scientific reports

Check for updates

OPEN

A robust deep learning framework for multiclass skin cancer classification

Burhanettin Ozdemir[1]✉ & Ishak Pacal[2,3]

Skin cancer represents a significant global health concern, where early and precise diagnosis plays a pivotal role in improving treatment efficacy and patient survival rates. Nonetheless, the inherent visual similarities between benign and malignant lesions pose substantial challenges to accurate classification. To overcome these obstacles, this study proposes an innovative hybrid deep learning model that combines ConvNeXtV2 blocks and separable self-attention mechanisms, tailored to enhance feature extraction and optimize classification performance. The inclusion of ConvNeXtV2 blocks in the initial two stages is driven by their ability to effectively capture fine-grained local features and subtle patterns, which are critical for distinguishing between visually similar lesion types. Meanwhile, the adoption of separable self-attention in the later stages allows the model to selectively prioritize diagnostically relevant regions while minimizing computational complexity, addressing the inefficiencies often associated with traditional self-attention mechanisms. The model was comprehensively trained and validated on the ISIC 2019 dataset, which includes eight distinct skin lesion categories. Advanced methodologies such as data augmentation and transfer learning were employed to further enhance model robustness and reliability. The proposed architecture achieved exceptional performance metrics, with 93.48% accuracy, 93.24% precision, 90.70% recall, and a 91.82% F1-score, outperforming over ten Convolutional Neural Network (CNN) based and over ten Vision Transformer (ViT) based models tested under comparable conditions. Despite its robust performance, the model maintains a compact design with only 21.92 million parameters, making it highly efficient and suitable for model deployment. The Proposed Model demonstrates exceptional accuracy and generalizability across diverse skin lesion classes, establishing a reliable framework for early and accurate skin cancer diagnosis in clinical practice.

Keywords Medical image analysis, Health, Skin cancer detection, ConvNeXtv2, Vision Transformer

As the body's primary defense against external threats, the skin plays a vital role in maintaining overall health and well-being[1]. It is the largest organ of the human body, serving as a protective barrier against harmful microorganisms, physical trauma, and harsh environmental factors[2]. Composed of three main layers: the epidermis, dermis, and hypodermis, the skin not only shields internal systems but also regulates temperature, facilitates vitamin D synthesis, and sustains hydration and electrolyte balance[3]. However, continuous exposure to external elements increases the risk of various conditions, including cancer, a disease marked by the abnormal and uncontrolled growth of cells[4]. Cancer can affect virtually any organ or tissue in the body, leading to severe health consequences and requiring timely intervention for effective management[5]. Among these, skin cancer represents a particularly pressing public health issue, influenced by factors such as prolonged sun exposure, genetic predisposition, and environmental triggers. Alarmingly, projections for 2024 in the United States estimate 108,270 new cases of skin cancer, with 13,120 related deaths[6].

Skin cancer is broadly categorized into two primary forms: melanoma and nonmelanoma[7]. Melanoma, known for its aggressive and life-threatening nature, accounts for the majority of fatalities associated with skin cancer[7,8]. Detecting melanoma at an early stage is essential for improving patient outcomes and expanding treatment options, as is the case with many other types of cancer[7–10]. Early detection not only allows for timely medical intervention but also significantly increases the likelihood of successful treatment, highlighting the importance of advancing diagnostic methods for skin cancer[8,11]. While traditional diagnostic approaches such

[1]Department of Operations and Project Management, College of Business, Alfaisal University, Riyadh 11533, Saudi Arabia. [2]Department of Computer Engineering, Faculty of Engineering, Igdir University, Igdir 76000, Turkey. [3] Department of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012 Nakhchivan, Azerbaijan. ✉email: bozdemir@alfaisal.edu

as visual inspection, dermoscopy, and biopsy remain fundamental, they are often invasive, time-intensive, and heavily reliant on specialized expertise[12]. These challenges, along with escalating healthcare costs, emphasize the critical need for innovative, accurate, and efficient diagnostic technologies to address the growing impact of skin cancer[13].

Traditional approaches for diagnosing skin cancer, such as physical examinations, dermoscopy, and biopsies, have been the foundation of clinical practice[14]. However, these methods are inherently invasive, time-intensive, and heavily reliant on specialized expertise, which limits their accessibility and overall efficiency. The increasing demand for precise and scalable diagnostic solutions has driven the adoption of advanced technological methodologies, including machine learning, deep learning, and computer-aided diagnosis (CAD) systems[15,16].

Machine learning, while effective in automating various diagnostic processes, is constrained by inherent challenges such as feature selection, limited generalizability, and reliance on manual input[17–19]. Conventional algorithms like support vector machines and k-nearest neighbors depend on predefined features, which often fail to encapsulate the complex and variable nature of medical imaging data[20]. To address these limitations, deep learning, a subset of artificial intelligence, has emerged as a transformative paradigm[21]. Unlike traditional machine learning methods, deep learning employs multilayered neural networks to autonomously extract patterns and features from raw data, resulting in significantly enhanced diagnostic accuracy[22–24]. This approach has shown exceptional promise in the field of medical imaging and cancer detection, with notable success in identifying cancers such as lung, breast, and skin, as well as in analyzing imaging modalities like MRI, CT, and histopathology data[25].

Within deep learning architectures, convolutional neural networks (CNNs) have established themselves as the cornerstone of medical image analysis[26,27]. By excelling in the extraction of spatial features, CNNs are particularly effective in identifying cancerous lesions and achieving high levels of accuracy in tasks like skin lesion classification and tumor segmentation[28]. Vision Transformers (ViTs), a more recent advancement, employ self-attention mechanisms to process image data and have demonstrated their capacity to handle large-scale datasets and extract intricate global patterns[29–32]. While CNNs are unrivaled in their ability to capture detailed local features, ViTs are highly adept at modeling broader, more complex relationships within image data, making these technologies complementary in skin cancer diagnosis[33–35].

Despite these advancements, several challenges persist. Traditional machine learning methods grapple with scalability and an overdependence on predefined features, while deep learning models face issues such as class imbalance, overfitting, and high computational resource demands[23,36]. Moreover, the scarcity of diverse, well-annotated datasets remains a critical bottleneck. Ethical considerations, along with the need for interpretable models, further complicate the clinical implementation of these technologies.

To address these challenges, this study introduces a hybrid framework that combines the strengths of CNNs and ViTs. This integrated approach utilizes the fine-grained feature extraction capabilities of CNNs alongside the global dependency modeling strengths of ViTs, creating a robust and efficient solution for skin lesion classification. The framework also incorporates advanced preprocessing techniques to address data imbalance and mitigate overfitting, thereby enhancing the reliability and generalizability of the model. The proposed framework demonstrates strong generalizability across diverse skin lesion types, addressing critical challenges in clinical settings and contributing to early and accurate skin cancer diagnosis.

The primary contributions of this study are as follows.

- This study proposes a novel hybrid deep learning model that integrates ConvNeXtV2[37] blocks for fine-grained feature extraction and separable self-attention mechanisms[38] for efficient global context modeling, tailored to the challenges of skin lesion classification.
- The model was trained and validated on the ISIC 2019[39–41] dataset, achieving state-of-the-art performance with 93.48% accuracy, 93.24% precision, 90.70% recall, and a 91.82% F1-score, outperforming over 20 state-of-the-art deep learning models, including more than ten CNN-based and ViT-based architectures, trained under the same conditions.
- Despite its robust performance, the model maintains a compact design with 21.92 million parameters, optimizing computational efficiency without compromising accuracy.

The following section offers a comprehensive review of the relevant literature, establishing a solid groundwork for the study. Subsequently, the materials and methods are described in detail in the third section to ensure methodological transparency. The fourth section focuses on presenting the experimental results, while the fifth section provides a thorough analysis and interpretation of these findings in the context of the research objectives.

Related works

The application of deep learning to skin cancer detection has attracted considerable interest, driven by the pressing need for diagnostic methods that are efficient, accurate, and scalable. As a branch of artificial intelligence, deep learning has transformed the field of medical imaging by enabling the analysis of intricate patterns and large datasets with minimal human involvement. Through automated feature extraction and classification, deep learning models have achieved exceptional results in identifying and categorizing skin lesions, establishing their critical role in the early diagnosis and treatment of cancer.

CNNs have emerged as one of the most effective architectures for medical image analysis, excelling in the extraction of detailed spatial features. This makes them particularly well-suited for tasks such as classifying skin lesions and segmenting tumors. ViTs, a more recent innovation, employ self-attention mechanisms to analyze image data. While CNNs are ideal for capturing localized features, ViTs are adept at modeling broader, global relationships, making the two approaches highly complementary in tackling the complexities of skin cancer detection. Hybrid models that merge the strengths of CNNs and ViTs have further enhanced the potential of

deep learning in this domain. By combining the fine-grained feature extraction of CNNs with the global pattern recognition capabilities of ViTs, these models are equipped to handle diverse and imbalanced datasets more effectively, achieving higher levels of accuracy, reliability, and generalizability.

The effectiveness of deep learning in skin cancer diagnosis has been validated by a wealth of studies, emphasizing its transformative impact on medical diagnostics[15,16,22]. Comparative reviews have analyzed various algorithms, shedding light on their advantages, limitations, and the importance of reliable, non-invasive diagnostic methods[23,36,42]. Comprehensive evaluations of deep learning-based segmentation have explored key factors such as dataset attributes, model architecture, and assessment criteria. Additionally, research has demonstrated the capacity of these models to detect skin cancer at early stages, highlighting the value of automated systems in improving diagnostic accuracy and efficiency[43]. The following are summaries of notable studies in this field.

