Diagnosing Skin Lesion Cancer and Melanoma Using Transfer Learning

Fahmid Bin Kibria*, Khan Abrar Shams†, Md Rafid Reaz‡
Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh
*fahmid.bin.kibria@g.bracu.ac.bd, †khan.abrar.shams@g.bracu.ac.bd, ‡md.rafid.reaz@g.bracu.ac.bd

I. Introduction

Skin cancer is a severe condition brought on by the body's melanocyte cells, which develop abnormally and have a propensity to multiply and migrate through lymph nodes to harm neighboring tissues. On the surface of the skin, the injured cells show themselves as a mole. They could or might not be cancerous. Melanoma, on the other hand, is categorized as cancer since it poses a greater hazard. Over 1 million people each month die from skin cancer worldwide, and there were 300,000 additional cases per month in 2018. The 19th most prevalent illness with the greatest mortality rate is melanoma. Although dealing with the high death rate has been challenging, recent advances in artificial intelligence and image processing have given us reason to hope that the survival rate will eventually rise. More significantly, CAD tools are quicker and easier to use than the clinical methods that are now being used.

A trained dermatologist must carry out a step-by-step process for diagnosis, which is fairly expensive in terms of time and work expended. However, the outcome of the diagnosis might vary depending on the dermatologist's level of training, and it's been said that the accuracy rate for correctly detecting skin lesions is under 80%. Because there are only a limited number of highly skilled dermatologists accessible worldwide, the figures are further dropped. Basal, squamous, and melanocyte are the three primary subtypes of skin cancer. The most typical type of skin cancer is basal cell carcinoma. Its development is quite modest, and does not spread to other body areas, but it has a propensity to come back.

Code link:
https://colab.research.google.com/drive/1UNX0fFTLbCKogtYeYfow4vrW9fFKVj15?usp=sharing

II. CNN Architecture Description

CNNs are effective artificial intelligence (AI) systems for image processing that employ deep learning to carry out both generative and descriptive tasks. They frequently use machine vision, which includes image and video identification, recommender systems, and natural language processing (NLP).

VGG19 has 19 layers that have already been trained and has a strong comprehension of the characteristics of a picture in terms of form, color, and structure. The extremely deep VGG19 has been trained on a massive amount of different pictures for challenging classification tasks. We used a pre-trained VGG19 model as illustrated in Fig. 1 and modified the output layer(last layer). Since the diagnosis problem require seven lesion classes, we stacked 7 units of dense layers and the flattened feature is passed through softmax activation function.

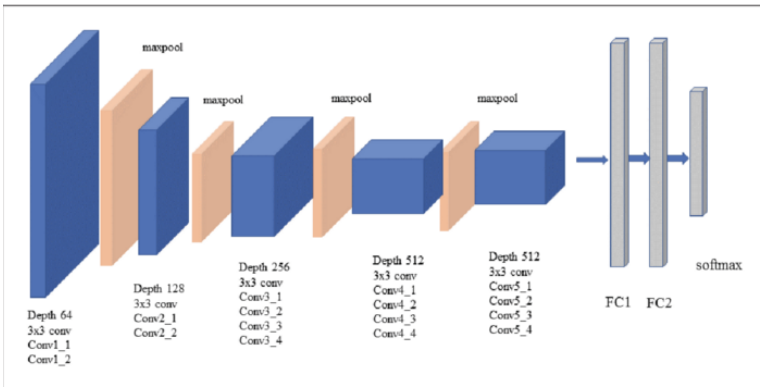

Fig. 1. VGG19 block diagram

III. Dataset Description

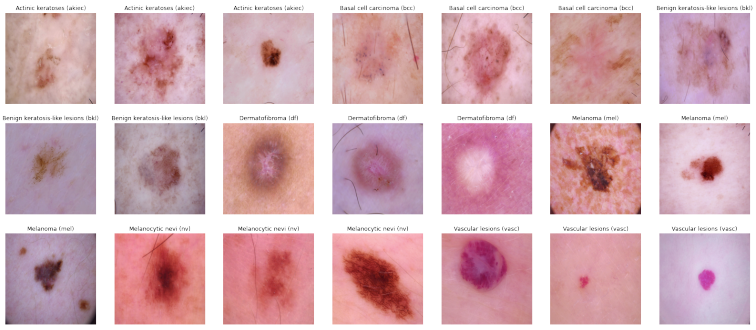

Fig. 2. Sample images of HAM10000

We used the HAM10000 dataset, which consists of a grand selection of thousands of multi-source dermoscopy images of common pigmented skin lesions. Among these are 115 images showing Dermatofibroma,142 images of Vascular Lesions, 327 pictures showing Actinic keratosis, 514 images of Basal cell carcinoma, 1099 pictures of Benign keratosis, 1113 images of Melanoma and 6705 images depicting Melanocytic nevi. This amounts to 10,015 dermoscopy images displaying

seven different types of skin cancer. Some sample images of skin cancer types from HAM10000 are represented in Fig .2 The frequency of images from seven classes are shown in fig.3

