Residual Learning Based Approach for Multi-class Classification of Skin Lesion Using Deep Convolutional Neural Network

V. N. Hemanth Kollipara[1](✉) and V. N. Durga Pavithra Kollipara[2]

[1] Vellore Institute of Technology, Vellore, India
hemanthkollipara95@gmail.com
[2] V R Siddhartha Engineering College, Vijayawada, India

Abstract. According to the Skin Cancer Foundation statistics, skin cancer is known to be the most common cancer in the United States and worldwide. By the age of seventy years, about twenty percent of Americans will have developed skin cancer due to exposure to radiation. Of all the types of skin cancers, melanoma is particularly deadly and responsible for most skin cancer deaths. Therefore, early detection is the key to survival. An automatic skin lesion diagnosis system can assist dermatologists since its challenging to differentiate between the different classes of skin lesions. In this paper, we propose a transfer learning based deep learning system using deep convolutional neural networks that leverage residual connections to perform the mentioned task with high accuracy. The HAM10000 dataset was utilized for training and testing the model and comparing its performance with other pre-trained models. This kind of automated classification system can be integrated into a computer- aided diagnosis (CAD) system pipeline to assist in the early detection of skin cancer.

Keywords: Skin lesion classification · ISIC 2018 · Convolutional neural networks · Transfer learning · Residual connections

1 Introduction

According to the World Health Organization statistics, 4 million skin cancers occur globally each year and are projected to only increase in the near future [1]. One in every five cancers diagnosed is skin cancer [2]. Early diagnosis is crucial as it shows a better survival rate, particularly in skin cancer. Skin cancer initially forms on the epidermal layer of skin where the cells grow abnormally and invade other tissues becoming noticeable by the naked eye. Exposure of skin to ultraviolet radiation is usually the cause of most skin cancers [3]. It alters the skin cell's DNA making the cell lose control over its growth, leading to cancer. Since skin cancer has become one of the significant causes of death, it is crucial to develop solutions for the early diagnosis of cancerous skin lesions before they become incurable.

The malignancy in the skin lesions is categorised primarily into non-melanoma and melanoma. When compared to other skin lesions, melanoma is one of the most

D. S. Guru et al. (Eds.): ICCR 2021, CCIS 1697, pp. 340–351, 2022.
https://doi.org/10.1007/978-3-031-22405-8_27

widespread and deadly. However, it is curable if it is detected in its early stage itself. However, malignant skin lesions are very much similar to benign skin lesions making it pretty hard to differentiate. Dermatologists can barely diagnose with an accuracy of 60% with their naked eyes. They can achieve accuracies of 75%-80% when trained and equipped with a dermatoscope since humans can make mistakes and dependency on the experience of the dermatologist. Dermoscopy is the process of imaging the pigmented skin lesions which showed significant improvement in skin cancer diagnosis compared to that of the naked eye. This brings us to the fact that computer-aided diagnosis (CAD) of dermoscopy images is capable of providing accurate preliminary diagnosis and automation of classifying the different skin lesions into their categories using computer vision.

With the current innovations in Artificial Intelligence developed over the past few decades, its applications in solving medical problems like a skin cancer diagnosis, particularly using deep neural networks, have enticed much attention. As diagnosis is a crucial factor, it is vital to employ the least error-prone method. With these advancements in artificial intelligence and the availability of data to train, employing deep neural networks for classification is an efficient method. A well-trained model can exceed an expert in classification, and this could help dermatologists. It is a dual benefit method with better accuracy and the benefits of automation. This work could change the process of diagnosing and help reduce the death rate due to skin cancers.

This paper proposes a transfer learning based approach using a pre-trained deep convolutional neural network model, a modified Inception-ResNetV2 that leverages residual connection to perform highly accurate classification of skin lesions from dermoscopy images and comparing its performance with other convolutional and transfer learning models.

