IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.24

Skin Cancer Classification Using Wine Transformer

T K Pradeepkumar
Assistant Professor/CSE
AMC Engineering College
Bannerghatta Rd,Bangalore
tkpradeep.it@gmail.com
https://orcid.org/0000-0001-8396-4283

Suguna.A
Assistant Professor/CSE
Sri Sairam College of Engineering
Bengaluru, India
sugunalecit@gmail.com
https://orcid.org/0009-0003-5310-3492

B.Jayakaruna
Assistant Professor/CSE
AMC Engineering College
Bannerghatta Rd,Bangalore
b.jayakaruna@gmail.com
https://orcid.org/0009-0000-4819-0585

K B Kavitha
PGT in Mathematics
Maharishi Vidya Mandir Senior Secondary School
Tiruvannamalai, India
kbkavithaindrani@gmail.com
https://orcid.org/0009-0008-3549-2267

Nithiya.S
Assistant Professor
Dept of Computer Science
Saradha Gangadharan College
Puducherry,India
nithiya.sgc@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.12/Issue11/NVAE10103
Research Article| Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.12/Issue11/NVAE10103
Received:22,October 2025,Revised:28,October 2025, Accepted:31, October 2025, Published Online: 21, November 2025.
https://www.ijirae.com/volumes/Vol12/iss-11/24.NVAE10103.pdf
Citation:Pradeepkumar,Suguna,Jayakaruna,Kavitha,Nithiya(2025),Skin Cancer Classification Using Wine Transformer,IJIRAE: International Journal of Innovative Research in Advanced Engineering, Volume 12, Issue 11 of 2025 pages 586-590 **doi:>**https://doi.org/10.26562/ijirae.2025.v1211.24
BibTeX Key: Pradeepkumar@2025Skin
IJIRAE papers should be cited as IJIRAE (International Journal of Innovative Research in Advanced Engineering, AM Publications, India 2025, ISSN 2349-2163, https://doi.org/10.26562/ijirae.2025.v1211.24 The journal's official abbreviation is IJIRAE. Orcid: https://orcid.org/0009-0004-9398-7488

Abstract: Skin lesions, which may indicate benign or malignant conditions, require accurate classification for the early detection and treatment of skin cancer. Traditional methods often struggle with extracting relevant features and handling imbalanced datasets, leading to reduced performance. This study presents a hybrid deep learning model for improved skin cancer classification using dermoscopic images. The proposed system combines the SwinTransformer with a custom CNN-based module called the Dense Group Shuffle Non-Local Attention (DGSNLA) network, which incorporates DenseNet169, Group Shuffle Depth-wise blocks, and enhanced non-local attention mechanisms. By fusing global and local features, the model enhances feature representation and improves classification accuracy. Evaluated on the HAM10000 dataset, the proposed approach achieves an accuracy of 94.21%, an F1-score of 96.64%, a precision of 96.62%,and a recall of 96.25%, demonstrating its effectiveness in skin lesion classification.

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.24

I.INTRODUCTION

Skin lesions are visible alterations in the skin's structure that differ from the surrounding tissue and may indicate underlying health conditions. These lesions can be present from birth or may develop later due to environmental exposure, genetic factors, or self-inflicted injuries such as scratching. They can be categorized as benign or malignant, with some potentially indicating the presence of skin cancer a condition that has become increasingly prevalent, with approximately 1.2 million cases reported globally in 2020.Notably, the incidence of melanoma, a serious form of skin cancer, has risen significantly across various regions. One of the primary contributors to skin cancer is prolonged exposure to ultraviolet (UV) radiation, which can lead to genetic mutations in skin cells such as keratinocytes and melanocytes. These mutations may disrupt normal cellular function, resulting in cancerous growths. Laboratory studies have shown that UV exposure can cause early cellular damage, marked by increased extracellular Lactate Dehydrogenase (LDH) release and heightened levels of inflammatory markers like IL-1β. Early detection of skin cancer can significantly improve patient outcomes, with survival rates increasing by up to95%when identified in the early stages. To identify skin cancer, dermatologists typically rely on visual examination supported by dermoscopy a non Invasive imaging technique that offers magnified views of the skin while minimizing surface glare. Dermoscopy has proven effective in the early diagnosis of skin cancer and has helped reduce mortality rates. However, visual inspection alone can lack precision and consistency, leading to longer diagnostic processes and potential errors. To assist in more standardized assessment, tools such as the ABCD rule (Asymmetry, Border, Color, Diameter) have been developed to help clinicians recognize melanoma characteristics more accurately. Despite these advances, distinguishing between different types of skin lesions remains a challenge, even with dermoscopic tools.

