Multiple Convolutional Neural Network for Skin Dermoscopic Image Classification

Yanhui Guo
Department of Computer Science
University of Illinois Springfield, Springfield，Illinois, USA
Email: yguo56@uis.edu

Amira S. Ashour
Department of Electronics and Electrical Communications Faculty of Engineering, Tanta University
Tanta, Egypt
Email: amirasashour@yahoo.com

Lei Si
Department of Computer Science
University of Illinois Springfield, Springfield，Illinois, USA
Email: lsi3@uis.edu

Deep P Mandalaywala
Department of Computer Science
University of Illinois Springfield, Springfield，Illinois, USA
Email: dmand3@uis.edu

Abstract— **Melanoma classification is a serious stage to identify the skin disease. It is considered a challenging process due to the intra-class discrepancy of melanomas, skin lesions' low contrast, and the artifacts in the dermoscopy images, including noise, existence of hair, air bubbles, and the similarity between melanoma and non-melanoma cases. To solve these problems, we propose a novel multiple convolution neural network model (MCNN) to classify different disease types in dermoscopic images, where several models were trained separately using an additive sample learning strategy. The MCNN model is trained and tested using the training and validation sets from the International Skin Imaging Collaboration (ISIC 2016), respectively. The classification accuracy and receiver operating characteristic (ROC) curve are used to evaluate the performance of the proposed method. The values of AUC (the area under the ROC curve) were used to evaluate the performance of the MCNN.**

Keywords— *Deep learning, conventional neural network, multiple model, skin cancer, dermoscopic images.*

I. INTRODUCTION

Skin abnormalities are the first indicator of different diseases including the skin cancer, which is one of the main health problems worldwide. There are several skin lesion types, including melanoma, basal cell carcinoma, melanocytic nevus, bermatofibroma, benign keratosis, actinic keratosis, and vascular lesion [1]. Early detection, identification and classification of the lesion's type play a significant role in the diagnosis and treatment strategy. Traditionally, the physicians depend on their visual experience to evaluate the patients' lesions case-by-case based on the local lesion patterns compared to the skin in the entire body [2]. However, the automated analysis and classification of the dermoscopic images provide accurate and timeless assessment of the skin lesion type according to the skin surface structures of the lesion regions [3].

Several classification techniques have been applied for skin lesion dermoscopic images classification, such as the artificial neural network (ANN) [4], support vector machine (SVM) [5, 6], bag-of-features model [7], decision tree [8], ensemble classification [9], and the fuzzy logic based classifiers [10]. Compared to statistical methods, the ANN provides reduced error rates. Thus, the ANN is designated and imitated in the computer aided diagnosis (CAD) systems to map a set of features/symptoms to a set of the probable diagnostic classes [11]. However, extra knowledge of the dermoscopic dataset cannot be attained from the neural network, where the parameters of the neural network model are not directly deduced. Additionally, the classification process of the dermoscopy images is still puzzling due to the presence of artifacts, noises, the complexity and variability of the skin lesion structures, illumination variation during the image capturing process, dense hairs, air bubble, and low contrast between normal skin and lesion, which perplex the lesion's border detection, feature extraction and classification processes. Along with the smooth transition in color between the background and the lesion regions that specifies the low contrast between the background and the pigmented lesion.

Consequently, researchers have devoted several classification techniques using deep learning (DL) in different applications. The DL algorithms have the machine learning architecture, which driven with the competency to handle large datasets of complex computations. For image classification, the Convolutional Neural Networks (CNNs) are considered the most prevalent architecture in several applications, especially in the medical image classification processes. The CNN has a significant conceptual framework, including weights sharing, local perception of the area, and time domain/ down-sampling space, which

guarantee a relative unchanged displacement, distortion and scaling characteristics.

For skin lesions classification, few researchers have applied the CNNs to exploit their dominant discrimination ability. There are several CNN designs, such as Inception (GoogleNet), AlexNet, QuocNet, BNInception-V2, and Inception-V3. For dermoscopy images classification, Codella *et al*. [12] employed pre-trained CNNs to extract deep features for further classification of the melanoma and atypical lesions as well as the classification of melanoma and non-melanoma. Kawahara *et al*. [13] applied an AlexNet based fully pre-trained CNN on dermoscopic skin images instead of using CNN from scratch. This CNN was pre-trained on natural image to transfer learning from other domains for reducing the time consuming in the training process with 81.8% classification accuracy. Esteva *et al*. [14] classified 129,450 skin lesion dataset using pre-trained CNN GoogleNet inception-V3 structure. Another study on skin lesion classification implemented a deep residual network using more than 50 layers [15]. This residual network comprised of stacked layers consisted of convolutional layer, rectified linear unit and batch normalization layer. In order to initialize the weights and the de-convolutional layers, a pre-trained network on ImageNet dataset model was used. The authors increased the dataset by rotating and cropping, major axis alignment and augmentation.

