RESEARCH ARTICLE

Transformer Enabled ResNet Based Automated Skin Cancer Detection System

Simran Negi[1] **and Aditya Joshi**[1,*]

[1] Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun 248002, India

Abstract

Early diagnosis plays a critical role in the successful skin cancer treatment. To classify skin lesions as benign or malignant, this study suggested a deep learning-enabled method that combines a Transformer module with ResNet50. At first ResNet50, a powerful convolutional neural network, to extract the image features and then enhance the model with a Transformer layer to improve the model accuracy is used. The final model is fine-tuned to achieve better accuracy. This approach shows improved classification results compared to the traditional model. The results signify that combining CNN-based feature extraction with Transformer-enabled global attention suggestively improves skin lesion classification, making the suggested framework a promising method for early and consistent skin cancer detection in intelligent healthcare systems.

Keywords: ResNet50, deep learning, convolutional neural networks, skin cancer.

Submitted: 20 December 2025
Accepted: 27 December 2025
Published: 31 December 2025

Vol. 1, **No.** 3, 2025.
10.62762/BISH.2025.444143

***Corresponding author:**
✉ Aditya Joshi
adi.joshi@geu.ac.in

1 Introduction

Skin cancer, especially Melanoma (MEL), Basal Cell Carcinoma (BCC), and Actinic Keratoses (AKIEC), can become life-threatening if not detected early. As per the World Health Organization (WHO), more than 5 million non-melanoma skin cancers and approximately 325,000 new melanoma cases are reported annually. Traditional diagnosis methods rely on expert dermatologists, which may consume time and be subject to human error. Automated detection systems are growing increasingly with more accurate results than humans.

Convolutional Neural Networks (CNNs) are a class of deep learning models widely emerged as the most effective tool for skin lesion classification [1], distinguishing between various skin lesions with high accuracy. These models perform feature extraction and learn from dermoscopic images and learn from dermoscopic images using large datasets. Deep learning-based models use labelled data for classification, and manually labelling thousands of images to train models is time-consuming; also, training the models takes a lot of time for large datasets, and that is why a transfer learning is used, where various models are pre-trained on large datasets. Moreover, transfer learning, where models are pre-trained for general tasks and then fine-tuned for specific tasks, has enhanced performance. Deep

Citation
Negi, S., & Joshi, A. (2025). Transformer Enabled ResNet Based Automated Skin Cancer Detection System. *Biomedical Informatics and Smart Healthcare*, 1(3), 149–154.

