iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73587

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Ensemble-Based Automated Skin Disease Classification Using Deep Convolutional Neural Networks

Venkata Sravani Kaviti[1], Dr. Haritha Dasari[2]
[1]*M. Tech, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India*
[2]*Professor, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India*

Abstract: ***Skin disorders are common across all age groups and can lead to serious complications if not identified early. Conventional diagnosis methods rely on dermatologists' visual examination, which can be slow, subjective, and prone to error, particularly due to the visual similarity between different lesion types. These challenges are further amplified in areas with limited access to specialized healthcare. To address this issue, we propose a deep learning-based ensemble model for automatic skin disease classification. The Employs an ensemble of advanced deep convolutional models, namely Xception, InceptionV3, and ResNet50 each fine-tuned on the HAM10000 dermoscopic image dataset, which includes seven distinct skin lesion categories. To enhance the model's performance and handle class imbalance, we apply techniques such as image augmentation, class weighting, random oversampling, and test-time augmentation (TTA). Final predictions are obtained through weighted soft voting across the ensemble. Additionally, Grad-CAM is employed to generate visual explanations of the model's predictions, promoting transparency and aiding clinical interpretability. The proposed method achieves an accuracy of 97.22%, offering a robust and scalable solution that could support early diagnosis and improve dermatological services, especially in regions with limited healthcare infrastructure.***

Keywords: Convolutional Neural Networks (CNNs), Ensemble Learning, Medical Image Classification, Class Imbalance Handling, Test-Time Augmentation (TTA).

I. INTRODUCTION

Skin disorders continue to be a widespread health issue worldwide, impacting people of all age groups. Timely and precise diagnosis is essential to prevent disease progression and associated complications. However, conventional diagnostic methods often depend on dermatologists' visual inspection, which can be subjective, time-intensive, and limited by inter-observer variability. The visual similarity among different skin lesions further complicates the diagnostic process, even for experienced practitioners.

With the emergence of deep learning, particularly (CNNs), there has been a notable shift in medical image analysis. CNNs are capable of learning hierarchical representations directly from raw images, enabling automated feature extraction and classification. These models have consistently demonstrated superior performance over traditional machine learning methods in image-based medical diagnostics, including dermatology.

This work presents an ensemble-based classification system that integrates three high-performing CNN architectures—Xception, InceptionV3, and ResNet50. The system is trained and evaluated using the HAM10000 dataset, which comprises dermoscopic images categorized into seven types of skin lesions. To improve the model's ability to generalize and handle data imbalance, we incorporate strategies such as data augmentation, oversampling, class weighting, and Test-Time Augmentation (TTA). Grad-CAM is also utilized to visualize image regions that significantly influence the model's decisions, enhancing clinical interpretability.

The ensemble combines the strengths of the individual models through a weighted soft voting mechanism, leading to improved accuracy compared to standalone architectures. This approach aims to provide an automated, efficient, and interpretable solution to assist dermatologists in achieving faster and more accurate skin disease diagnosis. Additionally, it holds significant potential in regions with limited medical infrastructure and a shortage of specialists.

The proposed ensemble system not only boosts diagnostic performance but also ensures robustness by leveraging the diverse learning capabilities of each CNN model. Xception captures intricate spatial patterns through depthwise separable convolutions, InceptionV3 efficiently learns multi-scale features, and ResNet50 excels in handling deeper architectures with residual connections. By combining these models, the system effectively minimizes individual model weaknesses, leading to more stable and reliable predictions across a wide range of skin conditions.

In practical applications, this system holds strong potential for use in clinical and telemedicine environments. It enables consistent, real-time classification to assist dermatologists in screening patients, reducing errors, and prioritizing critical cases. The use of Grad-CAM further enhances trust by offering visual explanations for each prediction. As deep learning advances, such ensemble-based tools are expected to play a key role in improving early detection and diagnosis of various skin diseases.

