Convolutional Neural Networks Based Study and Application for Multicategory Skin Cancer Detection

Paul Zhu*
College of Arts and Science
New York University
New York, United States of America
* Corresponding author: pz745@nyu.edu

***Abstract*—Skin cancer is a very common form of cancer that jeopardizes people's health. Like other cancers, early detection is crucial to its treatment. However, traditional methods for diagnosing skin cancer can be low in accuracy and often leads to unnecessary examination. In addition, some existing Machine Learning models for skin cancer detection can be limited as well for the small number of skin cancer categories they support. In this work, three types of Convolutional Neural Network (CNN) models are compared on a nine-class skin cancer classification task and the model with the highest accuracy is integrated into a web application. The three CNN models that are compared include VGG-16, VGG-19, and a self-designed network. Since the three models differ in their depth, the relationship between the depth and performance of a model within the scope of the dataset used was also explored. Test results showed that the most accurate model is VGG-19, which achieved 0.9290 in accuracy and 1.2842 in loss, making it a reliable method to assist skin cancer detection.**

Keywords-component; Convolutional Neural Network, VGG, skin cancer detection, web application

I. Introduction

Skin cancer refers to the abnormal growth of skin cells. According to the Centers for Disease Control and Prevention (CDC), skin cancer is the most common cancer in the United States [1]. More specifically, it is estimated that approximately 9,500 people in the U.S. are diagnosed with skin cancer every day [2]. Divided into many subcategories, skin cancer posts threats of different levels to human bodies. For example, Basal Cell Carcinoma (BCC) has an excellent prognosis, with a 100% survival rate for cases that have not spread to other sites [3]; in comparison, Melanoma causes most of the deaths from skin cancer in the U.S., given it only accounts for about 1% of all skin cancers diagnosed [4]. Regardless of the type of skin cancer, early detection is a crucial part of its treatment. Therefore, it would be beneficial to develop an effective application that assists the early detection of skin cancers.

In addition to statistical evidence, some problems with the existing pattern of skin cancer treatment also indicate the importance of the application mentioned above. Conventionally, the first step of skin cancer diagnosis is the visual examination of a dermatologist, which could result in low accuracy. For example, research [5] showed that among studies of 1339 suspicious skin lesions, 268 will have a visual inspection indicating melanoma is present. Of these, 185 will not be melanoma and will result in an unnecessary biopsy. Therefore, the twofold demand for a better skin cancer detection mechanism is made clear.

Fortunately, as the power of Artificial Intelligence (AI) has been applied to a variety of fields in recent years, researchers have made many attempts to use different Machine Learning algorithms and, in particular, Convolutional Neural Network (CNN) models to find better solutions for tasks in medical image analysis, e.g., rehabilitation training analysis, breast cancer classification and skin cancer detection [6][7]. More specifically, in the work by Alheejawi et al [8]., deep learning techniques are applied to divide the melanoma regions in H&E-stained histopathological images for better detection accuracy; in the work by Nersisson et al [9]., You Only Look Once (YOLO) Convolutional Neural Network is used to extract features that assist the detection of skin lesions. Both of the models mentioned above achieved an accuracy of around 90%.

While many of the recent Machine Learning models are scientific breakthroughs, limitations still exist. For instance, in the model proposed by Pham et al. [10], only two classifications, namely melanoma and non-melanoma, are presented. And while the model created by Junayed et al [11]. achieves high accuracy, no application is provided for a better user experience.

To solve the current limitations mentioned above, this work compared the performances of three CNN models, namely VGG16, VGG19, and a self-designed model, in the classification task of nine categories of skin cancers. Based on their performances, the model with the highest accuracy is integrated into a web application.

II. Method

A. Date Source and Preprocessing

The data used come from the online platform Kaggle [12]. There are a total of 6,857 images of nine kinds of skin cancer, namely Basal Cell Carcinoma, Actinic Keratosis, Dermatofibroma, Melanoma, Nevus, Seborrheic Keratosis, Squamous Cell Carcinoma, and Vascular Lesion. Each image is of size 600×450px. Some sample images of the nine kinds of skin cancers are attached in Fig. 1.