2 Related Work

The classification of skin cancer like melanoma detection has been here since the 1990s with the development of artificial neural networks but are very much limited to lower accuracy than human diagnosis. Later machine learning approaches were used for classification. The drawback with machine learning algorithms such as support vector machine (SVM), K-nearest neighbour (KNN) is that they require a lot of heavy image processing for feature extraction, feature selection from image samples to estimate characteristics of these skin lesions like size, texture, colour, shape making the classification tasks challenging. However, these approaches unfolded poor results because of the high visual similarity between melanoma and non-melanoma skin lesions.

In recent years, there have been various advancements in digital image-based AI techniques like the development of Deep Convolutional Neural Networks (CNNs), which have outperformed conventional image processing techniques in various tasks like image classification, segmentation, object detection, and many more. CNN's are being preferred in solving tasks involving complex image processing since these approaches do not require any feature engineering, unlike machine learning approaches, resulting in higher accuracy. With the advent of the release of substantial open medical datasets from various organizations, the amount of data available on dermoscopy skin lesion images is massive.

This makes using deep learning models like convolutional neural networks viable since they require extensive data for training.

3 Dataset

A public dataset from the International Skin Imaging Collaboration (ISCI) archive was obtained for this work known as HAM10000, which was published in 2018 [5]. It is quite a large dataset that consists of 10015 samples of multi-source dermatoscopic images of common pigmented skin lesions. (1) Melanocytic nevi, (2) Dermatofibroma, (3) Melanoma, (4) Actinic Keratosis, (5) Benign Keratosis, (6) Basal Cell Carcinoma, (7) Vascular Skin Lesion are the seven different categories of dermatoscopic images in this vast dataset, which are denoted by NV, DF, MEL, AKIEC, BKL, BCC, VASC, respectively (Fig. 1). All the images present in the dataset are of 600 × 450 pixel resolution. The dataset is highly imbalanced since the number of sample images in each class is not uniformly distributed which can be observed from Fig. 2. For example, the NV class itself is about 67% (6705) of all images in the dataset, whereas the DF class is only 1.1% (115). This unequal distribution needs to be handled since this induces unwanted bias of the model towards the class of dominant frequency of occurrences.

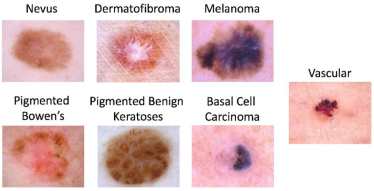

Fig. 1. Skin lesions images of the seven classes.

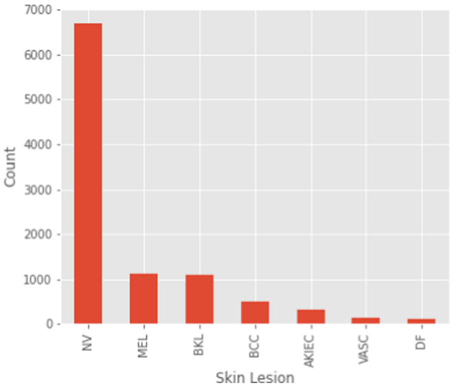

Fig. 2. Countplot of number of samples in each of the seven classes.

3.1 Data Preprocessing

Every image in the dataset is resized to a fixed dimension of 224 × 224. Each pixel is represented by a value between 0 and 255. Therefore, we normalize each pixel value by rescaling the images by a factor of 1/255. Finally, the processed dataset is divided into a training set (8012) and a validation set (2003) with a ratio of 80:20.

3.2 Data Augmentation

Since the training data is too imbalanced, augmenting the training set is required to avoid unwanted bias towards the majority class, making the model generalizable. In data augmentation, we employ some resampling techniques to augment the number of training data such that the total occurrences of each class in the dataset are almost the same. We can use several image augmentation techniques such as rotation, shear, zoom, scaling, horizontal and vertical flipping, height shifting, width shifting, brightness shifting (Fig. 3). This augmentation contributes to increasing the model's performance since the location of the skin lesion in the dermoscopy image is not fixed and differs from image to image.