Attallah proposed SCaLiNG, a CAD tool that combines compact CNNs and Gabor Wavelets (GW) to extract spatial, textural, and frequency features for skin cancer classification. By processing images through GW sub-bands and integrating features from multiple CNNs, SCaLiNG achieves a more detailed feature representation. A feature selection step further optimizes the model's performance. With an accuracy of 0.9170, SCaLiNG outperforms traditional single-CNN models, demonstrating its effectiveness[44]. Afza et al. proposed an innovative approach for classifying multiclass skin lesions by integrating deep learning feature fusion with extreme learning machines (ELM). Their framework includes steps for image enhancement, transfer learning-based feature extraction, hybrid optimization for feature selection, feature integration, and ELM-driven classification. When evaluated on the HAM10000 and ISIC2018 datasets, the method demonstrated superior accuracy of 93.40% and 94.36%, respectively, surpassing existing techniques in both performance and computational efficiency[45]. Akram et al. introduced a novel approach to enhance and streamline feature representation by integrating multiple deep learning models with an information-theoretic fusion strategy. To eliminate noise and redundancy, the method employs an entropy-driven binary bat selection algorithm, ensuring the retention of essential and distinctive features. Using transfer learning, features were extracted from Inception-ResNet V2, DenseNet-201, and Nasnet Mobile, followed by fusion and refinement to optimize the feature set. The approach demonstrated its effectiveness when tested on PH2, ISIC-2016, and ISIC-2017 datasets, highlighting its capability to deliver informative and unique feature representations[46]. Bibi et al. developed a deep learning framework for multiclass skin cancer detection, involving image preprocessing, feature extraction, selection, and classification. A novel luminance-based contrast enhancement improves image quality, while modified DarkNet-53 and DenseNet-201 models, optimized with a genetic algorithm, are used for feature extraction. Features are fused with a serial-harmonic mean method, and irrelevant data is eliminated using marine predator optimization guided by Reyni Entropy. The refined features are then classified using machine learning algorithms for accurate diagnosis[47]. Ozdemir and Pacal present a lightweight and mobile-friendly hybrid model designed to address challenges such as data imbalance and high model complexity. The model focuses on extracting critical features in the initial stages and enhances sensitivity in later stages by concentrating on diagnostically significant areas. Evaluated using the ISIC 2019 dataset, which includes eight highly imbalanced skin cancer classes, the model delivered strong performance with 93.60% accuracy, 91.69% precision, 90.05% recall, and a 90.73% F1-score[48]. Dillshad et al. proposed an advanced deep learning system for multiclass skin lesion classification. The method includes hybrid contrast enhancement for preprocessing and innovative data augmentation without traditional techniques. Features are extracted from MobileNetV2 and NasNet Mobile using transfer learning and fused through a dual-threshold serial approach. Optimized feature selection is achieved using the variance-controlled Marine Predator algorithm, followed by classification with machine learning, offering a precise and efficient solution[49]. Naeem et al. introduced SNC_Net, a hybrid model combining deep learning and handcrafted features to classify eight skin cancer types. Utilizing the ISIC 2019 dataset, the model employs a convolutional neural network (CNN) for training and validation. SNC_Net outperformed several baseline models, including EfficientNetB0 and ResNet-101, as well as state-of-the-art classifiers, showcasing its effectiveness in skin cancer detection[50]. Naeem and Anees introduced DVFNet, a deep learning approach for skin cancer detection using dermoscopy images. The method enhances image quality through anisotropic diffusion and extracts features by combining VGG19 and Histogram of Oriented Gradients (HOG). To address class imbalance in the ISIC 2019 dataset, SMOTE Tomek is applied. Segmentation highlights damaged skin areas, and a CNN performs multiclass classification using a feature vector map generated from HOG and VGG19 features[51].

Chanda et al. introduced DCENSnet, an ensemble of three deep convolutional neural networks (DCNNs) with customized dropout layers to enhance feature learning and achieve an optimal bias–variance trade-off. Evaluated on the HAM10000 dataset, the model demonstrated exceptional performance with a mean accuracy of 99.53%, along with high precision, recall, F1 score, and AUC for each class. This method outperforms state-of-the-art networks and shows significant potential for reliable computer-aided detection, classification, and analysis of malignant skin lesions, aiming to improve diagnostic and treatment accuracy[52]. Brancaccio et al. emphasize that the best outcomes for AI-based diagnostic tools are achieved through collaboration with human experts. While most studies assess AI performance in controlled environments, its real-world clinical effectiveness and implementation challenges remain unclear. The review explores AI's potential benefits and limitations for consumers, general practitioners, and dermatologists[53]. Pacal et al. advanced the Swin Transformer architecture by integrating hybrid shifted window-based multi-head self-attention (HSW-MSA), enhancing its capability to effectively handle overlapping skin cancer regions, extract intricate details, and address long-range dependencies, all while optimizing memory and computational efficiency. Furthermore, they substituted the conventional multi-layer perceptron (MLP) with a SwiGLU-based MLP, a refined version of the gated linear unit (GLU), to achieve superior accuracy, faster training, and improved parameter optimization[54]. Cheng et al. proposed a deep learning model for skin cancer classification, integrating convolutional neural networks for local feature extraction with an attention mechanism for global associations. Tested on the ISIC-2019 dataset,

the model outperformed state-of-the-art methods in Precision, Recall, and F1-scores, showing promise as a reliable diagnostic tool[55]. Attallah proposed "Skin-CAD," an explainable AI system for classifying dermoscopic skin cancer images as benign or malignant, with further subclassification into seven types. The system combines features from multiple CNNs, reducing dimensionality with PCA to optimize training and simplify computation. Key features are selected to enhance classification accuracy, and the LIME method provides interpretability for the predictions[56]. Riaz et al. conducted a comprehensive systematic literature review to assess the effectiveness of federated learning and transfer learning algorithms in detecting malignant skin cancer. The review analyzed performance metrics, including true positive rate, true negative rate, area under the curve, and accuracy, by examining 86 studies published between January 2018 and July 2023 from seven prominent databases. Additionally, the authors proposed a taxonomy that categorizes malignant and non-malignant skin cancer classes. The findings provided insights into the capabilities of FL and TL classifiers while highlighting the key limitations and challenges faced in recent advancements[57]. Naeem et al. developed SCDNet, a deep learning model combining VGG16 and CNNs for multiclass skin cancer classification, including melanoma, Basal Cell Carcinoma (BCC), and Benign Keratosis (BKL). Using the ISIC 2019 dataset, SCDNet achieved a 96.91% accuracy, outperforming state-of-the-art models such as ResNet50, AlexNet, VGG19, and Inception-v3, which achieved accuracies of 95.21%, 93.14%, 94.25%, and 92.54%, respectively. The study highlights SCDNet's superior performance in accurately detecting multiple skin cancer types[58].

Recent advancements in skin cancer detection have adopted a variety of innovative methods, combining sophisticated deep learning models with traditional feature extraction and optimization techniques. Approaches like integrating compact CNNs with Gabor wavelets or hybrid systems such as SCaLiNG have achieved enhanced accuracy by utilizing spatial, textural, and frequency features. Other models, including SNC_Net and DVFNet, merge handcrafted features with deep learning algorithms to address challenges like dataset imbalance and improve precision. Hybrid architectures that integrate CNNs for localized feature extraction and Vision Transformers (ViTs) for capturing global patterns represent a promising avenue, effectively tackling issues such as class imbalance, feature redundancy, and computational inefficiencies.

While these developments mark significant progress, obstacles persist, including computational demands, challenges in interpretability, and variability in datasets. Distinguishing itself from these methods, our proposed model combines the strengths of CNNs and ViTs within a unified hybrid framework. This approach not only resolves key limitations but also achieves superior accuracy, robustness, and scalability. By advancing the integration of cutting-edge techniques, our model sets a new standard for skin cancer diagnostics, demonstrating its potential as a transformative solution in the field.

Methodology

This study utilizes the ISIC 2019 dataset, a highly regarded and diverse benchmark resource widely used for advancing research in skin cancer detection. Encompassing a broad array of skin lesion types, this dataset serves as an ideal foundation for rigorously evaluating the performance and robustness of state-of-the-art diagnostic approaches. Our methodology is built upon a cutting-edge hybrid deep learning framework that seamlessly integrates ConvNeXtV2 blocks with separable self-attention mechanisms. This sophisticated architecture combines the precision of local feature extraction with the ability to capture global contextual patterns, resulting in unparalleled sensitivity and specificity in identifying and classifying skin cancers. Additionally, the model incorporates ViTs, augmented by advanced data augmentation techniques and transfer learning processes, to enhance its capacity to analyze intricate details and contextual relationships in dermoscopic imagery. To ensure the study's reproducibility and its contribution to the broader scientific community, we provide an in-depth description of the model's design, implementation, and training protocols. This comprehensive documentation is intended to inspire further advancements in cancer detection methodologies and drive innovation in the field of medical imaging.

Dataset

The ISIC 2019 dataset stands as one of the most comprehensive and influential publicly available resources for advancing research in skin cancer detection. Curated by the International Skin Imaging Collaboration, this dataset has become pivotal in the development of deep learning and AI-driven diagnostic solutions for skin cancer[40]. It provides an extensive collection of dermoscopic images, accompanied by detailed demographic and clinical metadata linked to skin lesion diagnoses. Widely utilized by researchers, the dataset supports diverse tasks such as early melanoma detection and the classification of various skin cancer types, with its structured training, validation, and testing subsets facilitating robust algorithm development. The ISIC 2019 dataset has been instrumental in driving innovation, enabling the development of state-of-the-art deep learning methodologies and establishing new benchmarks in the field. Figure 1 illustrates representative images from the dataset's diverse skin lesion categories.

