Skin Cancer Classification Model Based on VGG19 and Transfer Learning

Nour Abuared*, 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

2020 3rd International Conference on Signal Processing and Information Security (ICSPIS) | 978-1-7281-8998-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICSPIS51252.2020.9340143

***Abstract*—Skin cancer is a concerning health issue with yearly increasing numbers. Detecting and classifying cancer type is problematic, especially since patients have to undergo several diagnosis over lengthy periods of time, which hinders early treatment and survival chances. With the aid of digital image processing, features can be extracted to identify skin cancer and its different types. Convolutional Neural Networks (CNNs) recently emerged as powerful autonomous feature extractors, and they have high potential to achieve high accuracy with skin cancer diagnosis. In this paper, two cancer types in addition to one non-cancer type taken from Human Against Machine (HAM10000) dataset are classified using CNN model based on VGG19 and Transfer Learning technique. The training strategy is explained, tested, and evaluated by calculating the network's overall accuracy and loss.**

***Index Terms*—Skin Cancer, Image Classification, Convolutional Neural Network, Transfer Learning**

I. Introduction

Skin cancer is the most common type of cancer that affects humans. It is primarily detected visually, followed by dermoscopic analysis. Dermoscopic analysis is the study of skin legions using dermatoscope mainly to evaluate pigmented skin lesions without obstructing skin surface. Detecting and classifying cancer type in its early stages is a very crucial task for the patient's health and well-being. However, even with dermoscopy, the patient needs to pay several visits to the clinician in order to monitor the skin pigment and observe any changes. This is a lengthy process and it is prone to errors, which puts the patient's life at risk. Therefore, a faster, more reliable process is needed in order to detect and classify skin cancer pigments. In the recent years, Convolutional Neural Networks (CNNs) outperformed dermatologists in distinguishing between the different types of skin cancer [1]. Therefore, efforts have been exerted in this research direction to improve the accuracy of skin cancer classification. In this paper, a skin cancer classification strategy is proposed, implemented, and evaluated. Human Against Machine (HAM10000) dataset is used in order to test the performance of the proposed strategy. The CNN used for the classification is based on VGG19. The training procedure and parameters are explained and demonstrated. The rest of the paper is organized as follows: section II probes into the literature of skin cancer detection and classification, section III describes the dataset and preprocessing steps, section IV demonstrates the methodology of the training procedure and CNN parameters, section V illustrates and discusses the results, and finally, section VI concludes the paper.

II. Literature Review

There are different types of skin cancer. Melanoma is considered as the most dangerous one. According to the American Cancer Society's statistics, there is an annual increase of 53% in the melanoma cases reported in United States (US) [2] [3]. Traditional diagnostic methods of the skin cancer depend on the visual inspection by dermatologists, which is time consuming and the diagnostic accuracy is dependent on the professional experience of the dermatologist. Therefore, developing an automated Computer Aided Diagnoses (CAD) system that utilizes different image processing algorithms for the detection of skin cancer can be considered as an alternative solution of the visual inspection [4]. In the recent years, CNNs emerged as a powerful image classifier, and this is the current direction most researchers are taking. Hosny et al. [5] proposed skin lesion classification technique for classifying three different types of skin cancer; Melanoma, Common Nevus, and Atypical Nevus. The authors utilized pre-trained AlexNet and replaced the last layer of the deep CNN with a softmax layer for three skin cancer lesions, where the classification accuracy of 98.61% was achieved and the results show superior performance compared to other existing methods. Similarly, Younis et al. [6] developed skin cancer classification method using pre-trained MobileNet for the classification of HAM10000 dataset into seven skin cancer types: Melanoma, Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis-like Lesion, Dermatofibroma, Vascular Lesion, and Melanocytic Nevi. Their approach achieved classification accuracy of 97.07% and fast prediction time within 2-3 seconds. Other approaches that rely on similar strategies include Pham et al. [7], which presented Inception-v4 network with data augmentation to improve the Melanoma classification performance with accuracy of 89.2%. Additionally, Refianti et al. [8] used LeNet-5 Deep CNN for the classification of Melanoma images by increasing the number of training images and number of epochs. Also, Demir et al. [9] presented two different deep learning algorithms, which are Resnet-101 and Inception-v3, for the classification of skin lesions into two types malignant and benign. The performance of Inception-v3 model is 87.42%, which is superior to 84.09% achieved with Resnet-101 architecture. Some approaches rely on both

