Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

Article - e003

COMPARATIVE ANALYSIS OF TRANSFER LEARNING-BASED DEEP LEARNING MODELS FOR SKIN CANCER CLASSIFICATION USING THE HAM10000 DATASET

Harshal Hingarh[1], Dr. Lalji Prasad[2]

[1,2] Department of Advanced Computing Specialization, SAGE University, Indore, India

Received: 15/05/2026 **Revision Received**: 12/06/2026 **Accepted**: 22/06/2026

ABSTRACT

Skin cancer is among the most common forms of cancer worldwide, and early diagnosis is essential for improving patient survival rates. Recent advances in deep learning and computer vision have enabled the development of automated diagnostic systems capable of assisting dermatologists in skin lesion classification. This study presents a comparative analysis of transfer learning-based deep learning models for multiclass skin cancer classification using the HAM10000 dataset, which contains dermoscopic images belonging to seven different skin lesion categories. Five state-of-the-art convolutional neural network architectures, namely ResNet50, ResNet101, ResNet152, EfficientNetB0, and EfficientNetB7, were investigated. The models were trained using transfer learning with ImageNet pretrained weights and data augmentation techniques to improve generalization performance.

Experimental evaluation was conducted using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics. Among the evaluated models, EfficientNetB0 achieved the highest test accuracy of 75.45% and validation accuracy of 77.56%, while EfficientNetB7 obtained the highest ROC-AUC score of 94.08%, demonstrating excellent discriminative capability. ResNet50 achieved competitive performance with a validation accuracy of 74.83%, whereas ResNet152 exhibited comparatively lower performance. The results indicate that EfficientNet architectures provide superior feature extraction and classification performance for dermoscopic image analysis.

The findings demonstrate the effectiveness of transfer learning for automated skin cancer detection and highlight the potential of deep learning-based computer-aided diagnostic systems in supporting dermatologists for early and accurate skin lesion classification. Future work will focus on class imbalance handling, explainable artificial intelligence techniques, and ensemble learning approaches to further improve classification performance.

valuation, modelling, and prediction of client satisfaction are crucial business strategies that address customer dissatisfaction, promote retention, and enhance organisational reputation.

KEYWORDS: **:** Skin Cancer Detection, Deep Learning, Transfer Learning, HAM10000 Dataset, ResNet, EfficientNet, Dermoscopic Image Classification.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

1. INTRODUCTION

1.1 Background and Motivation

Skin cancer is one of the most prevalent and rapidly increasing forms of cancer worldwide. It occurs due to the abnormal growth of skin cells and is primarily caused by excessive exposure to ultraviolet (UV) radiation. Skin cancer can be categorized into melanoma and non-melanoma skin cancers. Among these, melanoma is considered the most dangerous type because of its high potential to spread to other parts of the body. Early detection and timely treatment significantly increase survival rates and reduce mortality associated with the disease.

Traditional diagnosis of skin cancer relies on clinical examination, dermoscopy, and histopathological analysis performed by experienced dermatologists. Although these methods are effective, accurate diagnosis often depends on the expertise and experience of medical professionals. Furthermore, manual examination can be time-consuming and may lead to inter-observer variability, especially when distinguishing between visually similar skin lesions. These limitations have motivated researchers to explore automated computer-aided diagnostic systems capable of assisting healthcare professionals in skin cancer detection.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have revolutionized medical image analysis. In particular, Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in image classification tasks due to their ability to automatically learn hierarchical features directly from images. Transfer learning has further improved the effectiveness of CNN-based models by utilizing knowledge learned from large-scale datasets such as ImageNet and adapting it to specific medical imaging applications.

Several deep learning architectures have been proposed for skin lesion classification, including VGGNet, InceptionNet, ResNet, DenseNet, MobileNet, and EfficientNet. Among these architectures, ResNet and EfficientNet have gained significant attention because of their strong feature extraction capabilities and efficient model design. ResNet utilizes residual connections to overcome the vanishing gradient problem in deep neural networks, while EfficientNet introduces compound scaling to balance network depth, width, and image resolution, resulting in improved performance with fewer parameters.