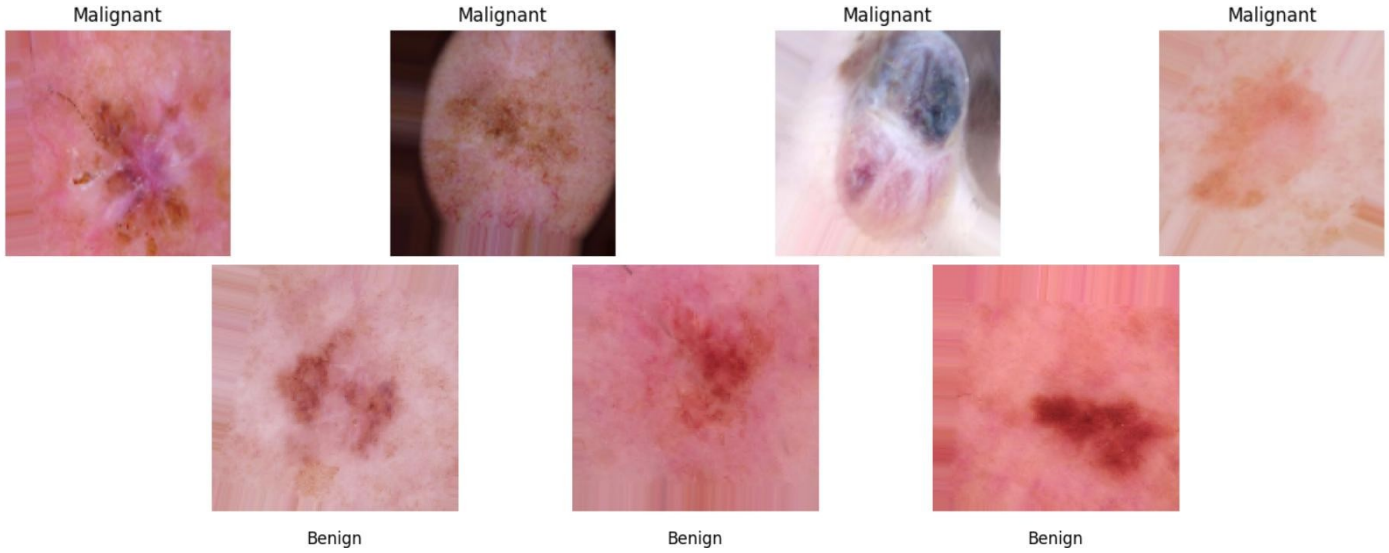

Figure 1. Malignant and Benign samples.

learning methods have demonstrated outstanding performance in the area of medical image diagnostics in recent years.

In this study, a deep learning-enabled method for detecting skin cancer from dermoscopic images is proposed. A combination of ResNet50, a powerful convolutional neural network that includes skip connections, which solves the Vanishing Gradient problem, and a Transformer module, which concentrates on extracting the global patterns from images is used. The model is trained to perform better on the task of classifying skin lesions. The Adam optimization technique, which is frequently employed in deep learning, also allowed for quicker convergence and more variable learning rates, which kept the model from becoming trapped in local minima and enhanced performance overall [2, 3].

For training and evaluation, the ISIC 2018 skin lesion dataset is utilized, which contains a large collection of labeled dermoscopic images representing 7 different types of skin lesions the samples are illustrated in Figure 1. Our goal is to develop an accurate and robust model that can assist dermatologists in detecting harmful lesions at an early stage.

The rest of the paper consists of the following sections: Section 2 covers the literature review, Section 3 consists of proposed methodology, Section 4 includes the experimental results, and Section 5 includes the conclusion.

2 Related Work

Several algorithms, including CNN, RNN, Decision Trees, and SVM, are employed in the paper [8] to identify oral and skin cancers using hospital datasets. Their model attained an accuracy of 92%. Another research [9] introduced a hybrid method combining CNN with KNN and SVM. Their model was tested on 640 images from the ISIC dataset and achieved an accuracy of 88.4%. This model struggled with outliers. Another recent paper [4] introduces a sophisticated deep learning architecture called the Multi-Scale Feature Fusion Framework (MFFF) for detecting multiple classes of skin lesions. The researchers used DenseNet201 as the base model, enhanced with three attention mechanisms - Channel Attention (CA), Soft Attention (SA), and Squeeze-and-Excitation Attention (SEA). Class imbalance and CNN models' lack of interpretability are the two main issues that the model attempts to solve. They worked on the ISIC 2018 dataset and achieved an outstanding accuracy of 97.15%. Another study [5] combines three powerful pre-trained models - VGG16, VGG19, and ResNet50 to enhance diagnostic performance. Features from all three models were concatenated and passed through fully connected layers using a sigmoid classifier. They performed training on a balanced dataset of 210 images (120 melanoma, 90 non-melanoma) and later augmented to 2000 images. Using the Adam optimizer, early stopping, and learning rate reduction, the model achieved an accuracy of 93.5%. Another study [6] develops a custom CNN model, "DermalDetect", that classifies skin lesions as benign or malignant. The model is built from scratch with 3 convolutional

Table 1. Summary of related work.