978-1-7281-8998-7/20/$31.00 ©2020 IEEE
Authorized licensed use limited to: Carleton University. Downloaded on June 01,2021 at 18:21:47 UTC from IEEE Xplore. Restrictions apply.

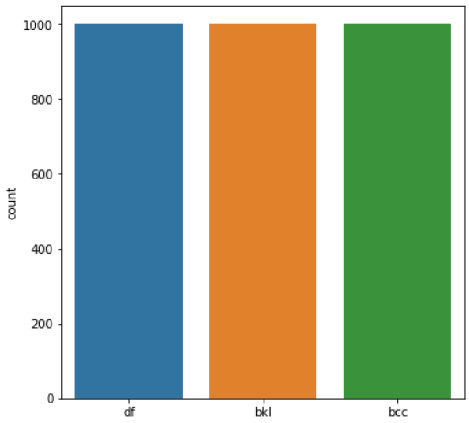

Fig. 1. Frequency histogram of skin cancer dataset.

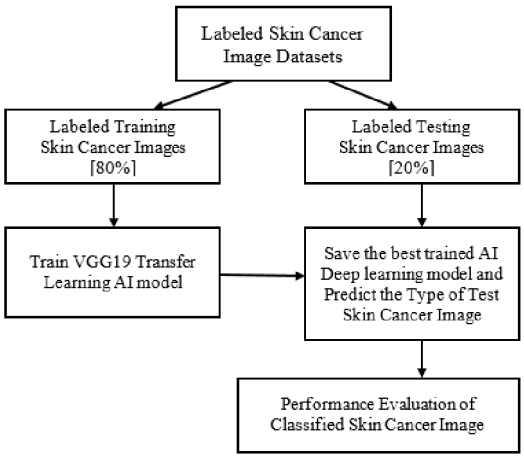

Fig. 2. Flowchart of applying TL to pre-trained VGG19 for skin cancer classification.

AI and human inspection, such as Hekler et al. [10]. The authors proposed hybrid approach of using both AI and human for classifying skin cancer into five different classes with accuracy rate of 82.95% compared to 81.59% achieved by CNN and 42.94% by human inspection individually. It can be concluded from the literature that this area of research requires further boost in classification accuracy, which is what this study attempts to achieve.

III. Dataset and Pre-processing

HAM10000 dataset was chosen for this study, which contains various types of skin cancer. Two types were selected from the dataset; Dermatofibroma (DF) and Basal Cell Carcinoma (BCC), in addition to one non-cancer type Benign Keratosis-like Lesions (BKL). However, imbalance exists in the dataset due to BKL being more common than the other two. Imbalance can affect the training process negatively and potentially cause overfitting. This is also a form of data bias, where one sample is more represented than the others. Therefore, augmentation is performed to increase DF and BCC types. The augmentation methods include crop, scale, contrast and brightness adjustment, horizontal flip, vertical flip, and combinations of these methods. After the augmentation, each skin cancer type has 1000 sample in the dataset, and the final size of the dataset is 3000. Figure 1 shows frequency histogram of the three types. The images in the dataset were resized to 64×64.