The HAM10000 dataset has become one of the most widely used benchmark datasets for automated skin lesion classification. It contains over 10,000 dermoscopic images categorized into seven clinically important skin lesion classes, including melanoma, basal cell carcinoma, benign keratosis, dermatofibroma, vascular lesions, actinic keratosis, and melanocytic nevi. The diversity and complexity of this dataset make it suitable for evaluating the effectiveness of deep learning models in real-world diagnostic scenarios.

1.2 Problem Statement

In this study, a comparative analysis of transfer learning-based deep learning architectures is conducted using the HAM10000 dataset. Five pretrained models, namely ResNet50, ResNet101, ResNet152, EfficientNetB0, and EfficientNetB7, are evaluated for multiclass skin lesion classification. Data augmentation techniques are applied to improve model generalization and reduce overfitting. Performance is assessed using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics.

1.3 Research Objectives

The primary objectives of this research are:

1. To develop an automated skin lesion classification framework using transfer learning.
2. To compare the performance of ResNet and EfficientNet architectures on the HAM10000 dataset.
3. To identify the most effective deep learning model for multiclass skin cancer classification.
4. To evaluate the applicability of deep learning techniques for supporting dermatologists in early skin cancer diagnosis

The outcomes of this study contribute to the growing field of AI-assisted healthcare and provide insights into the effectiveness of modern deep learning architectures for skin cancer detection.

2. LITERATURE REVIEW

Skin cancer classification using artificial intelligence has attracted significant research attention due to its potential to assist dermatologists in early diagnosis. Recent studies have demonstrated the effectiveness of deep learning models, particularly Convolutional Neural Networks (CNNs), in analyzing dermoscopic images and identifying malignant lesions with high accuracy.

Thwin and Park (2024) proposed a deep ensemble model for multiclass skin lesion classification. Their approach combined multiple deep learning architectures to improve classification robustness. Experimental results demonstrated improved performance compared to individual models; however, the study highlighted challenges related to cross-dataset generalization and dependency on dataset quality.

Choi et al. (2024) introduced an ABC ensemble framework integrating EfficientNet variants and SeReNeXt architectures. The proposed method improved balanced classification accuracy on ISIC-style datasets. Although the ensemble achieved strong performance, the increased computational complexity limited its suitability for resource-constrained environments.

Al Mahmud et al. (2024) developed SkinNet-14, a lightweight deep learning framework optimized for low-resolution dermoscopic images. Their results showed high classification accuracy with reduced training time, making the model computationally efficient. However, the authors noted that low-resolution images may omit important diagnostic features required for melanoma detection.

Behara et al. (2024) proposed a hybrid framework combining Active Contour Snake Models with Lightweight Attention-Guided Capsule Networks. The method improved lesion boundary extraction and classification accuracy. Nevertheless, performance degradation was observed in images containing artifacts such as hair and uneven illumination.

Nawaz et al. (2025) investigated the impact of class imbalance on skin lesion classification and employed augmentation techniques together with class-weighted training. Their findings indicated improved detection rates for minority classes; however, external validation remained limited.

Fiaz et al. (2025) introduced an explainable hybrid deep learning framework integrating lesion segmentation, multiclass classification, and Grad-CAM visualization. The approach enhanced

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

interpretability by providing visual explanations for predictions. However, the reliability of explanation maps in clinical practice requires further investigation.

Alruwaili and Mohamed (2025) proposed a fusion architecture combining EfficientNet and ResNet models for skin lesion classification. Their study reported improved stability and classification performance through model fusion, although increased model size created deployment challenges.

Recent studies consistently indicate that transfer learning-based CNN architectures outperform traditional machine learning methods. ResNet architectures effectively address vanishing gradient problems through residual learning, while EfficientNet models achieve superior accuracy by balancing network depth, width, and resolution. Despite these advancements, challenges related to class imbalance, dataset diversity, explainability, and clinical deployment remain unresolved.

