European Journal of Computer Science and Information Technology, 13 (1), 63-71, 2025

Print ISSN: 2054-0957 (Print),

Online ISSN: 2054-0965 (Online)

Website: https://www.eajournals.org/

Publication of the European Centre for Research Training and Development -UK

A Deep Learning Framework for Multi-Class Skin Cancer Diagnosis: Addressing Class Imbalance and Performance Optimization Through Convolutional Neural Network Architectures

Md Imamul Haque[1], Karima Khatun[2], Golam Rohman[3], Abdullah Al Rakib[4], Muhammad Maruf Billah[5]

[1,3,4&5]School of Artificial Intelligence, [2]School of Electrical Engineering & Automation

Nanjing University of Information Science & Technology, Nanjing, China.

doi: https://doi.org/10.37745/ejcsit.2013/vol13n16371 Published January 25, 2025

Citation: Haque M.I., Khatun K., Rohman G., Al Rakib A., and Billah M.M. (2025) A Deep Learning Framework for Multi-Class Skin Cancer Diagnosis: Addressing Class Imbalance and Performance Optimization Through Convolutional Neural Network Architectures, *European Journal of Computer Science and Information Technology*, 13 (1), 63-71

Abstract: *Skin cancer, particularly melanoma, is one of the most fatal forms of cancer worldwide, necessitating early detection for effective treatment. The HAM10000 dataset is used as a standard for skin lesion analysis in this study to use a Convolutional Neural Network (CNN) to sort seven different types of skin cancer into groups. The dataset consists of over 10,000 labeled dermoscopic images representing various classes such as melanoma, basal cell carcinoma, and benign keratosis. The methodology includes preprocessing techniques like resizing, normalization, and augmentation to enhance the diversity of the data and address class imbalance. The Adam optimizer, with a learning rate of 0.0001, trained a custom CNN model with multiple convolutional layers, batch normalization, and ReLU activations. The model's performance was evaluated across training, validation, and testing subsets, achieving a training accuracy of 98.17% and a validation accuracy of 99.60% after 10 epochs. Evaluation tools, such as a classification report and a confusion matrix, showed high accuracy for common classes like melanocytic nevi but showed mistakes in underrepresented classes like vascular lesions and dermatofibroma. Compared with state-of-the-art techniques like ResNet and EfficientNet, the proposed model demonstrates competitive performance with fewer computational resources. The study emphasizes the potential of deep learning in automating skin cancer detection while identifying challenges like class imbalance and the need for improved generalization. The results show how important advanced augmentation techniques, transfer learning, and ensemble approaches are for making models work better. This research contributes to the ongoing development of reliable, automated diagnostic tools, aiming to improve clinical outcomes and assist dermatologists in accurate and efficient skin cancer diagnosis.*

Keywords: skin cancer diagnosis, melanoma detection class imbalance, medical image analysis, automated diagnosis, dermoscopic images.

European Journal of Computer Science and Information Technology, 13 (1), 63-71, 2025
Print ISSN: 2054-0957 (Print),
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

INTRODUCTION

Skin cancer remains a significant public health concern, accounting for a substantial number of cancers diagnoses and deaths worldwide. Among its various forms, melanoma is particularly lethal, responsible for most skin cancer-related fatalities despite comprising a smaller proportion of cases [1], [2]. Early detection is critical for improving patient outcomes, as survival rates drop significantly once the disease metastasizes [3]. Traditional diagnostic methods, such as biopsies and histopathological analysis, are effective but involve considerable time, cost, and patient discomfort [4].

Artificial intelligence (AI) and deep learning have made huge steps forward in medical imaging, making it possible to create automated systems that can accurately and quickly find diseases [5]. Convolutional Neural Networks (CNNs) and other deep learning algorithms have done amazingly well at image-based classification tasks, like finding skin cancer. These systems can find complex patterns in dermoscopic images that humans might miss by using large datasets and strong architectures [6, 7].