IV. Methodology

VGG19 was first developed in [11], which is an enhanced version of VGG16. VGG19 is a deep CNN that consists of several convolutional layers and max pooling layers, known as feature extractors. These layers are followed by at least one fully connected layer, known as classifier. The size and numbers of the convolutional and fully connected layers are considered as a design choice determined by the architect of the CNN. The general architecture of VGG19 is seen in Figure 3. The input layer is set to size 64×64, and the output layer is replaced with softmax activation function that reflects one of the three cancer types. The autonomous feature extraction capability of VGG19 renders it easy to locate the features that distinguish each different cancer type without having to spend time inspecting them manually. This VGG19 model was utilized for skin cancer classification. By using a pre-trained VGG19 with fine-tuned parameters, Transfer Learning (TL) was applied. 80% of the dataset described in section III was used for training the network, and the remainder was used for testing. Out of the 2400 training images, 20% were used for validation to evaluate the network performance after each epoch. The network was trained over 100 epochs and a batch size of 50, with a learning rate of 0.01 The optimization function chosen for this network is Adam. After 100 epochs, the parameters of the model with the best performance were selected and used with the testing images in order to evaluate the overall performance of the network. A flowchart of the methodology described is depicted in Figure 2.

V. Results and Discussion

After training the network, it was tested on 600 images and the performance was evaluated using the overall accuracy and loss. The training and testing accuracy were 0.985 and 0.975, respectively. While the training and testing loss were 0.099 and 0.119, respectively. Figures 4 and 5 show the progression of accuracy and loss for training and validation from the start of the training until the final epoch. Table I summarizes the final results for both training and validation. It can be seen that the difference of the results between training and testing is not big, which means that the network is not overfitting. Moreover, the loss and accuracy stop fluctuating between epochs 60 and 70, which indicates that the network is stable. In order to further inspect the accuracy of the network, Figure 6 shows the confusion matrix. It can be observed that the vast majority of the predictions fall into the correct categories, whereas only a few incorrect predictions were made.

Figure 7 shows some of the sample results from the network. Each image is labeled with the network's prediction. The

Authorized licensed use limited to: Carleton University. Downloaded on June 01,2021 at 18:21:47 UTC from IEEE Xplore. Restrictions apply.

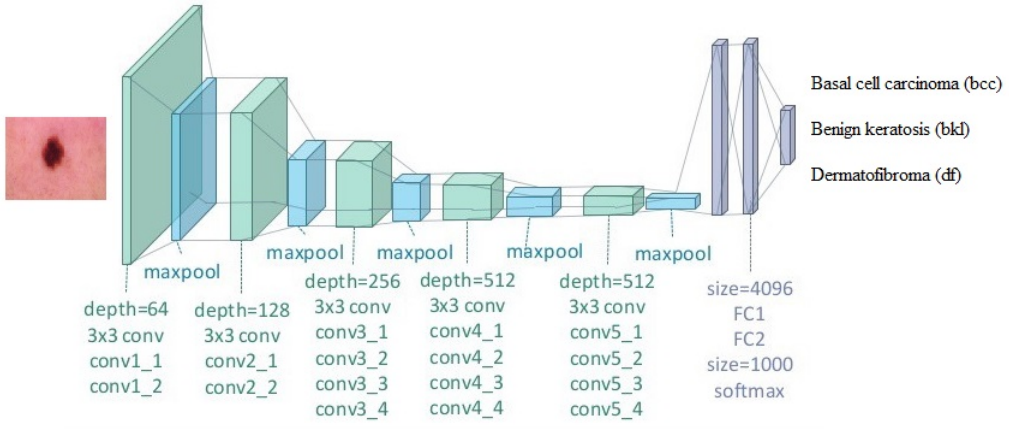

Fig. 3. VGG19 general architecture. [12]

TABLE I
RESULTS SUMMARY

Epoch	Training		Validation	
	Accuracy	Loss	Accuracy	Loss
25	0.9823	0.1094	0.9708	0.1264
50	0.9849	0.0997	0.9750	0.1188
100	0.9859	0.0991	0.9750	0.1185

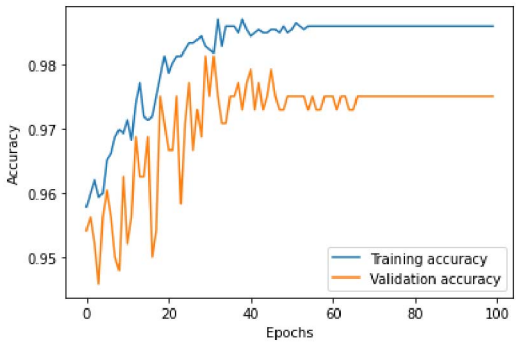

Fig. 4. Training and validation accuracy at each epoch..

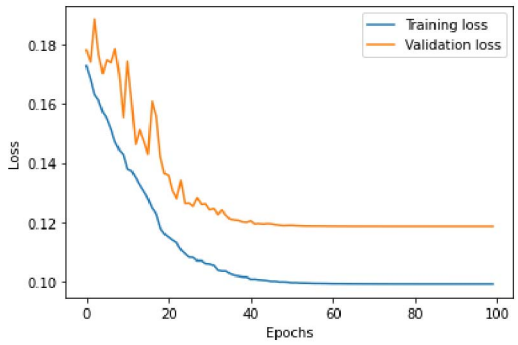

Fig. 5. Training and validation loss at each epoch.

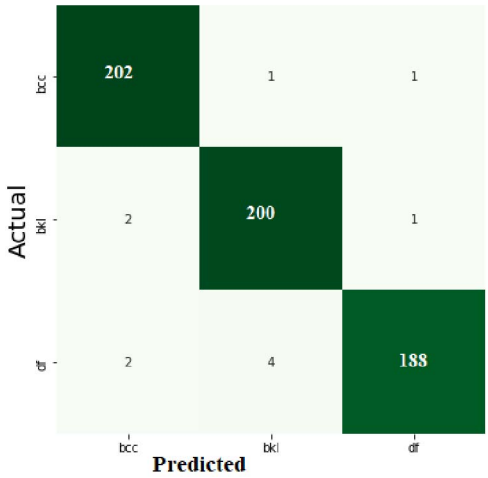

Fig. 6. Confusion matrix of VGG19.

prediction is correct if the predicted label matches the one in parenthesis.

VI. CONCLUSION AND FUTURE WORK

In this paper, a training strategy for classifying Dermatofibroma (DF), Keratosis-like Lesions (BKL), and Basal Cell Carcinoma (BCC) types of cancer was explained, demonstrated, and evaluated. VGG19-based CNN and TL proved to be powerful tools to aid skin cancer diagnosis with high accuracy. The overall accuracy and loss of the network indicate satisfactory outcome that can be improved further. The next steps in this research include covering a wider variety of skin cancer types. Additionally, further pre-processing steps can be taken to enhance the training accuracy further, such as hair removal.

REFERENCES

[1] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
[2] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," vol. 10, 2020.

Authorized licensed use limited to: Carleton University. Downloaded on June 01,2021 at 18:21:47 UTC from IEEE Xplore. Restrictions apply.

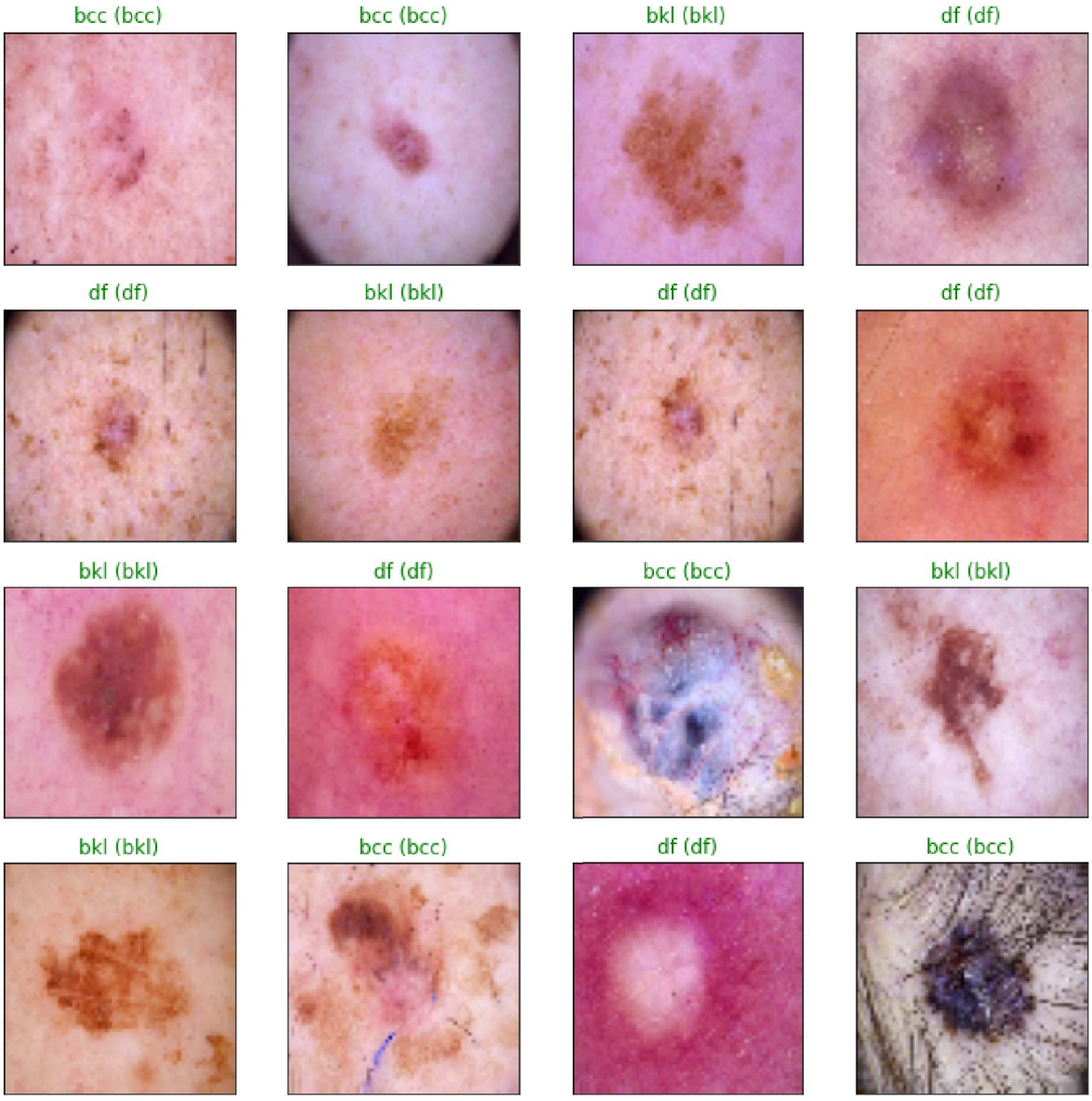

Fig. 7. Sample results from VGG19 with TL.

[3] 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.

[4] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, "Cancer diagnosis using deep learning: A bibliographic review," *Cancers*, vol. 11, p. 1235, 2019.

[5] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*, 2018, pp. 90–93.

[6] H. Younis, M. H. Bhatti, and M. Azeem, "Classification of skin cancer dermoscopy images using transfer learning," in *2019 15th International Conference on Emerging Technologies (ICET)*, 2019, pp. 1–4.

[7] T.-C. Pham, C.-M. Luong, M. Visani, and V.-D. Hoang, "Deep cnn and data augmentation for skin lesion classification," in *Asian Conference on Intelligent Information and Database Systems*, 2018, pp. 573–582.

[8] R. Refianti, A. B. Mutiara, and R. P. Priyandini, "Classification of melanoma skin cancer using convolutional neural network," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 10, pp. 409–417, 2019.

[9] 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.

[10] A. Hekler *et al.*, "Superior skin cancer classification by the combination of human and artificial intelligence," *European Journal of Cancer*, vol. 120, pp. 114 – 121, 2019.

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

[12] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," in *Computational Imaging III*, A. Mahalanobis, A. Ashok, L. Tian, and J. C. Petruccelli, Eds., vol. 10669, International Society for Optics and Photonics. SPIE, 2018, pp. 1–13.

Authorized licensed use limited to: Carleton University. Downloaded on June 01,2021 at 18:21:47 UTC from IEEE Xplore. Restrictions apply.