Skin Vision: A Deep Learning Approach for Skin Tone Classification

Dasuni Kawya
Dept. of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20197@eng.pdn.ac.lk

Ishan Thathsara
Dept. of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20211@eng.pdn.ac.lk

Imesh Malinga
Dept. of Computer Engineering
University of Peradeniya
Peradeniya, Sri Lanka
e20242@eng.pdn.ac.lk

***Abstract*—The aim of our project is to detect the skin tone of a person from an image. Cosmetology, dermatology and AL fairness are important applications, and they bring us the motivation to proceed with this project. Instead of going from scratch, transfer learning can give more accurate, lightweight solutions for these kinds of problems. MobileNetV2, which is a convolutional neural network (CNN) that's designed to run on mobile devices and used for image classification tasks like facial recognition and tagging photos, has used to feature extraction. The model takes the images as inputs after preprocessing. The defined model has trained on Kaggel Skin Tone Dataset. The testing can be done by using the images already existing in the validation set. After training 20 epochs, 86% accuracy level has achieved. Depending on the texture, the image will be classified into dark, light, mid-dark and mid-light. Model provides reliable classification but requires further tuning for real-world applications.**

***Index Terms*—Skin detection, deep learning, MobileNetV2, image classification**

I. Introduction

In many image processing and computer vision jobs that involves finding skin pixels or regions, skin detection is a pre-processing phase. Traditional methods use parametric models for skin pixel detection. In some color spaces, nonparametric models or skin cluster specified regions are used to detect skin tones. There are also other systems that recognize human shape features (hands, faces, and bodies) before looking for skin pixels, but most of the work looks for skin pixels first. Recently, a graph-based method was described in which the image is represented by a multilayer network, and the skin probability is then propagated over the graph. There are also neural network approaches that use the auto-encoder, such as adaptive neural networks, self-organizing maps, and deep-learning based methods. Despite the numerous algorithms, skin identification remains a difficult challenge due to a variety of factors such as changes in illumination, race and makeup skin color differences, and skin-like backdrops. In this study, we suggest a way to leverage pre-defined model MobileNetV2 further to perform well on a variety of picture for skin identification. In this approach, first we pre-process the images using normalization, augmentation and resizing for consistency. MobileNetV2's convolutional layers are used for feature learning. For the evaluation of the model, accuracy is analyzed. Other than that, precision-recall and confusion matrices are used.

II. Objectives

Human skin detection plays a key role in human-computer interactions. As we can see that there are so many new technological advancements happening such as biometric authentication in our smart phones for face detection, hand gesture recognition is a modern way of human computer interaction i.e., we can control our system by showing our hands in front of webcam and hand gesture recognition can be useful for all kinds of people. Based upon this idea of skin detection this paper is presented. This paper provides a detailed explanation of the procedure and methodologies for human skin detection or recognition.

III. Scope

For reliable or better human skin detection, our model is essential to have the ability to adapt to different human skin colors and lighting conditions. Even though there are methods to detect different skin colors, they are prone to false skin detection and are not able to cope with variety of human colors. In our approach, we consider more about the improvement of accuracy despite wide variety in ethnicity or illumination.

IV. Background

A. Previous Approaches

Previous approaches to skin tone classification relied on handcrafted features such as RGB histograms, Local Binary Patterns (LBP), and color spaces, but these methods struggled with lighting variations and real-world diversity. With the rise of deep learning, models like ResNet, VGG, and MobileNet have shown significant improvements in object detection and classification. Among them, MobileNetV2 stands out for its efficiency in edge computing and mobile applications. Being pretrained on ImageNet, it enables transfer learning, allowing the model to leverage existing feature extraction capabilities while adapting to skin tone classification tasks.

B. Existing System

We have gone through a model that uses MobileNetV2 as a feature extractor, leveraging its pretrained layers to identify patterns in skin tone images. The model processes input images of 224×224×3 dimensions and passes them through the convolutional layers of MobileNetV2. These layers remain frozen to retain their learned feature representations, such as edges, textures, and shapes. In that model, after feature extraction, the output is passed through a Global Average Pooling layer, fully connected Dense layer with 1024 neurons follows and finally a Softmax layer generates seven output probabilities, corresponding to different skin tone categories. This system was effective in distinguishing between multiple skin tone variations; however, it faced challenges in accurately differentiating similar skin tones due to a 7-class classification approach. Sometimes it results in overlapping categories, leading to misclassification.