The present study addresses these challenges by performing a comparative analysis of ResNet50, ResNet101, ResNet152, EfficientNetB0, and EfficientNetB7 using the HAM10000 dataset. The objective is to identify the most effective transfer learning architecture for multiclass skin lesion classification while evaluating performance through accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC metrics

2.1 Comparative Analysis of existing studies

Table 1 presents a structured comparative synthesis of representative works reviewed in this survey:

Table 1

Author (Year)	Method	Dataset	Key Findings	Limitation
Thwin & Park (2024)	Deep Ensemble	Dermoscopy Images	Improved robustness	Poor cross-dataset validation
Choi et al. (2024)	ABC Ensemble	ISIC	Better balanced accuracy	High computational cost
Behara et al. (2024)	Capsule Network + Snake Model	Skin Lesion Images	Better lesion segmentation	Sensitive to artifacts
Al Mahmud et al. (2024)	SkinNet-14	Dermoscopy Images	Efficient low-resolution classification	May miss fine lesion details
Nawaz et al. (2025)	CNN + Class Balancing	HAM10000	Improved minority class detection	Limited external validation
Fiaz et al. (2025)	XAI + CNN	Skin Lesion Images	Improved interpretability	Clinical reliability not fully validated
Alruwaili & Mohamed (2025)	EfficientNet + ResNet Fusion	Multi-class Dataset	Strong classification performance	Larger model size

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

3. RESEARCH GAP

Although numerous deep learning approaches have been proposed for skin cancer diagnosis, several limitations still exist. Most studies focus on maximizing classification accuracy while overlooking issues such as class imbalance, model generalization, and computational efficiency. Furthermore, many models are evaluated on a limited number of architectures, making it difficult to determine which transfer learning framework is most suitable for practical deployment.

Existing research has extensively investigated individual architectures such as ResNet, EfficientNet, MobileNet, and ensemble models. However, comprehensive comparative studies evaluating multiple ResNet and EfficientNet variants under identical experimental conditions remain limited. Additionally, minority classes in the HAM10000 dataset continue to exhibit lower classification performance due to severe class imbalance. To address these limitations, the present study performs a systematic comparison of ResNet50, ResNet101, ResNet152, EfficientNetB0, and EfficientNetB7 using identical preprocessing, augmentation, training, and evaluation procedures. The study aims to identify the most effective architecture for multiclass skin lesion classification and provide insights into the strengths and weaknesses of each model

4. IMPLEMENTATION DETAILS

4.1 **Proposed Framework**

The proposed framework for automated skin cancer classification consists of five major stages: dataset acquisition, data preprocessing, data augmentation, transfer learning-based model training, and performance evaluation. The overall workflow of the proposed system is illustrated in Figure 1.

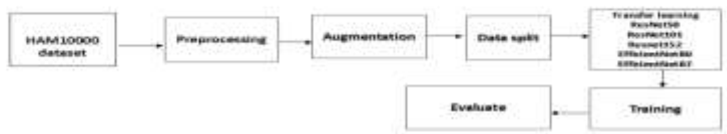


Figure 1. Proposed workflow for transfer learning-based skin cancer classification using the HAM10000 dataset

The objective of the framework is to classify dermoscopic skin lesion images into seven diagnostic categories using deep learning-based transfer learning techniques.

4.2 Dataset Description

The HAM10000 (Human Against Machine with 10,000 Training Images) dataset was used for experimental evaluation. The dataset contains 10,015 dermoscopic images collected from different clinical sources and is one of the most widely used benchmark datasets for skin lesion classification.

The dataset contains seven skin lesion categories:

Class	Description

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

akiec	Actinic Keratoses and Intraepithelial Carcinoma
bcc	Basal Cell Carcinoma
bkl	Benign Keratosis-like Lesions
df	Dermatofibroma
mel	Melanoma
nv	Melanocytic Nevi
vasc	Vascular Lesions

The HAM10000 dataset is highly imbalanced, with Melanocytic Nevi (nv) representing the majority class and Dermatofibroma (df) representing one of the minority classes.

The dataset was obtained from the Kaggle HAM10000 repository and loaded using the accompanying metadata file containing image identifiers and diagnostic labels.

The original HAM10000 dataset contains 10,015 dermoscopic images categorized into seven lesion classes. HAM10000 Dataset

4.3 Data Splitting

The dataset was divided into training, validation, and testing subsets using stratified sampling to preserve class distribution.

Dataset	Number of Images
Training Set	7010
Validation Set	1502
Test Set	1503
Total	10015

Stratified splitting ensured that each lesion category maintained approximately the same proportion across all subsets .