Figure 1 showcases five sample images per class from the ISIC 2019 dataset, providing an overview of the data distribution across categories. The dataset comprises 25,331 labeled dermoscopic images, classified into eight distinct skin lesion categories: Melanoma, Melanocytic Nevus (NV), BCC, Actinic Keratosis (AK), BKL, DF, Vascular Lesion (VASC), and Squamous Cell Carcinoma (SCC). The images exhibit considerable variation in resolution, ranging from 576×768 to 1024×1024 pixels, distributed across 101 unique resolution sizes. All images are in full color, utilizing three RGB channels. One significant challenge posed by this dataset is the severe class imbalance. For example, the NV class contains roughly 51 times more images than the VASC class, which can skew model training and adversely affect classification accuracy. Addressing this imbalance necessitates advanced techniques such as data augmentation, weighted loss functions, and sophisticated sampling methods. In response, this study adopts a CNN-ViT hybrid model, using the local feature extraction strengths of CNNs

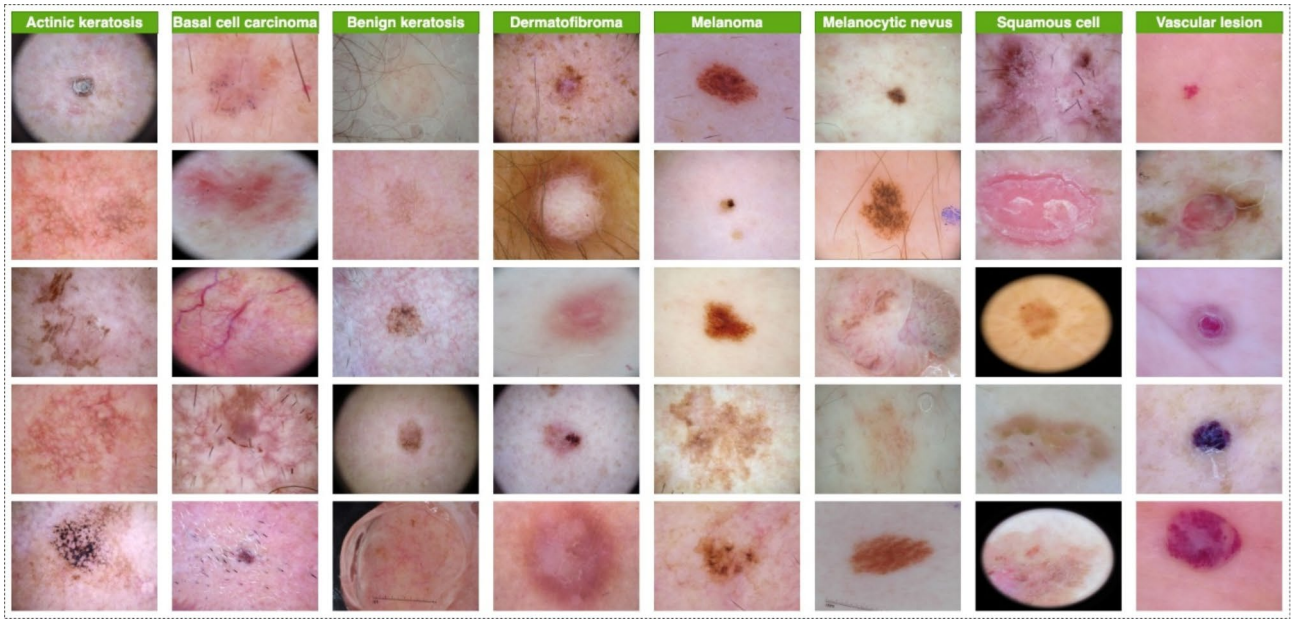

Fig. 1. Representative sample images from each class in the ISIC 2019 dataset.

and the global pattern recognition capabilities of ViTs. This integrated approach effectively addresses the dataset imbalance, delivering enhanced performance and robust classification of skin lesions.

Deep learning approaches

Deep learning, a branch of machine learning, is distinguished by its ability to identify intricate patterns and representations within data through its layered network architecture. These models automate feature extraction directly from raw data, eliminating the need for manual intervention. By employing hierarchical feature learning, deep learning algorithms uncover increasingly abstract and complex representations, achieving exceptional performance in domains such as image processing, natural language understanding, and audio analysis. Deep learning methodologies span both supervised and unsupervised learning paradigms. In supervised learning, models utilize labeled data to map inputs to outputs, making this approach highly effective for tasks such as classification and regression. Conversely, unsupervised learning deals with unlabeled data, uncovering latent structures for applications like clustering, dimensionality reduction, and feature learning. Both paradigms are essential for the adaptability and versatility of deep learning systems. Key architectures in deep learning include CNNs and ViTs. CNNs are instrumental in image analysis, utilizing convolutional layers to extract localized features, pooling layers to reduce dimensionality, and fully connected layers for predictions. Their hierarchical structure is optimized to capture spatial relationships within images. In contrast, ViTs employ self-attention mechanisms and positional embeddings to model global dependencies across entire image patches, offering a comprehensive perspective on spatial relationships. Together, CNNs and ViTs form the backbone of contemporary computer vision, each excelling in distinct yet complementary aspects of deep learning.

Proposed model

We introduce a hybrid model for skin lesion classification that integrates convolutional and attention-based frameworks to optimize both local and global feature representation. The model employs ConvNeXtV2 blocks in its early stages to efficiently capture hierarchical features, transitioning to separable self-attention layers in the later stages to effectively capture long-range dependencies. ConvNeXtV2 enhances conventional CNNs with larger kernels, depth-wise separable convolutions, and layer normalization, enabling the model to capture intricate patterns and fine details essential for distinguishing benign from malignant lesions. This improves the model's ability to detect subtle anomalies and classify skin cancers more accurately. In later stages, separable self-attention replaces standard self-attention, reducing computational complexity by decoupling spatial and channel dimensions while retaining global dependency capture. This mechanism focuses on diagnostically relevant regions, suppresses background noise, and enhances sensitivity and specificity, making the model highly effective for diverse skin cancer classifications. This approach combines the computational efficiency of convolutional networks with the adaptability of attention mechanisms, ensuring precise classification of skin lesion types. The architecture is structured into four stages, each beginning with a downsampling operation to decrease spatial dimensions while increasing the number of feature channels. In the first two stages, ConvNeXtV2 blocks focus on extracting fine-grained details, such as texture and pigmentation patterns in skin lesions. The final two stages incorporate separable self-attention layers, which efficiently aggregate global contextual information while minimizing computational overhead. The model processes input images of 224×224 pixels progressively, preserving critical low-level features in the early stages while building global representations in the later stages.

This hybrid design strikes a balance between computational efficiency and advanced representation learning, as depicted in Fig. 2.

Figure 2 displays the comprehensive architecture of the proposed model designed for automated skin cancer diagnosis. The proposed model combines convolutional and attention-based mechanisms in a unified architecture tailored for skin lesion classification. It is organized into four stages, each designed to progressively downsample input dermatologic images while increasing the number of feature channels. This hierarchical structure facilitates efficient extraction of both localized and global features, which are critical for accurately diagnosing various types of skin lesions. The configuration of the model consists of 3, 3, 9, and 12 layers across the four stages, with each stage contributing uniquely to feature representation and classification. Convolutional architectures have long been recognized for their ability to capture local spatial patterns efficiently. ConvNeXtV2, a modern convolutional architecture, was developed to address the limitations of earlier convolutional models by integrating advancements inspired by transformer architectures. It incorporates improvements such as depthwise convolution, layer normalization, and global response normalization, which together enhance its ability to capture hierarchical features while maintaining computational efficiency. These innovations are particularly relevant in skin lesion analysis, where fine-grained details such as pigmentation irregularities, texture variations, and lesion borders are essential for accurate classification. In the first stage, three ConvNeXtV2 blocks process the input dermatologic images to extract low-level features, laying the foundation for hierarchical feature representation. In the second stage, another three ConvNeXtV2 blocks refine these features while reducing the spatial dimensions and increasing the feature channels. This design ensures that critical dermatologic details are preserved and hierarchically organized, enabling the model to learn discriminative features required for identifying benign and malignant lesions.

Attention mechanisms have revolutionized deep learning by enabling models to capture long-range dependencies and global contextual relationships. Traditional self-attention mechanisms, however, are computationally expensive, particularly for high-resolution images. Separable self-attention was introduced as a computationally efficient alternative, designed to reduce the complexity of self-attention from $O(n^2)$ to $O(n)$ by focusing on interactions with a latent token instead of pairwise relationships among all tokens. This approach maintains the ability to model global dependencies while significantly reducing computational costs, making it ideal for resource-intensive tasks like skin cancer diagnosis. The third stage of the model consists of nine separable self-attention layers, which aggregate global contextual information efficiently. These layers capture high-level patterns such as asymmetry, irregular borders, and structural complexities that are critical for distinguishing between lesion types. In the fourth stage, twelve separable self-attention layers further refine the global feature representations, integrating them into a cohesive feature map that combines local and global information. This global perspective is essential for identifying subtle characteristics indicative of malignant skin lesions, such as melanoma and BCC.