V. Related Work

A. Proposed Approach

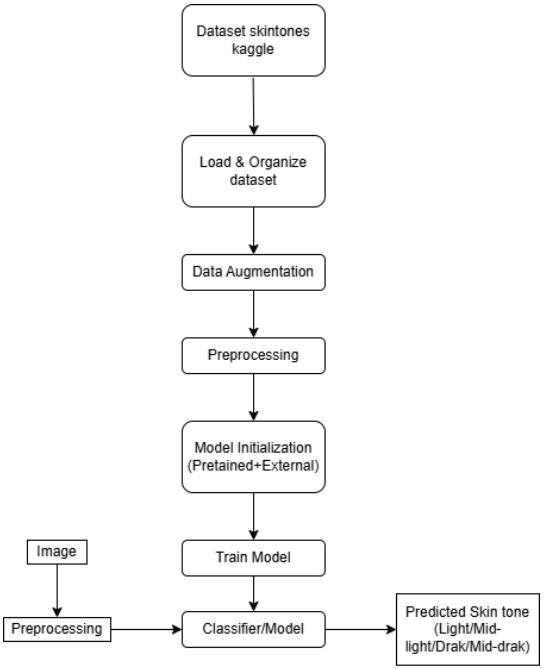

Fig. 1. High level architecture

B. Experiment

1) Experimental Setup: For our experiments, we used the Kaggle Skin Tone Dataset, which contains 40,000+ images categorized into four skin tone classes: Dark, Light, Mid-dark and Mid-light. We organized the dataset by splitting it into 80% training and 20% validation to ensure a balanced evaluation.

2) Preprocessing Steps: In addition to the existing system, we conducted experiments to optimize the dataset preprocessing and augmentation techniques to improve model generalization. The original approach used data augmentation parameters such as rotation (20°), zoom (20%), width and height shift (20%), and horizontal flipping, which helped in making the model more robust to lighting variations and slight transformations. However, to further enhance diversity and prevent overfitting, we modified several augmentation parameters. Our updated augmentation approach increased rotation range to 30° and width shifts to 30%, allowing the model to better adapt to images captured from different perspectives. We also introduced brightness adjustments (ranging from 0.7 to 1.2) to help the model recognize skin tones under varying lighting conditions. Additionally, we applied shear transformations (0.2) to introduce minor distortions that help improve feature extraction and set the fill mode to nearest to handle pixels outside the input boundaries. These modifications make the model more resilient to variations in real-world scenarios, ensuring improved classification accuracy across diverse datasets.

3) Model Architecture: Our model is built using MobileNetV2, a lightweight yet powerful deep learning architecture, leveraging transfer learning to extract meaningful features from skin tone images. MobileNetV2, pretrained on ImageNet, provides an efficient base for classification by retaining fundamental image features such as edges, textures, and shapes.

a) Base Model: We initialize MobileNetV2 with pretrained ImageNet weights, removing the top classification layer since it was originally trained for 1,000 classes (e.g., animals, objects). Instead, we freeze its convolutional layers to retain low-level feature extraction capabilities, making it well-suited for skin tone classification.

$$\begin{aligned}\text{Model} = \text{Input}(224, 224, 3)\\ \rightarrow \text{MobileNetV2 Layers (Frozen)}\\ \rightarrow \text{Custom Layers}\\ \rightarrow \text{Output}\end{aligned} \tag{1}$$

b) Modifications & Fine-Tuning: In the previous model, all MobileNetV2 layers were completely frozen, preventing further learning beyond its pretrained knowledge. In our approach, we experimented by unfreezing the last 10,5 and 3 layers, but those scenarios it leads to overfitting. At last, we decided to unfreeze the last layer to allow fine-tuning for skin tone-specific patterns.

