# Bias in Deep Learning Skin Cancer Detection: Parallel Residual Convolution Network Classification and Racial Bias Quantification

**Conference Paper** · October 2024

CITATIONS
0

READS
27

**2 authors**, including:

Md Biplob Hosen
University of Maryland, Baltimore County
**8** PUBLICATIONS **10** CITATIONS

SEE PROFILE

# Bias in Deep Learning Skin Cancer Detection: Parallel Residual Convolution Network Classification and Racial Bias Quantification

Sumaiya Binte Shahid
Institute of Information Technology,
Jahangirnagar University
Savar, Dhaka, Bangladesh
sbinteshahid@gmail.com

Sabbir Ahmed
Institute of Information Technology,
Jahangirnagar University
Savar, Dhaka, Bangladesh
sabbir.iit.ju@gmail.com

Mohammad Abu Yousuf
Institute of Information Technology,
Jahangirnagar University
Savar, Dhaka, Bangladesh
yousuf@juniv.edu

Md. Biplob Hosen
Institute of Information Technology,
Jahangirnagar University
Savar, Dhaka, Bangladesh
biplob.hosen@juniv.edu

Mohammad Ali Moni
AI and Cyber Futures Institute,
Charles Stuart University
NSW 2795, Australia
m.moni@uq.edu.au

## ABSTRACT

Globally, skin cancer remains one of the most popular and lethal forms of cancer, significantly affecting the death rate. Several studies have been carried out regarding the automatic identification and categorization of skin cancer, employing multiple datasets with varying image compositions. The discrepancies in how different skin tones, colors, and attributes are represented in the datasets cause this variation in image compositions, often known as racial bias. The observed variations significantly influence the development of machine-learning algorithms for skin cancer detection. We have successfully detected racial biases in two datasets. In addition, we have presented a deep convolutional neural network (DCNN) model that is intended for the thorough categorization of skin cancer in several classes. Leveraging the HAM10000 and ISIC-2019, we have developed a robust and accurate model. We have attained 98.55% classification accuracy for HAM10000 dataset and 92.71% classification accuracy for ISIC-2019 dataset. Furthermore, the model's performance is evaluated against several pre-trained transfer learning models to boost efficiency.

## CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; *Machine learning approaches*; Neural networks.

## KEYWORDS

Convolutional Neural Network, International Skin Imaging Collaboration, Dataset, Residual Network


*International Conference on Computing Advancements, October 2024, Dhaka, Bangladesh*


**ACM Reference Format:**
Sumaiya Binte Shahid, Sabbir Ahmed, Mohammad Abu Yousuf, Md. Biplob Hosen, and Mohammad Ali Moni. 2024. Bias in Deep Learning Skin Cancer Detection: Parallel Residual Convolution Network Classification and Racial Bias Quantification. In *ICCA 2024: International Conference on Computing Advancements.* , 8 pages.

## 1 INTRODUCTION

Skin cancer is indeed a typical category of cancer that usually occurs as a result of DNA damage. This damaged DNA leads to excessive growth of cells, resulting in the fast progression of the cancer. Other variables that are associated with it include smoking, alcohol consumption, infectious agents, allergies, viruses, workouts, changes in the surroundings, exposure to UV rays, and many more [1]. Melanoma is the most prevalent type of cancer by which millions of people are affected every year, imposing an immense burden on healthcare systems and societies alike. In the US, 5.4 million new instances of skin cancer are estimated annually [2].

Early identification is essential for skin cancer patients for the effective treatment and for decreasing the mortality rate. Timely diagnosis not only ensures prompt medical intervention but also empowers individuals with the opportunity for better prognoses and enhanced quality of life. Furthermore, early detection facilitates the implementation of less invasive treatment modalities, reducing the physical and emotional burden often associated with more advanced stages of the disease. The foremost way to diagnose skin cancer is via visual inspection by a dermatologist, which has an accuracy rate of about 60% [3]. The application of dermoscopy may significantly enhance this diagnosis accuracy, by up to 89%. While the accuracy of diagnosing melanoma is increased by dermoscopy, certain lesions, especially early melanomas, may still be difficult to identify because they do not have identifiable dermoscopic features. Though machine learning algorithms present a promising opportunity for automating the diagnosis process, they have certain drawbacks. One major drawback is their reliance on the training data. To precisely identify the characteristics and trends linked to skin cancer, these algorithms need large and varied datasets [4].

However, due to some problematic elements such as light reflections from the outermost layer of the skin, differences in lighting, and various shapes, colors and dimensions of the lesion, analysis of these images is quite difficult. Also, these elements result in racial bias. Racial bias in skin cancer datasets presents serious problems and moral dilemmas for skin cancer detection systems. The unequal representation of various racial and ethnic groups in these datasets is one noteworthy problem. People with darker skin tones may be misdiagnosed or delayed-diagnosed as a result of this imbalance, which can cause disparities in algorithmic performance. These challenges highlight the urgent need for focused efforts to identify and mitigate racial biases. Now-a-days, deep learning, particularly Deep Convolutional Neural Networks (CNNs), has been found to be remarkably effective at image identification tasks for their strong ability to reveal patterns and features [5]. Furthermore, these fields are heavily relying on transfer learning to use massive amounts of data and improve the correctness of the outcomes.