ConvNeXtv2-based block

To enable precise and robust autonomous detection of skin cancer, the initial block of the proposed model is built on the ConvNeXtV2 architecture, a state-of-the-art convolutional framework renowned for its computational

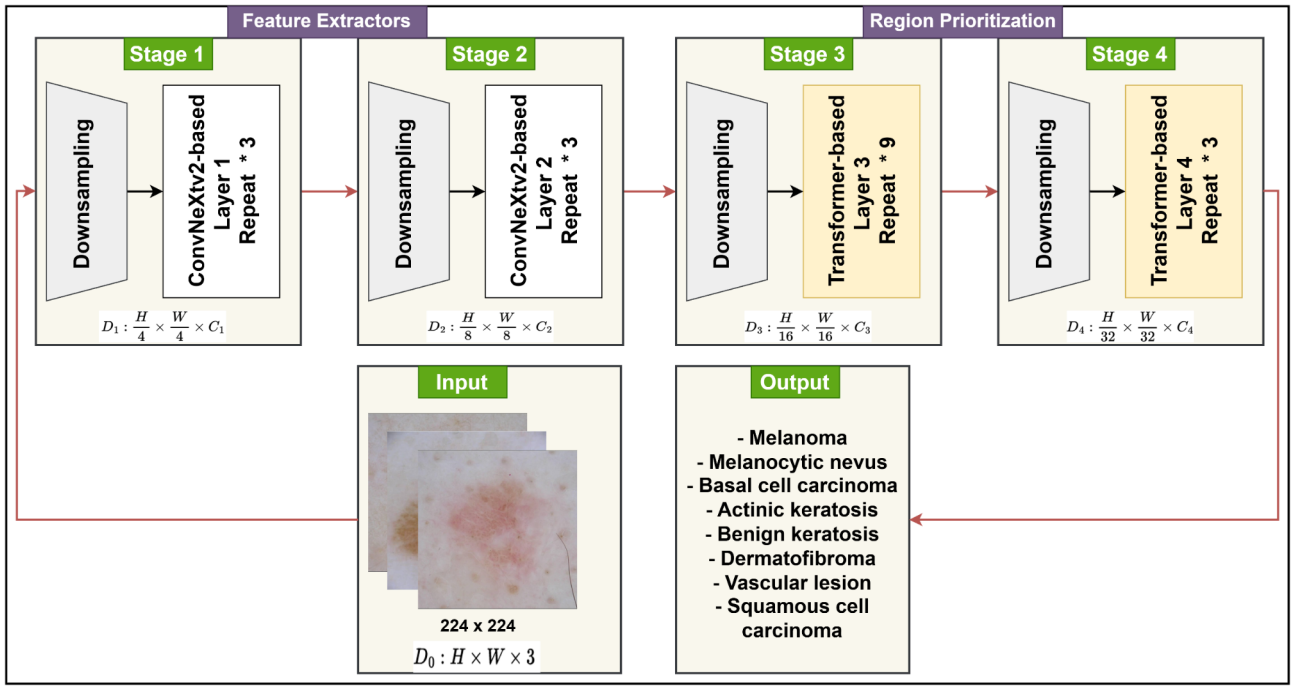

Fig. 2. The architecture of the Proposed Model for automated skin cancer diagnosis.

efficiency and capacity to extract meaningful features. This architecture is particularly suited for analyzing dermoscopic images, a key tool in diagnosing skin cancer. By incorporating ConvNeXtV2 blocks, the model ensures efficient extraction of hierarchical features while maintaining low computational costs, making it highly applicable for real-world medical imaging scenarios. Figure 3 illustrates the ConvNeXtV2-based block of the Proposed Model.

As depicted in Fig. 3, the foundation of the ConvNeXtV2 block lies in the depthwise convolution operation, a computationally efficient technique that processes each channel of the input feature map independently. For a given input feature map $X \in R^{H \times W \times C}$, where H, W, and C represent the height, width, and number of channels, respectively, depthwise convolution is applied separately to each channel:

$$X^c_{\text{out}} = K^c_{\text{depthwise}} * X^c \quad \forall\, c \in [1, C] \tag{1}$$

where X^c_{out} denotes the output for the *c-th* channel and $K^c_{\text{depthwise}}$ is the depthwise convolution kernel. This operation effectively captures localized spatial patterns crucial for identifying intricate features in skin lesions, such as subtle texture irregularities or color gradients, while reducing computational demands.

Following depthwise convolution, Layer Normalization (LN) is applied to stabilize the feature maps by standardizing their mean and variance across dimensions:

$$\widehat{X} = \frac{X - \mu}{\sigma} \tag{2}$$

where μ and σ represent the mean and standard deviation of X. This normalization ensures consistent gradient flow during training and enhances the model's ability to handle batch variability, particularly important when working with high-dimensional dermoscopic image data.

The normalized features are then passed through the GELU activation function, a smoother alternative to standard ReLU that improves gradient flow. This activation function facilitates the learning of non-linear relationships in the data, enabling the model to better capture subtle but diagnostically significant details in skin cancer images, such as asymmetry in lesion borders or variations in pigmentation.

A distinguishing feature of ConvNeXtV2 is its inclusion of Global Response Normalization (GRN), which optimizes inter-channel dependencies by normalizing feature responses across channels. GRN helps to reduce redundancy and ensures that each channel contributes uniquely to the representation. In medical imaging tasks like skin cancer detection, this is particularly beneficial, as it ensures that each channel in the feature map contributes meaningfully to the final prediction. The GRN is defined as:

$$X_{\text{GRN}} = \gamma \cdot \frac{X}{||X||_2} + \beta \tag{3}$$

where $||X||_2$ is the L2-norm of the input X and γ and β are learnable parameters. In the context of skin cancer detection, GRN enhances the model's ability to differentiate critical visual cues, such as irregularities in texture, changes in lesion structure, or atypical patterns indicative of malignancy, improving the diagnostic accuracy for distinguishing between benign and malignant lesions. To preserve essential information and stabilize the learning process, the ConvNeXtV2 block incorporates residual connections, allowing the input to bypass the convolutional and normalization layers and directly contribute to the output:

$$X_{\text{final}} = X_{\text{GRN}} + X_{\text{input}} \tag{4}$$

The inclusion of residual connections within the ConvNeXtV2 block plays a critical role in addressing the vanishing gradient problem, a common challenge in deep learning architectures. These connections allow the input features to bypass convolutional and normalization layers, directly contributing to the final output. This mechanism ensures that essential lower-level features, such as patterns related to lesion symmetry and

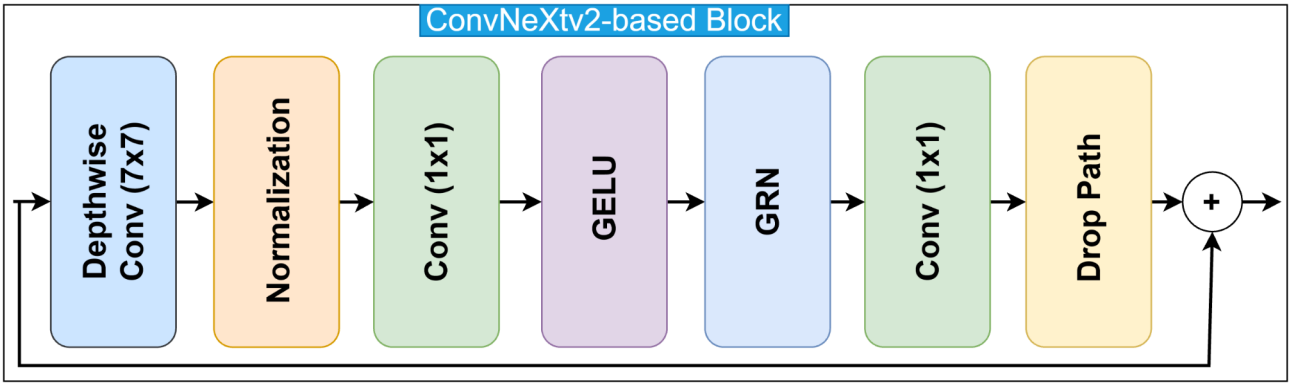

Fig. 3. ConvNeXtV2-based block of the Proposed Model.

edge structure, are preserved alongside higher-level abstract representations, enabling the model to learn a comprehensive hierarchy of features. The ConvNeXtV2 block serves as a foundational component in the proposed model for skin cancer diagnosis, demonstrating its efficacy in processing dermoscopic images. Its ability to extract both localized features, such as fine-grained edge details and texture irregularities, and global patterns, such as structural asymmetry and color variation, makes it particularly well-suited for the analysis of complex dermatological images. This dual capacity enables the model to effectively capture diagnostically significant characteristics, supporting the accurate identification of diverse skin cancer types, including melanoma, BCC, and SCC. By integrating ConvNeXtV2 blocks into the model architecture, a balance is achieved between computational efficiency and high diagnostic accuracy, positioning the model as a practical solution for various real-world applications, such as automated screening in clinical workflows and mobile diagnostic systems for remote healthcare delivery.

Transformer-based block

The proposed second block, based on a transformer architecture, is designed to facilitate a more efficient and accurate identification of contextual patterns in skin cancer detection. Transformer blocks consist of key components that effectively process both local and global contexts. As illustrated in Fig. 2, these blocks sequentially include normalization, separable self-attention, and channel-based Multi Layer Perceptron (MLP) operations. The process begins with the normalization of input features, which enhances model stability and prepares the features for the separable self-attention mechanism. Unlike traditional self-attention, separable self-attention captures global contextual information by utilizing a latent token for each feature instead of computing relationships between all features, thus reducing computational cost and improving efficiency. Following this, the output from the separable self-attention is normalized and passed through the channel-based MLP. The MLP independently processes each channel, applying non-linear transformations to extract richer and more meaningful feature representations. This structure is further reinforced by residual connections, which stabilize learning and maintain gradient flow. The transformer-based blocks in the proposed model, particularly in the 3rd and 4th stages, significantly contribute to autonomous skin cancer detection by enabling the model to learn high-level features and capture complex contextual relationships effectively. Figure 4 illustrates the Transformer-based block of the Proposed Model.

Separable self-attention was introduced in MobileViT-v2 in 2022 to address the high computational demands of traditional multi-headed self-attention (MHA), especially in applications that require real-time processing or deployment on devices with limited resources. The primary aim of this approach was to enhance the efficiency of ViTs in mobile and embedded systems as depicted in Fig. 4. In conventional MHA, attention scores are calculated between every pair of tokens, leading to a computational complexity of $O(k^2)$ where k represents the number of tokens. Separable self-attention, on the other hand, reduces this complexity to $O(k)$ by focusing on a single latent token, thereby making the process more computationally efficient.

