Advancing Dermatological Diagnosis with Multiclass Lesion Analysis

Enhancing Multiclass Skin Cancer Classification with a Fine-Tuned MobileNet Model

D. Sri Anjaneyam
Department of Information Technology
Institute of Aeronautical Engineering
Hyderabad, India
mr.anjaneyam@gmail.com

T. Yashaswini
Department of Information Technology
Institute of Aeronautical Engineering
Hyderabad, India
yashaswinithummanapally@gmail.com

D. Saiteja
Department of Information Technology
Institute of Aeronautical Engineering
Hyderabad, India
saitejachintu2003@gmail.com

***Abstract*—Skin cancer is the most common form of cancer globally, ranking as the 17th most prevalent cancer with 2-3 million non-melanoma and 132,000 melanoma cases worldwide. It arises due to the abnormal growth of skin cells, primarily caused by excessive exposure to ultraviolet rays. Early detection and accurate classification of skin lesions are crucial for effective treatment and improved patient outcomes. Given the challenges dermatologists face in accurately diagnosing skin cancer, there is an urgent need for an automated and efficient diagnostic system.**

This study presents an automated system for classifying skin lesions into seven different types of skin cancer using a Convolutional Neural Network (CNN) model, trained on the HAM10000 dataset, which includes 10,015 dermatoscopic images of various skin lesions. Our approach leverages the MobileNet architecture, renowned for its efficiency and effectiveness in image classification tasks. The MobileNet model, pre-trained on the extensive ImageNet dataset, is fine-tuned on the HAM10000 dataset to specialize in skin lesion classification by employing transfer learning. The model achieved an overall accuracy of 89.21% across the seven skin cancer classes, with top-2 and top-3 accuracies of 96.55% and 98.45%, respectively. The trained model demonstrated promising results, underscoring its potential as a reliable tool for assisting dermatologists in the early detection of skin cancer.

***Keywords*—Skin Cancer, CNN, MobileNet, HAM10000, Lesion Classification, ImageNet.**

I. Introduction

Skin cancer remains one of the most prevalent forms of cancer worldwide, with millions of new cases reported each year. Ultraviolet (UV) radiation exposure from the sun is a significant risk factor, leading to DNA damage in skin cells and resulting in mutations that can trigger cancerous growths. Early detection and treatment of skin cancer, particularly melanoma, are critical, as the prognosis is significantly better when the disease is caught at an early stage. However, the visual similarity between benign and malignant skin lesions poses a challenge even for experienced dermatologists, often leading to misdiagnosis and delayed treatment.

To address these challenges, advancements in artificial intelligence (AI) and deep learning have opened new avenues for improving skin cancer diagnosis. Convolutional Neural Networks (CNNs), a subset of deep learning, have shown remarkable efficacy in image recognition tasks, making them well-suited for medical image analysis. MobileNet, a lightweight CNN architecture, is particularly advantageous for this purpose due to its efficiency and ability to perform well on mobile and embedded devices. Unlike other models such as VGGNet and ResNet, which require significant computational resources, MobileNet is designed to optimize for both speed and accuracy. The model was initially pretrained on the extensive ImageNet dataset, which contains millions of images across thousands of categories, and then fine-tuned on the HAM10000 dataset, which includes images of various types of skin lesions. By leveraging the concept of transfer learning, MobileNet can effectively differentiate between various types of skin lesions with high accuracy.

Early and accurate detection of skin cancer, facilitated by MobileNet, can significantly improve patient outcomes. By reducing the likelihood of misdiagnosis and enabling timely treatment, MobileNet can help lower the mortality rate associated with melanoma and other skin cancers. The integration of MobileNet into clinical practice not only supports dermatologists in making better-informed decisions but also makes advanced diagnostic tools more accessible and practical for real-world medical applications, ultimately enhancing the overall effectiveness of skin cancer treatment.

II. Methodology

A. Dataset

The HAM10000 dataset was employed for training and validation in this study. HAM10000 is a benchmark dataset with over 50% of lesions confirmed by pathology. It consists of 10,015 dermoscopic images, categorized as follows:

(a) 6,705 Melanocytic nevi (NV) images
(b) 1,113 Melanoma (MEL) images
(c) 1,099 Benign keratosis-like lesions (BKL) images
(d) 514 Basal cell carcinoma (BCC) images
(e) 327 Actinic keratosis (AKIEC) images
(f) 142 Vascular lesions (VASC) images
(g) 115 Dermatofibroma (DF) images

Each image has a resolution of 600x450 pixels. Sample images of skin cancer types from HAM10000 are represented in Figure 1.

