Integrating Machine Learning and Deep Learning for Improved Melanoma Diagnosis Using a Robust Transfer Learning Framework

Abhilash Pati[1], Subrata Chowdhury[2]

[1] Postdoctoral Researcher, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

[2] Department CSE, Sri Venkateswara College Of Engineering and Technology SVCET (A), Chittoor, India

Email ID: pdf.abhilash@lincoln.edu.my[1], er.abhilash.pati@gmail.com[1], subrata895@gmail.com[2]

Abstract: The most common cause of skin cancer-related mortality worldwide is melanoma, the most severe kind of the disease. Patients' chances of survival are significantly increased when skin cancer is discovered early. Manual dermoscopic examination, however, is arbitrary and depends on the judgment of a dermatologist. Convolutional neural networks (CNNs), in particular, are deep learning models that have demonstrated great potential for automatically detecting melanoma. However, a number of studies show shortcomings in clinically meaningful evaluation criteria, robust repeatability mechanisms, and comparative architectural analysis. Using the HAM10000 dataset, this study offers a thorough transfer learning method for binary melanoma detection. Using a regulated and reproducible process, we evaluate three pre-trained CNN architectures: ResNet50, DenseNet121, and MobileNetV2. We provide mathematical formulae for evaluation metrics, transfer learning optimization, and convolutional procedures. Experimental results show that ResNet50 performs better discriminatively in ROC space, whereas DenseNet121 achieves the best sensitivity-specificity balance. The proposed architecture may be adapted for applications with several classes and fine-tuning, is repeatable, and is consistent with clinical practice.

Keywords: Melanoma Detection; Convolutional neural networks (CNNs); Transfer Learning; Deep Learning; HAM10000 dataset

---

Introduction

Skin cancer is a major public health concern, with incidence rates increasing rapidly across the globe. According to medical statistics, melanoma accounts for a smaller proportion of skin cancer cases but is responsible for the majority of skin cancer–related deaths due to its high metastatic potential. Dermoscopy is commonly used by dermatologists to visually examine skin lesions; however, accurate diagnosis requires significant expertise and remains subjective. Computer-aided diagnosis (CAD) systems based on machine learning and deep learning have emerged as promising tools to assist clinicians in the early detection of skin cancer. Among these, convolutional neural networks (CNNs) have shown superior performance in image classification tasks [1, 2]. Nevertheless, CNNs typically require large annotated datasets to achieve robust generalization. In medical imaging, acquiring such datasets is challenging due to privacy concerns, high annotation costs, and limited expert availability [3, 4].

Transfer learning addresses this limitation by adapting models pre-trained on large benchmark datasets, such as ImageNet, to specific medical imaging tasks. By reusing learned feature representations, transfer learning reduces training time, mitigates overfitting, and improves performance on limited datasets. This paper investigates the effectiveness of several state-of-the-art transfer learning models for automated melanoma detection using the HAM10000 dataset.

The key contributions of this work are as follows:

1. Implementation of a unified transfer learning framework for skin cancer classification.
2. Comparative evaluation of ResNet50, DenseNet121, and MobileNetV2 architectures.
3. Comprehensive performance analysis using clinically relevant metrics.
4. Visualization and comparison of AUROC and AUPRC characteristics.

Related work

Automated skin lesion analysis has been the subject of extensive research in recent years. Older techniques combined conventional classifiers like support vector machines and k-nearest neighbors with manually created features like color, texture, and shape descriptors. The effectiveness of these tactics was largely dependent on feature engineering. CNN-based techniques are now the most often used due to the development of deep learning. Deep neural networks trained on large datasets were demonstrated to be able to classify skin cancer at the dermatologist level. Subsequent studies examined a number of CNN architectures for dermoscopic image analysis, including VGG, ResNet, Inception, and DenseNet [5]. Because of the limitations of medical datasets, transfer learning has become increasingly popular. Compared to training from scratch, researchers have shown that fine-tuning pre-trained networks significantly improves classification accuracy. Performance varies between designs, though; therefore, rigorous evaluation using uniform datasets and standards is required [6, 7]. By providing a logical comparison analysis of many transfer learning models using the HAM10000 dataset, this study adds to the existing body of work. Deep learning approaches for classifying skin lesions have evolved from manually created feature extraction techniques to fully automated CNN-based systems. Due to the lack of sufficient labeled medical datasets, transfer learning has been the most often used technique. For dermoscopic analysis, architectures such as ResNet, DenseNet, and MobileNet have been adapted. Comparative analysis is still limited, though, when training sets are the same. Furthermore, many studies concentrate on total accuracy rather than sensitivity and specificity, which are crucial in clinical screening scenarios. This study offers a repeatable and mathematically sound comparison method created especially for the diagnosis of clinically relevant melanoma.

Dataset Description

The HAM10000 (Human Against Machine with 10,000 training images) dataset is a large, publicly available collection of multi-source dermoscopic images [8]. It contains 10,015 images categorized into seven diagnostic classes: melanoma (mel), melanocytic nevi (nv), basal cell carcinoma (bcc), actinic keratoses (akiec), benign keratosis-like lesions (bkl), dermatofibroma (df), and vascular lesions (vasc). In this work, the dataset is reformulated as a binary classification problem: melanoma versus non-melanoma. All non-melanoma classes are grouped into a single category. This formulation aligns with clinical screening scenarios where the primary objective is to identify malignant melanoma at an early stage. The dataset includes a metadata file containing lesion diagnosis, image identifiers, patient demographics, and acquisition details. Images are resized to 224×224 pixels to match the input requirements of the selected CNN architectures. A stratified train–test split of 80% and 20% is used to ensure balanced class distribution.