4.4 Data Preprocessing

Prior to model training, several preprocessing operations were applied:

1. Image paths were generated from image identifiers.
2. Images were resized to **224 × 224 pixels**.
3. Labels were converted into categorical format.
4. Pixel normalization was performed using the preprocessing functions associated with the pretrained networks.
5. Images were loaded dynamically using TensorFlow ImageDataGenerator.

These preprocessing steps ensured consistency across all experiments and improved training efficiency.

4.5 Data Augmentation

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

To improve model generalization and reduce overfitting, data augmentation techniques were applied to the training set.

The following augmentation operations were used:

Technique	Value
Rotation Range	20°
Width Shift	0.1
Height Shift	0.1
Zoom Range	0.1
Horizontal Flip	True
Vertical Flip	True

The augmentation process artificially increased data variability and improved the model's ability to learn robust lesion features.

4.6 Transfer Learning Models

Five pretrained deep learning architectures were evaluated:

4.6.1 ResNet50

ResNet50 is a 50-layer residual neural network that employs skip connections to alleviate the vanishing gradient problem and facilitate deeper network training.

4.6.2 ResNet101

ResNet101 extends the residual learning concept by increasing network depth to 101 layers, enabling more complex feature extraction.

4.6.3 ResNet152

ResNet152 is a deeper variant containing 152 layers. The architecture provides greater representational capacity but requires higher computational resources.

4.6.4 EfficientNetB0

EfficientNetB0 employs compound scaling to balance network depth, width, and input resolution. It is designed to achieve high classification performance with fewer parameters.

4.6.5 EfficientNetB7

EfficientNetB7 is the largest architecture within the EfficientNet family and offers enhanced feature extraction capability through increased scaling dimensions

4.7 Model Architecture

All transfer learning models were initialized using pretrained ImageNet weights.

A pretrained backbone network was employed as a feature extractor, followed by a Global Average Pooling layer. Two fully connected layers containing 512 and 256 neurons respectively were added with dropout regularization to reduce overfitting. Finally, a Softmax classification layer with seven output neurons was used to classify skin lesions into seven diagnostic categories.

Figure 2 illustrates the proposed transfer learning architecture used for skin lesion classification

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

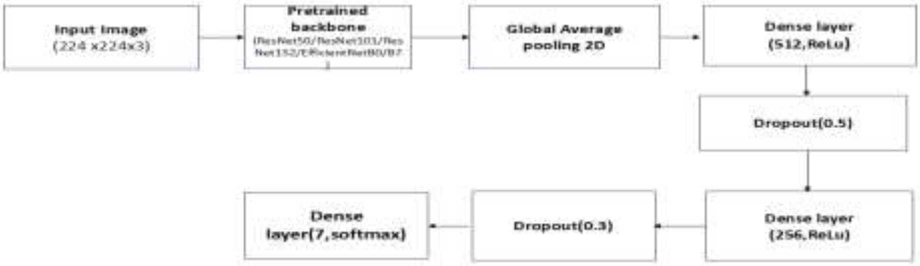


Figure 2.Model architecture

4.8 Training Configuration

The models were trained using TensorFlow/Keras on the Kaggle GPU environment.

Parameter	Value
Optimizer	Adam
Initial Learning Rate	0.0001
Batch Size	32
Epochs	30
Loss Function	Categorical Crossentropy
Activation Function	ReLU
Output Activation	Softmax

To improve convergence and avoid overfitting, the following callbacks were employed:

- ModelCheckpoint
- ReduceLROnPlateau
- EarlyStopping

Different transfer learning architectures exhibited different convergence behavior during training. EfficientNetB0 was trained using a two-stage strategy consisting of 20 epochs of transfer learning followed by 10 epochs of fine-tuning. EfficientNetB7 achieved stable convergence during the initial training phase and therefore was trained for 10 epochs without additional fine-tuning. The best-performing model weights were selected using ModelCheckpoint and EarlyStopping based on validation performance.

4.9 Reproducibility Details

To improve experimental reproducibility, all models were trained using a fixed random seed of 42. TensorFlow, NumPy, and Python random generators were initialized with the same seed to ensure consistent experimental results.