Unlike MHA, which calculates pairwise attention scores using dot products between all tokens, separable self-attention introduces a latent token L to capture global context. The attention mechanism in this case computes attention based solely on the interaction between the input tokens and this single latent token, rather than calculating pairwise attention across all tokens. This results in a significant reduction in computational cost. Specifically, the context scores cs are computed as:

$$cs = \text{softmax}\left(xW_I\right) \tag{5}$$

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:

$$cv = \sum_{i=1}^{k} cs\left(i\right) \cdot x_K\left(i\right), \quad cv \in R^d \tag{6}$$

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:

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

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

$$z = cv \odot x_V \tag{8}$$

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 $W_O \in R^{d\times d}$:

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

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

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

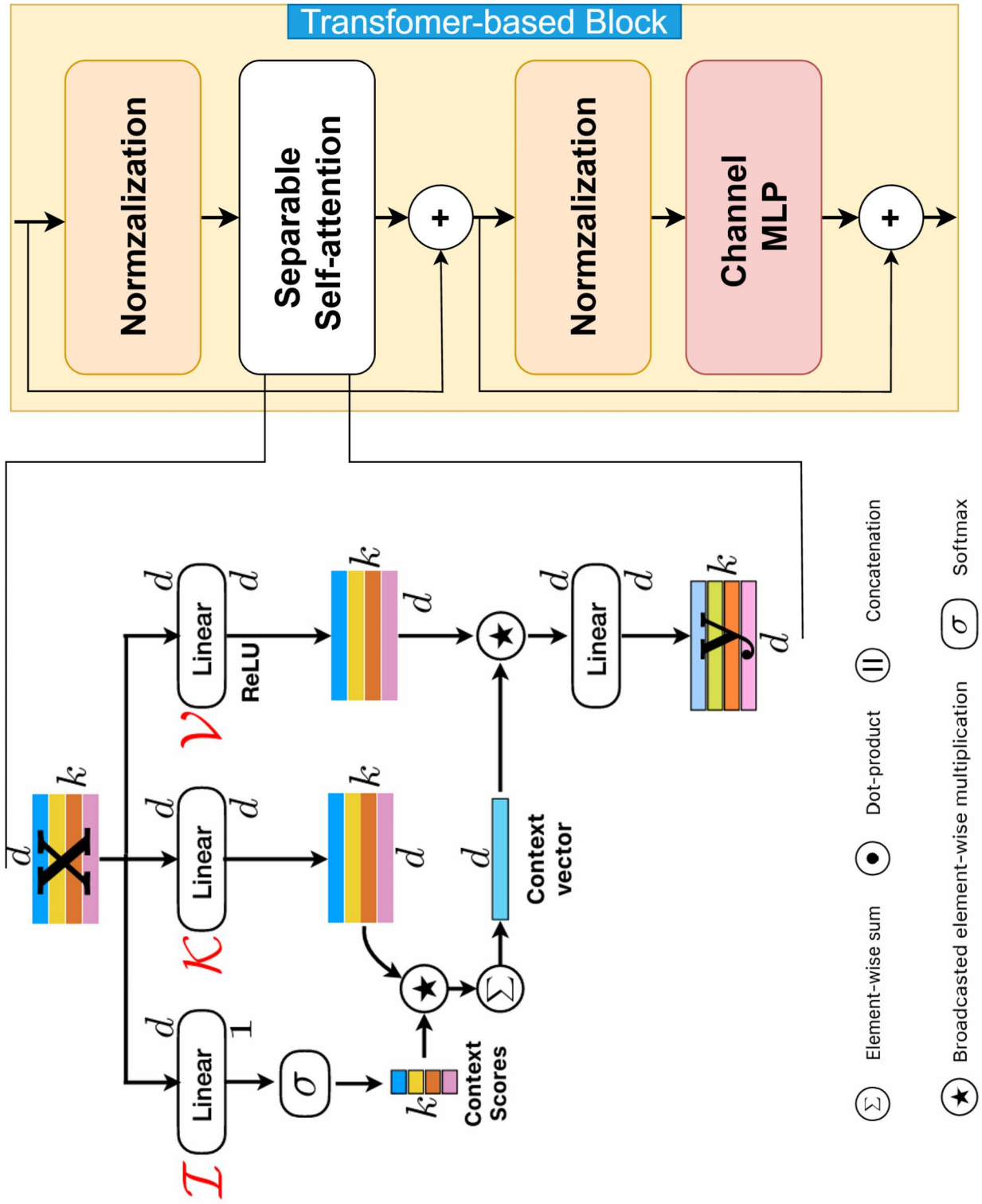

Fig. 4. Transformer-based block of the proposed model.

In the context of skin cancer detection, separable self-attention not only reduces the number of parameters, enhancing computational efficiency, but also improves the model's ability to capture complex relationships between different classes of skin lesions. Unlike traditional attention mechanisms that compute pairwise attention scores for all tokens, separable self-attention uses a single latent token to summarize global contextual information, enabling the model to effectively capture the nuances of skin cancer classes with fewer resources. This mechanism significantly enhances the model's ability to distinguish between malignant and benign lesions, as well as various subtypes of skin cancer, by focusing on global context while maintaining parameter efficiency.

By improving contextual understanding, separable self-attention allows models to make more informed decisions when analyzing dermoscopic images, effectively differentiating subtle visual cues in skin lesions and improving classification accuracy and robustness. The reduction in parameters ensures faster inference times, while better contextual representation leads to more accurate detection of complex lesion types, including melanoma, BCC, and SCC. This approach offers a practical and scalable solution for real-time, mobile-based diagnostic applications, balancing computational efficiency with the ability to capture intricate relationships in data, significantly enhancing the accuracy and speed of skin cancer detection systems in real-world settings.

Results and discussions

This section outlines the experimental framework, detailing the procedures and methodologies utilized to obtain the results. Key components include data preprocessing techniques, data augmentation strategies, the application of transfer learning, evaluation through performance metrics, and a comprehensive analysis of the results. Additionally, comparative assessments of deep learning models are presented to evaluate their effectiveness and reliability.

Experimental setup

All experiments were conducted on a Linux server running Ubuntu 24.04. The deep neural network models were developed and evaluated using a high-performance computing system equipped with a 14th-generation Intel Core i9 processor, an NVIDIA RTX 4090 GPU featuring 24 GB of GDDR6X memory, and 64 GB of DDR5 RAM. The latest stable version of the PyTorch framework, enabled with NVIDIA CUDA, supported the computational processes. Training and testing were carried out within this consistent environment, thereby ensuring identical conditions and parameter settings throughout the study.

Data processing and transfer learning

Effective data preprocessing plays a pivotal role in fine-tuning deep learning models, exerting a direct influence on their overall accuracy and reliability. This preparatory phase encompasses dividing the dataset into distinct training, validation, and test subsets, normalizing inputs, mitigating noise, and addressing outliers. Rather than relying on conventional two-way data splits or cross-validation strategies, this investigation adopts a three-way partitioning scheme. This methodological choice constitutes a notable departure from prevailing norms, facilitating more rigorous validation and enhancing the model's capacity to generalize effectively. By employing a three-way split, this study demonstrates a clear commitment to achieving precise and unbiased performance assessments, thereby distinguishing itself from earlier work. Moreover, the class distribution details of the ISIC 2019 dataset, presented in Table 1, underscore the necessity of meticulous preprocessing to effectively manage inherent variability and class imbalance.

Table 1 details the distribution of 25,331 images from eight lesion categories in the ISIC 2019 dataset, separated into training, validation, and testing subsets. Approximately 70% of the images are allocated to training, with 15% each reserved for validation and testing. This structured division provides a rigorous basis for developing models and evaluating their performance. Notably, class imbalance remains a concern. While the NV class includes 12,875 images, the DF class contains only 239, potentially undermining the model's ability to generalize to underrepresented classes. Techniques such as data augmentation or adjusted class weighting may help mitigate these disparities. By assigning images proportionally for each class across the three subsets, this consistent approach enhances the reliability and reproducibility of experimental findings, thereby informing both academic research and clinical practice.

Data augmentation is an essential methodology for enhancing both the efficacy and generalizability of deep learning models employed in skin cancer classification using the ISIC 2019 dataset. Given the uneven distribution of lesion types, a range of transformations, such as rotation, flipping, scaling, smoothing, mix-up, and color jittering, introduces a richer diversity of training examples. By encompassing a broader spectrum of visual presentations, this approach allows the Proposed Model to better adapt to previously unseen conditions and reduces overfitting. As a result, data augmentation significantly improves model robustness and accuracy, especially when dealing with underrepresented classes.

In parallel, transfer learning offers a strategic approach to overcoming challenges associated with class imbalance and limited data variability. By integrating pre-trained weights derived from large and heterogeneous

Class	Total	Training set (%70)	Validation set (%15)	Test set (%15)
Actinic keratosis (AK)	867	607	130	130
Basal cell carcinoma (BCC)	3323	2326	498	499
Benign keratosis (BKL)	2624	1837	394	393
Dermatofibroma (DF)	239	167	36	36
Melanoma (MEL)	4522	3165	678	679
Melanocytic nevus (NV)	12,875	9012	1931	1932
Squamous cell carcinoma (SCC)	628	440	94	94
Vascular lesion (VASC)	253	177	38	38
Total	25,331	17,731	3,799	3801

Table 1. Image counts for the three subsets in the ISIC 2019 dataset.

datasets like ImageNet, as demonstrated in prior research, the Proposed Model inherits a substantial foundation of visual features. This established knowledge base facilitates efficient adaptation to the ISIC 2019 dataset, accelerating convergence, diminishing computational demands, and enhancing the model's ability to differentiate subtle lesion variations. Together, data augmentation and transfer learning create a synergistic framework that leads to superior accuracy, reliability, and scalability in automated skin cancer detection efforts.

Training procedure

A methodical, stepwise training strategy was employed to optimize performance and uphold scientific rigor. Initially, an extensive array of online data augmentation methods—including scaling, smoothing, mix-up, color jitter, and flipping—was implemented to introduce greater variability into the training data. This approach aimed to improve the model's resilience to the intrinsic variability of skin lesion characteristics and minimize overfitting. To further enhance generalization and accelerate training, transfer learning was applied using ImageNet-pretrained weights, thereby utilizing the rich latent representations derived from a vast corpus of images.

Additionally, a Model Exponential Moving Average (EMA) was utilized to stabilize the learning process and refine the final parameter estimates. This smoothing technique ensured that the resulting model was more robust and consistently high performing. The choice of a 224×224 pixel resolution for input images followed established norms in both dermatological image analysis and the broader computer vision literature, ensuring an equitable basis for comparing various models and architectures.

