

Uma Abordagem de Classificação de Imagens Dermatoscópicas Utilizando Aprendizado Profundo com Redes Neurais Convolucionais

Alan Santos¹, Kelson Aires¹, Rodrigo Veras¹, Valeska Uchôa¹, Luís Santos¹

¹Núcleo de Computação – Universidade Federal do Piauí – UFPI

Caixa Postal 64.049 – 550 – Teresina – PI – Brasil

email: {kelson,rveras}@ufpi.edu.br, {alan.rafael18,valeskauchoa,luisteixeira}@gmail.com

Abstract. *The automatic identification of melanoma in dermoscopic images is still a problem difficult to handle. In this regard, this work proposes an approach to classify these images, using deep learning with convolutional neural networks. The Alexnet and VGG-F architectures were trained with dermoscopic images, and used as models to describe characteristics, through the last layers totally connected. The classification of the extracted data was performed by some traditional algorithms. On the best result of the approach, the classified k-nearest neighbours reached an accuracy of 91,5%, with kappa index of 0,83 and recall and precision higher than 0,9.*

Resumo. *A identificação automática do melanoma em imagens dermatoscópicas ainda é um problema de difícil solução. Nesse contexto, esse trabalho propõe uma abordagem de classificação dessas imagens, utilizando Aprendizado Profundo com Redes Neurais Convolucionais. As arquiteturas Alexnet e VGG-F foram treinadas com imagens dermatoscópicas, em seguida, utilizadas como modelo para descrição de características, através das últimas camadas totalmente conectadas. A classificação dos dados extraídos foi realizada por alguns algoritmos tradicionais. No melhor resultado da abordagem, o classificados K-Vizinhos mais Próximos atingiu uma acurácia de 91,5%, com índice kappa igual a 0,83, recall e precisão maiores que 0,9.*

1. Introdução

A incidência do câncer de pele do tipo melanoma é a mais agressiva e letal no Brasil, correspondendo a 9% de todos os casos de câncer registrados [INCA 2016]. O diagnóstico não invasivo dessa doença é realizado através do exame dermatoscópico, que produz imagens digitais das lesões. Essas imagens são analisadas posteriormente por um médico dermatologista, que determina as chances dessas lesões serem consideradas melanoma. Isso evita a indicação de incisões desnecessárias.

A forma mais utilizada pelos dermatologistas para identificar o melanoma em imagens dermatoscópicas é baseada no padrão visual ABCD (Assimetria, Borda, Cor e Diâmetro). As variações das características desse padrão permitem definir se uma lesão é ou não um câncer melanômico. O processo de identificação das características necessárias é uma tarefa que exige uma certa experiência do especialista, além do diagnóstico preciso.

Vários sistemas já foram desenvolvidos na literatura para ajudar o dermatologista no diagnóstico do melanoma. Essas ferramentas possibilitam uma identificação dos atributos relevantes da doença, favorecendo uma segunda opinião no diagnóstico. A maioria

dessas abordagens são inspiradas no padrão visual ABCD, com a identificação automática das características das lesões nas imagens dermatoscópicas. A classificação desses dados determina a presença ou ausência do melanoma [Maglogiannis and Doukas 2009]. Novas formas estão substituindo esse padrão visual por abordagens de Aprendizado Profundo, utilizando Redes Neurais Convolucionais (RNCs). Esse tipo de rede permite formas de transformação espacial em imagens, com o intuito de produzir apenas dados relevantes no aprendizado e classificação [Shi and He 2011].

O treinamento de Redes Neurais Convolucionais pode gerar modelos robustos que conseguem abstrair informações mais precisas na distinção de imagens em classes, por determinar apenas características espaciais relevantes [Shi and He 2011]. Nesse contexto, o objetivo desse trabalho é desenvolver uma abordagem de Aprendizado Profundo com essas redes para identificar o melanoma em imagens dermatoscópicas, treinando algumas arquiteturas e, em seguida, utilizando-as como modelos para extração de aspectos relevantes das lesões.