Manual classification is often subjective and prone to inconsistencies due to minimal differences between lesion types and intra-class variability. As a result, integrating Computer-Aided Diagnosis (CAD) systems has become increasingly important to enhance the accuracy and efficiency of skin cancer detection. CAD systems for skin cancer typically follow a multi-step process involving image preprocessing, lesion segmentation, feature extraction, and classification. Early CAD solutions, based on traditional machine learning methods, faced several limitations. These included difficulty handling complex image backgrounds and insufficient contrast, as well as limited accuracy due to reliance on manually crafted features. Such methods were also time-intensive and less adaptable to real-time clinical environments. The emergence of deep learning (DL) has transformed the landscape of skin cancer diagnosis by enabling automated detection of intricate patterns in medical images. DL models bypass manual feature extraction, offering faster and often more accurate recognition of skin abnormalities. Nevertheless, traditional DL approaches are not without short comings. They are prone to over fitting, especially when trained on imbalanced datasets, and may struggle to capture both spatial details and long-range feature relationships effectively. To address these issues, this research introduces a dual-track deep learning architecture for more reliable skin cancer classification. The first track utilizes the Swin Transformer to effectively capture global contextual information, while the second track incorporates a custom module Dense Group Shuffle Non-Local Attention (DGSNLA) which enhances local feature extraction.

II. RELATED WORKS

Skin cancer classification using automated systems has become a prominent area of research due to its critical role in early detection and improved patient outcomes. With the limitations of traditional handcrafted feature extraction and conventional machine learning techniques such as poor generalization, sensitivity to noise, and limited scalability modern studies increasingly rely on deep learning-based approaches. Several advanced architectures have emerged, each attempting to enhance classification accuracy, feature extraction, and computational efficiency.

In [1], Karthik et al. introduced a hybrid deep learning model that combines the Swin Transformer and the Dense Group Shuffle Non-Local Attention (DGSNLA) network for enhanced skin lesion classification. The Swin Transformer effectively models global dependencies by employing a hierarchical structure of shifted windows and localized attention, enabling the model to capture wide contextual relations without compromising spatial locality. In parallel, the DGSNLA block integrates DenseNet-169 for dense feature propagation with two novel attention modules: Group Shuffle Depthwise (GSDW) and Enhanced Non- Local Attention (ENLA), which together improve the model's sensitivity to fine-grained local and non-local patterns. The combination of global and local streams allows for a rich fusion of multi-scale features, leading to high classification performance on the HAM10000 dataset, with an accuracy of 94.21% and F1-score of 96.64%.

Calderón et al. in [2] proposed BILSK, a bilinear convolutional neural network approach. The model fuses the outputs from two independent convolutional streams via bilinear pooling, allowing it to model complex feature interactions. This approach boosts the model's ability to capture subtle differences in lesion textures and colors, which are essential for accurate classification. However, while bilinear pooling is powerful, it introduces a large number of parameters and high computational costs, making real-time or mobile implementation less feasible. In contrast, Mehmood et al. [3] developed SBXception, a lighter and broader version of the Xception architecture. Their goal was to reduce model depth while increasing width, which they achieved by decreasing the number of layers but increasing the number of filters. SBX ception retains strong feature representation ability through depth wise separable convolutions and performs well with a lower computational footprint. It is particularly suitable for low-resource environments, but its shallower architecture may struggle with deeper semantic understanding compared to transformer-based methods. A different hybrid approach was presented by Akilanda sowmya et al. in [4], where deep features extracted from pre-trained CNNs were fed into a collection of ensemble classifiers (such as SVM, Random Forest, and KNN).