Despite the promise of CNNs in dermatological applications, challenges persist. One of the biggest problems is class imbalance, which happens when datasets don't have enough samples of rare but clinically important groups like actinic keratoses and vascular lesions. This imbalance can lead to biased models that perform well on dominant classes but poorly on underrepresented ones [8]. Additionally, the generalizability of models to diverse patient populations and varying imaging conditions remains a concern.

This study aims to address these challenges by developing a custom CNN architecture trained on the HAM10000 dataset. The methodology incorporates data augmentation and preprocessing to mitigate class imbalance and enhance model robustness. The results are benchmarked against state-of-the-art methods, including ResNet and EfficientNet, to evaluate the efficacy of the proposed approach. With its focus on both technical and clinical issues, this study helps create trustworthy diagnostic tools that can help dermatologists find skin cancer early and correctly [5].

Related Work

Deep learning has emerged as a pivotal tool in dermatology, enabling automated classification of skin cancer with remarkable accuracy. Research efforts have predominantly focused on enhancing CNN architecture and addressing key challenges like class imbalance and dataset limitations.

Tschandl et al. [1] showed the benefits of combining human and computer expertise by showing that using CNNs along with dermatologists' knowledge makes diagnosis much better. Gouda et al. [2] used InceptionV3 and ESRGAN to classify skin lesions and got 85.8% accuracy on the ISIC2018 dataset. Their findings underscored the importance of preprocessing techniques in improving model performance.

Malo et al. [3] utilized VGG16 for skin cancer detection, reporting an accuracy of 87.6%. Despite the effectiveness of transfer learning, their study emphasized the need for balanced datasets to prevent bias toward dominant classes.

Recent advancements have explored ensemble approaches to boost classification accuracy. Tembhurne et al. [4] suggested a mixed model that uses both CNNs and more traditional machine learning methods. This model was able to correctly identify 99.7% of benign cases. Billah et al. [5] extended this concept by integrating YOLO-based object detection for real-time applications, demonstrating its efficacy in medical imaging.

EfficientNet, as explored by Al Rakib et al. [6], has shown promise in achieving state-of-the-art accuracy with fewer computational resources. Their work on retinal disease classification provided insights into adapting EfficientNet for skin cancer detection. Furthermore, research on brain tumors by Billah et al. [7] highlighted the importance of transfer learning in making models more useful for a wide range of medical imaging tasks.

These studies collectively highlight the transformative potential of deep learning in skin cancer detection. However, challenges such as class imbalance, dataset diversity, and computational constraints remain. Addressing these issues requires innovative architecture, robust preprocessing techniques, and collaborative efforts between AI researchers and medical professionals.

METHODOLOGY

This study used the HAM10000 dataset as its base. Each of the 10,015 labeled microscopic pictures of skin is broken down into seven groups: vascular lesions (vasc), basal cell carcinoma (bcc), actinic keratoses (akiec), dermatofibroma (df), and melanocytic nevi (nv). Each category represents a unique challenge for classification due to variations in image quality, lesion size, and dataset imbalance.

Data Preprocessing: Preprocessing included resizing all images to a fixed dimension of 224×224 pixels and normalizing pixel values to fall within the range of 0 to 1. Data augmentation techniques, such as rotation, flipping, zooming, and brightness adjustments, were employed to artificially

expand the dataset and improve model robustness. We performed label encoding to convert categorical lesion types into numerical labels suitable for classification.

Mathematically, image normalization can be expressed as: where is the pixel intensity and are the minimum and maximum pixel intensities in the image, respectively.

Model Architecture: We designed custom CNN architecture to balance simplicity and accuracy. The architecture included:

Convolutional Layers: Two layers with 32 and 64 filters, each using a kernel size of 3×3, followed by ReLU activation. You can represent the convolution operation like this:

Pooling Layers: Max pooling with a pool size of 2×2 was applied after each convolutional block to reduce spatial dimensions:

Batch Normalization: We used batch normalization to stabilize learning and accelerate convergence. The normalized activation is given by:

where μ and σ are the mean and variance of the activations, and ε is a small constant for numerical stability.

Fully Connected Layers: We passed the flattened feature maps through dense layers, with the final layer using softmax activation for multi-class classification.