Os dados extraídos com as arquiteturas sugeridas serão utilizados para determinar o diagnóstico do melanoma, através dos métodos convencionais de Máquina de Vetor de Suporte (*Support Vector Machine* - SVM), K-Vizinhos mais Próximos (*K-Nearest Neighbor* - KNN) e Perceptron Multicamadas (*MultiLayer Perceptron* - MLP), sugeridos por Maglogiannis et al. (2009)

As principais contribuições desse trabalho seriam a inclusão do Aprendizado Profundo em novas abordagens que envolvem imagens dermatoscópicas e o desenvolvimento de novas ferramentas de extração de características com Redes Neurais Convolucionais que sejam específicas para esse problema e que possam oferecer um ótimo desempenho computacional.

2. Trabalhos Relacionados

As RNCs com aprendizado hierárquico tem conduzido grandes avanços em muitas tarefas de análise que envolvem imagens médicas, incluindo a identificação de atributos relevantes para a classificação de lesões de pele. Alguns pesquisadores estudam essas redes, visando aproveitar a sua capacidade discriminativa para alcançar melhores resultados [Shi and He 2011].

Em Codella et al. (2015) foi proposta uma abordagem de classificação de imagens dermatoscópicas utilizando RNCs pré-treinadas. Esse método aborda uma combinação de Aprendizado Profundo, *sparse coding* e SVM para reconhecimento do melanoma. Os resultados obtidos nos testes com essas redes pré-treinadas atingiram uma acurácia de 72,3%. Essa abordagem considera apenas os recursos de treinamento das redes com imagens diversas.

Na proposta de Majtner et al. (2016) foi desenvolvida uma abordagem de reconhecimento de lesões de pele com uma combinação de redes convolucionais e recursos convencionais de classificação. Na etapa de extração de características essa proposta sugere o uso da arquitetura de convolução Alexnet pré-treinada. Nesse trabalho, as estimativas de dados são dadas pelas últimas camadas dessa arquitetura. Esses são classificados utilizando o algoritmo SVM, atingindo uma acurácia de 80,5%. Essa abordagem desconsidera as características espaciais do melanoma na extração de características, por utilizar modelos treinados com imagens diversas.

No trabalho de Kawahara et al. (2016) é sugerida uma classificação do melanoma sem realizar nenhum tipo de pré-processamento ou segmentação de imagens dermatoscópicas. Essa abordagem também utiliza uma estrutura de rede de convolução pré treinada para descrever características. O diferencial dessa abordagem é a aplicação de uma normalização em lotes de imagens antes de extrair seus recursos. Os resultados obtidos nos testes atingiram uma precisão de 81,8% na identificação das lesões.

As abordagens de classificação do melanoma que não utilizam as estruturas espaciais das lesões no treinamento das redes de convolução, não são capazes de lidar com um diagnóstico eficaz. Nesse sentido, a principal contribuição desse trabalho é o uso dessas características espaciais nesse treinamento, com objetivo de obter modelos treinados que possam oferecer um melhor descrição de imagens dermatoscópicas. Esses dados poderiam possibilitar uma melhor precisão na classificação das lesões.

3. Métodos

O processo de reconhecimento automático do melanoma é um desafio, pois existe um alto grau de similaridade visual entre as lesões de melanoma e não-melanoma. Essas semelhanças podem influenciar no reconhecimento da doença [Gutman et al. 2016]. Nesse sentido, o presente trabalho propõe uma forma de auxiliar na resolução dessa problemática, utilizando Redes Neurais Convolucionais para desenvolver um aprendizado na estrutura das lesões. O intuito é desenvolver modelos capazes de extrair informações relevantes. Nas próximas seções serão apresentadas algumas noções básicas sobre a proposta, as arquiteturas de convolução sugeridas e sua utilização para extração de características.