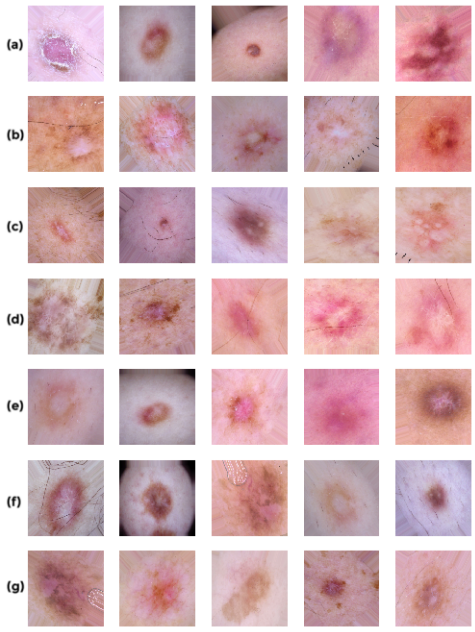

Fig. 1. Sample images of HAM10000 dataset for cancer types labeled above

B. Data Pre-processing

The pre-processing of skin lesion images was conducted using Keras ImageDataGenerator. The dataset contained 57 null entries for the 'Age' attribute, which were filled using the mean filling method. Dermoscopy images in the dataset were downscaled from their original resolution of 600x450 pixels to 224x224 pixels to make them compatible with the MobileNet model. The dataset, consisting of 10,015 images, was split into a training set and a validation set with 8,912 images and 1,103 images, respectively. To maintain authenticity in the validation process, images with no duplication in the training data were selected for the validation set. Additionally, data augmentation techniques, such as random rotations, horizontal and vertical flips, zoom-in and zoom-out transformations, and brightness and contrast adjustments, were applied using Keras ImageDataGenerator to increase the diversity of the training data and prevent overfitting.

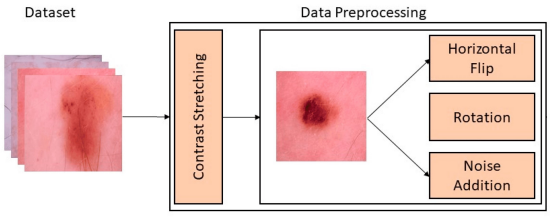

Fig. 2. Preprocessing of dataset

C. Model training

The training of the Convolutional Neural Network (CNN) model utilized the MobileNet architecture, leveraging its efficiency in image classification tasks. The MobileNet model was initialized with weights pre-trained on the ImageNet dataset, which includes millions of images across various categories. This pre-trained model served as the foundational network, and transfer learning techniques were employed to adapt it for the specific task of classifying skin lesions.

The HAM10000 dataset, comprising 10,015 dermoscopic images, was divided into a training set of 8,912 images and a validation set of 1,103 images. Each image was pre-processed to a resolution of 224x224 pixels to align with the input requirements of the MobileNet model.

During the training phase, the MobileNet architecture was fine-tuned on the HAM10000 dataset. The top layers of the pre-trained MobileNet model were replaced with a Global Average Pooling layer, a Dropout layer with a rate of 0.25 to mitigate overfitting, and a Dense layer with 7 units corresponding to the seven different skin cancer classes. This final layer used the softmax activation function to output class probabilities.

The model was compiled using the Adam optimizer with a learning rate of 0.01. The categorical cross-entropy loss function was employed to measure the model's performance, while metrics such as categorical accuracy, top-2 accuracy, and top-3 accuracy were monitored during training to evaluate classification performance.

The training process was conducted over 100 epochs, with early stopping and model checkpointing mechanisms implemented. Early stopping monitored the validation accuracy and halted training if no improvement was observed for a set number of epochs, while model checkpointing saved the best-performing model weights based on validation accuracy.

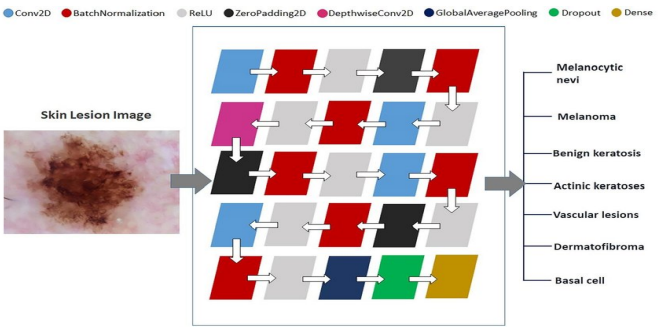

Fig. 3. MobileNet layers and processing

D. Evaluation metrics

The overall performance of the model was evaluated using several metrics: Accuracy, Micro Average Precision (MAP), Micro Average Recall (MAR), and Micro Average F1-score (MAF). The weighted averages for Recall, Precision, and F1-score were calculated using the following mathematical expressions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-score} = 2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right) \quad (4)$$

where,

TP: True Positives
TN: True Negatives
FP: False Positives
FN: False Negatives

III. Results

A. Hardware configurations

The calculations were performed on a Kaggle kernel with the following configurations:

- 4 CPU cores with 29GB RAM.
- 2 CPU cores with 16GB RAM (GPU P100).