To address class imbalance in the HAM10000 dataset, Class weights were computed from the training dataset for analysis purposes but were not applied during model training.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

Transfer learning was employed with ImageNet pretrained weights. During the initial training stage, the backbone layers were frozen and only the classification layers were trained. For EfficientNetB0, selected backbone layers were subsequently unfrozen for fine-tuning. ModelCheckpoint, EarlyStopping, and ReduceLROnPlateau callbacks were used to retain the best-performing model and reduce overfitting.

4.10 Evaluation Metrics

The trained models were evaluated using multiple performance metrics:

Accuracy

Accuracy measures the proportion of correctly classified samples.

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

Precision

Precision measures the proportion of correctly predicted positive samples.

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

Recall

Recall measures the ability of the model to identify positive samples.

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

F1-Score

The F1-score represents the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Precision, Recall, and F1-score were computed using weighted averaging across all seven lesion classes. Weighted averaging considers the number of samples in each class (class support) when calculating performance metrics. This approach is particularly suitable for the HAM10000 dataset because it contains significant class imbalance among skin lesion categories, providing a more representative evaluation of overall model performance.

ROC-AUC

Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values were used to assess the discriminative capability of the models.

Confusion Matrix

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

A confusion matrix was generated to visualize class-wise prediction performance and misclassification patterns.

Experimental Environment

The experiments were conducted using:

Component	Specification
Platform	Kaggle Notebook
GPU	NVIDIA Tesla T4
Framework	TensorFlow/Keras
Programming Language	Python
Libraries	NumPy, Pandas, Matplotlib, Scikit-Learn, TensorFlow

5. RESULT ANALYSIS AND PERFORMANCE EVALUATION

The performance of the proposed transfer learning models was evaluated using multiple metrics, including accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. These metrics provide a comprehensive assessment of the classification capability of each model on the HAM10000 dataset.

Model	Test Accuracy (%)	Validation Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
ResNet50	71.52	74.83	79.46	71.52	74.01	90.62
ResNet101	64.20	68.04	78.00	64.00	68.00	92.17
ResNet152	61.14	64.71	74.99	61.14	65.27	88.56
EfficientNetB0	75.45	77.56	74.99	75.45	74.01	90.62
EfficientNetB7	71.59	77.43	78.51	65.87	66.38	94.08

Table 2. Overall performance comparison (Weighted Precision, Weighted Recall, and Weighted F1-score)

5.1 Comparative Performance Analysis

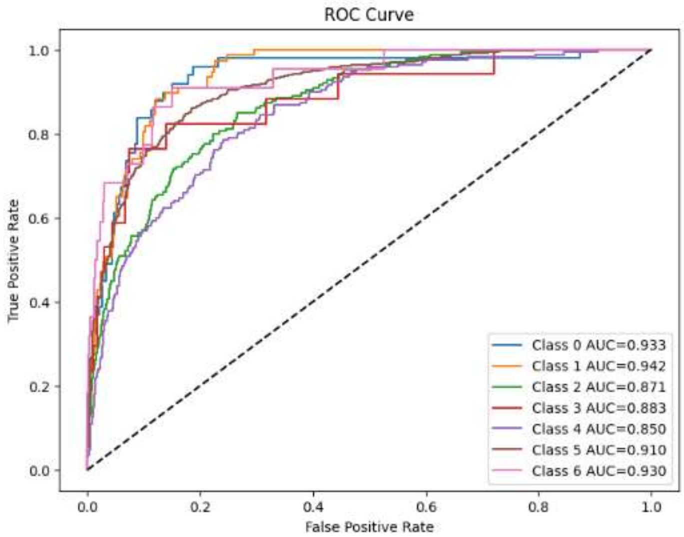
Table 2 presents the overall performance comparison of the investigated deep learning architectures.

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

The results demonstrate that EfficientNetB0 achieved the highest test accuracy of 75.45%, indicating superior classification performance among all evaluated models. EfficientNetB7 achieved the highest ROC-AUC value of 94.08%, reflecting excellent discriminative capability for skin lesion classification.

ResNet50 also demonstrated competitive performance with a validation accuracy of 74.83% and an F1-score of 74.01%. However, ResNet152 showed lower performance, suggesting that increasing model depth alone does not necessarily improve classification accuracy on the HAM10000 dataset.

