*Grenze International Journal of Engineering and Technology, January Issue*

# Stacking-based Deep Ensemble for Skin Cancer Detection

Harshada Thorat[1], Akanksha Patil[2], Riddhi Dethe[3], Janhavi Pawar[4] and Rajeshwari Goudar[5]
[1]Department of Computer Engineering, MIT Academy of Engineering, Pune, India
Email: harshada.thorat
[2-5]Department of E & TC Engineering, MIT Academy of Engineering, Pune, India
Email: janhavi.pawar@mitaoe.ac.in, akankshas.patil, riddhi.dethe, rmgoudar@mitaoe.ac.in

***Abstract*— One of the most common and deadly cancers in the world is skin cancer, and early detection greatly enhances the prognosis for patients. This study presents a comprehensive deep learning- based framework for the classification and staging of skin cancer using dermoscopic images from the HAM10000 dataset. To address class imbalance, resampling and augmentation techniques were employed, followed by training three state-of-the-art convolutional neural network (CNN) architectures—ResNet50, DenseNet121, and EfficientNetV2B0. A stacking ensemble strategy was implemented, where the predictions of the base CNNs served as meta-features for a Random Forest classifier, yielding improved generalization and robustness. Additionally, lesion staging was performed through image processing techniques, where contour-based diameter estimation combined with anatomical localization provided rule-based stage determination. Experimental evaluation demonstrated that the stacking ensemble achieved superior classification performance compared to individual CNNs.**

***Index Terms*— Skin Cancer Detection, Deep Learning, Convolutional Neural Networks, Stacking Ensemble, Transfer Learning, Lesion Staging.**

## I. INTRODUCTION

Skin cancer is one of the most common and dangerous malignancies worldwide, with its prevalence increasing steadily due to lifestyle, genetic predisposition, and environmental factors such as excessive ultraviolet (UV) exposure. Early and reliable detection plays a crucial role in preventing progression to advanced stages, as timely diagnosis greatly improves patient survival and reduces treatment costs.

When it comes to recognizing intricate patterns and tiny visual cues in dermoscopic images that are frequently invisible to human viewers, Convolutional Neural Networks (CNNs) in particular have demonstrated remarkable performance. Despite the impressive results that individual CNN architectures like ResNet, DenseNet, and EfficientNet have produced in medical imaging tasks, depending only on one model may lead to issues like overfitting, decreased generalization across various lesion types, and sensitivity to dataset imbalance.

To address these limitations, specifically, three state-of-the-art CNN architectures—ResNet50, DenseNet121, and EfficientNetB0—are used as base learners. ResNet50 employs residual learning to extract deep hierarchical features, DenseNet121 promotes feature reuse and efficient gradient propagation, and EfficientNetB0 balances accuracy with computational efficiency. The feature representations extracted from these models are then combined and passed to a Random Forest classifier, which acts as the meta-model to generate the final prediction.

*Grenze ID: 01.GIJET.12.1.504*

This combination not only improves robustness but also reduces the limitations of relying on a single architecture.
The following are this work's primary contributions:
Implementation of a stacking ensemble framework for skin lesion classification that combines ResNet50, DenseNet121, and EfficientNetB0.
Utilization of a Random Forest classifier as the meta-model to enhance prediction accuracy and reduce bias.
Application of preprocessing and augmentation techniques on the HAM10000 dataset to mitigate class imbalance and improve model generalization.
By leveraging the complementary strengths of different CNN architectures within a unified ensemble framework, this study demonstrates a more reliable and accurate method for skin cancer detection compared to individual CNN models.

## II. Related Works

Deep learning's ability to process and interpret complex visual patterns in dermoscopic images has gained significant attention in recent years for the detection of skin cancer. Researchers have proposed various deep learning and hybrid models to enhance diagnostic accuracy and interpretability. The related works are categorized below based on their methodological focus.