The paper primarily presents the following contributions:

- We have conducted a comprehensive analysis to identify and measure racial bias within the HAM10000, ISIC-2019, by examining factors such as inequitable representation of racial and ethnic communities. Our analysis revealed significant biases in the representation of different skin tones.
- A CNN-based classification model for skin cancer multiclassification has been proposed and evaluated on two individual datasets. A comparison was made between our classification model and three pre-trained transfer learning models, demonstrating competitive performance and highlighting areas for improvement in handling racial bias.
- We have illustrated our model by integrating it with well-known datasets like HAM10000 and ISIC-2019 into our multiclass classification and racial bias framework. Additionally, we have computed and shown that both the HAM10000 and ISIC-19 datasets exhibit racial bias.

Below is an outline of the remaining sections of this article: the most related works are presented in Section 2, materials and methodology are covered in Section 3, experimental results and discussion are addressed in Section 4 and Section 5 respectively, and we have concluded our work with further research ideas in Section 6.

## 2 LITERATURE REVIEW

The necessity of identifying cancerous lesions at an early stage has made skin cancer detection increasingly popular among researchers. Numerous studies have been conducted on this subject. This section will describe the main findings of some key studies.

Various researchers have employed machine learning algorithms to identify and categorize skin cancer. Ahmed Thaajwer et al. presented a research endeavor in a study that employed image processing mechanisms and SVM algorithms to develop a highly effective diagnostic system [6]. However, they were unable to find any images with dark skin, and their accuracy was not very satisfactory. Machine learning as well as image processing approaches were also used by M. Krishna Monika et al. to classify skin cancer [7], though the model might be biased in some cases.

An automated approach for diagnosing skin cancer based on Artificial Neural Networks (ANNs) was proposed by Masood et al. [8]. David Rofman et al. developed a model that could quickly identify non-melanoma skin cancer with high specificity and sensitivity,but their ANN model ignored skin lesion visuals [9].

Renato A. Krohling and Andre G. C. Pacheco proposed an attention-based deep learning model. The suggested solution was contrasted with two alternative combination methods: the MetaNet and feature concatenation [10]. In a work, Sana et al. provided a deep convolutional neural network (DCNN) and comprehensible deep learning framework for generating a multiclass skin cancer recognition system. With its clear, comprehensible visual explanations, this method improved accuracy and speed [11]. In a work by Mohammad Fraiwan et al., skin spot images were categorized into seven classes using raw deep transfer learning. An approach to extract spatial information with deep convolutional neural networks was offered by researchers in the publication referenced [12], with the goal of classifying skin tumors. However, concerns were expressed regarding the possibility of overfitting in the proposed model.

Hasan et al. presented a CNN architectural method to produce a binary classification of skin cancer systems. Several CNN models, including SVM, VGG16, and ResNet50 are implemented here. While VGG16 had the best precision (93.18%) among them, the accuracy of the other models was not significantly better [13].

## 3 MATERIALS AND METHODOLOGY

The detailed workflow of our work is given in Figure 1.

### 3.1 Dataset

The HAM10000 and ISIC 2019 datasets were the two datasets on which the experiment was run independently.

- HAM10000: The HAM10000 dataset contains 10,050 dermoscopic images representing seven different classes for detecting pigmented skin lesions [14]. The dataset is freely accessible through the ISIC archive. The following images are contained within the HAM10000 dataset: 327 actinic keratoses, 514 basal cell carcinomas, 1,099 benign keratoses, 115 dermatofibromas, 1,113 melanocytic nevi, 6,705 melanomas.
- ISIC 2019: The collection of ISIC 2019 has 25,331 images which display eight kinds of skin lesions which are melanoma, melanocytic nevus, BCC, AK, benign keratosis, dermatofibroma, vascular lesions, and SCC. The test dataset contains 8239 pictures, while the training dataset should have contained an extra outlier class [15]. Additionally, image metadata including the patient's location, age, and sex can be found in the ISIC 2019 dataset.

### 3.2 Data Preprocessing

The goal of image pre-processing is to enhance certain visual aspects that are important for additional processing and analysis or to eliminate undesirable flaws from the data.