3.1. Arquiteturas de Redes Neurais Convolucionais e Extração de Características

As RNCs são tipos de redes neurais artificiais que utilizam multicamadas complexas e profundas, que são capazes de reconhecer elementos em imagens. A principal camada dessas redes é a convolução. A sua função é aplicar máscaras nas imagem de entrada, com base em uma vizinhança de *pixels*. A saída produzida nessa operação são os filtros de convolução (matrizes) que armazenam os pesos das conexões entre os neurônios [Parkhi et al. 2015].

A convolução é representada por uma operador linear entre duas funções que produz uma terceira função. Essa representa as áreas de superposição das máscaras que são aplicadas na imagem. A Equação 1 denota uma convolução em relação a um tempo t ,

$$h(v) = (w * x)(t), \quad (1)$$

onde $h(v)$ é a combinação linear dos neurônio em relação a os pesos w na vizinhança x . A distinção de padrões de dados com essa operação é dada por campos receptivos locais (pequenas porções espaciais de *pixels*), que são representados por estruturas de neurônios localmente conectadas [Shi and He 2011].

O compartilhamento de pesos na camada de convolução garante que os filtros sejam aplicados em diferentes posições na imagem, diminuindo significativamente o número de dados a serem aprendidos [Kawahara et al. 2016].

A ativação é uma estrutura muito importante na camada de convolução. Essa realiza um ajuste entre um conjunto de neurônios $y_i(v)$, com uma função de ativação f

sobre uma convolução $h(v)$. Isso produz os mapas de características que armazenam as informações aprendidas dos filtros [Parkhi et al. 2015]. Essa ativação é representada pela Equação 2.

$$y_i(v) = f(h(v)) \quad (2)$$

Outra camada que também é muito importante nas RNCs é a *pooling*. Ela é responsável por reduzir a dimensionalidade dos mapas de características, diminuindo a largura e a altura dos mapas de características. A operação de *pooling* possibilita uma invariância espacial. O agrupamento de características, na maioria das arquiteturas de convolução, utilizam uma função de *Max pooling*. Essa determina o valor máximo de agrupamento em uma vizinhança x . A função *Max pooling* é descrita na Equação 3 [Simonyan and Zisserman 2014].

$$f(x) = \max(x, 0) \quad (3)$$

As próximas camadas das RNCs desempenham o papel de regressão das ativações. Em qualquer rede desse tipo, após a camada de *pooling*, é necessário ao menos uma camada totalmente conectada (*Fully Connected* - FC). Elas servem para criar caminhos de decisões, a partir dos filtros obtidos na camada anterior [Vedaldi and Lenc 2015].

A última camada das redes convolucionais também é totalmente conectada. Essa, por sua vez, é responsável por realizar a classificação dos dados. Nessa situação, uma função determina a identificação das saída em classes. Geralmente, a mais utilizada é a *Softmax*, que é representada na Equação 4,

$$y_i = \frac{e^{z_i}}{\sum_{j \in \text{classe}} e^{z_j}}, \quad (4)$$

onde y_i é probabilidade da saída de um neurônio com índice i , j os índices de saída da camada, e^{z_i} o vetor de saída do neurônio e e^{z_j} o vetor de saída da camada. A *Softmax* determina a probabilidade um conjunto de dados pertencer a uma classe de acordo com sua ocorrência [Parkhi et al. 2015].

O treinamento das RNCs é realizado na maioria dos casos com o *backpropagation*, que ajusta os pesos w dos neurônios pelo erro e , utilizando os componentes do vetor gradiente. A operação é representada na Equação 5,

$$w_{i+1} = w_i + \eta e_i X_i, \quad (5)$$

onde w_{i+1} é o peso do neurônio atual, w_i o peso do neurônio anterior, η o aprendizado da rede, e_i o erro e X_i a entrada [Vedaldi and Lenc 2015].

