Deep Convolutional Neural Network (DCNN) for Skin Cancer Classification

Nour Aburaed*, Alavikunhu Panthakkan†, Mina Al-Saad‡, Saad Ali Amin§ and Wathiq Mansoor¶
College of Engineering and IT, University of Dubai
Dubai, UAE
Email: *noaburaed@ud.ac.ae, †apanthakkan@ud.ac.ae, ‡malsaad@ud.ac.ae, §saamin@ud.ac.ae ¶wmansoor@ud.ac.ae

***Abstract*—Skin cancer is one of the most threatening types of cancer, with an increasing rates throughout the decade. Detecting and classifying skin cancer in its early stages provides better chances for treatment. In the recent years, Convolutional Neural Networks (CNNs) emerged as a powerful solution that aids the diagnosis of skin cancer. In this paper, Human Against Machine (HAM) 10000 dataset is used to demonstrate skin cancer classification strategy. VGG16, VGG19, and a Deep CNN proposed in this paper are implemented, trained, and evaluated. The dataset pre-processing steps and methodology are illustrated, and the network parameters and training process are explained. The performance of all three networks are compared in terms of the average overall accuracy and loss.**

***Index Terms*—Deep Learning, Convolutional Neural Networks, Skin Cancer, Classification, Biomedical Imaging, Image Processing**

I. Introduction

According to World Health Organization (WHO), cancer has been reported as one of the top causes of death [1], and according to statistical studies [2], more than 2 people die of skin cancer every hour in the US only. Several types of skin cancer have been increasing over the past decade. Around 3 million non-melanoma skin cancer, and 132,000 melanoma skin cancer are diagnosed globally every year [3]. The most common causes of skin cancer include Ultraviolet (UV) radiations, genetics, unhealthy lifestyle, and smoking. UV radiation has been reported as the most dominant cause for most skin cancer cases. With the rapid depletion of the ozone layers, more harmful UV radiations reach the surface of the earth, hence, the numbers of skin cancer cases are expected to rise. Early diagnosis of skin cancer gives great chances of successful treatment and preventing it from spreading to other organs. Dermoscopy, which is a noninvasive technique used to examine pigmented skin lesions at the surface level, can be used for early diagnosis. However, early diagnosis often proves to be a difficult task even for expert dermatologists. This difficulty arises from the fact that skin cancer takes various forms. For instance, it can appear as a form of swelling, and the challenge lies in detecting whether the swelling is benign or malignant. Moreover, skin cancer can take small forms that are difficult to detect with the naked eye, such as moles. Therefore, there is a need for more accurate and reliable detection. Hospitals have been utilizing image-based Computer Aided Diagnosis (CAD) systems due to their significant potential in screening and detecting melanoma in its early stages from Dermoscopic images. CAD systems treat these data from an image processing point of view by examining various features, such as texture, color, and shape in order to determine whether it is malignant or benign. In the case of positive skin cancer, determining the type is important because it helps to identify the most suitable treatment. Since there exists several possible types of cancer, this type of classification is highly non-linear. Therefore, Deep Convolutional Neural Networks (DCNNs) have the potential to classify them accurately, which has been proved in recent studies [4]–[13]. Mahbod et al. [5] used a hybrid approach for the purpose of improving the classification accuracy of skin lesions. They utilized Transfer Learning (TL) technique in multiple deep CNNs, such as AlexNet [14], VGG16 [15], and ResNet18 [16]. The aforementioned CNNs were used for feature extraction. Afterward, the extracted features were used for training different Support Vector Machine (SVM) classifiers. By combining the output from each classifier, a better classification performance is obtained. They tested their approach on ISIC 2016 and 2017 datasets combined, which consist of 2073 training images that include three different types of skin lesions; Malignant Melanoma (MM), Seborrheic Keratosis (SK), and Benign Nevi (BN). The authors evaluated their approach by computing the Area Under the ROC Curve (AUC), which is 90.69% on average. Ratul et al.[4] proposed a CAD system that can aid early detection of skin lesions. In this model, the authors applied dilated convolution instead of traditional convolution techniques, and utilized transfer learning in different pre-trained DCNN architectures, namely VGG16, VGG19, MobileNet [17], and Inception-v3 [18]. This approach was tested on Human Against Machine with 10000 training images (HAM10000) dataset. It was reported that the dilated Inception-v3 shows the highest overall classification accuracy of 89%. Some Dermoscopic images show artifacts that hinder the classification process, such as hair or a ruler placed next to the lesion to measure its diameter [19]. In [6], the authors developed deep learning model that utilized data purification to find and remove hair and rulers from Dermoscopic images, as well as applying data augmentation to solve the problem of data imbalance. ISIC 2017 and PH2 datasets were used to evaluate their approach. These two approaches show improvement in melanoma classification accuracy. Eliminating false negative and false positive cases is an issue of high importance in this field, and the percentage