To fit the images better for the models, original data sets are augmented in various ways. First, the size of the images is changed to 128×128px so that they can be processed by the models with a low number of trainable parameters. Then, the pixel values of the images are divided by 255 for normalization and the images are split into a training and testing subset respectively.

978-1-6654-7997-4/22/$31.00 ©2022 IEEE
DOI 10.1109/IWECAI55315.2022.00114

B. Proposed Approach

Given the task of image classification, CNN models are chosen because of their outstanding performance on image recognition tasks. CNN is a popular type of artificial neural network. It usually consists of convolutional layers, pooling layers, and fully connected layers. The “Convolution” part in the name CNN comes from the convolution operation in linear algebra.

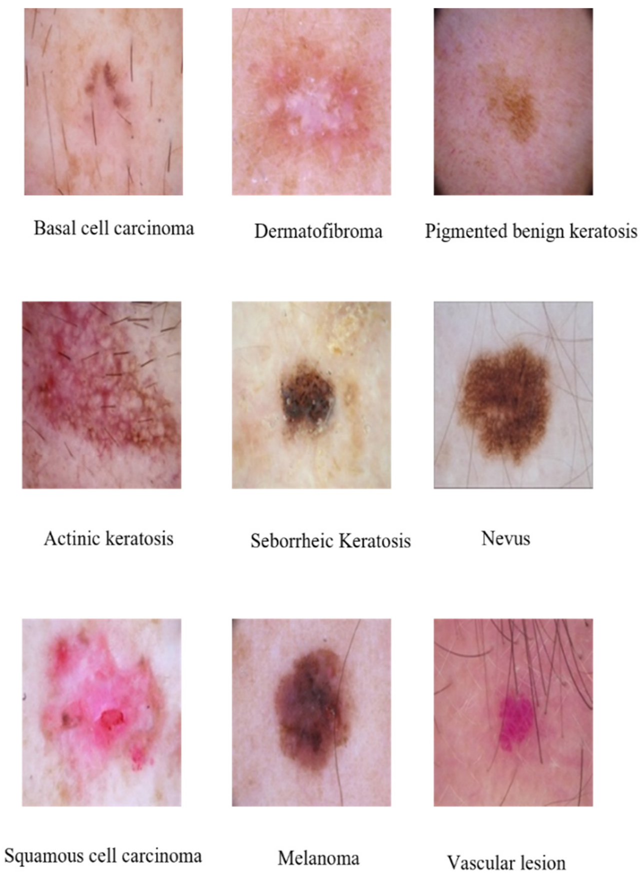

Figure 1. Sample images for the collected dataset.

To determine an ideal CNN network for the classification task, performances of the VGG models (both 16 and 19) built by the Visual Geometry Group at the University of Oxford [13] and a self-designed model are compared. VGG is chosen due to its remarkable accuracy achieved in the ImageNet Challenge 2014 and it is also a famous backbone that many studies used to extract features [14].

As for the self-designed model, it consists of 13 layers including rescaling, convolutional, pooling, and dense layers. The architecture of the proposed model is shown in Table I. To build the model layer by layer, the Sequential function provided by TensorFlow is used. Firstly, as its name suggests, the rescaling layer rescales the RGB channel values of the input images from the range [0, 255] to the range [0,1]. Then, four pairs of Conv2D and AveragePooling layers are used

TABLE I. THE ARCHITECTURE OF THE PROPOSED CNN MODEL

Layer (type)	Information for structure	
	Output Shape	***Param #***
rescaling_11 (Rescaling)	(None, 128, 128, 3)	0
conv2d_21 (Conv2D)	(None, 128, 128, 16)	448
average_pooling2d_15 (AveragePooling2D)	(None, 64, 64, 16)	0
conv2d_22 (Conv2D)	(None, 64, 64, 32)	4640
average_pooling2d_16 (AveragePooling2D)	(None, 32, 32, 32)	0
conv2d_23 (Conv2D)	(None, 32, 32, 64)	18496
average_pooling2d_17 (AveragePooling2D)	(None, 16, 16, 64)	0
conv2d_24 (Conv2D)	(None, 16, 16, 128)	73856
average_pooling2d_18 (AveragePooling2D)	(None, 8, 8, 128)	0
dropout_3 (Dropout)	(None, 8, 8, 128)	0
flatten_11 (Flatten)	(None, 8192)	0
dense_22 (Dense)	(None, 256)	2097408
dense_23 (Dense)	(None, 9)	2313