As arquiteturas de RNCs são criadas para adequar a uma quantidade de imagens e classes, proporcionando um maior reconhecimento de padrões. Vários estudos investigam o tamanho dessas arquiteturas e a sua eficácia. Nesse sentido, a *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) avalia abordagens de identificação e classificação que utilizam redes profundas. Uma das suas motivações é promover arquiteturas que obtêm grandes progressos [Russakovsky et al. 2015]. Em algumas das suas divulgações,

as arquiteturas de RNCs Alexnet e VGG-F aparecem com bons resultados. A Figura 1 apresenta um esboço dessas arquiteturas.

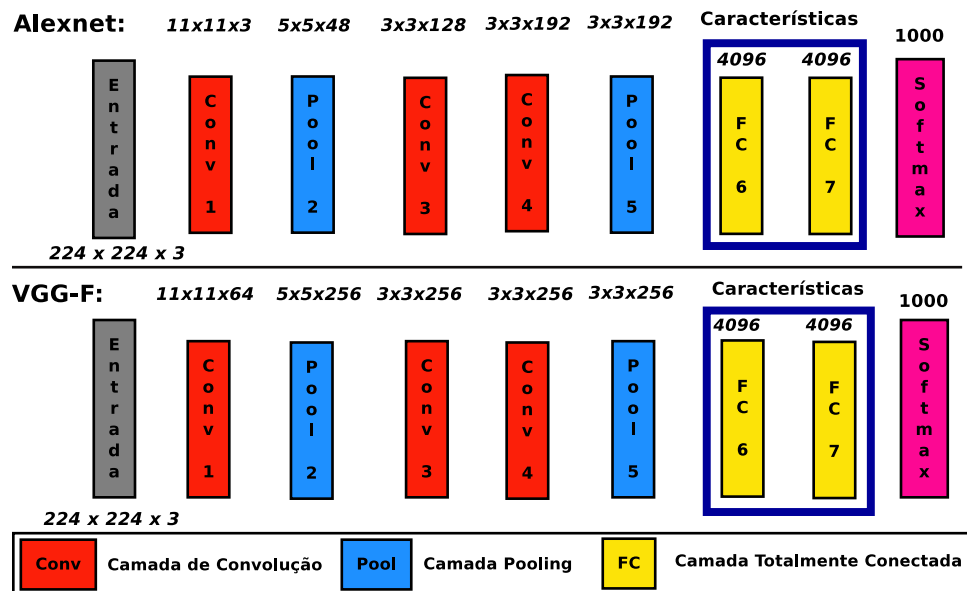


Figura 1. Arquiteturas de RNC.

A Alexnet foi desenvolvida por Krizhevsky et al. (2012) para reconhecimento de objetos diversos. A sua arquitetura é formada por oito camadas. As cinco primeiras camadas pertencem as operações de convolução e são localmente conectadas, já as três restantes são totalmente conectadas. A última camada totalmente conectada é a saída da rede, que é alimentada por uma operação de *Softmax*. Ela possui 1000 saídas, cada uma com rótulo de classe.

A primeira camada de convolução da Alexnet filtra a imagem de entrada com tamanho 224 x 224 x 3, utilizando uma máscara de tamanho 11 x 11 em 3 filtros por camada (11 x 11 x 3). Já a segunda camada recebe as matrizes de convolução e aplica máscaras com a função de *Max-pooling* de 5 x 5 x 48. Na terceira e na quarta camadas são realizadas outras operações de convolução, com máscaras de 3 x 3 x 128 e 3 x 3 x 192. Ao fim dessas convoluções, as matrizes de saída da camada anterior recebe a última operação de *Max-pooling* com uma máscara 3 x 3 x 192.

As camadas totalmente conectadas antes da *Softmax* armazenam as ativações a partir das última operação na camada *pooling*. A classificação dessa arquitetura é dada pelas melhores ativações que são utilizadas pela *Softmax*.

Já a arquitetura VGG-F (*Visual Geometry Group - Fast*) foi desenvolvida para o reconhecimento de faces por Simonyan et al. (2014). Ela é bem semelhante a Alexnet, contudo, a sua estrutura aplica mais filtros por camada.