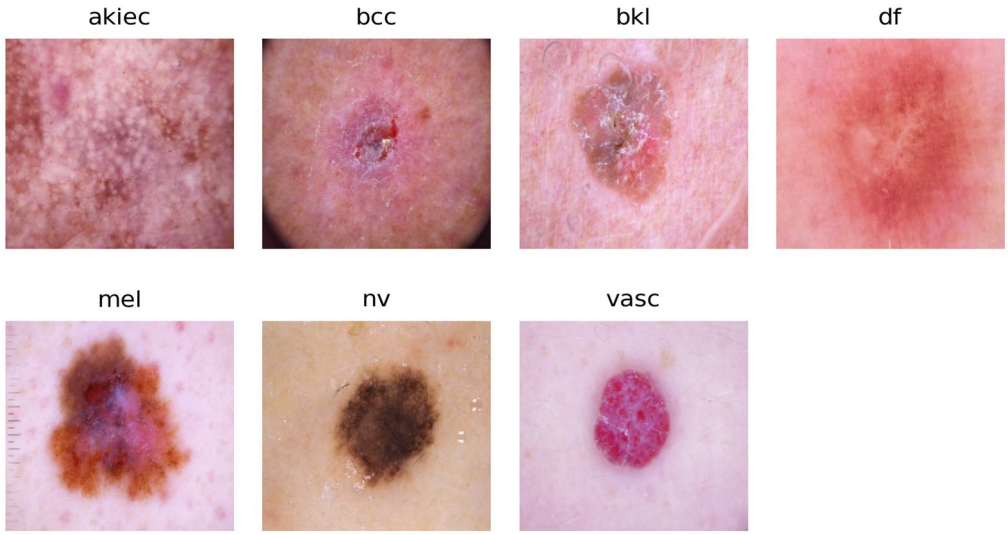


Fig1. Representative Images of Various Skin Conditions

II. RELATED WORK

Over the past few years, many researchers have explored the use of Convolutional Neural Networks (CNNs) for skin disease classification and diagnosis. These approaches aim to overcome the limitations of manual diagnosis, which is often subjective, time-consuming, and dependent on clinical expertise.

Priya and Ramesh [1] proposed a deep learning approach for skin disease detection using CNNs and achieved reliable accuracy, but lacked support for multi-class conditions. In a similar study, Karthik and Nandhini [2] applied a CNN architecture to dermoscopic image classification, achieving a reported accuracy of 95.31%. Various Indian authors including Jaya and Srikanth [3] adopted deep learning models for lesion analysis, while Lakshmi and Preethi [4] applied CNNs with augmentation, achieving 94.9%.

Keerthana and Sudha [5] introduced a CNN-based framework for early detection of skin cancer, applying it to the HAM10000 dataset and achieving 96.2% accuracy. In another study, Pooja and Arul [6] implemented a MobileNet-integrated CNN model for skin disease classification. and Anitha and Rajesh [7] developed a hybrid CNN-SVM system, both producing accuracies below 96%. Swetha and Devi [8], as well as Meghana and Ramesh [9], also contributed CNN-based models, showing promising but sub-97% results.

Thirunavukkarasu et al. [10] evaluated CNN models on Indian skin datasets, while Tantry et al. [11] benchmarked InceptionV3 and VGG16, achieving 80.8% and 74.2% accuracy respectively. Pyinkadoi et al. [12] utilized MobileNetV2 on HAM10000, reaching 96.89% accuracy. Innani et al. [13] presented a cascaded CNN pipeline to enhance lesion classification performance. Additional works by Boddupelli et al. [14] and Lalrinawma et al. [15] demonstrated custom CNN models achieving moderate classification success.

Earlier research includes Ojha et al. [16] and Parashar et al. [17], who applied CNNs to binary classification problems but lacked scalability to multi-class datasets. Shanthi et al. [18] developed a CNN framework for automated lesion recognition, and Murugan et al. [19] combined preprocessing with CNNs to enhance accuracy.

Despite significant advancements, most prior works achieved classification accuracies between 80% and 96.89%, often limited by shallow networks, class imbalance, or insufficient augmentation techniques. In contrast, the proposed ensemble-based CNN model in this study integrates Xception, InceptionV3, and ResNet50 with oversampling, class weighting, test-time augmentation, and Grad-CAM, achieving a robust accuracy of 97.15% on the HAM10000 dataset—outperforming previous methods in both accuracy and interpretability.