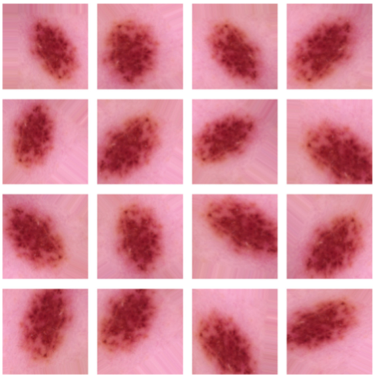

Fig. 3. Augmented images

4 Methodology

Deep Learning is a subset of Machine Learning in AI which consists of algorithms that are inspired by the functioning of the human brain called the neural networks, which consist of neurons connected to each other mimicking the human brain. These structures are called Neural Networks. It tells to computer to do as a human brain naturally does. In deep learning, there are several models such as Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Reinforcement Learning. However, there has been one algorithm that revolutionized computer vision and brought a lot of attention in tackling image classification problems, which is Convolution Neural Network (CNN) or ConvNets. Convolution Neural Network is a class of Deep Neural Networks that are useful to classify particular features from images and are mainly helpful in analyzing visual images. The various applications of CNN are image classification, image and video recognition, medical image analysis, image segmentation.

4.1 Baseline Model

In this approach, a baseline convolutional neural network is considered, which consists of four convolutional layers with [64, 64, 128, 256] filters, respectively, where each layer has a kernel size of 3 × 3. Then four pooling layers, each having a pool size 2 × 2, essentially halving the input dimensions, followed by a flattening layer, and two dense layers with a dropout rate of 0.5. The output dimension of the last dense layer is the number of classes (seven) in this classification.

4.2 Transfer Learning Model

Transfer Learning is a machine learning technique where the knowledge gained by the model while training for one task and applying that knowledge to another task that is a similar or downstream task of the first task for which the model is trained for the first time.

For a deep learning model to improve its generalization, it needs to be trained on enormous amounts of data which might consume tremendous computational resources and time. Hence transfer learning is a popular approach. Instead of training a deep model from scratch, pre-trained models leverage patterns learned from solving a previous problem applying the knowledge acquired from the previous task. This technique can be pretty helpful when a model needs to develop good generalization but is only limited to a small training dataset. These pre-trained models are usually trained on massive benchmark datasets and then can be used for a variety of downstream tasks.

In this paper, we have leveraged transfer learning for the classification of skin lesions by using state-of-the-art convolutional neural networks that are pre-trained on the Imagenet dataset like VGG16, Resnet50, DenseNet121, Efficient- NetB4. The architecture of these models is then modified according to the requirement of this particular classification. First, the top layer of these models is removed and then replaced with a custom pre-processing layer for rescaling and processing according to the models used. We replaced the last three layers with an average pooling layer that converts the two-dimensional convolutional layer to a single dimension [6]. Finally, added two fully connected dense layers along with drop layers in between to minimize overfitting. Softmax activation is used in the last dense layer as an activation function. All the layers in the architecture are unfrozen in order to retrain the entire model on the new training data and fine-tune it across all the layers for the model to learn new features and patterns.

4.3 Residual Connections Based Transfer Learning Approach Using Modified Inception-ResNetV2

Residual connections are a type of skip-connections where the gradients, instead of passing through non-linear activation functions, pass through the network layers directly. Introduced by He et al. in [12]. They help in the training of the model since non-linear functions sometimes cause vanishing or explosion of gradients. This results in very little or large updates to the weights in the neural network that cause the model to become unstable, leading to poor performance. We modified the Inception-ResNetV2 architecture, which belongs to the inception family but improved with the help of residual connections instead of the conventional filter concatenation stage. It combines the two architectures of Inception and Residual networks to obtain more solid performance but at the same time keeping the computational costs relatively low. It consists of a stem block, three sets of residual inception block modules with [5, 10] blocks of Inception-ResNetA, Inception-RetNetB, Inception-RetNetC modules, respectively, and subsequently pooling layer after each set of Inception-ResNet modules, all of which are connected sequentially. With a total of 164 layers, this deep convolutional network is capable of learning rich feature representation for broad categories of image data [14] (Fig. 4).