978-1-7281-6044-3/20/$31.00 ©2020 IEEE

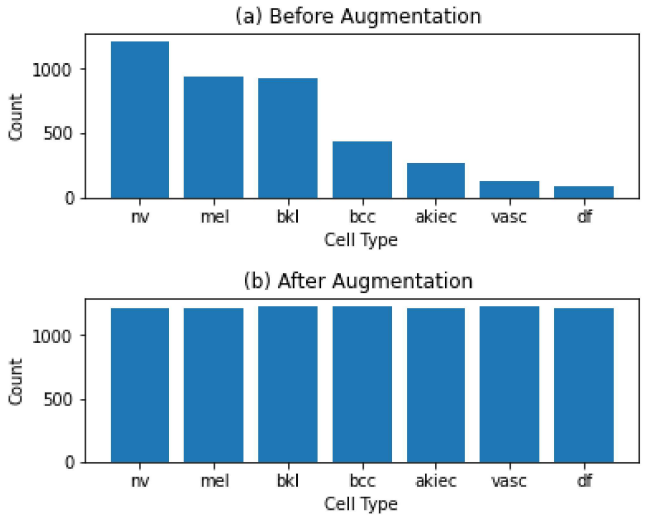

Fig. 1. Data size of each category in HAM10000 (a) before augmentation and (b) after augmentation.

of error needs to be minimized by as much as possible. Therefore, researchers always aim to further boost the accuracy of CNNs when detecting and classifying skin cancer types, especially when there are several possible types of cancer. In this paper, a Deep CNN is designed, implemented, trained, and evaluated on HAM10000 dataset after pre-processing it. Additionally, VGG16 and VGG19 are also implemented, trained, and evaluated for performance comparison purposes. The results are compared using the overall average accuracy and loss metrics. The rest of the paper is organized as follows: section II explores the dataset and pre-processing steps, section III explains the methodology and the training procedure of the propose DCNN in addition to VGG16 and VGG19, section IV illustrates the results of this study, and finally, section V summarizes the paper and draws the future direction of this research.

II. Dataset and Pre-processing

The dataset used in this study is HAM10000, which contains six types of cancer; Melanocytic Nevi (NV), Dermatofibroma (DF), Melanoma (MEL), Vascular Lesions (VASC), Basal Cell Carcinoma (BCC), and Actinic Keratoses (AKIEC), in addition to a non-cancer type Benign Keratosis-like Lesions (BKL). The dataset contains 10,015 images with their corresponding metadata, such as the patient's age, sex, and the location of the pigment. However, the dataset is highly imbalanced. Figure 1(a) shows that NV category has a significantly higher number of samples than all other categories. There is especially a severe lack in VASC and DF categories. This can affect the training process negatively and cause overfitting. To overcome this problem, several types of augmentation are applied to BCC, AKIEC, VASC, and DF. The augmentation methods include crop, scale, contrast and brightness adjustment, horizontal flip, vertical flip, and combinations of these

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 62, 62, 32)	896
max_pooling2d_1 (MaxPooling2	(None, 31, 31, 32)	0
conv2d_2 (Conv2D)	(None, 29, 29, 64)	18496
max_pooling2d_2 (MaxPooling2	(None, 14, 14, 64)	0
conv2d_3 (Conv2D)	(None, 12, 12, 128)	73856
max_pooling2d_3 (MaxPooling2	(None, 6, 6, 128)	0
flatten_1 (Flatten)	(None, 4608)	0
dense_1 (Dense)	(None, 512)	2359808
dense_2 (Dense)	(None, 7)	3591

Total params: 2,456,647
Trainable params: 2,456,647
Non-trainable params: 0

Fig. 2. Proposed DCNN architecture.

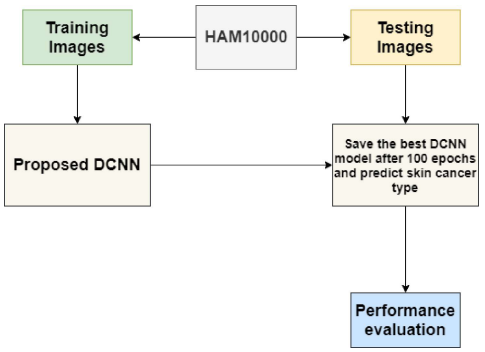

Fig. 3. Flowchart of DCNN training procedure.

methods. After augmenting the aforementioned categories and removing about 5000 samples from NV category, all categories have around 1000 samples. The final size of the dataset is 7182 with all categories balanced, as seen in Figure 1(b).