- Image Resizing: Image resizing is a vital initial step for image processing pipelines, enabling adaptability and accessibility. Image resizing frequently has two goals: to reduce file size and optimize quality for certain use cases. In the HAM10000 dataset, the input size provided in the work (28,28,3) resizes

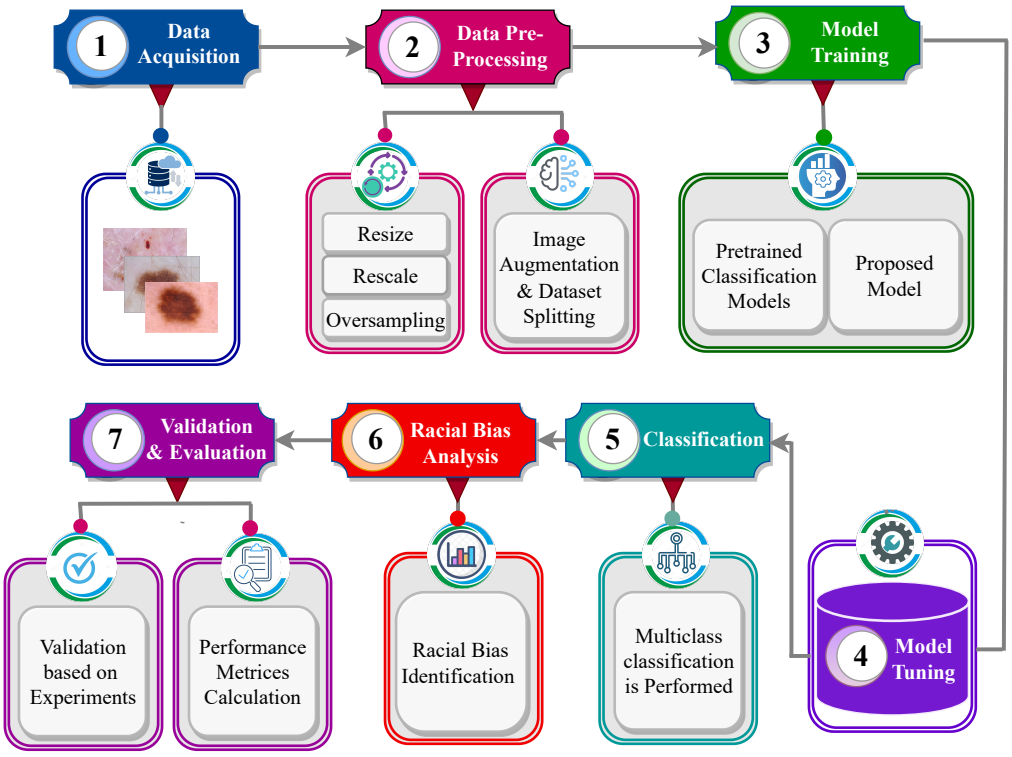


**Figure 1: System Workflow**

each image to a 28-pixel square (height and width) with three color channels—red, green, and blue.

- Image Rescaling: Rescaling of data is a typical preprocessing step in machine and deep learning. The mean pixel value of the total dataset has been subtracted from every pixel value. This aligns the data to nearly zero and reduces pixel bias. The resulting values are then divided by the standard deviation of the dataset.
- Oversampling: The oversampling method used in our experiment oversamples minority classes by randomly reproducing samples until each class is sufficiently represented. For this, the RandomOverSampler is used to balance sample distribution across classes.
- Image augmentation: Image augmentation involves applying multiple modifications to the original images, creating new copies of the images with updated features while keeping their underlying content. We have performed image augmentation activities such as rotation, shearing, zooming, horizontal flipping, ranging of height and width, and resizing on our ISIC 2019 dataset. The augmented dataset has been split into two parts: the training set and the testing-validation set, with an 80:20 ratio preserved.

## 3.3 Model Training

*3.3.1 Pre-trained Transfer Learning Models*

. **ResNet-152V2:** Training deep neural networks can be a challenging task. However, a new approach called residual learning can simplify the process. This approach involves restructuring the layers of the network so that they learn residual functions that reference the inputs to the layer, rather than learning functions that are not related to the inputs. This makes it easier to train networks that are much deeper than those previously used. It skips connections among layers during the training process.

**VGG-16:** The Visual Geometry Group at Oxford is in charge of developing the VGG frame- work, as its name suggests. It focuses on using deep convolutional layers to enhance accuracy. There are multiple versions of the VGG architecture, but for our purposes, we have proposed to fine-tune the VGG-16 architecture. With the VGG architecture, there is a combination of 13 convolutional layers in adding to 3 fully connected layers.

**EfficientNet-B7:** EfficientNet-B7 is a convolutional neural network design known for good computer vision performance. The largest and most powerful EfficientNet variant is B7, developed by Google AI researchers. Our scalable architecture balances model depth, width, and resolution to enhance performance across varied computational resources. EfficientNet B7 leverages squeeze-and-excitation blocks and deep separable convolution to capture complicated data patterns with minimal computational overhead.

