Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network

E. Nasr-Esfahani, S. Samavi, N. Karimi, S.M.R. Soroushmehr, M.H. Jafari, K. Ward, K. Najarian

***Abstract*— Melanoma, most threatening type of skin cancer, is on the rise. In this paper an implementation of a deep-learning system on a computer server, equipped with graphic processing unit (GPU), is proposed for detection of melanoma lesions. Clinical (non-dermoscopic) images are used in the proposed system, which could assist a dermatologist in early diagnosis of this type of skin cancer. In the proposed system, input clinical images, which could contain illumination and noise effects, are preprocessed in order to reduce such artifacts. Afterward, the enhanced images are fed to a pre-trained convolutional neural network (CNN) which is a member of deep learning models. The CNN classifier, which is trained by large number of training samples, distinguishes between melanoma and benign cases. Experimental results show that the proposed method is superior in terms of diagnostic accuracy in comparison with the state-of-the-art methods.**

I. Introduction

Melanoma, also referred to as malignant melanoma, is a type of skin cancer caused by abnormal multiplication of pigment producing cells that give color to the skin [1]. Despite significant death rate, early stage detected melanoma is curable in most cases [2]. Meanwhile differentiation between melanoma and other benign moles in their initial growth phases is a challenging task even for experienced dermatologists [3]. Computerized algorithms are being developed for this purpose. Some low complexity methods are designed, which are intended for running on tablets and smartphones, and can help non-specialists. But professional decision making, in this regard, requires sophisticated algorithms and equipment. There are various methods in dermatology such as ABCD (asymmetry, border irregularity, color patterns, and diameter) rule [4] and the seven-point checklist [5] that guide physicians in this task. Figure 1 shows the general structure of our proposed, highly complex, deep learning method.

Ebrahim Nasr-Esfahani is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran.

Shadrokh Samavi is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran. He is also with the Department of Emergency Medicine, University of Michigan, Ann Arbor, U.S.A.

Nader Karimi is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran.

S.M. Reza Soroushmehr and Kevin Ward are with the Emergency Medicine Department and Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, U.S.A.

Mohammad H. Jafari is with the Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran.

Kayvan Najarian is with the Department of Computational Medicine and Bioinformatics, Department of Emergency Medicine and the Michigan Center for Integrative Research in Critical Care, University of Michigan, Ann Arbor, U.S.A.

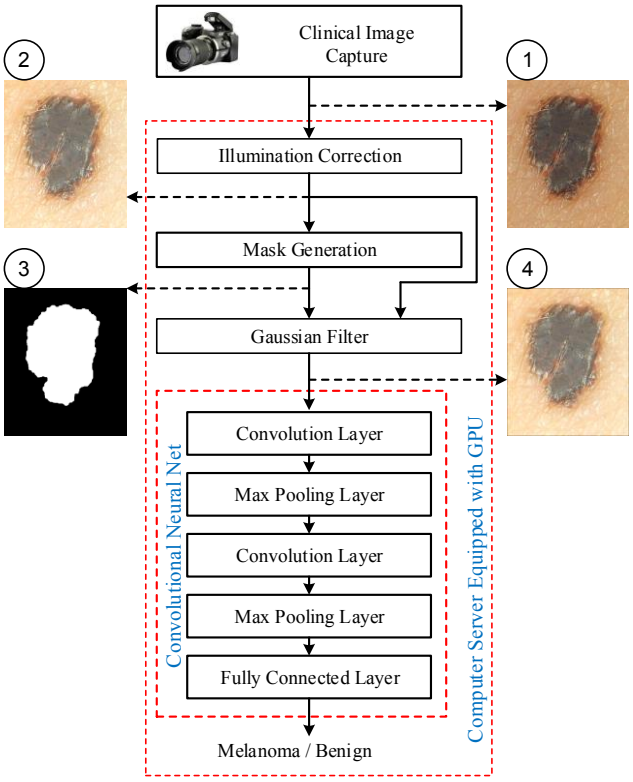

Figure 1. Block diagram of the proposed melanoma detection method for implementation on a GPU equipped server.

Automated analysis of pigmented lesions is a growing research topic that aims to develop tools for computer aided diagnosis of skin cancer [6]. Reviews of researches done in this field are given in [6] and [7].

