Research Article

Automated Skin Lesion Classification Using Efficientnet-B4 Deep Learning Architecture On The Ham10000 Dataset

Venkata Leela Kamal Challa*[1], P Vidya Sagar[1]

[1]Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India
Correspondence Email: 2401050026@kluniversity.in

ABSTRACT

Timely diagnosis of skin cancer is a key to longer life and better results of the treatment process; nevertheless, exploring dermoscopic images manually by the dermatologists is both time-consuming and has a high risk of diagnostic variability owing to the visual similarity of various types of skin lesions. In recent times, the development of deep learning has produced automated systems of image analysis as potentially useful tools in helping diagnose a medical illness. The current research provides a framework of deep learning-based classification of multiclass skin lesions with the help of the EfficientNet-B4 and HAM10000 models. The data set has dermoscopic images of seven classes of diagnoses, such as melanoma, basal cell carcinoma, benign keratosis, dermatofibroma, vascular lesion, actinic keratoses, and melanocytic nevus. The image preprocessing and augmentation methods are used to enhance the generalization and training efficiency of models whereas transfer learning is used to fine-tune the EfficientNet-B4 network to the task of classification. The provided model is experimentally tested to present the overall accuracy of 86.9% which proves the effectiveness of deep learning methods in assisting in early skin cancer screening and computerized dermatological diagnosis.

Keywords: Skin Cancer Detection, Dermoscopic Image Analysis, Deep Learning, EfficientNet-B4, HAM10000 Dataset, Multi-class Classification.

1. INTRODUCTION

Skin cancer ranks amongst the most prevalent types of cancer in the world and it is a significant public health agenda since its rate of incidence is constantly increasing. It mostly is seen when there is uncontrolled growth of abnormal cells of the skin due to over exposure of ultraviolet rays of the sun or other sources like tanning beds. The three commonest skin cancer are melanoma, basal cell carcinoma and squamous cell carcinoma, each having different severity and complexity of treatment. Among them, melanoma is deemed as the most perilous type due to its high probability of getting transferred to other body parts when it is not identified at an early age. Hence, prompt diagnosis and treatment are needed to enhance survival and lower healthcare cost per patient.

Dermoscopic imaging has now emerged as a valuable method of diagnosing skin lesions among the dermatologists. It enables a clear visualization of those skin structures which cannot be seen by naked eye with the help of which benign and malign lesions can be distinguished better. Nevertheless, dermoscopic images can be interpreted manually and mostly this process is relying on the abilities and experience of the dermatologists. Manual diagnosis can cause variability and a possible false diagnosis because the lesion type same cannot be identified by visual differences given that there are a great number of patients and the difference between the lesion types is subtle. These constraints show the importance of automated and trustworthy computer-aided diagnostic systems, which can help clinicians in the true determination of various forms of skin lesions.

Over the last few years, deep learning methods and specifically convolutional neural networks (CNNs) have proven to be incredibly effective at medical image processing challenges such as disease detection, classification, and segmentation. CNN-based networks can extract the complicated hierarchical-based features in images automatically and thus can be very useful in the classification of dermatological images. A number of studies have examined the use of deep learning architectures to analyze the skin lesion with publicly available dataset of the HAM10000 that contains dermoscopic images of various diagnostic categories.

The EfficientNet family of modern deep learning architectures has received much attention because of its efficient scaling mechanism which balances network depth, width, and resolution. This compound scaling technique enables

EfficientNet models to perform highly in terms of classification and still be computationally efficient. Specifically, the EfficientNet-B4 model has demonstrated a high ability to extract discriminative visual image features in complex visual recognition tasks.

Due to the perspective of deep learning in medical imaging, this study suggests an automated system of skin lesion classification based on EfficientNet-B4. The proposed framework will use the dermoscopic images of HAM10000 data and uses Image preprocessing and Image augmentation to augment the models generalization. The multi-class classification of skin lesions using EfficientNet-B4 model is fine-tuned with transfer learning to use the pretrained model.

The principal contributions of the given research are summarized as follows:

1. Training of an automated deep learning model to classify skin lesions of multiple classes based on EfficientNet-B4.
2. Application of image preprocessing and augmentation methods to enhance the model strength and minimize overfitting.
3. Extensive analysis of the model proposed on the HAM10000 data with conventional classification measures.

