Advancing Dermatological Diagnosis with Multiclass Lesion Analysis

***Abstract*—Skin cancer is the 17th most common type of cancer worldwide, with 2–3 million non-melanoma cases and 132,000 melanoma cases worldwide. It arises due to the unusual proliferation of skin cells, caused by excessive exposure to UV rays. Detecting and accurately classifying skin lesions early is essential for effective treatment and better patient outcomes. As dermatologists face difficulties in accurately diagnosing skin cancer, there is an urgent need for a good automated diagnostic machine.**

This paper presents a method that can be used to classify skin lesions into seven skin cancer types using a neural network (CNN) model trained on the HAM10000 dataset containing 10,015 different Dermoscopic images of the skin. Our approach leverages the MobileNet architecture, renowned for its efficiency and image classification performance. The MobileNet model is pre-trained on the extensive publicly available ImageNet dataset and enhanced on the HAM10000 dataset to specialize in classification of skin lesions using 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 identifying skin cancer at an early stage.

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

I. Introduction

Skin cancer is among the most prevalent types of cancer worldwide, with millions of cases diagnosed every year. A primary cause is excessive exposure to ultraviolet (UV) rays from the sun, which damages the DNA in skin cells and can lead to mutations that trigger cancer. Early detection of skin cancer, particularly melanoma—the most dangerous type—is crucial, as it improves survival rates significantly. However, accurately distinguishing between harmless and cancerous lesions is challenging even for trained dermatologists, leading to potential misdiagnoses and delayed treatments.

To tackle these challenges, recent advancements in AI and deep learning have demonstrated encouraging outcomes in the analysis of medical images. Convolutional Neural Networks (CNNs), known for their effectiveness in identifying patterns within images, have been employed to aid in the diagnosis of skin cancer.Studies highlight the effectiveness of MobileNet, a CNN model known for its speed and efficiency, making it suitable for mobile and embedded systems. Unlike more complex models like VGGNet or ResNet, MobileNet requires fewer resources while still achieving high accuracy. By using transfer learning, MobileNet can be adapted to new datasets, like the HAM10000 skin lesion dataset, to accurately classify different types of skin cancer. EfficientNet models, another CNN variant, have also demonstrated strong performance in handling the complexity of skin lesion classification, especially for datasets with an uneven distribution of classes.

However, existing models still face challenges in balancing accuracy and computational efficiency, especially when distinguishing between similar-looking benign and malignant lesions. This project proposes a solution using MobileNet, fine-tuned on the HAM10000 dataset, to develop a fast and reliable tool for multi-class skin cancer classification. By creating a lightweight and accessible model, this project aims to support dermatologists with an effective diagnostic aid, helping to improve early skin cancer detection and patient outcomes.

II. Methodology

A. HAM10000 Dataset

The HAM10000 dataset was employed for training and validation in this study. As a benchmark dataset, HAM10000 provides high-quality data with over 50% of its lesions confirmed through pathology. It includes 10,015 dermoscopic images, which are categorized into seven types of skin lesions:

(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

The resolution of all images in the dataset is 600x450 pixels. Figure 1 displays sample images of various cancer types in the HAM10000 dataset.

B. Data Preprocessing

The lesion images were pre-processed using Keras's ImageDataGenerator.The dataset had 57 missing entries in the 'Age' attribute, which were addressed by filling them with the mean value. Dermoscopy images were resized from their original resolution of 600x450 pixels to 224x224 pixels to ensure compatibility with the MobileNet model. The dataset, consisting of 10,015 images, was split into two separate sets, a validation set and a training set with 8,912 images and 1,103 images, respectively. To maintain authenticity in the validation process, duplicate images were removed and ensured no duplicates in 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