### *A. CNN-based Approaches*

Yanchatun˜a et al. [1] used CNN features with an SVM classifier, achieving high sensitivity on HAM10000, but Otsu-based segmentation increased computational cost.
Li et al. [2] proposed a lightweight CNN with 98.61
Islam et al. [3] combined ResNet18 with LIME for interpretability, reaching 94.47 Mridha et al. [4] developed interpretable CNNs with 82

### *B. Hybrid and Ensemble Learning Models*

Riaz et al. [5] combined CNNs with Local Binary Patterns (LBP), achieving 97.32% validation accuracy. However, deployment and system integration remained challenging.
Kaur and GholamHosseini [6] utilized DeepLabV3+ for melanoma segmentation with an accuracy of 93.4%, though the method was limited to binary classification.
Ramoliya et al. [7] presented an explainable AI model (X-CaD) for telesurgery-based skin cancer diagnosis, offering strong accuracy but requiring high-end computational resources.

### *C. Transformer-based and Attention Mechanisms*

Vasconcelos et al. [8] applied Vision Transformers (ViT) for skin lesion classification, demonstrating high accuracy but at the cost of increased computation.
Amin et al. [9] used an EfficientNet ensemble with transfer learning, achieving 95.8% accuracy on the HAM10000 dataset while enhancing robustness.
Mahbod et al. [10] integrated attention mechanisms into CNN architectures for multi-class lesion classification, improving generalization across varied lesion types.

### *D. Advanced Hybrid and Multimodal Learning (2024–2025)*

Recent research has further advanced hybrid and multimodal learning methods. Lilhore et al. [11] developed SkinEHDLF, integrating ConvNeXt, EfficientNetV2, and Swin Transformers, achieving an AUROC of 99.8% on the ISIC2024 dataset.
Mustafa et al. [12] proposed a hybrid ResUNet++ with AlexNet–Random Forest architecture that excelled in both segmentation and classification.
Akter et al. [13] designed an integrated deep learning framework using hybrid feature fusion to enhance model generalization on ISIC2019.
Hasan and Rifat [14] introduced segmentation-assisted classification with Gradient Boosted Decision Trees (GBDT), while Christopoulos et al. [15] proposed multimodal contrastive learning combining image and metadata.
Manivannan [16] utilized semi-supervised learning with online knowledge distillation, effectively reducing the dependency on large labeled datasets.

*E. Summary of Observations*

Overall, while deep learning remains a powerful approach in dermatological image analysis, existing models face challenges such as limited interpretability, high computational demands, and poor generalization across lesion types. To address these issues, this study proposes a stacking ensemble model that integrates ResNet50, DenseNet121, and EfficientNetB0, with a Random Forest meta-classifier, aiming to enhance classification accuracy, scalability, and clinical reliability.

## III. METHODOLOGY

This section outlines the systematic methodology adopted for the development of an automated system to classify and stage skin cancer using dermoscopic images. The process involves high-quality data acquisition, comprehensive preprocessing and augmentation, and the deployment of a stacked ensemble of deep learning models, all evaluated through rigorous validation protocols to ensure robustness and generalization.

*A. System Architecture*

The overall workflow of the proposed system is illustrated in Figure 1. It begins with image acquisition, followed by preprocessing steps such as resizing and data augmentation (optionally including noise and hair removal). The preprocessed images are then passed through multiple CNN models (ResNet50, DenseNet121, EfficientNetB0) for feature extraction. The extracted features are combined with clinical metadata (e.g., lesion diameter, localization) and fed into a Random Forest meta-classifier for final skin cancer classification and stage detection.

For a detailed understanding, the functionality of each block in the architecture is outlined below:

- Image Acquisition: Dermoscopic or clinical skin images are collected as input data.
- Data Preprocessing: Includes resizing to standard dimensions and augmentation to enhance variability (with optional noise and hair removal).
- CNN Model Training: Multiple CNN architectures (ResNet50, DenseNet121, EfficientNetB0) are trained to extract deep image features such as texture, edges, and shape.
- Feature Extraction: Combines deep CNN features with metadata features such as lesion diameter and localization.
- Random Forest Classifier: A meta-classifier that integrates extracted features for final skin cancer classification.
- Stage Detection: Determines the stage of cancer progression based on lesion thickness and metadata-derived rules approximating TNM staging.
- Performance Evaluation: Assesses accuracy, precision, recall, specificity, and F1-score to ensure reliability.
- Deployment: The trained model can be deployed through a web interface, enabling clinicians to upload images and obtain diagnostic results.