## 3.4 Proposed Model

In this article, a novel Convolutional Neural Network (CNN) architecture is proposed to enhance skin cancer diagnosis, incorporating residual and inception-like operations to boost model performance and robustness which is shown in Figure 2. The proposed

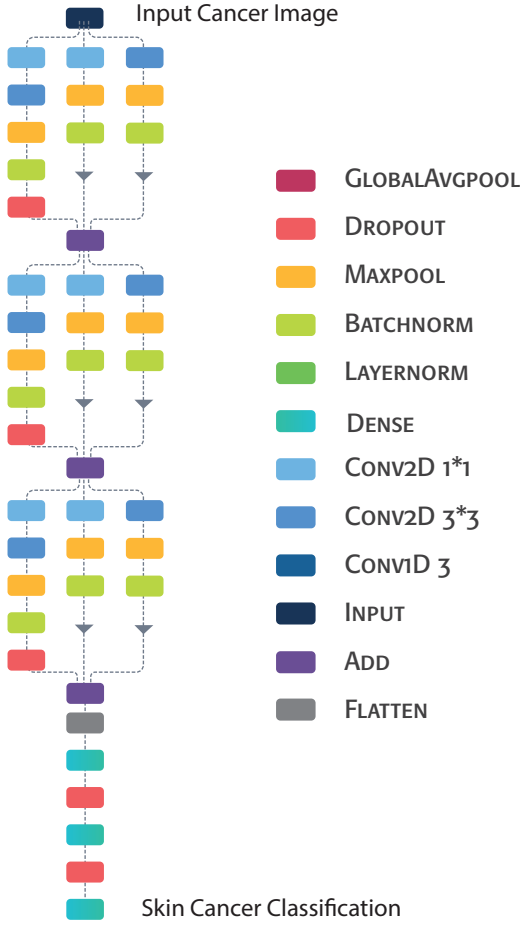


**Figure 2: Proposed Model**

model employs a hybrid block, referred to as `blockred`, which integrates three distinct convolutional strategies: inception modules, VGG-style convolutions with SqueezeNet principles, and Inception-ResNet techniques. The inception module uses a $1 \times 1$ convolution followed by max-pooling and batch normalization to capture local features efficiently. The VGG-style path applies consecutive $3 \times 3$ convolutions with batch normalization and dropout to enforce feature extraction depth and reduce overfitting. The Inception-ResNet path combines $3 \times 3$ and $1 \times 1$ convolutions with max-pooling and batch normalization to merge the benefits of residual connections with inception's multi-scale feature extraction.

Mathematically, the output of the `blockred` module is given by:

$$\text{output1} = \text{add}([x, y, z])$$

where $x$, $y$, and $z$ represent the outputs from the inception, VGG-SqueezeNet, and Inception-ResNet paths, respectively. The overall architecture starts with an initial convolutional layer with 32 filters of size $3 \times 3$ followed by three sequential `blockred` modules, increasing the filter count to 64 in the subsequent layers. After that, the feature maps are flattened and sent to fully connected layers that have 64 or 256 neurons, respectively. To reduce overfitting, dropout layers are inserted after each fully connected layer. Seven neurons and a softmax activation function form the final output layer. This innovative architecture, designed to leverage the strengths of multiple convolutional strategies, aims to improve the model's performance, thus addressing the crucial issue of racial bias in medical image analysis.

### 3.5 Racial Bias Analysis

The methodology involves processing a dataset of skin cancer images, standardized to dimensions of 28x28 pixels with three color channels (RGB). Two critical metrics are calculated for each image: average brightness and mode color intensity. Average brightness provides a measure of overall luminance, while mode color intensity captures the dominant skin tone by identifying the most frequent RGB values in each image.

Mathematically, the average brightness $B$ for an image $I$ with $N$ pixels is computed as:

$$B = \frac{1}{N} \sum_{i=1}^{N} I_i$$

where $I_i$ represents the intensity of the $i$-th pixel. For the mode color intensity, the image $I$ is reshaped into a 2D array where each row contains the RGB values of a pixel. The mode function is then applied along the RGB axis to determine the most frequently occurring color value. This process is described by:

$$\text{ModeColor} = \text{mode}(I.\text{reshape}(-1, 3), \text{axis} = 0)$$

where ModeColor denotes the dominant RGB values. By plotting these metrics, the distribution of brightness and color tones across the dataset is visualized, with horizontal reference lines corresponding to typical brightness and color intensity values for white, brown, and black skin tones. This visualization facilitates the identification of racial biases, ensuring that the CNN model is trained on a diverse and representative set of images, thus enhancing its generalizability and fairness.

## 4 EXPERIMENTAL ANALYSIS

This study has assessed the effectiveness of several transfer learning-based models, namely ResNet152V2, VGG-16, and EfficientNetB7, for categorizing skin cancer using the HAM10000 and ISIC-2019 datasets. The performances of the model were compared in order to ascertain which produced the best outcomes. Better performance matrices scores were demonstrated by our proposed model. Figure 3 provides a lucid illustration of the accuracy and loss curve of the implemented models. The training accuracy vs. validation accuracy curve and the training loss vs. validation loss of the proposed model for the HAM10000 datasets are presented in Figures 3 (a), (d), respectively. Other figures show the accuracy,loss curves for transfer learning models. We have also the accuracy and loss curves for ISIC-2019 dataset. To guarantee the fairness and resilience of the model, the study also looked into racial bias within the datasets. The results of this study will have a big impact on how trustworthy and objective skin cancer classification schemes are created.