c) Custom Layers:

$$\begin{aligned}\text{Model} = \text{Input}(224, 224, 3)\\ \rightarrow \text{MobileNetV2 Layers (Frozen)}\\ \rightarrow \text{GlobalAveragePooling2D}\\ \rightarrow \text{Dense}(512, \text{ReLU})\\ \rightarrow \text{Dropout}(0.2)\\ \rightarrow \text{Dense}(4, \text{Softmax})\end{aligned} \tag{2}$$

$$\rightarrow \text{Output} \tag{3}$$

The custom layers in our model play a crucial role in refining the extracted features for accurate skin tone classification. After passing through the MobileNetV2 base model, the feature maps are processed using Global Average Pooling (GAP), which reduces the spatial dimensions while preserving essential information. Additionally, we reduced the Dense layer size from 1024 to 512 neurons and introduced Dropout (0.2) to prevent overfitting by randomly "turning off" some neurons during training. And it helps regularize the model by preventing from memorizing training data. Finally, a Softmax activation layer with 4 outputs assigns probability scores to the four skin tone categories: Dark, Light, Mid-Dark, and Mid-Light.

d) Loss Function & Optimizer: To train the model, we use categorical cross-entropy loss, which measures the difference between the predicted and actual class probabilities. The Adam optimizer, with a learning rate of 0.001 ensures efficient weight updates for faster convergence. Additionally, to evaluate the model's performance, we track multiple metrics, including precision, recall, AUC (Area Under the Curve), and F1-score, which provide a comprehensive assessment of classification accuracy and model reliability.

4) Training Process: The training process begins with a forward pass, where input images are fed into the MobileNetV2 base model, followed by our custom layers for classification as above. The model processes the images and generates probability scores for each of the four skin tone categories. During training, the predicted labels are compared with the actual labels using the categorical cross-entropy loss function, which measures the error in classification. Since the MobileNetV2 base layers remain frozen, only the weights of the newly added layers are updated, ensuring that the model focuses on learning skin tone-specific features rather than relearning basic image features like edges and textures. To improve training efficiency and prevent overfitting, we implement several optimization strategies. Early stopping monitors validation loss and stops training if it does not improve for three consecutive epochs, ensuring that the model does not over-train on the dataset. Additionally, we use a learning rate scheduler to gradually reduce the learning rate when improvements plateau, allowing the model to fine-tune its learning process. A model checkpoint saves the best version of the model by tracking validation loss, ensuring that the optimal model is retained for evaluation. The model is trained for 40 epochs using augmented training data and validated on unseen images to measure accuracy, precision, and recall.

VI. Results & Evaluation

A. Quantitative Evaluation

The model's performance was evaluated using accuracy, precision, recall, AUC, and F1-score. After training for 40 epochs, the model achieved a validation accuracy of 86.06%, indicating its effectiveness in distinguishing between skin tone categories.

Validation Loss: 0.3399
Validation Accuracy: 86.06%
Validation Precision: 0.8631
Validation Recall: 0.8558
Validation AUC: 0.9792
Validation F1-Score: 0.8594

1) Accuracy: The training accuracy steadily increases and reaches 0.88 by the final epoch. The validation accuracy follows a similar trend, ending close to 0.86. The small gap between training and validation accuracy suggests minimal overfitting, indicating a well-balanced model.

2) Loss: The training loss decreases consistently, showing that the model is learning effectively. The validation loss also decreases but exhibits slight fluctuations in the middle epochs before stabilizing. This suggests minor variance issues, which could be mitigated with further regularization or fine-tuning.

3) Precision: Both training and validation precision improve over time, reaching approximately 0.88 and 0.86, respectively. The absence of major gaps between these values indicates that the model correctly identifies positive samples with a high level of confidence.

4) Recall: Training recall reaches 0.87, and validation recall closely follows the same trend.This suggests that the model is not sacrificing recall to boost precision, maintaining a good balance.

B. Confusion Matrix

A confusion matrix was generated to visualize classification performance and identify misclassifications between skin tone categories. The heatmap representation of the matrix helps in analyzing the areas where the model struggled, particularly between similar skin tones like light and mid-light.

C. Qualitative Evaluation

To further assess performance, sample images have been analyzed where the model made correct and incorrect predictions. The correctly classified images demonstrated the model's capability in distinguishing skin tones under varying lighting conditions. However, errors were observed in cases where lighting inconsistencies caused overlap between similar skin tones. These findings suggest that further dataset enhancement and fine-tuning may improve classification accuracy.