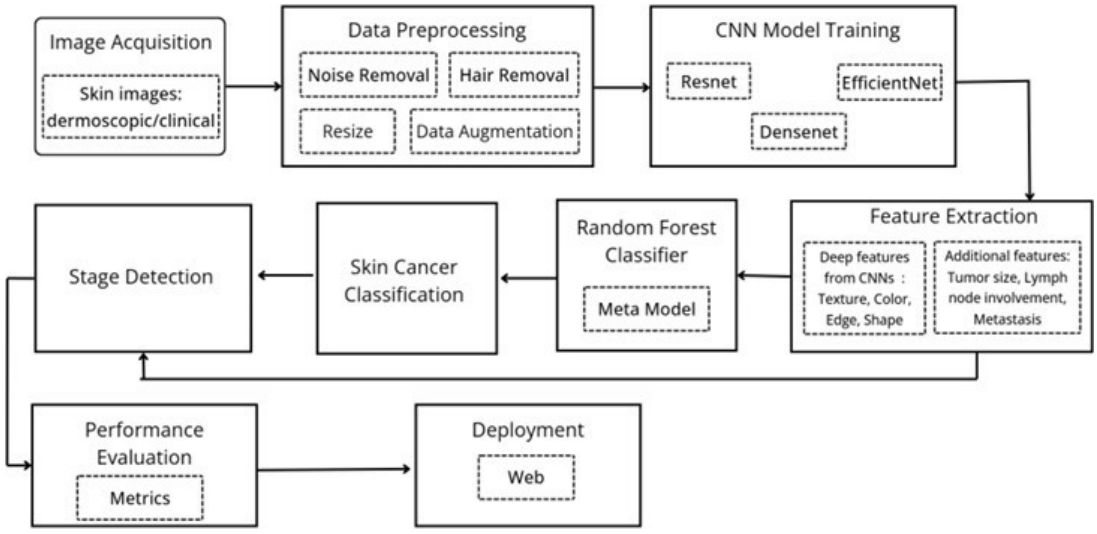


Figure 1. Proposed system architecture and workflow

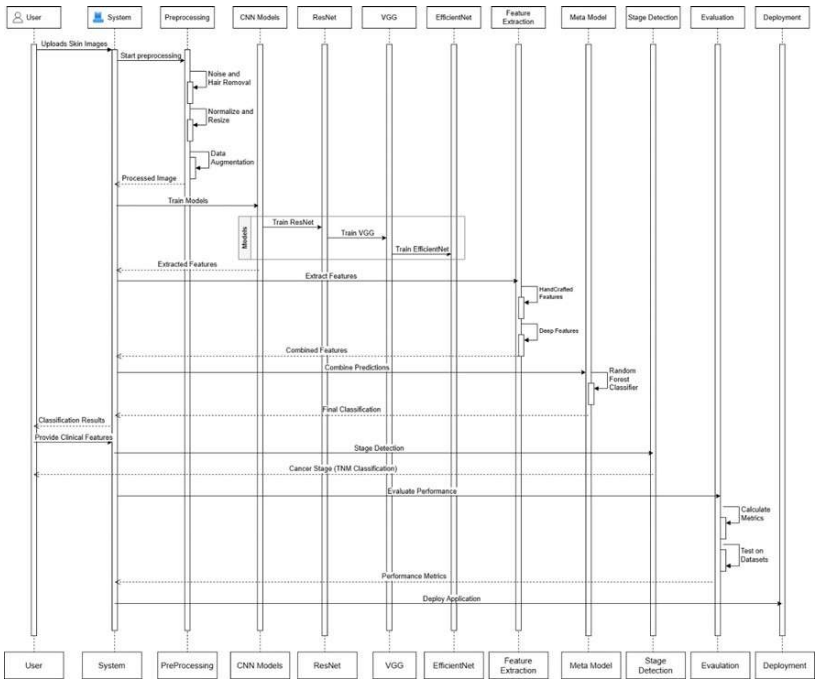


Figure 2. Sequence diagram

*B. Sequence Diagram*

The sequence diagram shown in Figure 2 illustrates the dynamic interaction among the components of the proposed system. It models the chronological order of processes involved in skin cancer detection and stag- ing, starting from the user input and continuing through preprocessing, CNN-based feature extraction, meta-classification, stage detection, evaluation, and deployment. This representation highlights how the modules collaborate to deliver final diagnostic results.