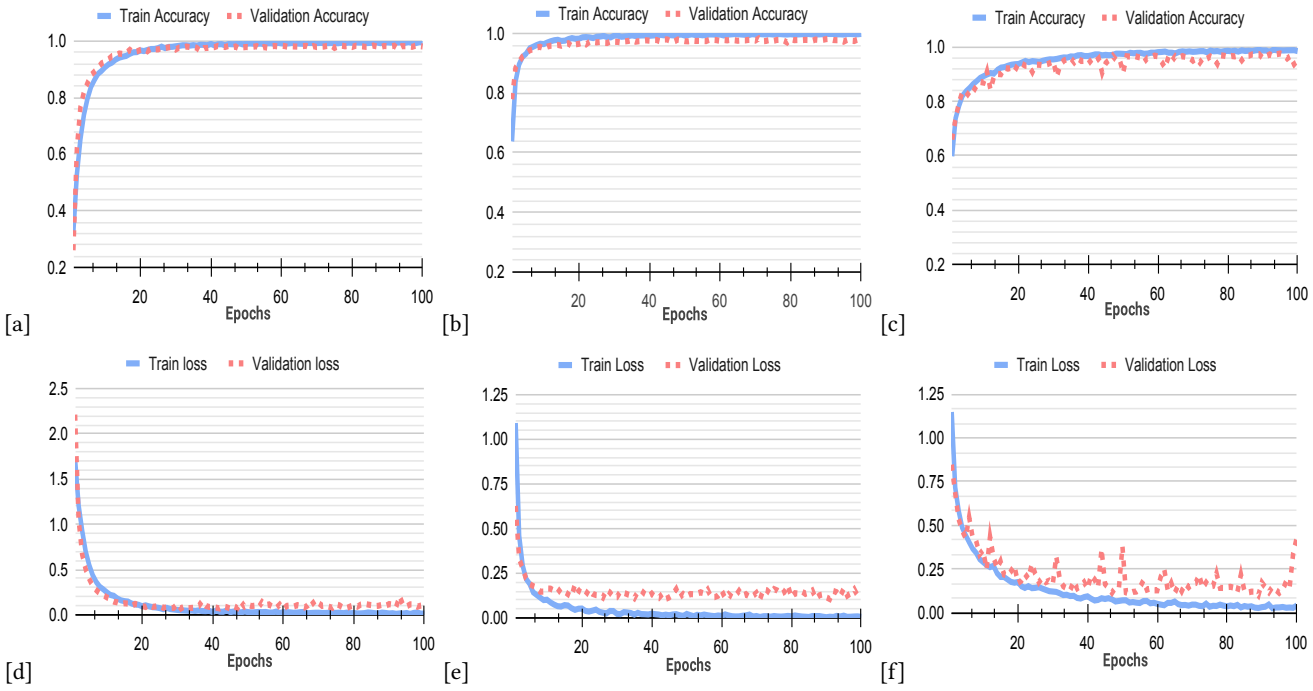


**Figure 3: (a) Proposed Model Accuracy of HAM10000 (b) Accuracy of ResNet152V2 for HAM10000 (c) Accuracy of VGG-16 for HAM10000 (d) Proposed Model Loss curve for HAM10000 (e) Loss curve of ResNet152V2 for Dataset for HAM10000 (f) Loss curve of VGG-16 for Dataset for HAM10000**

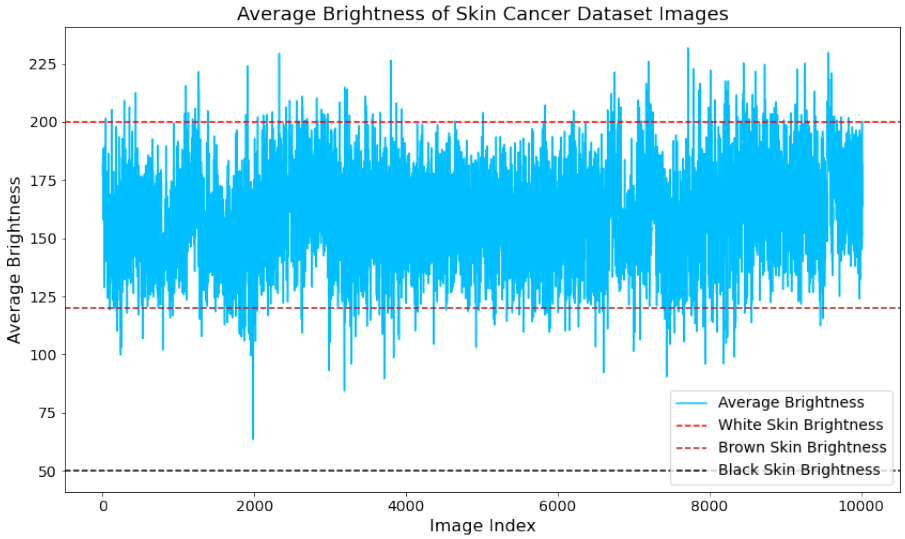


**Figure 4: A Comparison of Brightness in HAM10000 with average white, brown and black skin tone**

## 4.1 Racial Bias Estimation

The Figures 4 and 5 show the mode color intensity for images in the skin cancer dataset HAM10000 and ISIC2019. The axis-Y measures the average brightness, and the axis-X symbolizes the image