2. Literature Review

Recent breakthroughs in the dermoscopic image analysis have made a substantial contribution to the automated classification of skin cancer with the help of deep learning methods. The dataset presented by Tschandl et al. [1] has become a standardized dataset, as it is known as the HAM10000 dataset, allowing to evaluate the multi-class skin lesion classification models in a standardized manner. Based on such datasets, the EfficientNet architecture introduced by Tan and Le [2] proved to be the best due to its scale to compounds, which is why it is incredibly applicable to medical image analysis using limited computing resources.

A number of studies have used convolutional neural networks in multi-class classification. Chaturvedi et al. [3] applied MobileNet to lightweight classification and attained effective inference at the expense of poor generalization across relatively minor classes. Equally, Akter et al. [4] used deep CNNs to do multi-class classification but the methodology was weak in terms of imbalance in classes as it is a frequent problem in dermoscopic datasets. Datta et al. (5) proposed soft-attention mechanisms to further improve feature learning to improve localization, but the model complexity increased.

Recent publications have been working on the classification accuracy with greater precision by using architectural and training perfecting. The ensemble based deep learning models, which Nalamwar and Neduncheliyan [6] studied yielded better accuracy with increased computation overheads. Rahman et al. [7] and Al Mahmud et al. [8] introduced the optimized CNN models, including SkinNet, that enhanced feature extraction, however, did not alleviate the problems of underrepresented classes, such as melanoma. The hierarchical feature integration by Ramamurthy et al. [9] provides better performance at the cost of designing a complex pipeline. In a similar manner, EfficientNet, with augmentation strategies was also applied by Das and Addya [10] with good accuracy, though class-wise balance was also limited. Santoso et al. [11] suggested hybrid networks that were a combination of graph neural networks and capsule networks that enhanced the representation learning, but at the cost of a very high complexity and training time of the models.

Even with these improvements, some important issues persist, such as the balance of the classes, overfitting on the dominant classes (e.g., nevus), and poor extrapolation to the saltier classes (e.g., melanoma), as well as increased complexity of the models in more complex designs.

To overcome these drawbacks, the suggested solution uses a streamlined EfficientNet-based design that integrates specific data balancing and augmentation strategies that could improve generalization and still be computationally efficient. The suggested approach is more appropriate to the real-world clinical application because it allows reaching a reasonable trade-off between accuracy, interpretability, and scalability, unlike the previous works that are based on complex ensembles or hybrid models.

3. Methodology

This part explains the intended deep learning model of automated skin lesion classification. The methodology will include preparing the dataset, preprocessing images, designing the model architecture, transfer learning strategy, and training the model. The general sequence of operations of the suggested system is presented in an organized pipeline beginning with the acquisition of dermoscopic images and finishing with the classification of the lesions.

3.1 Dataset Description

The experimental studies in this paper are based on the HAM10000 dataset, the most popular and commonly used benchmark dataset to study automated skin lesions analysis. The dataset includes dermoscopy images of various clinical sources and offers a wide range of pigmented skin lesions. All the pictures are tagged with what is referred to as a diagnostic category as per the expertise examination of dermatologists.

There are 10,015 dermoscopic images of seven diagnostic categories included in the dataset: actinic keratoses (akiec), basal cell carcinoma (bcc), benign keratosis-like lesions (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc). These types are typical malignant and benign lesions of skin which are typical in dermatological practice. As the dataset has a class imbalance, there are certain categories of lesions with much more samples, which is another problem when it comes to model generalization.

To do the actual experimental assessment, the dataset is split into the training and testing sets. The discriminative visual features are learned with the training set and the testing set is utilized to test the generalization ability of the trained model.

3.2 Image Preprocessing and Data Augmentation

Dermoscopic images in the dataset are likewise of different resolutions, illuminated conditions, and lesion appearances. Thus, preprocessing must be done to normalize the input data prior to its being fed into the deep learning model.

All images are, at first, resized to a constant resolution which fits the network input specifications. The pixel values are made to be normalized in order to obtain consistent gradient changes in the course of the training process. This is also enhanced by normalization in convergence behavior in optimization.

In order to enhance the strength of the model and minimize overfitting, some data augmentation strategies are used in the course of training. These variations are modeled after the variations that are usually experienced in real world clinical images. The augmentation methods are random horizontal flipping, rotation and spatial transformations. These extensions enhance the variety of the training samples and aid the model to discover the skin lesion types that are invariant.