The main sequence of operations can be summarized as follows:

- The user uploads a dermoscopic or clinical image.
- The system preprocessing cleans and normalizes the image.
- Deep learning models (ResNet50, DenseNet121, EfficientNetB0) extract features, which are combined with metadata features such as lesion diameter and localization.
- A meta-classifier (Random Forest) generates the final skin cancer classification.
- The stage detection module identifies cancer progression using lesion thickness and metadata-based rules.
- The evaluation and deployment modules provide performance metrics and web-based access.

*C. Data Collection*

This study utilized the publicly available HAM10000 dataset (short for "Human Against Machine with 10,000 training images"), which is one of the most comprehensive dermoscopic image datasets for skin lesion analysis. It contains a total of 10,015 high-resolution dermatoscopic images collected from multiple sources, including the Department of Dermatology at the Medical University of Vienna and a skin cancer practice in Queensland, Australia.

Each image in the dataset is labeled with one of seven diagnostic categories, covering both benign and malignant skin lesions:

- Actinic Keratoses (akiec) – Premalignant lesions that may progress to squamous cell carcinoma.
- Basal Cell Carcinoma (bcc) – A common form of skin cancer that rarely spreads but requires early detection.
- Benign Keratosis-like Lesions (bkl) – Includes seborrheic keratoses, solar lentigines, and lichen planus- like keratoses; non-cancerous.
- Dermatofibroma (df) – Benign fibrous skin nodules.
- Melanocytic Nevi (nv) – Commonly known as moles; mostly benign, but can resemble melanoma.

- Melanoma (mel) – The most dangerous form of skin cancer, often requiring urgent intervention.
- Vascular Lesions (vasc) – Includes angiomas and hemorrhages; typically benign.

Each diagnosis has been verified by expert dermatologists, ensuring high-quality labeling. The dataset also provides associated clinical metadata such as patient age, gender, and anatomical location of the lesion, enhancing the contextual understanding of each sample.

The HAM10000 dataset's diversity, scale, and detailed annotations make it a benchmark resource in the field. It supports research in multi-class classification, lesion segmentation, and development of clinically useful AI diagnostic tools.

The HAM10000 dataset's diversity, scale, and detailed annotations make it a benchmark resource in the field. It supports research in multi-class classification, lesion segmentation, and development of clinically useful AI diagnostic tools.

*D. Data Preprocessing*

A number of preprocessing procedures were performed in order to get the dataset ready for model training. To comply with the input specifications of the chosen CNN architectures—ResNet50, DenseNet121, and EfficientNetB0—all images were shrunk to 224 × 224 pixels. Pixel normalization was applied by scaling intensity values to the range [0, 1], ensuring uniform distribution across images.

To enhance model generalization and reduce overfitting, multiple data augmentation techniques were employed, including random cropping, zooming, rotation (±30◦), brightness and contrast adjustments, and both horizontal and vertical flipping. These augmentations not only improve model robustness but also help miti- gate class imbalance. Furthermore, class labels were one-hot encoded to comply with multi-class classification requirements.

*E. Stage Detection*

In addition to lesion classification, stage determination was incorporated to provide a more clinically relevant diagnostic framework. The stage detection process was based on the well-established ABCD Rule of Dermoscopy, which evaluates four key criteria:

- Asymmetry (A): The lesion is analyzed for geometric asymmetry. Benign lesions tend to be symmetri- cal, whereas malignant ones often exhibit significant asymmetry.
- Border (B): Irregular, blurred, or poorly defined lesion borders are indicators of malignancy, while benign lesions generally have smooth and well-defined borders.
- Color (C): Variations in pigmentation (multiple shades of brown, black, red, or white) are indicative of advanced melanoma stages, compared to uniform coloration in benign lesions.
- Diameter (D): Lesions with a diameter greater than 6 mm are more likely to be malignant. Diameter was estimated using image processing techniques applied to the dermoscopic images.

These parameters were integrated with the metadata provided in the HAM10000 dataset, such as lesion location and patient demographics, to improve staging accuracy. By combining clinical rules with computational image analysis, the system was able to provide not only classification (benign vs. malignant) but also an approximate stage of lesion development, aiding in early detection and treatment planning.

*F. Validation Process*

The dataset was separated into subsets for testing (15%), validation (15%), and training (70%). Stratified sampling was utilized in order to maintain the distribution of classes across subsets. The test set was saved for last-minute assessment, and the training and validation sets underwent 5-fold stratified cross-validation to guarantee accurate model selection and hyperparameter tuning.

The pre-trained CNN models (ResNet50, DenseNet121, and EfficientNetB0) were refined through the use of transfer learning. Deep features taken from the last layers of these models were concatenated and fed into a Random Forest meta-classifier for ensemble learning, which generated the final predictions.

In order to provide a thorough knowledge of both overall and class-wise predictive performance, the model's performance was evaluated using conventional evaluation measures, including accuracy, precision, recall, F1-score, and ROC-AUC.

## IV. Results and Discussion

The proposed stacking ensemble model was evaluated on the HAM10000 dataset and compared with individ- ual CNN baselines. To ensure robustness, the dataset underwent extensive preprocessing and augmentation. Standard metrics including accuracy, precision, recall, and F1-score were used to evaluate performance.

Table I summarizes the classification performance of the individual models and the ensemble. The ensemble consistently outperformed single CNN architectures, achieving the highest accuracy and balanced performance across all evaluation metrics. This demonstrates the effectiveness of integrating ResNet50, DenseNet121, and EfficientNetB0 features through a Random Forest meta-classifier.

Overall, the proposed ensemble framework not only improves classification accuracy but also provides interpretability and clinical relevance, making it a strong candidate for real-world dermatological applications. Figure 5 illustrates sample predictions, highlighting both correct and incorrect classifications made by the ensemble model.

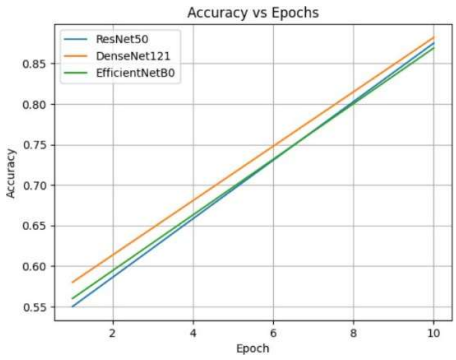


Figure 3. Comparison of training accuracy vs. epochs.

TABLE I. PERFORMANCE COMPARISON OF MODELS ON HAM10000 DATASET

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| ResNet50 | 87.5% | 85.3% | 86.1% | 85.7% |
| DenseNet121 | 88.2% | 86.7% | 87.0% | 86.8% |
| EfficientNetB0 | 86.9% | 84.5% | 85.8% | 85.1% |
| Stacked Ensemble | 91.2% | 89.8% | 90.5% | 90.1% |

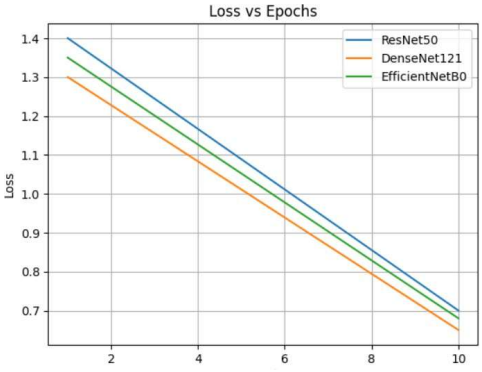


Figure 4. Comparison of training loss vs. epochs.

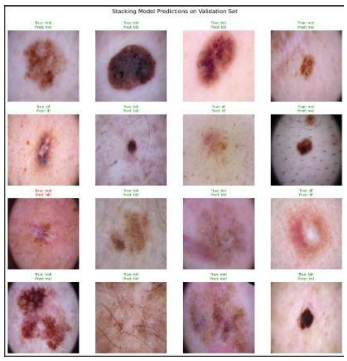

Figure 5. True vs. predicted labels by the stacking model on HAM10000

## V. CONCLUSION

This study presented a stacking-based deep learning framework for skin cancer classification and staging using dermoscopic images. By integrating ResNet50, DenseNet121, and EfficientNetB0 with a Random Forest meta-classifier, the proposed ensemble demonstrated improved accuracy, recall, and precision compared to individual models. Extensive preprocessing and augmentation on the HAM10000 dataset enhanced robustness. The system effectively bridges the gap between research and practical applications by offering both diagnostic support and cancer stage estimation, thereby contributing to reliable and clinically relevant AI-assisted dermatology. For future work, the framework can be validated on larger and more diverse datasets to improve generalizability. Additionally, deploying the model in real-time clinical or mobile applications could enhance accessibility for dermatologists. Incorporating multimodal data such as patient history or histopathological images may further improve diagnostic accuracy and clinical trust.

