Skin Lesion Classification Using Hybrid Deep Neural Networks

Amirreza Mahbod[*1,2], Rupert Ecker[2], and Isabella Ellinger[1]

[1]Institute of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria
[2]TissueGnostics GmbH, Vienna, Austria

Abstract

Skin cancer is one of the major types of cancers and its incidence has been increasing over the past decades. Skin lesions can arise from various dermatologic disorders and can be classified to various types according to their texture, structure, color and other morphological features. The accuracy of diagnosis of skin lesions, specifically the discrimination of benign and malignant lesions, is paramount to ensure appropriate patient treatment.

Machine learning-based classification approaches are among popular automatic methods for skin lesion classification. While there are many existing methods, convolutional neural networks (CNN) have shown to be superior over other classical machine learning methods for object detection and classification tasks.

In this work, a fully automatic computerized method is proposed, which employs well established pre-trained convolutional neural networks and ensembles learning to classify skin lesions. We trained the networks using 2000 skin lesion images available from the ISIC 2017 challenge, which has three main categories and includes 374 melanoma, 254 seborrheic keratosis and 1372 benign nevi images. The trained classifier was then tested on 150 unlabeled images. The results, evaluated by the challenge organizer and based on the area under the receiver operating characteristic curve (AUC), were 84.8% and 93.6% for Melanoma and seborrheic keratosis binary classification problem, respectively.

The proposed method achieved competitive results to experienced dermatologist. Further improvement and optimization of the proposed method with a larger training dataset could lead to a more precise, reliable and robust method for skin lesion classification.

1 Introduction

Skin cancer is the most common cancer type in the United States of America and it is estimated that one in five Americans will develop skin cancer in their lifetime. Among different types of skin cancer, malignant melanoma itself (the deadliest type) is responsible for 10000 deaths annually in the united states [1]. However, in case of diagnosis in early stages, it is completely curable [2]. Early diagnosis of the cancer type is also important since
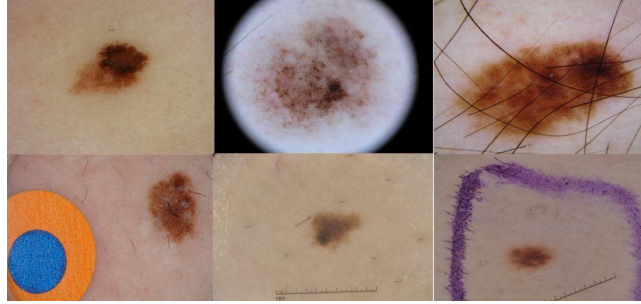
---

[*]amirreza.mahbod@tissuegnostics.com

Figure 1: Common artifacts in dermoscopic images. First row: a normal image, dark corner artifact and skin hair artifact from left to right, respectively. Second row: color chart artifact, ruler marker artifact and ink marker artifact from left to right, respectively. (Adapted from ISIC archive)

different treatments should be applied based on the type of cancer[3].

Dermoscopy has become the most important non-invasive tool for detection of melanoma and other pigmented skinlesions. In contrast to naked eye examinations, dermoscopy reveals subsurface structures of the skin. It makes use of handheld devices that extend optical light ray penetration beyond the skin surface and minimize surface reflection. [4]. This technique provides better differentiation between different lesion types based on their appearance and morphological features [5]. However, visual inspection of dermoscopic images by dermatologists relies on the experience of the physicians. Due to the difficulty and subjectivity of human interpretation, computerized analysis of dermoscopy images has become an important research area with the aim to reduce diagnostic errors [6]. While many computer-aided methods for lesion classification are based on lesion border detection and hand-crafted features, they lacked generalization capability due to high variations in dermoscopic images. These variations could be caused by different zooming, angles and lightening conditions [2],[7]. In addition, dermoscopic images suffer from various artifacts, which could have negative effect on classification accuracy. Fig. 1 shows some common artifacts. These problems could be overcome by using established pre-trained deep convolutional neural networks (CNN), which were trained on more that 1.4 million natural images (ImageNet dataset [8]) with various photographic conditions. These networks can be used as optimized feature extractors for dermoscopic images. While some research has been conducted to extract features using deep features for skin lesion classification [3],[9], they were limited by exploiting specific network architecture or using specific layer for extracting features. In this work, we hypothesize that using different architectures, extracting features from different layers and using ensemble learning on top of all, could lead to better classification with competitive precision to experienced dermatologist.

2   Method

The following subsections describe the algorithms and the dataset, which was used in the frame of this paper.