Após o treinamento de RNCs, as camadas de convolução armazenam os pesos ajustados para responder a padrões relevantes a cada uma das classes impostas. Essas redes passam a ser modelos de predição, que podem ser utilizadas em qualquer situação que envolva classificação de imagens. No entanto, cada modelo em questão possui sua peculiaridade, de acordo com o que pretende-se identificar.

As últimas camadas totalmente conectadas, com exceção da *Softmax*, mantém 4096 características referentes as classes de treinamento (ver na Figura 1). Essas características são uma espécie de melhores descrições espaciais da imagem, que são extraídas nas operações de convolução.

3.2. Ampliação da base dados com *Data Augmentation*

As operação de *Data Augmentation* são formas utilizadas para reduzir o *overfitting* em redes neurais profundas. Essa técnica cria vários exemplos sintéticos de um mesmo conjunto de dados, permitindo um melhor aprendizado que é oferecido com o aumento artificial de informações [Simard et al. 2003].

Em imagens, a indicação do uso de *Data Augmentation* é sugerida para bases pequenas com classes semelhantes. Essa operação permite um maior aproveitamento de informações que possam ser utilizadas para discriminar dados próximos [Parkhi et al. 2015]. A Figura 2 apresenta um exemplo de *augmentation* para imagens dermatoscópicas.

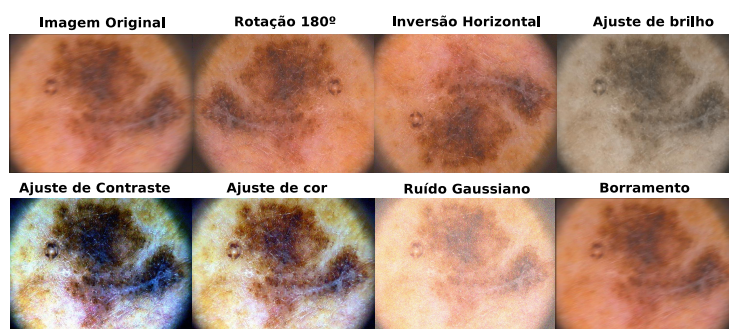


Figura 2. Exemplos de *augmentation* para imagens dermatoscópicas.

A argumentação em imagens é de extrema importância para o treinamento de Redes Neurais Convolucionais, já que as lesões melanômicas mantêm semelhanças significativas. Esse processo permite aproveitar o maior número de dados, através da aplicação de transformações no conjunto de treino. Além disso, esse conjunto de treino também é ajustado à arquitetura de convolução.

4. Experimentos e Resultados

A abordagem proposta nesse trabalho foi desenvolvida em MATLAB, utilizando a biblioteca MatConvNet [Vedaldi and Lenc 2015]. As redes Alexnet e VGG-F foram treinadas utilizando duas formas diferentes de treinamento nessa ferramenta, totalizando dois tipos de modelos para cada arquitetura. O primeiro tipo de treinamento não utiliza a operação de normalização em lotes (*Batch Normalization* - BN). Já no segundo, essa operação é realizada. A normalização em lotes ajusta o conjunto de treino de acordo com a média total de imagens, a média total de cores por canal e a covariância.

Os modelos de RNCs treinados nesse trabalho foram utilizados em partes, apenas as camadas totalmente conectadas (FC 6 e FC 7). Os dados extraídos dessas camadas serviram para a discriminação das lesões. Já a classificação desses dados foi realizada em duas classes (Melanoma e Não Melanoma) pela sugestão de classificadores implementados na ferramenta WEKA [Hall et al. 2009].

Os testes de classificação realizados nessa abordagem utilizaram o método estatístico *cross-validation 10-folds* para validação dos testes. Essa técnica particiona o conjunto de dados aleatoriamente em 10 subconjuntos de dados de mesmo tamanho, e um subconjunto é selecionado para a validação da classificação. Esse processo é realizado dez vezes e a estimativa produzida consiste na média de todos os resultados gerados.