All models were trained under identical hyperparameter configurations to maintain experimental uniformity. This included a learning rate of 0.01, a base learning rate of 0.1, a momentum value of 0.9, a weight decay of 2.0×10^{-5} and the use of stochastic gradient descent (SGD). The loss function employed was categorical cross-entropy, which is well-suited for multiclass classification tasks and ensures the model effectively learns to predict probabilities for each skin lesion category. Moreover, five warmup epochs, commencing with a learning rate of 1.0×10^{-5}, facilitated a smooth onset of the training dynamics. By rigorously standardizing these aspects of the training procedure, the study ensured reproducibility, comparability, and reliability of the resulting performance evaluations.

Results

This section presents a comprehensive performance analysis of thirty advanced deep learning architectures, including 10 modern CNNs and 20 leading vision transformers, on the ISIC 2019 dataset. In contrast to common practices that focus primarily on validation performance, this study emphasizes robust generalization by rigorously evaluating each model on an independent test set. Such an approach is crucial in clinical contexts like skin cancer detection, where reliable performance on previously unseen data is imperative. All models underwent thorough optimization procedures, including data augmentation, learning rate tuning, and regularization. While the validation phase informed hyperparameter selection and overfitting mitigation, the independent test evaluation ultimately determined each model's capacity to generalize beyond the training distribution, providing a more accurate measure of clinical applicability.

In this study, we utilized over 20 state-of-the-art architectures renowned for their effectiveness in medical image analysis tasks. The selected models include ResNetv2 [59], Res2NeXt[60], DenseNet[61], RexNet[62], MobileNetv3 [63], EfficientNetv2[64], Xception[65], InceptionNext[66], EfficientNet[67], and ConvNexTV2 [37], GhostNetv2 [68], ResMLP[69], PoolFormer, XCiT[70], DeiT[71], Swin[72], BeiTv2 [73], ViT[74], RepViT[75], PvTv2 [76], Tiny-ViT[77], GcViT[78]and PiT[79] These architectures were carefully chosen based on their proven performance in complex image recognition tasks and their suitability for medical applications such as skin lesion classification. Their diverse architectural designs provided a comprehensive basis for evaluating and comparing their capabilities in the context of this study.

The Proposed Model presents a refined deep learning architecture that integrates ConvNeXtV2 blocks with separable self-attention, effectively capturing both fine-grained features and broader contextual cues in dermatological imagery. This approach yields notable gains in accuracy, robustness, and generalization, establishing state-of-the-art performance across key evaluation metrics. Table 2 summarizes these findings, comparing the Proposed Model's outcomes with those of various CNN and ViT-based models on the ISIC 2019 dataset, and underscoring its pronounced efficacy.

Table 2 presents a comprehensive comparison of the Proposed Model's performance against a diverse array of CNN-based and ViT-based architectures on the ISIC 2019 dataset. These architectures, ranging from classic convolutional backbones like ResNetv2, DenseNet, and Xception to cutting-edge transformer models such as DeiT-Base, ViT-Base-Patch16, Swin-Base, and GcViT-Small, collectively provide a broad benchmark for evaluating state-of-the-art classification models in dermatological image analysis.

The Proposed Model achieves an accuracy of 0.9348, a precision of 0.9324, a recall of 0.9070, and an F1-score of 0.9182. These metrics surpass all competing models, including the strongest alternatives. For instance, GcViT-Small, a high-performing ViT-based model, attains a 0.9213 accuracy and a 0.8913 F1-score, while Swinv-Base, another top-tier transformer model, reaches a 0.9179 accuracy. ConvNeXtV2-Base, among the best-performing CNN variants, achieves a 0.9163 accuracy, still trailing behind the Proposed Model's results. Among the CNN-based approaches, ResNetv250 delivers the lowest accuracy (0.8493), indicating that not all convolutional models are equally adept at capturing the nuanced features of dermatological imagery. On the transformer side, XCiT-Small-Patch16 exhibits one of the lowest accuracies (0.8785) within the ViT-based group, suggesting that certain transformer designs may struggle with the complexity or diversity of skin lesions encountered in the ISIC 2019 dataset.

A key factor underlying the Proposed Model's success is its hybrid architecture. In the initial two stages, it employs ConvNeXtV2 blocks, which excel at capturing localized, low-level features pertinent to subtle textural patterns in skin lesions. In the subsequent two stages, the model integrates separable self-attention mechanisms

Model	Accuracy	Precision	Recall	F1-score
ResNetv250	0.8493	0.7809	0.7406	0.7571
Res2NeXt50	0.8798	0.8401	0.8144	0.8255
RexNet200	0.9040	0.8785	0.8596	0.8677
DenseNet121	0.8635	0.8126	0.7798	0.7932
Xception	0.8858	0.8677	0.8094	0.8345
GhostNetv2-100	0.8982	0.8615	0.8337	0.8440
MobileNetv3-large-075	0.8877	0.8442	0.8077	0.8249
EfficientNetv2-small	0.8858	0.8580	0.8148	0.8324
EfficientNet-B4	0.8827	0.8210	0.8035	0.8110
ConvNexTv2-base	0.9163	0.8968	0.8794	0.8873
InceptionNeXt-base	0.8929	0.8616	0.8308	0.8444
ResMLP-24	0.8885	0.8786	0.8171	0.8449
PoolFormer-M36	0.8948	0.8596	0.8231	0.8374
XCiT-Small-Patch16	0.8785	0.8390	0.7921	0.8123
DeiT-base	0.9034	0.8887	0.8359	0.8588
ViT-base-patch16	0.8961	0.8656	0.8336	0.8483
Swinv-base	0.9179	0.9049	0.8757	0.8893
BeiTv2-base	0.9090	0.8775	0.8731	0.8741
MViTv2-base	0.9071	0.8883	0.8670	0.8736
MaxViT-base	0.9084	0.8783	0.8723	0.8737
RepViT-m2	0.9061	0.8792	0.8669	0.8713
PvTv2-B2	0.9029	0.8754	0.8399	0.8532
Tiny-ViT-21 m	0.9082	0.8740	0.8724	0.8720
GcViT-small	0.9213	0.9127	0.8742	0.8913
PiT-base	0.9092	0.8952	0.8456	0.8675
Proposed model	0.9348	0.9324	0.9070	0.9182

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

rather than conventional self-attention. This design choice enhances the extraction of global contextual information while maintaining computational efficiency. By uniting robust convolutional feature extraction with advanced attention-driven representation, the Proposed Model effectively learns both fine-grained details and broader lesion characteristics essential for accurate classification. Notably, the table includes over twenty alternative models, encompassing a wide spectrum of design philosophies and complexity. The consistent use of a standardized input resolution and uniform training hyperparameters ensures that improvements cannot be attributed to discrepancies in experimental setups. Consequently, the superior metrics reported for the Proposed Model underscore its architectural strengths and its capacity to generalize effectively, setting a new benchmark for automated skin lesion classification performance on the ISIC 2019 dataset. Figure 5 illustrates the Proposed Model's performance relative to more than twenty state-of-the-art deep learning architectures by presenting their results in a line graph.

As seen in Fig. 5, The Proposed Model exhibits a clear performance advantage, achieving an accuracy of 0.9348. Among the models tested, GcViT-Small, with an accuracy of 0.9213, most closely approaches this performance level, indicating that certain advanced ViT-based architectures can rival top-tier results. At the lower end of the spectrum, ResNetv250 (0.8493) and XCiT-Small-Patch16 (0.8785) lag behind, suggesting that more CNNs as well as some transformer-based models, may require further refinement to match the leading methods on the ISIC 2019 dataset.

Figure 6 presents the confusion matrix depicting the class-specific performance of the Proposed Model on the ISIC 2019 dataset. Evaluating true positives (TP), false positives (FP), and false negatives (FN) for each lesion category provides a clearer understanding of the model's classification capabilities. Among the classes, VASC stands out with particularly strong performance. The model achieves a true positive count (TP) of 38, accompanied by negligible false negatives (FN = 0) and false positives (FP = 1), indicating an almost flawless distinction of this lesion type. Similarly, the DF class exhibits a promising outcome (TP = 29, FP = 0, FN = 7), suggesting the model's ability to effectively learn and recognize the distinct features of these lesions. Nevertheless, certain classes pose more challenges. melanoma (TP = 589, FP = 59, FN = 90) and BKL (TP = 349, FP = 42, FN = 44) present relatively higher FP and FN values, signaling difficulties in accurately differentiating these categories. Likewise, despite its large representation in the dataset, the NV class (FN = 62, FP = 91) still experiences a notable degree of misclassification. These findings highlight not only the classes in which the model demonstrates exceptional performance but also those requiring further refinement. Elevated FP and FN values for MEL, BKL, and NV indicate a need for targeted improvements in subsequent studies. Strategic approaches, such as enhanced data augmentation, class weighting, and alternative model architectures, may bolster the model's accuracy and enrich its generalization capacity.