Ref.	Task	Method	Dataset	Outcome	Limitations
[4]	Multi-class skin lesion	DenseNet201-based Multi-Scale Feature Fusion Framework (MFFF) with CA, SA, SEA	ISIC 2018 dataset	97.15%	Complex model
[5]	Skin lesion	VGG16, VGG19, ResNet50 (feature fusion)	210 images (balanced), augmented to 2000	93.5%	Small dataset
[6]	Binary skin lesion classifier	Custom CNN ("DermalDetect")	9000 dermoscopic images	96%	Limited to binary classification
[7]	Large-scale skin disease	GoogleNet Inception v3 (transfer learning)	135,550 images, 2,055 disease categories	Outperformed 21 dermatologists	Required High computation
[8]	Oral and skin cancer	CNN, RNN, Decision Tree, SVM	Hospital dataset	92%	Specific dataset
[9]	Skin cancer	Hybrid CNN + KNN + SVM	ISIC dataset with 640 images	88.4%	More Outliers
[10]	Skin lesion	ResNet50 + LSTM	ISIC2020 and HAM10000	95.72% and 94.23%	Complex hybrid model

layers followed by dense layers, batch normalization, dropout, and SoftMax for final classification. The model was trained on 9000 dermoscopic images and achieved an accuracy of 96%. Another study [7] used a pre-trained model, GoogleNet Inception v3, fine-tuned using transfer learning. Including both clinical and dermoscopic images, the dataset comprises 135,550 images categorized into 2,055 disease types. In tasks involving the detection of keratinocyte carcinoma and melanoma, the AI model performed better than 21 board-certified dermatologists. A combination of ResNet50 and LSTM is proposed to improve the accuracy of skin lesion classifier [10]. Table 1 summarizes the above-described studies.

3 Methodology

This section covers the creation of classification models, image pre-processing, and dataset description.

3.1 Dataset Description

The ISIC 2018 Task 3: Skin Lesion Classification Challenge Dataset, which comprises 10015 images divided into seven categories of skin lesions, including melanoma (MEL), basal cell carcinoma (BCC), and actinic keratoses (AKIEC). The images have been sourced from dermatology clinics worldwide via the International Skin Imaging Collaboration (ISIC) archive. The original dataset used in this project was a multiclass skin lesion dataset. To simplify the classification task and focus on identifying cancerous lesions, the dataset was restructured into a binary format. Specifically, the three malignant classes—melanoma (MEL), basal cell carcinoma (BCC), and actinic keratoses (AK) were grouped under the 'malignant' label, and the remaining classes were grouped as 'benign'.

3.2 Classification Model

Using a transfer learning methodology based on the ResNet50 architecture and pre-trained on the ImageNet dataset, the classification model was constructed. Initial training and fine-tuning were the two primary stages of the model development process. The ResNet50 model was used as the base, excluding its top layers. All base layers were frozen initially to prevent weight updates during the first training phase. A custom classification head, GlobalAveragePooling2D, a Dense layer and ReLU activation, and a Dropout layer for regularization were added to ResNet50. Then, to enhance feature learning, a multi-head attention block was implemented. The features were reshaped into sequences and then passed through a four-head Multi-Head Attention layer. The Output was normalized and passed through a small feedforward neural network, followed by a flattening layer. The final output layer returned a probability score for binary classification using a sigmoid activation.

Binary cross-entropy loss and the Adam optimizer were used to compile the model. Evaluation