4.1. Base de Imagens e Ampliação com *Data Argumentation*

Os experimentos realizados nesse trabalho utilizaram as bases PH² do Hospital Pedro Hispano (Matosinhos, Portugal) [Mendonça et al. 2013] que é composta por 200 imagens dermatoscópicas (160 de não melanoma e 40 melanoma) e a *International Skin Imaging Collaboration* (ISIC) [Gutman et al. 2016] que possui 900 imagens (173 melanoma e 887 não melanoma). Essas foram selecionadas por possuírem imagens diagnosticadas por dermatologistas.

Os testes realizados selecionaram 426 imagens (213 de melanoma e 213 de não melanoma) das bases. Os conjuntos utilizados foram definidos seguindo a proporção de 80 % para treino (340 imagens) e 20% para teste (86 imagens) que é sugerida por [Fleiss et al. 2013].

Os critérios utilizados para desenvolver as operações de *Data Augmentation* no conjunto de treino levou em consideração os estudos de [Chatfield et al. 2014] e [Parkhi et al. 2015], onde são aplicadas mudanças horizontais e verticais, rotações e translações, correções de cor, brilho e contraste, suavização, ruídos e recorte das regiões de interesse. No fim da execução dessa etapa foram produzidas 5.780 imagens (2.890 de melanoma e 2.890 de não melanoma) para treinamento das arquiteturas Alexnet e VGG-F.

4.2. Avaliação dos Resultados

A avaliação dos resultados é fundamental para definir o classificador que melhor representa um modelo de generalização do conhecimento. Em cada classificador é produzida uma matriz de confusão que exhibe as instâncias de dados diagnosticadas como corretas e incorretas. A partir disso, é possível determinar os cálculos que avaliam a sua eficácia.

As instâncias de dados podem ser classificadas em duas formas, denominadas de Verdadeiro Positivo (VP) e Falso Positivo (FP). Esses valores servem para determinar os parâmetros: acurácia (A), índice *kappa* (K), precisão (P), *recall* (R), *f-measure* (FM) e a área sobre a curva ROC (*Receiver Operating Characteristic*) [Codella et al. 2015].

A acurácia representa a porcentagem de casos classificados corretamente no conjunto de teste. Já o índice *kappa* mostra a concordância dos resultados obtidos e fornece uma ideia de quanto eles se afastam ou se aproximam do esperado, indicando a sua legitimidade. Os valores do índice *kappa* podem ser interpretados assim: < 0,4 (pobre) e > 0,75 (excelente). Quanto mais próximo de zero for o valor de *kappa*, maior a discordância dos dados [Fleiss et al. 2013].

A precisão indica as amostras classificadas corretamente como VP sobre o total de amostras (VP+FP), já o *recall* indica a proporção de amostras classificadas corretamente como positivas (VP) sobre todas as amostras efetivamente positivas. O *f-measure* é uma média ponderada de precisão e *recall*. Esse índice reflete a proporção de VP em relação às predições positivas (corretas e incorretas) somadas às amostras de classe positiva (efetivamente).

A área sobre a curva ROC exhibe o conflito entre a taxa de VP e FP de um classificador. Quanto maior a área sobre a curva ROC, mais preciso é o classificador. Essa métrica é utilizada para avaliar a capacidade discriminativa de um teste quanto a sua sensibilidade e especificidade [Fleiss et al. 2013].

4.3. Resultados de Classificação e Discussões

A partir dos métodos sugeridos, na Tabela 1 é apresentado os resultados obtidos para o classificador SVM, de acordo com as variações dos modelos de treinamento e seus respectivos dados extraídos das camadas FC 6 e FC 7, em todas as imagens do conjunto (426 imagens). Nela é possível observar uma acurácia de mais de 89% de instâncias classificadas corretamente, um índice *kappa* maior que 0,7, *recall* e precisão maior que 0,8. Esse resultado foi dado pela descrição da camada FC 7, com a operação de normalização, utilizando arquitetura VGG-F. Os resultados obtidos com a arquitetura Alexnet não foram satisfatórios.

Tabela 1. Resultados obtidos com o classificador SVM.