This method leverages transfer learning to extract high-level semantic features and improves generalization by aggregating decisions from multiple traditional classifiers. This strategy is particularly useful in the presence of class imbalance and small datasets. However, the added complexity of managing and training multiple classifiers increases the overhead and reduces scalability.Zhang et al. [5] proposed an innovative architecture that blends Gated Recurrent Unit (GRU) networks with the Improved Orca Predation Algorithm (IOPA). In this system, GRU s are applied to image patch sequences to model spatial dependencies as if they were temporal sequences, an unconventional yet creative use of RNNs in image classification. Meanwhile, IOPA dynamically adjusts hyperparameters and optimizes feature learning, which helps in achieving faster convergence and higher accuracy. Although promising, the combination of GRUs and a nature- inspired optimizer leads to a more complex training pipeline and may increase inference latency, limiting its real-time applicability.

While each of these models contributes unique innovations, they often focus on either local or global features, or make trade-offs between computational cost and accuracy. The proposed work builds upon and extends these prior efforts by unifying transformer-based global context modeling with attention-enhanced local feature extraction in a dual-path hybrid network. This design not only enhances these mantic understanding of dermoscopic images but also ensures fine-grained texture recognition, addressing limitations in earlier CNN-only or ensemble-based methods. Furthermore, the use of a single, end-to-end trainable architecture eliminates the complexities of ensemble models while retaining high precision, recall, and F1-scores making it highly suitable for clinical deployment and real-time diagnosis support.

The proposed system for skin cancer classification integrates advanced deep learning techniques to leverage both global and local feature extraction capabilities in a unified architecture. Drawing inspiration from prior research, including bilinear CNN models, lightweight Xception variants, ensemble-based classifiers, and GRU-optimized pipelines, this model synthesizes the most effective strategies into a streamlined, dual-pathway framework. The methodology begins with the preprocessing of dermoscopic images from the HAM10000 dataset, where each image is resized, normalized, and augmented using techniques such as horizontal and vertical flipping. This step not only standardizes the input dimensions but also addresses class imbalance, enhancing the generalizability of the model.

Once the data is prepared, it is fed into a two-track deep learning architecture. The first track utilizes a Swin Transformer, which processes images through shifted window-based self-attention mechanisms. This design enables the model to capture global contextual features and long-range dependencies efficiently by hierarchically aggregating spatial information. The second track comprises a Dense Group Shuffle Non-Local Attention (DGSNLA) network, which builds on a DenseNet-169 backbone and incorporates Group Shuffle Depthwise (GSDW) convolutions and Enhanced Non-Local Attention (ENLA) blocks. This module excels at extracting fine-grained local textures and capturing non-local interactions across the image. The outputs from both tracks are then fused to form a comprehensive feature representation that balances semantic richness with local detail. A final classification layer predicts the probability of each skin lesion class, outputting both the predicted label and a confidence score. By combining transformer-based global analysis with attention-enhanced local refinement, this hybrid model addresses the limitations of previous systems and achieves superior performance in terms of accuracy, precision, recall, and F1-score, making it suitable for real-world diagnostic applications.

