INTERNATIONAL JOURNAL OF CLINICAL & MEDICAL CASE STUDIES

Skin cancer classification using VGG-16

Abstract

Skin cancer with a high fatality rate is called melanoma. Due to the great degree of similarities among the many forms of skin lesions, a proper diagnosis cannot be made. Dermatologists can treat patients and save their lives by accurately classifying skin lesions in their early stages. This paper proposes a model for highly accurate skin lesion classification. The proposed model made use of transfer learning models known as VGG16. This model efficiently distinguished between benign and malignant cancerous skin lesions; those are the two distinct classes of skin diseases. The 1800 benign cancer images and 1498 malignant cancer images that were retrieved from the internet were considered for this proposed strategy. The VGG16 has obtained the recognition rates of 99.62% for training and 84.97% for validation.

Tanvir Mahmud[1,2]*; SA Sabbirul Mohosin Naim[3]; Prajoy Podder[4]

[1]*Department of Electrical and Electronic Engineering, Daffodil International University, Bangladesh.*
[2]*Department of Electrical and Computer Engineering, Lamar University, Beaumont, USA.*
[3]*Department of Electrical Engineering, School of Engineering, San Francisco Bay University, USA.*
[4]*Institute of Information and Communication Technology (IICT), Bangladesh Univewersity of Engineering and Technology(BUET), Bangladesh.*

*Corresponding author: **Tanvir Mahmud**

Department of Electrical and Electronic Engineering, Daffodil International University, Dhaka, Bangladesh.
Email: 0416312017@ict.buet.ac.bd

Introduction

A frequent type of cancer that is often caused by sun exposure is skin cancer [1]. Melanoma and nonmelanoma cases are classified according to the clinical category [1-3]. Even though melanoma is less common than nonmelanoma, it is the deadliest [2-5]. In 2019, the average death rate was 4740 males and 2490 females because of melanoma [1]. Indeed, the high cure rate of skin cancer mainly depends on early detection. In fact, dermatologists usually depend on dermoscopic images for melanoma detection and diagnosis. Despite this being a common method for diagnosis, it is not accurate. This is due to the high similarity of features between melanoma and nonmelanoma lesions. Additionally, image noise in the form of blood vessels and hair4 can lead to confusion. The accuracy of correctly diagnosing melanoma using visual examination is less than 80% [6].

To improve the diagnosis of melanoma, several automated systems have been proposed for identifying skin lesions to aid dermatologists. In this context, Machine Learning (ML) and deep learning methods were adopted. Various publications have classified melanoma and nonmelanoma lesions using ML methods. However, many limitations arise in terms of shallow training performance, extensive image processing, and variations in interclass and intraclass dermoscopic images [7].

Literature review

If skin cancer is found in its early stages, treatment may be given. The patient's skin observations can aid surgeons in making the best treatment options based on skin cancer images. Numerous studies on the identification and classification of skin cancer have been published in the literature. The paper published by Shahin et al. [3] has proposed the deep learning-based skin cancer classification. They used the HAM10000 dataset and

obtained 96.16% recognition accuracy for training, and for testing, they achieved 91.96% recognition accuracy, respectively. They compared the model with other models such as ResNet, AlexNet, VGG-16, MobileNet, DenseNet, etc.

In Jinnai et al. [4], they reported the work on skin cancer development for pigmented skin lesions by applying the deep learning method. From this, they achieved a highest recognition accuracy of 91.5%. The integrated design of deep features fusion-based skin cancer classification is reported by Amin et al. [5]. They used deep features and applied PCA on the features and obtained 99.00% recognition accuracy.

A seven-way skin cancer classification using MobileNet is reported by Chaturvedi et al. [6], using a pretrained model named MobileNet on the 2014 ImageNet Challenge dataset and obtained an overall accuracy of 83.10%. A review paper on skin cancer classification using various methods such as deep learning and CNN is given by Manne et al. [7].

An interpreted deep learning method to segment and classify non-melanoma skin cancer is reported by Thomas et al. [8], from which they obtained 97.9% recognition accuracy. Filali et al. [9] showed the effective use of handcrafted and CNN features for skin cancer classification. They used the PH2 dataset and obtained 98% recognition accuracy.

A novel approach is reported in the work of Singh et al. [10]. They proposed a transfer learning (TL) framework called Transfer Constituent Support Vector Machine (TrCSVM), through which they obtained an overall recognition accuracy of 98.82%.

Methods

Initially, this network was proposed by Simonyan et al. [12]. It is one of the best network models to date. Beginning with VGG11 and ending with VGG19, the VGG research group released a series of convolutional network models. A total of 1,800 benign cancer images and 1,498 malignant cancer images retrieved from the internet were taken into account for this proposed strategy [13].

To enhance classification results, the input dataset images were utilized for both training and testing the model. The proposed technique employed the ISIC dataset, which was also previously used for Skin Lesion Analysis towards the Detection of Melanoma in the ISBI 2016 challenge. A total of 3,297 images were collected, with 2,637 images (1,440 non-melanoma and 1,197 melanoma) used for training and 660 images (300 melanoma and 360 non-melanoma) for testing.

The programming language used for project development was Python, and the deep learning models were run using Google Colab. Input images must be preprocessed to reduce unwanted distortions and enhance image quality. This step is essential to prevent the deep CNN model from being misled during training. To reduce the effect of noise, a mean filter was applied to the entire dataset to make it noiseless. Additionally, a median filter was used because the input images contain important features at the edges. The mean filter removes noise without affecting these edges. To increase the number of training images, data augmentation techniques such as horizontal and vertical flipping were applied.

The primary goal of the VGG architecture group was to examine how the depth of convolutional networks affects the accuracy of models for large-scale image classification and recognition. For instance, VGG19 consists of 16 convolutional layers and 3 fully connected layers, while the smallest variant, VGG11, contains 8 convolutional layers and 3 fully connected layers. The final three fully connected layers remain consistent across all VGG variants.

The overall structure of VGG models includes five sets of convolutional layers, each followed by a MaxPooling layer. As depth increases—from VGG11 to VGG19—additional cascaded convolutional layers are added to each set. VGG16, for example, consists of 16 layers in total. The figure below illustrates the VGG16 architecture.

VGG16 is considered one of the most effective vision model architectures to date. It increases depth by using small (3 × 3) convolutional filters, significantly improving network performance. This architecture is particularly well-suited for large datasets.

In the experimental setup, skin cancer images were resized to 224 × 224 and used as input to the model. The dataset was split using an 80/20 ratio, with 80% for training and 20% for testing. The hyperparameters were configured as follows: maximum epochs = 25, mini-batch size = 128, and learning rate = 0.0001, using the Adam optimizer. Analyzing the size reduction after each MaxPooling layer was also essential.

In ImageNet—a dataset with over 14 million training images and 1,000 object classes—the VGG16 model achieved a test accuracy of 92.7%. It was a standout performer in the 2014 ILSVRC competition. VGG16 improved upon AlexNet by replacing large filters with a series of smaller 3×3 filters. For example, AlexNet used an 11-sized kernel in its first layer and a 5-sized kernel in the second layer.

The VGG model was originally trained for several weeks using NVIDIA Titan Black GPUs. It differs from earlier models in several ways. Notably, it uses the smallest 3×3 receptive field with a 1-pixel stride. The effect of a larger receptive field is achieved by stacking multiple small filters. This strategy improves decision functions and accelerates convergence by increasing the number of non-linear activation layers.

Additionally, the smaller convolutional filters used in VGG16 help reduce the risk of overfitting during training. A 3×3 filter size is considered optimal, as smaller filters are insufficient for capturing spatial information effectively in all directions.
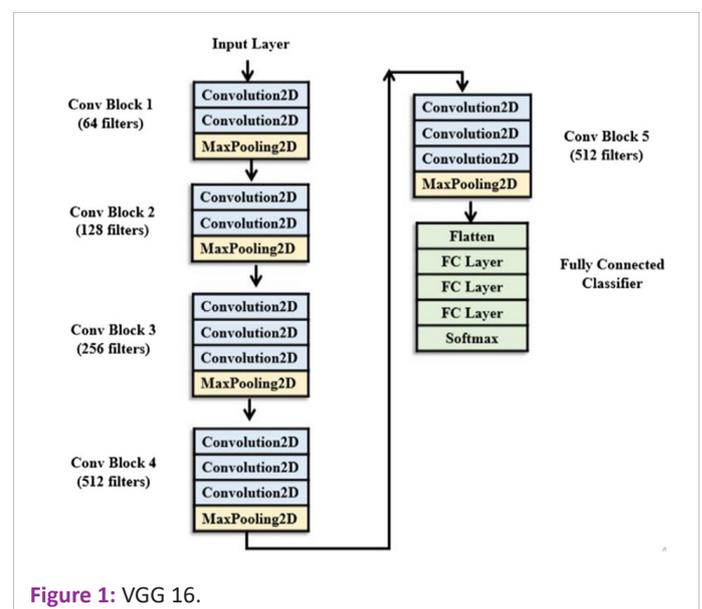

Figure 1: VGG 16.

Results

The ISIC dataset, a publicly accessible collection obtained through Kaggle, comprises 3297 images. For training, 2637 images are utilized, including 1440 non-melanoma and 1197 melanoma images, while 660 images (300 melanoma and 360 non-melanoma) are reserved for testing. Initially unlabeled, the input images undergo a labeling process. Figure 3 displays a selection of labeled random images from both melanoma and nonmelanoma classes, all resized to 224 × 224 pixels. Then, the data are normalized. Now, as shown in Figure 5, a model is built using the transfer learning concept [45,46], in which a VGG16 Deep CNN available in Keras is pretrained on the ImageNet dataset, which is the largest dataset with approximately 1.2 million number of images used for training, 50,000 images taken for validation, and 100,000 number of images taken for testing, divided into 1000 categories. Transfer learning could be a technique which allows to retrain the last layer of existing architecture which is employed to decrease the training time. As a result, all convolutional blocks are frozen, and only fully connected layers are trained on the new dataset. Before building and training the model, the convolutional layer is frozen. Freezing (by putting a layer on top of it) This prohibits the weights in a particular layer from being modifed during training (trainable=False) he model is been trained for 30 epochs and a learning rate of 0.002 with a batch size of 32 using the train images. For optimization, the Adam optimizer is used. As the convolution layers have been trained already onthe ImageNet dataset, the training time is low.
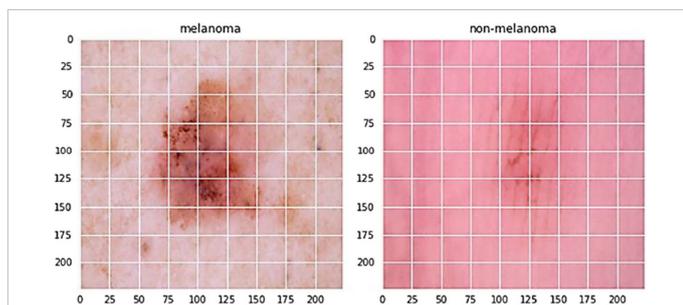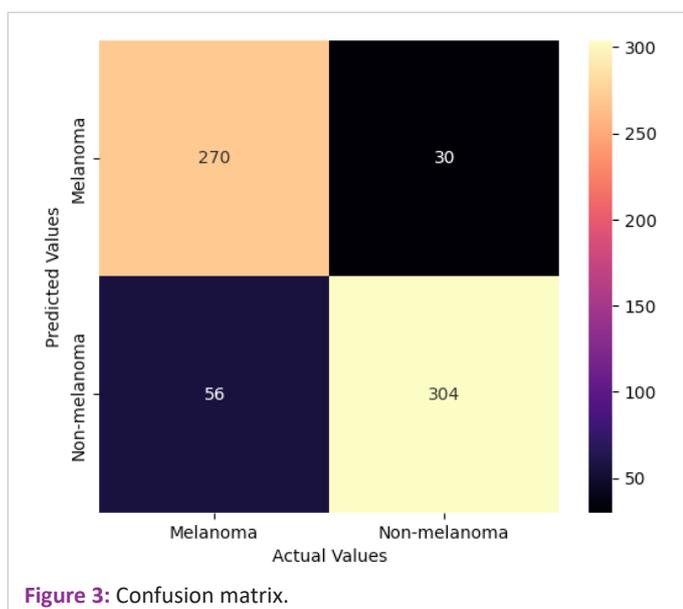

Figure 2: Labeled data.

Figure 3: Confusion matrix.

```
Classification Report:
              precision    recall  f1-score   support

    Melanoma     0.8282    0.9000    0.8626       300
Non-melanoma     0.9102    0.8444    0.8761       360

    accuracy                         0.8697       660
   macro avg     0.8692    0.8722    0.8694       660
weighted avg     0.8729    0.8697    0.8700       660
```

Figure 4: Classification report.

Table 1: Model evaluation.

Method	Accuracy	Recall	Precision
NASNet Mobile	85.42	85.35	85.44
NASNet Large	83.52	83.90	83.31
CNN [2]	81	81	
Ensemble CNN	83.6	64	
VGG	86.97	87.22	86.92

Conclusion

This study presents an efficient skin cancer classification framework utilizing the VGG16 deep convolutional neural network through a transfer learning approach. The model was trained and evaluated on a curated dataset comprising 3,297 dermoscopic images categorized into melanoma and non-melanoma classes. Extensive preprocessing, including noise reduction and data augmentation, enhanced the robustness of the input data. The proposed VGG16-based model achieved a training accuracy of 99.62% and a validation accuracy of 84.97%, outperforming several baseline and state-of-the-art architectures such as NASNet Mobile, NASNet Large, and conventional CNNs. These findings highlight the potential of transfer learning with deep CNNs in improving early and accurate diagnosis of skin cancer, particularly melanoma, which is crucial for reducing mortality rates. Future work will aim to further enhance classification performance by incorporating more diverse datasets and exploring hybrid ensemble models for improved generalization in real-world clinical settings.

Disclosure of conflict of interest: No conflict of interest to be disclosed.

References

1. Majumder S, Ullah MA. Feature extraction from dermoscopy images for melanoma diagnosis. SN Appl Sci. 2019; 1: 753.

2. Jaisakthi SM, Aravindan C, Appavu R. Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimed Tools Appl. 2023; 82: 15763–15778.

3. Li Z, et al. A classification method for multiclass skin damage images combining quantum computing and Inception-ResNet-V1. Front Phys. 2022; 10: 1120.

4. Abd Elaziz M, Dahou A, Mabrouk A, El-Sappagh S, Aseeri AO. An efficient artificial rabbits optimization based on mutation strategy for skin cancer prediction. Comput Biol Med. 2023; 163: 107154.

5. Ahmed KT, Rustam F, Mehmood A, Ashraf I, Choi GS. Predicting skin cancer melanoma using stacked convolutional neural networks model. Multimed Tools Appl. 2023.

6. Vestergaard ME, Macaskill PHM, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of

primary melanoma: A meta-analysis of studies performed in a clinical setting. Br J Dermatol. 2008; 159: 669–676.

7. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging. 2017; 36: 994–1004.

8. Saleh N, Hassan MA, Salaheldin AM. Skin cancer classification based on an optimized convolutional neural network and multi-criteria decision-making. Sci Rep. 2024; 14: 17323.

9. Ali MS, Miah MS, Haque J, Rahman MM, Islam MK. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Mach Learn Appl. 2021; 5: 100070.

10. Jinnai S, et al. The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules. 2020; 10: 1123.

11. Amin J, et al. Integrated design of deep features fusion for localization and classification of skin cancer. Pattern Recognit Lett. 2020; 131: 63–70.

12. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet. In: International Conference on Advanced Machine Learning Technologies and Applications. Singapore: Springer; 2020.

13. Manne R, Kantheti S, Kantheti S. Classification of skin cancer using deep learning, convolutional neural networks–Opportunities and vulnerabilities–A systematic review. Int J Mod Trends Sci Technol. 2020.

14. Thomas SM, et al. Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. Med Image Anal. 2021; 68: 101915.

15. Filali Y, et al. Efficient fusion of handcrafted and pre-trained CNNs features to classify melanoma skin cancer. Multimed Tools Appl. 2020; 79: 31219–31238.

16. Singh L, Janghel RR, Sahu SP. TrCSVM: a novel approach for the classification of melanoma skin cancer using transfer learning. Data Technol Appl. 2020.

17. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, Najarian K. Melanoma detection by analysis of clinical images using convolutional neural network. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016. p. 1373–1376.

18. Thurnhofer-Hemsi K, López-Rubio E, Domínguez E, Elizondo DA. Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting. IEEE Access. 2021; 9: 112193–112205.

19. Subramanian B, Muthusamy S, Thangaraj K, et al. A novel approach using transfer learning architectural models based deep learning techniques for identification and classification of malignant skin cancer. Wirel Pers Commun. 2024; 134: 2183–2201.