B. Exploratory Data Analysis

The graph(Figure 4) depicts the distribution of skin lesion types within the dataset, showcasing the frequency of Melanocytic Nevi, Melanoma, Benign Keratosis-like Lesions, Basal Cell Carcinoma, Actinic Keratoses, Vascular Lesions, and Dermatofibroma.

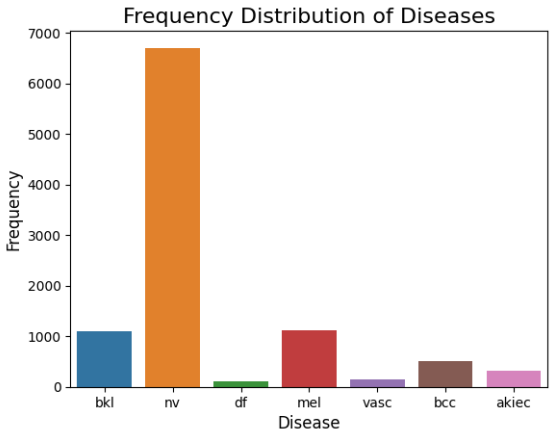

Fig. 4. Frequency distribution of diseases

The graph in Figure 5 illustrates the distribution of skin lesion locations by gender, providing valuable insights into the spatial prevalence across different anatomical regions.

The histogram in Figure 6 depicts the distribution of patient ages in the dataset, highlighting the demographic spread of individuals across various age groups affected by skin lesions.

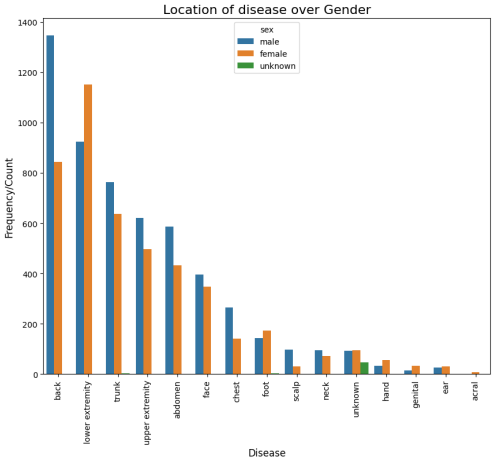

Fig. 5. Location of lesions

Fig. 6. Age of patients

C. Model Evaluation

The model was validated on 1103 unseen images from the validation set. We measured the micro and weighted averages for precision, recall, and f1-score to evaluate the MobileNet model's performance on these images. The results showed a Weighted Average of 89%, 89%, 89%, and a Micro Average of 76%, 67%, 70% for Precision, Recall, and F1-score, respectively. The MobileNet model achieved the highest precision, recall, and f1-score for Melanocytic Nevi. Table 1 presents the Multi-Class Classification Report, detailing the Micro and Weighted Averages for Precision, Recall, and F1-Score.

Table 2 represents the comparison of the current study with other related previous works. The majority of previous work is done on two or three classes, and their accuracies and recall vary between approximately 66 percent to 81 percent and 60 percent to 76 percent, respectively. In previous studies, classification accuracies using CNN models have varied significantly. For example, study [2] reported accuracies of 48.9% and 55.4% for nine classes. Another study [1] achieved a classification accuracy of 75.1% for ten classes using a Multi-track CNN. Additionally, [3] reported accuracies of 70%, 76%, 74%, and 67% for seven classes using InceptionResNetV2,

TABLE I
CLASSIFICATION REPORT

Classes	Precision	Recall	F1 Score
Actinic Keratosis	0.58	0.47 3	0.52
Basal Cell Carcinoma	0.68	0.86	0.76
Benign Keratosis	0.67	0.61	0.64
Dermatofibroma	1.00	0.50	0.67
Melanoma	0.47	0.46	0.46
Melanocytic Nevi	0.95	0.96	0.96
Vascular Lesions	1.00	0.85	0.92
Micro Average	0.76	0.67	0.70
Weighted Average	0.89	0.89	0.89

PNASNet-5-Large, SENet154, and InceptionV4, respectively. In recent study [4], accuracies of 83.15%, 91.36%, 95.84% were achieved using MobileNet for seven classes.