D. Performance Visualization

Training and validation accuracy, loss, precision, recall, AUC, and F1-score over epochs have been plotted. The accuracy and loss curves indicate stable training progress, while the precision and recall metrics provide a deeper understanding of class-wise performance. The AUC and F1-score highlight the model's overall classification effectiveness.

E. Model Evaluation and Predictions

The model was evaluated on the validation dataset, and key performance metrics such as validation loss, accuracy, precision, recall, AUC, and F1-score were recorded. Additionally, predictions were made on the validation set, and a classification report was generated to analyze the model's

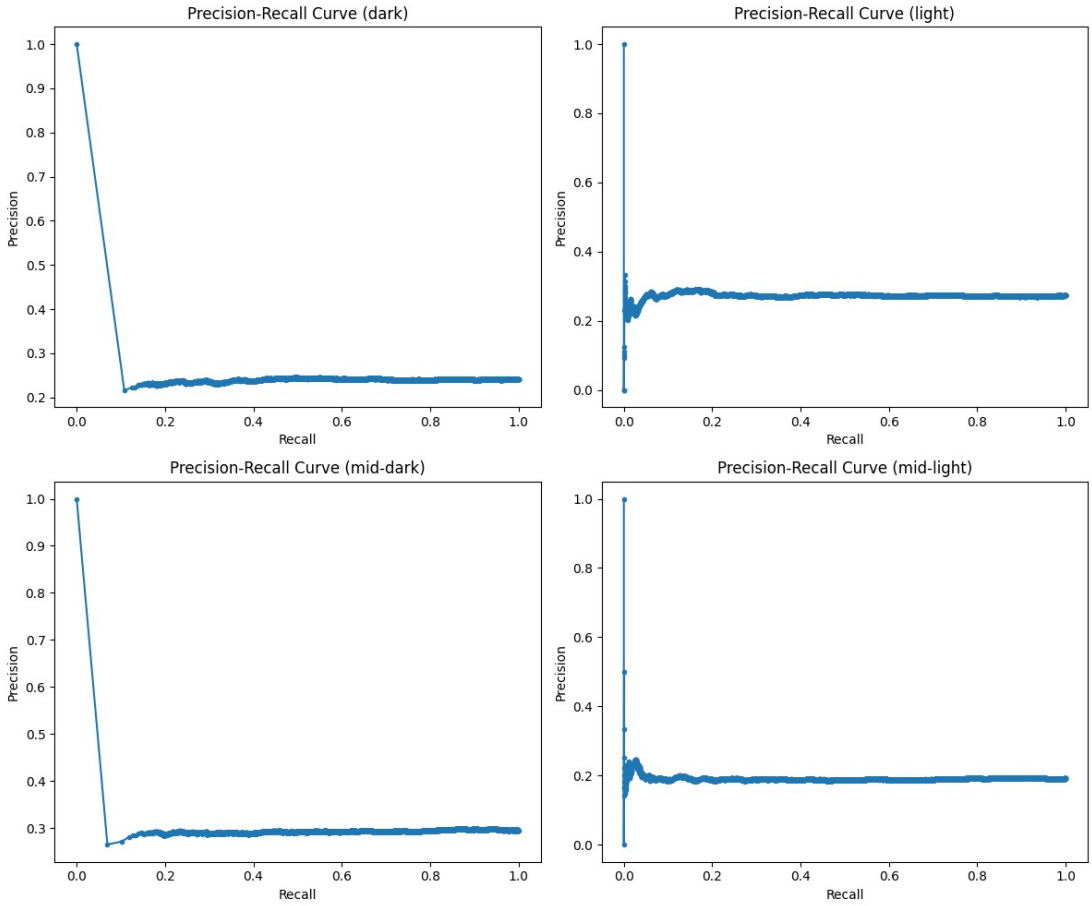

Fig. 2. Precision-Recall Curves

performance across all four skin tone categories. The report indicates that certain skin tones had higher misclassification rates, highlighting areas for further improvement.