RNC	BN	FC	A	K	Taxa VP	Taxa FP	P	R	FM	Área ROC	Matriz de Confusão			
											Melanoma (a)		NoMelanoma (b)	
Alexnet	sim	6	71,3	0,427	0,714	0,286	0,716	0,714	0,713	0,653	163	50	72	141
		7	67,7	0,352	0,676	0,324	0,682	0,676	0,673	0,676	124	89	49	164
	não	6	81,9	0,638	0,819	0,181	0,821	0,819	0,819	0,819	183	30	47	166
		7	75,11	0,502	0,751	0,249	0,761	0,751	0,749	0,751	181	32	74	139
VGG-F	sim	6	82,6	0,652	0,826	0,174	0,841	0,826	0,824	0,826	198	15	59	154
		7	89,9	0,798	0,899	0,101	0,899	0,899	0,899	0,876	191	22	21	192
	não	6	70,18	0,403	0,702	0,298	0,712	0,702	0,698	0,702	173	40	87	126
		7	83,1	0,662	0,831	0,169	0,836	0,831	0,830	0,831	190	23	49	164

Na Tabela 2 é apresentado os resultados obtidos com o classificador KNN. Nessa fase de teste, a mesma camada citada anteriormente obteve uma acurácia de mais de 91% na identificação de instâncias corretas, um índice *kappa* maior que 0,8, *recall* e precisão maiores que 0,9. Nesse mesmo cenário, a Alexnet obteve seu primeiro resultado satisfatório com a camada FC 6, implementando a operação de normalização. Isso fez com que esse modelo atingisse uma acurácia de 87,5%, um índice *kappa* maior que 0,7 *recall* e precisão maior que 0,8.

Tabela 2. Resultados obtidos com o classificador KNN.

RNC	BN	FC	A	K	Taxa VP	Taxa FP	P	R	FM	Área ROC	Matriz de Confusão			
											Melanoma (a)		NoMelanoma (b)	
Alexnet	sim	6	87,5	0,751	0,876	0,124	0,876	0,876	0,876	0,891	188	25	28	185
		7	85,4	0,708	0,854	0,146	0,856	0,854	0,854	0,875	190	23	29	174
	não	6	74,41	0,488	0,744	0,256	0,757	0,744	0,741	0,750	135	78	31	182
		7	71,59	0,431	0,716	0,284	0,728	0,716	0,712	0,713	128	85	36	177
VGG-F	sim	6	86,8	0,737	0,869	0,131	0,869	0,869	0,869	0,860	183	30	26	187
		7	91,5	0,831	0,915	0,085	0,916	0,915	0,915	0,943	190	23	13	200
	não	6	74,41	0,744	0,256	0,758	0,758	0,744	0,741	0,739	134	79	30	183
		7	73,0	0,460	0,730	0,270	0,743	0,730	0,726	0,728	131	82	33	180

Por fim, na Tabela 3 é apresentado os resultados obtidos com o classificador MLP. Com este classificador a abordagem Alexnet obteve o melhor resultado sem a operação de normalização, utilizando a camada FC 6. Esse teste atingiu resultados de acurácia de 89,6% e índice *kappa* maior que 0,7. Já a VGG-F, com a normalização utilizando a camada FC 7, obteve uma acurácia de 85,3%, um índice *kappa* maior que 0,7, *recall* igual a 0,9 e precisão maior que 0,8.

Tabela 3. Resultados obtidos com o classificador MLP.