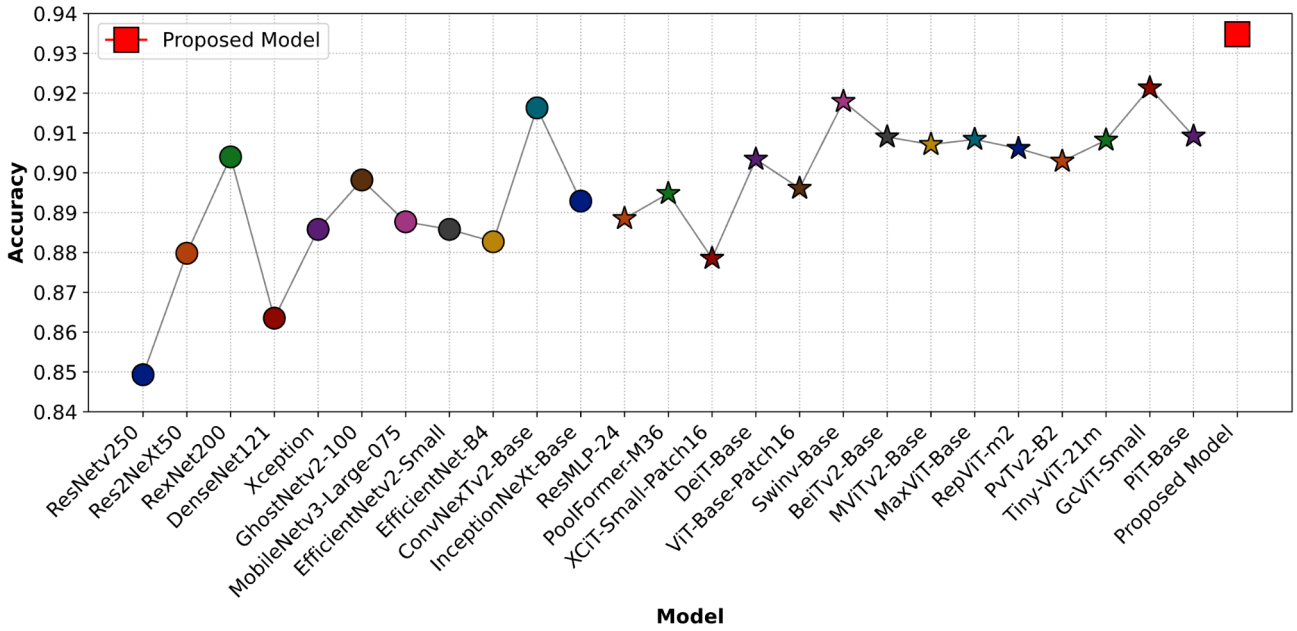

Fig. 5. Accuracy and F1-score metric for all models, including CNN and ViT-based models, and Proposed Model.

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

In this section, we perform ablation studies to assess the individual contributions of ConvNeXtV2 blocks and separable self-attention mechanisms within the Proposed Model. These experiments aim to isolate the effects of each architectural component, providing a detailed analysis of their impact on the model's ability to accurately classify skin lesions. The study involves evaluating variations of the Proposed Model, including configurations that utilize only ConvNeXtV2 blocks or only separable self-attention mechanisms. By comparing these simplified versions against the full model, we analyze how each component influences key performance metrics, including accuracy, precision, recall, and F1-score. This approach enables a comprehensive understanding of the role each architectural innovation plays in improving diagnostic performance, as shown in Table 3.

Table 3 summarizes the results, comparing the performance of various configurations in terms of parameter count, accuracy, precision, recall, and F1-score. These configurations include a baseline model, a model incorporating ConvNeXtV2 blocks in Stages 1 and 2, a model with separable self-attention in Stages 3 and 4, and the full Proposed Model combining both components. The Baseline Model, with 24.30 million parameters, achieved an accuracy of 90.21%, precision of 86.12%, recall of 83.40%, and an F1-score of 84.58%. While these metrics demonstrate reasonable performance, they also highlight the limitations of the baseline configuration in effectively capturing and modeling the complex patterns present in skin lesion images.

The incorporation of ConvNeXtV2 blocks in Stages 1 and 2 resulted in a slight increase in parameters to 26.14 million, leading to a notable improvement in performance. The accuracy increased to 91.19%, precision to 87.19%, recall to 86.27%, and F1-score to 86.52%. This enhancement can be attributed to the ConvNeXtV2 blocks' ability to capture fine-grained local features, particularly in the initial stages of the model, which are critical for distinguishing between visually similar skin lesion types. The substitution of standard self-attention with separable self-attention in Stages 3 and 4 reduced the parameter count to 20.12 million, emphasizing computational efficiency. Despite this reduction, the model achieved an accuracy of 91.05%, precision of 87.37%, recall of 86.47%, and an F1-score of 86.68%. These results underline the effectiveness of separable self-attention in prioritizing diagnostically relevant regions while maintaining performance. Finally, the Proposed Model, integrating ConvNeXtV2 blocks in Stages 1 and 2 and separable self-attention in Stages 3 and 4, demonstrated superior performance with 21.92 million parameters. The model achieved an accuracy of 93.48%, precision of 93.24%, recall of 90.70%, and an F1-score of 91.82%, significantly outperforming the baseline and ablation configurations. This improvement highlights the complementary roles of ConvNeXtV2 and separable self-attention in enhancing feature extraction and global context modeling, respectively. Figure 7 illustrates the comparative performance metrics (Accuracy, Precision, Recall, and F1-Score) of the models evaluated in this study, including the Baseline Model, ConvNeXtV2 (Stages 1 & 2), Separable Self-Attention (Stages 3 & 4), and the Proposed Model.

Discussion

The hybrid model introduced in this study, which integrates ConvNeXtV2 blocks and separable self-attention mechanisms, presents a significant advancement in the domain of skin lesion classification. By combining the localized feature extraction capabilities of ConvNeXtV2 and the efficient global dependency modeling of separable self-attention, the model effectively addresses the challenges of distinguishing between benign and malignant skin lesions. The approach yielded superior performance across critical metrics, including a 93.48%

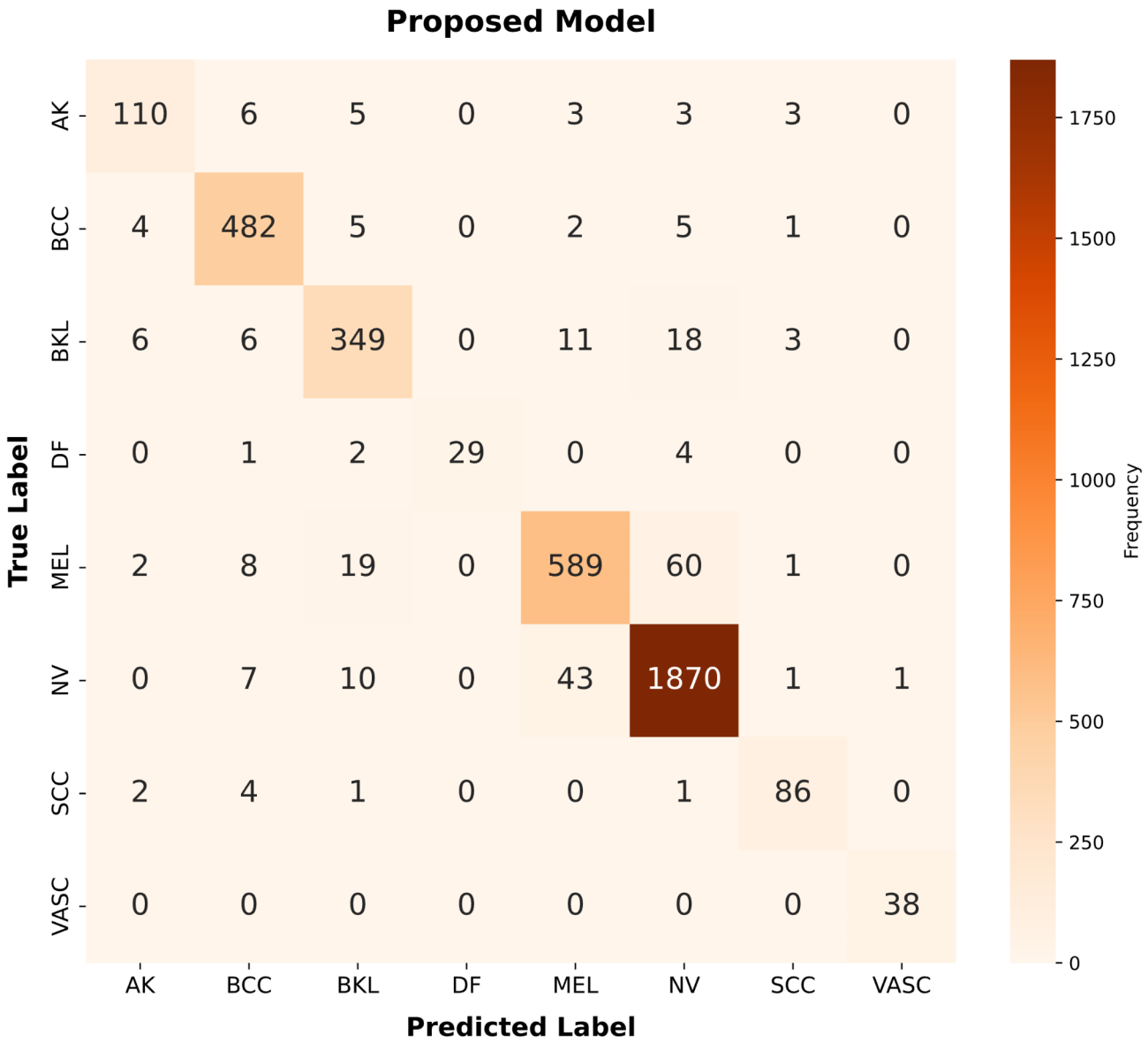

Fig. 6. The confusion matrix showing the class-specific performance of the proposed model.

Model	Params(M)	Accuracy	Precision	Recall	F1-score
Baseline model	24.30	0.9021	0.8612	0.8340	08458
ConvNeXtv2-base (in stage 1 and stage 2)	26.14	0.9119	0.8719	0.8627	0.8652
Separable self-attention (in stage 3 and stage 4)	20.12	0.9105	08737	0.8647	0.8668
Proposed model (ConvNeXtV2 + separable self-attention)	21.92	0.9348	0.9324	0.9070	0.9182

Table 3. Ablation study on the ımpact of each block in the proposed model using the ISIC 2019 dataset.

accuracy, 93.24% precision, 90.70% recall, and a 91.82% F1-score. These outcomes demonstrate the robustness and reliability of the hybrid architecture, surpassing over 20 state-of-the-art deep learning models evaluated under comparable conditions.