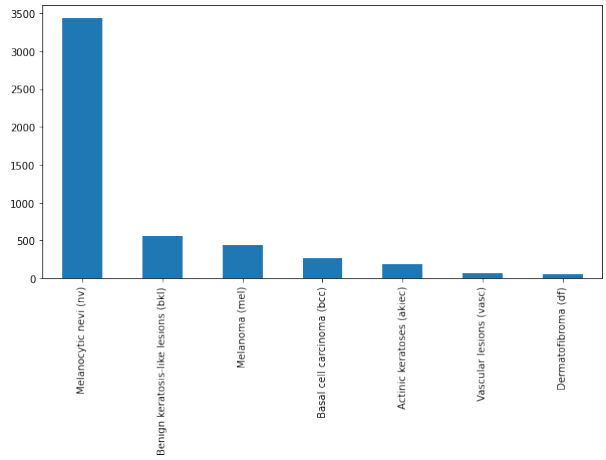

Fig. 3. Frequency of images from each class

IV. Experimental Results and Discussion

	precision	recall	f1-score	support
Melanocytic Nevi(nv)	0.84	0.91	0.87	677
Melanoma(mel)	0.34	0.26	0.29	104
Benign keratosis-like lesions(bkl)	0.41	0.42	0.41	114
Basal cell carcinoma(bcc)	0.35	0.27	0.31	44
Actinic keratoses (akiec)	0.30	0.24	0.27	37
Vascular lesions(vasc)	0.60	0.21	0.32	14
Dermatofibroma (df)	0.00	0.00	0.00	10
Accuracy			0.71	1000
Macro Average	0.41	0.33	0.35	1000
Weighted Average	0.69	0.71	0.70	1000

Fig. 4. Classification Report

The following evaluation metrics are used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1Score = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision} \tag{4}$$

$$Sensitivity(TPR) = \frac{TP}{TP + FN} \tag{5}$$

Fig. 4 represents the classification report of our model where precision, recall and F1-score for each of the seven lesion classes are evaluated using the equations (3), (2) and (4) respectively. The accuracy defined in equation (1) is the ratio of the model's correct predictions to the total number of predictions. Precision is used to determine the proportion of correct identifications. The recalls figure out what percentage of true positives(TPs) are correctly detected. Similarly, F1-Score metric measures the number of occurrences identified correctly by the learning model. Lastly, Support metric represents the number of samples true positives for each class.

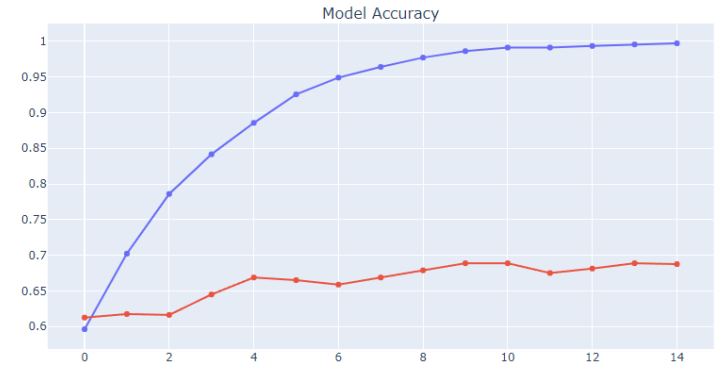

Fig. 5. Model Accuracy Curve

Fig. 6. Model Loss Curve

We achieved a training accuracy of 99.72% and test accuracy of 71% using our transfer learned model **TLVGG19**. Due to having limitations in computational power(used NVIDIA Tesla K80 GPU of Google Colab), we had to train only 2000 randomly selected dermoscopic images from original HAM10000. Thus, it is visible from the accuracy and loss curves shown in Fig. 5 and Fig. 6 that our model overfitted since training accuracy dominates testing accuracy.

V. Summary

The number of people diagnosed with skin cancer is increasing day by day. Recent progress in the field of deep learning facilitated medical professionals to automate the diagnosis with high accuracy. In future work, we would like to achieve higher test accuracy and reduce overfitting of the proposed model keeping in mind of the computational resource limitations. Therefore, this will ensure lesser risks in medical practices.