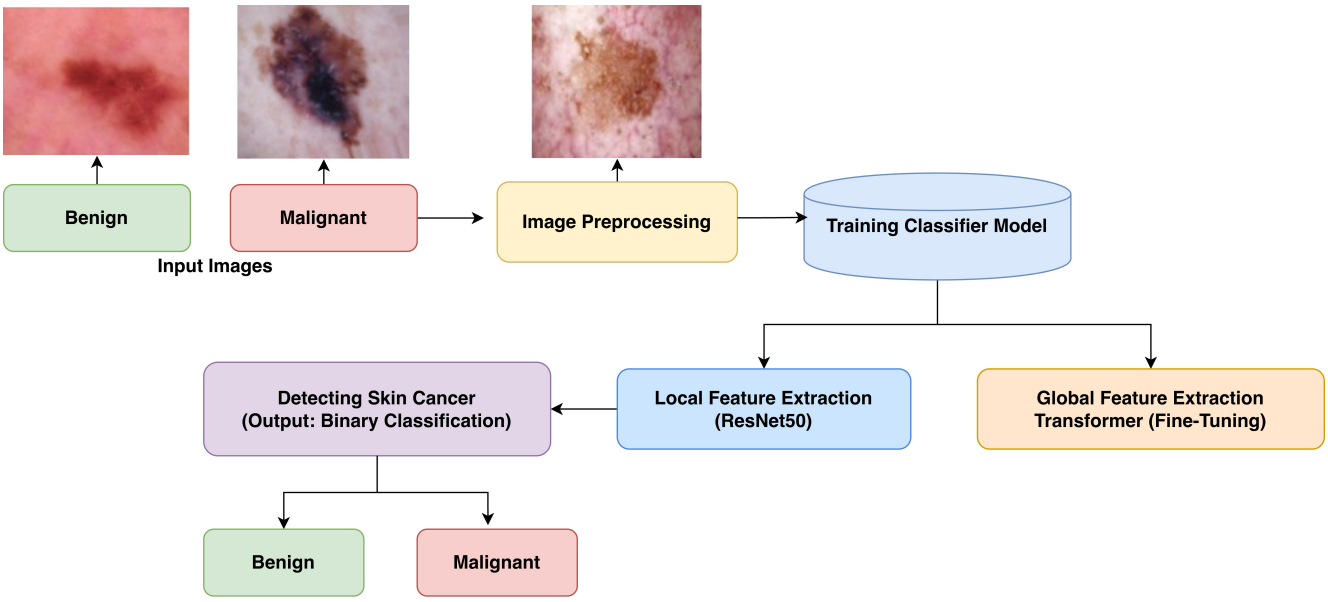

Figure 2. Proposed workflow.

metrics included accuracy. Three callbacks were employed: ModelCheckpoint to save the best model, ReduceLROnPlateau to lower the learning rate on a performance plateau, and Early stopping to halt training early if validation loss stopped improving. During the fine-tuning stage, a portion of the ResNet50 model—specifically the upper layers—was made trainable, allowing the network to adapt to the new dataset. The lower layers remained fixed to preserve features learned from ImageNet. The model was recompiled with a lower learning rate (1e-5) and trained for an additional 15 epochs to fine-tune high-level features. The proposed workflow is given in Figure 2.

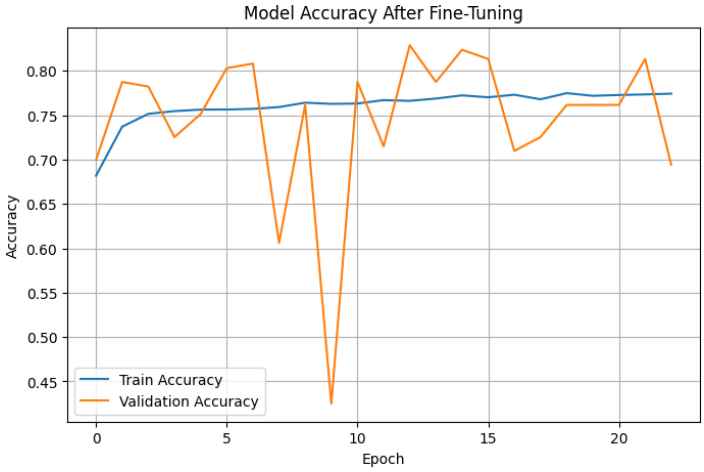

Figure 3. Training vs Validation enabled comparison of Accuracy.

4 Experiments

The model was trained over 30 epochs in the initial phase and further fine-tuned for 35 additional epochs. After both training phases were finished, the training accuracy grew steadily until it reached 76.73 percent.

Additionally, as shown in Figure 3, validation accuracy was approximately 82.9%. As can be seen in Figure 4, the model was learning successfully because training loss decreased steadily while validation loss first increased and then levelled out. The model was tested on unseen images after the training phase. It achieved a final accuracy of 80.89%, reflecting the model's optimal performance after fine-tuning. After initial fluctuations, loss values stabilized, indicating that the training approach improved generalization ability and prevented overfitting. The model's prediction is depicted in Figure 5.

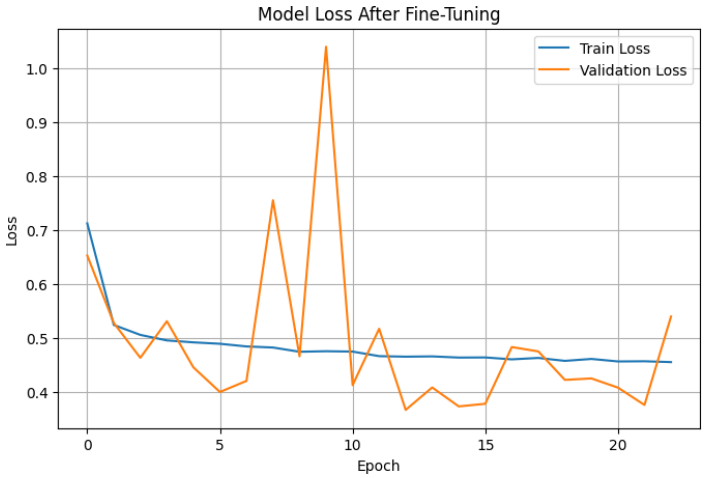

Figure 4. Training vs Validation enabled comparison of Loss.

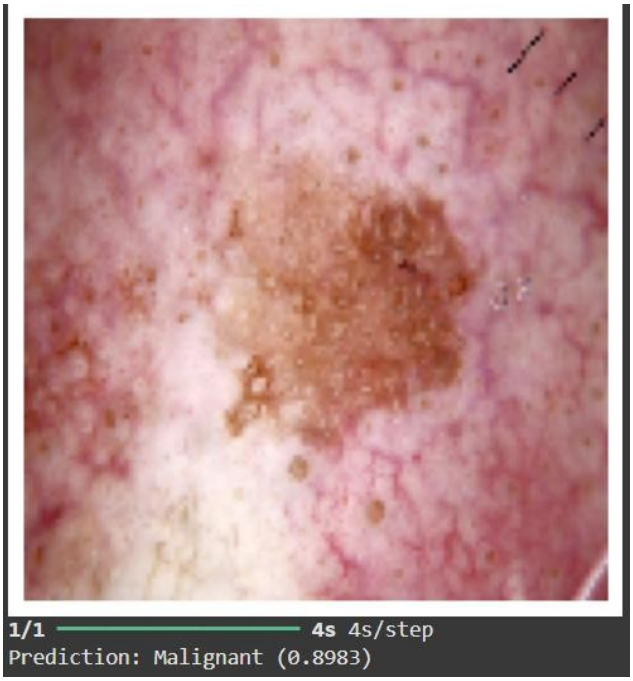

Figure 5. Model prediction.

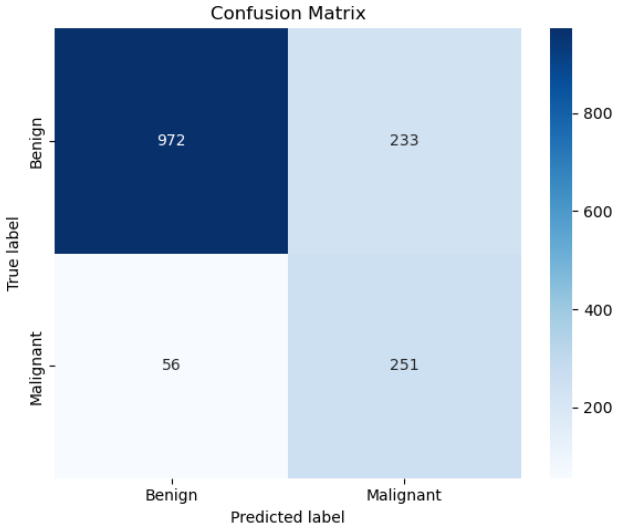

Figure 6. Confusion matrix.