TABLE I
MODEL EVALUATION METRICS

Class	Precision	Recall	F1-Score	Support
Dark	0.24	0.23	0.24	1728
Light	0.27	0.34	0.30	1954
Mid-Dark	0.29	0.29	0.29	2115
Mid-Light	0.20	0.13	0.15	1369
Accuracy		0.26		7166
Macro Avg	0.25	0.25	0.24	7166
Weighted Avg	0.25	0.26	0.25	7166

VII. CONCLUSION & FUTURE WORK

A. Findings

Deep learning had been successfully developed and trained for skin tone classification using MobileNetV2. The use of data augmentation significantly improved the model's ability to handle variations in lighting and angles. Additionally, fine-tuning specific layers of MobileNetV2 helped optimize performance and enhanced feature extraction for better classification.

B. Limitations

Despite its effectiveness, the model struggles under poor lighting conditions, leading to occasional misclassifications. Another limitation is dataset bias, as some skin tones have fewer samples, affecting the model's ability to generalize well across all categories.

C. Future Work

To improve accuracy, we plan to expand the dataset by collecting more real-world images with diverse lighting and skin tones. Further fine-tuning of MobileNetV2 by unfreezing its last 5-10 layers can help improve feature learning. Additionally, we aim to deploy the model as a web or mobile application for real-time skin tone classification, making it accessible for practical use.

REFERENCES

[1] Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing.
[2] malakalaabiad, "Skin Tone Classification," Kaggle.com, Sep. 07, 2024. https://www.kaggle.com/code/malakalaabiad/skin-tone-classification#7.-Evaluate-the-Model (accessed Mar. 12, 2025).
[3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks.
[4] DucNguyen168, "Dataset Skin Tone," Kaggle.com, 2024. https://www.kaggle.com/datasets/ducnguyen168/dataset-skin-tone (accessed Mar. 12, 2025).