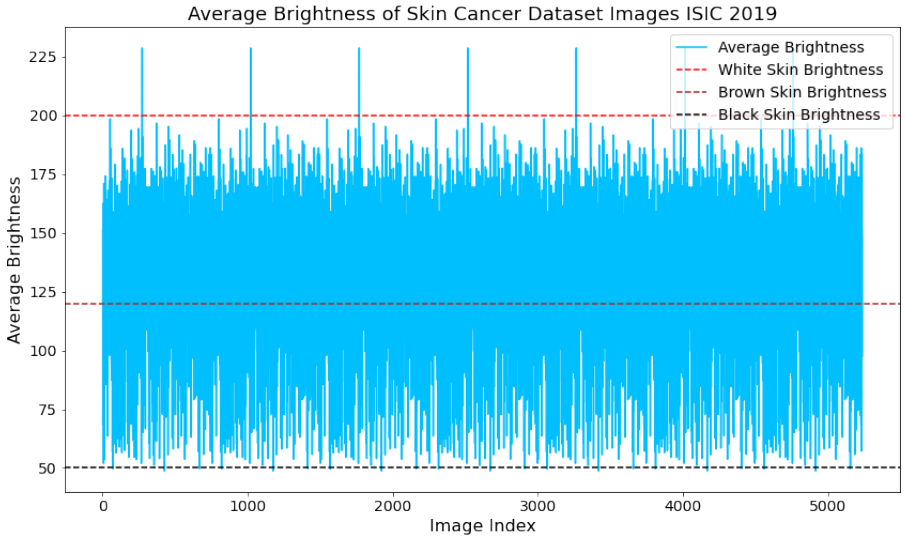


**Figure 5: A Comparison of Brightness in ISIC 2019 with average white, brown and black skin tone**

index. Horizontal reference lines show typical brightness values for white (200), brown (120), and black (50) skin tones. The average brightness values range between 50 and 225, with a concentration around 200. This trend indicates an overrepresentation of images with higher brightness levels, corresponding to lighter skin tones. The underrepresentation of images with lower brightness values further underscores the dataset's racial bias. However The Figures 5 have lower brightness and higher range of brightness than HAM10000's datasets, although, some racial bias is also present in the dataset. The Figure 6 depicts the average brightness for the same 10,000 images but here the HAM10000 and 2019 images are included. The Y-axis represents the mode color intensity, while the X-axis denotes the image index. Horizontal reference lines indicate typical mode color intensity values for white (200), brown (120), and black (50) skin tones. The data points show significant variability, with a clustering around the 200 intensity mark, suggesting a bias towards lighter skin tones. This indicates that the dataset is predominantly composed of images with higher mode colour intensity, highlighting a potential racial bias.

## 4.2 Performance Demonstration and Confusion Matrix

Performance Matrices are indicators that provide a basic and easy evaluation of the efficacy of a classification model. The performance matrix formulas are provided in Equation 1 - 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4}$$

The efficacy of an algorithm for classification is evaluated utilizing a table known as a confusion matrix. Through comparison with the actual findings, it summarizes the model's predictions. In our work, both of the datasets feature a confusion matrix. Here, this confusion matrix, which shows the amount of accurate and inaccurate predictions for each class, is shown in Figure 7 and provides a thorough representation of the models' classification performance for the HAM10000 dataset.

## 4.3 Comparision with other existing Models

Using the HAM10000 and ISIC-2019 datasets, Table 1 compares our suggested model with other models that are currently in use. The findings of Imran Iqbal et al., Juan Pablo et al., HC Reis et al., and MK Monika et al. were surpassed by our proposed model, which obtained F1 score of 0.9855, accuracy of 0.9855, precision of 0.9852 and recall of 0.9858 for the HAM10000 dataset, also F1 score of 0.9345, accuracy of 0.9271, precision 0.9295 and recall of 0.9237 for the ISIC-2019 dataset. This comparison shows how well our proposed model performs in skin cancer classification tasks, as evidenced by its superior performance matrices.

# 5 DISCUSSION

The primary objective of this study was to tackle the issue of accurately classifying skin cancer by employing deep convolutional neural networks (DCNNs), and concurrently identifying and correcting any racial biases present in the datasets used to train these models. The accuracy rates of our proposed model were exceptionally high, attaining 98.58% on the HAM10000 dataset and 92.71% on the ISIC-2019 dataset. The classified classes are presented in Figure 7. These results serve as evidence of the model's efficacy in

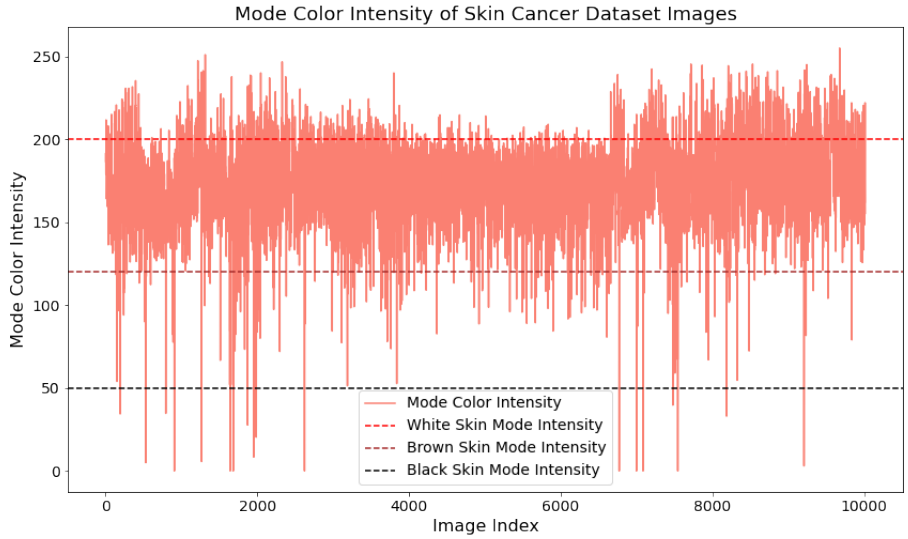


**Figure 6: A Comparison of Color Tone in HAM10000 and ISIC 2019 with average white, brown and black skin tone**

**Table 1: Comparison with Existing Models**

| Reference | HAM10000 | | | | ISIC-2019 | | | |
|---|---|---|---|---|---|---|---|---|
| | F1 Score | Accuracy | Precision | Recall | F1 Score | Accuracy | Precision | Recall |
| Imran Iqbal et al.[16] | 0.8911 | 0.8875 | 0.8943 | 0.8844 | 0.8975 | 0.8958 | 0.9015 | 0.8958 |
| Juan Pablo et al.[17] | 0.98 | 0.98 | 0.98 | 0.98 | 0.93 | 0.91 | 0.91 | 0.93 |
| HC Reis et al.[18] | 0.9139 | 0.9459 | 0.9288 | 0.9356 | 0.9010 | 0.9189 | 0.9255 | 0.9074 |
| MK Monika et al.[7] | - | 0.9704 | - | - | - | 0.9083 | - | - |
| Implemented ResNet152V2 | 0.9785 | 0.9785 | 0.9785 | 0.9795 | 0.9219 | 0.9199 | 0.9229 | 0.9155 |
| Implemented VGG-16 | 0.9755 | 0.9750 | 0.9800 | 0.9710 | 0.9017 | 0.9176 | 0.9117 | 0.9015 |
| **Proposed Model** | 0.9855 | 0.9855 | 0.9858 | 0.9852 | 0.9345 | 0.9271 | 0.9295 | 0.9237 |

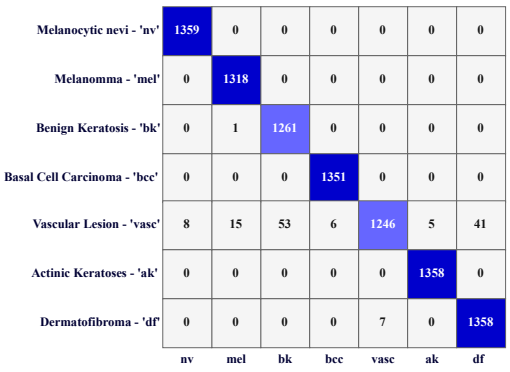


**Figure 7: Confusion Matrix for HAM10000 Dataset**

differentiating various categories of skin cancer. The demonstrated efficacy of our DCNN model highlights the considerable potential that deep learning methods exhibit when applied to medical image analysis. By comparing three well-known transfer learning models (ResNet152V2, VGG16, and EfficientNetB7), the best model was found to be ResNet152V2, which consistently outperformed the other two and demonstrated the advantages of using pre-trained weights for improved performance. The relevant results for both dataset has mentioned in Table 1. Our study's investigation into racial bias in the datasets was crucial since it demonstrated the existence of bias and highlighted the necessity of more inclusive and representative data gathering procedures in order to ensure equal healthcare results. For enhancing the reliability and beauty of skin cancer classification systems, future research should concentrate on expanding the diversity of datasets, implementing cutting-edge methods like attention mechanisms, and combining multimodal data. Additionally, attention must be directed toward racial bias mitigation. Our research not only underscores the effectiveness of deep convolutional neural networks (DCNNs), specifically ResNet152V2, for categorizing skin cancer but also draws attention to the crucial

concern of racial bias. We advocate for the utilization of diversified datasets and sophisticated modeling techniques to guarantee equitable and precise healthcare solutions for all demographics.

## 6 CONCLUSION

This paper used a deep convolutional neural network (DCNN) to classify skin cancer into multiple categories. It attained notable accuracy rates of 98.58% on the HAM10000 data and 92.71% on the ISIC-2019 dataset. A comparative analysis of three transfer learning models—ResNet152V2, VGG16, and EfficientNetB7—revealed that of effectiveness with ResNet152- V2. The analysis further disclosed substantial racial bias within the datasets, underscoring the imperative for data acquisition practices that are more inclusive in nature. By showcasing the capabilities of DCNNs in medical image analysis, this research emphasizes the criticality of tackling racial bias in order to establish healthcare solutions that are fair.

## REFERENCES


[1] Md Shahin Ali, Md Sipon Miah, Jahurul Haque, Md Mahbubur Rahman, and Md Khairul Islam. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Machine Learning with Applications*, 5:100036, 2021.

[2] Azhar Imran, Arslan Nasir, Muhammad Bilal, Guangmin Sun, Abdulkareem Alzahrani, and Abdullah Almuhaimeed. Skin cancer detection using combined decision of deep learners. *IEEE Access*, 10:118198–118212, 2022.

[3] Riccardo G Borroni, Vincenzo Panasiti, Mario Valenti, Luigi Gargiulo, Giuseppe Perrone, Roberta Dall'Alba, Clarissa Fava, Francesco Sacrini, Luca L Mancini, Sofia AAM Manara, et al. Long-term sequential digital dermoscopy of low-risk patients may not improve early diagnosis of melanoma compared to periodical handheld dermoscopy. *Cancers*, 15(4):1129, 2023.

[4] Harsh Bhatt, Vrunda Shah, Krish Shah, Ruju Shah, and Manan Shah. State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review. *Intelligent Medicine*, 3(03):180–190, 2023.

[5] Atika Akter, Nazeela Nosheen, Sabbir Ahmed, Mariom Hossain, Mohammad Abu Yousuf, Mohammad Ali Abdullah Almoyad, Khondokar Fida Hasan, and Mohammad Ali Moni. Robust clinical applicable cnn and u-net based algorithm for mri classification and segmentation for brain tumor. *Expert Systems with Applications*, 238:122347, 2024.

[6] MA Ahmed Thaajwer and UA Piumi Ishanka. Melanoma skin cancer detection using image processing and machine learning techniques. In *2020 2nd international conference on advancements in computing (ICAC)*, volume 1, pages 363–368. IEEE, 2020.

[7] M Krishna Monika, N Arun Vignesh, Ch Usha Kumari, MNVSS Kumar, and E Laxmi Lydia. Skin cancer detection and classification using machine learning. *Materials Today: Proceedings*, 33:4266–4270, 2020.

[8] Ammara Masood, Adel Ali Al-Jumaily, and Tariq Adnan. Development of automated diagnostic system for skin cancer: Performance analysis of neural network learning algorithms for classification. In *Artificial Neural Networks and Machine Learning–ICANN 2014: 24th International Conference on Artificial Neural Networks, Hamburg, Germany, September 15-19, 2014. Proceedings 24*, pages 837–844. Springer, 2014.

[9] David Roffman, Gregory Hart, Michael Girardi, Christine J Ko, and Jun Deng. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Scientific reports*, 8(1):1701, 2018.

[10] Andre GC Pacheco and Renato A Krohling. An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, 25(9):3554–3563, 2021.

[11] Sana Ulla and Mohammad Abu Yousuf. Mscc: Multi-class skin cancer classification and interpretable deep learning systems. In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, pages 735–740. IEEE, 2024.

[12] Rehan Ashraf, Iqra Kiran, Toqeer Mahmood, Ateeq Ur Rehman Butt, Nafeesa Razzaq, and Zobia Farooq. An efficient technique for skin cancer classification using deep learning. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–5. IEEE, 2020.

[13] Mohammed Rakeibul Hasan, Mohammed Ishraaf Fatemi, Mohammad Monirujjaman Khan, Manjit Kaur, Atef Zaguia, et al. Comparative analysis of skin cancer (benign vs. malignant) detection using convolutional neural networks. *Journal of Healthcare Engineering*, 2021, 2021.

[14] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

[15] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap. Analysis of the isic image datasets: Usage, benchmarks and recommendations. *Medical image analysis*, 75:102305, 2022.

[16] Imran Iqbal, Muhammad Younus, Khuram Walayat, Mohib Ullah Kakar, and Jinwen Ma. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. *Computerized medical imaging and graphics*, 88:101843, 2021.

[17] Juan Pablo Villa-Pulgarin, Anderson Alberto Ruales-Torres, Daniel Arias-Garzon, Mario Alejandro Bravo-Ortiz, Harold Brayan Arteaga-Arteaga, Alejandro Mora-Rubio, Jesus Alejandro Alzate-Grisales, Esteban Mercado-Ruiz, M Hassaballah, Simon Orozco-Arias, et al. Optimized convolutional neural network models for skin lesion classification. *Computers, Materials & Continua*, 70(2), 2022.

[18] Hatice Catal Reis, Veysel Turk, Kourosh Khoshelham, and Serhat Kaya. Insinet: a deep convolutional approach to skin cancer detection and segmentation. *Medical & Biological Engineering & Computing*, pages 1–20, 2022.