III. Methodology

The design of the proposed DCNN consists of a pair of convolutional layer and max pooling layer repeated 3 times, followed by a flattening layer and two densely connected layers. The kernel size for all convolutional layers is 3×3 with stride value 1, and the kernel size for all max pooling layers is 2×2 with stride value 1. ReLu is used as an activation function in the hidden layers. Since the output layer conveys seven different categories, softmax is used as an activation function. This DCNN was trained from scratch, and the general architecture of it is depicted in Figure 2. Next, a pre-trained VGG16 architecture was implemented and slight modifications were introduced with parameter fine-tuning. This VGG16 was not trained from scratch, and TL technique was used instead. VGG16 consists of a pair of 2 convolutional layers and a pooling layer repeated twice, followed by a pair of 3 convolutional layers and a pooling layer repeated 3 times, and finally, 3 densely connected layers. VGG19 is an enhanced version of VGG16, providing more depth in the network's layers. Similar to the procedure followed with VGG16, VGG19's parameters were also fine-

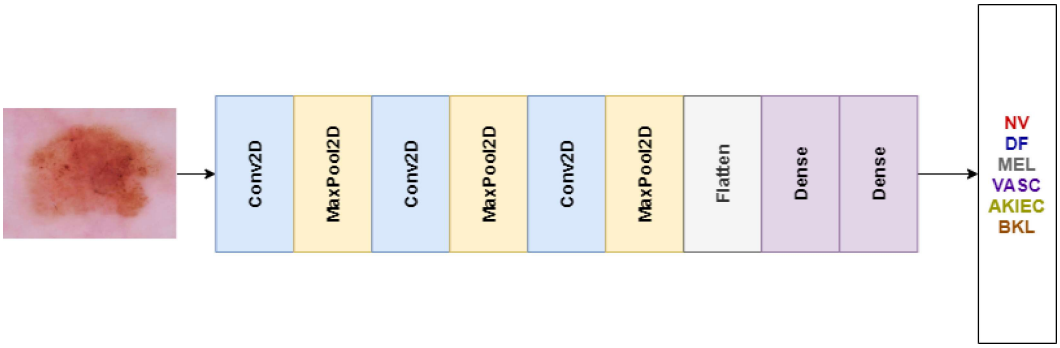

Fig. 4. DCNN for skin cancer classification.

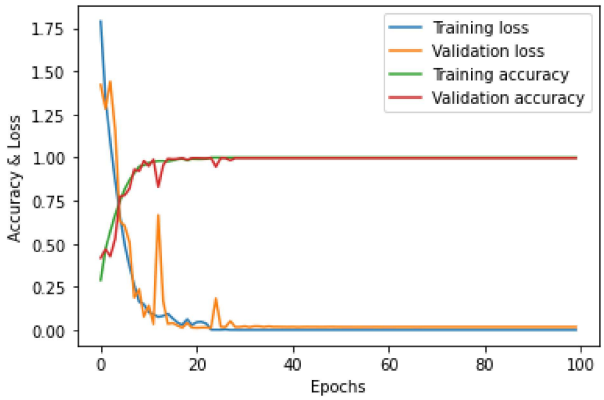

Fig. 5. Proposed DCNN accuracy and loss progression throughout 100 epochs of training.

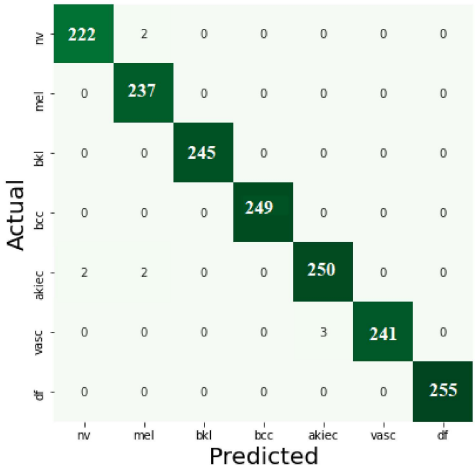

Fig. 6. Proposed DCNN confusion matrix.

tuned prior to training it using TL. All networks were trained using augmented HAM10000 dataset. 80% of the dataset was used for training, and the remaining for testing. From the training data, 20% were used for validation of the proposed model while training. All three networks' parameters were set to 100 epochs, 0.01 learning rate, and RMSprop optimizing algorithm. At the final epoch, the final network parameters are extracted from the epoch that provided the best training accuracy and loss. A general flowchart of the methodology is shown in Figures 3 and 4.

IV. Results and Discussion

The training and testing overall accuracy and loss of VGG16, VGG19, and the proposed DCNN are summarized in Table I. The proposed DCNN shows superiority in terms of both overall average accuracy and loss. The small difference between training and testing accuracy and loss is an indication that the network is not overfitting. Additionally, Figure 5 shows the progress of VGG16 training up to epoch 100. The accuracy increases steadily until it reaches the asymptotic of 100%. Conversely, the loss decays rapidly below 0.2. Furthermore, the confusion matrix for DCNN in Figure 6 illustrates that the vast majority of the predicted results fall under the correct categories. Some of the predicted results are shown in Figure 7.

