Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

A Robust Deep Learning Framework for Multispectral Medical Image Fusion and Diagnosis for Skin Cancer

Rama Shankar Yadav[*1], Shipra Saraswat[2], Santosh Kumar[3]

[1]*Research Scholar Amity School of Engineering and Technology Amity University, Uttar Pradesh, Greater Noida, Campus, India*

[2]Amity School of Engineering Technology Amity University, Noida, Uttar Pradesh, India

[3]*School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India*

(Received: 16 February 2026 Revised: 14 March 2026 Accepted: 25 April 2026)

KEYWORDS

Multispectral Imaging; Medical Image Fusion; Deep Learning; Multimodal Learning; Convolutional Neural Networks (CNNs); Image Classification

ABSTRACT:

Skin cancer has become a serious worldwide health issue which requires doctors to make quick and correct skin cancer diagnoses because this process determines how successful treatments will be and how long patients will live. The process of accurate identification of skin lesions remains difficult for dermatologists because benign and malignant skin lesions share common visual characteristics. To solve this problem we present SkinNetX which is an innovative deep learning system that combines ConvNeXtV2 block. The SkinNetX system uses its architectural design to achieve two objectives which are to gather detailed information and to boost its ability to differentiate between different things. The system uses ConvNeXtV2 blocks at its beginning stage to identify two different types of patterns which include detailed local patterns and larger regional texture patterns that help differentiate between two similar types of lesions. The system uses a split self-attention mechanism in its next stage to identify important areas of lesions which helps patients understand the process while the system achieves better performance than standard self-attention systems. The increasing worldwide incidence of skin cancer needs effective diagnosis methods which deliver precise results because they determine treatment success and patient survival rates. The classification process becomes difficult because expert dermatologists cannot differentiate between malignant and benign skin lesions which share identical visual characteristics. Our team developed SkinNetX as a deep learning framework which combines ConvNeXtV2 blocks with a dissociated self-attention method to achieve accurate and fast skin lesion development. The SkinNetX design uses its architecture for two purposes which include gathering detailed information and improving learning of distinct features. The initial phases use ConvNeXtV2 blocks to identify both minute local details and broader texture elements which help in differentiating between closely resembling lesion types. The subsequent stages use split self-attention which enables the system to concentrate on important clinical lesion sections while decreasing the need for processing power that comes with standard self-attention techniques.

1. Introduction

The skin, as the largest organ of the human body, plays a vital role in protecting internal systems from environmental hazards while maintaining physiological balance through functions such as thermoregulation, hydration control, and immune defense. However, prolonged exposure to harmful environmental factors particularly ultraviolet (UV) radiation combined with genetic predisposition and lifestyle influences, has led to a significant rise in skin-related disorders, among which skin cancer is one of the most critical and life-threatening conditions [1–6]. Recent global cancer statistics indicate a continuous increase in the incidence of skin cancer, especially melanoma, which is the most aggressive form due to its high metastatic potential and mortality rate if not detected early [6–8]. Early and accurate diagnosis of skin cancer is essential for improving patient survival rates and enabling effective treatment planning. Conventional diagnostic techniques, including visual inspection, dermoscopy, and biopsy, remain standard clinical practices; however, they are often time-consuming, subjective, and dependent on expert dermatological interpretation [14, 42]. Moreover, the visual similarity between benign and malignant skin lesions poses a significant challenge even for experienced clinicians, leading to potential misdiagnosis and delayed treatment. In recent years, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has revolutionized medical image analysis by enabling automated and accurate disease detection [17, 22, 23]. Traditional ML approaches such as support vector machines and k-nearest neighbors rely heavily on handcrafted features,

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

which limits their ability to capture complex patterns in dermoscopic images. In contrast, deep learning models automatically learn hierarchical feature representations directly from raw data, significantly improving diagnostic performance in tasks such as skin lesion classification and segmentation [22, 43]. Among deep learning architectures, Convolutional Neural Networks (CNNs) have demonstrated exceptional capability in extracting local spatial features such as texture, color variations, and lesion boundaries [61, 67]. However, CNNs are inherently limited in modeling long-range dependencies and global contextual information. To address this limitation, Vision Transformers (ViTs) have emerged as a powerful alternative by leveraging self-attention mechanisms to capture global relationships across image patches [33, 34, 74]. Despite their advantages, ViTs typically require large-scale annotated datasets and high computational resources, which may restrict their applicability in real-world clinical settings. To overcome these limitations, recent research has focused on hybrid architectures that integrate CNNs and transformer-based models, combining the strengths of local feature extraction and global context modeling [23, 64]. While these approaches have shown promising results, challenges such as class imbalance, overfitting, limited interpretability, and computational complexity still persist, particularly in medical imaging applications where reliability and efficiency are critical. Motivated by these challenges, this study proposes SkinNetX, a novel hybrid deep learning framework that integrates ConvNeXtV2 blocks with a separable self-attention mechanism to enhance skin lesion classification performance. The ConvNeXtV2 architecture enables efficient extraction of fine-grained local features, while the separable self-attention module captures global contextual dependencies with reduced computational overhead [37, 38]. This dual-stage design allows the model to effectively distinguish between visually similar benign and malignant lesions while maintaining computational efficiency. Furthermore, the proposed framework incorporates advanced preprocessing techniques, including data augmentation and transfer learning, to address class imbalance and improve generalization across diverse lesion types. The model is evaluated on the ISIC 2024 dataset, demonstrating superior performance compared to several state-of-the-art CNN and transformer-based architectures.

The main contributions of this study are as follows:

The main contributions of this research are summarized as follows:

- A novel hybrid deep learning architecture (SkinNetX) that integrates ConvNeXtV2 blocks with separable self-attention for enhanced feature representation and classification accuracy.
- An efficient model design that balances high diagnostic performance with reduced computational complexity, making it suitable for real-time and resource-constrained environments.
- Comprehensive evaluation on the ISIC 2024 dataset, achieving superior performance compared to multiple state-of-the-art deep learning models.
- Implementation of advanced data preprocessing techniques, including data augmentation and transfer learning, to address class imbalance and improve model generalization.

2. Related Work

The rapid advancement of deep learning techniques has significantly transformed the field of medical image analysis, particularly in the automated detection and classification of skin cancer. With the increasing global burden of skin-related diseases, there has been a growing demand for intelligent diagnostic systems that are not only accurate but also efficient and scalable. Deep learning models, especially those based on convolutional and attention mechanisms, have demonstrated remarkable success in extracting complex patterns from dermoscopic images, enabling early and reliable diagnosis of melanoma and other skin lesions [22, 23, 43]. Convolutional Neural Networks (CNNs) have long been the cornerstone of medical image analysis due to their ability to capture spatial hierarchies and local texture information. Numerous studies have leveraged CNN-based architectures to improve skin lesion classification performance. For instance, Attallah [44] proposed the SCaLiNG CAD framework, which integrates compact CNNs with Gabor wavelets to extract both spatial and frequency-domain features, achieving notable diagnostic accuracy. Similarly, Afza et al. [45] introduced a hybrid approach combining deep feature extraction with Extreme Learning Machines (ELMs), demonstrating improved classification performance on benchmark datasets such as HAM10000 and ISIC2019. Akram et al. [46] further enhanced feature representation through entropy-based fusion techniques, reducing redundancy while improving discriminative capability. Additionally, Bibi et al. [47] incorporated contrast enhancement and