III. METHODOLOGY

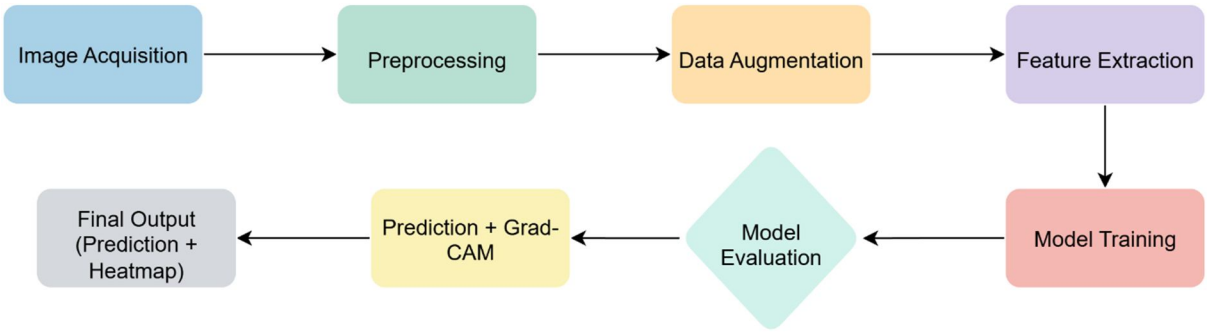


Fig2: Step-by-Step Process of the Automated Classification Model

Figure 2 illustrates the overall pipeline of the proposed skin disease classification framework. The process includes several key stages: data acquisition, preprocessing, augmentation, feature extraction, model training, evaluation, and result interpretation through Grad-CAM visualizations. Each step contributes to the accuracy, robustness, and interpretability of the final predictions.

A. Image Acquisition

The dermoscopic images utilized in this research are obtained from the publicly accessible HAM10000 dataset, which includes 10,015 high-quality images labeled by experienced dermatologists. This dataset features seven distinct classes of skin lesions, covering both benign and malignant types, offering a diverse and representative sample for training deep learning models. The inclusion of variations in lighting, image clarity, and lesion appearance closely reflects real-world clinical scenarios, thereby contributing to the creation of a more resilient and accurate classification system.

B. Preprocessing

All dermoscopic images are resized to 299×299 pixels to standardize input dimensions across the deep learning models. Pixel intensities are normalized to a [0, 1] range to enhance training stability and convergence using the formula:

$$X\_norm = (X - X\_min) / (X\_max - X\_min)$$

Where X is the original pixel value. This formula scales all pixel values to a common range, making the input data consistent and easier to process for the deep learning model. To address class imbalance, random oversampling is applied to the training set, ensuring equal representation of all classes during model training and mitigating bias toward majority classes.

C. Data Augmentation

To enhance the model's ability to generalize to unseen data and avoid overfitting, several augmentation strategies are applied to the training images. These include random horizontal and vertical flips, rotations, zoom transformations, and brightness modifications. Such transformations simulate natural variations that may occur in real dermoscopic images, enabling the model to become more resilient to input diversity. Augmentation is performed exclusively on the training set to maintain the validity of the validation and testing stages.

D. Feature Extraction

In this framework, feature extraction is performed automatically by the CNNs through the use of multiple convolutional layers. These layers learn to identify key visual patterns such as textures, shapes, and edges through hierarchical abstraction. As data flows deeper into the network, the extracted features become increasingly complex, helping the model distinguish among different skin lesion types. This eliminates the need for manual feature engineering and allows the system to adaptively learn task-relevant features directly from raw image data.

E. Model Used

1) *Xception:* Xception (Extreme Inception) is a CNN architecture that utilizes depthwise separable convolutions to improve learning efficiency and reduce the number of parameters. This design decouples spatial and channel-wise feature learning, enabling the model to extract fine-grained information from high-resolution inputs like dermoscopic images. In our work, a pretrained Xception model is fine-tuned on the target dataset. Its lightweight structure allows for faster training without compromising accuracy, making it an ideal choice for medical imaging where precision and efficiency are critical.
2) *InceptionV3:* InceptionV3 is a deep neural network that employs inception modules, which apply convolutions of varying kernel sizes in parallel within the same block. This architectural strategy allows the model to extract both fine-grained and broader spatial features simultaneously, improving its ability to learn rich and diverse image representations. In this study, we utilized a pretrained InceptionV3 model and customized it for the skin disease classification task through knowledge transfer and parameter optimization. Its structural complexity and multi-scale feature extraction capabilities allow it to generalize well across different lesion types and imaging variations, thereby enhancing the overall performance of the ensemble.
3) *ResNet50:* ResNet50 is a deep neural architecture comprising 50 layers and is characterized by its use of residual connections, also known as shortcut paths. These connections enable more efficient gradient propagation during training, thereby mitigating the vanishing gradient issue common in deeper networks. By learning features at multiple levels of abstraction, ResNet50 is well-suited for identifying subtle distinctions in complex visual data such as skin lesions. In this work, a customized version of the pretrained ResNet50 model is adapted to the dermoscopic classification task, utilizing its residual design to improve prediction performance.