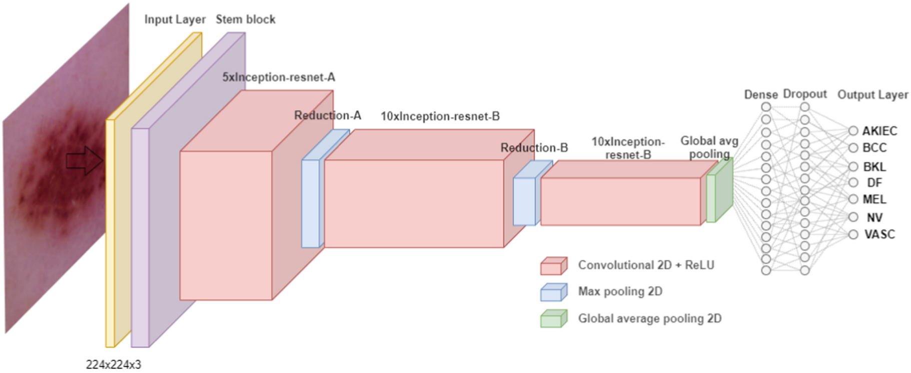

Fig. 4. Model architecture

A pre-trained model of this architecture is obtained, which is trained on more than a million images from the Imagenet dataset is taken as a base model, and modifications are done to retrain the model on the skin lesion classification problem. The top layer is replaced with a custom input layer with an input size of 224 × 224. A global average pooling layer is used to convert the learned features to a single dimension, which are then connected to two fully connected dense layers incorporated with dropout layers (0.5) and kernel regularizer applied to minimize overfitting. The final dense layer of seven nodes with softmax activation. The model is trained with all the layers unfrozen with approximately 54 million trainable parameters. The class imbalance inherently in the dataset is handled by compensating by using class-weighted learning while training. Here the majority classes are assigned less weightage, and more weight is given to minority classes such that the loss function penalizes them accordingly. Categorical cross-entropy is utilized as the loss function and Adam optimizer for training for the network. Callbacks are employed during training to reduce the learning rate dynamically upon learning stagnation based on metrics like validation loss, accuracy, and early stopping to terminate the training of the model if there is no improvement of the model for certain epochs. All the models are implemented using the Tensorflow framework and Keras library. The training is performed over Google Colab.

5 Results and Discussion

This section will summarize the results obtained using convolutional neural network using a confusion matrix and some evaluation metrics widely used for classification tasks. The evaluation metrics are Accuracy, Precision, Recall, F1- score. We evaluate the performance of the models obtained from the simulations and discuss the results obtained and the visualization of features extracted by some of the layers in the model.

$$Accuracy = \frac{True\,Positive + True\,Negative}{True\,Positive + False\,Positive + True\,Negative + False\,Negative}$$

$$Precision = \frac{True\,Positive}{True\,Positive + False\,Positive}$$

$$Recall = \frac{True\,Positive}{True\,Positive + False\,Negative}$$

$$F1\_Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right)$$

5.1 Baseline Models

The baseline convolutional neural network achieved an accuracy of 71% after being trained for over 40 epochs with a batch size of 64 before early stopping to prevent overfitting. The precision is 73%, recall being 71% with an F1-score of 71%. From Fig. 7, which shows the confusion matrix of this baseline CNN, it is pretty apparent that the model is biased towards the NV class and struggles to classify AKIEC, BKL, DF, MEL classes. With the performance of this model, it is evident that a much complex architecture is required.

