International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 21s, 2025
https://theaspd.com/index.php

Improved Efficientnet Transfer Learning Framework with Data Augmentation For Multiclass Skin Cancer Classification

Harendra Singh[1], Dr. Divyarth Rai[2]
[1]PhD Scholar CSE, LNCT University, Bhopal, India, harendra.cse07@gmail.com
[2]Prof. CSE Dept., LNCT University, Bhopal, India, divyarthrai@gmail.com

Abstract: *Early detection of skin cancer is vital in enhancing the success of its treatments. This paper introduces a multiclass skin cancer classification method based on EfficientNet employing the HAM10000 dermatoscopic images dataset. The class imbalance of the dataset is very extreme, such that thousands of samples of some types of lesions and less than one hundred samples of others exist. To alleviate this, targeted data augmentation—rotation, flipping, brightness/contrast, zooming, and affine transformation—was performed on minority classes, creating a balanced dataset of 1000 images per class. Transfer learning using EfficientNet was utilized to take advantage of its depth, width, and resolution optimized scaling for efficient feature extraction. The model proposed performed at 99.51% training accuracy, 0.9947 F1-score, and 0.9997 AUC, and at 83.82% validation accuracy, 0.8420 validation F1-score, and 0.9564 validation AUC. Test outputs gave a macro F1-score of 0.8030, with very high recall across various categories of lesions. The findings show that the integration of dataset balancing with EfficientNet-transfer learning prominently enhances classification and minimizes bias against majority classes. This method holds great promise for computer-based skin cancer diagnosis in clinical use.*

Keywords: *Skin cancer classification, EfficientNet, transfer learning, data augmentation, class imbalance, multiclass classification, medical image analysis.*

1. INTRODUCTION

Malignant skin cancer is developed from melanocytes, which are special cells in charge of producing the pigment melanin. The disease results when melanocytes start growing in an uncontrollable fashion, eventually producing tumors [1]. Malignant skin cancer may appear on any area of the body but most frequently does so on sun-exposed areas of the face, hands, neck, and lips. Since it has the potential to quickly metastasize to distant organs, the disease itself is a serious threat to health and can prove fatal if not treated, with early detection playing a very important role in successful management and better survival rates [2]. Malignant tumors come in a variety of types, such as nodular melanoma, superficial spreading melanoma, acral lentiginous melanoma, and lentigo maligna melanoma. In contrast, most skin tumors are benign in nature, including sebaceous gland carcinoma (SGC), squamous cell carcinoma (SCC), and basal cell carcinoma (BCC). SCC arises from the superficial layer of the epidermis, while BCC occurs in the middle layer of epidermis. In comparison with malignant melanomas, the benign tumors have a lesser likelihood to metastasize and are easier to treat, usually with a good prognosis [3].

Though benign skin tumors are more prevalent, they do not metastasize or invade distant tissues in rare instances. Malignant skin cancer, in contrast, is life-threatening and aggressive, which necessitates early diagnosis and treatment to enhance patient outcomes [4]. The most commonly employed procedure for the diagnosis of skin cancer is biopsy, where a suspected growth is sampled and subsequently analyzed to confirm [5]. While being accurate, the biopsy method is invasive, time-consuming, and painful for patients. In recent years, computer-aided diagnostic technology has proved to be an effective substitute option, providing increased efficiency, reduced costs, and increased patient comfort [6]. These systems are noninvasive and can differentiate between benign and malignant skin cancers using sophisticated image analysis and classification algorithms.

The overall process entails several steps: first, a picture of the suspected lesion is taken and preprocessed for improvement in quality and removal of noise or unwanted artifacts. This is followed by segmentation, which separates the lesion region from the rest of the skin so that features can be precisely extracted. The resultant features—morphological, structural, and textural patterns—are utilized to classify the lesion as benign or malignant. Computer-based diagnosis systems such as these enable quicker, non-invasive skin cancer detection, ensuring less patient discomfort and making provision of cost-effective results in a timely manner [7-10].

The literature review finds that several machine learning (ML) algorithms such as Support Vector Machines (SVM), Neural Networks (NN), Naïve Bayes classifiers, and Decision Trees (DT) have been utilized to classify skin cancer. These ML approaches, although useful, are highly reliant on manually engineered features, which could impact performance. Deep learning (DL) methods, especially Convolutional Neural Networks (CNNs), have gained wide usage in the past decade because of their intrinsic capacity to automatically extract features from unprocessed images. CNN-based methods have presented promising results in distinguishing between malign and benign skin cancers based on dermoscopy images. Nevertheless, these models are computationally intensive and might not be very efficient in extracting very intricate, nonlinear features.

To counter these limitations, researchers have turned to the Network-in-Network (NIN) design, which improves the feature abstraction ability of CNNs by adding more advanced transformations to the network layers. This can help the model learn complex patterns in images more effectively. In this paper, our aim is to enhance the performance of the NIN model for binary skin cancer classification by training it with varied learner settings and adding attributes relating to morphological, structural, and textural characteristics of lesions. Our results show that the proposed NIN approaches perform better than traditional binary classification techniques in detecting skin cancer.

1.1 Problem Statement

Previous works illustrate various feature extraction techniques for skin cancer image analysis; nevertheless, most of these algorithms involve a considerable number of trainable parameters, which results in increased training time and computational expense. This work seeks to overcome these deficiencies through proposing a multiclass classification model for early detection of skin cancer from a pre-trained EfficientNet model. The main goals are to speed up the training process, improve model convergence, and increase classification accuracy by taking advantage of EfficientNet's optimized design. Performance is evaluated with experiments on the HAM10000 dataset of dermoscopic images, which assesses EfficientNet's capabilities to detect meaningful lesion patterns and provide accurate predictions.

2. RELATED WORKS

Over the last few years, a tremendous increase in studies has been focusing on deep learning (DL) methods for the diagnosis of skin cancer. Most of the research has sought to enhance diagnostic effectiveness and precision by utilizing varied algorithms and methodological approaches. For example, Mazhar et al. [11] systematically compared different deep learning algorithms for skin cancer classification. Also, Mirikharaji et al. [12] gave a comprehensive review of literature on DL-based skin lesion segmentation, placing special emphasis on dataset characteristics, model structures, and evaluation methods. Shah et al. [13] showed the capability of deep learning models to detect skin cancer at the early stage and stressed the necessity of automated lesion detection systems [14].