Recently there has been an emerging trend for automatic diagnosis of melanoma using conventional digital cameras [8-14]. This can be applicable in web based and mobile application as a telemedicine tool and also as a supporting system that assists physicians. Computerized diagnosis is a rising need due to increasing rate of incident, subjectivity of procedure and time and cost expenses [11].

A decision support system was proposed by Alcon *et al.* that takes advantage of image processing techniques for evaluation of skin lesions [8]. It also uses some background information of patient before the diagnosis is made. Cavalcanti *et al.* proposed an automatic classification system that consists of preprocessing, segmentation and feature extraction steps. In order to reduce miss rate of melanoma cases (false negative cases), a two stage classifier has been proposed that rechecks the lesions labeled as benign [10]. Works of [9] and [10] try to estimate the four criteria of the

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

ABCD rule by extracting a set of low level features. Amelard *et al.* proposed a set of high level features that intuitively describes ABCD characteristics [11]. Giotis *et al.* proposed a computer aided system that performs some preprocessing for enhancing the quality of the image [12]. Then, some color and texture features of lesions are extracted automatically. Afterward, a set of attributes are provided by the examining physician and the final diagnosis is made by voting among observed and automatically extracted descriptors. Examples of commercial applications that exist in this field can be found in [13] and [14].

Deep learning approaches have shown promising results in some applications such as natural language processing [15], speech recognition and recently in computer vision areas such as object tracking [16], object detection [17], and image classification [18]. Deep learning mechanisms can learn set of high level features from low level ones and gain high accuracy for classification applications without the need for extracting handcrafted features. Especially there has been a trend for taking advantage of superior power of deep learning methods in medical imaging tasks [19-21]. Convolutional neural network (CNN) is a type of deep learning method where trainable filters and pooling operations are applied on the raw input images, extracting set of complex high level features automatically [22].

In this paper we aim to take advantage of deep learning methods to form an automatic diagnosis system for melanoma detection. For this purpose the input digital images, usually subjected to noise and illumination effects, are preprocessed. This preprocessing effectively helps CNN extract discriminative features from the images. Afterward, the images are fed to a CNN architecture to classify the input as melanoma or benign. In order to deal with the limitations of the training dataset and relatively low number of images, some approaches such as rotation, cropping, and resizing of dataset images are considered that increase the number of samples. The experimental results show that the proposed system can outperform comparable state-of-the-art methods.

The rest of this paper is organized as follow. In Section II, the proposed method is explained in details. Experimental results of quantitative evaluation of our method are presented in Section III. Section IV concludes the paper.

II. Proposed Method

In this section the proposed method is explained in details. The proposed system, as shown in the block diagram of Fig. 1, consists of a preprocessing step. The goal of this step is to reduce artifacts that could mislead the convolutional neural network. Preprocessed images are fed into second stage which is a CNN. Details of these two stages are explained in the following.

A. Preprocessing

Images of skin surface, even those taken by professional digital cameras, usually contain illumination and noise effects that should be eliminated. These effects are the result of non-uniform lighting, and reflections of incident light from skin surface. To reduce effects of these misleading factors on CNN's training and classification, firstly an illumination correction step is performed on input images. This step is similar to the algorithm of [12], where the illumination effects are detected as sharp changes in the saturation and value channels of the HSV color space. Thus, the illumination effects are discarded by excluding a specific range of gradients. This is done without destroying the real edges of the original image.

Another factor that should be considered is that an input image contains both healthy (normal skin) and lesion parts. This can mislead the training of CNN. Texture of the healthy areas is an irrelevant criterion for melanoma detection. Meanwhile, cropping healthy areas could cause loss of information, such as color difference between lesion and patient's normal skin. Such information could be a discriminative clue. For this purpose, a segmentation mask is produced by applying a k-means classifier ($k = 2$) on the preprocessed image. This mask is further enhanced by some morphological operations. The attained mask extracts the lesion's region.

For reducing the effects of the normal skin's texture on the classification process, we use the segmentation mask for smoothing the area outside of the lesion. For this aim, a Gaussian filter (with $\sigma = 2$) is applied on the normal parts of the skin based on the information of the segmentation mask.