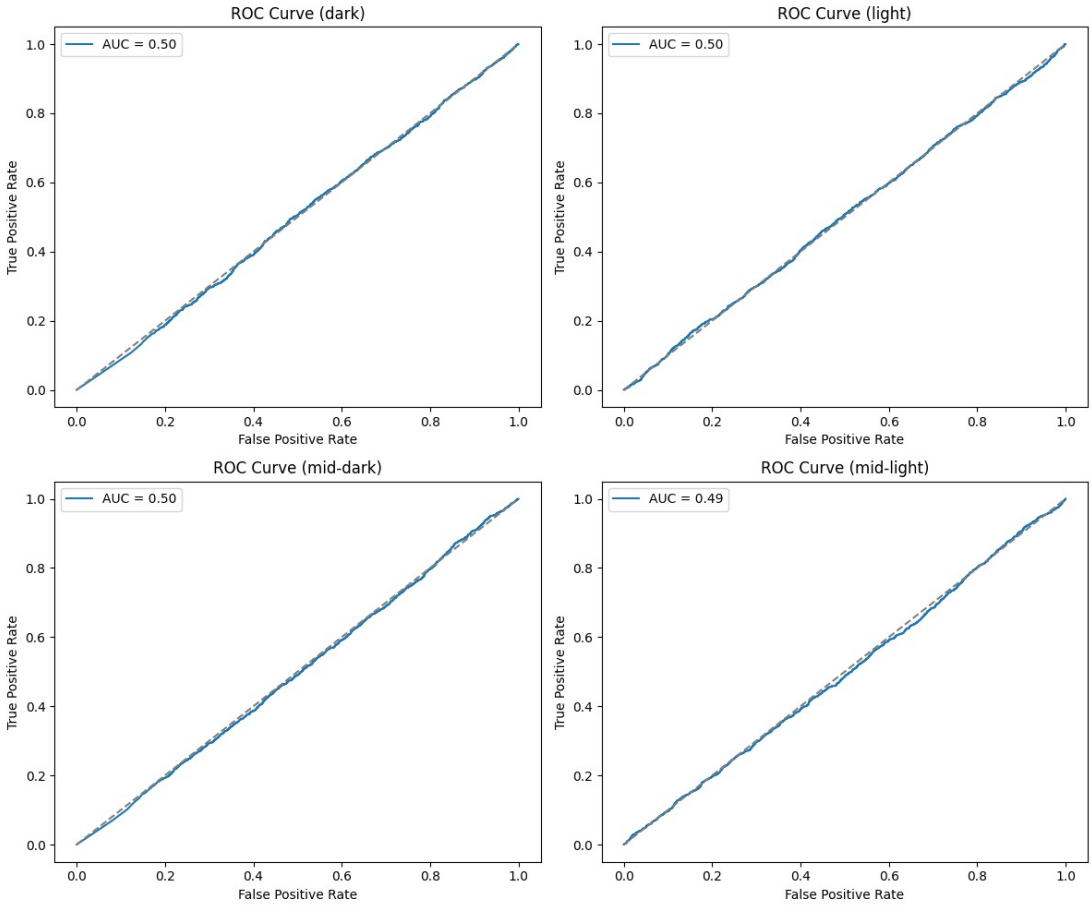

Fig. 3. ROC Curves

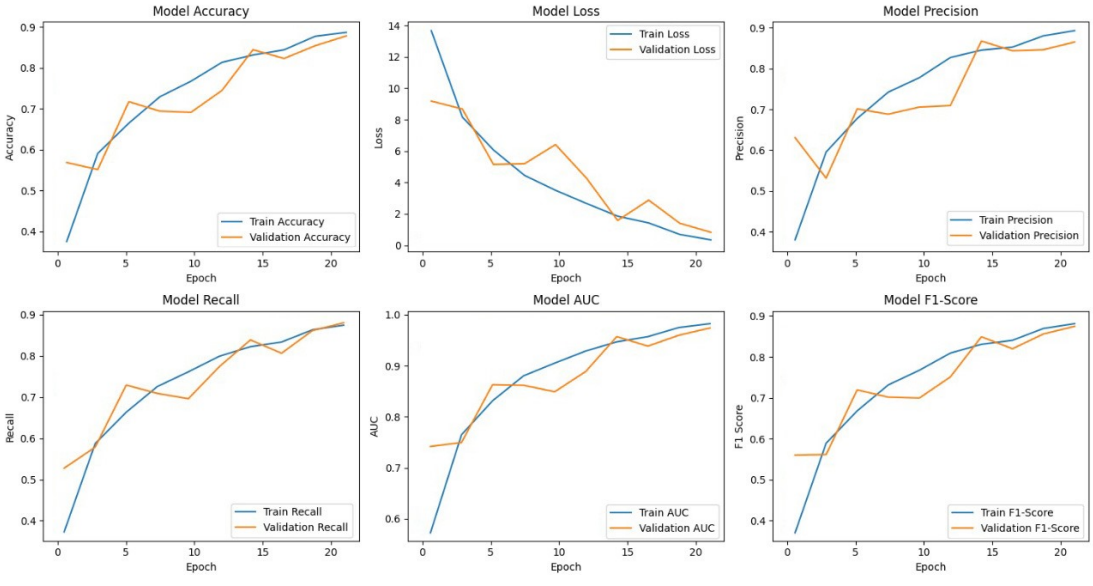

Fig. 4. Performance Visualization

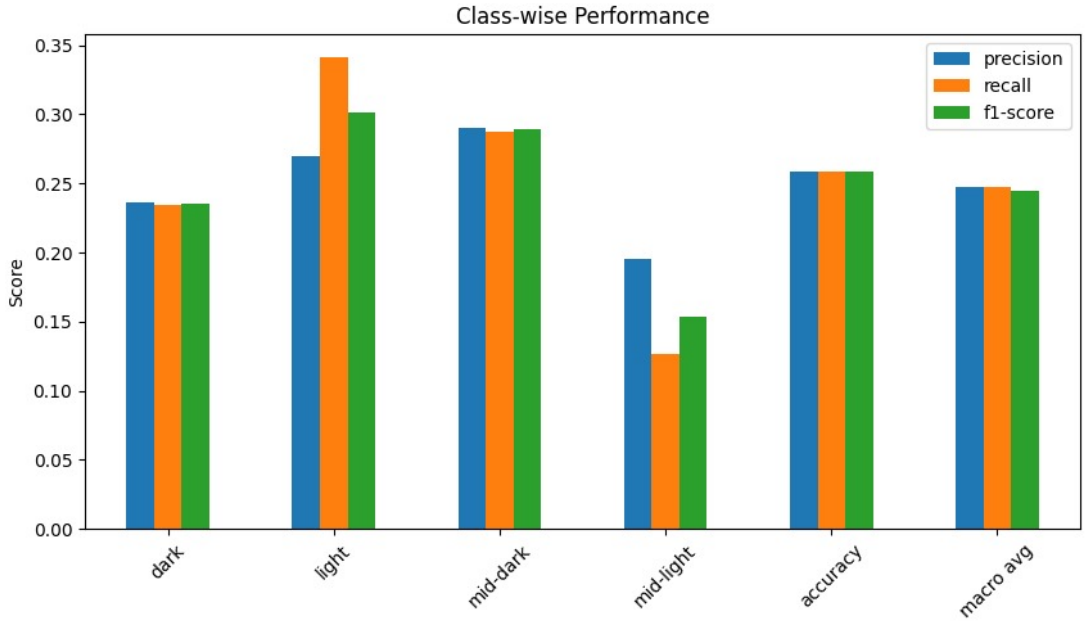

Fig. 5. Class-wise Performance