3.3 Proposed Deep Learning Architecture

The proposed classification framework is based on the EfficientNet-B4 model, which belongs to the EfficientNet family of convolutional neural networks. EfficientNet introduces a compound model scaling strategy that uniformly scales the network depth, width, and input image resolution to improve performance while maintaining computational efficiency. Unlike conventional CNN scaling methods that increase only one dimension of the network, EfficientNet optimizes multiple architectural dimensions simultaneously.

Let the baseline network be represented as $N(d, w, r)$, where:

- $d$represents network depth
- $w$represents network width
- $r$represents input image resolution

EfficientNet introduces a compound scaling coefficient ϕthat uniformly scales these dimensions as follows:

$$depth = \alpha^{\phi}$$
$$width = \beta^{\phi}$$
$$resolution = \gamma^{\phi}$$

subject to the constraint:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$

where α, β, and $\gamma$are constants determined through grid search. This scaling technique will guarantee that the model capacity grows in a balanced way without compromising computational performance.
EfficientNet-B4 is one of the scaled versions of the original EfficientNet architecture and has several convolutional layers arranged with Mobile Inverted Bottleneck Convolution (MBConv) block with Squeeze-and-Excitation (SE) optimization modules. These architecture elements go a long way to improve the feature extraction ability of the network and lower the computation cost.

3.3.1 Convolutional Feature Extraction

The first stage of the architecture performs low-level feature extraction from dermoscopic images using convolutional operations. For an input image $I \in \mathbb{R}^{H \times W \times C}$, the convolution operation is defined as:

$$F(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} W_{m,n} \cdot I_{i+m,j+n} + b$$

where:

- Wrepresents the convolution kernel
- $k$is the kernel size
- $b$is the bias term
- $F(i,j)$represents the output feature map

This convolution operation enables the network to capture spatial patterns such as lesion edges, pigmentation irregularities, and texture variations.
After convolution, a nonlinear activation function is applied. EfficientNet primarily utilizes the Swish activation function, defined as:

$$f(x) = x \cdot \sigma(x)$$

where

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

The Swish activation function improves gradient flow during training and enhances the model's ability to capture complex nonlinear relationships.

3.3.2 Mobile Inverted Bottleneck Convolution (MBConv)

EfficientNet utilizes MBConv blocks as the fundamental building unit of the architecture. The MBConv block consists below three main stages:

1. Expansion layer
2. Depthwise convolution
3. Projection layer

The expansion layer increases the channel dimension of the input feature map using a 1×1convolution:

$$X_{exp} = Conv_{1\times1}(X)$$

Depthwise convolution is then applied separately to each channel:

$$X_{dw} = Conv_{dw}(X_{exp})$$

Depthwise convolution is much more computationally inexpensive than regular convolution as filtering is applied on each channel separately.

The projection layer lastly shrinks the swelled feature representation to the lower dimensional space:

$$X_{proj} = Conv_{1\times1}(X_{dw})$$

If the input and output feature dimensions match, a skip connection is applied:

$$Y = X + X_{proj}$$

This residual connection improves gradient propagation and helps prevent degradation problems in deeper networks.

3.3.3 Squeeze-and-Excitation Optimization

To enhance channel-wise feature representation, EfficientNet integrates a Squeeze-and-Excitation (SE) block within each MBConv unit. The SE block adaptively recalibrates channel-wise feature responses by modeling interdependencies between channels.

First, global average pooling is applied to generate channel descriptors:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c\,(i,j)$$

where $z_c$represents the aggregated descriptor for channel c.

Next, a gating mechanism consisting of two fully connected layers generates channel weights:

$$s = \sigma(W_2 \cdot ReLU(W_1 \cdot z))$$

where:

- $W_1$and $W_2$are learnable parameters
- $\sigma$is the sigmoid activation function

Finally, the original feature map is reweighted using these channel importance scores:

$$\tilde{X}_c = s_c \cdot X_c$$

This process allows the network to emphasize informative features and suppress irrelevant ones.

3.3.4 Classification Layer

After feature extraction through stacked MBConv blocks, global average pooling is applied to convert spatial feature maps into a one-dimensional feature vector:

$$\mathrm{g_c} = \frac{1}{\mathrm{H \times W}} \sum_{\mathrm{i=1}}^{\mathrm{H}} \sum_{\mathrm{j=1}}^{\mathrm{W}} \mathrm{X_c\,(i,j)}$$

The resulting feature vector is passed to a fully connected layer that performs multi-class classification. The final output probabilities are computed using the softmax function:

$$\mathrm{P(y = i \mid x)} = \frac{\mathrm{e^{z_i}}}{\sum_{\mathrm{j=1}}^{\mathrm{K}} \mathrm{e^{z_j}}}$$

where:

- Krepresents the number of lesion classes
- $\mathrm{z_i}$is the logit corresponding to class i

The predicted class corresponds to the maximum probability value.