5.2 ROC Curve Analysis



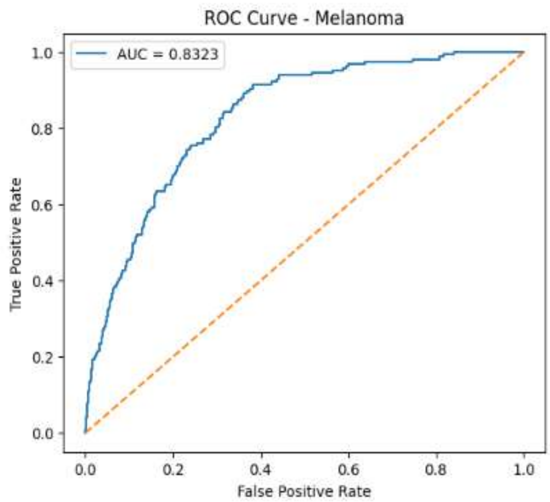


Figure 3. Roc curve (EfficientNetB0) **Figure 4.Roc curve (EfficientNetB7)**

(One-vs-Rest ROC Curve for Melanoma)

The ROC curve analysis demonstrated excellent classification performance, particularly for EfficientNetB7, which achieved an AUC score of 94.08%.

An AUC value greater than 0.90 is generally considered excellent in medical image classification and indicates strong diagnostic reliability. The ROC results confirm that the proposed deep learning framework can effectively distinguish between different skin lesion categories.

5.3 Class-wise Performance Analysis

Table 3. Class-wise Performance Metrics of the EfficientNetB0 Model

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

Class	Precision	Recall	F1-score
akiec	0.75	0.12	0.21
bcc	0.49	0.36	0.42
bkl	0.44	0.58	0.50
df	0.33	0.12	0.17
mel	0.53	0.38	0.44
nv	0.85	0.93	0.89
vasc	1.00	0.18	0.31

Table 3 presents the class-wise performance of the EfficientNetB0 model on the HAM10000 dataset. The model achieved its best performance on the Melanocytic Nevi (nv) class, obtaining a precision of 0.85, recall of 0.93, and F1-score of 0.89. The Benign Keratosis-like Lesions (bkl) class showed moderate performance with an F1-score of 0.50. The Melanoma (mel) and Basal Cell Carcinoma (bcc) classes achieved F1-scores of 0.44 and 0.42, respectively. Lower performance was observed for Actinic Keratoses and Intraepithelial Carcinoma (akiec), Dermatofibroma (df), and Vascular Lesions (vasc), primarily due to class imbalance and the limited number of training samples available for these categories. These results indicate that while the model performs well on majority classes, further improvements are required to enhance recognition of minority lesion classes.

Table 4. Class-wise Performance Metrics of the EfficientNetB7 Model

Class	Precision	Recall	F1-score
akiec	0.83	0.10	0.18
bcc	0.61	0.18	0.28
bkl	0.37	0.30	0.33
df	0.00	0.00	0.00
mel	0.49	0.21	0.29
nv	0.77	0.97	0.86
vasc	0.33	0.09	0.14

Table 4 presents the class-wise performance metrics of the EfficientNetB7 model on the HAM10000 dataset. The model achieved its highest performance for the Melanocytic Nevi (nv) class with a precision of 0.77, recall of 0.97, and F1-score of 0.86. Moderate performance was observed for the Benign Keratosis-like Lesions (bkl) and Melanoma (mel) classes. The Basal Cell Carcinoma (bcc) class achieved an F1-score of 0.28, while the Actinic Keratoses and Intraepithelial Carcinoma (akiec) class obtained an F1-score of 0.18. The model struggled to classify Dermatofibroma (df) and Vascular Lesions (vasc) due to the limited number of samples and class imbalance present in the dataset. Overall, the results indicate that the model performs well on majority classes but requires further optimization to improve recognition of minority skin lesion categories.