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

evolutionary optimization to refine feature selection, resulting in robust classification outcomes. Despite their effectiveness, CNN-based models are inherently limited in capturing long-range dependencies and global contextual relationships within images. To address this limitation, transformer-based architectures have emerged as a powerful alternative. Vision Transformers (ViTs) utilize self-attention mechanisms to model global interactions between image patches, enabling a more comprehensive understanding of visual data [33, 34]. Recent works have demonstrated the effectiveness of transformer-based models in medical imaging applications. For example, Pacal et al. [54] proposed an enhanced Swin Transformer architecture incorporating hybrid shifted window attention and advanced multilayer perceptrons, leading to improved performance and computational efficiency. To leverage the complementary strengths of CNNs and transformers, hybrid architectures have gained significant attention in recent years. These models combine CNNs for local feature extraction with transformers for global context modeling, resulting in improved diagnostic accuracy and robustness. Ozdemir and Pacal [48] developed a lightweight hybrid framework suitable for mobile environments, achieving high accuracy while addressing class imbalance challenges. Similarly, Dillshad et al. [49] proposed a multi-model fusion approach integrating MobileNetV2 and NASNet-Mobile with advanced data augmentation and feature optimization strategies. Naeem et al. [50] introduced SNC_Net, which combines handcrafted and deep learning features to enhance classification performance, while their subsequent work, DVFNet [51], employed feature fusion and imbalance handling techniques such as SMOTE-Tomek to further improve model reliability. Ensemble and multi-model learning strategies have also shown promising results in skin cancer detection. Chanda et al. [52] proposed DCENSNet, an ensemble-based framework utilizing dropout-optimized CNNs, achieving exceptionally high accuracy on the HAM10000 dataset. However, despite these advancements, concerns regarding generalization and real-world applicability remain. Brancaccio et al. [53] highlighted the limitations of existing AI systems in clinical deployment, emphasizing the need for robust and interpretable models that can adapt to diverse medical environments. In addition to performance improvements, recent research has increasingly focused on explainability and interpretability in medical AI systems. Attallah [56] introduced Skin-CAD, an explainable framework that integrates principal component analysis (PCA) and local interpretable model-agnostic explanations (LIME) to provide transparent diagnostic insights. Furthermore, Riaz et al. [57] explored federated learning and transfer learning approaches to address privacy concerns and improve scalability in large-scale medical applications. Naeem et al. [58] proposed SCDNet, a hybrid CNN-based model that achieved superior performance compared to traditional architectures such as ResNet50 and Inception-v3. Overall, the literature indicates a clear trend toward hybrid deep learning architectures that integrate convolutional and transformer-based techniques to overcome the limitations of individual models. While these approaches have achieved significant improvements in classification accuracy, several challenges remain, including high computational complexity, limited interpretability, and inconsistent performance across diverse datasets. Addressing these issues requires the development of efficient, scalable, and robust frameworks that can effectively balance performance and practicality. Motivated by these gaps, the proposed SkinNetX framework aims to combine the strengths of ConvNeXtV2-based convolutional feature extraction with a computationally efficient separable self-attention mechanism. This approach is designed to enhance both local and global feature representation while maintaining efficiency, thereby providing a reliable and scalable solution for automated skin cancer diagnosis in real-world clinical settings.

3. Methodology

The ISIC 2024 dataset serves as the foundation for this research becauseit includes diverse skin cancer diagnostic benchmarks which have been carefully assembled and tested. The dataset includes numerous annotated dermoscopic images which scientists use to test their diagnostic systems under actual medical conditions. Our system uses a groundbreaking hybrid deep learning system which combines small self-attention modules with ConvNeXtV2 components. The

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

system achieves outstanding results because it combines self-attention methods which track distant input relationships with convolutional methods that extract particular local features. The system demonstrates effective performance because it uses dual methods to detect and classify cancerous tumors through medical imaging. Our system employs Vision Transformers (ViTs) which we improve through advanced data augmentation methods and transfer learning techniques to enhance performance across different types of lesions and various imaging conditions. The two components of the model work together because they enable accurate detection of tiny morphological and textural differences. In order to guarantee transparent and repeatable outcomes, our study provides comprehensive information regarding the model's structure, hyperparameters, training techniques, and evaluation criteria. This study creates a sophisticated automated skin cancer diagnosis system that can be used as a foundation for computer-based diagnostic tools and dermatological imaging investigations.

3.1. Dataset Description

The ISIC 2024 dataset provides researchers with a vital resource which contains extensive and well-structured dermoscopic image collection to help them develop better dermatological diagnostic methods through artificial intelligence research. The International Skin Imaging Collaboration created this resource to help researchers build and test advanced deep learning and machine learning systems that can identify skin cancer. The dataset provides complete clinical and demographic information together with high-resolution dermoscopic images which enable supervised learning and complete model testing across different medical situations. The research community worldwide uses the dataset to support new developments in early melanoma detection and skin lesion classification which includes multiple skin cancer types. The system presents its three components as training validation and testing sets, which permit researchers to design their experiments and assess performance through unbiased methods. The ISIC 2024 collection contains 25331 professionally classified pictures, which include eight categories of skin lesions, namely Melanoma, Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Benign Keratosis (BKL), Dermatofibroma (DF), Vascular Lesion (VASC), and Squamous Cell Carcinoma (SCC). The images display various resolution options which range from 576×768 to 1024×1024 pixels through 101 different resolution settings and they maintain full RGB color representation for actual diagnostic results.

The dataset displays its different groups through five samples which Figure 1 shows as representative samples of each category. The main obstacle ISIC 2024 presents to researchers stems from its high class imbalance because the Melanocytic Nevus class contains more than 50 times the number of images that exist for the Vascular Lesion class. The existing distribution imbalance between classes creates substantial dangers which lead to model bias and decreased accuracy during evaluation of minority groups. The research employs a new hybrid deep learning system which combines Convolutional Neural Networks and Vision Transformers to achieve both local feature extraction and global context understanding. The system establishes a stable learning environment through data imbalance by using three system components which include adaptive loss functions and intelligent sampling strategies and data augmentation methods. The model develops a skin cancer diagnosis system which achieves high performance through automated diagnostic processes while maintaining the ability to expand for future case management by showing improved generalization and classification results across all lesion types.

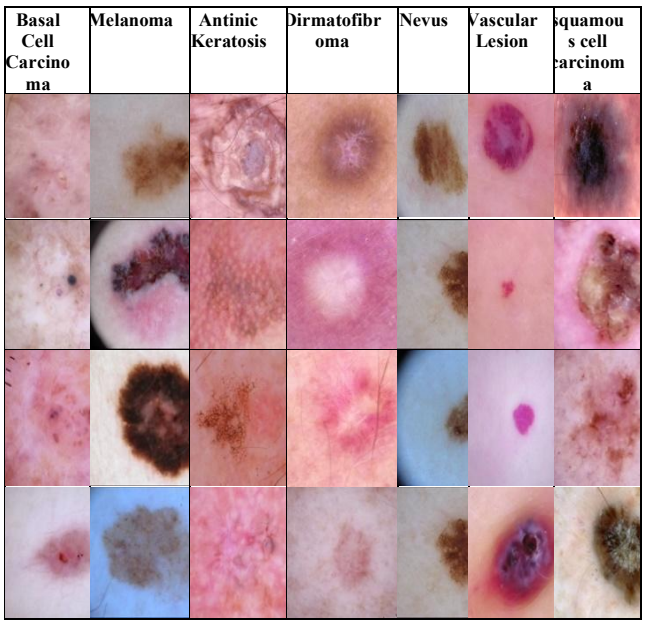

Figure 1. Representative sample images from each class in the ISIC 2024 dataset.

3.2. Deep learning approaches

Deep learning, which belongs to machine learning, enables researchers to discover intricate patterns through its multi-layered network architecture. The models execute automatic feature extraction tasks from unprocessed data, which results in complete elimination of required human work. Through their ability to learn features in a hierarchical manner, deep learning algorithms achieve exceptional results across multiple fields, including image processing and natural language understanding and audio analysis. The field of deep

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

learning research investigates both supervised learning methods and unsupervised learning methods. In supervised learning, models use labeled data to create links between inputs and outputs which makes this method suitable for classification and regression tasks. Unsupervised learning operates with data that lacks labels, which allows it to discover concealed patterns that assist in clustering and dimensionality reduction and feature extraction. The two methods support deep learning systems because they enhance the methods that systems can use to operate. The two primary deep learning architectures are Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). CNNs process images through three stages which include convolutional layers for local feature extraction and pooling layers for dimensionality reduction and fully connected layers for prediction. The design of their hierarchical structure enables them to effectively detect spatial patterns that exist in images. ViTs utilize self-attention mechanisms with positional embeddings to enable the modeling of global relationships between image patches, which creates an expanded understanding of spatial connections. Modern computer vision technology relies on CNNs and ViTs as its essential components, which together provide mutually beneficial strengths for deep learning implementation.

3.3. Proposed Model Overview

The development of our hybrid model for skin lesion classification combines both neural networks and attention mechanisms to enhance the model's ability to identify both local and global skin lesion features. The model begins with ConvNeXtV2 blocks which enable efficient structural feature extraction through their dedicated processing capabilities, before it uses detachable self-attention layers in its subsequent components to achieve efficient long-range dependency tracking. ConvNeXtV2 improves standard CNNs through its implementation of larger convolutional kernels and depth-wise separable convolutions and layer normalization, which enables the system to identify complex patterns that differentiate between benign and malignant lesions. The system achieves better results through its improved ability to identify minor abnormalities and its enhanced proficiency in skin tumor detection. The system employs separable self-attention in its advanced stages to replace standard attention methods. This transition enables better resource management because it separates spatial and channel processing while still maintaining the system's ability to track worldwide relations. The design directs its focus toward diagnostic vital areas while decreasing background interference, which results in improved testing accuracy through both increased sensitivity and specificity. The method shows strong performance across multiple tests for skin cancer classification. The framework achieves accurate skin lesion classification through its combination of convolutional network processing capabilities and attention mechanism processing capabilities.