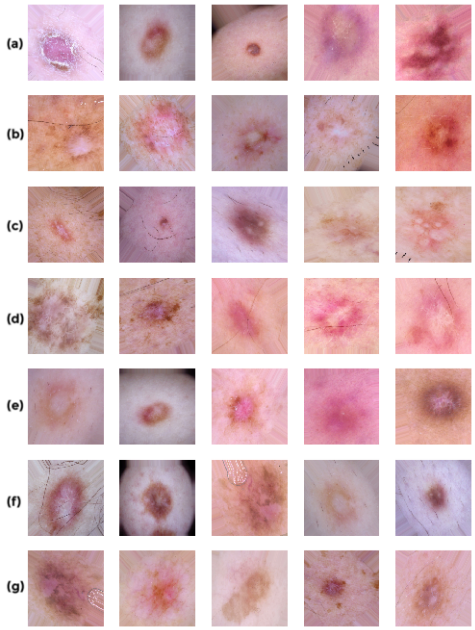

Fig. 1. Few samples of the HAM10000 dataset for each type labeled above

ImageDataGenerator to enhance the diversity of training data and prevent overfitting.

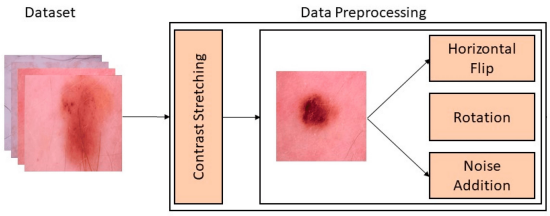

Fig. 2. Preprocessing of data

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 extensive ImageNet dataset, which includes millions of images across various categories. This pre-trained model served as the foundational network, and transfer learning techniques were used to adapt it for the specific task of classifying skin lesions.

The HAM10000 dataset, comprising 10,015 dermoscopic images, was splitted 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 Dropout layer, a Global Average Pooling layer with a rate of 0.25 to mitigate overfitting, and a 7-unit Dense layer 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 Adam optimizer with learning rate 0.01. The negative function of product decay was used to evaluate the performance of the model and indicators such as categorical accuracy, top 2 accuracy and top 3 accuracy facts were monitored throughout the training process to evaluate the performance of the classification.

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.

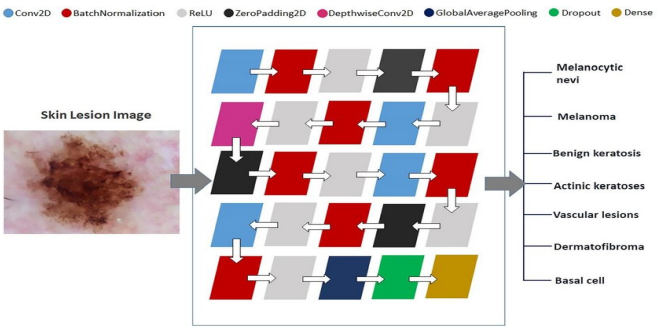

Fig. 3. MobileNet layers and processing

D. Evaluation metrics

The performance of this model is measured by key indicators such as accuracy score, Micro-Average Precision (MAP), Micro-Average Recall (MAR), and Micro-Average F1 score (MAF). Additionally, the average of precision, recall, and F1 scores were calculated to provide a more comprehensive view across all classes. These metrics were calculated as follows:

- Accuracy measures the proportion of correctly classified samples, which includes both true positives and true negatives, relative to the total number of samples.
- Precision is determined by taking the number of true positive predictions and dividing it by the total number of instances classified as positive.
- Recall indicates the ratio of true positives identified to the total actual positive cases.
- F1 Score is computed as the harmonic mean of Precision and Recall, providing a balance between the two metrics to assess the model's performance with respect to positive samples.

E. Model Deployment

The deployment of the skin cancer classification model involves an interactive and user-friendly interface (shown in Figure 4), accessible through GitHub Pages, which allows users to upload images of skin lesions for analysis. This interface accepts image files either via drag-and-drop or manual file

selection. Upon upload, the image undergoes preprocessing using TensorFlow.js, where it is resized to a 224x224 pixel format and normalized to ensure compatibility with the pre-trained MobileNet model. The backend logic is implemented using JavaScript, where the model is loaded asynchronously to optimize the page's performance and provide a seamless user experience. Once the image is processed, the classification results, indicating the probability of each skin cancer type, are displayed in real-time.