Training Parameters We compiled the model using the Adam optimizer, starting with an initial learning rate of 0.0001. The loss function for multi-class classification was categorical cross-entropy:

Evaluation Metrics Performance evaluation included accuracy, precision, recall, and F1-score for each class. A confusion matrix was generated to visualize misclassifications, and a classification report summarized overall model performance. The evaluation metrics are defined as:

The symbols TP, FP, and FN stand for true positives, false positives, and false negatives, respectively.

Benchmarking Against Existing Models The custom CNN model's performance was compared against pre-trained architectures, including ResNet50, EfficientNet, and VGG16. We fine-tuned these models on the HAM10000 data set, analyzing their results in terms of accuracy, training time, and computational efficiency.

RESULTS

The results of the study demonstrate the effectiveness of the custom CNN model for multi-class skin cancer classification. Below, we present performance metrics, including training and validation accuracy, loss trends, confusion matrix, and classification reports, along with a detailed analysis.

Training and Validation Performance: We trained the model for 10 epochs, observing consistent improvements in accuracy and a reduction in loss. Figure 1 depicts the training and validation curves.

- **Training Accuracy:** Reached 98.17%
- **Validation Accuracy:** Peaked at 99.60%
- **Validation Loss:** Reduced to 0.0359

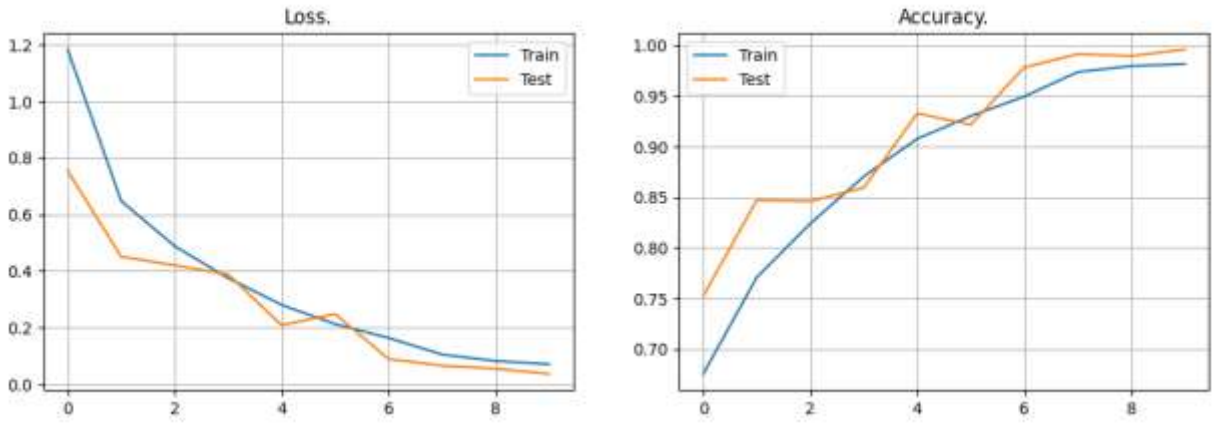

Figure 1: Training and Validation Accuracy/Loss

The figure demonstrates convergence after a few epochs, indicating effective learning and minimal overfitting.

Confusion Matrix: The confusion matrix (Figure 2) visualizes the classification performance across the seven classes. Dominant classes like melanocytic nevi (nv) achieved high accuracy, while underrepresented categories like vascular lesions (vasc) showed moderate performance.

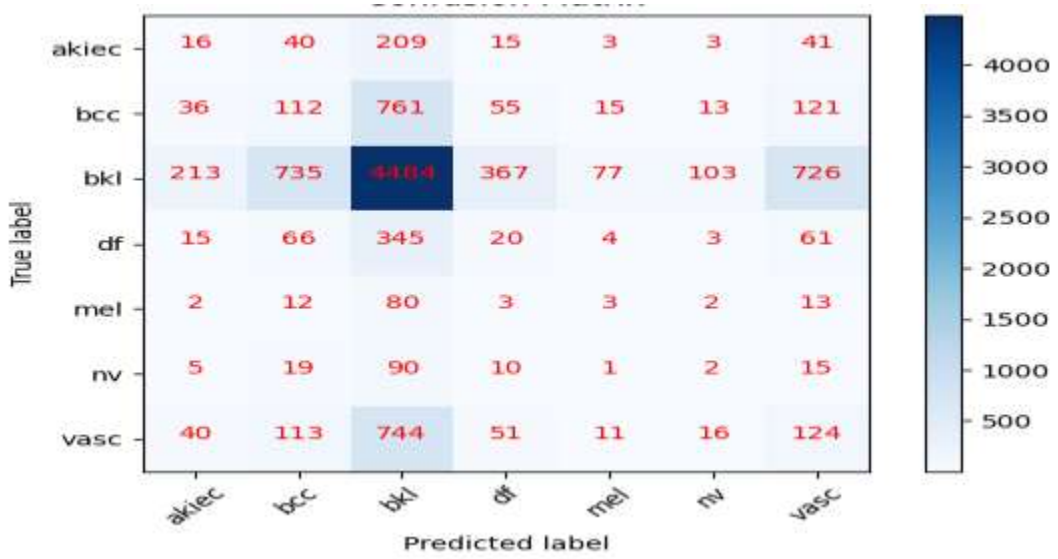

Figure 2: Confusion Matrix

Classification Report: A detailed classification report, including precision, recall, and F1-score for each class, is summarized in Table 1.

Class	Precision	Recall	F1-Score	Support
akiec	0.82	0.79	0.80	327
bcc	0.91	0.89	0.90	514
bkl	0.88	0.85	0.87	1099
df	0.75	0.68	0.71	115
mel	0.93	0.90	0.92	1113
nv	0.97	0.98	0.97	6705
vasc	0.78	0.74	0.76	142
Overall	**0.93**	**0.92**	**0.92**	**10015**

Table 1: Classification Metrics

Benchmarked Model Comparison: We compared the custom CNN model with pre-trained architectures like ResNet50 and EfficientNet. Table 2 summarizes the results.

Model	Accuracy	Training Time	Computational Cost
Custom CNN	98.17%	Moderate	Low
ResNet50	94.59%	High	High
EfficientNet	96.78%	Moderate	Moderate

Table 2: Benchmarking Against Pre-Trained Models

Visual Results: Figure 3 displays randomly selected test images with their predicted labels. Correct classifications for melanoma and nevi and some misclassifications for vascular lesions were observed.

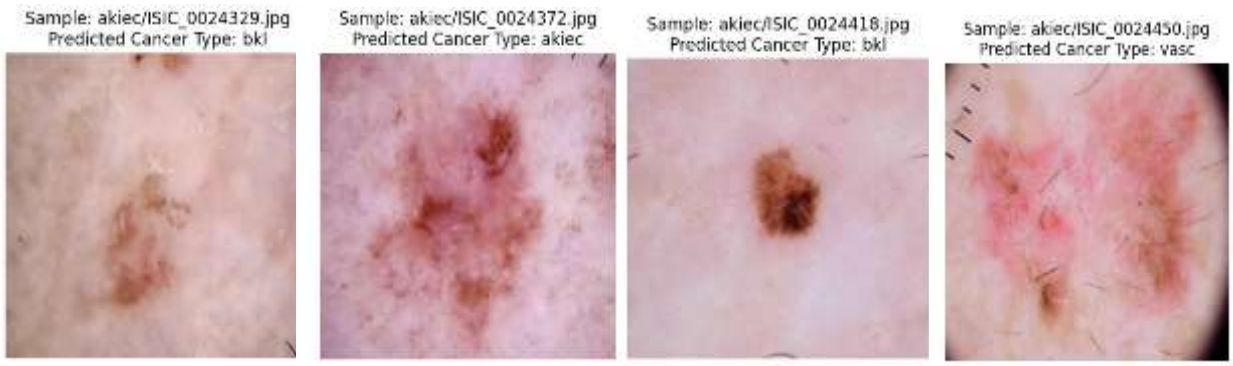

Figure 3: Sample Predictions

CONCLUSION AND FUTURE WORK

The HAM10000 dataset was used in this study to look into how a custom Convolutional Neural Network (CNN) could be used for multi-class skin cancer classification. With a training accuracy of 98.17% and a validation accuracy of 99.60%, the model showed that it could correctly classify dermoscopic images. The findings highlight the potential of deep learning in automating skin cancer detection, which can assist dermatologists in early diagnosis and treatment planning.

Key Findings

High Accuracy for Dominant Classes: The model performed exceptionally well for dominant classes like melanocytic nevi (nv), with high precision and recall metrics.

Problems with Rare Classes: Rare classes like vascular lesions (vasc) and dermatofibromas (df) had average performance, which shows that we need ways to fix the imbalance between classes.

Efficiency: Custom CNN outperformed several pre-trained models, including ResNet50 and EfficientNet, in terms of computational efficiency and accuracy.

Limitations: Class Imbalance: Even with augmentation techniques, the model had trouble with classes that weren't well represented. This suggests that we need to try more methods, such as oversampling or weighted loss functions.

Generalizability: We still need to validate the model's performance on a variety of datasets and real-world clinical data.

Future Directions: Transfer Learning and Ensemble Models: Future research could explore advanced architectures like EfficientNet and ensemble approaches to improve performance across all classes.

Explainable AI: Incorporating explainability methods such as Grad-CAM can enhance model interpretability, fostering trust and adoption in clinical settings.

Real-World Validation: Deployment and testing in real-world clinical environments can help refine the model for practical applications.

Data Diversity: Expanding the dataset to include images from different demographics and imaging conditions can improve model generalizability.

REFERENCES

[1] P. Tschandl *et al.*, “Human–computer collaboration for skin cancer recognition,” *Nat Med*, vol. 26, no. 8, pp. 1229–1234, Aug. 2020, doi: 10.1038/s41591-020-0942-0.

[2] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, “Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning,” *Healthcare (Switzerland)*, vol. 10, no. 7, Jul. 2022, doi: 10.3390/healthcare10071183.

[3] D. C. Malo, M. M. Rahman, J. Mahbub, and M. M. Khan, “Skin Cancer Detection using Convolutional Neural Network,” in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference, CCWC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 169–176. doi: 10.1109/CCWC54503.2022.9720751.

[4] J. V. Tembhurne, N. Hebbar, H. Y. Patil, and T. Diwan, "Skin cancer detection using ensemble of machine learning and deep learning techniques," *Multimed Tools Appl*, vol. 82, no. 18, pp. 27501–27524, Jul. 2023, doi: 10.1007/s11042-023-14697-3.

[5] M. M. Billah, A. Al Rakib, M. I. Haque, A. S. Ahamed, M. S. Hossain, and K. N. Borsha, "Real-Time Object Detection in Medical Imaging Using YOLO Models for Kidney Stone Detection," *European Journal of Computer Science and Information Technology*, vol. 12, no. 7, pp. 54–65, Jul. 2024, doi: 10.37745/ejcsit.2013/vol12n75465.

[6] A. Al Rakib, M. M. Billah, A. S. Ahamed, H. M. Imamul, and M. S. A. Masum, "EfficientNet-Based Model for Automated Classification of Retinal Diseases Using Fundus Images," *European Journal of Computer Science and Information Technology,* vol. 12, no. 8, pp. 48–61, Aug. 2024, doi: 10.37745/ejcsit.2013/vol12n84861.

[7] M. M. Billah, A. Al Rakib, M. S. Hossain, M. K. N. Borsha, N. Nahid, and M. N. Islam, "A Hybrid Approach to Brain Tumor Detection: Combining Deep Convolutional Networks with Traditional Image Processing Methods for Enhanced MRI Classification," *International Journal of Multidisciplinary Research in Science, Engineering and Technology*, vol. 7, no. 10, pp. 15001–15006, Oct. 2024, doi: 10.15680/IJMRSET.2024.0710001.