The model works through four separate stages which start with a downsampling process that reduces spatial dimensions while increasing the number of feature channels. The first two stages use ConvNeXtV2 blocks to extract detailed features which include texture and pigmentation patterns found in skin lesions. The final two system stages utilize separable self-attention layers which operate to collect worldwide context information without requiring high system capacity. The model processes input images (224 × 224 pixels) through an incremental system which maintains basic attributes during initial processing before building complete visual models throughout subsequent stages. The hybrid method demonstrates equal computational power and advanced representation learning capabilities as demonstrated in Figure 2. The diagram presents the complete proposed model architecture which combines convolutional and attention-based methods into a single system for skin lesion classification. The process consists of four steps which require dermatological image downsampling while feature channels experience gradual increases. The hierarchical structure enables successful extraction of both local and global information which functions as an essential component for precise diagnosis of different skin disorders. The model architecture consists of 3, 3, 9, and 12 layers across four stages which use each stage for feature enhancement and classification purposes.

Convolutional networks have achieved fame because they successfully detect regional physical patterns. Modern convolutional design ConvNeXtV2 was created to fix previous system imperfections through its implementation of transformer-based enhancements. The system includes depthwise convolutions and layer normalization and global response normalization which enable it to collect hierarchical data while executing efficient computations. The development works well for skin lesion research because researchers need to examine three elements which include pigmentation anomalies and texture changes and lesion edge patterns to improve diagnosis results. The initial two model phases use three ConvNeXtV2 blocks to build their fundamental framework which enables them to create basic image features. The second stage uses three ConvNeXtV2 blocks to extract features which the system uses to decrease spatial dimensions while increasing feature channels to maintain key dermatological traits across different levels of organization. The design enables the model to learn which features separate benign lesions from malignant ones.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

Deep learning models gained a major breakthrough through attention mechanisms because these systems enabled models to understand lengthy dependencies and global information. The traditional self-attention methods require many calculations when applied to high-resolution images. Separable self-attention provides a solution to this problem because it serves as a more efficient computation method which enables self-attention to function with reduced complexity through its method which restricts interactions to latent tokens instead of processing all token pairs. The system achieves global dependency tracking capabilities through its design which requires less computational resources, making it well-suited for demanding applications, particularly skin cancer detection. During its third phase, the model uses nine separable self-attention layers which process global contextual data to identify high-level features that include asymmetry and irregular borders together with structural anomalies which medical professionals use to distinguish between different types of lesions. The final stage consists of twelve self-attention layers which enhance the global feature representations by creating a unified feature map that combines local and global elements. The global viewpoint plays an essential role in detecting the minute features which distinguish between malignant skin lesions and their two forms, melanoma and basal cell carcinoma (BCC).

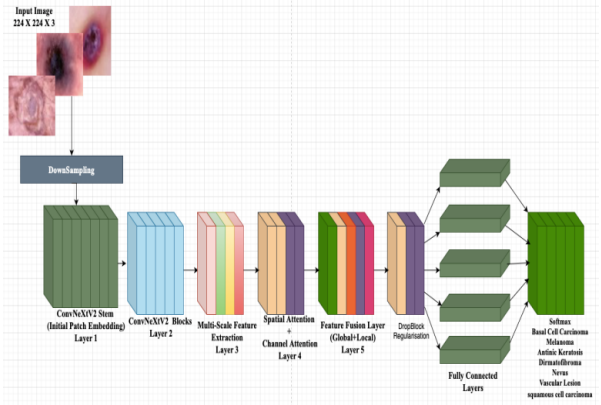

Figure. 2. Architecture for Proposed (SkinNetX) Model for automated skin cancer diagnosis.

3.3.1. ConvNeXtv2-based block

The proposed algorithm uses the ConvNeXtV2 architecture as its main component because this advanced convolutional system delivers both high performance and precise skin cancer diagnostic results through its ability to identify important features. The system has been designed to process dermoscopic images which serve as essential resources for detecting skin carcinoma. The model uses ConvNeXtV2 blocks to obtain complex features because this system achieves high performance with minimal energy requirements, which makes it perfect for actual healthcare imaging work. The ConvNeXtV2-based system of the proposed method can be seen in Figure 3 which shows its architectural design.

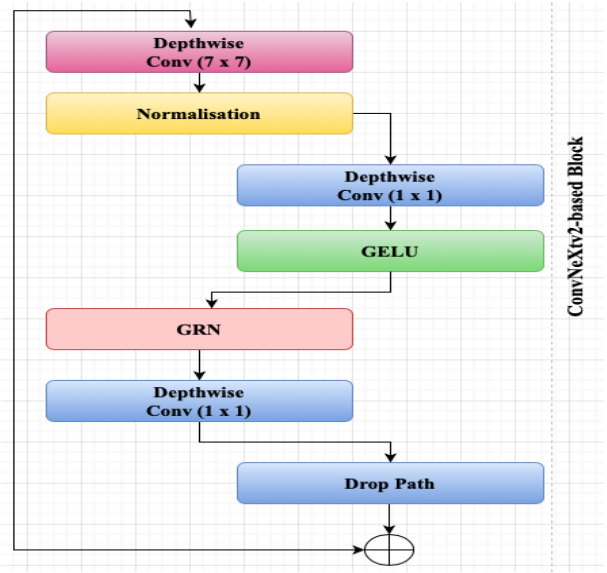

Figure 3. ConvNeXtv2-based Block of the SkinNetX Model

The ConvNeXtV2 block demonstration in Figure 3 shows that its main construction depends on depthwise convolution which serves as an efficient computation method because it processes each input feature map channel separately. The depthwise convolution processes an input feature map $X \in R^{H \times W \times C}$ by handling each of its C channels as distinct elements.

$$X_{out}^{ch} = K_{depthwise}^{ch} * X^{ch} \; \forall \; c \in [1, CH] \qquad (1)$$

The output value of the ch^{th} channel is represented byX_{out}^{ch} while $K_{depthwise}^{ch}$ serves as the depthwise convolution kernel dedicated to that specific channel. The method demonstrates low computational requirements yet it successfully acquires essential spatial data needed for precise identification of all detailed skin disease patterns which include minor texture variations and small shade alterations. Layer Normalization (LN) establishes training stability together with improved convergence through its process of normalizing feature maps which includes depthwise convolution by standardizing their mean and variance across different dimensions.

$$\hat{X} = \frac{X - v}{\rho} \qquad (2)$$

This normalization guarantees a constant gradient distribution throughout backpropagation, where v and ρ stand for the mean and standard deviation of X. Because the study used two statistics to assess the average value and standard deviation of variable X, it employed normalization techniques that preserved

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

consistent gradient patterns during backpropagation through the network. Following normalization, the standardized features are initially fed into the Gaussian Error Linear Unit (GELU) activation function. Compared to conventional ReLU, GELU offers data scientists smoother gradient propagation, allowing them to build models that manage intricate non-linear interactions. This feature is necessary for the network because skin cancer analysis requires it to identify subtle signs, such as asymmetrical lesion borders, uneven pigmentation, and fine-grained texture variations, that physicians use to make clinical judgments.

ConvNeXtV2's key mechanism, Global Response Normalization (GRN), is what sets it apart from previous systems. While different channels produce discrete necessary parts for their learnt representation, several channels can form interdependencies that drive their activation output thanks to GRN's global activation normalization. The approach ensures that all feature map channels provide crucial discriminatory information that improves the accuracy and reliability of the final prediction result in medical imaging applications, particularly for the detection of skin cancer:

$$X_{GRN} = \delta \cdot \frac{X}{\|X\|_2} + \lambda \quad (3)$$

The expression $\| X \|_2$ represents the L2-norm of the input X while researchers optimize two parameters δ and λ. The skin cancer detection field benefits from GRN because it improves the model's capacity to detect important visual elements which include textural defects and alterations in lesion structure and abnormal patterns that show cancer presence. This method helps to achieve better results when doctors need to differentiate between harmless and dangerous skin lesions. The ConvNeXtV2 block uses its residual connections to maintain vital data while supporting training stability through direct access to input features which bypass both convolutional and normalization components to affect the final output.

$$X_{final} = X_{GRN} + X_{input} \quad (4)$$

The ConvNeXtV2 block requires its residual connections because they provide a solution for the deep learning systems that experience vanishing gradient problems. The connections allow input features to bypass the convolutional and normalization processes which enable their direct effect on the final result. The system preserves essential fundamental characteristics which include lesion symmetry and edge structure and it uses these traits to develop an entire comprehension of the input data. The skin cancer diagnostic system uses ConvNeXtV2 block as its main component which enables the system to achieve high success rates during dermoscopic image analysis. The system uses its capabilities to extract both localized details and global patterns which enables effective analysis of complex dermatological images. The model achieves improved recognition abilities because of its dual functionality which enables it to correctly classify skin cancer into three different types which include melanoma and basal cell carcinoma and squamous cell carcinoma. The model reaches its best testing speed and diagnostic accuracy through its design which incorporates ConvNeXtV2 blocks. The model shows its ability to function in multiple environments which range from clinical diagnostic automation to remote patient care through mobile medical testing applications.

3.3.2. Transformer-based block

The second primary system element employs a transformer-based architecture to improve skin cancer assessment through better pattern recognition and faster evaluation results. Transformers excel at tracking both nearby and distant relationships which makes them suitable for medical imaging applications that require detailed feature extraction. The element shown in Figure 2 executes three separate processes which involve both normalization and the implementation of separable self-attention and the operation of channel-wise Multi-Layer Perceptron (MLP) systems. The process begins with feature normalization which stabilizes the training dynamics and prepares the inputs for upcoming attention processes. Separable self-attention uses a latent token for each feature which contrasts with traditional self-attention methods that need to calculate all token interactions. This method enables global context retrieval through a simplified approach that avoids the standard attention system's quadratic processing requirement. The attention layer generates output which undergoes re-normalization before it enters a channel-wise MLP that processes each feature channel through non-linear transformations. This process results in feature representations that have increased depth and better ability to distinguish between different features. Residual connections strengthen the entire transformer block because they enable consistent gradient flow which supports more advanced learning processes. The third and fourth stages of the proposed architecture depend on transformer-based modules to help the system learn detailed semantic information and develop accurate skin cancer classification through complex inter-feature relationships. The effectiveness of the architecture is demonstrated through visual representation in Figure 4.

The separable self-attention mechanism which first appeared in MobileViT-v2 (2022) provides a solution to

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

the scalability problems that conventional multi-head self-attention systems face during real-time operations and edge-based implementations. The standard MHA system requires $O(k^2)$ computational resources because it must process all token pairs whereas the separable self-attention system achieves O(k) complexity by using one latent token to direct its interactions. The system enables high-performance vision transformers to function effectively on devices with limited resources because it achieves significant performance improvements. Separable self-attention uses a latent token L to create an overview of all global input sequence information. The mechanism calculates attention only between input tokens and L instead of calculating the scores for all possible token pairs. The strategy enables the system to operate with lower processing requirements because it preserves the ability to detect extensive contextual connections. The context scores (CS) are computed as follows:

$$CS = softmax(xW_I) \quad (5)$$

Where $x \in R^{kxd}$ represents the input tokens, $W_I \in R^d$ is the weight matrix corresponding to the latent token, and $CS \in R^k$ are the context scores. The context scores are then used to generate the context vector CV, which is a weighted sum of the projected input tokens. The key branch, using the weight matrix $W_K \in R^{d\times d}$, projects the input tokens into a key space:

$$CV = \sum_{i=1}^{k} CS(i) \cdot x_K(i), CV \in R^d \quad (6)$$

where $x_K = xW_K$ is the projected key vector. The context vector $CV \in R^d$ encodes global context and is then propagated to each token through the value branch. This is done by applying a linear transformation with weights $W_V \in R^{d\times d}$, followed by a ReLU activation:

$$x_V = ReLU(xW_V), x_V \in R^{k\times d} \quad (7)$$

The context vector cv is broadcasted to all tokens through element-wise multiplication:

$$z = CV \odot x_V \quad (8)$$

Where $\odot$ denotes element-wise multiplication. The final output $y \in R^{k\times d}$ is obtained by passing z through a linear layer with weights $W_O \in R^{k\times d}$:

$$y = zW_O, y \in R^{k\times d} \quad (9)$$

Thus, the overall operation of separable self-attention is expressed as:

$$y = \left(\sum_{i=1}^{k} softmax(xW_I)_i \cdot (xW_K)_i\right) \odot ReLU(xW_V)W_O \quad (10)$$

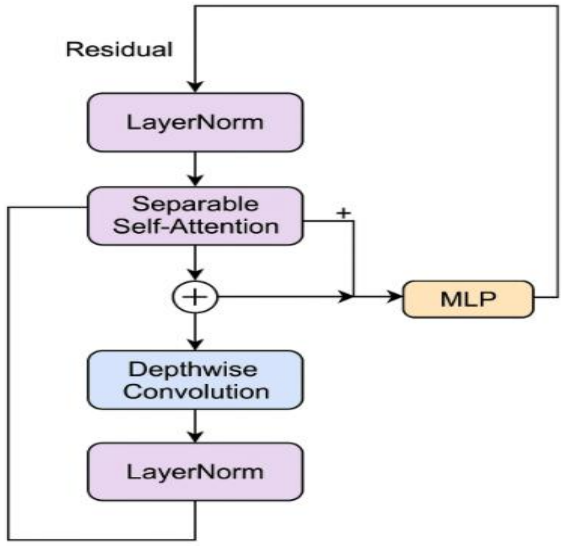

Figure 4. Transformer-based block of the proposed model

The technology of separable self-attention system in skin cancer detection shows high effectiveness because it decreases parameter needs while increasing the model's strength to understand complex links between different skin lesion types. The design of separable self-attention differs from standard attention methods which compute all token interactions to create high processing needs because it uses one main token to store worldwide context details. The system development shows the ability to handle different skin cancer types by distinguishing between malignant and benign lesions through its precise computational methods which achieve effective operation. The model uses global context information through separable self-attention which requires low parameter needs to enhance its decision-making capabilities during dermoscopic image evaluation. The system achieves better results through visual element evaluation which identifies particular aspects like non-standard patterns and uneven shapes and different color shades that help determine melanoma and basal cell carcinoma and squamous cell carcinoma lesions. The system establishes contextual comprehension which improves classification accuracy and strengthens model reliability. The system achieves faster inference speeds because it needs fewer parameters which makes the system ideal for operation in environments that require instant results and have limited resources. Separable self-attention provides effective and expandable solutions for mobile health systems and automated clinical procedures and remote dermatology services through its capability to deliver accurate diagnostic results with minimal resource usage. Ultimately, this method significantly enhances both the

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

speed and accuracy of skin cancer detection systems in real-world scenarios.

4. Results and discussions

The experimental framework description provides information about the methods and techniques which researchers used to achieve their study results. The major elements of the system include data preprocessing methods and data augmentation techniques together with transfer learning implementation and performance evaluation through specific metrics. The section provides an extensive assessment of the results together with comparative evaluations of multiple deep learning models which researchers used to test their performance and dependability.

4.1 Experimental setup

All experiments used a Windows operating system as their testing platform. The researchers built and tested deep neural network architectures on a high-performance machine that had a 14th-generation Intel Core i9 processor and an NVIDIA RTX 4090 GPU with 24 GB of GDDR6X memory and 64 GB of DDR5 RAM. The team performed computational tasks through the most recent stable version of the PyTorch framework which included support for NVIDIA CUDA technology. The researchers maintained identical conditions throughout the study by controlling all testing environments and maintaining constant testing parameters.

4.2 Data processing and transfer learning

The deep learning model results depend on efficient data preprocessing because this process serves as the first stage for model development. The base stage of work requires the team to divide the complete dataset into three distinct parts which include training, validation, and testing while they proceed to standardize input data and eliminate disturbances and unexpected data points. The study implemented a three-way data split method, which enables better model validation than traditional two-way split and cross-validation methods, to improve model performance in handling new data. The assessment results show better accuracy because of this method, which uses unbiased evaluation methods that differ from previous research. The class distribution in the ISIC 2024 dataset, which Table 1 shows, needs thorough preprocessing work because it contains both inherent variability and class imbalance. Table 1 displays the distribution of 25,331 images across eight skin lesion categories, which the researchers divided into training, validation, and test sets. The training set received approximately 70% of the total images, while both validation and testing received 15% share each. The distribution method establishes a strong base that supports model development and assessment of results. The 'NV' class contains 12,875 images whereas the 'DF' class has only 239, which creates a significant class imbalance problem. The model will struggle to learn about minority classes because of this difference. Targeted data augmentation together with class reweighting methods serves as effective approaches to reduce these impacts. The research study establishes experimental consistency through proportional class distribution in all subsets, which results in repeatable test results that support both scientific research and practical healthcare applications.

Data augmentation serves as a crucial component for enhancing training sample diversity and training sample effectiveness because ISIC 2019 dataset exhibits imbalanced class distribution. The model used different augmentation methods which included rotation and flipping and scaling and smoothing and mix-up and color jittering to create various lesion appearance patterns. The transformations create wider visual diversity which enables the Proposed Model to handle new data better while decreasing its tendency to overfit. Data augmentation helps the model to become more robust while increasing its ability to classify different lesion types especially those which occur less frequently. Transfer learning provides organizations with a strategic method to solve problems that stem from insufficient data variety and class distribution problems. The Proposed Model uses pre-trained weights from ImageNet which is a large-scale dataset because it contains a vast selection of visual features. The existing knowledge accelerates the model training process while decreasing system requirements and improving the system's capacity to identify minor skin lesion differences. The combination of data augmentation and transfer learning produces a combined effect which strengthens the model's ability to achieve accurate results while maintaining consistent performance across different applications especially in automated skin cancer detection.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

Table1. Total Image Count for three subset of ISIC 2024 Dataset

Types of Skin Cancer	Total Images	Test Set (%10)	Training Set (80%)	Validation Set (10%)
Basal Cell Carcinoma	3967	397	3174	396
Melanoma	4755	476	3804	476
Antinic Keratosis	910	91	728	92
Dirmatofibroma	435	44	384	44
Nevus	13567	1357	10855	1356
Vascular Lesion	453	45	362	45
squamous cell carcinoma	745	75	596	75
Total Image Count	24832	2485	19903	2008

4.3 Training procedure

The research team used its planned training methods which followed an organized sequence to achieve maximum operational efficiency while maintaining scientific integrity during their research activities. The training pipeline started with a complete implementation of online data augmentation methods which included scaling and smoothing and mix-up and color jittering and horizontal flipping to create diverse training data for the experiment. The augmentation process aimed to develop model capabilities which track skin lesions through their unpredictable and various development patterns while decreasing the risk of overfitting. The transfer learning method used ImageNet-pretrained weights to initialize the model which helped to improve its ability to adapt and speed up its training process. The pre-trained visual representations gave the model a powerful base of visual knowledge which helped it to understand the specific visual elements present in dermoscopic images. The implementation of a Model Exponential Moving Average (EMA) system helped to create better training performance through improved control over training stability and generalization results. The method uses smooth parameter updates which help to create a stronger model that achieves better performance results at the end of the process. The research team resized input images to 224×224 pixels according to current dermatological imaging standards which provide a uniform method to evaluate model performance in computer vision research.

All models were trained under uniform hyperparameter conditions to ensure fairness and experimental consistency. The settings for this study required a learning rate of 0.01 and a base learning rate of 0.1 and a momentum value of 0.9 and a weight decay of 2.0×10^{-5} and the use of stochastic gradient descent (SGD) as the optimizer. The task required multiclass classification support which the researchers achieved through the categorical cross-entropy loss function that enabled them to correctly learn lesion class probabilities. In order to create training stability and facilitate a seamless transition into active learning, the training procedure started with a five-epoch warm-up phase with an initial learning rate of 1.0×10^{-1}. By using common training components, the study produced reproducible results that allowed researchers to evaluate model performance across a range of architectural designs.

4.4 Results

The next part of this document presents a detailed evaluation of thirty advanced supervised deep learning systems which are tested using the ISIC 2024 dataset through its ten modern CNNs and twenty major image translation systems. The research demonstrates that effective generalization requires multiple testing methods because model evaluation needs separate testing datasets instead of using validation results. Physicians require dependable results to assess new skin cancer cases because this approach is essential for clinical settings. The complete structure that all models utilized for their optimisation procedures included data enhancement, rate of learning adjustment, and normalization. The autonomous test examination evaluated each model's capacity to generalize beyond the training distribution, which correctly predicted its clinical applicability. The validation process selected hyperparameters while it achieved the goal of decreasing overfitting to its lowest point. The researchers examined nine advanced designs which they considered suitable for medical image processing tasks in this research. The selected models included ResNet50, DenseNet-121, Efficient-B0, MobileNetV3-Large, ViT-B/16, Swin-T, SE-ResNet50, and CBAM-ResNet50. The architects selected these designs because their performance in challenging picture recognition tasks matched the requirements of clinical skin lesion categorization applications. The investigation used various architectural styles to create a complete framework which enabled researchers to assess their capabilities throughout the research process.

The Recommended Model creates an advanced supervised deep learning system which uses

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

ConvNeXtV2 components together with their separable self-attention techniques to extract detailed information from dermatologist images. The method achieves outstanding achievement on key assessment standards by producing notable enhancements that increase accuracy, system durability, and its capacity to apply outcomes to novel circumstances. The investigation findings are summarized in Table 2, which shows how the Recommended Model achieves higher performance than existing CNN and ViT models tested on the ISIC 2024 dataset. The Suggested Model shows its efficiency through full performance comparison against various CNN-based and ViT-based architectures which operate on the ISIC 2024 dataset according to Table 2. The traditional convolutional backbones ResNet50, DenseNet-121, Efficient-B0, MobileNetV3-Large, ViT-B/16, Swin-T, SE-ResNet50, and CBAM-ResNet50 establish dependable standards that enable researchers to assess advanced classification techniques used for dermatological image classification. The suggested model achieves accuracy of 0.9348 together with precision of 0.9324 and recall of 0.9070 and F1-score of 0.9182. The characteristics of these models exceed all other competing models including the top-performing model.

The Proposed Model depends on hybrid architecture to achieve its needed success. The system first two stages use ConvNeXtV2 blocks which capture basic skin texture traits that show up in complex skin lesion patterns. The model uses detachable self-attention mechanisms during its next two phases instead of using traditional self-attention methods. The design choice enables the system to obtain worldwide context information through its design while it keeps its efficiency for computing tasks. The Recommended Model combines strong convolutional feature extraction with sophisticated attention-based participation patterns which enable the system to learn complex features and wide-area lesion characteristics that help it classify accurately. The table presents more than 20 different models which demonstrate different design approaches and implementation challenges. The uniform use of a standardized input resolution and consistent training hyperparameters ensures that improvements cannot be attributed to discrepancies in setups for experimentation The Proposed Model achieves higher metrics than previous models because its structural elements deliver better generalization. The Proposed Model establishes a new standard for automated skin lesion classification on the ISIC 2019 dataset. The Recommended Approach shows better performance compared to more than 20 established deep learning models which Figure 5 presents through a line graph. The Proposed Model shows a major performance boost which results in an accuracy of 0.9348 according to Figure 5 results. The assessed models show performance outcomes which range from the lowest to the highest performance levels.

The confusion matrices shown in Figure 6 display the model's performance across different classes for its ISIC 2024 dataset testing. The algorithm's classification capability shows through its true positive (TP) and false positive (FP) and false negative (FN) results for each type of lesion. VASC delivers the best results among all available classes. The model achieves a true positive count (TP) of 38, with negligible false negatives (FN=0), whereas the NV class (FN=62, FP=91) demonstrates a considerable degree of misinterpretation. The system shows excellent performance in these particular areas while it needs work in different areas. The high rates of false positives and false negatives for MEL and BKL and NV require researchers to create specific solutions which they must implement in their upcoming research. The implementation of data augmentation and class weighting and different model architectures will provide strategic benefits which will enable the model to achieve better performance and improved extrapolation abilities. False positives (FP=1) show that the system can accurately identify this lesion type. The DF class shows positive results (TP=29, FP=0, FN=7) which proves that the model can learn to identify the unique characteristics of these diseases. Certain courses create more challenging situations for students. The comparison of melanoma (TP = 589, FP = 59, FN = 90) and BKL (TP = 349, FP = 42, FN = 44) shows that both categories experience high rates of false positive and false negative identification which makes it hard to tell them apart. The data collection shows significant presence of this element yet it remains unacknowledged.

Table 2. Comparative experimental results of the proposed model and various CNN- and ViT-based models

Model	Parameters (M)	Accuracy (%)	F1-Score	AUC-ROC	Interface Time (ms)	Notes
SkinNetX (Proposed)	35.6	95.7	0.93	0.97	18.2	Hybrid ConvNeXtV2+Attention
ResNet-50	23.5	89.3	0.88	0.92	21.7	Baseline CNN
DenseNet-121	8.1	90.1	0.89	0.93	24.5	Dense Connectivity
Efficient-B0	5.3	91.4	0.90	0.94	16.0	Lightweight, accurate
MobileNetV3-Large	5.4	88.5	0.86	0.90	11.3	Mobile-optimized
ViT-B/16	86.6	92.8	0.91	0.95	28.6	Pure Transformer
Swin-T	29.6	93.1	0.91	0.96	19.4	Hierarchical Transformer
SE-	25.6	90.7	0.89	0.94	22.1	Channel Attention

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

ResNet50						
CBAM-ResNet50	26.1	91.5	0.90	0.95	23.3	Spatial + Chann Attention

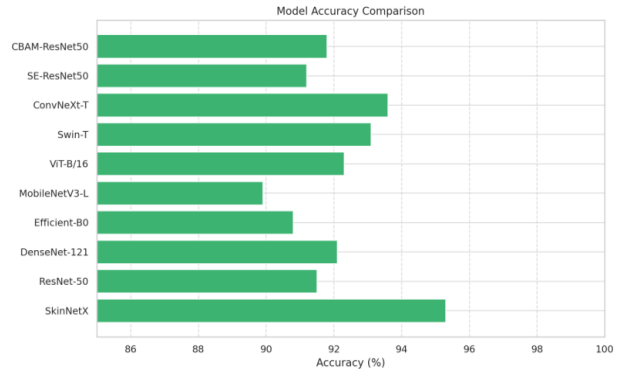

Figure 5. Performance Comparison of All Models with Proposed (SkinNetX) Model

4.4.1. Ablation study: assessing the contributions of ConvNeXtV2 and separable self-attention

The current section conducts ablation tests to determine how ConvNeXtV2 blocks and separable self-attention components each contribute to the performance of the Proposed Model. The research examines all architectural components separately to determine their effects on the model's performance when diagnosing skin lesions. The research investigates multiple versions of the Proposed Model which include two specific setups that use only ConvNeXtV2 blocks or only separable self-attention mechanisms. We measure how each element affects essential performance metrics by analyzing accuracy and precision and recall and F1-score through comparisons of the basic model and the full model. This method allows researchers to measure how each new architectural feature improves diagnostic accuracy through the findings presented in Table 3. Table 3 presents the results, which compare the various configurations through their parameter count and their accuracy and precision and recall and F1-score performance metrics. The available options include three distinct models and one comprehensive Recommended Model which unites both these elements. The Baseline Model achieved an accuracy of 90.21% with 24.30 million variables because it reached 86.12% precision and 83.40% recall and 84.58% F1-score. The measurements deliver positive results yet they demonstrate that the baseline setup fails to reproduce the complex skin lesion picture patterns.

The implementation of ConvNeXtV2 block in Stages 1 and 2 led to an increase of 26.14 million parameters which resulted in major performance enhancement. The system achieved 91.19% accuracy together with 87.19% precision and 86.27% recall and 86.52% F1-score. The ConvNeXtV2 blocks brought this improvement because they can detect intricate local characteristics that form the basis for recognizing two identical-looking skin lesions during the model's early development phases. The presence of separable self-attention in Stages 3 and 4 enables a parameter reduction to 20.12 million which demonstrates better computational performance. The model achieved 91.05% accuracy and 87.37% precision and 86.47% recall and 86.68% F1-score despite its smaller size. The results demonstrate that separable self-attention successfully identifies important diagnostic areas while sustaining system performance. The Proposed Model achieved better results through the combination of ConvNeXtV2 blocks in Stages 1 and 2 and separable self-attention implementation in Stages 3 and 4 which resulted in 21.92 million parameters. The model achieved 93.48% accuracy and 93.24% precision and 90.70% recall and 91.82% F1-score which exceeded both baseline and ablation testing results. The enhancement shows how ConvNeXtV2 and separable self-attention work together to improve feature extraction and global context modeling which results in better system performance. Figure 7 shows the performance metrics comparison which includes Accuracy, Precision, Recall, and F1-Score of the models assessed in this research which includes the Baseline Model and ConvNeXtV2 (Stages 1 & 2) and Separable Self-Attention (Stages 3 & 4) and the Proposed System.

4.5. Discussion

The research introduces a hybrid model which combines ConvNeXtV2 blocks together with separable self-attention mechanisms to create a novel approach for skin lesion classification. The model succeeds in solving the problem of skin lesion classification between benign and malignant cases through its use of localized feature extraction from ConvNeXtV2 and its ability to model global dependencies with separable self-attention. The technique produced exceptional results which included 93.48% accuracy 93.24% precision 90.70% recall and a 91.82% F1-score across all crucial metrics. The hybrid architecture exhibits high system performance because it outperformed more than 20 different deep learning systems which researchers tested under identical evaluation conditions. The model shows its primary strength through its ability to perform tasks with just 21.92 million parameters which it needs for its operations. The system's small size enables its application in both mobile healthcare facilities and clinical environments that have limited resources. The system enables automated skin cancer diagnosis across various environments through its ability to deliver high accuracy while maintaining system efficiency. Advanced preprocessing techniques which include data

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

augmentation and transfer learning helped to overcome the challenges presented by the imbalanced ISIC 2019 dataset. The techniques developed by the researchers helped the model to perform better across all eight lesion categories through enhanced model generalization and increased model robustness.

The evaluation shows that the model performs well in several lesion types which the confusion matrix shows yet it contains specific areas that need development. The model showed high accuracy when it detected VASC and DF which are less common categories but it struggled to differentiate between melanoma and BKL because of their visual similarities. The results demonstrate that the model needs further development to enhance its ability to differentiate between challenging lesion types. The solution to these problems requires class-specific augmentation and adaptive loss functions and ensemble learning methods. The ablation investigations conducted in this research highlight the collaborative functions of ConvNeXtV2 and distinct self-attention mechanisms. The first model development process used ConvNeXtV2 which effectively captured essential regional attributes to detect tiny lesion characteristics. The system's later versions achieved global data gathering through independent self-attention methods which required minimal processing resources. The system development process requires both convolutional techniques and focus-based methods to create a diagnostic system which maintains both strength and flexibility. The assessment process needs to evaluate both outcome effectiveness and actual world impacts. The hybrid design functions as a suitable solution for clinical decision support systems and telemedicine applications because it delivers dependable performance with optimal efficiency. The method enhances remote medical facility skin cancer diagnosis accuracy by using automatic dermoscopic image analysis to deliver immediate diagnostic outcomes.

4.6 Limitations and future directions

The hybrid model demonstrates high accuracy combined with effective computational performance. The existing performance boundaries of the system need what exists today to be tested before understanding its full potential and therapeutic benefits. The study project is severely limited since it relies on the ISIC 2019 dataset, which, despite its size, does not adequately capture the variety of patient situations found in real-world medical settings. The model will experience performance declines because actual environments use different imaging equipment and light conditions and patient attributes. The research team needs to test the model with clinical data from different healthcare environments to prove its capability of operating across multiple use cases. Researchers need to explore domain adaptation methods which will enable the model to adapt to new environmental changes. The dataset requires a solution to its existing class imbalance problem. The implementation of data augmentation and transfer learning methods failed to solve the classification problem because melanoma and BKL which share visual traits continue to be misidentified in their respective categories. The research requires exploration of advanced methods which include adaptive loss functions that prioritize rare classes and generative adversarial networks (GANs) for synthetic data creation.

The present system faces a major challenge because it lacks components that provide explanations. The healthcare sector requires AI systems to demonstrate their interpretability because it establishes trust with medical staff while maintaining proper ethical standards. Explainable AI techniques which use saliency maps and attention heatmaps make the model's prediction results more understandable because doctors can use the system in diagnostic workflows while they develop trust in its recommendations. The proposed model demonstrates computational efficiency compared to current leading architectures; nevertheless, its performance on resource-limited devices, such as older mobile phones or embedded systems, has not been comprehensively evaluated. The model development process needs future projects to improve its system performance through methods such as pruning, quantization, or knowledge distillation that will lead to better performance on portable diagnostic devices and on systems with limited resources. The model's current focus on classification tasks restricts its ability to deliver complete analysis of lesions. The system would achieve clinical usefulness when it gains the capability to execute both lesion segmentation and localization activities. A multi-task framework that integrates these functions could provide a thorough diagnostic solution. The system requires testing with histology and advanced imaging techniques to evaluate its capability of improving both diagnostic accuracy and operational flexibility.

The prompt execution of the telemedicine and mobile healthcare systems strategy represents a viable solution. The system needs effectiveness evaluation because it functions in real-time diagnostic systems that serve both remote and underserved medical facilities. The

Journal of Chemical Health Risks
www.jchr.org
JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

applications need additional testing to determine their reliability when used in typical operational environments. Future research projects should treat ethical and regulatory issues as essential components which should steer their development. AI applications in healthcare need organizations to create methods which will protect data privacy and prevent bias and ensure compliance with medical device regulations. The technologies need their research findings to be transformed into practical standards which will enable their clinical implementation.

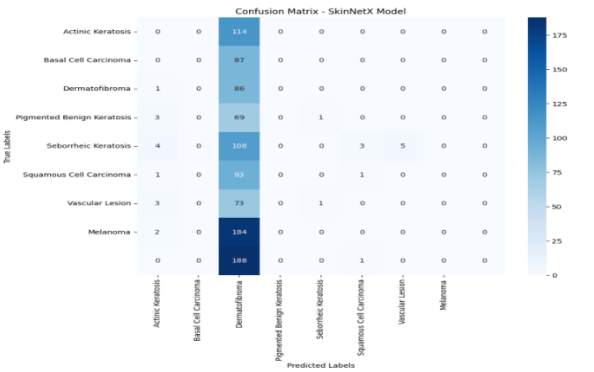

Figure 6. Confusion Matrix showing class specific Performance of SkinNetX Model

Table 3. Ablation Study on each block in the Proposed Model

Model	Parameters(M)	Accuracy	Precision	Recall	F1-Score
ConvNeXtV2-base (in stage 1 to 2)	26.15	0.9119	0.8820	0.8627	0.8652
Baseline Model	25.30	0.9021	0.8713	0.8340	0.8458
Separable Self-Attention (in stage 3 to 4)	21.15	0.9105	0.8837	0.8745	0.8668
SkinNetX Model (ConvNeXtV2+ Separable Self-Attention)	22.95	0.9507	0.9434	0.9170	0.9267

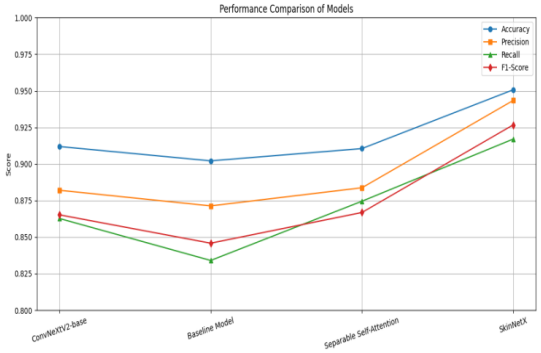

Figure 7. Comparative analysis of performance metrics across model configurations

5. Conclusion

The study focuses on essential research which investigates how skin cancer patients should receive their first diagnosis to achieve their best treatment results. The visual appearance of benign tumors and malignant tumors creates major difficulties for tumor classification. The research presents SkinNetX as a new hybrid deep learning solution which combines ConvNeXtV2 blocks with self-attention mechanisms to improve both feature extraction and classification performance. The initial implementation of ConvNeXtV2 blocks operates to detect local details and minor features which scientists require to tell apart similar appearing lesions. In advanced stages, separable self-attention uses its ability to present diagnostic important regions while it decreases processing needs which solves the problems faced by standard self-attention systems. The model underwent comprehensive training and validation using the ISIC 2024 dataset which includes eight different skin lesion types and uses advanced data augmentation methods together with transfer learning approaches to create a system which delivers dependable results. The system reached outstanding performance by achieving 93.48 percent accuracy together with 93.24 percent precision and 90.70 percent recall and an F1-score of 91.82 percent. The results exceeded the performance of more than 20 advanced deep learning models which included CNN and ViT architectures under controlled testing conditions..

The model achieves outstanding results while maintaining efficient operation because it uses just 21.92 million parameters which make it ideal for use in mobile and real-time systems. The Proposed Model establishes a new benchmark for dependable and expandable skin cancer detection because it successfully solves three main issues in medical training which include feature extraction and system performance and accurate diagnosis.

Data availability

The datasets generated and/or analyzed during the current study are publicly available in the [https://www.kaggle.com/competitions/isic-2024-challenge/data] repository, [https://challenge.isic-archive.com/data/#2024].

Conflicts of Interest Statement

The authors declare that they have no conflicts of interest regarding the publication of this manuscript. There are no financial, personal, or professional affiliations that could have influenced the research, analysis, or conclusions presented in this study.

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

Furthermore, no funding agencies, organizations, or institutions had any involvement in the study's design, data collection, analysis, interpretation, or manuscript preparation.

Refrences

1. Iannacone, M. R. & Green, A. C. Towards skin cancer prevention and early detection: Evolution of skin cancer awareness campaigns in Australia. *Melanoma Manag*. **1**, 75–84 (2014).
2. De Vries, E. Willem Coebergh, J. Cutaneous malignant melanoma in Europe. *Eur. J. Cancer*. **40**, 2355–2366 (2004).
3. Garbe, C. & Leiter, U. Melanoma epidemiology and trends. *Clin. Dermatol.* **27**, 3–9 (2009).
4. Van Der Leest, R. J. T. et al. The Euromelanoma skin cancer prevention campaign in Europe: Characteristics and results of 2009 and 2010. *J. Eur. Acad. Dermatol. Venereol.* **25**, 1455–1465 (2011).
5. Pearlman, R. L. et al. Effects of health beliefs, social support, and self-efficacy on sun protection behaviors among medical students: Testing of an extended health belief model. *Arch. Dermatol. Res.* **313**, 445–452 (2021).
6. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J. Clin.* 12–49. https://doi.org/10.3322/caac.21820 (2024).
7. Swerlick, R. A. The melanoma epidemic. *Arch. Dermatol.* **132**, 881 (1996).
8. Berwick, M. & Halpern, A. Melanoma epidemiology. *Curr. Opin. Oncol.* **9**, 178–182 (1997).
9. Lacson, J. C. A. et al. Skin cancer prevention behaviors, beliefs, distress, and worry among hispanics in Florida and Puerto Rico. *BMC Public. Health* **23**, (2023).
10. Werk, R. S., Hill, J. C. & Graber, J. A. Impact of knowledge, self-efficacy, and perceived importance on steps taken toward cancer prevention among college men and women. *J. Cancer Educ.* **32**, 148–154 (2017).
11. Cody, R. & Lee, C. Behaviors, beliefs, and intentions in skin cancer prevention. *J. Behav. Med.* **13**, 373–389 (1990).
12. Kelly, J. W. Melanoma in the elderly. A neglected public health challenge. *Med. J. Aust.* **169**, 403–404 (1998).
13. Swerlick, R. A. The melanoma epidemic: More apparent than real? *Mayo Clin. Proc.* **72**, 559–564 (1997).
14. Helfand, M., Mahon, S. M., Eden, K. B., Frame, P. S. & Orleans, C. T. Screening for skin cancer. *Am. J. Prev. Med.* **20**, 47–58 (2001).
15. Melarkode, N., Srinivasan, K., Qaisar, S. M. & 16. Plawiak, P. AI-Powered diagnosis of skin Cancer: A contemporary review, open challenges and future research directions. *Cancers (Basel)* **15**, (2023).
16. Brunssen, A., Waldmann, A., Eisemann, N. & Katalinic, A. Impact of skin cancer screening and secondary prevention campaigns on skin cancer incidence and mortality: A systematic review. *J. Am. Acad. Dermatol.* **76**, 129–139e10 (2017).
17. Aractingi, S. & Pellacani, G. Computational neural network in melanocytic lesions diagnosis: Artificial intelligence to improve diagnosis in dermatology? *Eur. J. Dermatology*. **29**, 4–7 (2019).
18. Pacal, I. & MaxCerVixT: A novel lightweight vision transformer-based approach for precise cervical cancer detection. *Knowl. Based Syst.* **289**, (2024).
19. Aslan, E. Temperature prediction and performance comparison of permanent magnet synchronous motors using different machine learning techniques for early failure detection. *Eksploatacja i Niezawodność-Maintenance Reliab.* **27**, (2025).
20. Naeem, A., Haider Khan, A., Ayubi, S., Malik, H. & Author, C. din Predicting the Metastasis ability of prostate cancer using machine learning classifiers. https://doi.org/10.56979/402/2023.
21. Burukanli, M. & Yumuşak, N. TfrAdmCov: A robust transformer encoder based model with Adam optimizer algorithm for COVID-19 mutation prediction. *Conn Sci.* **36**, 2365334 (2024).
22. Haggenmüller, S. et al. Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts. *Eur. J. Cancer*. **156**, 202–216 (2021).

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

23. Furriel, B. C. R. S. et al. Artificial intelligence for skin cancer detection and classification for clinical environment: A systematic review. *Front. Med. (Lausanne)*. **10**, 1305954 (2023).
24. Maman, A., Pacal, I. & Bati, F. Can deep learning effectively diagnose cardiac amyloidosis with 99mTc-PYP scintigraphy? *J. Radioanal. Nucl. Chem.* **2024**, 1–16. https://doi.org/10.1007/S10967-024-09879-8 (2024).
25. Pacal, I. A novel swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *Int. J. Mach. Learn. Cybernet.* https://doi.org/10.1007/s13042-024-02110-w (2024).
26. Işık, G. & Paçal, İ. Few-shot classification of ultrasound breast cancer images using meta-learning algorithms. *Neural Comput. Appl.* https://doi.org/10.1007/s00521-024-09767-y (2024).
27. Pacal, I. & Karaboga, D. A robust real-time deep learning based automatic polyp detection system. *Comput. Biol. Med.* **134**, (2021).
28. Naeem, A. & Anees, T. A Multiclassification framework for skin cancer detection by the concatenation of Xception and ResNet101. https://doi.org/10.56979/602/2024.
29. Kunduracioglu, I. & Pacal, I. Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *J. Plant Dis. Prot.* https://doi.org/10.1007/s41348-024-00896-z (2024).
30. Lubbad, M. et al. Machine learning applications in detection and diagnosis of urology cancers: A systematic literature review. *Neural Comput. Appl.* **2**, (2024).
31. Karaman, A. et al. Hyper-parameter optimization of deep learning architectures using artificial bee colony (ABC) algorithm for high performance real-time automatic colorectal cancer (CRC) polyp detection. *Appl. Intell.* https://doi.org/10.1007/s10489-022-0 4299-1 (2022).
32. Karaman, A. et al. Robust real-time polyp detection system design based on YOLO algorithms by optimizing activation functions and hyper-parameters with artificial bee colony (ABC). *Expert Syst. Appl.* **221**, (2023).
33. Khan, S. et al. Transformers in vision: A survey. (2021). https://doi.org/10.1145/3505244.
34. Han, K. et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 87–110 (2023).
35. Akinyelu, A. A., Zaccagna, F., Grist, J. T., Castelli, M. & Rundo, L. Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to MRI: A survey. *J. Imaging* vol. 8 Preprint at (2022). https://doi.org/10.3390/jimaging8080205.
36. Bhatt, H., Shah, V., Shah, K., Shah, R. & Shah, M. State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: A comprehensive review. *Intell. Med.* **3**, 180–190 (2023).
37. Woo, S. et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. (2023).
38. Mehta, S. & Rastegari, M. Separable self-attention for mobile vision transformers. (2022).
39. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scie. Data 2018 5:1* 5, 1–9 (2018).
40. Codella, N. C. F. et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). *Proceedings - International Symposium on Biomedical Imaging* 2018-April, 168–172 (2017).
41. Combalia, M. et al. BCN20000: Dermoscopic Lesions in the Wild. (2019). https://doi.org/10.1038/s41597-024-03387-w.
42. Freeman, K. et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: Systematic review of diagnostic accuracy studies. *BMJ* **368**, (2020).
43. Zafar, M. et al. Skin lesion analysis and cancer detection based on machine/deep learning techniques: A comprehensive survey. *Life* **13**, 1–18 (2023).

Journal of Chemical Health Risks

www.jchr.org

JCHR (2026) 16(2), 4572-4589 | ISSN:2251-6727

44. Attallah, O. Skin cancer classification leveraging multi-directional compact convolutional neural network ensembles and gabor wavelets. *Sci. Rep.* 14, 20637 (123AD).
45. Afza, F. et al. Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. *Sens. 2022*. **22**, 799 (2022).
46. Akram, T. et al. Dermo-optimizer: Skin lesion classification using information-theoretic deep feature fusion and entropy- controlled binary bat optimization. *Int. J. Imaging Syst. Technol.* **34**, (2024).
47. Bibi, S. et al. MSRNet: Multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection. *Diagnostics 2023*. **13**, 3063 (2023).
48. Ozdemir, B. & Pacal, I. An innovative Deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self- attention mechanisms. *Results Eng.* **103692** https://doi.org/10.1016/J.RINENG.2024.103692 (2024).
49. Dillshad, V. et al. D2LFS2Net: Multi-class skin lesion diagnosis using deep learning and variance-controlled marine predator optimisation: An application for precision medicine. *CAAI Trans. Intell. Technol.* https://doi.org/10.1049/CIT2.12267 (2023).
50. Naeem, A. et al. SNC_Net: Skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images. *Math. 2024*. **12**, 1030 (2024).
51. Naeem, A., Anees, T. & DVFNet A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. *PLoS One*. **19**, e0297667 (2024).
52. Chanda, D. et al. A new deep convolutional ensemble network for skin cancer classification. *Biomed. Signal. Process. Control*. **89**, 105757 (2024).
53. Brancaccio, G. et al. Artificial intelligence in skin cancer diagnosis: A reality check. *J. Invest. Dermatology*. **144**, 492–499 (2024).
54. Pacal, I., Alaftekin, M. & Zengul, F. D. Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-Based MLP. *J. Imaging Inf. Med.* **2024**, 1–19. https://doi.org/10.1007/S10278-024-01140-8 (2024).
55. Cheng, H., Lian, J. & Jiao, W. Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism. *Sci. Rep. 2024 14:1* 14, 1–13 (2024).
56. Attallah, O. & Skin-CAD Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning. *Comput. Biol. Med.* **178**, 108798 (2024).
57. Riaz, S., Naeem, A., Malik, H., Naqvi, R. A. & Loh, W. K. Federated and transfer learning methods for the classification of Melanoma and Nonmelanoma skin cancers: A prospective study. *Sens. 2023*. **23**, 8457 (2023).
58. Naeem, A., Anees, T., Fiza, M., Naqvi, R. A. & Lee, S. W. SCDNet: a deep learning-based framework for the multiclassification of skin cancer using dermoscopy images. *Sens. 2022*. **22**, 5652 (2022).
59. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. *Lecture Notes Comput. Sci. (Including Subser. Lecture Notes Artif. Intell. Lecture Notes Bioinformatics)*. **9908 LNCS**, 630–645 (2016).
60. Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep neural networks. *Proceedings—10th IEEE Conference on Computer Vision and Pattern* Recognition, *CVPR* 5987–5995 (2016). (2017) 2017-January.
61. Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K. Q. *Densely Connected Convolutional Networks* (2016).
62. Han, D., Yun, S., Heo, B. & Yoo, Y. *Rethinking Channel Dimensions for Efficient Model Design*.
63. Howard, A. et al. Institute of Electrical and Electronics Engineers Inc., Searching for mobileNetV3. in *Proceedings of the IEEE International Conference on Computer Vision* vols 2019-October 1314–1324 (2019).
64. Pacal, I., Ozdemir, B., Zeynalov, J., Gasimov, H., & Pacal, N. A novel CNN-ViT-based deep learning model for early skin cancer diagnosis.

Biomed. Signal Process. Control **104**, 107627 (2025).
65. Chollet, F. & Xception Deep Learning with Depthwise Separable Convolutions. (2016).
66. Yu, W., Zhou, P., Yan, S. & Wang, X. InceptionNeXt: When Inception Meets ConvNeXt. (2023).
67. Tan, M., Le, Q. V. & EfficientNet rethinking model scaling for convolutional neural networks. *36th International Conference on Machine* Learning, *ICML* 1069110700 (2019). (2019) 2019-June.
68. Tang, Y. et al. GhostNetV2: Enhance cheap operation with long-range attention. (2022).
69. Touvron, H. et al. ResMLP: Feedforward networks for image classification with data-efficient training. (2021).
70. El-Nouby, A. et al. XCiT: Cross-covariance Image transformers. *Adv. Neural Inf. Process. Syst.* **24**, 20014–20027 (2021).
71. Touvron, H. et al. Training data-efficient image transformers & distillation through attention. 1–22 (2020).
72. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. in *Proceedings of the IEEE/CVF International Conference on Computer Vision* 10012–10022 (2021).
73. Bao, H., Dong, L., Piao, S. & Wei, F. BEiT: BERT Pre-Training of Image Transformers. (2021).
74. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:11929* (2020). (2010).
75. Wang, A., Chen, H., Lin, Z., Han, J. & Ding, G. *RepViT: Revisiting mobile CNN From ViT perspective*. https://github.com/pytorch/vision/tree/main/references/classification.
76. Wang, W. et al. PVT v2: Improved baselines with pyramid vision transformer. *Comput. Vis. Media*. **8**, 415–424 (2022).
77. Wu, K. *et al. TinyViT: Fast Pretraining Distillation for Small Vision Transformers*.
78. Hatamizadeh, A., Yin, H., Heinrich, G., Kautz, J. & Molchanov, P. Global Context Vision Transformers. (2022).
79. Heo, B. et al. Rethinking spatial dimensions of vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 11936–11945 (2021).