Explainable Skin Cancer Classification Using Deep Feature Fusion and Ensemble Learning

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: Skin cancer is one of the most dangerous and rapidly increasing forms of cancer worldwide, where early melanoma detection significantly improves survival rates. However, accurate diagnosis using dermoscopic images remains challenging due to class imbalance, visual similarity among lesions, and limited annotated medical datasets. This study proposes an explainable hybrid skin cancer classification framework integrating deep feature fusion with ensemble machine learning techniques for melanoma detection using the HAM10000 dataset. Pre-trained convolutional neural networks, namely ResNet50 and MobileNetV2, are employed as deep feature extractors, while Support Vector Machine (SVM) and XGBoost classifiers are utilized for classification. Synthetic Minority Oversampling Technique (SMOTE), image augmentation, and transfer learning strategies are incorporated to improve robustness and generalization. Furthermore, explainable artificial intelligence (XAI) techniques, including LIME and SHAP, are integrated to enhance interpretability and clinical reliability. Experimental results demonstrate that the SVM classifier achieved the highest classification accuracy of 80.09%, specificity of 86.97%, and AUROC of 0.8641, while XGBoost achieved superior sensitivity and precision-recall performance. The proposed framework can support dermatologists in reliable melanoma screening and can be extended for multiclass skin lesion classification.

Keywords: Skin Cancer; Deep Learning; Transfer Learning; Explainable AI; Ensemble Learning.

Introduction

Melanoma is the most deadly kind of skin cancer because of its high potential for metastasis, making it one of the most common diseases worldwide. Recent medical reports state that UV radiation, genetic vulnerability, and environmental exposure all contribute to the annual rise in melanoma incidence. Reducing death rates and increasing patient survival depend heavily on early diagnosis. Conventional diagnosis methods mainly rely on the visual analysis of dermoscopic pictures and the knowledge of dermatologists [1, 2]. Nevertheless, manual diagnosis is frequently subjective, laborious, and prone to inter-observer variability. Automated medical image processing has been greatly enhanced by recent developments in artificial intelligence, especially deep learning and transfer learning. The ability of Convolutional Neural Networks (CNNs) to extract intricate visual patterns from dermoscopic images has been shown to be remarkably successful. Despite these developments, a number of issues are yet

unsolved [3, 4]. Numerous current research studies use a single CNN architecture without doing a comparative analysis, have issues with class imbalance, and lack the explainability mechanisms necessary for clinical adoption. Furthermore, the majority of models ignore clinically important parameters like sensitivity and specificity in favor of just increasing accuracy. This study suggests an explainable skin cancer classification system that makes use of ensemble machine learning and deep feature fusion to address these issues. To increase robustness and interpretability, the suggested approach combines conventional machine learning classifiers with transfer learning-based feature extraction. Furthermore, Explainable Artificial Intelligence (XAI) methods like SHAP and LIME are used to improve credibility and make transparent forecasts in clinical settings. The key contributions of this work are summarized as follows:

1. A hybrid skin cancer classification framework combining transfer learning and ensemble machine learning is proposed.
2. Deep features extracted from ResNet50 and MobileNetV2 are fused for improved representation learning.
3. SVM and XGBoost classifiers are comparatively evaluated for melanoma detection.
4. SMOTE and data augmentation techniques are applied to address dataset imbalance.
5. Explainable AI methods, including LIME and SHAP, are incorporated for interpretable clinical decision-making.
6. The proposed framework achieves strong AUROC and specificity performance on the HAM10000 dataset.