A public challenge on Skin Lesion Analysis Toward Melanoma Detection has been hosted by the International Skin Imaging Collaboration (ISIC) in 2016 as reported in [16] to classify the dermoscopic images into only two classes. Matsunaga *et al*. [17] participated in this challenge (part 3) for skin lesion classification by using an ensemble of CNNs. AdaGrad and RMSProp techniques have been used to fine tune the CNNs. The results depicted that Area Under Curve (AUC) of value 0.953 has been achieved to classify the seborrheic keratosis (SK). However, González-Díaz [18] achieved 0.965 AUC value using the CNN, where several networks have been designed to provide the lesion area identification, lesion segmentation and final diagnosis. In the same challenge, Menegola *et al*. [19] used CNN by employing the ResNet-101 and Inception-v4, which were pre-trained on ImageNet, where 0.943 AUC value has been achieved. These preceding studies established that using pre-trained deep learning approaches to classify dermoscopic images is a promising domain.

This paper applied the proposed model training and testing phases in the classification task of the ISIC2016 dataset, there are 900 and 379 images in training and testing sets, respectively. In the ISIC 2016 dataset, there are two categories including benign and malign. Considering the more challenges in the ISIC2016 classification task, multiple models are necessary to be employed to enhance the performance of individual model methods. At the same time, the multiple model based methods can reduce the training time with fewer epochs on each model.

This work proposed a novel multiple convolution neural network model (MCNN) to classify the dermoscopic images. Transfer learning is used to train multiple models on different samples. The samples are selected according to their scores' performance at the previous models.

The structure of the remaining sections is as follows. In section II, the methodology of the proposed model is introduced. Afterwards, the experimental results are interpreted and finally the conclusions are presented in sections III and IV, respectively.

II. METHODOLOGY & PROPOSED METHOD

In the proposed multiple convolutional neural network (MCNN) model, multiple networks are trained using an additive training strategy to adjust the learning samples for different CNN models. The idea of this strategy is to improve the model's performance by repeated training of the model on wrongly classified samples, which is consistent to the human's recognition procession. In the prediction stage, the classification results of the testing samples are determined by the model that has the highest prediction score.

A multiple DCNNs framework is constructed by cascading multiple networks with same structures as:

$$MDCNN = \left\{ CNN_1, CNN_2, ..., CNN_N \right\} \quad (1)$$

where N is the total number of DCNN in the framework.

To speed up the training stage, a transfer learning is employed, and multiple models are initialized based on a pre-trained CNN model, such as AlexNet, VGG16, GoogleNet, and ResNet. The MDCNN is trained using an incremental learning strategy to improve the networks' performance. The next DCNN is trained using the same samples as previous one and enhanced to learn the samples which are not performed well in the previous DCNN. In this way, the next DCNN will overcome the poor performance of the previous one.

$$Sample(k+1) = \left\{ Sample(k), IncreSample(k) \right\} \quad (2)$$

where $Sample(k)$ and $Sample(k+1)$ are the sample for the kth and $(k+1)^{th}$ DCNN, respectively. Moreover,

IncreSample(k) are the samples, which are wrongly classified by the kth DCNN.

This training procedure is applied iteratively until all models are trained using the added samples with poor performance from the previous models.

In testing stage, the input image is classified by a serial of MCNN models. The final classification decision is determined according to the model with the highest prediction score, which can be expressed as follows:

$$L(i) = \arg\max\left\{S_m^n(i)\right\} \tag{3}$$

where $S_m^n(i)$ is the prediction score to category n of the mth model. The final classification result of each sample is obtained from the model with the highest performance score on it.

Accordingly, the proposed MCNN approach can be summarized as follows.

Algorithm: Proposed multiple convolutional neural network

> **Step 1: Train** the first model using the training set for limited iterations
>
> **Step 2: Identify** some wrongly classified samples from the previous training and duplicate samples in the training set of the trained current model
>
> **Step 3: Determine** the samples whose results are less than the threshold (using trial-and-error) and duplicate these samples in the training set
>
> **Step 4: Train** the next model using the additive sample set
>
> **Step 5: Repeat** the previous steps from step 2 until no more wrongly classified and additive samples
>
> **Step 6: Predict** the samples and determine their classification result according to their performance score of different models

---

III. EXPERIMENTAL RESULTS

A. Dataset

The available part of the ISIC2016 dataset [20] including 900 images for training has been used for training, and 379 are used to evaluate the proposed MCNN model.

B. Experiment on classification

In the training stage, five ResNet50 networks are employed for transferring learning and trained for 5 epochs using the reinforcement sample learning strategy.

The classification accuracy and receiver operating characteristic (ROC) curve is used to evaluate the performance of the classifier, which is a useful tool for organizing the classifiers and visualizing their performance. Compared with scalar measures, such as the accuracy, error rate, and the error cost; the ROC graphs can provide richer measure by calculating the AUC (the area under the ROC curve) [21].

Figs. 1 and 2 demonstrate the classification confusion matrix on training and testing sets, respectively. Figs. 3 and 4 show the ROC curve of classification results on malign and benign cases in training set, respectively.
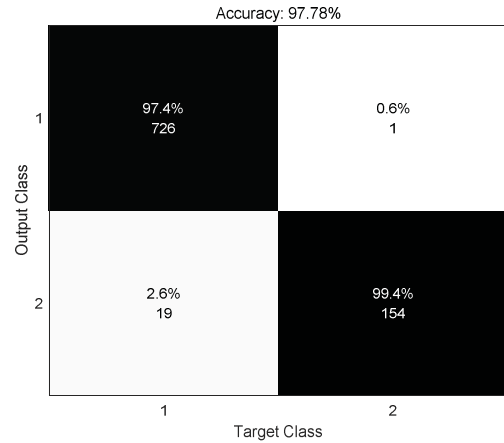

Fig. 1. Confusion matrix of classification results in training set
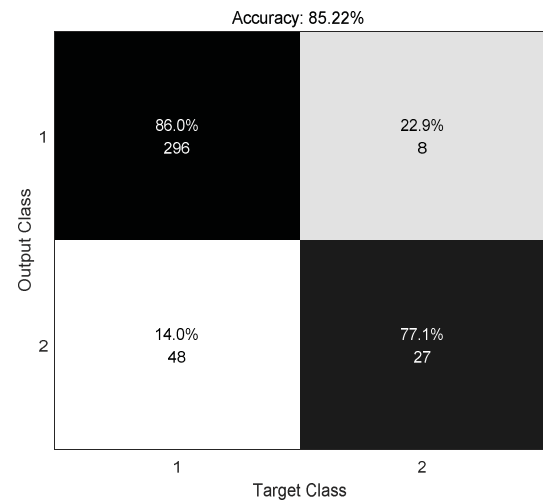

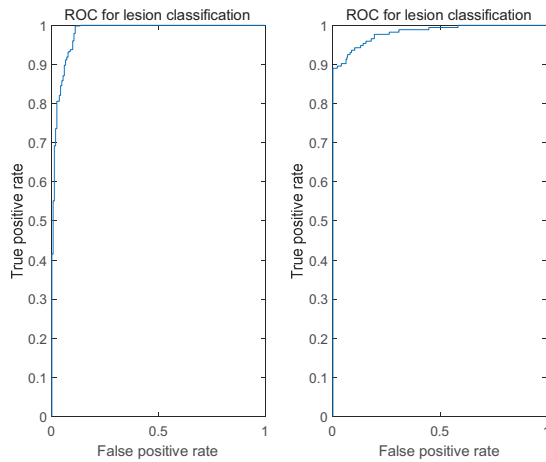Fig. 2. Confusion matrix of classification results in testing set

Fig. 3. ROC curve of classification results on malign and benign cases in training set
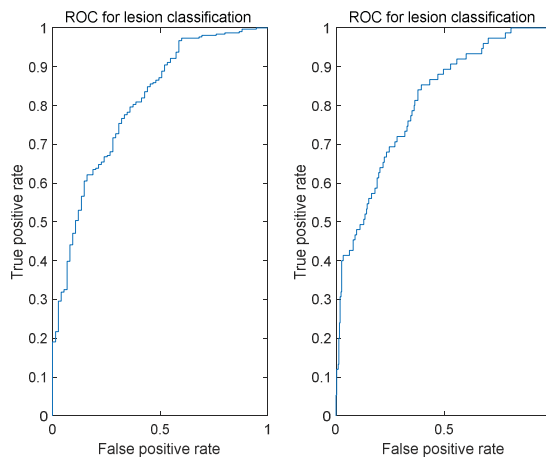

Fig. 4. ROC curve of classification results on malign and benign cases in testing set

Figs 1 and 2 illustrate the accuracy and AUC for the training set are 97.78% and 0.98, and 85.22% and 0.81 for the testing set, respectively.

IV. CONCLUSIONS

Deep learning procedures proved their efficiency in the different medical image processing applications. These DL networks enabled the progress of the medical diagnostic imaging systems. Since the skin lesion classification problem is a challenging task owing to the special characteristics of the skin lesions and their variety of types, predominantly the detection problem of early melanoma, DL based approach was proposed in the current work to solve the classify dermoscopic image enclosing two skin lesion types.

In this work, a multiple convolutional neural network is proposed to classify malign and benign of the skin lesion in dermscopic images. The proposed DCNN was trained using additive training procedure for adjusting the learning samples for different CNN models. This training procedure was repeated to improve the model's performance on wrongly classified samples. The ISIC 2016 dataset was used to train and validate the proposed deep learning model. The achieved results established the effectiveness of the proposed model with AUC as 0.98 and 0.81 on training and testing sets.

REFERENCES

[1] S. S. Han, M. S. Kim, W. Lim, G.H. Park, I. Park, and S. E. Chang, "Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm,". Journal of Investigative Dermatology, vol. 138, no. 7. Pp. 1529- 1538, 2018.

[2] J. Gachon, P. Beaulieu, J. F. Sei, J. Gouvernet, J. Claudel, M. Lemaitre, M., and J.J. Grob, "First prospective study of the recognition process of melanoma in dermatological practice," Archives of dermatology, vol. 141, no. 4, pp. 434-438, 2005.

[3] M. E. Celebi, H.A. Kingravi, B. Uddin, H. Iyatomi, Y.A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images,"Computerized Medical Imaging and Graphics, vol. 31, no. 6, pp. 362-373, 2007.

[4] G. Kaur, and E. N. Aggarwal, "Supervised Classification of Dermoscopic Images Using Gaussian Mixture Model and Artificial Neural Network," Imperial Journal of Interdisciplinary Research, vol. 2, no.8, 2016.

[5] X. Yuan, Z. Yang, G. Zouridakis, N. Mullani, "SVM-based texture classification and application to early melanoma detection," EMBS'06. 28th Annual International Conference of the IEEE in Engineering in Medicine and Biology Society, 2006. (pp. 4775-4778).

[6] M. A. Wahba, A. S. Ashour, S. Napoleon, M. M. A. Elnaby, and Y. Guo, "Combined empirical mode decomposition and texture features for skin lesion classification using quadratic support vector machine,' Health information science and systems, vol. 5, no. 1, 2017.

[7] C. Barata, J. S. Marques, and J. Rozeira, "Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model," In International Symposium on Visual Computing, 2013, (pp. 40-49). Springer, Berlin, Heidelberg.

[8] L. K. Ferris, J. A. Harkes, B. Gilbert, D. G. Winger, K. Golubets, O. Akilov, and M. Satyanarayanan, "Computer-aided classification of melanocytic lesions using dermoscopic images," Journal of the American Academy of Dermatology, vol. 73, no. 5, pp. 769-776, 2015.

[9] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," IEEE transactions on medical imaging, vol. 36, no. 3, pp. 849-858, 2017.

[10] B. Salah, M. Alshraideh, R. Beidas, F. Hayajneh, "Skin cancer recognition by using a neuro-fuzzy system," Cancer informatics, vol. 10, 2011.

[11] H. B. Burke, "Artificial neural networks for cancer research: outcome prediction," In Seminars in Surgical Oncology, (1994, January). (Vol. 10, No. 1, pp. 73-79). New York: John Wiley & Sons, Inc.

[12] 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 (pp. 118-126), 2015, Springer, Cham.

[13] Kawahara, J., BenTaieb, A., & Hamarneh, G. (2016, April). Deep features to classify skin lesions. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on* (pp. 1397-1400). IEEE.

[14] 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, 2017.

[15] A.Blum, R. Hofmann-Wellenhof, H. Luedtke, U. Ellwanger, A. Steins, S. Roehm, and H. P. Soyer, "Value of the clinical history for different users of dermoscopy compared with results of digital image analysis,"Journal of the European Academy of Dermatology and Venereology, vol. 18, no. 6, pp. 665-669, 2004.

[16] N.C.Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC), 2017.

[17] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble, International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI), 2017.

[18] I. G. Díaz, Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions, 2016, *arXiv preprint arXiv:1703.01976*.

[19] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. Avila, and E. Valle, "RECOD titans at ISIC challenge 2017," 2017, *arXiv preprint arXiv:1703.04819*.

[20] https://challenge.kitware.com/#phase/5667455bcad3a56fac786791

[21] T.Fawcett, "An introduction to ROC analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861-874, 2006.