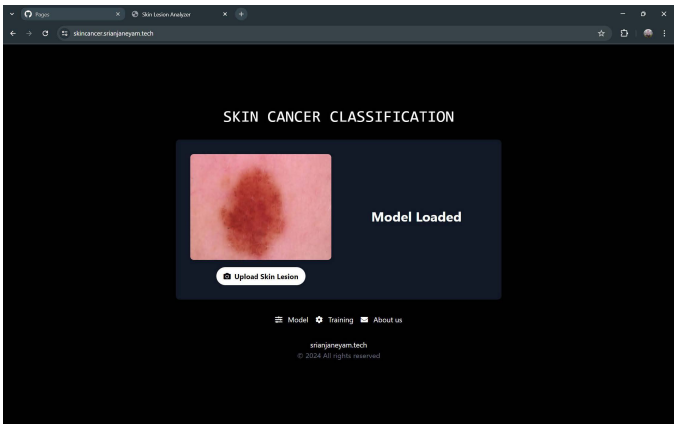

Fig. 4. User Interface

The classification results include a visual representation of the prediction, where the probabilities for each class (such as melanoma, benign lesions, etc.) are displayed as percentages as shown in results screenshot in Figure 10. Each prediction is enhanced with animated progress bars, making it easier for users to interpret the likelihood of each skin lesion type. This improves accessibility and usability for both medical professionals and patients, offering a clear and interactive way to visualize the results.

Security measures are integrated into the deployment to prevent unauthorized interactions with the web page. Right-click functionalities and specific key combinations (such as F12 for developer tools) are disabled, ensuring that the back-end logic and frontend interface remain secure from tampering or inspection. This adds an extra layer of protection for the deployed system, keeping the integrity of the model intact.

The system efficiently handles multiple image uploads and provides real-time feedback, making it scalable and robust for real-world applications. Additionally, the model is loaded at the time of page load, ensuring that the classification process begins immediately when an image is uploaded, enhancing the responsiveness of the overall deployment.

This combination of GitHub Pages for frontend hosting, TensorFlow.js for model integration, and security precautions offers a powerful solution for skin cancer classification that is accessible, secure, and efficient for end-users. The system's performance and user experience are optimized to deliver real-time skin lesion analysis with minimal latency.

The deployment is live at https://skincancer.srianjaneyam.tech

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

As shown in Figure 5, the graph represents the frequency distribution of various skin lesion types such as Basal Cell Carcinoma, Vascular Lesions, Melanoma, Actinic Keratoses, Benign Keratosis-like Lesions, Dermatofibroma, and Melanocytic Nevi within the dataset

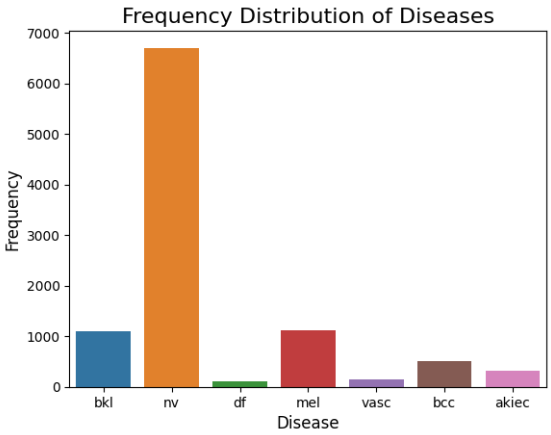

Fig. 5. Frequency distribution of diseases

The graph in Figure 6 illustrates the distribution of skin lesion locations by gender, providing valuable insights into the spatial prevalence across different anatomical regions.