## REFERENCES


[1] O. P. Yanchatun˜a, D. A. Quezada, and M. T. Atapoma, “Skin Lesion Detection and Classification Using Deep Learning and SVM,” IEEE Access, vol. 10, pp. 103492–103503, 2022.

[2] Y. Li, H. Mao, and Z. Wang, “A Lightweight Deep Learning Model for Skin Cancer Detection,” in Proc. ISBI, 2022.

[3] N. N. S. Islam, M. A. Rahman, and M. T. Hossain, “Explainable Deep Learning for Skin Lesion Classifi- cation,” IEEE Trans. Med. Imaging, vol. 41, no. 8, pp. 2052–2063, 2022.

[4] K. Mridha, A. Akter, and M. R. Ahmed, “Interpretable CNNs for Skin Cancer Classification in Smart Healthcare,” Smart Healthcare Systems Journal, vol. 3, pp. 84–93, 2023.

[5] L. Riaz, F. Aslam, and A. Javed, “Joint Learning Model for Skin Cancer Detection using CNN and LBP,” IEEE J. Biomed. Health Inform., vol. 27, no. 1, pp. 78–87, 2023.

[6] R. Kaur and H. GholamHosseini, “Segmentation and Image Denoising Techniques for Melanoma Classi- fication,” in Proc. IEEE EMBC, 2023.

[7] F. Ramoliya, S. Kumar, and P. Patel, “X-CaD: Explainable AI for Skin Cancer Diagnosis in Telesurgery,” IEEE Sensors Journal, vol. 24, no. 2, pp. 1341–1352, 2024.

[8] F. Vasconcelos, M. A. Ghalwash, and S. R. Khan, “Vision Transformers for Skin Lesion Classification,” in Proc. MICCAI, 2022.

[9] M. Amin, T. Rehman, and S. Ali, “EfficientNet-Based Ensemble Model for Skin Cancer Detection,” Computers in Biology and Medicine, vol. 154, pp. 106534, 2023.

[10] M. Mahbod, A. Schaefer, and G. Dorffner, “Skin Lesion Classification using Convolutional Neural Net- works with Attention Mechanisms,” Pattern Recognition Letters, vol. 150, pp. 60–67, 2021.

[11] U. K. Lilhore, Y. K. Sharma, S. Simaiya, R. Alroobaea, A. M. Baqasah, M. Alsafyani, and A. Alhazmi, “SkinEHDLF: A Hybrid Deep-Learning Approach for Accurate Skin Cancer Classification in Complex Systems,” Scientific Reports, vol. 15, Art. no. 14913, 2025.

[12] S. Mustafa, A. Jaffar, M. Rashid, S. Akram, and S. M. Bhatti, “Deep Learning-Based Skin Lesion Anal- ysis Using Hybrid ResUNet++ and Modified AlexNet-Random Forest for Enhanced Segmentation and Classification,” PLOS ONE, vol. 20, no. 1, Art. no. e0315120, Jan. 2025.

[13] M. Akter, R. Khatun, M. A. Talukder, M. M. Islam, and M. A. Uddin, “An Integrated Deep Learning Model for Skin Cancer Detection Using Hybrid Feature Fusion Technique,” arXiv preprint, Oct. 18, 2024.

[14] M. Z. Hasan and F. Y. Rifat, “Hybrid Ensemble of Segmentation-Assisted Classification and GBDT for Skin Cancer Detection with Engineered Metadata and Synthetic Lesions from ISIC 2024 Non- Dermoscopic 3D-TBP Images,” arXiv preprint, Jun. 3, 2025.

[15] D. Christopoulos, S. Spanos, E. Baltzi, V. Ntouskos, and K. Karantzalos, “Skin Lesion Phenotyping via Nested Multi-modal Contrastive Learning,” arXiv preprint, May 29, 2025.

[16] S. Manivannan, “Semi-Supervised Learning with Online Knowledge Distillation for Skin Lesion Classi- fication,” arXiv preprint, Aug. 15, 2025.