to extract features and reduce the parameters. The numbers of output filters are chosen to be 16, 32, 64, and 128 respectively in the four Conv2D layers and the height and width of the four 2D convolution windows are 3×3. For the activation function, ReLU is used to avoid the vanishing gradient problem and to speed up the training process. Among the pooling methods, average pooling is chosen since it can retain features of input images better than the max pooling. After the four pairs of Conv2D and AveragePooling layers, a dropout layer is added with a dropout probability of 0.3 to reduce the number of parameters and prevent overfitting. Lastly, a flatten layer and two dense layers are applied. The last dense layer has 9 neurons, matching the number of skin cancer types.

C. Implementation Details

Since the task involves both comparing the three models and integrating the model with the highest accuracy into a web application, two sets of parameters are used. On the one hand, smaller batch size and epoch can speed up the training process without affecting the rank of the accuracy of the three models. Therefore, batch size=16 and epoch=5 as the first set of parameters is used. On the other hand, after an ideal model is determined, larger batch sizes and epochs is important for more accurate prediction. Therefore, batch size=40 and epoch=30 is used as the second set of parameters. In both scenarios, Adam as the optimizer and sparse categorical cross-entropy as the loss function are chosen. The default learning rate is 0.0001 and the evaluation indicator is accuracy.

D. Application

For the application part, the proposed model is deployed as a web application. The homepage of the web application is shown in Fig.2.

The main tool used is Flask, a lightweight web framework written in Python. The HTML component of my application is rather simple, just a message prompting the user to upload the cancer image to be classified and an input form for uploading the image. One point worth noticing is that the

Figure 2. Homepage image

image uploaded by the user needs to be processed before it is sent to the model for prediction. Specifically, it needs to be expanded in dimension by adding batch size as one of its dimensions so that it can be accepted by the model.

III. Result and Discurssion

A. Comparison of the Three Models

Under epoch=5 and batch size=16, the training accuracy and loss of the three models are shown in Table 2. It can be observed that there is a significant gap in the performance between the self-designed CNN model and those of the VGG family. While the accuracy of the two VGG models is on the same level, VGG-19 outperforms VGG-16 by a narrow margin. Therefore, after retraining the VGG-19 model under epoch=22 and batch size=16, it is chosen to be integrated into the web application.

From Table 2, it can be observed that the training accuracy increases as the depth of the model increases. This result is somewhat intuitive. The first few layers of a CNN model usually target low-level features of an image, such as curves and edges. As the number of convolutional layers increases, the model becomes more capable of extracting higher-level, more abstract features that are composed of low-level features. Given the complicated, nine-category classification task, where the cancer images are rich in features like shape and color and yet might not be that distinct from one another (e.g., Sample images of Nevus and Melanoma in Fig.1), the ability to extract and distinguish high-level features is essential for high accuracy. Therefore, a shallow model like the self-designed network yields poor performance whereas VGG-19 shows high accuracy.

B. Web Application

upon opening the application, the user will see the homepage shown in Fig. 2. To upload an image for diagnosis, the user clicks the choose file button and uploads their image. The upload window is shown in Fig. 3. After a successful upload, the application jumps to another webpage and produces a result

TABLE II. Training Accuracy and Loss of the Three CNN Models

Performance	**Models**			
	Self-designed CNN model (epoch=5)	***VGG-16 (epoch=5)***	***VGG-19 (epoch=5)***	***VGG-19 (epoch=22)***
Training accuracy	0.5351	0.8377	0.8539	0.9290
Training loss	1.2124	1.5206	1.4909	1.2842

based on the proposed model that achieves the best accuracy. One example result is shown in Fig.4.

To generate a prediction, the trained model is firstly loaded into the Python program using a function called load_model in Keras. Then, after the model produces probabilities for each skin cancer category, the cancer type that has the highest probability value is chosen to be the predicted result and it will be presented to the user.

In summary, the workflow of the proposed model is shown in Fig. 5.