A key strength of the model lies in its computational efficiency, achieved with a parameter count of just 21.92 million. This compact design makes it an attractive option for real-time clinical applications, including resource-constrained environments and mobile healthcare platforms. The architecture's ability to balance high accuracy with computational efficiency positions it as a promising tool for automating skin cancer diagnosis in diverse settings. Additionally, the adoption of advanced preprocessing techniques, such as data augmentation

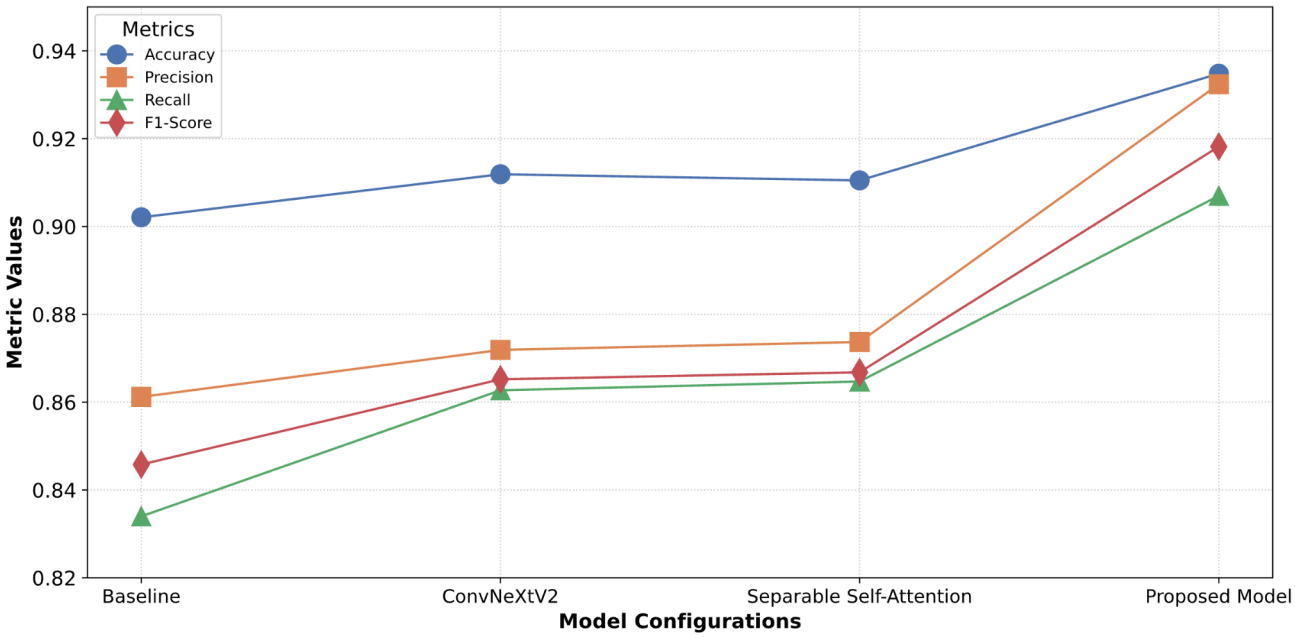

Fig. 7. Comparative analysis of performance metrics across model configurations.

and transfer learning, mitigated the limitations posed by the imbalanced ISIC 2019 dataset. These strategies not only enhanced the model's generalizability but also improved its robustness across the eight lesion categories.

A detailed examination of the model's performance across different lesion classes, as depicted in the confusion matrix, revealed notable strengths and areas for improvement. While the model exhibited exceptional accuracy in identifying underrepresented categories, such as VASC and DF, it faced challenges in classifying categories with high visual similarity, such as melanoma and BKL. These findings suggest that further refinement of the model is required to enhance its discriminative capabilities for complex lesion types. Employing class-specific augmentation, adaptive loss functions, or ensemble learning could provide avenues to address these challenges.

The ablation studies conducted as part of this research highlight the complementary contributions of ConvNeXtV2 and separable self-attention mechanisms. ConvNeXtV2 demonstrated its efficacy in the initial stages of the model by capturing intricate local features essential for differentiating subtle lesion characteristics. In contrast, separable self-attention mechanisms, integrated into the later stages, effectively aggregated global contextual information while maintaining computational efficiency. This synergy underscores the value of combining convolutional and attention-based mechanisms to achieve a robust and adaptive diagnostic framework.

Beyond its quantitative performance, the model's practical implications merit attention. The hybrid architecture's efficiency and reliability make it a viable candidate for integration into telemedicine systems and clinical decision support tools. By enabling automated, real-time analysis of dermoscopic images, the model has the potential to expand access to accurate skin cancer diagnosis, particularly in resource-limited healthcare settings.

Limitations and future directions

While the proposed hybrid model has demonstrated exceptional accuracy and computational efficiency, several limitations should be acknowledged, and future research directions must be explored to maximize its clinical applicability and robustness. One significant limitation of this study is the reliance on the ISIC 2019 dataset, which, despite its comprehensiveness, may not fully capture the variability present in real-world clinical settings. Differences in imaging technologies, lighting conditions, and patient populations can lead to potential discrepancies in model performance when deployed outside the controlled experimental environment. Future research should aim to validate the model using datasets from diverse clinical environments, ensuring that it generalizes well across varied contexts. Domain adaptation techniques could also be explored to further enhance the model's adaptability.

Another challenge lies in addressing the class imbalance within the dataset. Although data augmentation and transfer learning were employed to mitigate this issue, misclassification rates remain higher for certain categories, such as melanoma and BKL, which exhibit high visual similarity. To overcome this, future studies should investigate advanced techniques such as generative adversarial networks (GANs) for synthetic data generation or adaptive loss functions that place greater emphasis on underrepresented classes.

The absence of explainability features is another notable limitation. In clinical practice, the interpretability of AI models is essential for building trust among healthcare professionals and ensuring ethical implementation. Incorporating explainable AI (XAI) techniques, such as saliency maps or attention heatmaps, could make

the model's predictions more transparent, thereby facilitating its integration into diagnostic workflows and increasing clinician confidence in its recommendations.

While the proposed model is computationally efficient compared to other state-of-the-art architectures, its performance on resource-constrained devices such as older mobile phones or embedded systems has not been extensively evaluated. Future efforts should focus on optimizing the model using techniques like pruning, quantization, or knowledge distillation to reduce its computational footprint, enabling broader deployment in low-resource environments or portable diagnostic tools.

The current focus of the model on classification tasks limits its utility in providing comprehensive lesion analysis. Expanding its capabilities to include lesion segmentation or localization could significantly enhance its clinical applicability. A multi-task framework that integrates these functionalities could provide a holistic diagnostic solution. Additionally, exploring its utility with other diagnostic modalities, such as histopathology or advanced imaging techniques, could extend its versatility and accuracy.

Real-time application of the model in telemedicine and mobile healthcare systems represents another promising direction. Evaluating its performance in live diagnostic scenarios, especially in remote or underserved regions, could demonstrate its practical value in improving healthcare accessibility and outcomes. Such applications would benefit from further validation to ensure reliability under real-world conditions.

Lastly, ethical and regulatory considerations should be an integral part of future research efforts. Ensuring data privacy, addressing potential biases, and complying with medical device regulations are essential for the responsible deployment of AI in healthcare. Establishing clear guidelines for the clinical adoption of such technologies will be crucial to bridging the gap between research and practical implementation.

Conclusion

This study tackles the critical challenge of early and accurate skin cancer diagnosis, a significant global health concern where timely detection plays a crucial role in improving treatment outcomes and patient survival rates. The inherent visual similarities between benign and malignant lesions present substantial classification difficulties. To address these challenges, an innovative hybrid deep learning model is proposed, integrating ConvNeXtV2 blocks and separable self-attention mechanisms to enhance feature extraction and classification performance. ConvNeXtV2 blocks, employed in the initial stages, are designed to effectively capture intricate local features and subtle patterns, which are essential for distinguishing between visually similar lesion types. In the later stages, separable self-attention is utilized to prioritize diagnostically significant areas while reducing computational complexity, overcoming the inefficiencies of traditional self-attention mechanisms.

The model underwent rigorous training and validation on the ISIC 2019 dataset, comprising eight distinct skin lesion categories, using advanced data augmentation and transfer learning techniques to ensure robustness and reliability. The architecture demonstrated outstanding performance, achieving 93.48% accuracy, 93.24% precision, 90.70% recall, and a 91.82% F1-score. These results surpass those of more than 20 state-of-the-art deep learning models, including CNN-based and ViT-based architectures, under standardized experimental conditions.

Despite its exceptional performance, the model is computationally efficient, with only 21.92 million parameters, making it highly suitable for deployment in real-time and mobile applications. By addressing the key challenges of feature extraction, computational efficiency, and classification accuracy, the Proposed Model sets a new standard for reliable and scalable skin cancer diagnosis, offering significant potential for clinical implementation and improved patient outcomes.

Data availability

The datasets generated and/or analyzed during the current study are publicly available in the [https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-images-for-classification] repository, [https://challenge.isic-archive.com/data/#2019]

Received: 5 October 2024; Accepted: 4 February 2025
Published online: 10 February 2025

References

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. & 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).
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/s10489-022-04299-1 (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).
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* 10691–10700 (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).

Author contributions

IP.: Conceptualization, Methodology, Software, Investigation, Data curation, Validation, Supervision, Writing – review & editing. BO.: Conceptualization, Data curation, Investigation, Reviewing, Validation, Writing – review & editing.

Funding

This study was funded by Alfaisal University, which supports research initiatives aimed at advancing knowledge and innovation in alignment with its commitment to academic excellence.

Declarations

Competing interests

The authors declare no competing interests.

Consent to participate

No formal consent to participate was required for this work as it did not involve interactions with human subjects or the collection of sensitive personal information.

Consent to publish

This study did not use individual person's data.

Ethics approval

No ethics approval was required for this work as it did not involve human subjects, animals, or sensitive data that would necessitate ethical review.

Additional information

Correspondence and requests for materials should be addressed to B.O.