Table 1. Results obtained from different model architectures

Model	Accuracy	Precision	Recall	F1-Score
Baseline CNN	71.11%	73%	71%	71%
VGG16	78.98%	78%	78%	78%
DenseNet121	81.18%	81%	81%	81%
EfficientNetB4	86.59%	87%	86%	86%
ResNet50	87.45%	88%	87%	87%
Modified Inception-ResNetV2	90.07%	89%	89%	89%

5.2 Transfer Learning Models

The results from Table 1 show that all the transfer learning models have out-performed the baseline model by a considerable margin and can generalize quite well. ResNet50 and EfficientNetB4 achieved almost similar performance with an accuracy of 86% and 87%, respectively, ResNet has a marginal lead over the EfficientNet model. VGG16 resulted in the lowest accuracy of all the transfer learning standing at 79%. Comparing the F1-score of all these models, ResNet achieves the top results reaching 87%. From the confusion matrices of all the models from Figs. 8, 9, 10 and 11, it can be observed that most models are able to differentiate quite well between seven classes.

5.3 Modified Inception-ResNetV2

Modified Inception-ResNetV2 clearly outperformed all the other models with an accuracy of 90% and precision, recall, and F1-score hitting 89%, 89%, 89%, respectively. This model is trained for 100 epochs with a batch size of 64. The Adam optimizer is initialized with an initial learning rate of 1e–4. Callbacks such as reduce learning rate on plateau, and early stopping are applied while training to avoid overfitting (Fig. 6). The performance of this model is compared with all other models with visualization in Fig. 13 (Figs. 5 and 12).