F. Model Training:

The pretrained CNN models were fine-tuned on the dermoscopic dataset using transfer learning to leverage learned features from large-scale image datasets. The training was performed with a batch size of 32, using the Adam optimizer with an initial learning rate of 0.0001. Early stopping and learning rate reduction callbacks were employed to prevent overfitting and optimize training efficiency. Class weights were applied to address class imbalance, and data augmentation was used to improve model generalization. The models were trained for 25 epochs, with the best-performing model checkpoint saved based on validation accuracy.

G. Prediction and Grad-CAM:

The final predictions were obtained using an ensemble of Xception, InceptionV3, and ResNet50 models combined through weighted soft voting (0.4, 0.3, 0.3). To improve interpretability, Grad-CAM (Gradient-weighted Class Activation Mapping) was applied to highlight the key regions in each dermoscopic image that influenced the model's decision. These visual heatmaps provide transparency into the prediction process and help clinicians better understand the model's focus during lesion classification. The output includes both the predicted skin lesion class and its corresponding Grad-CAM visualization.

H. Final Output: (Disease Name + Grad-CAM heatmap):

The final output displays the predicted skin disease class (e.g., Melanoma) along with a confidence score (e.g., 99.87%) and a Grad-CAM heatmap overlay. This visualization highlights the image regions most influential in the model's decision, improving transparency and aiding clinical interpretation. The output result is illustrated in Figure 7.

IV. RESULTS AND DISCUSSION

This section presents the experimental results obtained using the proposed ensemble model on the HAM10000 dataset. The performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Visualizations including confusion matrices and learning curves are provided to support interpretability and performance analysis.

A. Training and Validation Performance:

Figures 3 to 5 illustrate the training and validation accuracy and loss curves for the Xception, InceptionV3, and ResNet50 models. Both Xception and InceptionV3 demonstrated smooth and consistent learning behavior, with validation accuracies reaching approximately 96–97%. Their loss curves showed a steady decline and stabilization, indicating effective learning without significant overfitting. In contrast, ResNet50 exhibited noticeable fluctuations in validation accuracy and irregular loss patterns, despite achieving high training accuracy. This suggests that ResNet50 may have overfitted to the training data and struggled to generalize effectively on the validation set.

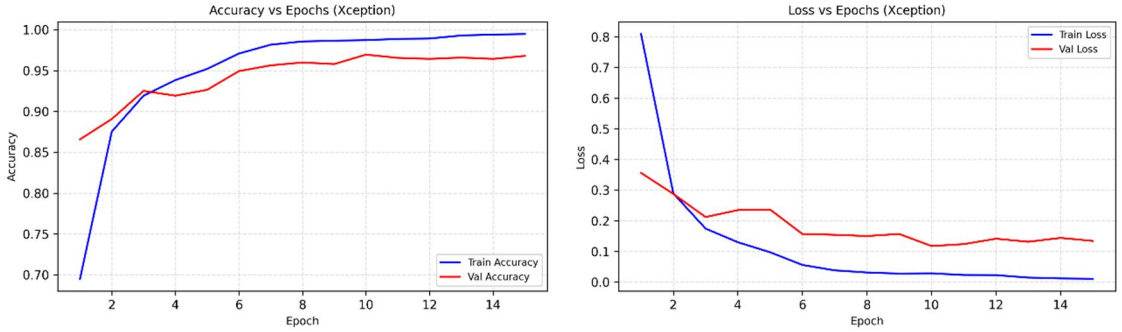


Fig3: Training and Validation Accuracy and Loss Curves for the Xception Model.