In our study, we achieved a categorical accuracy of 89.21%, a top-2 accuracy of 96.55%, a top-3 accuracy of 98.45%, and a recall of 89% using MobileNet. Our seven-class skin cancer classification method outperformed these previously proposed computer-aided diagnosis systems in terms of both accuracy and recall. Furthermore, our method is more efficient, benefiting from the faster processing capabilities and lightweight architecture of MobileNet.

TABLE II
COMPARISON RESULTS OF THE CURRENT STUDY WITH PREVIOUS RELATED WORK

Source	Year	Classifier	No. of Classes	Accuracy %
[2]	2016	Multi-tract CNN	Ten	75.1
[1]	2017	CNN	Three	69.4
		CNN-PA		72.1
		CNN		48.9
		CNN-PA	Nine	55.4
[35]	2019	InceptionResnetV2	Seven	70.0
		PNASNet-5-Large		76.0
		SENet154		74.0
		InceptionV4		67.0
3	2019	MobileNet	Seven	83.15(cat)
				91.36 (top2)
				95.84 (top3)
	2024	Current Study	Seven	89.21(cat)
				96.55 (top2)
				98.45 (top3)

D. Confusion Matrix

The confusion matrix for our model, evaluated for seven classes without normalization, shows the performance comparison between true and predicted labels for each image in the validation set. The model achieved the best results for Melanocytic Nevi, accurately predicting 850 out of 883 images. Basal Cell Carcinoma was correctly identified for 30 out of 35 images. Benign Keratosis proved to be challenging, with only 54 correct predictions out of 88 images, indicating potential confusion with other classes. The overall results highlight areas for improvement, particularly in the accurate classification of Benign Keratosis and Melanoma. Figure 7 shows the confusion matrix.

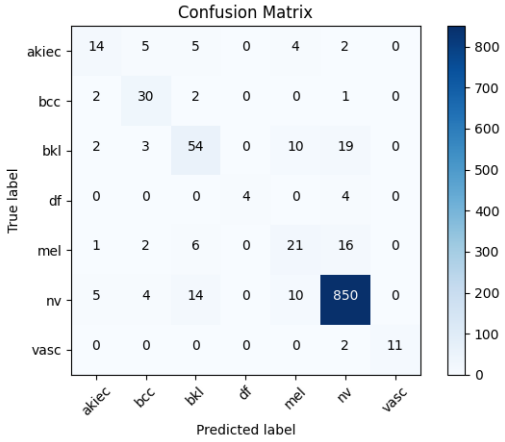

Fig. 7. Confusion matrix.

E. Loss and accuracy curves

The training process of our model was monitored using loss and accuracy curves, which provide a visual representation of the model's performance over time. The loss curve (Fig.8) shows the change in the loss function's value during training and validation, indicating how well the model is learning the underlying patterns in the data. A steadily decreasing loss curve signifies that the model is improving its predictions.

Similarly, the accuracy curve (Fig. 8) tracks the accuracy of the model on the training and validation datasets. An increasing accuracy curve suggests that the model's predictive performance is getting better. Ideally, the training and validation curves should converge, indicating that the model is not overfitting or underfitting.