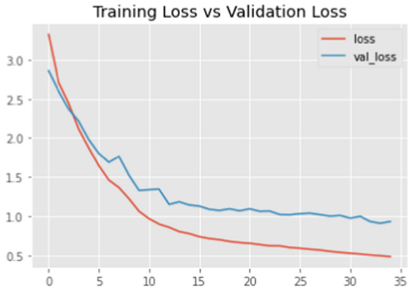

Fig. 5. Loss vs Epoch

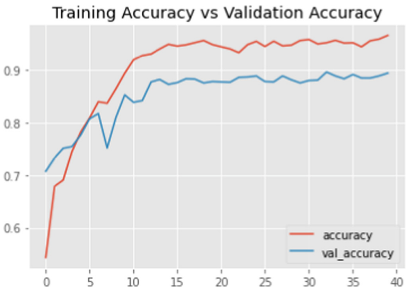

Fig. 6. Accuracy vs Epoch

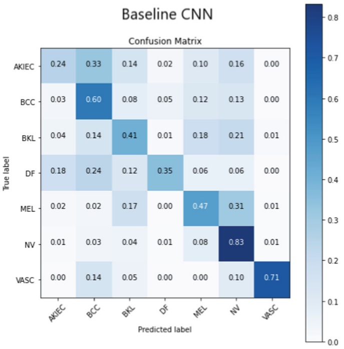

Fig. 7. Confusion matrix of base line model

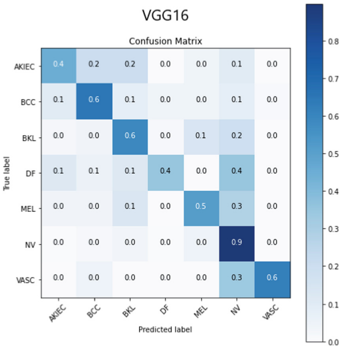

Fig. 8. Confusion matrix of VGG16 model

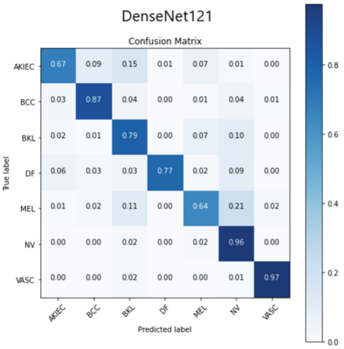

Fig. 9. Confusion matrix of DenseNet121 model

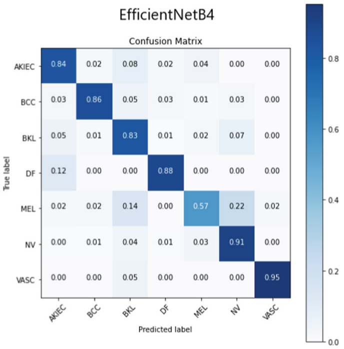

Fig. 10. Confusion matrix of EfficientNetB4 model

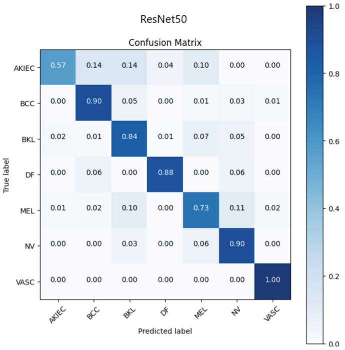

Fig. 11. Confusion matrix of ResNet50 model

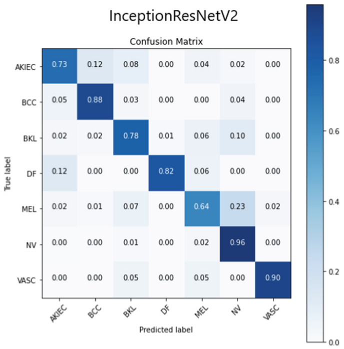

Fig. 12. Confusion matrix of modified inception-ResNetV2

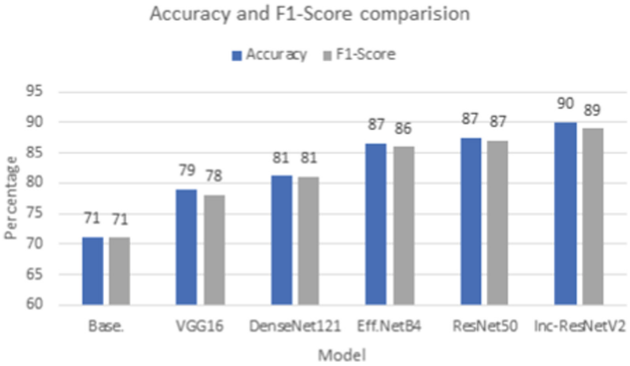

Fig. 13. Comparison of accuracy and F1-score across different models

5.4 Model Interpretation

The tf-keras-vis library has been utilized to produce visualizations of the convolutional filter layers, convolutional layer outputs, activation maps. These visuals aim to understand where the model is focusing its attention to perform the task. Figure 14 represents some of the convolutional filters which are applied to extract the convolved features that are passed to the following layers in the network. Figure 15 represents the features that are extracted from a sample image at different layers in the network. The GradCAM algorithm [16] is applied to visualize the regions that trigger the activations, i.e., the areas that contribute the most in producing the output. The yellow indicates the most attention, and the violet indicates the least, as seen from Fig. 16. In most cases, the model looked only at the lesion region to produce its output, but in some cases, GradCAM visuals show that the model focuses on regions outside lesions to determine the output.