The proposed Skin Cancer Classification System is built on a robust modular framework designed to process dermoscopic images efficiently and accurately. It begins with data ingestion, utilizing the HAM10000 dataset, which includes 10,015 dermoscopic images categorized into seven lesion types. In the preprocessing module, the input images are resized to a fixed resolution, normalized to standardize pixel intensity values, and augmented using horizontal and vertical flipping techniques. These preprocessing steps enhance the dataset's diversity, address class imbalance, and improve the generalization capability of the model. Following preprocessing, the images are fed into a dual-track feature extraction module. The first track uses the Swin Transformer, a hierarchical vision transformer that applies shifted window-based multi-head self-attention. This architecture effectively captures long-range dependencies and contextual relationships within the image, making it ideal for extracting global features. In parallel, the second track employs the Dense Group Shuffle Non-Local Attention (DGSNLA) network. This track leverages DenseNet-169 to perform deep feature extraction, while Group Shuffle Depth wise (GSDW) convolutions capture spatial details, and Enhanced Non-Local Attention (ENLA) blocks model non- local contextual dependencies. Together,these tracks ensure the extraction of both coarse semantic and fine-grained spatial features. The features from both tracks are merged in a feature fusion module, which concatenates and refines the global and local feature representations into a unified, enriched feature vector. This fused representation is passed to a classification module comprising a fully connected layer with softmax activation to predict the lesion class being nonmalignant accompanied by confidence scores. To enhance classification reliability, especially in the presence of imbalanced class distributions, the model is trained using focal loss, which emphasizes hard- to-classify samples.

Hyperparameter tuning is conducted using Ray Tune with grid search, allowing optimization of key parameters such as learning rate, dropout rate, batch size, weight decay, and number of epochs. This tuning ensures better convergence and generalization across diverse lesion types. Training is performed on GPU (NVIDIA T4), with early stopping to avoid overfitting.

IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 12, Issue 11, November 2025 https://www.ijirae.com/archives
https://doi.org/10.26562/ijirae.2025.v1211.24

The system achieves high performance on the HAM10000 test set, with an accuracy of 94.21%, precision of 96.62%, recall of 96.25%, and F1-score of 96.64%.Afterclassification,thesystemoutputsthepredicted lesion type and confidence level, which can assist dermatologists in making informed clinical decisions, particularly in early detection scenarios where diagnostic accuracy is critical.

1. Data Collection

The process begins with acquiring the HAM10000 dataset, which contains over 10,000dermoscopicimagesfrom seven distinct classes of skin lesions. Images are resized to uniform dimensions, normalized to standardize pixel intensities, and augmented using horizontal and vertical flipping. These preprocessing steps improve the robustness of the model by reducing over fitting, handling class imbalance, and enhancing the diversity of training data.

2. Feature Extraction

This step involves extracting visual features using two parallel path ways. The first track uses the Swin Transformer, which applies a window-based self-attention mechanism with shifted windows to model long-range dependencies and global context. The second track utilizes the DGSNLA network, combining DenseNet-169 for deep hierarchical features, GSDW for localized spatial representation, and ENLA blocks for non-local contextual learning. This dual approach ensures comprehensive feature learning.

3. Feature Fusion

The global features from the Swin Transformer and the local features from the DGSNLA network are concatenated in this step. The fusion creates a unified feature vector that captures both broad semantic and fine-grained visual information, improving the discriminative power of the representation for classification.

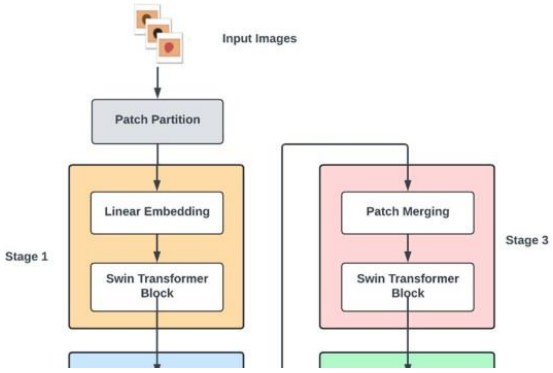
4. Lesion Classification

The fused feature vector is input into a fully connected layer followed by a softmax activation function. This layer predicts the class of the skin lesion (e.g., benign or malignant) and provides a confidence score. To better handle imbalanced data, focal loss is used during training, which puts more focus on hard-to-classify and minority class samples.

5. Model Training

Training is conducted on a GPU-enabled system using NVIDIAT4 for efficient computation. Hyperparameters such as learning rate, batch size, dropout rate, and epochs are fine tuned using Ray Tune with grid search. Early stopping is also implemented to avoid over fitting and ensure model generalization.

6. Evaluation The model's performance is validated using standard metrics such as accuracy, precision, recall, and F1-score. Achieving an accuracy of 94.21% and an F1-score of 96.64%, the system demonstrates strong effectiveness. The final output includes the lesion type and its associated confidence score, which dermatologists can use for reliable clinical decision-making in early skin cancer diagnosis. The faculty of our respective institutions for their continuous encouragement and guidance.



EXPERIMENT RESULT

The proposed hybrid deep learning model, integrating the Swin Transformer and DGSNLA network, demonstrated highly effective performance on the HAM10000 dataset. After comprehensive training and hyperparameter tuning, the model achieved a classification accuracy of 94.21%, indicating a strong ability to correctly identify skin lesion categories. Additionally, it recorded a precision of 96.62%, reflecting the model's capability to minimize false positives, and a recall of 96.25%, highlighting its efficiency in correctly detecting positive cases. The F1-score reached 96.64%, showing a balanced trade-off between precision and recall. These metrics confirm the model's robustness and reliability in classifying both benign and malignant skin lesions. The high confidence scores provided by the model further support its use as a diagnostic aid for dermatologists, especially in early and accurate detection of skin cancer.

CONCLUSION

The proposed hybrid deep learning model represents a significant advancement in the field of automated skin cancer diagnosis. By seamlessly integrating the global attention capabilities of the Swin Transformer with the fine-grained local and non-local feature extraction strengths of the DGSNLA network, the system effectively overcomes traditional limitations in dermoscopic image classification.

Enhanced with robust preprocessing methods, focal loss to manage class imbalance, and meticulous hyperparameter tuning, the model achieves a high classification accuracy of 94.21%, along with impressive precision and recall metrics. These results validate the model's reliability and effectiveness in assisting dermatologists with early and accurate skin cancer detection. Beyond its technical merits, this hybrid framework lays the ground work for broader adoption of AI-driven tools in clinical environments, promising to reduce diagnostic errors, accelerate decision-making, and ultimately improve patient outcomes.

ACKNOWLEDGEMENT

We would like to extend our sincere thanks to all those who supported us throughout the development of this research. Special appreciation goes to our mentor.

REFERENCES

1. R.Karthik,R.Menaka,S.Atre,J.Cho,andS.V.E. Easwaramoorthy, "A Hybrid Deep Learning Approach for Skin Cancer Classification Using Swin Transformer and Dense Group Shuffle Non-Local Attention Network,"*IEEE Access*, vol. 12, pp. 158040–158051, Nov. 2024.
2. C.Calderón,A.DelaOssa,M. U.Pérez,andM.A. Salinas, "BILSK: A bilinear convolutional neural network approach for skin lesion classification," *Comput. Methods Programs Biomed. Update*, vol. 1, Jan. 2021, Art. no. 100036.
3. A.Mehmood,M.Yasmin,A.Khan,andH.A.Jalab, "SBX ception: Ashal lower and broader xception architecture for efficient classification of skin lesions,"Cancers,vol.15, no. 14, p. 3604, Jul. 2023, doi: 10.3390/cancers15143604.
4. G. Akilandasowmya, N. Anusha, and V. Suresh, "Skin Cancer Diagnosis: Leveraging Deep Hidden Features and Ensemble Classifiers for Early Detection and Classification," *Biomed. Signal Process. Control*, vol. 88, Feb. 2024, Art. no. 105306.
5. L. Zhang, Y.Zhao,X. Chen,andJ.Wang, "ADeep Learning Outline Aimed at Prompt Skin Cancer Detection Utilizing Gated Recurrent Unit Networks and Improved Orca Predation Algorithm,"Biomed. Signal Process. Control, vol. 90, Apr. 2024, Art. no. 105858.