RNC	BN	FC	A	K	Taxa VP	Taxa FP	P	R	FM	Área ROC	Matriz de Confusão			
											Melanoma (a)		NoMelanoma (b)	
											(a)	(b)	(a)	(b)
Alexnet	sim	6	84,7	0,694	0,847	0,153	0,847	0,847	0,847	0,862	180	33	32	181
		7	85,6	0,713	0,857	0,143	0,859	0,857	0,857	0,932	174	39	22	191
	não	6	89,6	0,793	0,897	0,103	0,900	0,897	0,896	0,898	201	12	32	181
		7	80,0	0,600	0,800	0,200	0,801	0,800	0,800	0,901	143	40	45	168
VGG-F	sim	6	84,5	0,690	0,845	0,155	0,845	0,845	0,845	0,875	181	32	34	179
		7	85,3	0,713	0,857	0,143	0,857	0,857	0,857	0,881	187	26	35	178
	não	6	70,18	0,403	0,702	0,298	0,712	0,702	0,698	0,702	173	40	87	126
		7	73,0	0,460	0,730	0,270	0,743	0,730	0,726	0,728	131	82	33	180

Os valores de precisão e *recall* obtidos nos melhores testes de classificação foram maiores de 0,8, como visto nas Tabelas 1, 2 e 3, e indicam eficiência na definição de Verdadeiros Positivos no conjunto de amostras. No entanto, o classificador KNN obteve o melhor resultado entre todos, com os dados da camada FC 7 na arquitetura VGG-F, utilizando a normalização. Nesse caso, os valores de *recall* e precisão foram superiores a 0,9.

Ao analisar as matrizes de confusão para as três classificações de dados, notou-se que as instâncias que indicam a ausência do melanoma tem mais chances de serem classificadas corretamente em relação às instâncias que correspondem à presença do melanoma.

5. Conclusões e Trabalhos Futuros

Neste trabalho foi apresentada uma abordagem de classificação de imagens dermatoscópicas utilizando arquiteturas de RNCs treinadas com esse tipo de imagem. Em seguida, esses modelos foram utilizados para extrair características espaciais. Os dados das últimas camadas foram utilizados como descritores, e classificados com algoritmos tradicionais. Os resultados obtidos comprovam a eficácia da abordagem desenvolvida, onde a arquitetura VGG-F com a normalização, utilizando a camada FC 7, manteve os melhores dados do conjunto de imagens. Esses foram melhores discriminados pelo classificador KNN, com uma acurácia de 91,5%, índice *kappa* de 0,83, *recall* de 0,91 e precisão de 0,91.

Em trabalhos futuros, novos treinamentos em outras arquiteturas RNCs deverão ser realizados. Pretende-se utilizar um maior conjunto de imagens para verificar a eficácia da utilização de aprendizado profundo na classificação do melanoma.

Referências

- Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*.
- Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., and Smith, J. R. (2015). Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In *Machine Learning in Medical Imaging*, pages 118–126. Springer.
- Fleiss, J. L., Levin, B., and Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
- Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., and Halpern, A. (2016). 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). *arXiv preprint arXiv:1605.01397*.

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- INCA (2016). Instituto nacional do câncer - estimativas do câncer para o ano de 2016. [urlhttp://www.inca.gov.br/estimativa/2016/](http://www.inca.gov.br/estimativa/2016/). Acessado em 01-05-2016.
- Kawahara, J., BenTaieb, A., and Hamarneh, G. (2016). Deep features to classify skin lesions. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1397–1400. IEEE.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Maglogiannis, I. and Doukas, C. N. (2009). Overview of advanced computer vision systems for skin lesions characterization. *Information Technology in Biomedicine, IEEE Transactions on*, 13(5):721–733.
- Majtner, T., Yildirim-Yayilgan, S., and Hardeberg, J. Y. (2016). Combining deep learning and hand-crafted features for skin lesion classification. In *Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on*, pages 1–6. IEEE.
- Mendonça, T., Ferreira, P. M., Marques, J. S., Marcal, A. R., and Rozeira, J. (2013). Ph 2-a dermoscopic image database for research and benchmarking. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 5437–5440. IEEE.
- Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *BMVC*, volume 1, page 6.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.
- Shi, Z. and He, L. (2011). Current status and future potential of neural networks used for medical image processing. *Journal of multimedia*, 6(3):244–251.
- Simard, P. Y., Steinkraus, D., and Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2, ICDAR '03*, pages 958–, Washington, DC, USA. IEEE Computer Society.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Vedaldi, A. and Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM.