IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

# Automated Skin Lesion Classification Using Deep Learning on the HAM10000 Dataset

**Ali Osman, Abdelrahman Magdi, Eyad Yehia, Omar Elhawwari**

**ABSTRACT** This project aims to enhance the automated diagnosis of pigmented skin lesions by leveraging deep learning techniques, specifically Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), and VGG models on the HAM10000 dataset. The dataset, consisting of 10,015 dermatoscopic images spanning seven classes of skin cancer, offers a diverse and comprehensive foundation for developing robust machine learning models. By implementing image augmentation techniques and the Adam optimizer, we seek to improve the accuracy and generalizability of our predictive models. The study demonstrates the potential of CNNs in medical image classification and provides insights into the challenges and future directions in this domain.

**Keywords** Automated Diagnosis, Pigmented Skin Lesions, Deep Learning, Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), HAM10000 Dataset, Image Augmentation, Adam Optimizer, Medical Image Classification, Skin Cancer, Melanocytic Nevi, Melanoma, Benign Keratosis-like Lesions, Basal Cell Carcinoma, Actinic Keratoses, Vascular Lesions, Dermatofibroma, Histopathology, Evaluation Metrics, Accuracy, Precision, Recall, F1-Score, AUC-ROC, Class Imbalance, Data Preprocessing, Model Development, Training and Evaluation, Error Analysis, Transfer Learning, Clinical Validation, Machine Learning Models, Dermoscopy Images, Image Classification, Medical Image Analysis, Feature Detection, Deep Neural Networks, InceptionResNetV2, EfficientNet B0, VGG16, Transfer Learning, Class Activation Maps, Telemedicine, Dataset Diversity, Overfitting, CNN Model Architecture, Pooling Layers, Dropout Layers, Fully Connected Layers, Training Dataset

## I. INTRODUCTION

Skin cancer is a prevalent and potentially fatal disease where early detection is crucial for effective treatment. Automated classification of skin lesions can significantly aid dermatologists by providing reliable second opinions and reducing diagnostic errors. The HAM10000 dataset, a large collection of dermatoscopic images from diverse populations, serves as an excellent resource for training machine learning models to classify skin lesions into seven distinct categories.

## II. Problem Statement

The primary objective of this project is to develop and evaluate deep learning models capable of accurately classifying dermatoscopic images into one of seven categories: melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma. We aim to address the challenges posed by the small size and lack of diversity in available datasets, thereby improving the automated diagnosis of pigmented skin lesions.

### A. Dataset Description

The HAM10000 dataset comprises 10,015 dermatoscopic images, each associated with one of seven skin lesion classes. These images are collected from various populations using different acquisition methods, ensuring a diverse dataset. Over 50% of the lesions are confirmed through histopathology, with the remaining verified by follow-up examinations, expert consensus, or in-vivo confocal microscopy. The dataset is publicly available under the CC BY-NC-SA 4.0 license.

### B. Literature Review

Automated skin lesion classification has been a significant research focus, with various approaches leveraging machine learning and deep learning. Early methods relied on handcrafted features and traditional classifiers, such as support vector machines (SVMs) and decision trees. These methods utilized features like color, texture, and shape descriptors to distinguish between different types of skin lesions. For instance, a study by Codella et al. (2015) utilized handcrafted features combined with SVMs, achieving moderate accuracy but facing limitations in handling complex patterns in lesion images.

In contrast, recent advancements in deep learning, particularly convolutional neural networks (CNNs), have revolutionized this field. CNNs automatically learn hierarchical feature representations directly from the raw pixel values of images, which has significantly improved classification performance. Studies, such as those by Esteva et al. (2017) and Tschandl et al. (2018), demonstrated that CNNs could achieve remarkable accuracy in image classification tasks, including skin lesion detection, often surpassing the performance of dermatologists in specific scenarios.

However, challenges such as class imbalance, data augmentation, and overfitting remain prevalent. Class imbalance, where certain types of lesions are underrepresented, can lead to biased models that perform poorly on rare classes. Data augmentation techniques, like random rotations, flips, and color variations, are often employed to artificially increase the diversity of the training data and mitigate overfitting. Studies by Haenssle et al. (2018) and Brinker et al. (2019) have explored various data augmentation strategies to enhance the robustness of deep learning models.

This project builds upon these findings, aiming to develop a robust model using the HAM10000 dataset. The HAM10000 dataset, which includes a diverse set of dermatoscopic images, addresses some of the challenges related to data diversity and class imbalance. By employing advanced CNN architectures and incorporating effective data augmentation techniques, this project seeks to improve upon the current state-of-the-art in skin lesion classification. Comparing to previous papers, our approach not only leverages the latest deep learning advancements but also systematically addresses the common pitfalls identified in prior research, striving for a more generalized and accurate skin lesion classification model.

## III. Proposed Model Architecture

Our proposed model architecture primarily utilizes CNNs due to their proven efficacy in image classification tasks. The CNN model comprises multiple convolutional layers followed by pooling layers, dropout layers for regularization, and fully connected layers. Additionally, we explore the use of ANNs and VGG models to compare performance. Image augmentation techniques such as rotation, zoom, and horizontal flipping are employed to enhance the dataset's diversity and prevent overfitting. The Adam optimizer is used for training due to its efficiency and effectiveness in handling sparse gradients.

## IV. Methods

- Data Preprocessing: The images are resized to a uniform size, normalized, and augmented using techniques like rotation, zoom, and flipping.

- Model Development: Three models are designed: a CNN model with multiple convolutional, pooling, and fully connected layers, an ANN model, and a VGG model known for its deep architecture and performance in image classification tasks.

- Training: The models are trained using the Adam optimizer with the training data split into training and validation sets.

- Evaluation: The models are evaluated using metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC).

## V. Evaluation Metrics and Results

The performance of the models is assessed using various metrics:

- Accuracy: The proportion of correctly classified images.

- Precision and Recall: Evaluating the trade-off between false positives and false negatives.

• F1-Score: The harmonic mean of precision and recall, providing a balanced measure.

• AUC-ROC: Assessing the model's ability to distinguish between classes.

Preliminary results indicate that the CNN model outperforms the ANN and VGG models, achieving higher accuracy and better generalization on the validation set. The image augmentation and Adam optimizer significantly contribute to the model's performance.

## VI. Error Analysis

An in-depth error analysis is conducted to identify common misclassifications and their underlying causes. Factors such as class imbalance, image quality, and similarity between certain lesion types are considered. Strategies to mitigate these issues, including enhanced augmentation techniques and class-specific adjustments, are proposed.

## VII. CONCLUSION

This project demonstrates the potential of CNNs in the automated classification of skin lesions using the HAM10000 dataset. While promising results are achieved, further improvements can be made by addressing class imbalance and exploring advanced architectures like transfer learning. Future work includes expanding the dataset, incorporating additional modalities, and deploying the model in real-world clinical settings for validation.

## APPENDIX

Appendixes, **********.

## REFERENCES

• Codella, N., Rotemberg, V., Tschandl, P., et al. (2018). Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). arXiv:1902.03368.

• Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data, 5, 180161. doi:10.1038/sdata.2018.161.