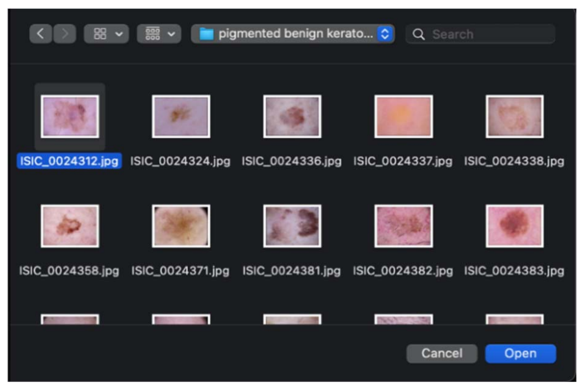

Figure 3. Image uploading

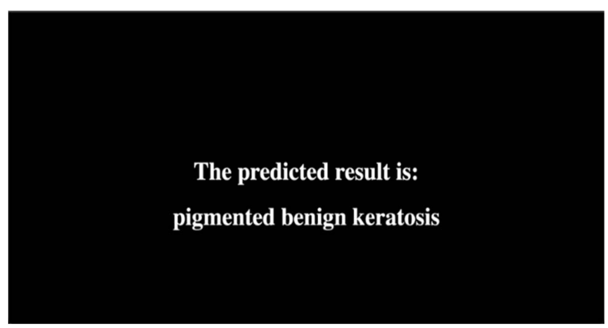

Figure 4. Example result image

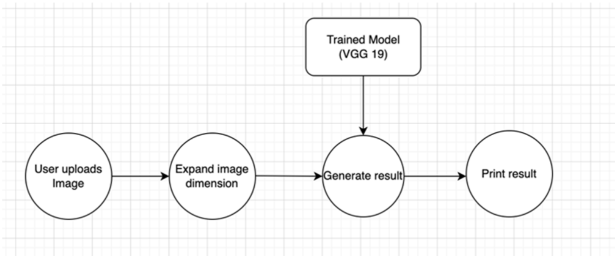

Figure 5. Workflow of the proposed model

IV. CONCLUSION

To summarize, performances of three CNN models, namely VGG-16, VGG-19, and a self-designed model, are compared on a multiclass skin cancer classification task. Empirical evidence suggests that VGG-19 has the best performance with a training accuracy of 0.9290 and a training loss of 1.2842. The VGG-19 model is integrated into a web application to assist the diagnosis of skin cancer. In the future, adding more functionalities to the web application is planned for a better user experience.

REFERENCES

[1] Basic Information About Skin Cancer. Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/skin/basic_info/index.htm 2021.

[2] SKIN CANCER. American Academy of Dermatology Association. https://www.aad.org/media/stats-skin-cancer, 2021.

[3] What is the prognosis of basal cell carcinoma (BCC)?. Medscape. https://www.medscape.com/answers/276624-100130/what-is-the-prognosis-of-basal-cell-carcinoma-bcc, 2021.

[4] Melanoma: Statistics. Cancer.Net. https://www.cancer.net/cancer-types/melanoma/statistics, 2021.

[5] How accurate is visual inspection of skin lesions with the naked eye for diagnosis of melanoma in adults?. Cochrane. https://www.cochrane.org/CD013194/SKIN_how-accurate-visual-inspection-skin-lesions-naked-eye-diagnosis-melanoma-adults, 2018.

[6] Y. Qiu, et al. "Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training." Biomedical Signal Processing and Control, vol. 72, 103323, 2022.

[7] Z. Wang, et al. "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features." IEEE Access 7, 105146-105158, 2019.

[8] S. Alheejawi, et al. "An efficient CNN based algorithm for detecting melanoma cancer regions in H&E-stained images" 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021.

[9] R. Nersisson, et al. "A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model" Arab J Sci Eng 46, 9797-9808, 2021.

[10] T. Pham, et al. "Deep CNN and data augmentation for skin lesion classification." Asian Conference on Intelligent Information and Database Systems. Springer, Cham, 2018.

[11] M. Junayed, et al. "A deep CNN model for skin cancer detection and classification.", International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 71-80, 2021.

[12] Augmented. Kaggle. https://www.kaggle.com/midhunkumar/augmented, 2021.

[13] K. Simonyan, and Z. Andrew. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[14] A. Sengupta, et al. "Going deeper in spiking neural networks: VGG and residual architectures." Frontiers in neuroscience 13, vol. 95, 2019.