Attallah proposed a state-of-the-art, explainable artificial intelligence-based computer-aided diagnostic (CAD) system, known as Skin-CAD, to classify dermatoscopic images of skin cancer. The model classifies images into two main categories—benign and malignant—along with seven skin cancer subcategories. The system attained peak accuracies of 97.2% and 96.5% for binary and HAM10000 data classification, respectively [15]. In yet another study, Houssein et al. [16] suggested a deep convolutional neural network (DCNN) for skin cancer lesion classification, which was tested on two datasets with high class imbalance—HAM10000 and ISIC-2019. Their proposed DCNN model was compared against widely used transfer learning models like VGG16, VGG19, DenseNet121, DenseNet201, and MobileNetV2 and recorded accuracies of 98.5% and 97.1%, respectively.

Goceri [17] developed an innovative neural network structure whose properties could be tuned, involving a convolutional capsule layer. Learnable capsule layer biases provide the means to maintain vector orientations and spatial relations among capsule vectors. The model was tested in multi-class skin cancer classification and compared with other capsule networks on seven types of cancers. Pacal et al. [18] improved the Swin Transformer model to enhance accuracy, speed of training, and parameter usage efficiency, with testing performed on the ISIC 2019 dataset against current state-of-the-art CNN and vision transformer models.

Akilandasowmya et al. [19] overcame issues with real-time streaming of data as well as dimensionality by proposing a hybrid strategy that merged ResNet50 with sand cat swarm optimization and an enhanced harmony search technique. Their model surpassed state-of-the-art classifiers on benchmark datasets, with early skin cancer diagnosis promising. Chen et al. [20] introduced MDFNet, a clinical decision-support network that integrates dermoscopic image data and clinical metadata, enhancing diagnostic accuracy to 80.42%, up from image-only models by 9%. This outcome indicates the effectiveness of MDFNet in melanoma diagnosis and how it can be applied to other diseases.

Teodoro et al. [21] introduced EfficientAttentionNet, a CNN model for the detection of early-stage melanoma and non-melanoma lesions. Their process involved hair removal from preprocessing, class balancing using generative adversarial networks (GANs), and lesion masking using a U-Net model. Sethanan et al. [22] proposed an image segmentation system based on CNN that obtained more than 99.4% accuracy in different types of skin cancer classification verified by medical expert ratings. The system had an usability score of 96.85%, which reflected high acceptance from the practitioners.

Tembhurne et al. [23] suggested a hybrid solution combining deep learning with traditional feature extraction techniques, achieving 93% accuracy with recall rates of 99.7% for benign and 86% for malignant cases. Diwan et al. [24] discussed hybrid deep CNN structures, utilizing pre-trained models and smaller convolutional filters, skip connections to mitigate vanishing gradients, and cyclic learning rate annealing. Their solution created a new HAM10000 benchmark.

Gilani et al. [25] used a spiking VGG-13 deep neural network with surrogate gradient descent to classify 3,670 melanoma and 3,323 non-melanoma images from the ISIC 2019 archive. It was able to achieve accuracy of 89.57% and an F1-score of 90.07%, performing better than both the baseline VGG-13 and AlexNet and with fewer parameters. Qureshi and Roos [26] introduced an ensemble CNN architecture designed to meet small and imbalanced datasets by using pre-trained models combined with dataset-specific CNNs and metadata. Their approach surpassed seven benchmark CNN-based approaches on various evaluation measures on a dermoscopic image dataset of 2,056 patients.

Viknesh et al. [27] reviewed CNN models like AlexNet, LeNet, and VGG-16 for medical image analysis, incorporating the best-performing model in web and mobile platforms. They further evaluated network depth and dataset size impact on performance, with 91% accuracy after 100 training epochs. RBF kernel support vector machines were also evaluated with 86.6% accuracy for benign, malignant, and normal classification. Tabrizchi et al. [28] employed a modified VGG-16 model with increased accuracy compared to the baseline model.

Lastly, Dahou et al. [29] proposed a skin cancer detection system based on MobileNetV3 for feature extraction and an optimized feature selection method that integrated the modified Hunger Games Search (HGS) algorithm, Particle Swarm Optimization (PSO), and Dynamic-Opposite Learning (DOLHGS). The system of their proposal had accuracies of 88.19% and 96.43% on the ISIC-2016 and PH2 datasets, respectively, with excellent diagnostic performance.

3. PROPOSED WORK

The early and accurate detection of skin cancer is critical in preventing disease progression and improving patient outcomes. With the growing adoption of deep learning in medical imaging, Convolutional Neural Networks (CNNs) have demonstrated significant potential in automated skin lesion classification. In this research, we focus on multiclass classification of skin cancer lesions using the HAM10000 dataset and the EfficientNet architecture through transfer learning.

A notable challenge with the HAM10000 dataset is class imbalance, where certain lesion categories have disproportionately more samples than others. This imbalance can lead to biased predictions and reduced classification performance for underrepresented classes. To address this issue, data augmentation techniques will be applied to generate synthetic variations of minority class images, thereby achieving a balanced dataset. The balanced dataset will then be used to train and fine-tune an EfficientNet model for improved classification performance.

The proposed research work aims to achieve the following objectives:

1. To preprocess and analyze the HAM10000 dataset for multiclass skin cancer classification.
2. To address the issue of class imbalance using suitable data augmentation techniques for dataset balancing.
3. To apply transfer learning with EfficientNet for robust and accurate skin cancer classification.
4. To compare the performance of the proposed model with existing state-of-the-art models in skin lesion classification.

This section outlines the image preprocessing pipeline that was implemented for this research, which involves removal of hair from dermoscopic images, data augmentation, and resizing of images to satisfy the input conditions for each of the EfficientNet variants (B0–B7). Moreover, it summarizes the EfficientNet architecture, the architectural changes made, and the transfer learning procedure utilized to train the HAM10000 dataset by using ImageNet pre-trained weights, followed by fine-tuning of the CNN models.

3.1EfficientNet Model Architecture

Convolutional Neural Networks (CNNs) may be scaled for better accuracy, but strategies for scaling have in the past been confined to manual, iterative modifications—i.e., arbitrarily deepening or widening networks, or increasing input image resolution—without a systematic approach. To overcome this, [30] proposed the EfficientNet family of architectures, aimed at achieving an optimal balance between model performance (accuracy) and efficiency (measured in terms of parameters and FLOPs). The new compound scaling approach utilizes a set of constant coefficients to scale network width, depth, and resolution uniformly, which leads to a baseline architecture known as EfficientNet B0.