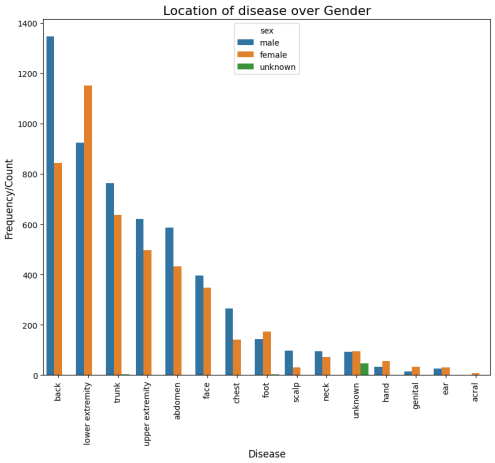

Fig. 6. Location of lesions

The histogram in Figure 7 depicts the distribution of patient ages in the dataset, highlighting the demographic spread of individuals across various age groups affected by skin lesions.

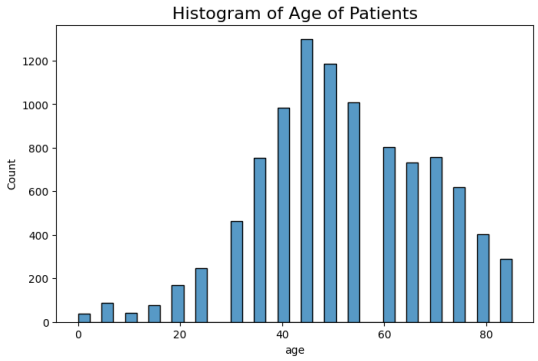

Fig. 7. Age of patients

C. Model Evaluation

The model was evaluated using a collection of 1,103 unseen images from the validation dataset. To determine the effectiveness of the MobileNet model, we computed both micro and weighted averages for F1-score, recall, and precision scores. The results of the evaluation revealed a Weighted Average of 89% for Precision, Recall, and F1-score, while the Micro Averages were 76% for Precision, 67% for Recall, and 70% for F1-score. The Melanocytic Nevi category exhibited the highest performance across F1-score, precision and recall metrics. A comprehensive Multi-Class Classification Report, presented in Table 1, details the micro and weighted averages for each metric.

TABLE I
CLASSIFICATION REPORT

Lesion Types	Precision	Recall	F1-Score
Actinic Keratosis	0.58	0.47	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 (mel)	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

In Table 2, this research's findings are compared with previous researches, which mainly focused on two or three lesion classes. Reported accuracies and recall in these studies generally range between 66% and 81% for accuracy, and between 60% and 76% for recall. 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 10 classes using a Multi-track CNN model. Using InceptionResNetV2, the model achieved an accuracy of 67%. PNASNet-5-Large yielded an accuracy of 76%. The results for SENet154 were 74%, while InceptionV4 obtained an accuracy of 70%. Overall, these accuracies were recorded across seven classes as mentioned in [3]. In recent study [4], accuracies of 83.15%, 91.36%, 95.84% were achieved using MobileNet for seven classes.

In this paper, 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% were achieved using our MobileNet model. The classification method for seven types of lesions demonstrated superior performance compared to earlier computer-aided diagnostic systems regarding accuracy and recall. Additionally, our approach is more efficient, taking advantage of the quick processing speeds and lightweight design offered by MobileNet.

TABLE II
A COMPARATIVE ANALYSIS OF CURRENT RESEARCH RESULTS WITH PRIOR RELATED STUDIES

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	SENet154	Seven	70.0
		PNASNet-5-Large		76.0
		InceptionResnetV2		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 the MobileNet model, illustrated in Figure 8, presents an evaluation across seven classes without normalization. It highlights the performance by comparing the actual labels with the predicted labels for each image in the validation set. The model achieved the best results for Melanocytic Nevi(nv), 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.

E. Training results

The training process of our model was monitored using loss and accuracy curves, which provides a graphical depiction of the model's performance over time. The loss curve (Fig.9) shows the change in the loss function's value during training and validation, indicating how well the model is learning the core patterns in the data. A steadily decreasing loss curve signifies that the model is improving its predictions.