Related work

In recent years, a lot of studies have been done on automated skin lesion analysis. Previous methods integrated manually generated features like color, texture, and form descriptors with traditional classifiers like support vector machines and k-nearest neighbors. Feature engineering had a major role in these strategies' efficacy. Deep learning has made CNN-based methods the most popular. It has been shown that dermatologists can classify skin cancer using deep neural networks trained on big datasets. Several CNN architectures, such as VGG, ResNet, Inception, and DenseNet, were investigated in further research for dermoscopic image processing [5]. Transfer learning has grown in popularity due to the limitations of medical datasets. Researchers have demonstrated that fine-tuning pre-trained networks greatly increases classification accuracy when compared to training from scratch. However, performance differs across designs, necessitating a thorough examination with consistent datasets and standards [6, 7]. This paper contributes to the body of work by offering a logical comparison analysis of numerous transfer learning models using the HAM10000 dataset. From manually developed feature extraction methods to fully automated CNN-based systems, deep learning algorithms for skin lesion classification have been developed. Transfer learning has been the most popular method because there aren't enough annotated medical datasets. Architectures like ResNet, DenseNet, and MobileNet have been modified for dermoscopic analysis. However, when training sets are identical, comparative analysis is still constrained. Additionally, a lot of research focuses on overall accuracy rather than sensitivity and specificity, which are important in situations involving clinical screening. Specifically designed for the detection of clinically relevant melanoma, this study provides a reproducible and mathematically solid comparison procedure.

Dataset Description

A sizable, publicly accessible set of multi-source dermoscopic pictures is called the HAM10000 (Human Against Machine with 10,000 training images) dataset [8]. Melanoma (mel), melanocytic nevi (nv), basal cell carcinoma (bcc), actinic keratoses (akiec), benign keratosis-like lesions (bkl), dermatofibroma (df), and vascular lesions (vasc) are the seven diagnostic classifications into which it comprises 10,015 photos. The dataset is reframed in this work as a binary classification problem: melanoma versus non-melanoma. There is only one category for all non-melanoma classes. This formulation is consistent with clinical screening circumstances where early detection of malignant melanoma is the main goal. A metadata file including lesion diagnosis, image identification, patient demographics, and acquisition information is included in the collection. To meet the input requirements of the chosen CNN architectures, images are scaled to 224 by 224 pixels. To guarantee a balanced distribution of classes, a stratified train-test split of 80% and 20% is employed.

Method, Experiments, and Results

The several phases of the suggested framework are intended to enhance the precision and comprehensibility of skin cancer categorization. To guarantee consistency throughout model training, dermoscopic images first go through preprocessing, which includes scaling and normalization. Synthetic Minority Oversampling Technique (SMOTE) is used in conjunction with data augmentation techniques like flipping, rotation, and zooming to address dataset imbalance and enhance generalization capabilities. Two pre-trained convolutional neural network architectures, ResNet50 and MobileNetV2, initialized with ImageNet weights, are then used to carry out deep feature extraction. The transfer learning method trains bespoke classification layers for feature learning while freezing the convolutional base layers. To improve discriminative representation, the retrieved deep features are then merged and dimensionality reduced. Two machine learning techniques, Support Vector Machine (SVM) and XGBoost, are used and compared for classification. Additionally, the system incorporates Explainable Artificial Intelligence (XAI) approaches like SHAP and LIME to enhance clinical interpretability and find significant deep features that contribute to melanoma prediction. With a batch size of 32, a learning rate of 1e-4, an Adam optimizer, a Binary Cross-Entropy loss function, and five training epochs, the experiments were carried out using Python in the Kaggle Notebook environment. Several performance indicators, such as accuracy, precision, sensitivity (recall), specificity, F1-score, Area Under the Receiver Operating Characteristic Curve (AUROC), and Area Under the Precision-Recall Curve (AUPRC), were used to assess the suggested models.

Table 1. Tentative Performance Comparison on HAM10000 Dataset

Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score
SVM	0.8009	0.7184	0.6637	0.8697	0.6900
XGBoost	0.7799	0.6638	0.6906	0.8247	0.6769

Figure 1 presents the AUROC comparison between SVM and XGBoost classifiers. Figure 2 demonstrates the Precision-Recall curves for both classifiers. Figure 3 shows LIME-based feature interpretability for melanoma prediction. Figure 4 presents SHAP summary plots illustrating feature importance and model interpretability.