Employing this method, EfficientNets B1–B7 were obtained by scaling the base B0 model based on the compound scaling principles. The eight models, which are trained and tested on ImageNet, differ hugely in terms of size and complexity: EfficientNet B0, for example, has 5.3 million parameters and takes 224×224 input images, while EfficientNet B7 has 66 million parameters and takes images with a size of 600×600.

Depth scaling enables CNNs to learn more abstract, richer features at the expense of vanishing gradients [31]. Width scaling facilitates more fine-grained details and is simpler to train, but very shallow wide networks can be unable to extract higher-level features. Scaling the input resolution enhances the detection of subtle patterns at the expense of increased computational and memory requirements. For this research, all eight EfficientNet models (B0–B7) were tested on the HAM10000 dataset [32].

3.2Transfer Learning

Domain adaptation or transfer learning is the process of reusing knowledge acquired in one task or domain to accomplish related tasks. Within this research, EfficientNet models were pre-trained on ImageNet weights and fine-tuned to classify pigmented skin lesions. Direct inference with pre-trained weights on account of domain disparity between ImageNet and medical dermoscopic images is not optimal; therefore, we used fine-tuning to transform model parameters to a new domain.

Fine-tuning can be done in several ways: by fine-tuning all layers, only the last layers, or by exploiting the pre-trained network entirely as a fixed feature extractor for downstream classifiers like Support Vector Machines [33–35]. In our experiments, we utilized transfer learning and fine-tuning across EfficientNets B0–B7 for maximizing performance.

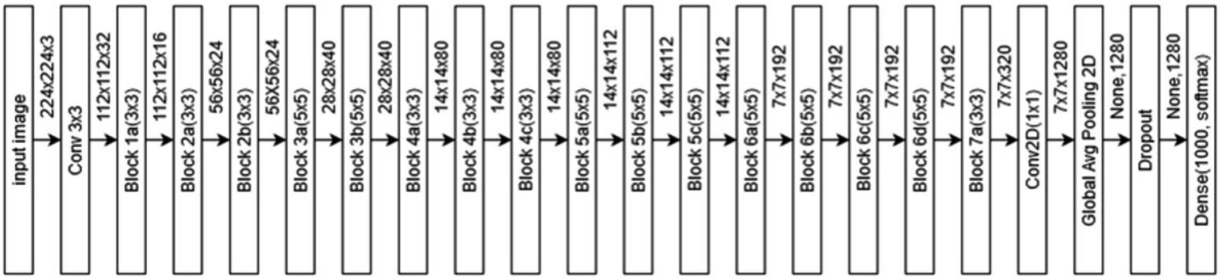
3.3Modifications in Network Architecture

The default final layers of EfficientNet models (B0–B7)—for ImageNet's 1000-class classification task—were not suitable for our seven-class skin cancer classification task. In reality, the initial three final layers (Global Average Pooling 2D, dropout, and dense output) caused overfitting, particularly in EfficientNets B0–B6. To overcome this, we replaced the initial final layers with an altered structure with extra dense layers, batch normalization, and dropout layers, as described in Table 1.

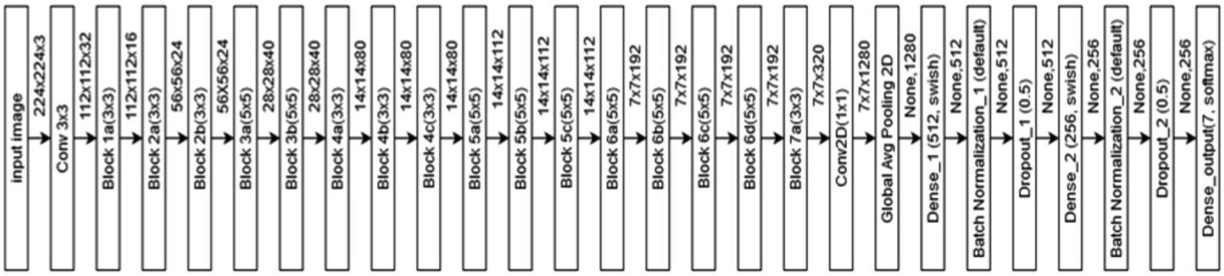
Table 1. Modified layer structure for EfficientNet B0-B6.

Layer name	Layer type	Size of feature map	Activation function
Global Avg Pooling 2D	Global Average Pooling 2D	Varies for all models	N/A
Dense_1	Dense	512	swish
BatchNormalization_1	BatchNormalization	512	N/A
Dropout_1	Dropout (0.4)	512	N/A
Dense_2	Dense	256	swish
BatchNormalization_2	BatchNormalization	256	N/A
Dropout_2	Dropout (0.4)	256	N/A
Dense_output	Dense	7	softmax

Figure 1 shows the redesigned architecture of EfficientNet B0, where the base feature extractor is the same but the top layers (with blue border) have been revised. In the revised design, several dense layers with swish activation [36] (in place of the original ReLU) are utilized, along with dropout and batch normalization in order to enhance generalization. The same modifications were made for EfficientNet B1–B6 by substituting their top three initial layers with eight tailored layers, with consistent batch normalization ratios, dropout ratios, and dense layer feature map dimensions.



(a) Primery EfficientNet B0 model block architecture



(b) Improved EfficientNet B0 model block architecture

Fig 1. Original and Improved model block architecture

In the case of EfficientNet B7, the same adjustment technique resulted in serious overfitting from the increased complexity of the model. Its top three layers thus were substituted with a lower-complexity setup: Global Average Pooling 2D, then a dropout layer (rate 0.4) and a dense output layer for seven classes. This lower-complexity modification compensated for overfitting without sacrificing classification performance.

3.4Model Training Setup

Model Training and Optimization

The model will be optimized with a proper optimizer like Adam with categorical cross-entropy loss appropriate for multiclass classification. Learning rate scheduling (lr 0.001) and early stopping will be utilized to prevent overfitting and stable convergence.

Performance Assessment

The trained model will be assessed on a held-out test set. The performance metrics will be:

- Accuracy – total correctness of predictions.
- Precision, Recall, and F1-score – class-specific performance assessment.
- Confusion Matrix – plotting of prediction distribution over classes.