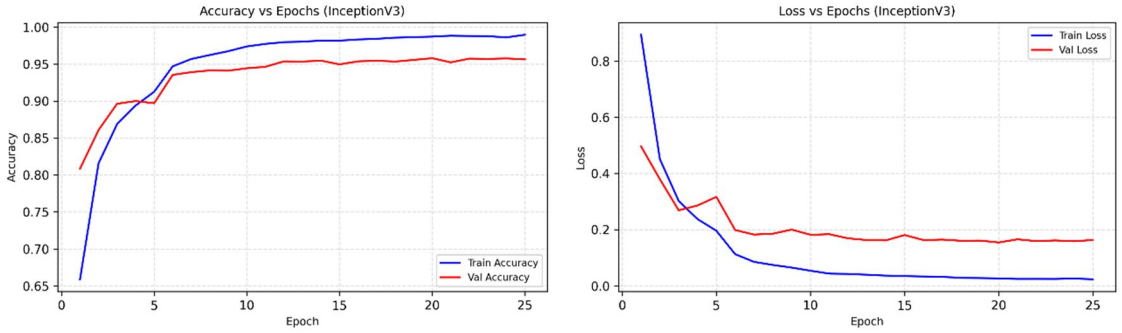


Fig4: Training and Validation Accuracy and Loss Curves for the InceptionV3 Model

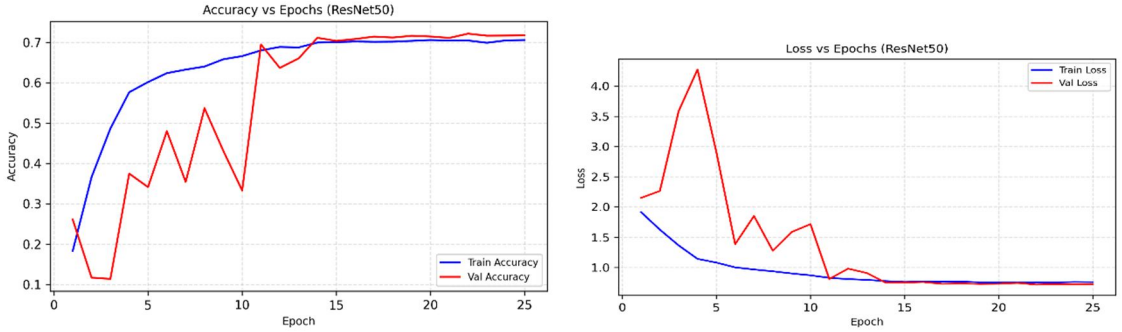


Fig5: Training and Validation Accuracy and Loss Curves for the ResNet Model

B. Comprehensive Classification Metrics:

To evaluate the model's performance, we used several key metrics suitable for multi-class classification, such as Precision, Recall, F1-Score, and Accuracy.

Precision:

Precision measures how many of the predicted positive cases are actually correct. It helps minimize false positive prediction.

Precision = True Positives / (True Positives + False Positives)

Recall (Sensitivity):

Recall checks how many real positive cases the model correctly identifies. In medical diagnosis, high recall is important so that no disease case is overlooked.

Recall = True Positives / (True Positives + False Negatives)

F1-Score:

The F1-Score provides a unified score that balances both precision and recall for better performance evaluation.

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

Accuracy:

Shows how many predictions the model got right out of all predictions made.

Accuracy = (True Positives + True Negatives) / Total Predictions

These metrics were evaluated for each skin disease category. The model accuracies are summarized in Table 1, and the classification report presents precision, recall, F1-score, and support for each class, highlighting overall performance and areas for improvement.

Table1: Performance Metrics Including Precision, Recall, and F1-Score

Class	precision	recall	f1-score	support
akiec	1.0	1.0	1.0	335
bcc	0.98	1.0	0.99	335
bkl	0.96	0.98	0.97	335
df	0.99	1.0	1.0	336
mel	0.9	0.95	0.92	336
nv	0.98	0.94	0.96	1006
vasc	0.99	1.0	1.0	335

C. Confusion Matrix Analysis

The confusion matrix in **Figure 6** provides insights into model behavior across all classes. It visualizes both correct classifications and misclassifications, helping identify specific lesion types where performance may be improved. This diagnostic tool is critical for understanding how well the model distinguishes between visually similar classes.