5.4 Confusion Matrix Analysis

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

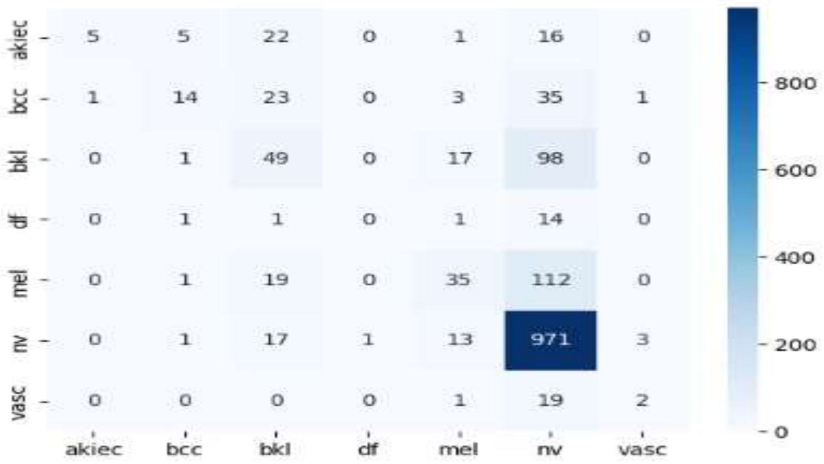


Figure 5. confusion matrix(EfficientNetB0)

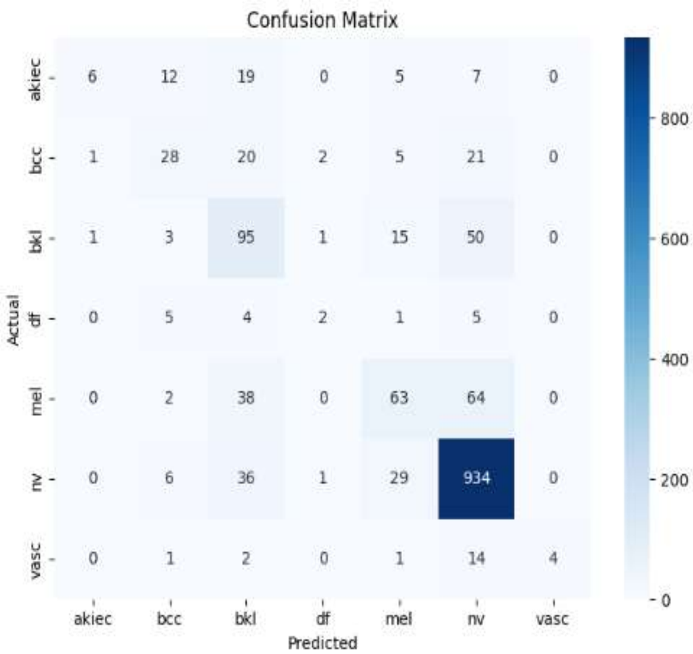


Figure 6.confusion matrix(EfficientNetB7)

The confusion matrix analysis revealed that:

- Melanocytic Nevi (nv) achieved the highest classification accuracy due to its large number of training samples.

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

- Melanoma (mel) and Benign Keratosis (bkl) occasionally exhibited misclassification because of visual similarities between lesion patterns.
- Dermatofibroma (df) and Vascular lesions (vasc) demonstrated lower recall values due to severe class imbalance.

These findings indicate that dataset imbalance remains one of the primary challenges affecting multiclass skin lesion classification.

6. DISCUSSION

An interesting observation from the experiments is that the deepest residual architecture, ResNet152, did not outperform ResNet50. Despite having significantly more layers and parameters, ResNet152 achieved only 61.14% test accuracy compared to 71.52% achieved by ResNet50. This finding suggests that increasing model depth beyond a certain point may not yield substantial performance improvements for the HAM10000 dataset and may instead introduce optimization challenges and overfitting risks. The results indicate that architectural efficiency and feature utilization are more important than network depth alone for skin lesion classification.

6.1 Strength and Clinical Contribution

The proposed research offers several important contributions:

1. A comprehensive comparison of five state-of-the-art transfer learning architectures was performed under identical experimental conditions.
2. The study evaluates both residual learning and compound scaling approaches, providing valuable insights into their effectiveness for skin lesion classification.
3. The proposed framework achieved a maximum classification accuracy of 75.45% and ROC-AUC of 94.08%, demonstrating reliable diagnostic performance.
4. The framework can serve as a computer-aided diagnostic system to assist dermatologists in early skin cancer detection.
5. The use of transfer learning significantly reduces training time and computational requirements compared to training deep neural networks from scratch.

Clinically, such systems can support medical professionals by providing rapid preliminary assessments and reducing diagnostic workload.