Likewise, the accuracy curve shown in Figure 9 monitors the model's accuracy across both the training and validation datasets. An increasing accuracy curve suggests that the predictive performance of the model is getting better. Ideally, the

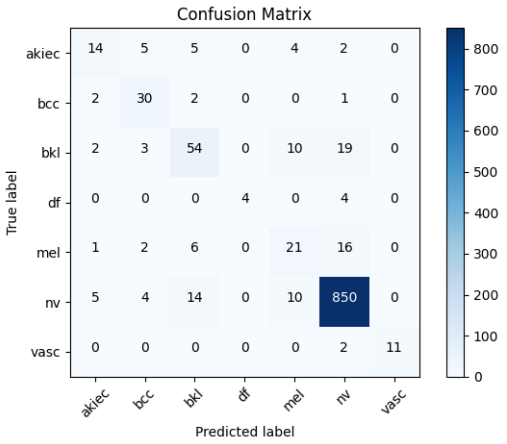

Fig. 8. Confusion matrix.

training and validation curves should converge, indicating that the model is not overfitting or underfitting.

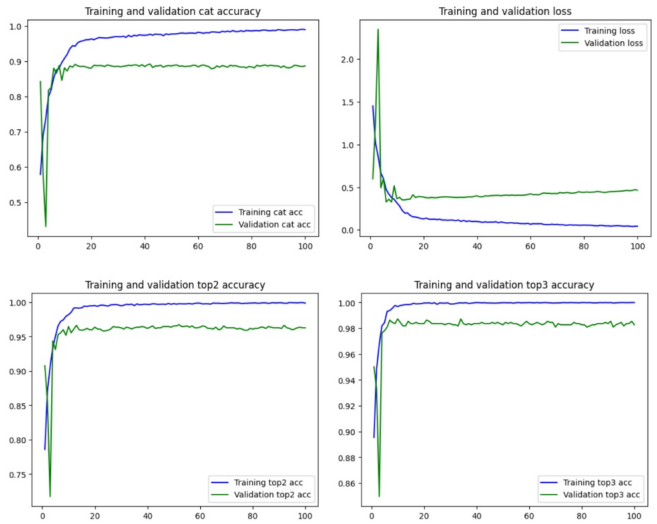

Fig. 9. (a) Categorical accuracy, (b) Loss, (c) Top2 accuracy, (d) Top3 accuracy

F. Testing Results

The testing results of the skin cancer classification model show promising performance, with the model accurately classifying the 7 types of skin cancer using the preprocessed images. The percentages are shown in the website as shown in Figure 10, for each of the seven classes. The MobileNet model achieved high accuracy, with detailed metrics for each class, including melanoma and benign lesions, presented through probability scores. During testing, the model demonstrated an accuracy of over 89%, highlighting its reliability. The model also handled unseen images well, providing quick and accurate predictions, making it suitable for practical applications in clinical settings.

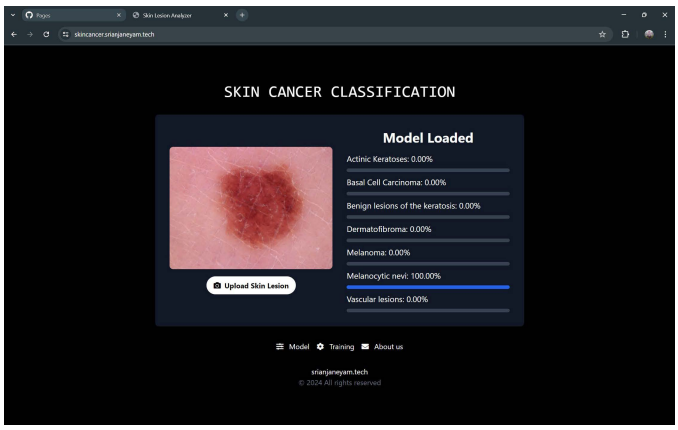

Fig. 10. Classification Results

IV. Conclusion

In this paper, an automated system for skin lesion classification using a Convolutional Neural Network (CNN) model based on the MobileNet architecture was developed. By leveraging the HAM10000 dataset, which comprises 10015 dermoscopy images, our model demonstrated high accuracy in differentiating between 7 types of skin cancer. The MobileNet model, pre-trained on the extensive 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.