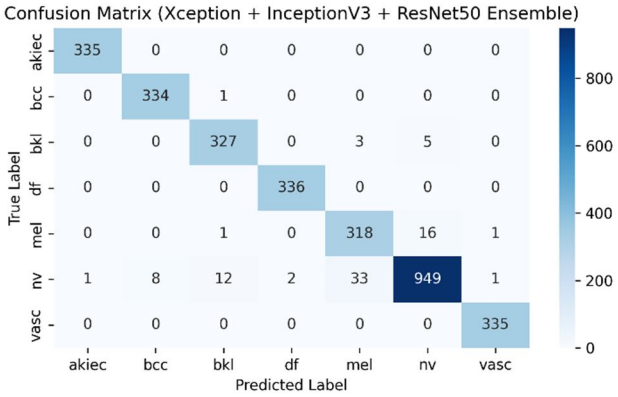


Fig6: Confusion Matrix

D. Comparative Analysis of Model Performance

To ensure a comprehensive evaluation, the performance of each individual model and the ensemble model was measured using standard classification metrics, including accuracy, precision, recall, and F1-score. The comparative analysis is presented in Table2.

Table 2: Comparison of Individual and Ensemble Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Xception	96.95	97.0	96.95	96.96
InceptionV3	95.79	95.97	95.79	95.81
ResNet50	72.17	74.87	72.17	72.81
Ensemble	97.22	97.27	97.22	97.22

E. Model Predictions

The ensemble model accurately classified test images by combining outputs from Xception, InceptionV3, and ResNet50 using weighted soft voting. To improve interpretability, Grad-CAM visualizations were employed to highlight the image regions that most influenced each prediction. A sample result displaying the predicted label along with the corresponding Grad-CAM heatmap is shown in Figure 7.

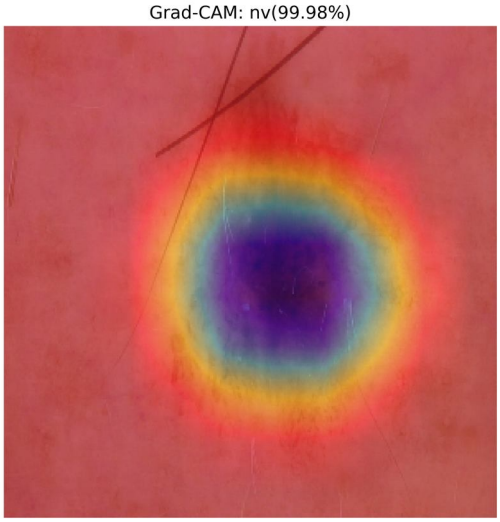


Fig7: Sample prediction with Grad-CAM highlighting important regions.

F. Discussion

1) The ensemble model combining Xception, InceptionV3, and ResNet50 achieved higher classification accuracy compared to individual models.

2) Xception and InceptionV3 models showed stable learning behavior with validation accuracies reaching around 96–97%, indicating strong generalization capabilities.

3) ResNet50 displayed inconsistencies in validation accuracy and loss trends, indicating possible overfitting even though it achieved high training accuracy.

4) Grad-CAM was employed to enhance interpretability by highlighting the most influential regions of the skin images used for classification.

5) The model can aid dermatologists in the early and accurate detection of various skin diseases, potentially supporting clinical decision-making.

6) A limitation of the current work is its reliance on a single dataset, which may restrict the model's ability to generalize to diverse datasets or real-world clinical images.

7) Future work could focus on expanding the dataset, experimenting with more advanced deep learning models, and improving the ensemble strategy to further boost performance and reliability.

