo, assim, uma melhoria na capacidade de reconhecimento e classificação do modelo.

Portanto, foi utilizada uma rede pré-treinada, VGG16, como base para a realização da transferência de aprendizagem. Ela apresenta seus pesos treinados na ImageNet, que é uma extensa base de dados de imagens, contendo uma variação enorme de amostras e classes. Dessa forma, deve-se congelar as camadas iniciais da rede. O congelamento tem como objetivo, garantir que as características mais gerais pré-aprendidas pela rede não sejam perdidas, assegurando que a rede tenha um melhor aprendizado das características mais específicas condizentes às nossas bases de dados.

Após a rede base, é construída uma nova rede com a finalidade de aprender características específicas do problema de diagnóstico de melanoma. Para este estudo em andamento, foi proposto uma arquitetura simples.

A mesma é composta por camadas de *pooling* [Zeiler and Fergus 2013] com o objetivo de reduzir a dimensionalidade do mapa de características gerado pela rede pré-treinada, de *Dropout* [Baldi and Sadowski 2014] para aumentar a generalização da rede e camadas convolucionais e totalmente conectadas para extrair novas características e realizar a classificação respectivamente. Todas as redes foram treinadas com classificação baseada na função Sigmoid e com ativações internas usando a função *Rectified Linear Unit* (ReLU) [Xu et al. 2015].

Tabela 1. Variações de arquiteturas testadas e avaliadas

| Arquitetura | Pooling | Qtd Convolução | Qtd Densas | Dropout |
|-------------|---------|----------------|------------|---------|
| Modelo 1 | Média | 0 | 1 | 0,5 |
| Modelo 2 | Média | 1 | 1 | 0,3 |
| Modelo 3 | Média | 0 | 2 | 0,4 |
| Modelo 4 | Média | 0 | 2 | 0,5 |

A Tabela 1 apresenta combinações de arquiteturas testadas afim de aferir o desempenho no diagnóstico de melanoma. Para primeira etapa de compilação, foram utilizados o otimizador RMSProp [Mukkamala and Hein 2017] e a função de perda *Cross Entropy* [De Boer et al. 2005], entretanto, na segunda, alterou-se o otimizador da rede para *Stochastic Gradient Descent* (SGD) [Klein et al. 2009].

3. Resultados

Para realizar os experimentos, utilizou-se *hold-out* para dividir a base em 5 *splits* diferentes, onde cada um apresenta 3 partes distintas, sendo elas: treino, validação e teste. Cada parte apresenta uma porcentagem das imagens existentes na base, de forma que os valores correspondentes equivalem a 100% das amostras. Os percentuais usados foram: 80% para treino, 10% para teste e 10% para validação. Então a etapa de balanceamento foi realizada usando apenas os dados de treinamento de cada *split*. Em seguida as arquiteturas foram construídas e testadas atingindo os resultados apresentados na Tabela 2.

Tabela 2. Métricas obtidas na validação dos modelos.

| | Acurácia (%) | Precisão (%) | F1 (%) | Sensibilidade (%) | Especificidade (%) |
|----------|--------------|--------------|--------|-------------------|--------------------|
| Modelo 1 | 85,00 | 83,92 | 84,16 | 50,00 | 93,75 |
| Modelo 2 | 90,00 | 90,00 | 90,00 | 75,00 | 93,75 |
| Modelo 3 | 90,00 | 93,33 | 90,67 | 100,00 | 87,50 |
| Modelo 4 | 95,00 | 96,00 | 95,20 | 100,00 | 93,75 |

Dessa forma, podemos perceber que o acréscimo de camadas densas interfere positivamente nos valores de sensibilidade, justificado pelos modelos 3 e 4 contidos na Tabela 2, diferentemente do observado nos modelos 1 e 2.

Com base nos dois últimos modelos presentes na Tabela 2, notamos que não se faz necessário o uso de camadas de convolução para a obtenção dos melhores valores de acurácia e sensibilidade, entretanto, a variação do índice de *dropout* resultou em melhorias relevantes, visto que, ao aumentar no modelo 4, em relação ao modelo 5, tivemos um ganho de 6,25% na especificidade.

Assim, temos o modelo 4 como o de melhor resultado, visto que, além de apresentar um alto valor de acurácia, explicita um alto valor de sensibilidade, sendo este o mais importante. Pois, apresentar a capacidade de reconhecer a doença em um paciente de risco, julga-se mais importante do que apontar um falso positivo.

4. Conclusões e passos futuros

O presente trabalho em andamento apresentou uma técnica para transferência de aprendizado e tratamento de bases não balanceadas, com foco no diagnóstico de câncer de pele

do tipo Melanoma. Tendo como objetivo, dar suporte aos profissionais, por meio de triagens e auxiliando nas decisões, a fim de que os mesmos concentrem esforços em casos corretos.

A metodologia proposta neste trabalho está em andamento e os próximos passos consistem em otimizar automaticamente, via *Tree-structured Parzen Estimator* (TPE) a escolha da melhor arquitetura (coleção de camadas e hiperparâmetros) após a rede base e ainda testes com a base ISIC-Archive.

Referências

- Image preprocessing. <https://keras.io/preprocessing/image/>. Accessed: 2018-12-05.
- Baldi, P. and Sadowski, P. (2014). The dropout learning algorithm. *Artificial intelligence*, 210:78–122.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- De Boer, P.-T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of operations research*, 134(1):19–67.
- Klein, S., Pluim, J. P., Staring, M., and Viergever, M. A. (2009). Adaptive stochastic gradient descent optimisation for image registration. *International journal of computer vision*, 81(3):227.
- Mendonça, T., Ferreira, P. M., Marques, J. S., Marçal, A. R. S., and Rozeira, J. (2013). Ph2 - a dermoscopic image database for research and benchmarking. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2013:5437–40.
- Mukkamala, M. C. and Hein, M. (2017). Variants of rmsprop and adagrad with logarithmic regret bounds. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2545–2553. JMLR. org.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Soares, H. B. (2008). *Análise e classificação de imagens de lesões da pele por atributos de cor, forma e textura utilizando máquina de vetor de suporte*. PhD thesis, Universidade Federal do Rio Grande do Norte.
- Xu, B., Wang, N., Chen, T., and Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.
- Zeiler, M. D. and Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks. *arXiv preprint arXiv:1301.3557*.