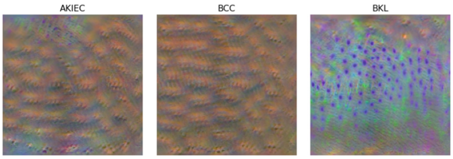

Fig. 14. Convolutional filters

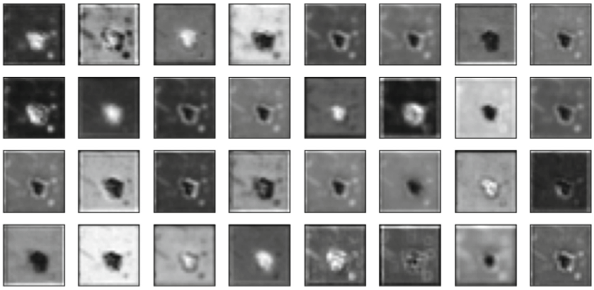

Fig. 15. Output of convolutional layer

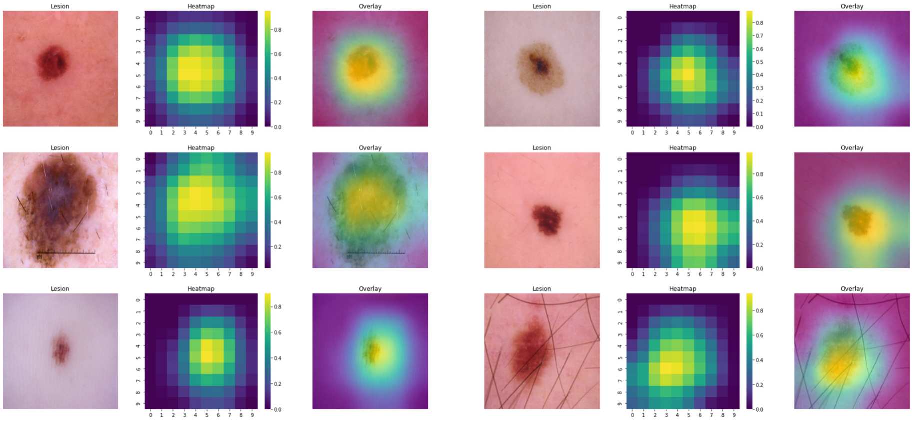

Fig. 16. GradCAM visualizations

6 Conclusion

In this paper, we have taken up the task of multi-class classification of skin lesions from dermatoscopic images in the HAM10000 dataset using deep convolutional neural networks, alleviating the need for complex feature engineering. We leveraged transfer learning by using pre-trained models and modified the Inception-ResNetV2 architecture to the required problem. We handled the class imbalance problem using data augmentation and weighted classes that appropriately compensate for the loss function. Overfitting has been abated by using a global average pooling layer, incorporating dropout layers, and applying kernel regularizer. Together with all of this, the model achieved an accuracy of 90.08% with an F1-score of 89%, outperforming the rest. Compared with the rest, ResNet50 also yielded satisfactory results, with its accuracy standing at 87%. However, the proposed approach poses its own drawbacks since the size of the model is quite large, lowering its feasibility to be used on mobile devices. The model performance can be further improved by expanding the dataset by collecting more images and balancing the samples per class.

References

1. The Skin Cancer Foundation. https://www.skincancer.org
2. World Cancer Research Fund International. https://www.wcrf.org/dietandcancer/skin-cancer-statistics/
3. World Health Organisation. https://www.who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer
4. ISIC 2018: International Skin Imaging Collaboration Data Archive
5. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161
6. Le, D.N.T., Hieu, X., Le, L.T. Ngo, H.T.: Transfer learning with class-weighted and focal loss function for automatic skin cancer classification." arXiv preprint arXiv:2009.05977 (2020)
7. Gupta, H., Bhatia, H., Giri, D., Saxena, R., Singh, R.: Comparison and Analysis of Skin Lesion on Pretrained Architectures (2020)
8. Chaturvedi, S.S., Gupta, K., Prasad, P.S.: Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using mobilenet. In: Hassanien, A.E., Bhatnagar, R., Darwish, A. (eds.) AMLTA 2020. AISC, vol. 1141, pp. 165–176. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-3383-9_15
9. Kassem, M.A., Hosny, K.M., Mohamed, M.F.: Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. IEEE Access **8**, 114822-114832 (2020)
10. Sagar, A., Jacob, D.: Convolutional neural networks for classifying melanoma images. bioRxiv (2020)
11. Mohamed, E.H., El-Behaidy, W.H.: Enhanced skin lesions classification using deep convolutional networks. In 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), pp. 180–188. IEEE (2019)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
13. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
14. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on Artificial Intelligence (2017)
15. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convo- lutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)