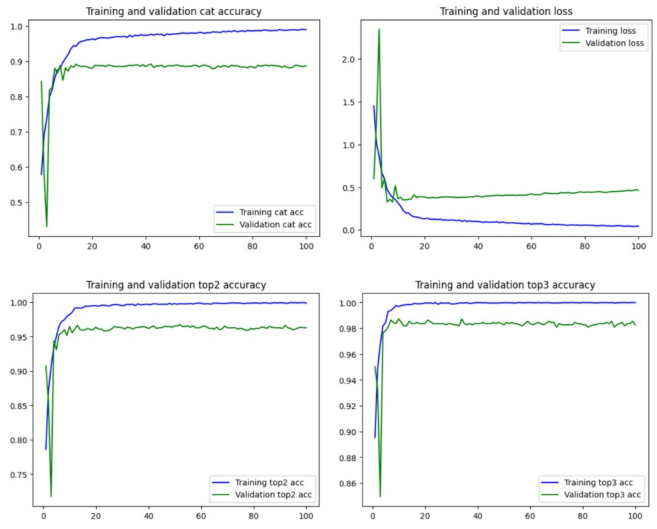

Fig. 8. (a) Categorical accuracy, (b) Loss, (c) Top2 accuracy, (d) Top3 accuracy

IV. Conclusion

In this project, we developed an automated system for skin lesion classification using a Convolutional Neural Network (CNN) model based on the MobileNet architecture. By leveraging the HAM10000 dataset, which comprises 10,015 dermoscopic images, our model demonstrated high accuracy in differentiating between seven types of skin cancer. The MobileNet model, pre-trained on the ImageNet dataset and fine-tuned on our specific task, achieved a categorical accuracy of 89.21%, top-2 accuracy of 96.55%, and top-3 accuracy of 98.45%. These results outperform previous studies, highlighting the effectiveness of our approach in early and accurate skin cancer detection. The confusion matrix analysis showed that Melanocytic Nevi were most accurately identified, while there were challenges in distinguishing Benign Keratosis. Overall, this study underscores the potential of using lightweight CNN architectures like MobileNet for real-time and efficient skin cancer diagnosis, providing significant support to dermatologists in clinical settings.

References

[1] Radiation: Ultraviolet (UV) radiation and skin cancer, World Health Organization; https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer

[2] World Cancer Report 2014, International Agency for Research on Cancer, 2014; https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014

[3] Skin Cancer, National Cancer Institute; https://www.cancer.gov/types/skin

[4] Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the US Population, 2012; https://doi.org/10.1001/jamadermatol.2015.1187

[5] World Health Organization, Skin Cancers, WHO-2017

[6] Prevention and Early Detection Strategies for Melanoma and Skin Cancer by Howard K. Koh, MD; Alan C. Geller; Donald R. Miller; https://doi.org/10.1001/archderm.1996.03890280098014

[7] Parkin DM, Mesher D, Sasieni P. Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010. Br. J. Cancer. 2011;105(2): S66-S69. https://doi.org/10.1038/bjc.2011.486

[8] Canadian Cancer Society. Risk factors for melanoma skin cancer, https://www.cancer.org/cancer/melanoma-skin-cancer/causes-risks-prevention/risk-factors.html; 2018

[9] Cancer facts & figures 2016. Atlanta, American Cancer Society 2016. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2016.html; 2016

[10] Neville JA, Welch E, Leffell DJ. Management of nonmelanoma skin cancer in 2007. Nat. Clin. Pract. Oncol. 2007; 4(8):462–469. https://doi.org/10.1038/ncponc0883

[11] Binder M, Schwarz M, Winkler A et al. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. Arch. Dermatol. 1995; 131(3):286–291. https://doi.org/10.1001/archderm.1995.01690150050011

[12] Argenziano G et al. Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. J. Am. Acad. Dermatol. 2003; 48(5):679–693. https://doi.org/10.1067/mjd.2003.281

[13] Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet. Oncol. 2002 Mar;3(3):159-65. https://doi.org/10.1016/S1470-2045(02)00679-4

[14] Mnih V et al. Human-level control through deep reinforcement learning. Nature. 2015; 518:529-533. https://doi.org/10.1038/nature14236

[15] Russakovsky O et al. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 2015; 115(3):211–252. https://doi.org/10.1007/s11263-015-0816-y

[16] Esteva A et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542(7639):115–118. https://doi.org/10.1038/nature21056

[17] Stern RS. Prevalence of a History of Skin Cancer in 2007. Arch. Dermatol. 2010; 146(3):279–282. https://doi.org/10.1001/archdermatol.2010.4

[18] Masood A, Al-Jumaily AA. Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. Int. J. Biomed. Imaging; 2013:323268. http://dx.doi.org/10.1155/2013/323268

[19] Khosla A et al. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 2015; 115(3);211–252. https://doi.org/10.1007/s11263-015-0816-y

[20] M. A. A. Milton, "Automated Skin Lesion Classification Using Ensemble of Deep Neural Networks in ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection Challenge," Jan. 2019.

[21] Howard AG et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv Prepr. arXiv https://arxiv.org/abs/1704.04861. 2017.

[22] Image Preprocessing - Keras Documentation. Keras, Available: https://keras.io/preprocessing/image/; 2019

[23] Kaggle: Your Home for Data Science. Available: https://www.kaggle.com/

[24] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-sources dermatoscopic images of common pigmented skin lesions. Sci. Data. 2018; 5:180161. https://doi.org/10.1038/sdata.2018.161

[25] Pandas. Working with missing data — pandas 0.22.0 documentation, Available: https://pandas.pydata.org/pandas-docs/stable/user_guide/missing_data.html;2019. [Accessed: 31-Mar-2019].