6.2 Real-World Application Domains

The proposed skin lesion classification framework can be applied in several real-world healthcare scenarios:

Dermatology Clinics

The model can assist dermatologists by providing preliminary lesion classification before clinical examination.

Telemedicine Platforms

Patients can upload dermoscopic images remotely, allowing AI-assisted screening and consultation.

Rural Healthcare Centers

Regions with limited access to dermatology specialists can benefit from automated diagnostic support systems.

Interdisciplinary Journal of AI, Machine Learning & Data Science (IJAIMLDS)
ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

Mobile Healthcare Applications

Integration into smartphone-based diagnostic applications can facilitate early detection and patient awareness.

Clinical Decision Support Systems

Hospitals and healthcare institutions can incorporate the proposed framework into existing diagnostic workflows to improve efficiency and consistency.

6.3 Limitations and Future Directions

Despite promising results, several limitations remain.

Limitations

1. The HAM10000 dataset is highly imbalanced, resulting in lower classification performance for minority classes.
2. The study utilized a single dataset, limiting the assessment of model generalization across diverse populations.
3. Deep learning models function as black-box systems, making interpretation of predictions challenging.
4. Clinical metadata such as patient age, gender, and lesion location were not incorporated into the classification process.
5. Computational requirements increase significantly for deeper architectures such as EfficientNetB7.

Future Directions

Future research can focus on:

- Advanced data balancing techniques.
- Ensemble learning approaches.
- Explainable Artificial Intelligence (Grad-CAM, SHAP).
- Vision Transformer (ViT) architectures.
- Multi-modal learning using image and clinical metadata.
- Cross-dataset validation using HAM10000 and ISIC datasets.

These improvements may further enhance diagnostic accuracy and clinical applicability.

7.CONCLUSION

This study presented a comparative analysis of transfer learning-based deep learning architectures for multiclass skin cancer classification using the HAM10000 dataset. Five pretrained models, namely ResNet50, ResNet101, ResNet152, EfficientNetB0, and EfficientNetB7, were evaluated using standardized preprocessing, augmentation, and training procedures.

Experimental results demonstrated that EfficientNetB0 achieved the highest classification accuracy of **75.45%**, while EfficientNetB7 achieved the highest **ROC-AUC score of 94.08%**. ResNet50 provided competitive performance with a test accuracy of **71.52%**, whereas ResNet152 achieved a test accuracy of **61.14%** and an AUC of **88.56%**. These findings indicate that

ISSN: 3139-3527 | July-September-Issue, Vol. 1, No. 3 (2026) | DOI: 10.66261/nsxq3430

EfficientNet architectures provide superior feature extraction and classification capability for dermoscopic image analysis compared with deeper residual networks.

The study confirms that transfer learning provides an effective solution for automated skin lesion analysis and can serve as a valuable decision-support tool for dermatologists. The obtained results highlight that architectural efficiency and optimized feature utilization are more important than simply increasing network depth for skin lesion classification tasks.

Although challenges related to class imbalance, minority-class detection, and model interpretability remain, the proposed framework demonstrates significant potential for assisting early skin cancer detection and improving diagnostic efficiency. Future work will focus on explainable artificial intelligence techniques, ensemble learning approaches, Vision Transformer architectures, and multi-modal frameworks incorporating clinical metadata.

The outcomes of this research contribute to the advancement of AI-assisted healthcare systems and provide a foundation for the development of reliable, accurate, and clinically deployable skin cancer diagnosis frameworks.

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.

[2] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of CVPR, pp. 770–778, 2016.

[4] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” Proceedings of ICML, pp. 6105–6114, 2019.

[5] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” ICLR, 2015.

[6] F. Chollet, Deep Learning with Python. Manning Publications, 2021.

[7] H. Hingarh and L. Prasad, “Review on Machine Learning and Deep Learning Approaches for Automated Skin Cancer Diagnosis,” 2025.

[8] M. Thwin and J. Park, “Deep ensemble learning for skin lesion classification,” IEEE Access, 2024.

[9] Y. Choi et al., “ABC Ensemble Framework for Skin Lesion Analysis,” Expert Systems with Applications, 2024.

[10] A. Al Mahmud et al., “SkinNet-14: Lightweight Deep Learning for Skin Cancer Classification,” Computers in Biology and Medicine, 2024.