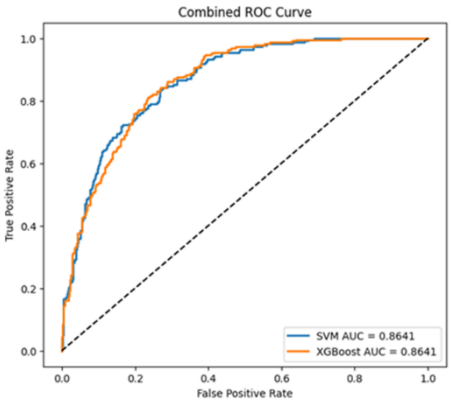

Figure 1: AUROC Plot

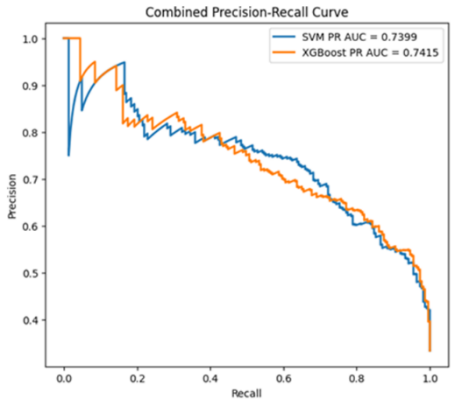

Figure 2: AUPRC Plot

Feature	Value
Feature_685	0.20
Feature_1703	-0.96
Feature_3105	-0.03
Feature_14	-1.16
Feature_3239	-0.06
Feature_1556	-0.78
Feature_3021	-0.04
Feature_2676	-0.04
Feature_2889	-0.08
Feature_900	-0.11

Figure 3: LIME Explainability

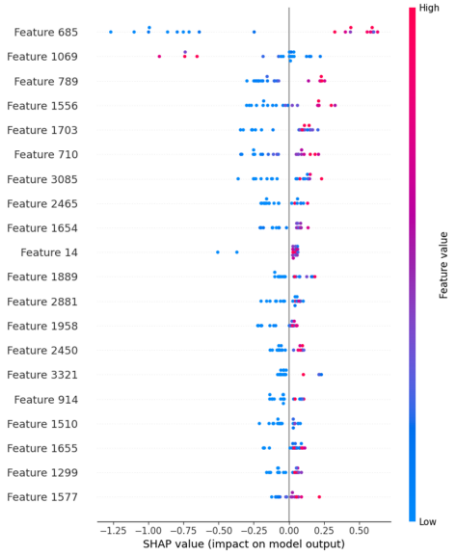

Figure 4: SHAP Explainability

Discussions

Deep feature extraction and ensemble machine learning enhance melanoma classification, according to experimental findings. SVM classifiers distinguished between benign and malignant lesions with higher specificity and classification accuracy. Better melanoma detection is predicted by XGBoost's higher sensitivity, which is clinically significant since false negatives increase mortality and delay treatment. After the LIME and SHAP explainability study, the suggested method is more dependable. Transparency and clinical decision-making are enhanced by the identification of significant deep features. Explainable AI can increase the trust of medical professionals and encourage the use of dermatological diagnosis systems. Despite its limitations, the suggested paradigm demonstrated promise. In contrast to clinical diversity, this study employed a tiny dataset. We only looked at binary classification. Advanced ensemble approaches, external dataset validation, and multiclass lesion classification should be the main topics of future research.

Conclusions

A hybrid system was created to accurately and eloquently classify dermoscopic images by combining deep feature fusion, Support Vector Machine (SVM), XGBoost, and explainable artificial intelligence. The SVM classifier outperformed XGBoost in terms of sensitivity and precision-recall for automated skin cancer detection, with 80.09% accuracy and 86.97% specificity. By determining the most crucial deep features for melanoma prediction, LIME and SHAP additionally enhanced model interpretability, increasing clinical dependability and transparency. The study only employed one dataset for binary classification, despite these encouraging findings. In order to enhance diagnostic performance and resilience, subsequent phases involve multiclass skin lesion categorization, validation of external clinical datasets, optimization of deeper convolutional neural network designs, and incorporation of more sophisticated ensemble learning techniques.

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).