TABLE I
Results Summary

	VGG16	VGG19	Proposed DCNN
Training accuracy(%)	98	96	99
Testing accuracy(%)	96	94	99
Training Loss	0.12	0.19	0.03
Testing Loss	0.16	0.20	0.02

V. Conclusion and Future Work

In this paper, skin cancer classification strategy was demonstrated using HAM10000 dataset. After solving the imbalance between the dataset's categories, a DCNN was implemented, trained, and evaluated. Furthermore, the performance of this DCNN was compared to the performances of VGG16 and VGG19. Both networks were fine-tuned, trained using TL, and evaluated. The result comparison between all three networks demonstrate that DCNN shows superiority compared to VGG16 and VGG19. Additionally, the proposed DCNN model did not overfit; the testing and training results were

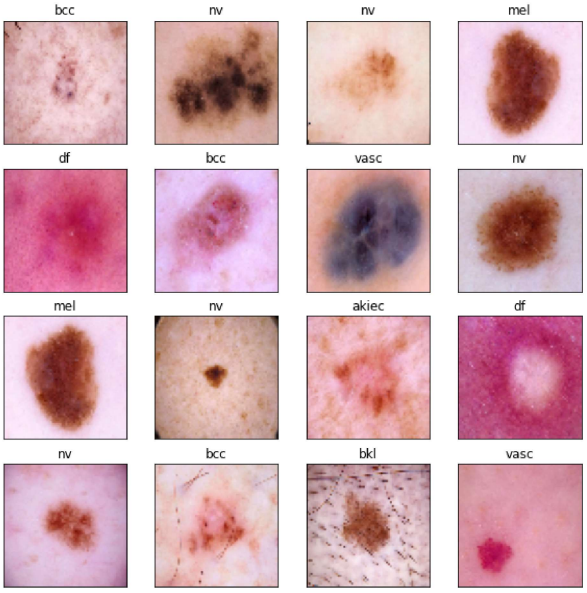

Fig. 7. Sample results from the proposed DCNN.

almost consistent. The next steps in this research include testing the networks on other skin cancer datasets, such as the ones provided in ISIC challenge. The aim is to make the network reselient to all types of Dermoscopic images. Furthermore, the metadata provided in HAM10000 dataset were not utilized. Exploiting additional information along with the images can potentially boost the performance and robustness of the network. These possibilities will be investigated as a future direction of this research.

REFERENCES

[1] "Cancer," https://www.who.int/news-room/fact-sheets/detail/cancer, Sep 2018.

[2] "Skin cancer facts & statistics," https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/, Jul 2020.

[3] "Ultraviolet (uv) radiation and skin cancer," https://www.who.int/news-room/q-a-detail/ultraviolet-(uv)-radiation-and-skin-cancer, Oct 2017.

[4] A. R. Ratul, M. Hamed Mozaffari, W.-S. Lee, and E. Parimbelli, "Skin lesions classification using deep learning based on dilated convolution," *bioRxiv*, 2019.

[5] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 1229–1233.

[6] D. Bisla, A. Choromanska, R. S. Berman, J. A. Stein, and D. Polsky, "Towards automated melanoma detection with deep learning: Data purification and augmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2720–2728.

[7] H. Alquran, I. A. Qasmieh, A. M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, and F. Hasayen, "The melanoma skin cancer detection and classification using support vector machine," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 2017, pp. 1–5.

[8] P. Bumrungkun, K. Chamnongthai, and W. Patchoo, "Detection skin cancer using svm and snake model," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–4.

[9] D. Moldovan, "Transfer learning based method for two-step skin cancer images classification," in *2019 E-Health and Bioengineering Conference (EHB)*, 2019, pp. 1–4.

[10] A. Demir, F. Yilmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3," in *2019 Medical Technologies Congress (TIPTEKNO)*, 2019, pp. 1–4.

[11] A. G. C. Pacheco, A.-R. Ali, and T. Trappenberg, "Skin cancer detection based on deep learning and entropy to detect outlier samples," 2019.

[12] M. Z. Alom, T. Aspiras, T. M. Taha, and V. K. Asari, "Skin cancer segmentation and classification with nabla-n and inception recurrent residual convolutional networks," 2019.

[13] M. A. Kadampur and S. Al Riyaee, "Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images," *Informatics in Medicine Unlocked*, vol. 18, p. 100282, 2020.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Very deep convolutional networks for large-scale image recognition," *Communications of the ACM*, vol. 60, no. 6, 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015.

[19] H. Mahmoud, M. Abdel-Nasser, and O. A. Omer, "Computer aided diagnosis system for skin lesions detection using texture analysis methods," in *2018 International Conference on Innovative Trends in Computer Engineering (ITCE)*, 2018, pp. 140–144.