2.1   Dataset

The available images from ISIC 2017 challenge were used for training purposes. This dataset is composed of 2000 color dermoscopic skin images with corresponding labels. Three different skin lesion types are found in this dataset including 374 melanoma images, 254 seborrheic keratosis and 1372 benign nevi images with various sizes (from 1022×767 to 6748×4499), photographic angles and lightening conditions. Some images from this dataset include artefacts as shown in Fig. 1. Another set of 150 skin images was provided by the challenge organizers. The class labels of these test images were concealed from the participants for final evaluation of the performance of each proposed method by the challenge organizers.

2.2   Preprocessing

We applied three preprocessing steps on the images. First, we normalized the images by subtracting the mean RGB value of ImageNet dataset as suggested in [10] since the pre-trained networks were optimized for those images. Next, the images were resized to appropriate size using bicubic interpolation to be fed to the networks (227×227 and 224×224). Finally, we augmented the images to have a bigger training set. For data augmentation we rotated the images by 0, 90, 180 and 270 degree and then applied horizontal flipping to each image. From the derived artificial data set, we used 80% of the images randomly for training the classifiers.

2.3   Deep learning models and classification

The presented deep feature extractor used two pre-trained networks. We used pre-trained AlexNet architecture [10] and VGG-16 architecture [11] as feature extractor. We explore extracting features from different layers of pre-trained models to see how they could affect classification results. The features were mainly extracted from the last fully connected (FC) layers. We used first and second fully connected layers (referred as FC6 and FC7 with 4096 dimensions) and also the concept detector layer (referred as FC8 with 1000 dimension). These features along with corresponding labels (skin type) were then used to train multi-class non-linear support vector machine (SVM) classifiers. We trained different classifiers for each network and then averaged the class score to have the final classification result.

2.4   Evaluation

To evaluate the classification results, we mapped SVM scores to probabilities using logistic regression as described in [12]. Since the classifiers were trained for a multi classification problem (3 classes including melanoma, seborrheic keratosis and benign nevi classes), we combined the scores to have the results for two binary classification problems, which were melanoma classification and seborrheic keratosis classification. Classifier performance, in terms of accuracy (ACC) and area under the receiver operating characteristics curve (AUC) were measured for 150 test images.

3   Results

The summarized classification results on the test images are shown in Table 1. To calculate the accuracy, a probability of 50% was used as threshold. All experiments were performed on a laptop computer with NVIDIA GTX 840M 2GB , 6 GB installed memory and Intel Core i7-4500U 1.8 GHz CPU. The main implementation of the algorithm was done with MATLAB (version 2016b) and MatConvNet framework for CNN. The run-time of the implemented algorithm for training SVM and extracting features was around 70 minutes and for testing on all test images it was around 5 minutes.

Table 1: Summarized results. In the table, M stands for Melanoma and SK stands for seborrheic keratosis.

Network	Feature Layer	M-ACC	SK-ACC	M-AUC	SK-AUC	Avg AUC
AlexNet	FC8	83.3	88.0	78.4	90.3	84.3
ALexNet	All FC	84	89.3	82.8	94.9	88.9
VGG-16	FC8	82.7	79.3	78.8	81	79.9
VGG-16	All FC	85.3	88.0	81.9	92.8	87.3
Fusion	All FC	84.7	84.7	84.8	93.6	89.2

4   Discussion & Conclusion

In this work, we demonstrated how using pre-trained deep architectures, which were originally trained for natural images, can be used for dermoscopic image classification with competitive accuracy to experienced dermotologists. As the results of Table 1 suggested, fusing networks results and extracting features from different layers of the network could increase both accuracy and AUC. Extracting features from other pre-trained networks and from other layers is an open field for further research. In addition, using the

same approach on a larger training data set could potentially lead to a better predictive model.

References

[1] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the US population, 2012," *JAMA dermatology*, vol. 151, no. 10, pp. 1081–1086, 2015.

[2] G. Schaefer, B. Krawczyk, M. E. Celebi, and H. Iyatomi, "An ensemble classification approach for melanoma diagnosis," *Memetic Computing*, vol. 6, no. 4, pp. 233–240, 2014.

[3] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on.* IEEE, 2016, pp. 1397–1400.

[4] G. Argenziano, H. P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, and M. Delfino, "Dermoscopy: a tutorial," *EDRA, Medical Publishing & New Media*, p. 35, 2002.

[5] A. Steiner, M. Binder, M. Schemper, K. Wolff, and H. Pehamberger, "Statistical evaluation of epiluminescence microscopy criteria for melanocytic pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 29, no. 4, pp. 581–588, 1993.

[6] M. G. Fleming, C. Steger, J. Zhang, J. Gao, A. B. Cognetta, and C. R. Dyer, "Techniques for a structural analysis of dermatoscopic imagery," *Computerized medical imaging and graphics*, vol. 22, no. 5, pp. 375–389, 1998.

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE, 2009, pp. 248–255.

[9] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *International Workshop on Machine Learning in Medical Imaging.* Springer, 2015, pp. 118–126.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[12] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.