B. CNN Architecture

In this section, details of the proposed CNN are discussed. Extracting an effective and discriminative feature set, which precisely differentiate between various classification groups, is a challenging task. In the one hand, there is a pitfall that by using a large set of features we may feed some incoherent traits to the network. On the other hand, by using a small set of features there is a possibility of missing some proper descriptors. Hence, automatic feature extraction systems could be utilized to achieve a discriminative feature set based on labeled training set, without the need for definition of handcrafted feature extraction procedures.

In this paper CNN, as a deep learning framework, is used for automatic detection of melanoma. CNNs take advantage of a set of powerful convolve-filters. They can examine various structures in input images. Hence, in utilization of CNN, the input is the image itself and the network automatically extracts appropriate aspects of the image.

Conventional CNNs usually contain several convolve and pooling layers and the last layer is made by a fully connected layer. Convolve layers filter the input image by a set of kernels. Usually each convolve layer is followed by a pooling layer. By selecting the maximum or mean values, in each defined window, the pooling layer reduces the size of the feature map. It is done for the purpose of recognizing some general patterns in the images. These general patterns are perceptible in resized image.

The used CNNs in this paper consist of two convolving layers with a 5×5 kernel. There are 20 feature maps in the first convolution layer and 50 feature maps in the second convolution layer. There is one pooling layer after each convolution layer. Outputs of these four layers are fed to a 2-layer fully connected stage respectively having 100 and 2 neurons. This 2-layer network forms the final diagnosis results with a linear transfer function. The proposed configuration can be seen in Fig. 2.

The images of the dataset, after removal of noise and illumination artifacts, are fed to the proposed CNN.

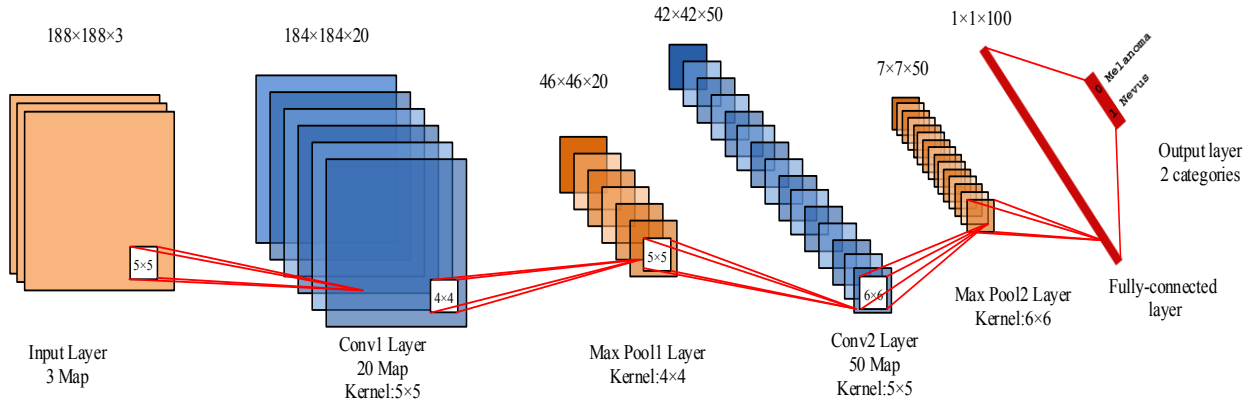

Figure 2. Artichecture of the proposed CNN.

Usually there should be a large number of samples for proper training of any CNN. However, due to difficulties in collection and labeling of images the existinfigg datasets usually have limited number of images for detection of melanoma from non-dermoscopic images. Hence, we should cope with the limited training set. For this aim there are some approaches for automatic expansion of the training data. For example an image is cropped from its top-left corner by cropping 5 percent from the top and from the left side of the image. Then, the new image is scaled back to the original size. This cropping method could be also done from three other corners. The same thing could be done with 10 percent cropping. This results in 8 images besides the original image. Afterward, these 9 images are rotated by 0, 90, 180 and 270 degrees. Therefore, 36 synthetized images are derived from each input image and the size of training set will be multiplied by 36. It should be noticed that all input images are resized to 188×188 pixels. The same approach is applied for the testing data, where 36 versions of each image are produced by cropping and rotation. The final diagnosis for any input image is made by voting among classification results of its 36 synthesized versions.