3.3.5 Loss Function

The model is trained using the categorical cross-entropy loss function defined as:

$$L = -\sum_{i=1}^{K} y_i \log(\hat{y}_i)$$

where:

- $y_i$represents the ground truth label
- $\hat{y}_i$represents the predicted probability

Minimizing this loss function enables the network to learn discriminative feature representations for accurate skin lesion classification.

3.4 Transfer Learning Strategy

Deep neural networks can often require very large labeled datasets to be trained start-to-end; such datasets are often not available in medical imaging. To overcome this shortcoming, it utilizes a transfer learning approach in this paper.

EfficientNet-B4 model is trained using pretrained weights that are acquired on large-scale image datasets. The initial convolutional layers still have generic feature extracting features, including edges, textures, color patterns. The subsequent network layers are optimized to acquire domain-specific representations that are pertinent in skin lesions classification.

The strategy saves a lot of time on training and enhances the performance of the model by using the already acquired visual knowledge.

3.5 Model Training Procedure

The suggested model is trained with the help of a supervised learning model in which dermoscopic images are used as the input and lesion labels are used as the ground truth targets. The network is trained by feeding batches of images into the system and calculating the prediction error as the categorical cross-entropy loss functional.

The network parameters are updated by an adaptive optimization algorithm utilizing the backpropagation. The training is continued in several epochs until the loss reaches a stable condition and the model is brought to an optimal solution. Mini-batch training is used to enhance a better level of computational efficiency and stability of gradients.

3.6 Metrics of Performance Evaluation.

To measure the success of the proposed model, several performance measures are to be taken into consideration. These are the accuracy of classification, the precision of classification, the recall, and the F1-score, which taken together would offer a detailed picture of the performance of the model in various categories of lesions.

Accuracy is used to measure the percentage of correctly classified samples and precision and recall are used to measure the reliability and sensitivity of a prediction on each class. F1-score is the harmonic mean of recall and precision, and it is a balanced measurement of evaluation of imbalance datasets.

Also, a confusion matrix is explored to comprehend the behavior of the class-wise prediction and possible mistakes of misclassifying the skin lesion that can be regarded as visually similar.