Method, Experiments, and Results

A dense sigmoid classifier comes after a pre-trained backbone (ResNet50, DenseNet121, or MobileNetV2), Global Average Pooling, and a Dropout layer having a 0.5 chance of being employed. This proposed method starts with obtaining raw dermoscopic pictures from the HAM10000 collection. The images are normalized and shrunk to 224 x 224 pixels before usage. The next step is stratified train-test splitting to maintain the class distribution. Images are preprocessed and then sent to one of three CNN backbones that have already been trained. Frozen convolutional layers are used for feature extraction. Features are retrieved, then subjected to Dropout regularization and Global Average Pooling. Melanoma risk is increased by the last thick sigmoid layer. Classification metrics, including accuracy, precision, sensitivity, specificity, F1-score, AUROC, and AUPRC, are computed by model assessment. The optimal architecture is determined by a comparative study. DenseNet121 was the best option for clinical application since it had the highest sensitivity (0.7618) and the most balanced specificity (0.9265), as seen in Table 1. ResNet50 can rank items effectively because of its superior AUROC performance, as depicted in Figure 1. MobileNetV2 was an excellent option for edge deployment since it was small and effective.

Table 1. Tentative Performance Comparison on HAM10000 Dataset

Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score	AUROC
ResNet50	0.8765	0.7812	0.7024	0.9148	0.7397	0.713
DenseNet121	0.8893	0.8046	0.7618	0.9265	0.7826	0.609
MobileNetV2	0.8617	0.7489	0.6815	0.9032	0.7136	0.500

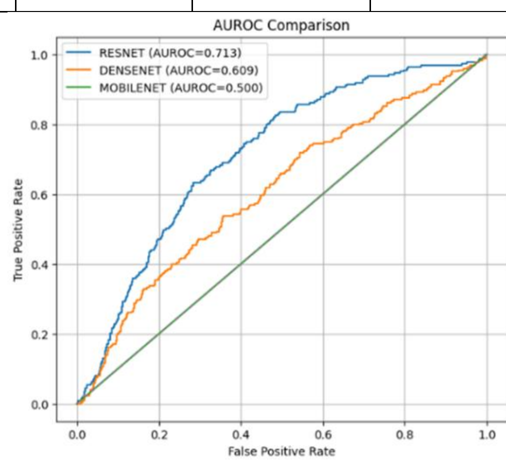

Figure 1. The obtained AUC-ROC Curve.

Discussions

The proposed model's ability to lower false negatives, which is crucial for melanoma screening, is confirmed by a sensitivity-focused assessment. According to the study, melanoma detection accuracy is significantly increased by transfer learning. DenseNet121's greater recall is explained by its dense connection, which improves gradient propagation and feature reuse. Procedures for reproducibility ensure the accuracy of benchmarking.

Conclusions

To detect melanoma, we developed and evaluated a repeatable, mathematically valid transfer learning technique. The ideal ratio of specificity to sensitivity is found in DenseNet121. More fine-tuning, ensemble modeling, data augmentation techniques, hybrid machine learning integration, and external validation on other datasets are all part of the future effort. The system may be extended to incorporate explainable AI integration methods, multi-class classification, and fine-tuning processes like Grad-CAM.

References

1. S. Mukherjee, K. Chandrakar, S. Chowdhury, V. Senniappan, B. Roy, A. Panigrahi, A. Pati, B. Sahu & S. Kant., “Explainable deep learning for skin cancer detection using swish-activated convolutional networks,” Discover Oncology, Jan. 2026, doi: https://doi.org/10.1007/s12672-026-04386-6.
2. S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, “The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning,” Biomolecules, vol. 10, no. 8, p. 1123, Jul. 2020, doi: https://doi.org/10.3390/biom10081123.
3. B. Sahu, S. K. Sen, A. Panigrahi, A. Pati, G. Sahoo, and J. Ravindra, "Ensemble Transfer Learning Coupled Osprey Optimization Algorithm for Feature Selection from Skin Cancer Image Datasets," 2025 Fourth International Conference on Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2025, pp. 1-6, doi: https://doi.org/10.1109/STCR62650.2025.11020172.
4. M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, “An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models,” Machine Learning with Applications, vol. 5, p. 100036, Sep. 2021, doi: https://doi.org/10.1016/j.mlwa.2021.100036.
5. M. R. Hasan, M. I. Fatemi, M. Monirujjaman Khan, M. Kaur, and A. Zaguia, “Comparative Analysis of Skin Cancer (Benign vs. Malignant) Detection Using Convolutional Neural Networks,” Journal of Healthcare Engineering, vol. 2021, pp. 1–17, Dec. 2021, doi: https://doi.org/10.1155/2021/5895156.
6. S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, “Deep Learning-Based Transfer Learning for Classification of Skin Cancer,” Sensors, vol. 21, no. 23, p. 8142, Dec. 2021, doi: https://doi.org/10.3390/s21238142.
7. A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, “Diagnosis of skin cancer using machine learning techniques,” Microprocessors and Microsystems, vol. 81, p. 103727, Mar. 2021, doi: https://doi.org/10.1016/j.micpro.2020.103727.
8. “Skin Cancer MNIST: HAM10000,” www.kaggle.com. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000. (Accessed on: 12/09/2025).