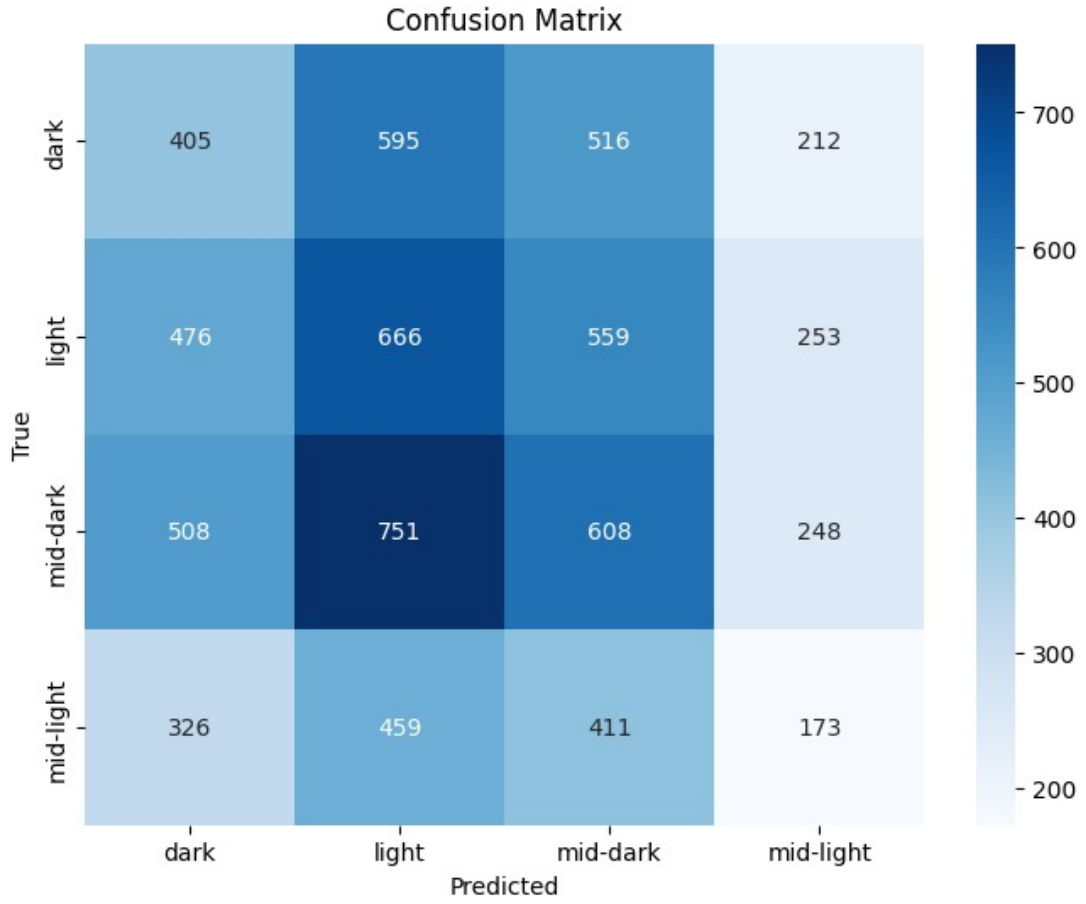

Fig. 6. Confusion Matrix

Fig. 7. Predicted Skin Tone Class: Mid-Light

Fig. 9. Predicted Skin Tone Class: Light

Fig. 8. Predicted Skin Tone Class: Dark

Fig. 10. Predicted Skin Tone Class: Mid-Dark