To better understand the model's predictions, a confusion matrix generated on the test dataset as depicted in Figure 6, which compares the predicted labels with the actual ground truth labels for both benign and malignant cases. The model was able to recognize 251 cancerous skin lesions accurately. While 233 benign lesions were mistakenly classified as malignant, 972 benign lesions were correctly identified as non-cancerous. However, it is expected that 56 malignant lesions will be benign. The model evaluation metrics and their outcome is given in Table 2.

Table 2. Model evaluation metric.

Metric	Value (%)
Sensitivity	81.76%
Specificity	80.66%
Accuracy	80.89%
F1-Score	63.45%

5 Conclusion

This research paper presents a binary classification model that aids in the early detection of skin cancer. An extra transformer block was added to a ResNet50-based model to enhance feature learning. The model achieved a test accuracy of 80.89%. This performance demonstrates that even with a lightweight and relatively lower-complexity architecture, effective classification is possible. Looking ahead, future work will focus on improving malignant detection by exploring more advanced attention mechanisms and increasing dataset diversity. The ultimate goal is to make skin cancer detection more accessible, accurate, and fast, empowering early diagnosis and better patient outcomes.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

This work utilized publicly available, fully anonymized datasets. No human subjects were involved, and no new data were collected from participants. Therefore, ethical approval was not required.

References

[1] Ranjan Kumar, H. S., Gireesh Babu, C. N., Vijay, C. P., Raju, K., Santhosh Kumar, K. L., Prabhavathi, K., & Puttegowda, K. (2025). Deep learning-based automated classification of skin lesions using CNN and computer vision. *SN Computer Science, 6*(7), 846. [Crossref]

[2] Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[3] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., ... & Zhou, Y. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

[4] Rath, K. T., Muduli, D., Kumar, G. K., Ray, D. S., & Sharma, S. K. (2024, June). Enhancing Skin Cancer Diagnosis using Dermoscopic Images: A Feature Fusion based Deep Learning Approach. In *2024 15th International Conference on Computing Communication and Networking Technologies* (*ICCCNT*) (pp. 1-6). IEEE. [Crossref]

[5] Roy, P., Efat, A. H., Hasan, S. M., Srizon, A. Y., Hossain, M. R., Faruk, M. F., & Al Mamun, M. (2024, September). Multi-Scale Feature Fusion Framework Based on Attention Integrated Customized DenseNet201 Architecture for Multi-Class Skin Lesion Detection. In *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications* (*PEEIACON*) (pp. 496-501). IEEE. [Crossref]

[6] Xin, C., Liu, Z., Zhao, K., Miao, L., Ma, Y., Zhu, X., ... & Chen, H. (2022). An improved transformer network for skin cancer classification. *Computers in Biology and Medicine, 149*, 105939. [Crossref]

[7] Himel, G. M. S., Islam, M. M., Al-Aff, K. A., Karim, S. I., & Sikder, M. K. U. (2024). Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system. *International Journal of Biomedical Imaging, 2024*(1), 3022192. [Crossref]

[8] Sharma, G., & Chadha, R. (2022, April). Skin Cancer and Oral Cancer Detection using Deep Learning Technique. In *2022 IEEE International Conference On Distributed Computing And Electrical Circuits And Electronics* (*Icdcece*) (pp. 1-5). IEEE. [Crossref]

[9] Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020, September). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In *2020 5th international conference on advanced technologies for signal and image processing* (*ATSIP*) (pp. 1-5). IEEE. [Crossref]

[10] Padhy, S., Dash, S., Kumar, N., Singh, S. P., Kumar, G., & Moral, P. (2025). Temporal integration of ResNet features with LSTM for enhanced skin lesion classification. *Results in Engineering, 25*, 104201. [Crossref]

Aditya Joshi received the PhD. degree in Computer Science and engineering from Graphic Era Deemed to be University, Dehradun 248002, India, in 2025. His research area includes Machine Learning and Computer Vision. He has authored and published several research papers in reputable journals indexed in Scopus, SCI. (Email: adi.joshi@geu.ac.in)