III. Experimental Results

In this section the proposed method is evaluated on a publically available dataset of skin lesion images [23]. This dataset consists of 170 non-dermoscopic images (70 melanoma, 100 nevus) from the digital image archive of the Department of Dermatology of the University Medical Center Groningen (UMCG). The proposed system is implemented on a server with Intel Core i7-4790K processor, 32 GB of RAM, and two NVIDIA GeForce GTX Titan X GPU cards with scalable link interface (SLI).

The training set of CNNs must be sufficient enough. This dataset of 170 images is increased to 6120 original and synthesized images. To perform training and testing the dataset is split into two randomly selected groups. A 80% - 20% ratio is used where 80% of the dataset images are randomly selected for training and the rest is used for test while there is no overlap between the test and train samples. The training data is fed to a network with a batch size of 64. The network is trained through 20,000 iterations. To make the results independent from the selected training and test data, the procedure of learn and test is repeated 50 times and the mean values of different groups of results are reported.

For quantitative evaluation of the performance of the proposed system, five commonly used metrics in classification problems are measured. These metrics are defined as follow:

$$sensitivity = \frac{true\ detected\ melanoma\ cases}{all\ melanoma\ cases},$$

$$specifity = \frac{true\ detected\ non\_melanoma\ cases}{all\ non\_melanoma\ cases},$$

$$PPV = \frac{true\ detected\ melanoma\ cases}{detected\ melanoma\ cases},$$

$$NPV = \frac{true\ detected\ non\_melanoma\ cases}{detected\ non\_melanoma\ cases},$$

$$accuracy = \frac{true\ detected\ cases}{all\ cases},$$

For comparing the proposed method with other existing methods, three works that have reported their results on the same dataset are studied [12], [13] and [24]. Work of [24] is one of the early efforts in this field used as a baseline. Work of [13] is an example of existing commercial applications and method of [12] makes its final diagnosis in a semi supervised framework, where the opinion of a physician who has examined the lesion is involved. For the fairness of comparison, only the automatically extracted descriptors of the work of [12] are reported here. The evaluation results are shown in Table 1. As can be seen, the proposed method has the highest accuracy with respect to other state-of-the-art methods. These results show the effective power of utilizing a deep learning framework. However, some false classified images are shown in Fig. 3. In the left column are benign moles labeled as melanoma by our classifier and the right column are samples of missed melanoma cases.

IV. Conclusion

Non-dermoscopic images taken by digital cameras have been serving as a tool for melanoma detection in

telemedicine. In this paper a computational complex method based on deep learning was implemented that used clinical images. This system was capable of detecting melanoma cases from benign ones. We were able to increase the accuracy of the system by sending images through illumination correction that increased the discrimination capability of the system. For training, we used an available small dataset. By cropping, scaling, and rotating of images the number of images was increased. Our proposed method left the process of feature extraction to CNN while traditional learning approaches try to extract features from data. Experimental results showed our better accuracy, as compared to other detection algorithms.

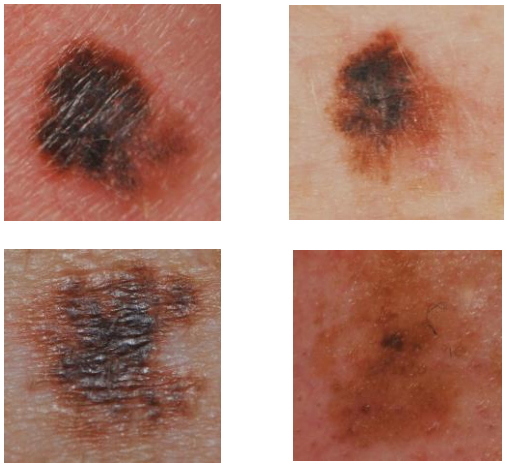

Figure 3. Samples of wrong diagnosed images, left column false positives and right column false negative cases.

TABLE I. QUANTITATIVE COMPARISON OF DIAGNOSTIC RESULTS, BEST RESULTS ARE BOLD.

Methods	Metrics				
	Sensitivity	*Specificity*	*PPV*	*NPV*	*Accuracy*
[24]	0.46	**0.87**	0.70	0.71	0.70
Spotmole [13]	**0.82**	0.57	0.56	0.83	0.67
MED-NODE texture descriptor [12]	0.62	0.85	0.74	0.77	0.76
MED-NODE color descriptor [12]	0.74	0.72	0.64	0.81	0.73
Proposed	0.81	0.80	**0.75**	**0.86**	**0.81**

REFERENCES

[1] Skincancer.org, "Melanoma - SkinCancer.org," 2016. [Online]. Available: http://www.skincancer.org/skin-cancer-information/.

[2] American Cancer Society, "Cancer Facts & Figures 2016," American Cancer Society, Atlanta, GA, USA, 2016.

[3] A. F. Jerant, J. T. Johnson, C. Sheridan and T. J. Caffrey, "Early detection and treatment of skin cancer," *American family physician*, vol. 62, no. 2, pp. 357-386, 2000.

[4] F. Nachbar, W. Stolz, T. Merkle, A. B Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. B.-Falco, and G. Plewig, "The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[5] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comp. of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563-1570, 1998.

[6] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial intelligence in medicine*, vol. 56, no. 2, pp. 69-90, 2012.

[7] I. Maglogiannis and C. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721-733, 2009.

[8] J. F. Alcon, C. Ciuhu, W. Ten Kate, A. Heinrich, N. Uzunbajakava, G. Krekels, D. Siem and G. De Haan, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE Journal of Selected Topics in Signal Processing*, vol.3, no. 1, pp.14-25, 2009.

[9] P. G. Cavalcanti and J. Scharcanski, "Automated prescreening of pigmented skin lesions using standard cameras," *Computerized Medical Imaging and Graphics*, *Elsevier*, vol. 35, no. 6, pp. 481-491, 2011.

[10] P. G. Cavalcant, J. Scharcanski and G.V. Baranoski, "A two-stage approach for discriminating melanocytic skin lesions using standard cameras," *Expert Systems with Applications*, Elsevier, vol. 40, no. 10, pp. 4054-4064, 2013.

[11] R. Amelard, J. Glaister, A. Wong and D. Clausi, "High-level intuitive features (hlifs) for intuitive skin lesion description," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 820-831, 2015.

[12] I. Giotis, N. Molders, S. Land, M. Biehl, M. F. Jonkman and N. Petkov, "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images," *Expert Systems with Applications*, Elsevier, vol. 42, no. 19, pp. 6578-6585, 2015.

[13] C. Munteanu and S. Cooclea, "Spotmole – melanoma control system," 2009. Available: http://www.spotmole.com/

[14] Health Discovery Corporation, "Worlds first svm-based image analysis iphone app for melanoma risk assessment, melapp," launched by health discovery corporation, 2011. Available: http://www.healthdiscoverycorp.com/pr/july06_11.

[15] J. Hirschberg and C. D. Manning, "Advances in natural language processing," *Science,* vol. 349, no. 6245, pp. 261-266, 2015.

[16] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems*, 2013, pp. 809-817.

[17] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2147-2154.

[18] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012*, 2012, pp. 3642-3649.

[19] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated mass detection in mammograms using cascaded deep learning and random forests," in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, pp. 1-8, 2015.

[20] L. Siqi, L. Sidong, C. Weidong, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *IEEE International Symposium on Biomedical Imaging (ISBI), 2014*, pp. 1015-1018, 2014.

[21] Suk, Heung-Il, and Dinggang Shen, "Deep learning-based feature representation for AD/MCI classification," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer Berlin Heidelberg, 2013. 583-590.

[22] J. Shuiwang, X. Wei, Y. Ming, and Y. Kai, "3D Convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, no. 1, pp. 221-231, 2013.

[23] Available: http://www.cs.rug.nl/~imaging/databases/melanoma_naevi/.

[24] E. Zagrouba and W. Barhoumi, "A preliminary approach for the automated recognition of malignant melanoma, " *Image Analysis & Stereology,* vol. 23, no. 2, pp. 121-135, 2004.