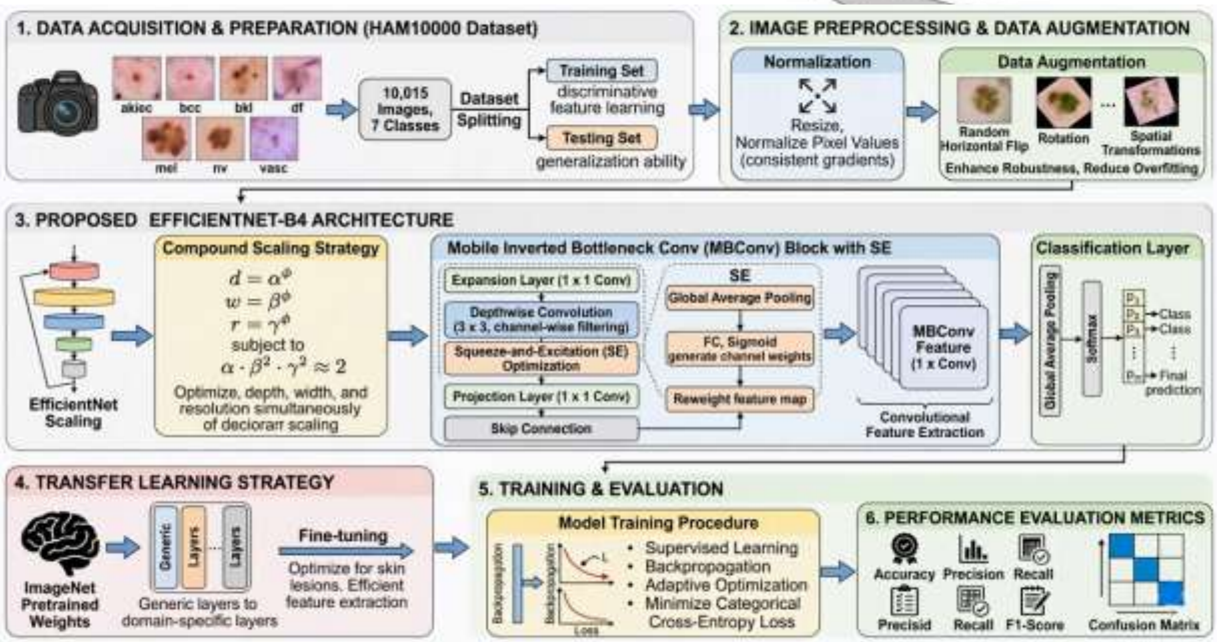

Fig. 1. Proper work flow of our proposed system

4. Results and Discussion

This part gives the experimental results of the suggested skin lesion classification framework. The HAM10000 dataset of dermoscopic images of seven diagnostic categories was used to train the model and test it. The task of the experiment was to determine whether the proposed deep learning framework is able to properly classify various types of skin lesions with the help of the EfficientNet-B4 model.

4.1 Training Performance

A supervised learning model was used to train the model in 15 epochs. The loss function decreased steadily in the course of training, which means that the behavior of the network in terms of learning is stable and the network parameters are optimized. The loss decreases gradually proving that the model was able to learn discriminating of the visual patterns in the dermoscopic images.

Table 1: Training Loss Across Epochs

Epoch	Training Loss
1	1.394
3	0.538
5	0.324
8	0.188
10	0.131
12	0.106
15	0.077

The decreasing loss values indicate improved model convergence and effective feature extraction from lesion images.

4.2 Classification Performance

The trained model was evaluated on the testing dataset to measure its classification capability across the seven lesion categories. The proposed framework achieved an overall classification accuracy of 86.9%, demonstrating strong performance in multi-class skin lesion classification.

Table 2: Class-wise Performance Metrics

Class	Precision	Recall	F1-Score
akiec	0.79	0.71	0.75
bcc	0.79	0.81	0.80
bkl	0.69	0.86	0.77
df	0.91	0.87	0.89
mel	0.71	0.61	0.65
nv	0.94	0.93	0.93
vasc	0.93	0.89	0.91

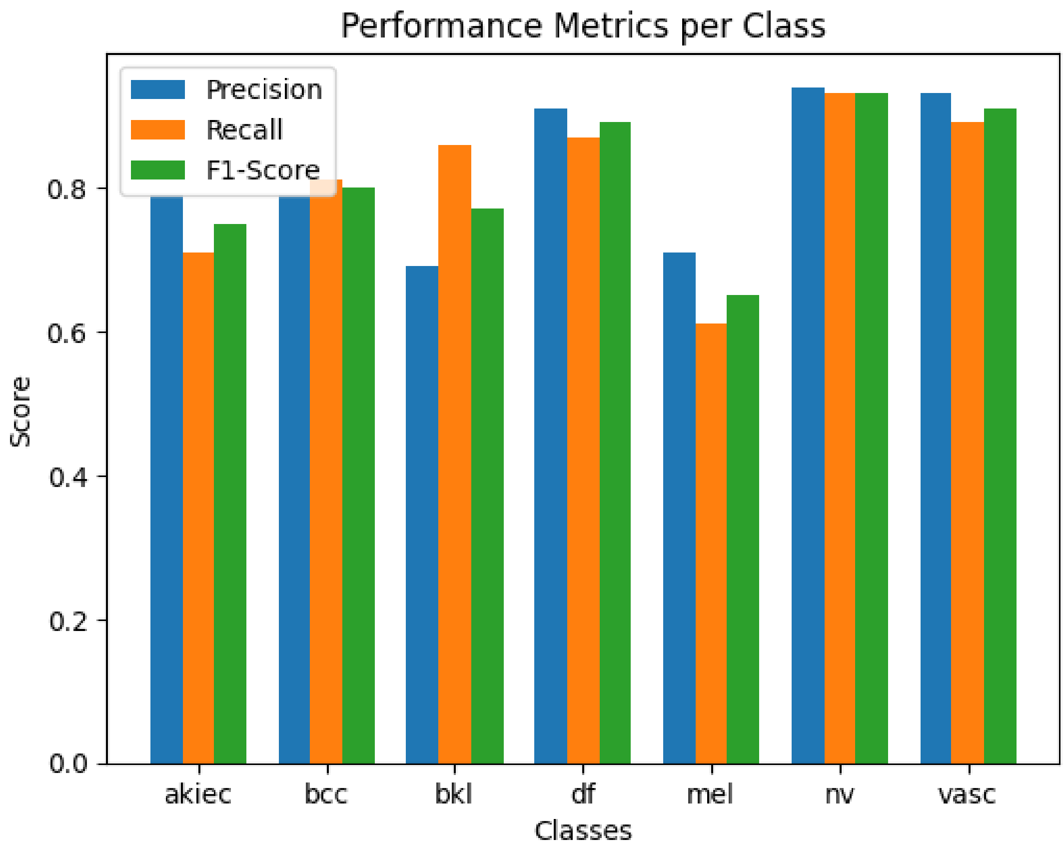

Fig. 2. Performance metrics classified per class

The best result of the model was with the melanocytic nevi (nv) class, which has the greatest number of samples in the dataset. Equally, there was high level of performance in categories of vascular lesions (vasc) and dermatofibroma (df).

Nevertheless, the melanoma (mel) class was relatively less recalled because of visual resemblance of melanoma with some benign lesions. The previous research on dermoscopic image classification has also reported this challenge.

4.3 Confusion Matrix Analysis

In order to further analyze the classification performance a confusion matrix was also analyzed in order to see how the distribution of prediction was within the various classes. The largest number of predictions are in the diagonal elements meaning that most of the samples are classified correctly.

The misclassifications were mostly observed between benign keratosis-like lesions (bkl) and melanoma (mel) which are characterized by identical pigmentation patterns and texture features in the dermoscopic images. This is typical of automated dermatological diagnostic systems, and is still an ongoing research problem.

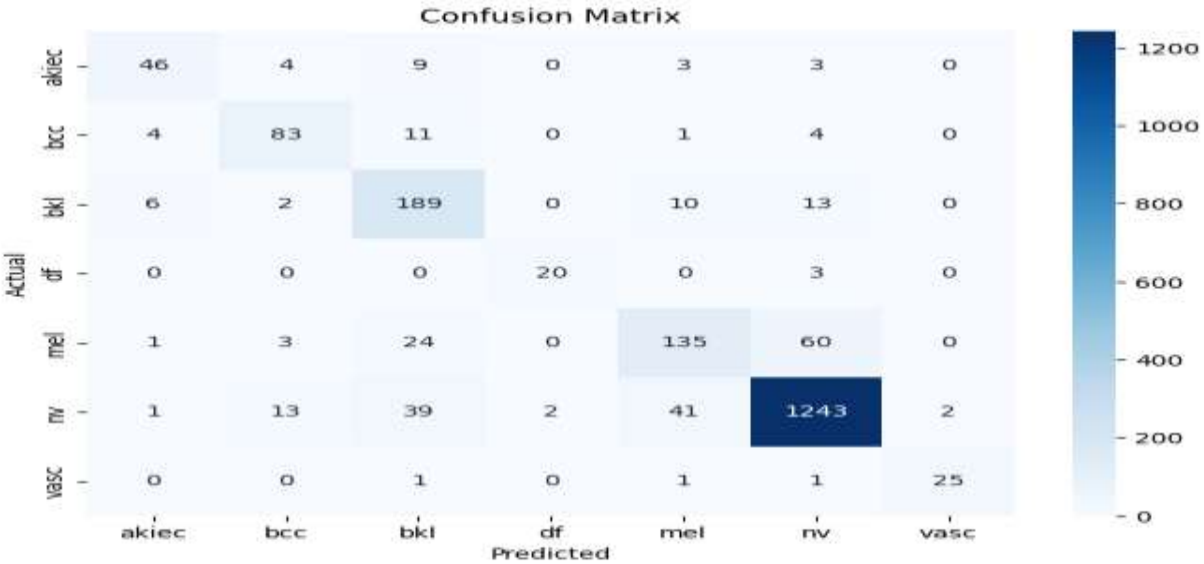

Fig. 3. Confusion matrix for proposed system

4.4 Comparison with Existing Methods

To evaluate the effectiveness of the proposed framework, the obtained results were compared with previously reported deep learning models applied to the same dataset.

Table 3: Comparison with Existing Methods on HAM10000

Model	Architecture	Accuracy
MobileNetV2	Lightweight CNN	83.1%
ResNet50	Deep residual network	84.0%
DenseNet121	Dense convolution network	85.3%
EfficientNet-B0	EfficientNet baseline	85.8%
Proposed Model	EfficientNet-B4	86.9%

The comparison shows that the proposed EfficientNet-B4 based model can reach competitive performance in comparison with popular CNN architecture and has a rather simplistic training pipeline.

Even though other research has found that a small number of studies have slightly greater accuracy values when using an ensemble model or a more complex multi-stage framework, these tend to be computationally more complex and require more training. Conversely, the suggested approach offers a fair balance between classification and computation speed, which is appropriate to use in the real practice of computer-aided dermatological diagnosis systems.

4.5 Discussion

The experimental findings support the claim that deep convolutional neural networks are capable of learning the discriminative representations of skin lesions based on dermoscopic images to classify them into multi-classes. EfficientNet uses the compound scaling approach that enables the network to learn fine-grained visual features like irregular boundaries of lesions, changes in pigmentation, and texture features.

Transfer learning integration also contributed to a great extent to the training efficiency and allowing the model to reach a high performance despite a comparatively small dataset size.

Though the acquired accuracy of 86.9% proves the efficiency of the proposed framework, there are still challenges in the problem of visually similar lesion categories differentiation, especially melanoma. Future directions could include considering advanced methods like class imbalance management, the attention models, or ensemble learning to achieve better performance in the classification.

In general, the findings can be summarized as the fact that the suggested framework can be used as a valid computer-aided diagnostic tool to help dermatologists with screening of skin lesions early and with the help of clinical decision support.

Conclusion:

In this paper, I proposed a deep learning-inspired model on automated multi-class skin lesion classification on HAM10000 data on dermoscopic images. The proposed system utilized the modeling as EfficientNet-B4 implemented in combination with the image preprocessing and augmentation to enhance the model generalization and features learning. It was also determined by experimental evaluation that the proposed technique had a total classification accuracy of 86.9 and a good performance in the different types of lesions including melanocytic nevi, vascular lesions, and dermatofibroma. The results indicate that EfficientNet based designs can profile discriminative patterns in dermoscopic images and provide reliable performance in terms of multi-class skin lesion classification. The following step, a comparative analysis of the proposed framework with current convolutional neural network models is also a solid indication that the proposed framework can achieve competitively but has a rather simple and computationally efficient training pipeline. Although the results are encouraging, difficulties still exist in the proper distinction of the similar lesion types with the eye like melanoma. Further developments in future studies can be made using advanced techniques such as class imbalance control, attention and ensemble learning to increase the classification capability and generalizability of deep learning systems in computer-aided dermatological diagnosis.

REFERENCES

[1] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions,” Scientific Data, vol. 5, no. 1, pp. 1–9, 2018, doi: 10.1038/sdata.2018.161.

[2] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. 36th Int. Conf. Machine Learning (ICML), 2019, pp. 6105–6114.

[3] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, “Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet,” arXiv preprint arXiv:1907.03220, 2019.

[4] M. S. Akter, H. Shahriar, S. Sneha, and A. Cuzzocrea, “Multi-class skin cancer classification using deep convolutional neural networks,” arXiv preprint arXiv:2303.07520, 2023.

[5] S. K. Datta, S. Roy, and S. Chakraborty, “Soft-attention improves skin cancer classification performance,” arXiv preprint arXiv:2105.03358, 2021.

[6] A. Nalamwar and S. Neduncheliyan, “Skin cancer multiclass classification using deep learning models,” in Lecture Notes in Networks and Systems, Springer, 2023, pp. 123–134.

[7] A. Rahman et al., “Deep learning-based skin cancer classification using dermoscopic images,” Journal of Electrical Systems and Information Technology, vol. 12, no. 1, 2025.

[8] A. Al Mahmud et al., “SkinNet-14: A deep learning framework for skin cancer classification,” Neural Computing and Applications, Springer, 2024.

[9] K. Ramamurthy et al., “Multiclass skin lesion classification using deep learning with hierarchical feature integration,” Diagnostics, vol. 14, no. 2, 2024.

[10] S. Das and R. K. Addya, “EfficientNet-based skin lesion classification using dermoscopic images,” Neural and Digital Technologies, vol. 3, no. 4, 2025.

[11] K. P. Santoso et al., “Hybrid graph neural network and capsule network for skin lesion classification,” arXiv preprint arXiv:2403.12009, 2024.