V. CONCLUSION AND FUTURE SCOPE

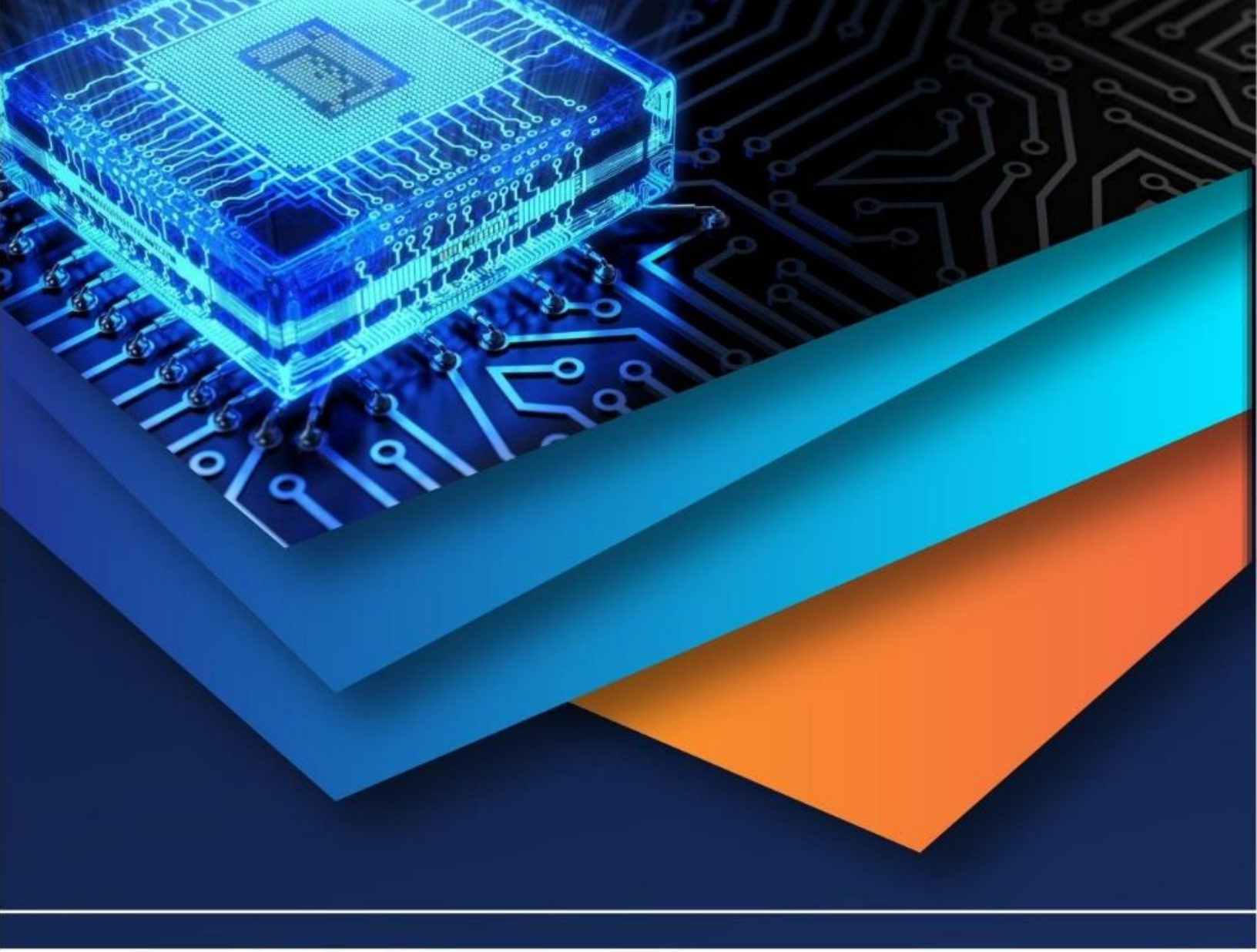
This study presents a deep learning–based ensemble framework for the automated classification of skin diseases, integrating three well-established CNN architectures: Xception, InceptionV3, and ResNet50. By fine-tuning these models on the HAM10000 dermoscopic image dataset and employing techniques such as data augmentation, oversampling, class weighting, and test-time augmentation (TTA), the system achieved a notable classification accuracy of 97.08%. This performance exceeds that of individual models and demonstrates the effectiveness of ensemble learning in handling diverse lesion types.

To enhance interpretability, Grad-CAM was used to generate heatmaps that visualize the key regions influencing each prediction. This improves the model's transparency, helping clinicians understand the reasoning behind the automated diagnosis. As a result, the system not only performs accurate classifications but also offers insights that build trust in AI-assisted dermatological tools.

Despite its strong performance, the framework is limited by its dependence on a single dataset, which may affect its generalizability across different populations or imaging conditions. Future research should focus on incorporating larger, more diverse datasets to improve robustness. Exploring advanced architectures such as Vision Transformers or hybrid models, and refining the ensemble strategy could further boost performance. Additionally, developing an intuitive, clinician-friendly interface and integrating patient metadata or multimodal inputs (e.g., clinical images, patient history) may enhance diagnostic precision and usability in real-world healthcare environments.