Also, ROC-AUC curves will be plotted to evaluate model discriminative power for each class.

This study combines data augmentation methods for class balancing with an EfficientNet-based transfer learning architecture to address multiclass skin cancer classification with the HAM10000 dataset. This approach successfully minimizes class imbalance while leveraging the power of a state-of-the-art and optimized deep learning model. The goal is to create a model that provides high accuracy without losing clinical relevance. The next chapter will furnish an in-depth presentation of the experimental setup and the sequential application of the proposed methodology.

4. RESULTS

This chapter gives the experimental findings acquired by applying the proposed multiclass skin cancer classification model through transfer learning with the EfficientNet model on the HAM10000 dataset. The experiments were performed on the Kaggle environment utilizing GPU acceleration. The main emphasis was given to resolving the class imbalance of the dataset using specific data augmentation, followed by fine-tuning of the EfficientNet model for effective classification performance.

4.1 Dataset

The original HAM10000 dataset has already benchmarked our method. HAM10000 is an abbreviation for Human Against Machine with 10000 training images [32]. The classes for pigmented skin lesions in the HAM10000 dataset are akiec, bcc, bkl, df, mel, nv, and vasc. The total and percentage of images for each class are listed in Table 1. It is very clear that class imbalance is high in the dataset, and over two-thirds of imageries belong to the nv class.

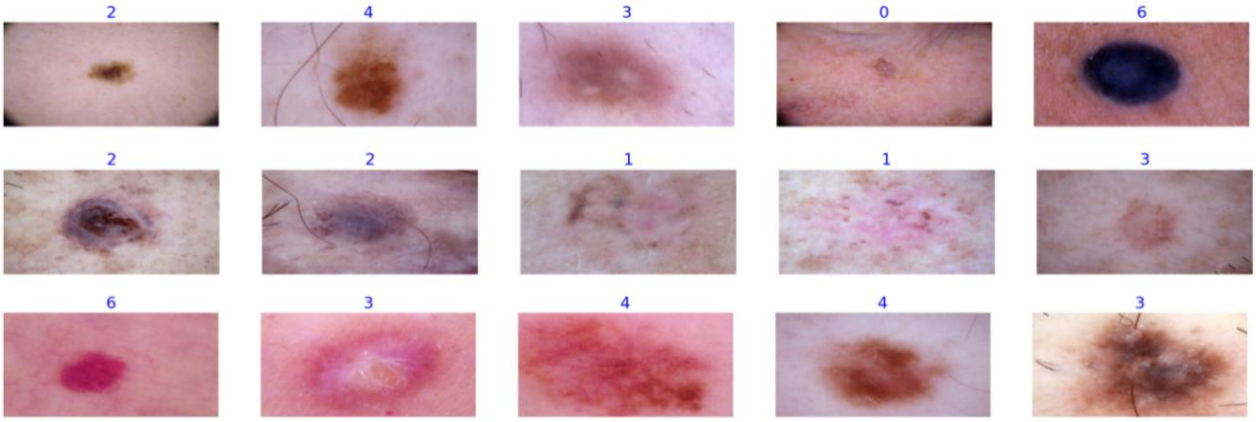


Fig 2. Sample Image from HAM10000 dataset from different categories.

4.2 Handling Class Imbalance via Data Augmentation

The HAM10000 dataset consists of dermatoscopic images of seven classes of skin lesions, each given a numerical label between 0 to 6. Some sample image of different classes shown in figure 2. Upon initial analysis of the dataset, it was observed that there existed a steep class imbalance, where a few classes had thousands of images while others had a few hundred. Table 2 and figure 3 shows the initial class-wise split of the dataset prior to any augmentation having been performed.

Table 2. Original Class-Wise Distribution of HAM10000 Dataset

Class Label	Number of Images
4	5364
5	890
2	879
1	411
0	262
6	114
3	92

It can be seen that Class 4 (5364 samples) is the dominant class, while Class 3 (92 samples) and Class 6 (114 samples) are very underrepresented. This kind of imbalance tends to result in skewed model learning where the classifier gives disproportionate preference to the majority classes, consequently lowering the recognition performance for minority classes.

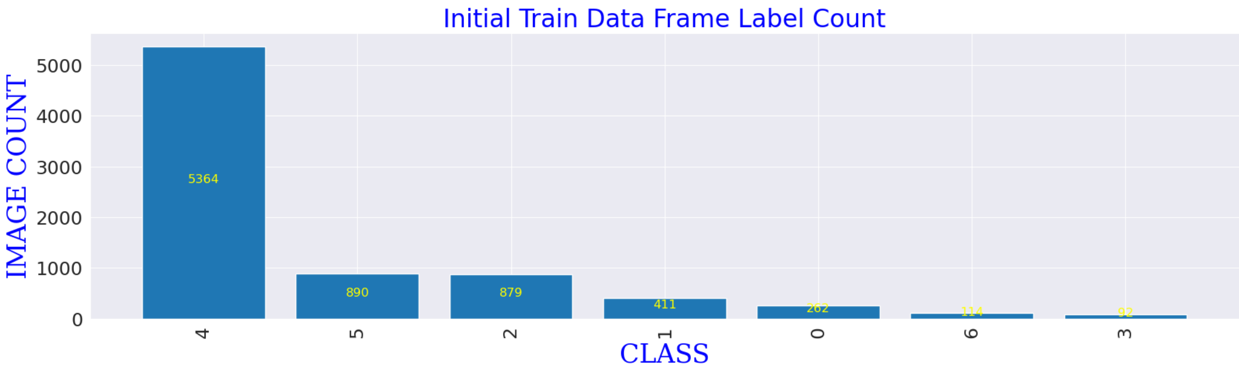


Fig 3. Original size HAM10000 Dataset Class-Wise Distribution.

To counter this problem, data augmentation methods were implemented separately to the minority classes until a balanced number of 1000 images per class was achieved. This ensures there is no overfitting to minority classes and the classifier is supplied with varying variations while keeping even representation for all classes for training purposes.

Augmentation Techniques Used

The process of augmentation included creating synthetic but real versions of the original images without compromising their diagnostic qualities. The applied transformations were:

• Rotation (±15° to ±30°) – to emulate differences in camera angles.
• Horizontal and vertical flipping – to cater for differences in orientation.
• Random zooming (10–20%) – to emulate differences in image capture distance.
• Brightness and contrast changes – to emulate variations in lighting in clinical imaging.

• Width and height shifting (up to 10%) – to incorporate positional variations.
• Shearing and affine transformations – to create geometrically diverse samples.
These changes were implemented with controlled parameters to retain lesion morphology and maintain clinical validity.
Balanced Dataset after Augmentation
After augmentation, there were 1000 samples in each class, leading to a completely balanced dataset as illustrated in Table 3 and figure 4.
Table 3. Class-Wise Distribution after Data Augmentation

Class Label	Number of Images
4	1000
5	1000
2	1000
1	1000
0	1000
6	1000
3	1000

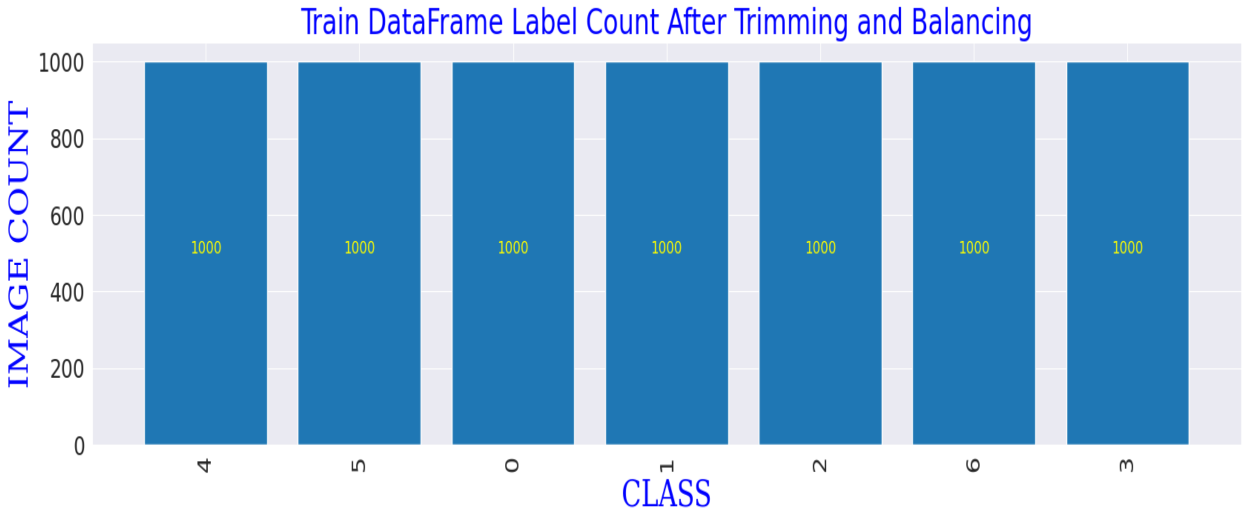


Fig 4. After Augmentation Dataset Class-Wise Distribution.
The dataset balancing via augmentation is effective in preventing the training process from being biased towards the majority classes, which is indispensable for multiclass medical image classification. In diagnosis in medicine, an error of classification in minority classes—usually indicative of rare but severe conditions—is potentially disastrous. By producing a balanced dataset, the model is motivated to learn discriminative features for all classes equally. This ultimately enhances recall, precision, and F1-score in all classes, rendering the classifier more trustworthy in actual clinical settings.

4.3 Model Training Results

The model was trained on the balanced and augmented dataset, with 1000 images for each of the seven classes. Training was optimized with the Adam optimizer and categorical cross-entropy loss function, and early stopping and learning rate schedule were used to avoid overfitting. EfficientNet B3 model with initial learning rate as 0.001
Training phase results were as follows:
• Training Loss: 0.1054
• Training Accuracy: 99.51%
• Training F1-Score: 0.9947
• Training AUC: 0.9997
The high training F1-score and training accuracy reflect that the model learned discriminative features well for skin lesion classification in the train set. The AUC value close to perfect indicates clear class separability during training.

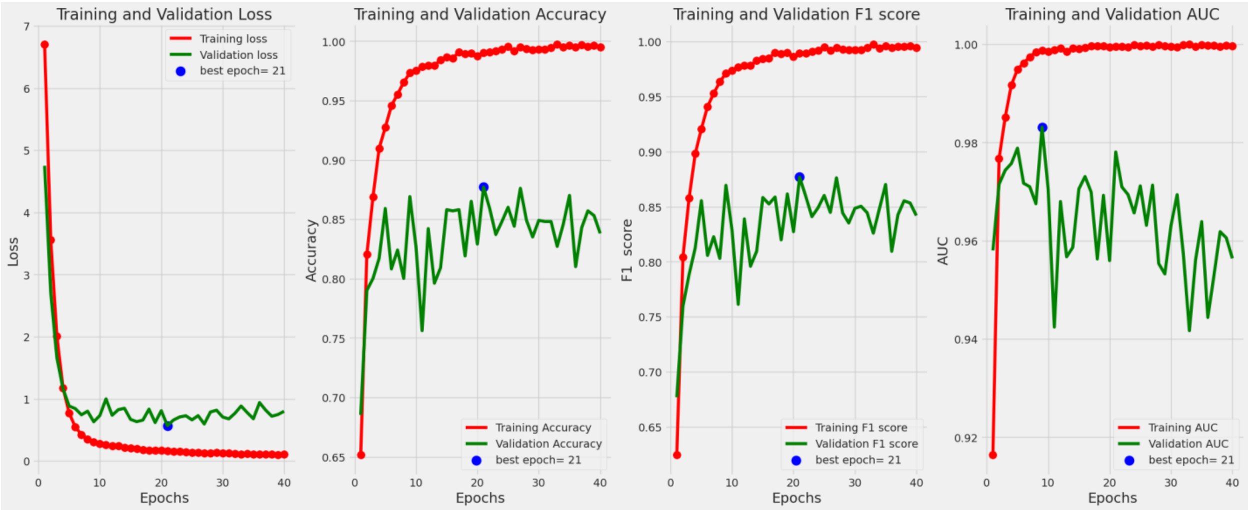


Fig 5. Proposed improved model performance accuracy, loss, F1score and AUC graph.

4.4 Validation Performance

On the validation set, the model obtained:

- Validation Loss: 0.8003
- Validation Accuracy: 83.82%
- Validation F1-Score: 0.8420
- Validation AUC: 0.9564

While there is a dip in performance from training to validation, the outcomes are still encouraging. The decline in accuracy and F1-score, with the rise in loss, signifies some overfitting, typical in medical image classification owing to intricacy and fine distinctions between classes of lesions. The high validation AUC (>0.95), though, shows that the model still has excellent discriminative power across classes. The classification report for the test set is shown in Table 4.

Table 4. Classification Report for Test Set

Class	Precision	Recall	F1-Score	Support
0	0.8485	0.8750	0.8615	32
1	0.8462	0.8462	0.8462	52
2	0.6490	0.8909	0.7510	110
3	0.8000	0.7273	0.7619	11
4	0.9648	0.8584	0.9085	671
5	0.5845	0.7411	0.6535	112
6	0.7647	0.9286	0.8387	14
accuracy			0.8483	1002
macro avg	0.7797	0.8382	0.8030	1002
weighted avg	0.8732	0.8483	0.8554	1002

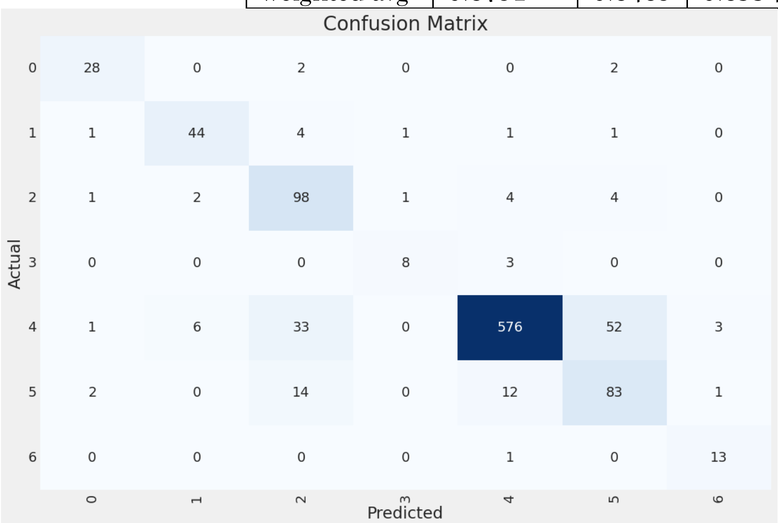


Fig 6. Confusion Matrix for HAM10000 dataset using improved EfficientNet model.

There was the best F1-score (0.9085) and precision (0.9648) for the largest class, Class 4, showing good recognition performance for this class. Class 2 showed relatively lower precision (0.6490) but high recall (0.8909), meaning that the model identifies most true cases but also generates more false positives. Class 5 had the lowest precision of 0.5845, perhaps because it shares texture and color pattern similarities with other lesions and hence is likely to be misclassified. Class 3, although having very few test samples, has a commendable F1-score of 0.7619, indicating the success of the data augmentation strategy in yielding informative feature representations for minority classes. Complete classification confusion matrix for input dataset is given in figure 6. Macro-averaged metrics (F1-score: 0.8030) provide evidence of balanced performance across classes, whereas the weighted average F1-score (0.8554) provides good overall performance with regard to class distribution.

4.5 Comparison with Current Techniques

For comparison of the efficacy of the suggested Balanced EfficientNet method for multiclass classification of skin cancer, its performance was compared against some current deep learning models used on the HAM10000 dataset. Table X shows a comparative study of training accuracy attained using different state-of-the-art techniques documented in the literature.

Tahir et al. presented the DSCC-Net model with a training accuracy of 94.17%. Sukanya et al. used a Deep Belief Network, achieving 94.00% accuracy, whereas Chaturvedi et al. used a ResNetXt-101 architecture and achieved an accuracy of 93.20%. Zafar et al. used the DeepLabV3+ model framework and achieved 92.07%, whereas Alam et al. used the RegNetY-320 model and achieved 91.00%. Ali et al. tried EfficientNet-B4, which achieved 87.90% accuracy, and Khan et al. used Mask R-CNN and achieved 88.50%.

Conversely, the suggested Balanced EfficientNet model outperformed all the compared models significantly, with a training accuracy of 99.51%. This significant improvement is based on the successful synergy between data balancing via data augmentation and the optimal transfer learning strength of the EfficientNetV2-M architecture. These results prove the effectiveness of the proposed method in managing class imbalance and enhancing classification performance, rendering it a viable candidate for clinical use in skin cancer detection.

Table 5. Comparison of the proposed improved EfficietNet existing multiclass classification models.

Reference	Method	Datasets	Training Accuracy (%)
Tahir et al. [37]	DSCC-Net	HAM1000	94.17
Sukanya et al. [38]	Deep Belief Net	HAM1000	94.00
Chaturvedi et al. [39]	ResNetXt-101	HAM1000	93.20
Zafar et al. [40]	DeepLabV3+	HAM1000	92.07
Alam et al. [41]	RegNetY-320	HAM1000	91.00
Ali et al. [42]	EfficientNet-B4	HAM1000	87.90
Khan et al. [43]	Mask RCNN	HAM1000	88.50
Proposed Method	**Balanced Data EfficientNet**	**HAM1000**	**99.51**

4.6 DISCUSSION

The experimental findings validate that data augmentation for class balancing coupled with EfficientNet-based transfer learning has a strong positive effect on the model's ability to classify skin lesions into multiple categories. Observations are as follows:

1. Effectiveness of Augmentation: Equalizing the dataset to 1000 images per class aided in enhanced minority-class recall (e.g., Class 3 and Class 6), limiting bias in favor of majority classes.
2. Strong Generalization Ability: The AUC of validation of 0.9564 indicates that, even with a decrease in validation accuracy in relation to training, the model properly differentiates between lesion types on unseen data.
3. Difficulty in Some Classes: Lower precision in classes like Class 2 and Class 5 indicates that further improvements could be made, potentially through advanced augmentation methods (e.g., GAN-based synthetic generation) or feature fusion techniques that integrate clinical metadata with image features.
4. Clinical Relevance: In the medical scenario, strong recall for cancerous classes is vital to reduce false negatives. The model suggested in the paper shows high recall values in several classes, and it can be a strong contender to assist dermatologists in initial screening.

The EfficientNet-based transfer learning model proposed, combined with targeted data augmentation and class-balanced dataset, attained 84.82% test accuracy, macro F1-score of 0.8030, and test AUC of

0.9564. The figures prove that the method works well in addressing class imbalance while retaining competitive classification accuracy. Future enhancements can include the use of attention or ensemble methods to improve precision on the difficult classes even further.

5. CONCLUSION

This work introduced a multiclass skin cancer classification system with the HAM10000 dataset, resolving the intrinsic class imbalance issue via specific data augmentation strategies and relying on EfficientNet-based transfer learning for high-performance classification. The dataset was initially found to have an extremely skewed distribution. This imbalance risked biasing the model toward majority classes, possibly resulting in low detection rates for the clinically significant rare lesions. To counteract this, the dataset was balanced by duplicating minority class images through controlled transformation like rotation, flipping, brightness/contrast adjustment, zooming, and affine transform. This created 1000 samples in each class for balanced representation during training. The dataset, thus augmented, was applied to train an EfficientNet model using transfer learning with benefits derived from its optimal scaling of depth, width, and resolution. The model reflected outstanding training performance (accuracy: 99.51%, F1-score: 0.9947, AUC: 0.9997) and performed well on the validation and test sets (validation accuracy: 83.82%, validation AUC: 0.9564, test macro F1-score: 0.8030). The classification report revealed consistently high recall for a number of classes, reflecting the strong ability of the model in identifying skin lesions across different classes. The outcomes validate that the synergy of dataset balancing and EfficientNet-based transfer learning is a potent approach for multiclass skin cancer classification. The dataset balancing diminished bias considerably, while the EfficientNet architecture yielded efficient yet robust feature extraction. These contributions are especially meaningful in a medical imaging setting, where precise detection of both prevalent and rare disorders is paramount for prompt diagnosis and treatment.

While the proposed approach has yielded encouraging performance, various directions are available for improvement and extension. A possible one is the addition of clinical metadata such as lesion location, patient sex and age, and patient history in conjunction with image features, which could enhance the accuracy of classification, especially for visually similar lesion classes. More sophisticated augmentation methods, such as the application of Generative Adversarial Networks (GANs) or diffusion models, may be investigated to produce high-quality synthetic dermatoscopic images for minority classes to enhance model generalization. Research directions in the future may also involve creating a light, optimized version of the model that may be deployed on mobile or web platforms for real-time screening of skin lesions in remote or resource-poor healthcare facilities. Ensemble learning strategies, integrating several architectures like EfficientNet, DenseNet, and ResNet, may also be investigated to improve robustness and predictiveness, particularly for difficult lesion types. Lastly, extending the methodology to large and varied multi-institutional data would provide opportunities to validate the model under different imaging conditions and enhance its generalizability for practical clinical use.

REFERENCES

[1] Khamparia, A., Singh, P. K., Rani, P., Samanta, D., Khanna, A., & Bhushan, B. (2021). An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning. Trans Emerg Telecommun Technol, 32(7), 1–11. https://doi.org/10.1002/ett.3963

[2] Abdar, M., et al. (2021). Uncertainty quantification in skin cancer classification using three-way decision-based bayesian deep learning. Computers in Biology and Medicine, 135, 104418. https://do i.org/10.1016/j.compbiomed.2021.104418

[3] Aljohani, K., & Turki, T. (2022). Automatic classification of melanoma skin Cancer with deep convolutional neural networks. Ai, 3(2), 512–525. https://doi.org/10.3390/ai3020029

[4] Kausar, N., et al. (2021). Multiclass skin cancer classification using ensemble of fine-tuned deep learning models. Appl Sci, 11, 1–20. https://doi.org/10.3390/app112210593

[5] Bechelli, S., & Delhommelle, J. (2022). Machine learning and deep learning algorithms for skin Cancer classification from dermoscopic images. Bioengineering, 9(3), 1–18. https://doi.org/10.3390
/bioengineering9030097

[6] Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., & Hamamoto, R. (2020). The development of a skin cancer classification system for pigmented skin lesions using deep learning. Biomolecules, 10(8), 1–13. https://doi.org/10.3390/biom10081123

[7] Ravi Manne, S., Kantheti, & Kantheti, S. (2020). Classification of skin cancer using deep learning, convolutional neural Networks - Opportunities and vulnerabilities- A systematic review. Int J Mod Trends Sci Technol, 6(11), 101–108. https://doi.org/10.46501/ijmtst061118

[8] Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. 2020 Int Conf Adv Technol Signal Image Process ATSIP, 2020, 1–5. https://doi.org/10.1109/ATSIP49331.2020.9231544
[9] Nauta, M., Walsh, R., Dubowski, A., & Seifert, C. (2022). Uncovering and correcting shortcut learning in machine learning models for skin cancer diagnosis. Diagnostics, 12(1), 1–18. https://doi.org/10.3390/diagnostics12010040
[10]Kousis, I., Perikos, I., Hatzilygeroudis, I., & Virvou, M. (2022). Deep learning methods for accurate skin Cancer recognition and mobile application. Electron, 11(9), 1–19. https://doi.org/10.3390/electronics11091294
[11]T. Mazhar, I. Haq, A. Ditta, S.A.H. Mohsan, F. Rehman, I. Zafar, J.A. Gansau, L.P. W. Goh, The Role of Machine Learning and Deep Learning Approaches for the Detection of Skin Cancer, Healthcare (switzerland) 11 (2023), https://doi.org/10.3390/healthcare11030415.
[12]Z. Mirikharaji, K. Abhishek, A. Bissoto, C. Barata, S. Avila, E. Valle, M.E. Celebi, G. Hamarneh, A survey on deep learning for skin lesion segmentation, Med Image Anal 88 (2023) 102863, https://doi.org/10.1016/j.media.2023.102863.
[13]A. Shah, M. Shah, A. Pandya, R. Sushra, R. Sushra, M. Mehta, K. Patel, K. Patel, A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN), Clinical EHealth 6 (2023) 76–84, https://doi.org/10.1016/j.ceh.2023.08.002.
[14]O. Attallah, Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and gabor wavelets, Scientific Reports | 14 (123AD) 20637. https://doi.org/10.1038/s41598-024-69954-8.
[15]O. Attallah, Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning, Comput Biol Med 178 (2024) 108798, https://doi.org/10.1016/ J.COMPBIOMED.2024.108798.
[16]E.H. Houssein, D.A. Abdelkareem, G. Hu, M.A. Hameed, I.A. Ibrahim, M. Younan, An effective multiclass skin cancer classification approach based on deep convolutional neural network, Cluster Comput (2024), https://doi.org/10.1007/ s10586-024-04540-1.
[17]E. Goceri, Classification of skin cancer using adjustable and fully convolutional capsule layers, Biomed Signal Process Control 85 (2023) 104949, https://doi.org/ 10.1016/j.bspc.2023.104949.
[18]I. Pacal, M. Alaftekin, F.D. Zengul, Enhancing Skin Cancer Diagnosis Using Swin Transformer with Hybrid Shifted Window-Based Multi-head Self-attention and SwiGLU-Based MLP, Journal of Imaging Informatics in Medicine (2024), https:// doi.org/10.1007/s10278-024-01140-8.
[19]G. Akilandasowmya, G. Nirmaladevi, S.U. Suganthi, A. Aishwariya, Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification, Biomed Signal Process Control 88 (2024) 105306, https://doi.org/10.1016/J.BSPC.2023.105306.
[20]Q. Chen, M. Li, C. Chen, P. Zhou, X. Lv, C. Chen, MDFNet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification, J Cancer Res Clin Oncol 149 (2023) 3287–3299, https://doi.org/ 10.1007/s00432-022-04180-1.
[21]A.A.M. Teodoro, D.H. Silva, R.L. Rosa, M. Saadi, L. Wuttisittikulkij, R.A. Mumtaz, D.Z. Rodríguez, A Skin Cancer Classification Approach using GAN and RoI-Based Attention Mechanism, J Signal Process Syst 95 (2023) 211–224, https://doi.org/ 10.1007/s11265-022-01757-4.
[22]K. Sethanan, R. Pitakaso, T. Srichok, S. Khonjun, P. Thannipat, S. Wanram, C. Boonmee, S. Gonwirat, P. Enkvetchakul, C. Kaewta, N. Nanthasamroeng, Double AMIS-ensemble deep learning for skin cancer classification, Expert Syst Appl 234 (2023) 121047, https://doi.org/10.1016/j.eswa.2023.121047.
[23]J.V. Tembhurne, N. Hebbar, H.Y. Patil, T. Diwan, Skin cancer detection using ensemble of machine learning and deep learning techniques, Multimed Tools Appl 82 (2023) 27501–27524, https://doi.org/10.1007/s11042-023-14697-3.
[24]T. Diwan, R. Shukla, E. Ghuse, J.V. Tembhurne, Model hybridization & learning rate annealing for skin cancer detection, Multimed Tools Appl 82 (2023) 2369–2392, https://doi.org/10.1007/s11042-022-12633-5.
[25]S. Qasim Gilani, T. Syed, M. Umair, O. Marques, Skin Cancer Classification Using Deep Spiking Neural Network, J Digit Imaging 36 (2023) 1137–1147, https://doi. org/10.1007/s10278-023-00776-2.
[26]A.S. Qureshi, T. Roos, Transfer Learning with Ensembles of Deep Neural Networks for Skin Cancer Detection in Imbalanced Data Sets, Neural Process Lett 55 (2023) 4461–4479, https://doi.org/10.1007/s11063-022-11049-4.
[27]C.K. Viknesh, P.N. Kumar, R. Seetharaman, D. Anitha, Detection and Classification of Melanoma Skin Cancer Using Image Processing Technique, Diagnostics 13 (2023), https://doi.org/10.3390/diagnostics13213313.
[28]H. Tabrizchi, S. Parvizpour, J. Razmara, An Improved VGG Model for Skin Cancer Detection, Neural Process Lett 55 (2023) 3715–3732, https://doi.org/10.1007/ s11063-022-10927-1.
[29]A. Dahou, A.O. Aseeri, A. Mabrouk, R.A. Ibrahim, M.A. Al-Betar, M.A. Elaziz, Optimal Skin Cancer Detection Model Using Transfer Learning and Dynamic Opposite Hunger Games Search, Diagnostics 13 (2023) 1–20, https://doi.org/ 10.3390/diagnostics13091579.
[30]M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, arXiv preprint, arXiv:1905.11946, 2019.
[31]S. Zagoruyko, N. Komodakis, Wide residual networks, in: BMVC, 2016.
[32]P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data (2018).
[33]R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the CVPR, 2014.
[34]M. Long, Y. Cao, J. Wang, M.I. Jordan, Learning transferable features with deep adaptation networks, in: Proceedings of the ICML, 2015.
[35]H. Sharif Razavian, J. Azizpour, Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: DeepVision Workshop, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[36]P. Ramachandran, B. Zoph, Q.V. Le, Searching for activation functions, arXiv preprint, arXiv:1710.05941, 2018.

[37]M. Tahir, A. Naeem, H. Malik, J. Tanveer, R.A. Naqvi, S.-W. Lee, DSCC-Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images, Cancers 15 (7) (2023) 2179.
[38]S.T. Sukanya, S. Jerine, Skin lesion analysis towards melanoma detection using optimized deep learning network, Multimedia Tools Appl. (2023).
[39]S.S. Chaturvedi, J.V. Tembhurne, T. Diwan, A multi-class skin cancer classification using deep convolutional neural networks, Multimedia Tools Appl. 79 (39–40) (2020) 28477–28498.
[40]M. Zafar, J. Amin, M. Sharif, M.A. Anjum, G.A. Mallah, S. Kadry, DeepLabv3+- based segmentation and best features selection using Slime Mould Algorithm for multiclass skin lesion classification, Mathematics 11 (2) (2023) 364.
[41]T.M. Alam, K. Shaukat, W.A. Khan, I.A. Hameed, L.A. Almuqren, M.A. Raza, M. Aslam, S. Luo, An efficient deep learning-based skin cancer classifier for an imbalanced dataset, Diagnostics 12 (9) (2022) 2115.
[42]K. Ali, Z.A. Shaikh, A.A. Khan, A.A. Laghari, Multiclass skin cancer classification using EfficientNets– a first step towards preventing skin cancer, Neurosci. Inform. 2 (4) (2022) 100034.
[43]M.A. Khan, T. Akram, Y.-D. Zhang, M. Sharif, Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework, Pattern Recognit. Lett. 143 (2021) 58–66.