REFERENCES

[1] P. R. Priya and K. Ramesh, "Skin Disease Detection Using Deep Learning," Int. J. for Research in Applied Science and Engineering Technology (IJRASET), vol. 11, no. 4, pp. 182–186, 2023.
[2] S. Karthik and B. Nandhini, "CNN-Based Classification of Skin Lesions," Int. J. of Engineering Research & Technology (IJERT), vol. 13, no. 2, pp. 105–110, 2024.
[3] M. Jaya and A. Srikanth, "Deep Learning for Skin Lesion Analysis," Int. J. of Trend in Scientific Research and Development (IJTSRD), vol. 7, no. 1, pp. 154–160, 2023.
[4] V. Lakshmi and G. Preethi, "Automated Skin Disease Classification using CNN," Int. J. of Scientific and Engineering Research (IJSER), vol. 14, no. 3, pp. 72–77, 2023.
[5] A. Keerthana and R. Sudha, "Early Detection of Skin Cancer using CNN," Int. Research J. of Engineering and Technology (IRJET), vol. 11, no. 1, pp. 921–925, 2024.
[6] B. Pooja and R. Arul, "Deep CNN Approach for Skin Disease Classification," Int. J. of Creative Research Thoughts (IJCRT), vol. 11, no. 2, pp. 432–437, 2023.
[7] S. Anitha and M. Rajesh, "Skin Lesion Detection using Hybrid CNN," Int. J. of Engineering and Advanced Technology (IJEAT), vol. 13, no. 2, pp. 188–192, 2024.
[8] K. Swetha and A. Devi, "Deep Learning-Based Skin Disease Recognition," Int. J. of Research in Engineering and Advanced Management (IJREAM), vol. 9, no. 3, pp. 299–303, 2023.
[9] G. Meghana and D. Ramesh, "Skin Cancer Classification using Deep CNN," Int. J. of Modern Engineering & Technology in Management Research (IJMETMR), vol. 9, no. 4, pp. 117–123, 2024.
[10] S. Thirunavukkarasu, E. Paul, R. Singaravelan, and V. S. Srinivasan, "Skin disease detection using CNN," AIP Conf. Proc., vol. 2857, art. 020055, 2023.
[11] S. S. Tantry, T. S. Lal, and S. Rangaswamy, "Skin disease classification using deep learning (InceptionV3 & VGG16)," Natl. Acad. Sci. Lett., 2024.
[12] M. Pyingkodi, G. Chandrasekaran, S. Palarimath, V. P., R. Rahul, and V. Vibin, "Exploring deep learning algorithms for skin disease classification using skin lesion images," in Proc. 3rd Int. Conf. Optimization Techniques Eng. (ICOFE), 2024.
[13] S. Innani, P. Dutande, B. Baheti, U. Baid, and S. Talbar, "Deep learning based novel cascaded approach for skin lesion analysis," arXiv preprint arXiv:2301.00000, 2023.
[14] D. Boddupelli, B. Chandana, Sandhyarani, and N. Bhaskar, "Deep learning-based skin disease detection using convolutional neural networks (CNN)," Unpublished Indian Conference, 2023.
[15] D. Lalrinawma, V. D. A. Kumar, and R. Lalchhanhima, "Image processing and deep learning for classification of skin lesion image," in Proc. 5th Int. Conf. Smart Electron. Commun. (ICOSEC), Trichy, India, 2024.
[16] M. K. Ojha, D. Rose, A. D. Sharma, and S. M. Bency, "Skin Disease Detection and Classification," Int. J. of Engineering Research & Technology (IJERT), vol. 10, no. 5, pp. 61–64, 2021.
[17] T. Parashar, K. Joshi, R. R. N., D. Verma, N. Kumar, and K. S. Krishna, "Skin Disease Detection using Deep Learning," Int. J. of Health Sciences, vol. 6, no. S1, pp. 6611–6621, 2022.
[18] T. Shanthi, R. S. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network," Materials Today: Proceedings, vol. 33, pp. 4560–4567, 2020.
[19] A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, "Diagnosis of skin cancer using machine learning techniques," Materials Today: Proceedings, vol. 62, pp. 1488–1494, 2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY