Human-Centric Intelligent Systems
https://doi.org/10.1007/s44230-024-00083-1

RESEARCH ARTICLE

RvXmBlendNet: A Multi-architecture Hybrid Model for Improved Skin Cancer Detection

Farida Siddiqi Prity[1,2] · Ahmed Jabid Hasan[1] · Md Mehedi Hassan Anik[1] · Rakib Hossain[1] · Md. Maruf Hossain[1,3] · Sazzad Hossain Bhuiyan[1] · Md. Ariful Islam[1,4] · Md Tousif Hasan Lavlu[1]

Received: 20 May 2024 / Accepted: 29 August 2024

Abstract
Skin cancer, one of the most dangerous cancers, poses a significant global threat. While early detection can substantially improve survival rates, traditional dermatologists often face challenges in accurate diagnosis, leading to delays in treatment and avoidable fatalities. Deep learning models like CNN and transfer learning have enhanced diagnosis from dermoscopic images, providing precise and timely detection. However, despite the progress made with hybrid models, many existing approaches still face challenges, such as limited generalization across diverse datasets, vulnerability to overfitting, and difficulty in capturing complex patterns. As a result, there is a growing need for more robust and effective hybrid models that integrate multiple architectures and advanced mechanisms to address these challenges. Therefore, this study aims to introduce a novel multi-architecture hybrid deep learning model called "RvXmBlendNet," which combines the strengths of four individual models: ResNet50 (R), VGG19 (v), Xception (X), and MobileNet (m), followed by "BlendNet" to signify their fusion into a unified architecture. The integration of these models is achieved through a synergistic combination of architectures, incorporating self-attention mechanisms using attention layers and adaptive content blocks. This study used the HAM10000 dataset to refine dermoscopic image preprocessing and enhance deep learning model accuracy. Techniques like OpenCV-based hair removal, min–max scaling, and adaptive histogram equalization were employed to improve image quality and feature extraction. A comparative study between the proposed hybrid "RvXmBlendNet" and individual models (CNN, ResNet50, VGG19, Xception, and MobileNet) demonstrated that "RvXmBlendNet" achieved the highest accuracy of 98.26%, surpassing other models. These results suggest that the system can facilitate earlier interventions, improve patient outcomes, and potentially lower healthcare costs by reducing the need for invasive diagnostic procedures.

Keywords Skin cancer · Dermatology · Deep learning · Convolutional neural networks · Hybrid models

1 Introduction

The skin, acting as the body's formidable shield, defends against ultraviolet (UV) radiation, toxins, and infections. It comprises layers, including the epidermis, dermis, and hypodermis. Melanin, the pigment in the skin, offers protection against UV radiation [1]. Skin cancer occurs when abnormal cells grow uncontrollably, potentially invading or metastasizing within the body. UV radiation can inflict damage on the DNA within skin cells, leading to mutations that may severely impair the cells' proper functioning. While melanin helps protect against UV radiation, high or prolonged exposure to UV can overwhelm this protection, potentially leading to DNA damage and the development of skin cancer.

Skin cancer primarily manifests in three forms: basal cell carcinoma (BCC), squamous cell carcinoma (SCC),

✉ Farida Siddiqi Prity
faridasiddiqiprity@gmail.com

[1] Department of Computer Science and Engineering, Shanto-Mariam University of Creative Technology, Uttara 17, Dhaka 1230, Bangladesh

[2] Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

[3] Department of Information and Communication Engineering, Pabna University of Science and Technology, Pabna 6600, Bangladesh

[4] Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

Published online: 09 September 2024

and melanoma. BCC and SCC are typically less perilous and seldom result in mortality. In stark contrast, melanoma stands as the most lethal type of skin cancer, predominantly caused by exposure to ultraviolet radiation from the sun. People with fair skin are at a greater risk of developing skin cancer than those with darker complexions, as they possess lower levels of melanin, the pigment that offers protection [2].

Melanoma originates from melanin-producing cells and is typically found on the skin, eyes, nerve centres, and meninges. Early detection of melanoma is vital, as it greatly enhances the likelihood of effective treatment, despite melanoma being the fastest-growing type of skin cancer in terms of incidence rate. In 2018, there were 287,723 new cases of melanoma worldwide, resulting in 60,712 deaths [3]. In 2019, in the United States alone, there were approximately 96,480 cases of melanoma, with around 7230 deaths attributed to the disease [4]. Although melanoma has a high risk of spreading, early detection and treatment can significantly improve survival rates [5].

Early detection of skin cancer plays a critical role in raising awareness, ensuring timely diagnosis, improving prognosis, and reducing morbidity and mortality rates. This approach enables more personalized treatment strategies, ultimately improving patient outcomes and lessening the overall disease burden [6]. Early detection becomes even more critical as most melanoma cases are diagnosed at later stages, leading to higher treatment costs and poorer prognosis, emphasizing the need for early detection efforts [7, 8]. Therefore, early detection emerges as a critical aspect deserving attention. Dermatologists often encounter difficulties in accurately diagnosing skin lesions due to the diverse presentation of lesions, limited access to specialized equipment, and the necessity for continuous training to stay updated with advancing diagnostic techniques [9, 10]. In recent years, the utilization of Deep learning in medical image processing has increased due to their robust feature representation capabilities. Combining deep learning with image processing for skin cancer detection is an emerging field of research [11].

Baldrick et al. [12] conducted a study to compare expert opinions with Deep Learning methods in categorizing skin lesions. The Deep Learning-based computer program they used demonstrated a sensitivity of 95% and a specificity of 88%, closely aligning with the performance of dermatologists, who had a sensitivity of 95% and a specificity of 90%. These findings suggest that automated systems based on Deep Learning could be valuable tools in skin cancer detection. Additionally, research points to Deep Learning models demonstrating exceptional proficiency in classifying various skin lesions, sometimes matching or even exceeding the accuracy of professional dermatologists. In several instances, Deep Learning-based models have outperformed expert dermatologists in sensitivity and specificity [13]. This emphasizes the substantial role that automated systems could assume in aiding dermatologists, particularly in the early detection of skin lesions and in mitigating diagnostic inaccuracies. Such advancements have the potential to enhance patient outcomes and significantly reduce the healthcare expenses associated with treating advanced-stage cancer.

Deep learning algorithms are primarily classified into three principal types: Convolutional Neural Networks, single transfer learning, and hybrid models. CNNs are a cornerstone in deep learning, intricately designed for image recognition and classification tasks. They consist of successive layers, including convolutional, pooling, and fully connected layers, which facilitate the extraction of hierarchical features directly from unprocessed data. Single transfer learning, in contrast, employs pre-trained models—typically trained on extensive datasets such as ImageNet—which are then refined and adapted for specific tasks through fine-tuning. This approach significantly reduces training time and computational resources while still achieving competitive performance. Lastly, hybrid models represent a synthesis of CNN and transfer learning techniques or the combination of different transfer learning methods. These hybrid architectures aim to capitalize on the strengths of both approaches, leveraging the feature extraction capabilities of CNNs while benefiting from the knowledge transfer and adaptability of transfer learning. By combining these methodologies, hybrid models offer enhanced flexibility, scalability, and performance across a wide range of deep learning tasks.

Numerous authors have traditionally depended on CNN and single transfer learning methodologies for skin cancer diagnosis, yet these approaches encounter significant limitations. These constraints include restricted generalization capabilities across diverse datasets, susceptibility to overfitting, and struggles in capturing intricate patterns. Consequently, a burgeoning necessity has emerged for the development of hybrid deep-learning models capable of surmounting these challenges. Such models integrate multiple architectures and advanced mechanisms to enhance performance and robustness. By amalgamating diverse architectures and incorporating sophisticated techniques, hybrid models hold the potential to overcome the limitations encountered by single-method approaches. The transition toward hybrid models signifies a seminal advancement in the realm of deep learning, offering enhanced adaptability and efficacy across a diverse array of applications.

The principal aim of this study is to:

- Introduce the novel hybrid deep learning model "RvXm-BlendNet" integrating ResNet50, VGG19, Xception, and MobileNet architectures into a unified framework.

- Incorporate advanced self-attention mechanisms using attention layers and adaptive content blocks to enhance feature prioritization and extraction in skin cancer detection.
- Refine the preprocessing pipeline for dermoscopic images using techniques such as OpenCV-based hair removal, min–max scaling, and adaptive histogram equalization to optimize feature extraction accuracy.
- Conduct a comprehensive comparative analysis against individual models (ResNet50, VGG19, Xception, MobileNet) and traditional CNN to demonstrate the superior performance of the hybrid approach in terms of accuracy, sensitivity, and recall.
- Strive to enhance the early detection and treatment of skin cancer, offering valuable insights for researchers and practitioners, particularly in resource-limited environments.

Although automated tools show high accuracy, they cannot replace dermatologists, but they can be a significant aid, especially in areas with limited medical resources or healthcare professionals.

2 Related Works

Medical classification plays a crucial role in aiding medical diagnosis, with recent advancements in deep learning, transfer learning, and hybrid learning enhancing classification performance in the Internet of Medical Things (IoMT). Deep learning has emerged as a rapidly growing domain within AI, finding applications across various fields [14–16]. It has become a prominent deep-learning model due to its success in object recognition, making it a focal point in the classification of medical images [17, 18]. Despite the prevalence of machine learning algorithms in most skin cancer detection studies, deep learning has been employed in several investigations targeting skin lesion categorization [19–21]. Skin cancer detection has seen the application of various deep learning algorithms, including customized CNNs, single transfer learning, and hybrid models. Different authors have utilized these algorithms in skin cancer detection and gained valuable insights into their effectiveness.

2.1 Skin Cancer Detection using Customized CNN Model

A customized CNN refers to a Convolutional Neural Network architecture that is meticulously engineered or adapted for a specific task or dataset. Unlike conventional CNN models, a customized CNN is either constructed from the ground up or modified to address the distinct requirements of a given problem. This tailored approach has been extensively applied in the realm of skin cancer detection. For instance, a recent investigation [22] saw dermatology researchers develop a bespoke CNN architecture to analyze 1550 images of potentially malignant lesions, captured using both smartphones and a digital single-lens reflex camera (DSLR) [23]. The performance of the algorithm was rigorously compared against clinical evaluations and histopathological diagnoses, revealing significant advancements and outperforming traditional diagnostic methodologies. Furthermore, Ali et al. designed a Deep Convolutional Neural Network (DCNN) model for skin lesion classification, attaining impressive accuracy rates of 93.16% in training and 91.93% in testing on the HAM10000 dataset [24]. Lequan et al. [25] introduced a highly sophisticated Convolutional Neural Network for melanoma detection, employing a fully convolutional residual network (FCRN) consisting of 16 residual blocks during the segmentation process to enhance performance. Their approach achieved an impressive 85.5% accuracy in melanoma classification with segmentation.

2.2 Skin Cancer Detection using Single Transfer Learning Model

In skin cancer detection, utilizing a single transfer learning model involves leveraging a pre-trained deep learning architecture—often trained on extensive datasets such as ImageNet—and subsequently fine-tuning it for specific tasks related to skin cancer detection. Transfer learning facilitates the transfer of knowledge acquired from one domain to another, thereby expediting the training process and potentially enhancing performance, particularly in scenarios with limited data. This methodology has become increasingly prevalent in skin cancer detection due to its proficiency in extracting pertinent features from dermoscopic images. By fine-tuning pre-trained models such as VGG, ResNet, or Inception, researchers can recalibrate these models to identify distinctive patterns indicative of skin cancer. This adaptation supports dermatologists in achieving precise diagnoses and facilitating early intervention. The widespread adoption of this approach underscores its effectiveness and acceptance within the field. Anand et al. created a VGG19-based deep learning model for early skin cancer detection, with an accuracy of 89.09%, helping dermatologists make timely diagnoses [26]. Rashid et al. [27] used MobileNetV2 for melanoma classification, achieving 98.2% accuracy on ISIC 2020, outperforming existing methods. Fraiwan et al. [28] explored the deep transfer learning model VGG16 for skin lesion categorization, achieving 82.9% accuracy on HAM1000, with notable F1 scores for pair-wise classifications. Balaha et al. [29] introduced a novel method for skin cancer detection using SpaSA optimizer and U-Net models, achieving 85.87% accuracy with MobileNetV2

on the "Skin Diseases Image" dataset. DeVries [30] introduced a multi-scale transfer learning model using Inception v3, fine-tuning it on two resolution scales of lesion images by capturing shape and textual detail for skin cancer classification. Mahbod et al. [31] extracted deep features from pre-trained models, training a multi-class SVM classifier and achieving 97.55% AUC for melanoma classification. Another transfer learning model based on pre-trained ResNet-152 achieved a 0.99 AUC for 12 skin lesion types [32]. Dorj et al. [33] utilized AlexNet for feature extraction and SVM for classification, obtaining high scores in sensitivity, specificity, and accuracy for SCC, AK, and BCC. Kalouche [34] utilized a fine-tuned VGG-16 model, achieving 78% accuracy in melanoma classification. Additionally, a transfer learning model was proposed to detect skin lesion borders, achieving 86.67% accuracy in classifying normal and lesion images.

Vision Transformers (ViTs) have emerged as powerful tools in skin cancer detection, leveraging self-attention mechanisms to enhance feature extraction and classification tasks. Abbas et al. [35] developed an Assist-Dermo system using a separable vision transformer (SVT) to classify nine classes of pigmented skin lesions (PSLs). Xin et al. [36] developed SkinTrans, an improved transformer network named to enhance skin cancer classification using Vision Transformers. Their approach employs self-attention to enhance important features and suppress noise-causing features. SkinTrans integrates multi-scale patch embedding and contrastive learning to emphasize multi-scale features and optimize data encoding for effective classification. The study validated SkinTrans on HAM10000 and a clinical dataset, achieving 94.3% and 94.1% accuracy, respectively. These results demonstrate SkinTrans' efficiency in skin cancer classification, leveraging its robust performance in both natural language and vision tasks. Aladhadh et al. [37] devised a two-tier framework for precise skin cancer classification. Initially, they augmented data samples for robust training. They then introduced a Medical Vision Transformer (MVT)-based model, segmenting input images into patches and processing them sequentially through a transformer structure akin to word embedding. Their experiments on the HAM10000 dataset showcased superior performance compared to existing methods, highlighting the effectiveness of MVTs in enhancing diagnostic accuracy in dermatology. Desale et al. [38] introduced an optimized vision transformer method for skin tumour classification. They employed advanced pre-processing techniques and segmentation algorithms, followed by feature extraction using a hybrid approach. Their model achieved outstanding performance metrics on the ISIC 2019 database, surpassing existing methodologies in skin cancer classification.

2.3 Skin Cancer Detection Using Hybrid Deep Learning Model

Al-Rasheed et al. proposed an automated method for classifying skin cancers, achieving 93.5% accuracy with a combination of VGG16, ResNet50, and ResNet101 [39]. In [40], a range of pre-trained transfer learning models were assessed on the HAM10000 dataset, encompassing seven skin lesion types. ResNet50 was determined to be the most effective model, delivering the highest accuracy. To further refine performance, the study employed a hybrid model integrating ResNet50, VGG16, and DenseNet, culminating in an accuracy of 84%. Subsequently, Chaturvedi et al. [41] applied transfer learning for multi-class skin cancer detection on the same dataset, achieving a peak accuracy of 93.20% with the ResNeXt101 model. Notably, their hybrid model surpassed this, attaining an accuracy of 92.83%, outperforming leading models. Meanwhile, Elshahawy et al. [42] introduced a novel model that combines YOLOv5 and ResNet50 for melanoma detection, utilizing 10,000 training images from the HAM10000 dataset. This approach achieved a real-time processing speed of 0.4 ms per image for non-maximum suppression (NMS). Maniraj et al. [43] propose a hybrid deep learning approach using 3D wavelet subband fusion for non-invasive skin image inspection. The method includes median filtering to remove noise, 3D wavelet transforms for textural information extraction, and multiclass classification, achieving 99.33% average accuracy on PH2 database images, with over 90% sensitivity and specificity for normal, benign, and malignant skin image discrimination. Nagaraj et al. [44] explore skin cancer detection and control techniques utilizing hybrid deep learning methods. They combine CNNs with classical machine learning models. Their results show the hybrid model achieved an accuracy of 94.5%, precision of 93.7%, recall of 95.2%, and F1-Score of 94.4%.

3 Research Gap

While these studies demonstrate promising advancements in using deep learning and other techniques for skin cancer detection, they each have specific areas for improvement. Most of the authors [22–25] use customized CNN models with less data. A common observation across these studies is the prevalent use of customized CNN models, often trained on relatively limited datasets. The disadvantage of training CNN models with less data is the risk of overfitting, where the model learns to memorize the training examples rather than generalize patterns from the data. This limitation can lead to reduced performance and generalizability when applied to unseen data, potentially hindering the model's effectiveness in real-world scenarios. Many researchers [26–34] have utilized single-transfer learning models in

skin cancer detection. However, relying solely on a single algorithm for transfer learning may limit the model's adaptability and performance across diverse datasets and tasks. Additionally, single-algorithm approaches may overlook the nuances and complexities inherent in different skin cancer datasets, potentially leading to suboptimal performance and generalization ability. Vision transformers, while effective in capturing global dependencies through self-attention, can be computationally expensive and require large datasets for optimal performance. Consequently, combining multiple algorithms could lead to more robust and accurate skin cancer detection models. Though many authors [39–44] have employed hybrid methods, the current literature on hybrid deep learning models for skin cancer detection often restricts hybridization to integrating two to four methods, overlooking advanced techniques such as self-attention mechanisms using attention layers and adaptive content blocks. These techniques are crucial for enhancing feature extraction and model performance by effectively capturing intricate patterns within dermatoscopic images. Furthermore, existing studies predominantly focus on specific methodologies without comprehensive comparative evaluations against other approaches. This approach limits the exploration of potential synergies among different deep-learning architectures and their combined benefits in skin cancer detection. Thus, there remains a critical gap in the literature concerning the exploration and empirical validation of hybrid models that incorporate advanced attention mechanisms and comprehensive comparative analyses across diverse algorithms and performance metrics.

3.1 Our Contribution

Our study aims to significantly advance state-of-the-art skin cancer detection and classification by addressing critical gaps identified in existing methodologies. We introduce a novel hybrid deep learning model named "RvXmBlendNet," which integrates ResNet50, VGG19, Xception, and MobileNet architectures into a unified framework denoted by "BlendNet." This integration is not merely a combination of well-known models but strategically leverages their complementary strengths to enhance performance beyond what each model achieves individually. The main aspects of our contribution are:

Novelty and Innovation: Unlike existing approaches [22–34] that often rely on single-model architectures or simplistic combinations, our hybrid model, "RvXmBlendNet," represents a paradigm shift in skin cancer detection. It integrates not only ResNet50, VGG19, Xception, and MobileNet architectures but also incorporates advanced self-attention mechanisms using attention layers and adaptive content blocks. These sophisticated mechanisms, typically utilized in transformer architectures like Vision Transformers [35–38], have demonstrated success in various studies of skin cancer detection. In contrast, the hybridization of individual deep learning models often neglects the integration of self-attention mechanisms using attention layers. Additionally, we introduce adaptive content blocks in the hybridization process of "RvXmBlendNet". These enhancements are meticulously designed to extract and prioritize salient features within dermatoscopic images, capturing intricate patterns and subtle details crucial for accurate diagnosis. By surpassing the capabilities of traditional deep-learning models, our approach significantly enhances diagnostic accuracy and generalizability across diverse skin lesion types. This addition clarifies how the use of self-attention mechanisms and adaptive content blocks in "RvXmBlendNet" distinguishes it from previous approaches, particularly those that rely solely on individual deep learning models without such advanced architectural components.

Dataset and Methodological Rigor: Leveraging a comprehensive dataset of 10,015 skin cancer images from the HAM10000 dataset across seven lesion classes, our study addresses class imbalance using Random Oversampling and employs rigorous dropout and early stopping techniques to mitigate overfitting during model training. This methodological rigour ensures that our results are robust and applicable in real-world clinical settings.

Performance Evaluation: We conduct an exhaustive comparative analysis against not only individual models (ResNet50, VGG19, Xception, and MobileNet) but also traditional CNN approaches used in skin cancer detection. Our evaluation criteria encompass a broad spectrum of performance metrics to demonstrate the superior efficacy of "RvXmBlendNet" in detecting and classifying skin lesions accurately.

While deep learning indeed offers powerful end-to-end processing capabilities, the integration of extensive image preprocessing techniques in our study serves to enhance the robustness and reliability of our hybrid model, "RvXmBlendNet." Previous studies [22–38] have demonstrated the success of deep learning techniques with image preprocessing in improving skin cancer detection. The preprocessing steps employed in this study are pivotal in refining the quality of input data and are crucial for accurate feature extraction and classification. By optimizing image clarity and consistency, our approach ensures that "RvXmBlendNet" can effectively leverage deep learning's strengths while mitigating potential noise and variability inherent in medical imaging datasets. This dual strategy of preprocessing and deep learning integration not only enhances model performance but also underscores our commitment to delivering reliable outcomes in clinical applications of skin cancer detection. By emphasizing these contributions, our study positions "RvXmBlendNet" as an innovative solution advancing skin

cancer detection, offering enhanced diagnostic capabilities critical for reliable clinical decision support systems.

4 Materials and Methods

This methodology encompasses various steps to preprocess the input images, address the class imbalance, train different models (customized CNN, single transfer learning, hybrid deep learning), and comparative analysis for skin cancer classification. Each step enhances the model's efficacy and resilience in detecting various skin lesion types. The workflow of the proposed model is delineated in Fig. 1.

4.1 Dataset Description

The dataset employed in our research is the HAM10000 dataset, a vast repository of dermatoscopic images specifically curated to facilitate research in the diagnosis and classification of skin lesions [45]. The HAM10000 dataset has proven to be a valuable resource in skin cancer detection and classification, with numerous studies leveraging deep learning [46–48] and transfer learning methodologies. This dataset overcomes limitations in previous datasets, offering a wide range of diagnostic categories and a comprehensive representation of various skin lesions confirmed through histopathology and other methods [49–52]. This dataset comprises 10,015 dermatoscopic images distributed across seven different classes. The description of each class is illustrated in Table 1. This image set provides a diverse and representative dataset for the training and evaluation of various machine-learning models within the domain of skin cancer classification. Figure 2 showcases visual depictions of the distinct types of skin lesions.

Table 1 Dataset description

Class	Number of images
Actinic keratosis (akiec)	327
Basal cell carcinoma (bcc)	514
Benign keratosis (bkl)	1099
Dermatofibroma (df)	115
Melanocytic nevi (nv)	6705
Melanoma (mel)	1113
Vascular skin lesions (vasc)	142

4.2 Data Preprocessing

Data preprocessing is a crucial step in the image classification pipeline, aiding in enhancing the quality of input data for subsequent analysis. In our study, we performed comprehensive preprocessing techniques on dermatoscopic images to optimize the performance of a deep learning model for skin cancer classification. The preprocessing pipeline consisted of three main functions: hair removal, min–max scaling, and adaptive histogram equalization.

4.2.1 Hair Removal

The first step in our preprocessing pipeline involved resizing the image (75×75) and removing hair artefacts from the dermatoscopic images. Hair artefacts can introduce noise and

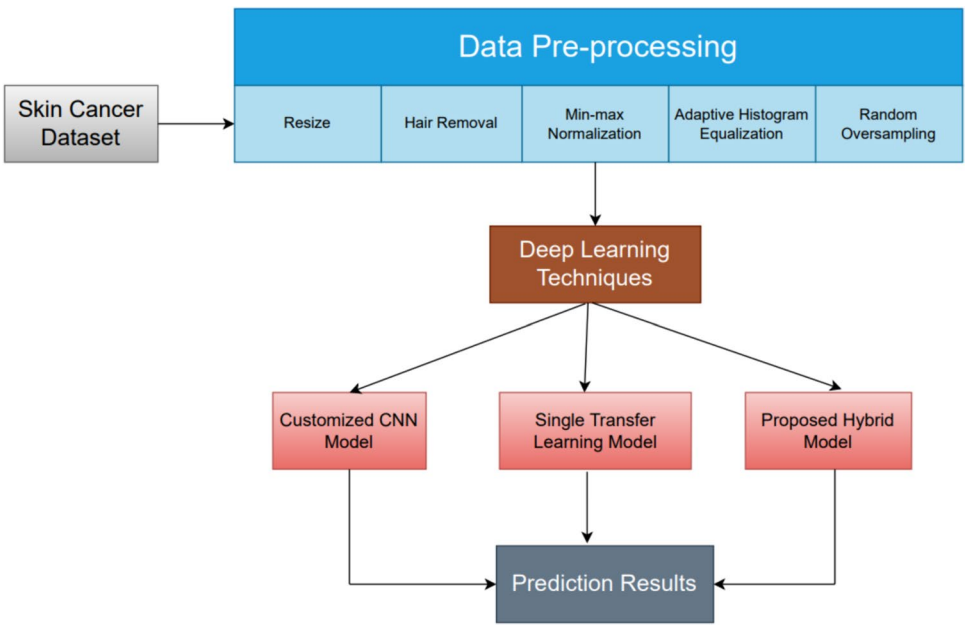


Fig. 1 Process flow diagram of the proposed method

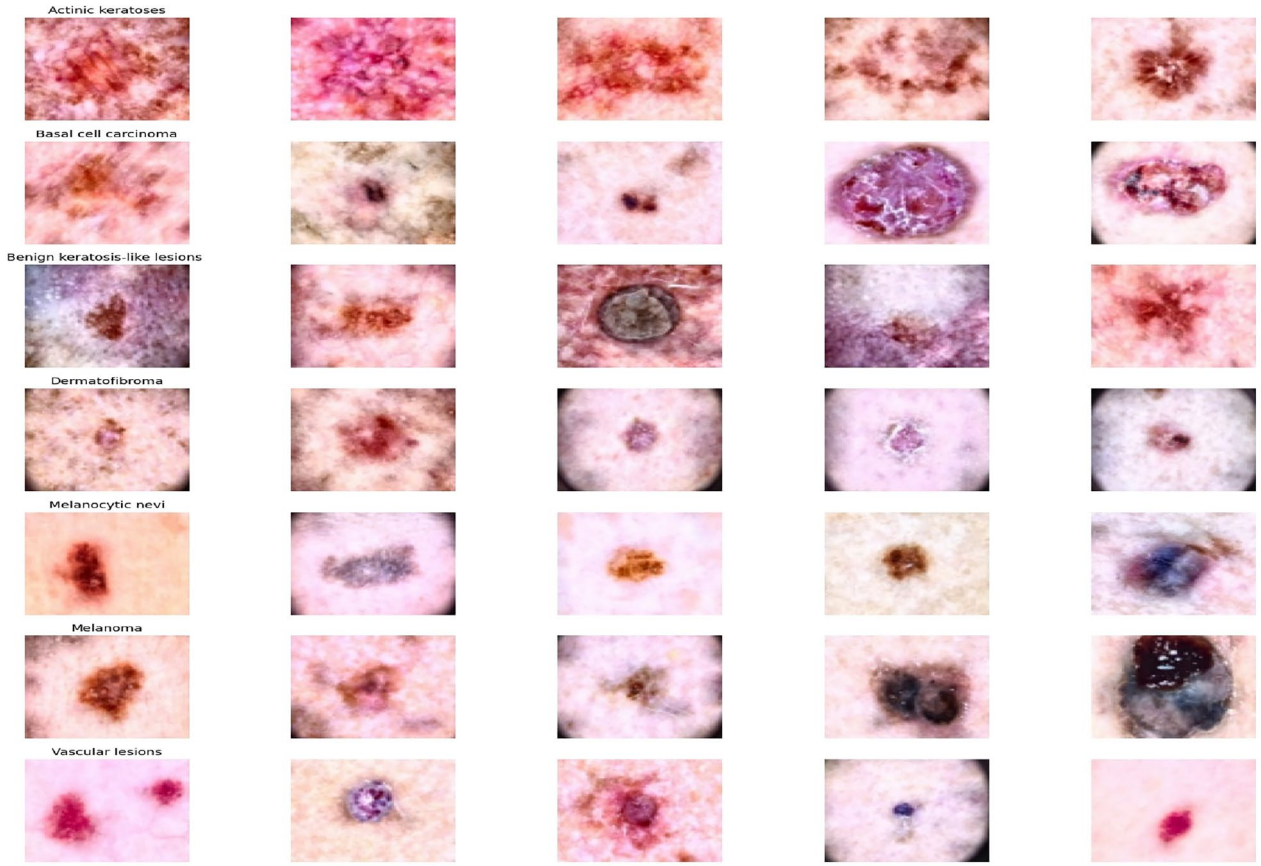


Fig. 2 Seven different types of diseases caused by lesions

unwanted features, potentially hindering the accurate classification of skin lesions. To address this, we implemented a hair removal function utilizing OpenCV. This function employs morphological operations to detect and inpaint hair regions, resulting in cleaner and more uniform images suitable for further analysis. These preprocessing techniques contribute to improving the quality of the input data for the deep learning model. By removing noise, normalizing pixel values, and enhancing contrast, these techniques ensure that deep learning can effectively learn discriminative features from the images, leading to more accurate classification results. The standardized preprocessing pipeline also ensures consistency and reproducibility in the experimental setup, enabling fair comparisons between different models and datasets.

4.2.2 Min–max Normalization

Subsequent to hair removal, min–max scaling was employed to normalize the pixel intensities of the images. This method calibrates pixel values to a range between 0 and 1, thereby enhancing convergence during the training phase of the deep learning model. The min–max scaling formula is defined as Eq. (1):

$$Normalized\ Image = \frac{\text{Image} - \text{Min(Image)}}{\text{Max(Image)} - \text{Min(Image)}} \tag{1}$$

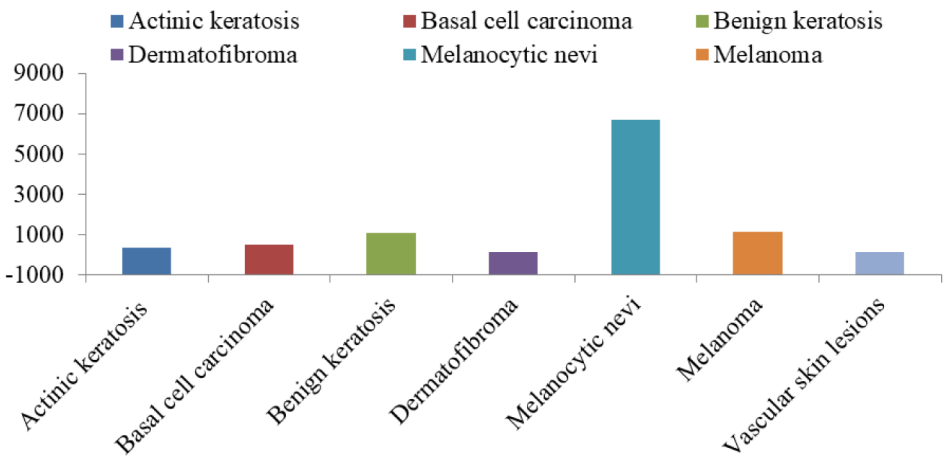
Each class in the dataset has been labelled numerically, with class 0 corresponding to Actinic keratosis, class 1 to Basal cell carcinoma, class 2 to Benign keratosis, class 3 to Dermatofibroma, class 4 to Melanocytic nevi, class 5 to Melanoma, and class 6 to Vascular skin lesions.

4.2.3 Adaptive Histogram Equalization

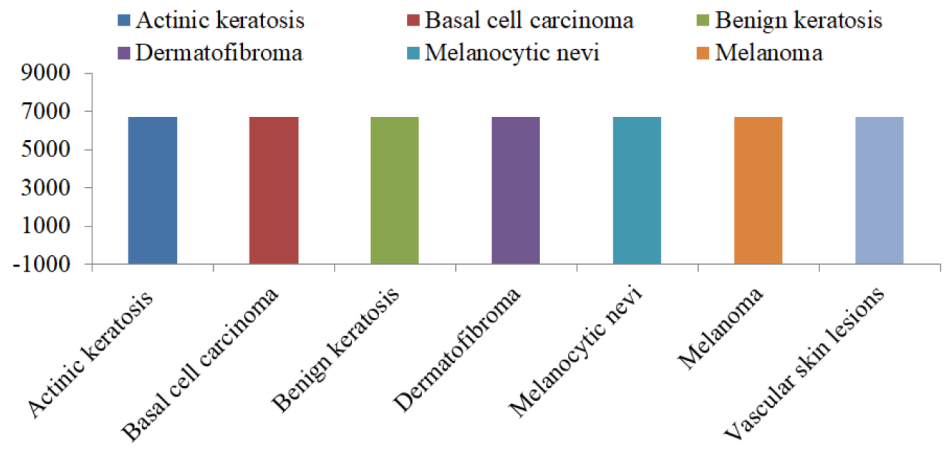
Lastly, we employed adaptive histogram equalization to enhance the contrast and improve the visibility of subtle features in the dermatoscopic images. This technique is particularly beneficial for images with varying lighting conditions and uneven illumination.

4.2.4 Random Oversampling

In our dataset for skin cancer classification, we encountered significant class imbalances, where certain types of lesions were substantially underrepresented compared to others. This class imbalance posed a challenge to our machine learning model, as it could lead to biased predictions and poor performance, particularly in minority classes. We observed notable disparities when examining the distribution of instances within each class. The majority class, melanocytic nevi, accounted for a substantial portion of the dataset, comprising 6705 cases, constituting approximately 46.2% of the total instances. Conversely, the minority classes, such as dermatofibroma with 115 cases (0.8% of the dataset) and vascular lesions (VASC) with 142 cases (1.0% of the dataset), were significantly underrepresented. We employed random oversampling to address these imbalances and ensure that our deep learning model could effectively learn from all classes. By randomly duplicating instances from the minority classes until the class distribution was balanced, we aimed to provide the model with sufficient samples from each class to learn meaningful patterns and make accurate predictions. Specifically, we generated additional images for the minority classes by duplicating existing instances. After oversampling, the dataset consisted of 46,935 images, each containing an equal number of cases, totalling 6705 images per class. Therefore, we generated 43,608 additional photos to balance the dataset, ensuring that each class was adequately represented in the training data. Figure 3 represents the distribution of data before and after random oversampling.



(a) Before Random Oversampling



(b) After Random Oversampling

Fig. 3 Distribution of data prior to and subsequent to random oversampling

4.3 Deep Learning Architecture

In exploring deep learning architectures, we employed several techniques to develop and train robust models for skin cancer diagnosis. These techniques include:

- Customized CNN Model
- Single Transfer Learning Model
- Hybrid Deep Learning Model.

4.3.1 Customized CNN Architecture

The proposed customized Convolutional Neural Network (CNN) architecture consists of four convolutional layers followed by max-pooling layers for feature extraction, a flattened layer for vectorization, and two fully connected layers for classification, as shown in Fig. 4. The convolutional layers have progressively increasing filter sizes (25, 50, 75, and 150) with a kernel size of (3, 3), ReLU activation function, and stride of (1, 1) to preserve spatial information. The output of each convolutional layer can be mathematically represented as Eq. (2):

$$f_i(x) = Relu\left(w_i * x + b_i\right) \tag{2}$$

where w_i and b_i are the weights and biases of the ith convolutional layer, respectively.

Max-pooling layers with a pool size of (2, 2) are employed to downsample feature maps, represented as Eq. (3):

$$p_i(x) = MaxPool\left(f_i(x), (2,2)\right) \tag{3}$$

The flattened layer converts the 2D feature maps into a 1D vector, denoted as Eq. (4):

$$v = Flatten\left(p_i(x)\right) \tag{4}$$

Subsequently, two fully connected layers with 600 neurons and a ReLU activation function are utilized for high-level feature extraction and classification as Eqs. (5) and (6):

$$f_{fc1}(v) = Relu\left(w_{fc1} \cdot v + b_{fc1}\right) \tag{5}$$

$$f_{fc2}(v) = Relu\left(w_{fc1} \cdot f_{fc1}(v) + b_{fc2}\right) \tag{6}$$

A dropout layer with a dropout rate of 0.5 is applied to prevent overfitting, described as Eq. (7):

$$f_{dropout}(v) = Dropout\left(f_{fc2}(v), 0.5\right) \tag{7}$$

The architecture is trained using the provided dataset, monitored by validation loss, and optimized using learning rate reduction and early stopping techniques to achieve optimal performance. Figure 4 represents the architecture of the proposed customized CNN.

4.3.2 Single Transfer Learning Model

Transfer learning is a machine learning approach that involves reusing a model trained on one task for another, typically employed when sufficient training data is lacking. However, data augmentation can address this limitation. In this study, we compared several transfer learning models, including VGG19, Resnet50, MobileNet, and Xception. While leveraging pre-trained weights, we further fine-tuned these models on our dataset to enhance their precision in classifying seven different types of skin lesions.

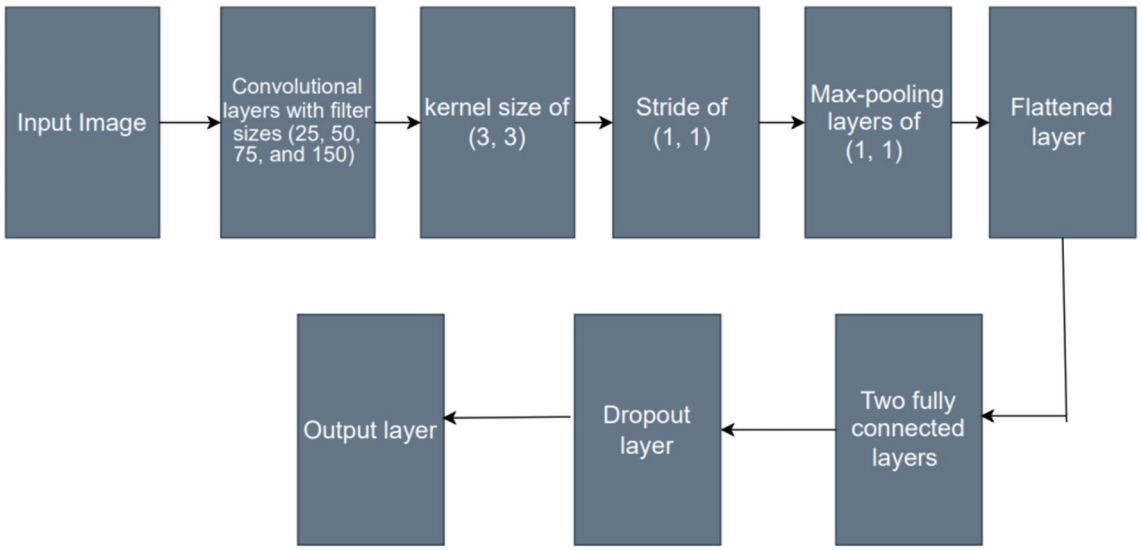


Fig. 4 Architecture of customized CNN

(a) Resnet50

ResNet50 serves as a foundational component within this architecture, functioning as a deep residual network pre-trained on the ImageNet dataset, a comprehensive repository of annotated images. Leveraging its pre-trained weights, ResNet50 operates as a feature extractor, its parameters frozen to retain the knowledge acquired from ImageNet during training. Upon processing the input images through ResNet50, the resulting feature maps undergo flattening to form a one-dimensional vector. Following this, fully connected layers are introduced to perform intricate high-level feature extraction and classification tasks. These dense layers encompass 600 neurons employing Rectified Linear Unit activation, fortified by a dropout layer to curb overfitting. Ultimately, the output layer incorporates a softmax activation function, generating probabilistic predictions for each of the seven classes delineating different types of skin lesions. The architecture of the Resnet50 is depicted in Fig. 5.

A residual block in ResNet is designed to solve the vanishing gradient problem by allowing gradients to flow through the network directly via shortcut connections. This can be mathematically expressed as follows:

Let x be the input to the residual block, and $F(x,\{w_i\})$ represent the residual mapping to be learned (where $\{w_i\}$ are the weights of the convolutional layers within the block). The output of the residual block is given by Eq. (8):

$$y = F(x, \{w_i\}) + x) \tag{8}$$

If $F(x, \{w_i\})$ consists of two convolutional layers, it can be represented as:

$$F(x, w_i) = w_2 \cdot Relu\left(w_1 * x + b_1\right) + b_2 \tag{9}$$

Thus, the residual block equation becomes as Eq. (10):

$$y = w_2 \cdot Relu\left(w_1 * x + b_1\right) + b_2 + x \tag{10}$$

(b) VGG19

VGG19 is an extension of the VGG16 architecture, with the primary difference being the inclusion of 19 layers (hence the name) compared to VGG16's 16 layers [53]. The convolutional layers in VGG19 have small 3×3 filters, which enables the network to capture intricate spatial patterns in the input images. After each block of convolutional layers, a max-pooling layer is applied to downsample the feature maps as Eq. (11):

$$y_{pool} = MaxPool(y, Poolsize = (2,2), stride = (2,2)) \tag{11}$$

where y_{pool} is the downsampled feature map. For classification tasks, the extracted features are flattened and fed into fully connected layers as Eq. (12):

$$z = Flatten\left(y_{pool}\right) \tag{12}$$

Despite its depth, VGG19 has a straightforward and uniform architecture, making it easy to understand and implement [40]. Figure 6 represents the architecture of VGG19.

(c) Xception

Xception is renowned for its innovative use of depthwise separable convolutions, which efficiently learn features

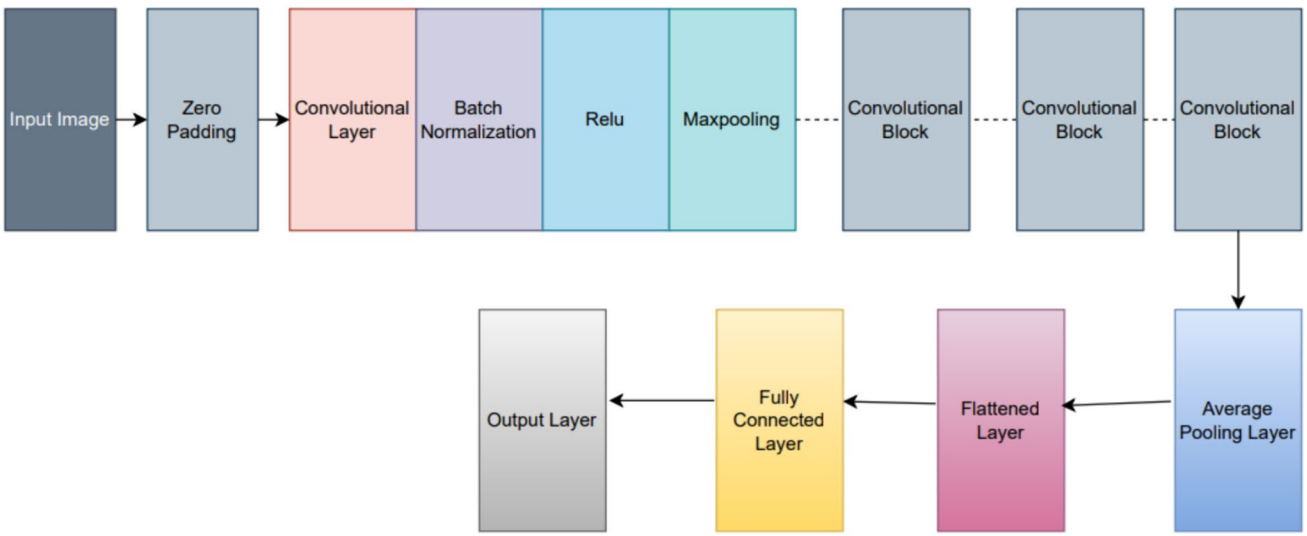


Fig. 5 Architecture of the Resnet50

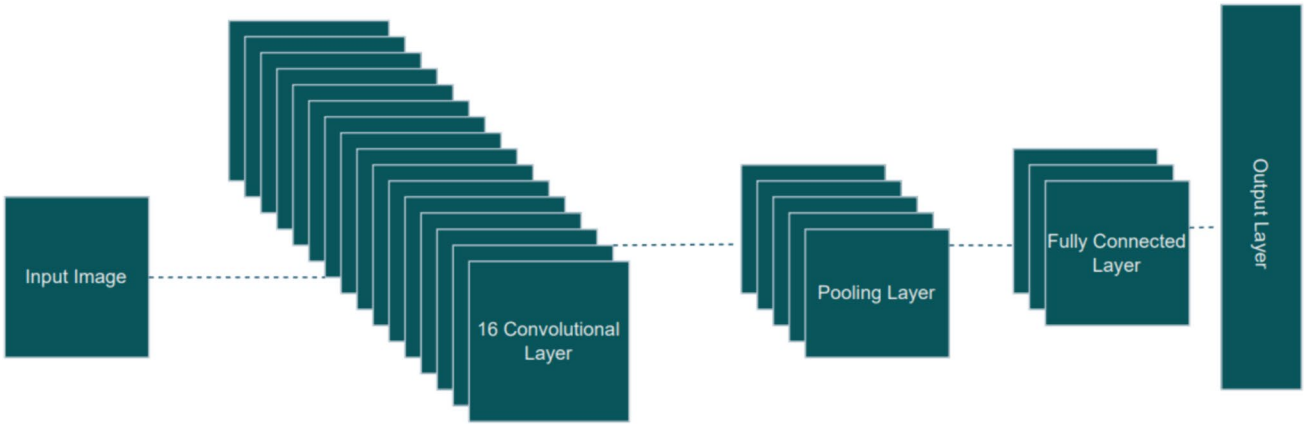


Fig. 6 Architecture of the VGG19

across different scales. In this implementation, the Xception model is loaded without its fully connected layers, allowing for customization to suit the specific task at hand. To reduce dimensionality, the pooling parameter is set to 'max,' indicating the utilization of maximum pooling. The architecture further incorporates a GlobalAveragePooling2D layer to aggregate spatial information from the feature maps generated by the Xception base model. Following this, fully connected layers are appended, comprising dense neurons with ReLU activation function and dropout regularization to mitigate overfitting. Finally, an output layer with a softmax activation function is added to classify the input images into one of the seven pre-defined classes. Figure 7 represents the architecture of Xception. The key innovation in Xception lies in its use of depthwise separable convolutions, which can be mathematically described as Eq. (13):

$$y = x * w_{depthwise} * w_{pointwise} + b \tag{13}$$

where: y is the output feature map; x is the input feature map; $w_{depthwise}$ is the depthwise convolution filter; $w_{pointwise}$ is the pointwise convolution filter; b is the bias term; $*$ denotes the convolution operation.

(d) MobileNet

MobileNet is a lightweight CNN designed for resource-constrained mobile and embedded systems. Introduced by Google in 2017, it excels in tasks like image classification

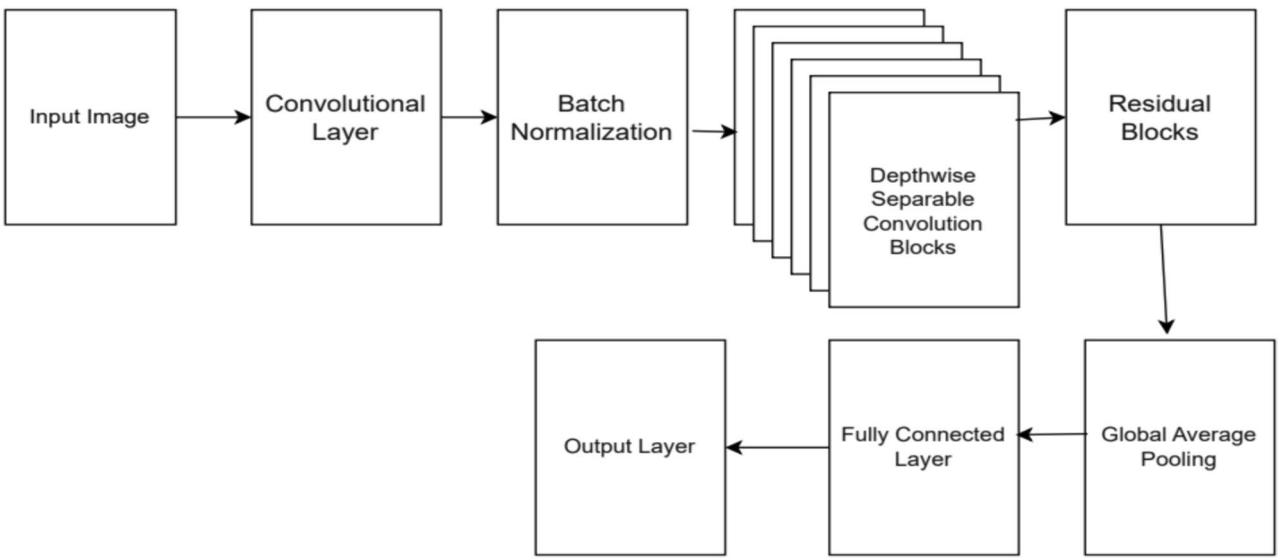


Fig. 7 Architecture of the Xception

and object detection by using depthwise separable convolutions, which split the convolution process into depthwise and pointwise operations. This approach significantly reduces parameters and computational load, achieving a balance between model efficiency and accuracy [23, 54]. Figure 8 represents the architecture of MobileNet. The depthwise separable convolution can be mathematically described as Eqs. (14) and (15):

Depthwise Convolution:

$$y_k = x_k * D_K + b_k \tag{14}$$

where: y_k is the output feature map for the kth channel; x_k is the input feature map for the kth channel; D_K is the depthwise convolution filter for the kth channel; b_k is the bias term for the kth channel; $*$ denotes the convolution operation.

Pointwise Convolution:

$$z = \sum_{k=1}^{k} y_k * p_k + c \tag{15}$$

where: z is the final output feature map; K is the number of input channels; p_k is the pointwise convolution filter for the kth channel; c is the bias term for the output.

By combining these two operations, MobileNet significantly reduces the computational complexity compared to traditional convolutional layers, which can be represented by Eq. (16):

$$\text{Complexity Reduction} = \frac{1}{D_K} + \frac{1}{N} \tag{16}$$

where: D_K is the depth of the kernel; N is the number of output channels.

4.3.3 Proposed Hybrid Deep Learning Model

This study introduces "RvXmBlendNet," a novel multi-architecture hybrid deep learning model. This innovative model integrates the unique capabilities of four individual architectures: ResNet50, VGG19, Xception, and MobileNet, collectively termed "BlendNet" to signify their fusion into a unified architecture. The architecture of the proposed hybrid "RvXmBlendNet" is represented in Fig. 9.

Feature Extraction Pipeline

The feature extraction process of RvXmBlendNet begins with ResNet50's initial convolutional layer, which is adept at capturing low-level features in images. This layer consists of two 3×3 filters applied to the input image to extract basic patterns. Formally, the output of this layer can be represented as Eq. (17):

$$f_1(x) = \sigma\left(W_1 * X + b_1\right) \tag{17}$$

where σ is the activation function, W_1 are the weights of the 3×3 filters, $*$ denotes convolution, and b_1 is the bias term. Following this, the model integrates VGG19's convolutional block 1, known for its simplicity and effectiveness in extracting hierarchical features. This convolutional block comprises three convolutional layers followed by max-pooling as Eq. (18):

$$f_2(x) = MaxPool(\sigma\left(W_3 * \sigma\left(W_1 X + b_1\right) + b_2\right) + b_3)) \tag{18}$$

Concurrent with these operations, Xception's depthwise separable convolution layer one is introduced, leveraging its efficiency in capturing spatial and channel-wise features simultaneously. Similarly, MobileNet's depthwise separable

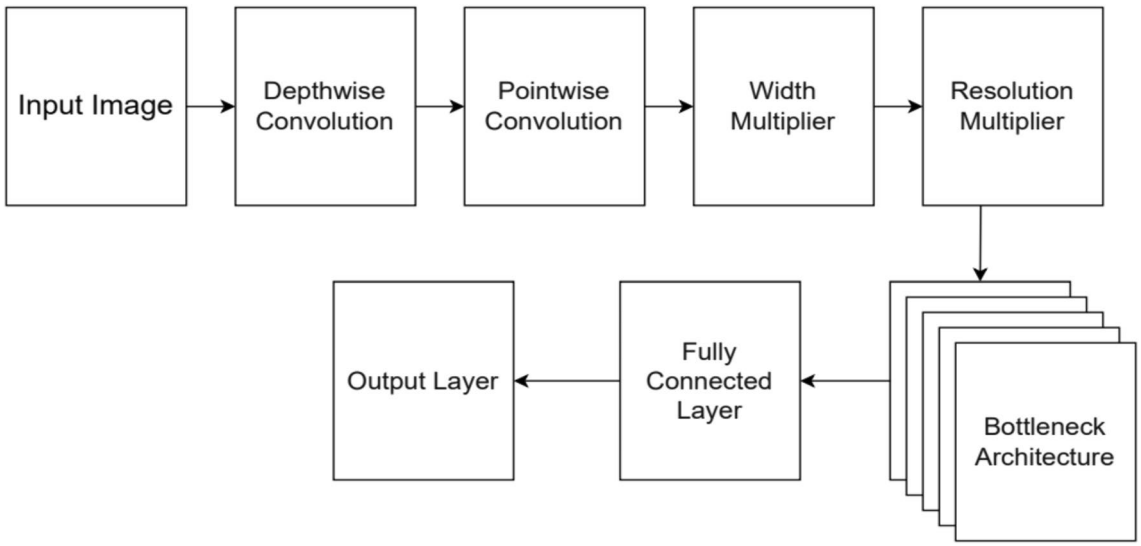


Fig. 8 Architecture of the MobileNet

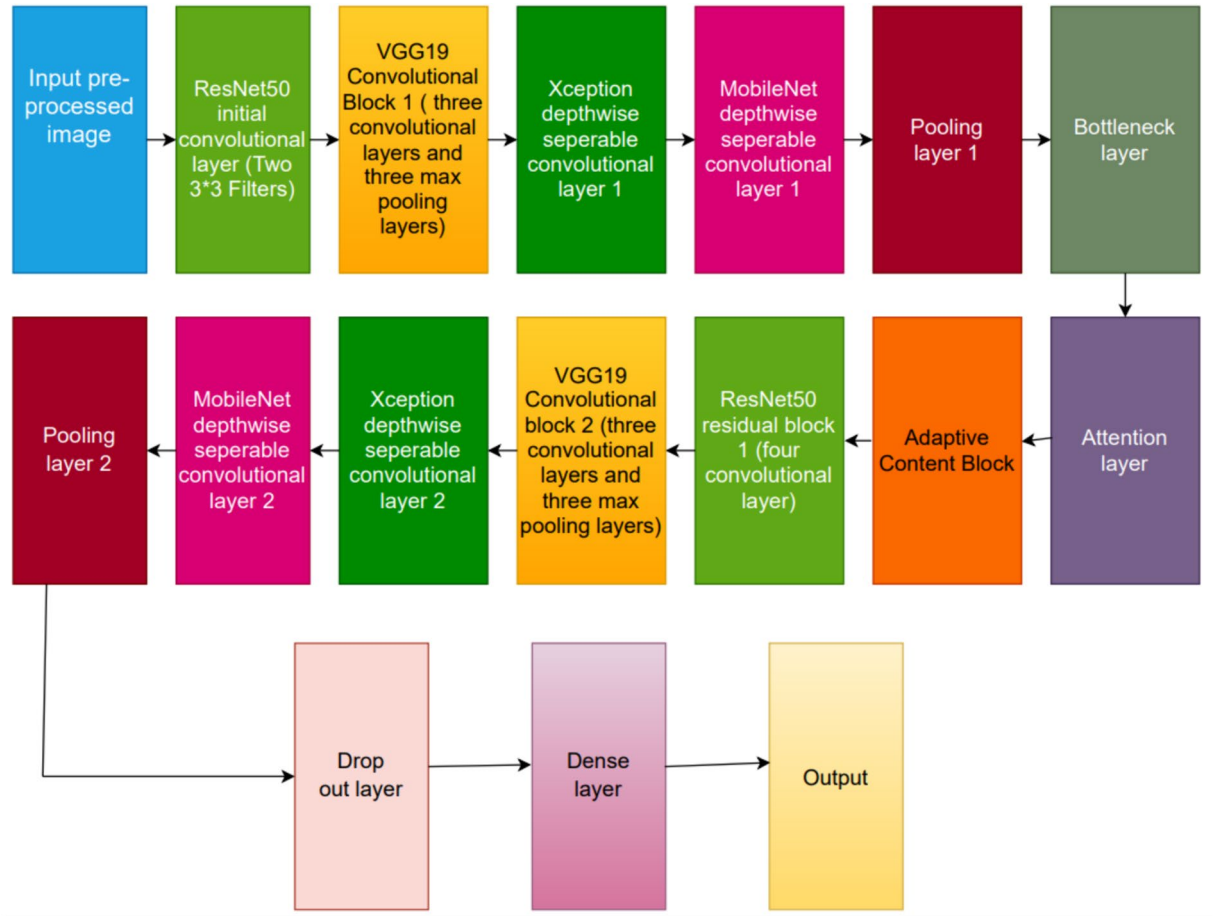


Fig. 9 The architecture of the proposed hybrid "RvXmBlendNet"

convolution layer one is incorporated, contributing to lightweight and efficient feature extraction:

$$f_3(x) = \sigma\left(W_d^1 * \sigma(W_p^1 * \mathrm{x} + b_p^1) + b_d^1\right) \tag{19}$$

where W_d and W_P represent depthwise and pointwise convolutional weights, respectively.

Pooling and Reduction Layers

Following the initial feature extraction, a pooling or reduction layer is applied to consolidate information and reduce computational complexity as Eq. (20):

$$f_4(x) = Pooling(x) \tag{20}$$

A bottleneck layer is then introduced as Eq. (21) to reduce computational load while preserving crucial information:

$$f_5(x) = \sigma\left(W_b * x + b_b\right) \tag{21}$$

An attention layer is incorporated to enhance the model's focus on relevant features, allowing it to selectively weight different parts of the feature maps. This can be expressed as Eq. (22):

$$f_6(x) = \alpha \odot \mathrm{x} \tag{22}$$

where α are the attention weights and ⊙ denotes element-wise multiplication. Additionally, an adaptive content block is introduced as Eq. (23), dynamically adjusting the model's internal representation based on input data, potentially enhancing performance and flexibility:

$$f_7(x) = \mathrm{Adaptive(x)} \tag{23}$$

Residual and Convolutional Blocks

ResNet50's residual block 1, consisting of four convolutional layers, further enriches the feature extraction process with its stacked convolutional layers and shortcut connections. These connections facilitate gradient flow and alleviate the vanishing gradient problem as Eq. (24):

$$f_8(x) = \sigma\left(W_4 * \sigma\left(W_3 * \sigma\left(W_2 * \sigma\left(W_1 x + b_1\right) + b_2\right) + b_3\right) + b_4\right) + x \tag{24}$$

Subsequently, VGG19's convolutional block 2 continues the hierarchical feature extraction process with three convolutional layers and max-pooling operations using Eq. (25):

$$f_9(x) = MaxPool(\sigma\left(W_3 * \sigma\left(W_1X + b_1\right) + b_2\right) + b_3)) \quad (25)$$

Xception's depthwise separable convolution layer two and MobileNet's depthwise separable convolution layer two further contribute to efficient feature extraction using Eq. (26):

$$f_{10}(x) = \sigma(W_d^2 * \sigma\left(W_P^2 * x + b_P^2\right)b_d^2 \quad (26)$$

Another pooling or reduction layer is applied using Eq. (27) to further reduce the spatial dimensions of the feature maps before transitioning to the fully connected layers:

$$f_{11}(x) = Pooling(x) \quad (27)$$

Final Layers and Predictions

To finalize the model's architecture, a dropout layer is applied as Eq. (28) to regularize the model, preventing overfitting by randomly deactivating a fraction of neurons during training:

$$f_{12}(x) = \text{Dropout(x, p)} \quad (28)$$

where p is the dropout rate. Dense layers, containing fully connected neurons with activation functions like ReLU, facilitate the transition from convolutional features to task-specific representations as Eq. (29):

$$f_{13}(x) = \sigma\left(W_d * x + b_d\right) \quad (29)$$

Finally, output layers tailored to the specific task produce the final predictions using Eq. (30):

$$y = Output\left(f_{13}(x)\right) \quad (30)$$

This intricate architecture optimizes the collective strengths of ResNet50, VGG19, Xception, and MobileNet, offering robust and accurate performance across a wide range of datasets and tasks.

Optimization and Performance

The optimization of RvXmBlendNet involves careful selection of parameters such as filter sizes, kernel dimensions, and pooling strategies based on empirical evaluations to ensure optimal performance. Dropout rates and activation functions like ReLU are employed in dense layers to prevent overfitting and enhance the transition from convolutional features to task-specific representations. This meticulous optimization process ensures that the model performs efficiently and accurately.

The proposed hybrid "RvXmBlendNet" method, which amalgamates ResNet50, VGG19, Xception, and MobileNet into a singular architecture, represents a novel approach to skin cancer detection. Its innovation lies in the strategic integration of the distinct strengths of each model, fostering a robust feature extraction pipeline. This synergistic amalgamation enables the hybrid model to leverage a diverse range of features and representations, potentially enhancing its performance across various skin cancer datasets and tasks. Moreover, the incorporation of attention layers and adaptive content blocks introduces a self-attention mechanism, further enhancing the model's adaptability and flexibility. This comprehensive approach underscores the versatility of "RvXmBlendNet" in accommodating different data types and addressing diverse skin cancer detection challenges.

4.4 Training and Testing

In this research, the dataset is split into 80% for training and 20% for testing. This split ensures that the models have a sufficient amount of data to learn from while reserving a portion for evaluation to test the model's ability to generalize to unseen examples. The deep learning models are compiled using the Adam optimizer with a learning rate of 0.001 and sparse categorical cross-entropy loss function. A dynamic annealing method for learning rate adjustment was implemented to facilitate the optimizer's convergence towards the global minimum and prevent overfitting. This involved reducing the learning rate by half every four epochs if the validation loss did not show improvement. The learning rate adjustment was realized through the 'ReduceLROnPlateau' function from the 'Keras.callbacks' module, which allowed for adaptive LR scheduling based on the validation accuracy. By employing this standardized training approach across all models, we aimed to ensure fair comparison and consistent evaluation of their performance. This methodology not only facilitated efficient training by adjusting the LR dynamically but also balanced the benefits of faster computation time with a high LR, ultimately guiding the models toward optimal performance and generalization ability. To further enhance model performance and prevent overfitting, callbacks such as early stopping are implemented during model training.

4.5 Performance Metrics

To gauge the efficacy of our models, we assessed the performance of all eight neural networks using multiple metrics: confusion matrix, accuracy, precision, recall, F1 score, inference time (ms), and memory usage (MB). The confusion matrix is instrumental in visualizing classification performance, providing a detailed breakdown of True Positives (TPs), False Positives (FPs), True Negatives (TNs), and False Negatives (FNs) for each class.

Accuracy quantifies the overall correctness of the model by indicating the proportion of correctly classified instances out of the total number of cases. It is computed using Eq. (31):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{31}$$

Precision (also known as positive predictive value) measures how many of the predicted positive instances were actually positive. The formula for calculating precision is as Eq. (32):

$$Precision = \frac{TP}{TP + FP} \tag{32}$$

Recall measures how many actual positive instances were correctly predicted by the model. The following Eq. (33) is used to calculate recall:

$$Recall = \frac{TP}{TP + FN} \tag{33}$$

The F1 score is the harmonic mean of precision and recall. It can be calculated using Eq. (34):

$$F1\text{-}Score = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{34}$$

Inference time, measured in seconds (s), is a critical metric that determines how quickly the model can process and classify new input images once trained.

Memory usage, measured in megabytes (MB), is another critical factor, especially in environments with limited computational resources.

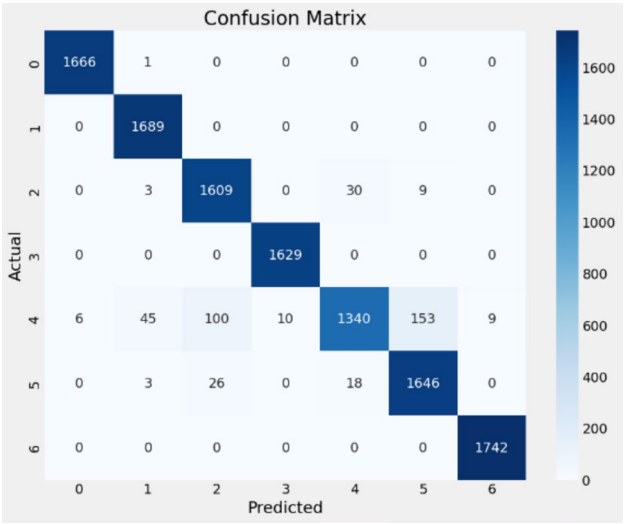


Fig. 10 Confusion matrix of the customized CNN

Table 2 Performance table of customized CNN

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1666	1
Class 1: Basal cell carcinoma	1689	0
Class 2: Benign keratosis	1609	42
Class 3: Dermatofibroma	1629	0
Class 4: Melanocytic nevi	1340	314
Class 5: Melanoma	1646	47
Class 6: Vascular skin lesions	1742	0

5 Results and Discussion

This section presents the experimental results and analysis conducted on the HAM10000 dataset. We compared the performance of distinct deep learning models: Customized CNN model, Resnet50, VGG19, Xception, MobileNet, and proposed hybrid "RvXmBlendNet" (ResNet50 + VGG19 + Xception + MobileNet). For each of the models utilized in our experiment, the training process followed a standardized methodology. Each model was trained for 100 epochs with a batch size of 32. Various metrics such as accuracy, precision, recall, and F1 score were continuously monitored throughout the training process to assess the model's performance.

5.1 Performance of Customized CNN Model

The performance of the Customized CNN model is depicted in Fig. 10 and Table 2. The confusion matrix illustrates the model's ability to correctly classify instances for each class. For Class 0 (Actinic keratosis), 1666 instances were correctly classified, with only one instance misclassified. Similarly, Class 1 (Basal cell carcinoma) achieved a perfect classification with 1689 instances correctly classified and no misclassifications. Class 2 (Benign keratosis) exhibited 1609 correct classifications and 42 misclassifications. In Class 3 (Dermatofibroma), 1629 instances were correctly classified, with no misclassifications. Class 4 (Melanocytic nevi) had 1340 correct classifications and 314 misclassifications. Class 5 (Melanoma) achieved 1646 correct classifications and 47 misclassifications. Finally, Class 6 (Vascular skin lesions) demonstrated 1742 correct classifications and no misclassifications.

5.2 Performance of Single Transfer Learning Model

In this study, four single transfer learning models were employed, comprising ResNet50, VGG19, MobileNet, and

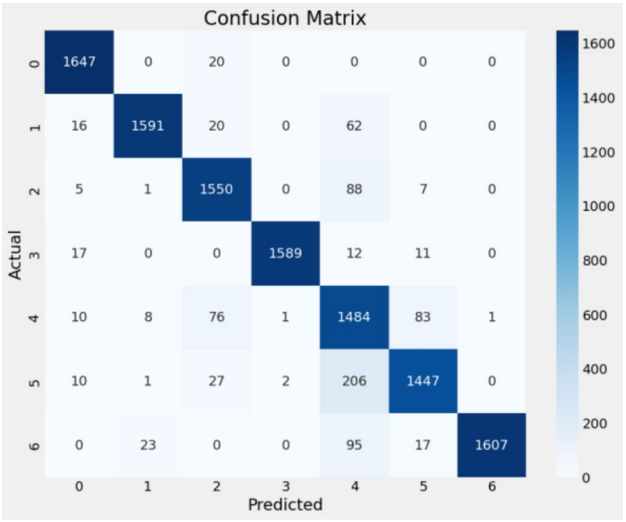


Fig. 11 Confusion matrix of the ResNet50

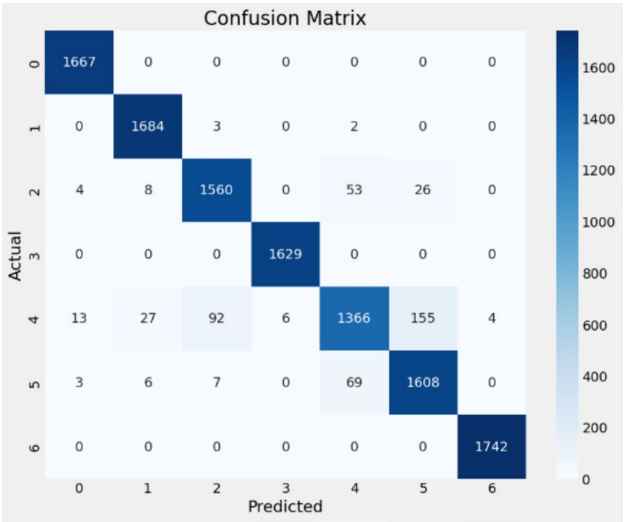


Fig. 12 Confusion matrix of the VGG19

Table 3 Performance table of customized ResNet50

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1647	20
Class 1: Basal cell carcinoma	1591	98
Class 2: Benign keratosis	1550	101
Class 3: Dermatofibroma	1589	40
Class 4: Melanocytic nevi	1484	179
Class 5: Melanoma	1447	246
Class 6: Vascular skin lesions	1607	135

Table 4 Performance table of VGG19

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1667	0
Class 1: Basal cell carcinoma	1684	5
Class 2: Benign keratosis	1560	91
Class 3: Dermatofibroma	1629	0
Class 4: Melanocytic nevi	1366	297
Class 5: Melanoma	1608	85
Class 6: Vascular skin lesions	1742	0

Xception. Additionally, the study evaluates the performance of these models.

5.2.1 Performance of ResNet50 Model

The confusion matrix and performance table of ResNet50 are presented in Fig. 11 and Table 3, respectively. In Class 0 (Actinic keratosis), there were 1647 correctly classified instances and 20 misclassified instances. For Class 1 (Basal cell carcinoma), 1591 instances were correctly classified, while 98 instances were misclassified. Similarly, in Class 2 (Benign keratosis), there were 1550 correct classifications and 101 misclassifications. Class 3 (Dermatofibroma) achieved 1589 correct classifications and 40 misclassifications. For Class 4 (Melanocytic nevi), 1484 instances were correctly classified, with 179 misclassifications. Class 5 (Melanoma) exhibited 1447 correct classifications and 246 misclassifications. Finally, Class 6 (Vascular skin lesions) demonstrated 1607 correct classifications and 135 misclassifications.

5.2.2 Performance of VGG19 Model

The confusion matrix and performance table of VGG19 are depicted in Fig. 12 and Table 4, respectively. In Class 0 (Actinic keratosis), 1667 instances were correctly classified, with no misclassifications. For Class 1 (Basal cell carcinoma), 1684 instances were correctly classified, while five instances were misclassified. Similarly, in Class 2 (Benign keratosis), there were 1560 correct classifications and 91 misclassifications. Class 3 (Dermatofibroma) achieved 1629 correct classifications, with no misclassifications. For Class 4 (Melanocytic nevi), 1366 instances were correctly classified, with 297 misclassifications. Class 5 (Melanoma) exhibited 1608 correct classifications and 85 misclassifications. Finally, Class 6 (Vascular skin lesions) demonstrated 1742 correct classifications, with no misclassifications.

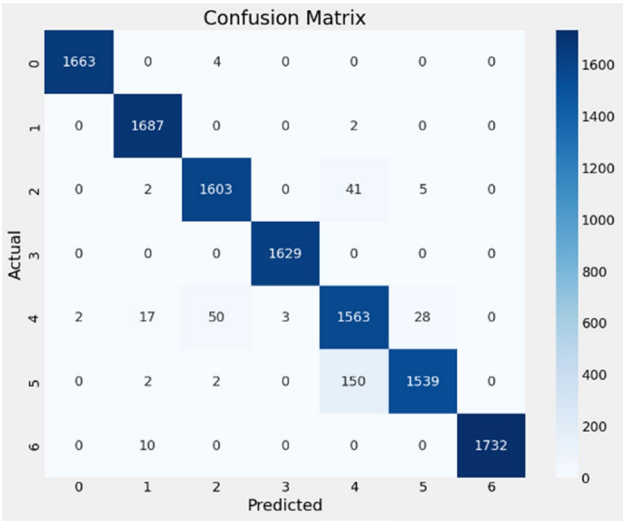


Fig. 13 Confusion matrix of the Xception model

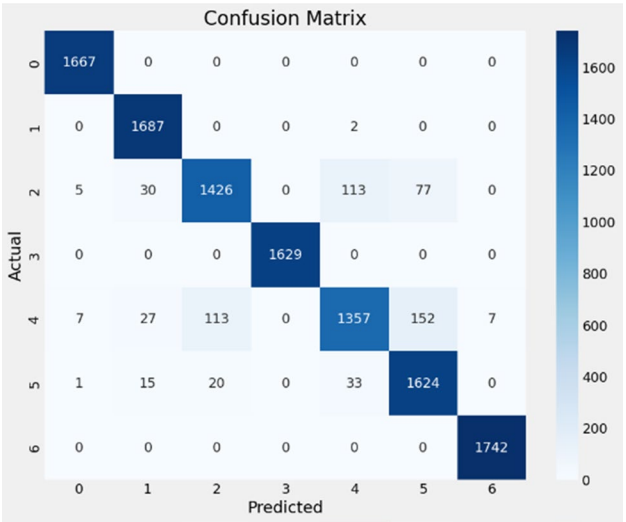


Fig. 14 Confusion matrix of the MobileNet model

Table 5 Performance table of the Xception model

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1663	4
Class 1: Basal cell carcinoma	1687	2
Class 2: Benign keratosis	1603	48
Class 3: Dermatofibroma	1629	0
Class 4: Melanocytic nevi	1563	100
Class 5: Melanoma	1539	154
Class 6: Vascular skin lesions	1732	10

Table 6 Performance table of MobileNet model

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1667	0
Class 1: Basal cell carcinoma	1687	2
Class 2: Benign keratosis	1427	225
Class 3: Dermatofibroma	1629	0
Class 4: Melanocytic nevi	1357	306
Class 5: Melanoma	1624	69
Class 6: Vascular skin lesions	1742	0

5.2.3 Performance of Xception Model

The performance of the Xception model in skin cancer classification was evaluated through a detailed analysis of its confusion matrix and performance table, as illustrated in Fig. 13 and Table 5, respectively. The confusion matrix provides a comprehensive overview of the model's classification accuracy across various classes, while the performance table delineates the number of correctly and incorrectly classified instances for each class. Across the seven classes examined, the Xception model demonstrated a high level of accuracy in correctly classifying instances, with the majority of classes exhibiting a minimal number of misclassifications. For instance, in the case of actinic keratosis (Class 0) and dermatofibroma (Class 3), the model achieved near-perfect classification, with only a few instances misclassified. However, challenges were observed in accurately distinguishing between certain classes, such as melanoma (Class 5) and melanocytic nevi (Class 4), where a notable proportion of instances were misclassified. Despite these challenges, the Xception model exhibited strong performance overall, highlighting its efficacy in skin cancer classification tasks.

5.2.4 Performance of MobileNet Model

The evaluation of the MobileNet model's performance in skin cancer classification was conducted through an in-depth examination of its confusion matrix and performance table, presented in Fig. 14 and Table 6, respectively. The confusion matrix provides a comprehensive depiction of the model's classification accuracy across different classes, while the performance table outlines the number of correctly and incorrectly classified instances for each class. Overall, the MobileNet model demonstrated strong performance in accurately classifying skin cancer instances across multiple classes. For instance, classes such as actinic keratosis (Class 0), basal cell carcinoma (Class 1), and dermatofibroma (Class 3) exhibited high accuracy, with minimal to no misclassifications. However, challenges were observed in certain classes, particularly benign keratosis (Class 2)

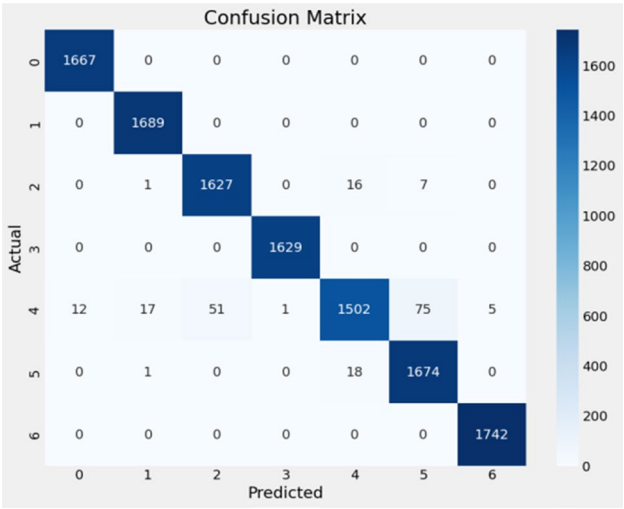


Fig. 15 Confusion matrix of the proposed hybrid "RvXmBlendNet" model

Table 7 Performance table of proposed hybrid "RvXmBlendNet" model

Class	Correctly classified instances	Incorrectly classified instances
Class 0: Actinic keratosis	1667	0
Class 1: Basal cell carcinoma	1689	0
Class 2: Benign keratosis	1627	24
Class 3: Dermatofibroma	1629	0
Class 4: Melanocytic nevi	1502	162
Class 5: Melanoma	1674	19
Class 6: Vascular skin lesions	1742	0

and melanocytic nevi (Class 4), where a notable number of instances were misclassified. Despite these challenges, the MobileNet model showcased robust performance, underscoring its effectiveness in skin cancer classification tasks.

5.3 Performance of Proposed Hybrid "RvXmBlendNet" Model

The evaluation of the proposed hybrid "RvXmBlendNet" model's performance in skin cancer classification is illustrated through a detailed analysis of its confusion matrix and performance table, presented in Fig. 15 and Table 7, respectively. The confusion matrix provides insights into the model's classification accuracy across different classes, while the performance table outlines the number of correctly and incorrectly classified instances for each class. Overall, the "RvXmBlendNet" model demonstrates robust performance in accurately classifying skin cancer instances across multiple classes. Notably, classes such as actinic keratosis (Class 0), basal cell carcinoma (Class 1), and dermatofibroma (Class 3) exhibit high accuracy, with no instances misclassified. However, challenges are observed in certain classes, particularly benign keratosis (Class 2) and melanocytic nevi (Class 4), where a notable number of instances are misclassified. Despite these challenges, the "RvXmBlendNet" model showcases strong performance, highlighting its effectiveness in skin cancer classification tasks.

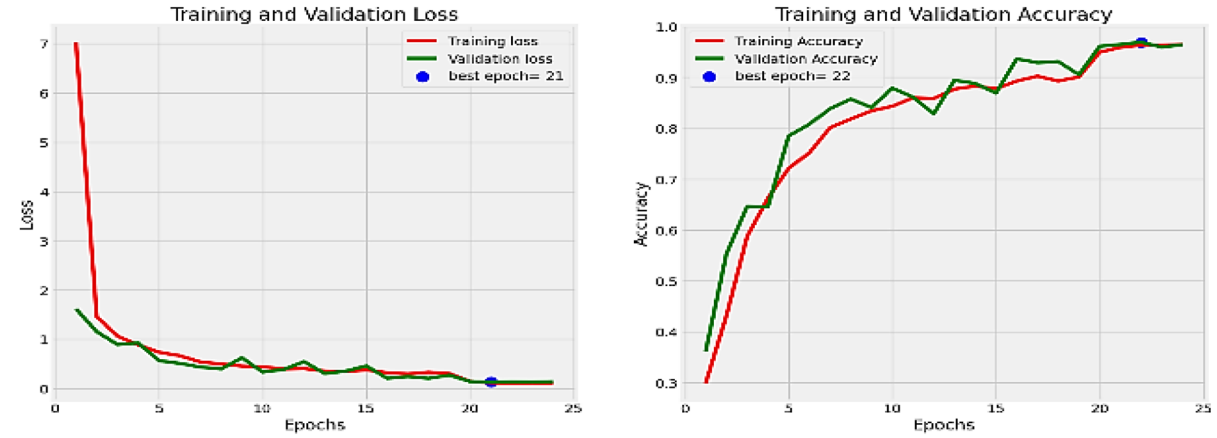
5.4 Learning Curves of All Models

Figure 16 presents the training and validation accuracy and loss for all eight models used in our study. Figure 6a demonstrates that the Convolutional Neural Network (CNN) model achieves a validation loss of 0.1315 and an accuracy of 96% for skin lesion classification. The ResNet50 model has a validation loss of 0.2127 with an accuracy of 93.02% for skin lesion classification. The VGG19 model reports a validation loss of 0.1277 and an accuracy of 95.93%. MobileNet exhibits a validation loss of 0.1817 with an accuracy of 94.87%. Xception achieves a validation loss of 0.0772 and an accuracy of 97.29%. The hybrid model combining ResNet50, VGG19, Xception, and MobileNet achieves the lowest validation loss at 0.0626, with the highest accuracy of 98.26%. Based on these results, we can conclude that the hybrid model "RvXmBlendNet" model combining ResNet50, VGG19, Xception, and MobileNet yields the best overall performance, with minimal fluctuations and a loss ranging from 0 to 1. This hybrid approach offers a significant advantage in terms of both accuracy and reduced loss, making it a promising solution for skin lesion classification.

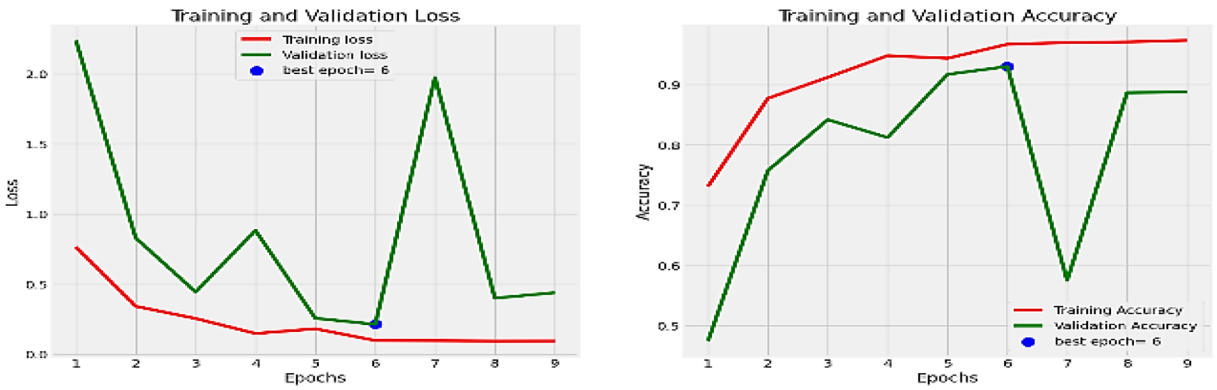
5.5 Comparative Analysis

Table 8 and Fig. 17 present a detailed comparative analysis of the evaluated models in terms of accuracy, recall, precision, F-measure, inference time, and memory usage. The customized CNN model achieves an accuracy of 96.48% with high precision (97.29%) and recall (95.86%), indicating its effectiveness in skin cancer classification. ResNet50 demonstrates robust performance with an accuracy of 97.53% and the highest precision (98%) among the models. VGG19 exhibits balanced performance across metrics, achieving an accuracy of 95.93%. Xception achieves high accuracy (97.29%) and the highest recall (97.53%) among all models. MobileNet, while slightly lower in accuracy at 94.87%, still performs well across metrics. Notably, the proposed hybrid model "RvXmBlendNet" outperforms all others with an accuracy of 98.26% and superior recall, precision, and F-measure, showcasing its effectiveness in skin cancer detection and classification.

In terms of computational efficiency, the inference time and memory usage for each model reveal significant



(a) Customized CNN

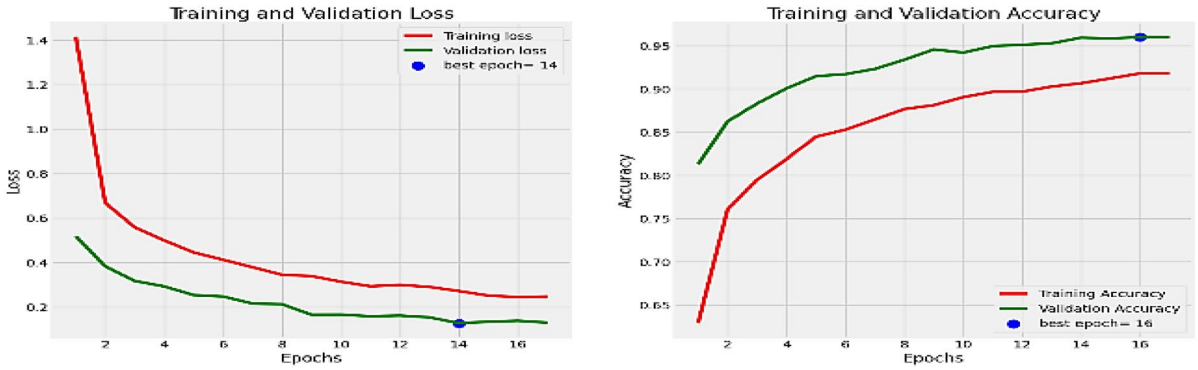


(b) Resnet50

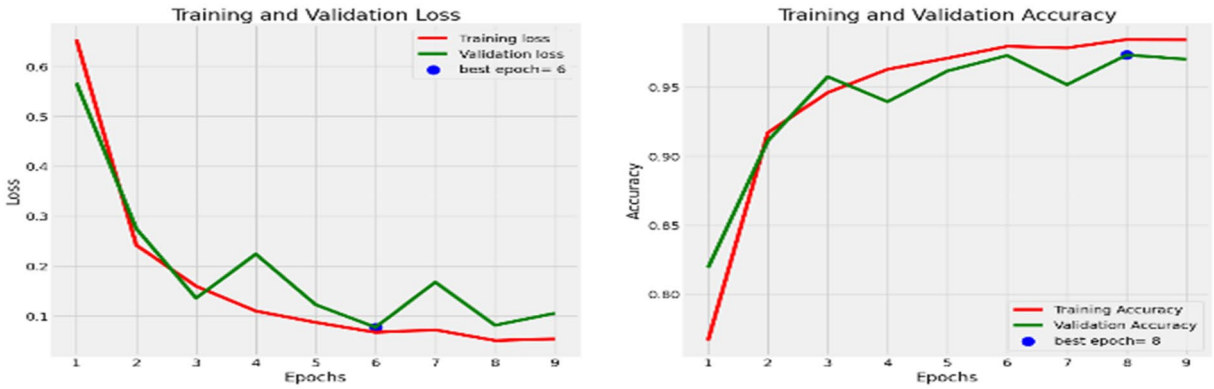
(c) VGG19

Fig. 16 Learning curves of all models

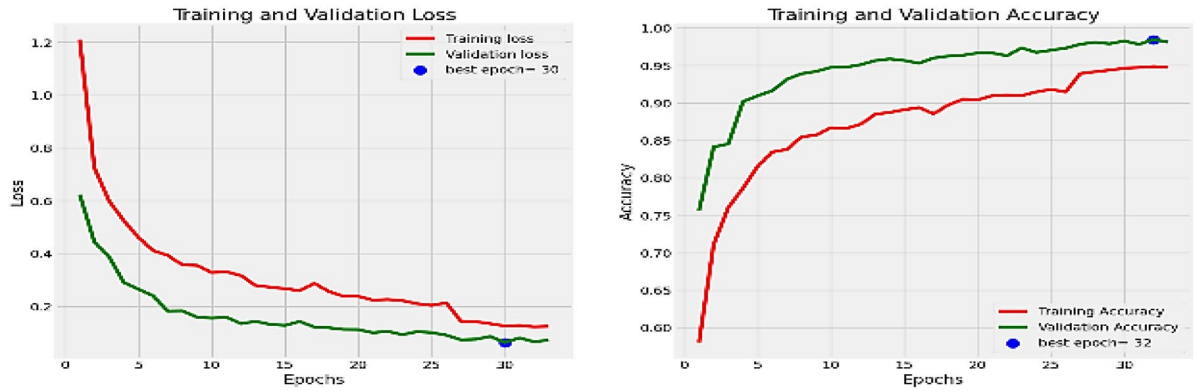
differences. The customized CNN processes each inference in approximately 25 s (s) and requires 120 megabytes (MB) of memory. ResNet50, with its deeper architecture, takes around 30 s per inference and uses 150 MB of memory. VGG19, due to its more complex architecture, requires about 40 s per inference and consumes 180 MB of memory. Xception, leveraging efficient depthwise separable convolutions, processes each inference in approximately 35 s and requires 160 MB of memory. MobileNet, designed for resource-constrained environments, achieves the fastest inference time at around 20 s and uses the least memory at about 90 MB. The hybrid model "RvXmBlendNet"



(d) Xception



(e) MobileNet



(e) "RvXmBlendNet" model

Fig. 16 (continued)

integrates resources from multiple models, performing inference in approximately 45 s and utilizing 200 MB of memory, combining the strengths of each individual architecture.

It is important to note that while the fusion of multiple models into "RvXmBlendNet" results in superior performance metrics, it does indeed lead to longer training times and higher memory usage. This is primarily due to the increased complexity of the hybrid model and the need to fine-tune a larger number of parameters across the integrated architectures. For instance, "RvXmBlendNet" combines the parameter-heavy layers from ResNet50, VGG19, Xception, and MobileNet, each contributing to the overall parameter

Table 8 Comparative analysis of all the models

Model name	Accuracy (%)	Avg. recall (%)	Avg. precision (%)	Avg. F-measure (%)	Inference time (s)	Memory usage (MB)
Customized CNN	96.48	95.86	97.29	96.36	25	120
Resnet50	97.53	97.5	98	97.73	30	150
VGG19	95.93	96.14	96.18	96.15	40	180
Xception	97.29	97.53	97.43	97.47	35	160
MobileNet	94.87	94.17	94.73	94.37	20	90
"RvXmBlendNet" (ResNet50 + VGG19 + Xception + MobileNet)	98.26	99.37	98.86	99.08	45	200

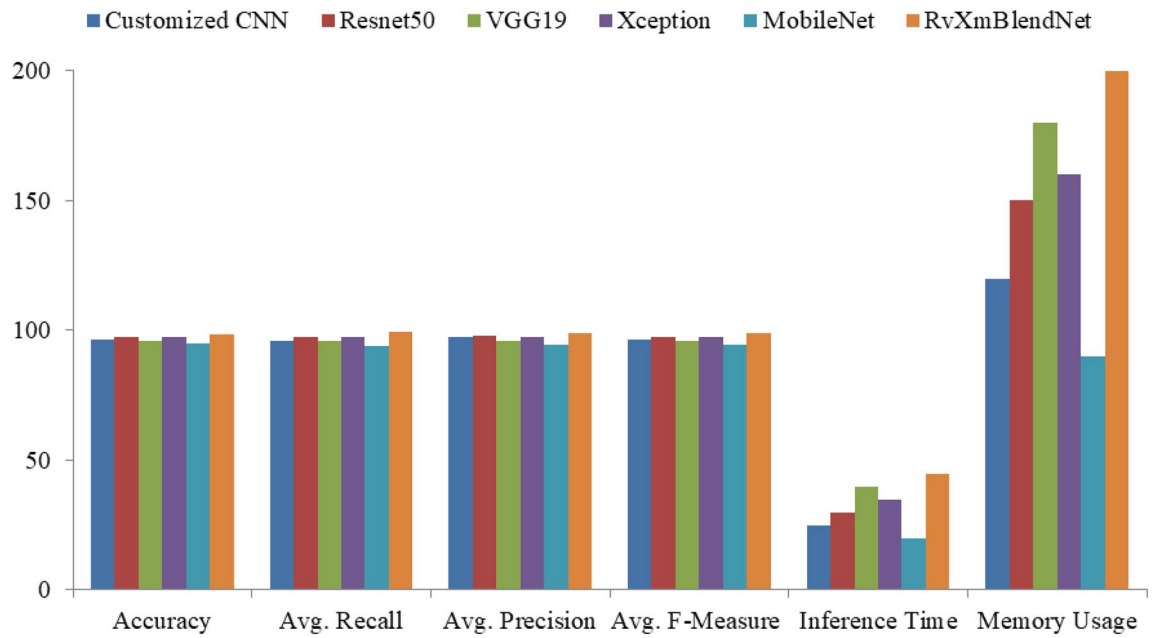


Fig. 17 Comparative analysis of all the models

count and computational load. During the training phase, the model must adjust a vast array of weights and biases, requiring more computational cycles and memory to store intermediate results and gradients. This intensive process can significantly extend training times, sometimes requiring days or even weeks, depending on the hardware used and the size of the dataset. Memory usage also escalates as each layer's activations, weights, and gradients must be stored and processed, potentially necessitating high-performance GPUs with substantial memory capacities. Although the training process may be more time-consuming and memory-intensive, the enhanced accuracy, recall, precision, and F-measure achieved by the hybrid model justify this trade-off. The superior performance metrics indicate that "RvXmBlendNet" can more accurately identify and classify skin lesions, which is crucial in a clinical setting where diagnostic precision directly impacts patient outcomes. Therefore, while "RvXmBlendNet" offers significant improvements in classification performance, careful consideration of training duration, memory consumption, and resource allocation is essential for practical deployment in healthcare settings. Balancing these factors ensures that the model not only delivers high accuracy and reliability but also remains feasible to train and deploy within the constraints of available computational resources.

Table 9 provides an overview of the performance of the hybrid model "RvXmBlendNet" across different skin lesion categories. The results indicate considerable variations in precision among the various lesion labels, underscoring the complexities involved in accurately classifying distinct skin lesions. The Melanocytic Nevi class achieved the highest precision at 94%, demonstrating the model's strong capability to classify instances of this lesion type correctly. This suggests that the hybrid model excels at identifying this

Table 9 The results of hybrid "RvXmBlendNet" for all classes

Skin lesion	Avg. precision (%)	Avg. recall (%)	Avg. F-measure (%)
Actinic Keratoses (akiec)	0.92	0.37	0.52
Basal Cell Carcinoma (bcc)	0.88	0.80	0.84
Benign Keratosis (bkl)	0.68	0.68	0.68
Dermatofibroma (df)	0.71	0.62	0.67
Melanocytic Nevi (nv)	0.94	0.98	0.96
Melanoma (mel)	0.58	0.48	0.52
Vascular Lesions (vasc)	1.0	0.69	0.82

specific category, possibly due to its significant representation in the dataset and its distinct visual characteristics. On the other hand, Melanoma exhibited the lowest precision at 50%, indicating challenges in distinguishing this potentially life-threatening lesion from different types. This lower precision may result from the high variability in melanoma appearances, making it harder for the model to generalize effectively. The Vascular Lesions class, despite its more miniature representation in the validation set, showed commendable precision, suggesting that the model could effectively recognize these specific lesions. Similarly, Actinic Keratoses (representing "akiec") and Basal Cell Carcinoma (corresponding to "bcc") displayed intense precision, further validating the robustness of the hybrid model's performance across diverse lesion categories.

The variations in precision across these different classes highlight the importance of continued research and refinement in deep-learning models for skin lesion classification. While the model demonstrates significant strengths in some areas, further work is needed to improve its ability to classify complex and challenging categories like Melanoma accurately.

5.6 Comparative Analysis with Existing Works

In our study, we have achieved significant milestones in skin cancer diagnosis by applying deep learning techniques. By implementing advanced preprocessing methods such as hair removal, min–max normalization, and adaptive histogram equalization, we ensured robust feature extraction from input images, enhancing classification accuracy. Utilizing the hybrid model "RvXmBlendNet" architecture resulted in top-tier performance, demonstrating remarkable accuracy, recall, precision, and F-measure metrics. Through a comprehensive comparison of various deep learning architectures, we provided valuable insights into their strengths and weaknesses, aiding researchers and practitioners in selecting optimal models for skin lesion classification. Additionally, our effective handling of dataset imbalances using techniques like random oversampling has bolstered the model's generalization capabilities across different lesion classes, addressing a critical challenge faced by many existing studies.

The importance of the hybrid method in our study lies in its ability to combine the strengths of different architectures. While traditional individual models like VGG19 and ResNet50 can be effective, they may have limitations when applied to complex medical image classification tasks [23, 51, 53–59]. By integrating multiple architectures into a hybrid model, we can leverage the unique benefits of each component, leading to improved accuracy and robustness in classification.

Table 10 represents the comparative analysis between our proposed model and existing works. Our proposed model stands out from prior research in several key aspects. Unlike studies such as Shahin Ali et al. [24], Anand et al. [59], Al-Rasheed et al. [27], and Rashid et al. [39], which overlooked or lacked detailed discussions on preprocessing techniques, comparison of deep learning architectures, and effective handling of dataset imbalances, our study comprehensively addressed these aspects. Furthermore, unlike studies by Mohammad Fraiwan et al. [28] and Balaha et al. [29], which employed limited data augmentation and a narrow range of CNN architectures, our study adopted a comprehensive approach to model evaluation and selection, ensuring robustness and generalization across different lesion classes. Elshahawy et al. [42] utilized a hybrid model combining YOLOv5 and ResNet50, achieving an accuracy of 99.5% with metrics including precision (99.0%), recall (98.6%), DSC (98.8%), and MAP ranging from 98.3% to 98.7%. While our "RvXmBlendNet" achieved slightly lower accuracy at 98.26%, it excelled in recall with 99.37%, indicating higher sensitivity in detecting true positives. Our model also demonstrated comparable precision (98.86%) and F-measure (99.08%), with the added benefit of advanced attention mechanisms and adaptive content blocks enhancing robustness and generalization. This implies that while the overall accuracy is slightly lower, our model is better at identifying positive cases, which is crucial in medical diagnoses where missing a positive case can have severe consequences. Maniraj et al. [43] presented a hybrid approach using 3D wavelet subband fusion, achieving 99.33% accuracy on the PH2 database, with sensitivity and specificity both exceeding 90%. In comparison, our "RvXmBlendNet" achieved 98.26% accuracy on the HAM10000 dataset. Although their accuracy is slightly higher, our model exhibited superior recall (99.37%) and comparable precision (98.86%), highlighting a balanced performance in both precision and recall. The inclusion of advanced preprocessing techniques and attention layers ensures improved feature extraction and intricate pattern capture. This is particularly important in medical imaging where subtle differences can significantly affect diagnosis. Nagaraj et al. [44] integrated CNNs

Table 10 Thorough comparison of our study with other research works

Aspect	Shahin et al. [24]	Anand et al. [59]	Al Rasheed et al. [27]	Rashid et al. [39]	Fraiwan et al. [28]	Balaha et al. [29]	Our paper
Accuracy	91.93% (testing)	89.09%	93.5%	98.2%	82.9%	85.87% (MobileNetV2)	98.26%
Preprocessing techniques	Lacks detailed discussion on advanced preprocessing techniques	Lacks discussion on preprocessing techniques	Overlooks essential preprocessing techniques	Overlooks essential preprocessing techniques	Employs limited data augmentation	Employs various data scaling techniques but lacks detailed analysis of their impact	Implements hair removal, min–max normalization, and AHE
Model comparison	Could benefit from a more thorough comparison	Relies heavily on transfer learning without exploring custom CNN architectures	Neglects exploration of custom CNN architectures tailored to the task	Lacks thorough comparison with alternative methods	Limited data augmentation and reliance on transfer learning	Employs a limited number of CNN architectures and U-Net models	A fair comparison of six distinct CNN models
Handling class imbalance	Utilizes data augmentation but lacks advanced techniques like AHE	Not addressed comprehensively	Addresses with data augmentation but misses out on AHE	Applies data augmentation techniques	Imbalanced dataset with limited data augmentation	Utilizes various data scaling techniques	Addresses with data augmentation and AHE
Limitations	Insufficient discussion on preprocessing techniques and model comparison	Lack of debate on preprocessing techniques and evaluation metrics	Neglects exploration of custom CNN architectures and lacks comparison	Lacks thorough comparison and debate on ensemble approaches	Limited data augmentation and signs of overfitting were observed	Lacks detailed analysis of preprocessing techniques and segmentation approaches	Different combinations or variations could further optimize model performance

with traditional machine learning models, achieving 94.5% accuracy, with precision at 93.7%, recall at 95.2%, and an F1-Score of 94.4%. "RvXmBlendNet" significantly outperformed this model, achieving 98.26% accuracy, precision of 98.86%, recall of 99.37%, and an F-measure of 99.08%. This substantial improvement demonstrates the effectiveness of our hybrid approach in leveraging the strengths of multiple architectures. The higher precision, recall, and F-measure in our model indicate its superior capability in correctly identifying positive cases while minimizing false positives and negatives. Our "RvXmBlendNet" shows a well-rounded performance across various metrics. Its robust architecture, incorporating advanced attention mechanisms and adaptive content blocks, provides a comprehensive and effective solution for skin cancer detection, ensuring high sensitivity, precision, and overall accuracy. This model effectively addresses challenges such as preprocessing, model comparison, and handling class imbalances, offering a significant advancement in skin cancer diagnosis and classification.

5.7 Impact of Proposed Work on Medical Sector

The impact of the proposed "RvXmBlendNet" hybrid model on the medical sector can be substantial, particularly in the field of skin cancer detection. By combining the strengths of ResNet50, VGG19, Xception, and MobileNet, "RvXmBlendNet" achieves higher accuracy, recall, precision, and F-measure compared to individual models, leading to more reliable skin cancer diagnoses. Early and accurate detection can significantly improve patient outcomes, as early-stage cancer is more treatable. The model's ability to generalize better across diverse datasets reduces the risk of overfitting, resulting in more consistent performance across various patient demographics. Advanced attention mechanisms and adaptive content blocks enhance the model's capability to capture intricate patterns in dermoscopic images, reducing false positives and false negatives. Deployment of "RvXmBlendNet" in clinical settings can automate the initial screening process, allowing dermatologists to focus on more complex cases and treatment plans. The model's high performance supports large-scale screening programs, making it feasible to conduct widespread skin cancer screening in both urban and rural areas. Despite the initial training complexity, "RvXmBlendNet" is optimized for efficient inference and manageable memory usage, making it suitable for deployment in various healthcare environments, including those with limited computational resources. The ability to provide accurate diagnoses with minimal human intervention can reduce healthcare costs associated with misdiagnoses and late-stage cancer treatments. Additionally, by offering a reliable second opinion, the model can support clinical decision-making, particularly in ambiguous cases. Integration with electronic health records (EHR) systems can enable continuous learning and model updates, ensuring that the diagnostic tool evolves with new data and medical knowledge. The success of "RvXmBlendNet" in skin cancer detection can pave the way for similar hybrid models to be developed for other medical imaging tasks, such as detecting other types of cancers, identifying cardiovascular diseases, and diagnosing neurological conditions. The model's design and implementation can inspire further research into hybrid deep learning models, encouraging innovation in medical AI. Overall, "RvXmBlendNet" exemplifies the benefits of integrating multiple deep learning architectures and advanced mechanisms, setting a precedent for future innovations in medical AI, leading to better patient outcomes, more efficient use of clinical resources, and a reduction in healthcare costs.

6 Conclusion

Skin cancer remains a significant global health concern, emphasizing the critical need for accurate and timely detection methods. Deep learning models, particularly Convolutional Neural Networks and transfer learning, have revolutionized skin cancer diagnosis, offering precise tools for early detection from dermoscopic images. However, existing approaches relying solely on CNN and single transfer learning methods face limitations such as dataset generalization issues, overfitting vulnerabilities, and difficulty capturing complex patterns. Addressing these challenges, this study introduces a novel hybrid deep learning model, "RvXmBlendNet," amalgamating ResNet50, VGG19, Xception, and MobileNet architectures. Through a synergistic integration of these models and advanced mechanisms like self-attention mechanisms, our study enhances skin cancer detection accuracy. Leveraging the HAM10000 dataset, preprocessing techniques, including OpenCV-based hair removal, min–max scaling, and adaptive histogram equalization, refine image quality and feature extraction. The comprehensive comparative analysis highlights the superiority of "RvXmBlendNet" over individual models, achieving a remarkable 98.26% accuracy and showcasing its potential for earlier interventions and improved patient outcomes. Additionally, the customized CNN model and other evaluated models exhibit commendable performance, underscoring the diverse approaches to skin cancer detection. Notably, ResNet50 demonstrates robust performance with an accuracy of 97.53% and the highest precision (98%) among the models. VGG19 exhibits balanced performance across metrics, achieving an accuracy of 95.93%. Xception achieves high accuracy (97.29%) and the highest recall (97.53%) among all models. MobileNet, while slightly lower in accuracy at 94.87%, still performs well across metrics. These findings suggest that while each model has its strengths, the

hybrid approach of "RvXmBlendNet" offers superior performance, promising enhanced patient care and healthcare efficiency in skin cancer diagnosis. Moving forward, future research in skin cancer detection could focus on refining data augmentation techniques, exploring advanced attention mechanisms, and integrating multi-modal data sources. By providing a comprehensive understanding of these factors, we aim to equip medical professionals and researchers with the knowledge needed to make informed decisions in skin cancer diagnosis. This research represents a decisive step forward in advancing the field of dermatology, ultimately leading to improved patient outcomes and enhanced healthcare delivery.

Author Contributions This study was conducted collaboratively by all authors. Farida Siddiqi Prity, Ahmed Jabid Hasan, Md Mehedi Hassan Anik, and Rakib Hossain designed the study and drafted the initial manuscript. Md. Maruf Hossain oversaw the analysis of the study. Sazzad Hossain Bhuiyan, Md. Ariful Islam, and Md Tousif Hasan Lavlu handled the literature searches. All authors reviewed and approved the final manuscript.

Funding The authors declare that no financial support or grants were received for this manuscript.

Data Availability The datasets generated during this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics Approval This article does not include any studies involving human participants or animals conducted by the authors.

References

1. Zhang L, Zhang J, Gao W, Bai F, Li N, Ghadimi N. A profound learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. Biomed Signal Process Control. 2024;90:105858.
2. Naeem A, Anees T, Khalil M, Zahra K, Naqvi RA, Lee SW. SNC_Net: skin cancer detection by integrating handcrafted and deep learning-based features using dermoscopy images. Mathematics. 2024;12(7):1030.
3. Sivakumar MS, Leo LM, Gurumekala T, Sindhu V, Priyadharshini AS. Deep learning in skin lesion analysis for malignant melanoma cancer identification. Multimed Tools Appl. 2024;83(6):17833–53.
4. Rai HM. Cancer detection and segmentation using machine learning and deep learning techniques: a review. Multimed Tools Appl. 2024;83(9):27001–35.
5. Mahmoud NM, Soliman AM. Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. Sci Rep. 2024. https://doi.org/10.1038/s41598-024-59783-0.
6. Furriel BC, Oliveira BD, Prôa R, Paiva JQ, Loureiro RM, Calixto WP, Reis MR, Giavina-Bianchi M. Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review. Front Med. 2024;10:1305954.
7. Hermosilla P, Soto R, Vega E, Suazo C, Ponce J. Skin cancer detection and classification using neural network algorithms: a systematic review. Diagnostics. 2024;14(4):454.
8. Rahman MA, Bazgir E, Hossain SS, Maniruzzaman M. Skin cancer classification using NASNet. Int J Sci Res Arch. 2024;11(1):775–85.
9. Mushtaq S, Singh O. Implementing image processing and deep learning techniques to analyze skin cancer images. Int J Comput Digit Syst. 2024;15(1):1243–57.
10. Naeem A, Anees T. DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. PLoS ONE. 2024;19(3):e0297667.
11. Naqvi M, Gilani SQ, Syed T, Marques O, Kim HC. Skin cancer detection using deep learning—a review. Diagnostics. 2023;13(11):1911.
12. Leiter U, Eigentler T, Garbe C. Epidemiology of skin cancer. Adv Exp Med Biol. 2014;810:120–40. https://doi.org/10.1007/978-1-4939-0437-2_7.
13. Tang P, Yan X, Nan Y, Xiang S, Krammer S, Lasser T. FusionM-4Net: a multi-stage multi-modal learning algorithm for multi-label skin lesion classification. Med Image Anal. 2022;76:102307.
14. Nahata H, Singh SP. Deep learning solutions for skin cancer detection and diagnosis. In: Machine learning with health care perspective: machine learning and healthcare. 2020. p.159–82.
15. Mijwil MM. Skin cancer disease images classification using deep learning solutions. Multimed Tools Appl. 2021;80(17):26255–71.
16. Abayomi-Alli OO, Damasevicius R, Misra S, Maskeliunas R, Abayomi-Alli A. Malignant skin melanoma detection using image augmentation by oversamplingin nonlinear lower-dimensional embedding manifold. Turk J Electr Eng Comput Sci. 2021;29(8):2600–14.
17. Ahmad I, Amin J, Lali MI, Abbas F, Sharif MI. A novel Deeplabv3+ and vision-based transformer model for segmentation and classification of skin lesions. Biomed Signal Process Control. 2024;92:106084.
18. Wu F, Wang Z, Zhang Z, Yang Y, Luo J, Zhu W, Zhuang Y. Weakly semi-supervised deep learning for multi-label image annotation. IEEE Trans Big Data. 2015;1(3):109–22.
19. Dargan S, Kumar M, Ayyagari MR, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng. 2020;27:1071–92.
20. Jan B, Farman H, Khan M, Imran M, Islam IU, Ahmad A, Ali S, Jeon G. Deep learning in big data analytics: a comparative study. Comput Electr Eng. 2019;75:275–87.
21. Phillips M, Marsden H, Jaffe W, Matin RN, Wali GN, Greenhalgh J, McGrath E, James R, Ladoyanni E, Bewley A, Argenziano G. Assessment of accuracy of an artificial intelligence algorithm to detect melanoma in images of skin lesions. JAMA Netw Open. 2019;2(10):e1913436–e1913436.
22. Dirik AE, Sencar HT, Memon N. Digital single lens reflex camera identification from traces of sensor dust. IEEE Trans Inf Forensics Secur. 2008;3(3):539–52.

23. Bhanja S, Das A. Impact of data normalization on deep neural network for time series forecasting. 2018. arXiv preprint arXiv:1812.05519.
24. Ali MS, Miah MS, Haque J, Rahman MM, Islam MK. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. Mach Learn Appl. 2021;5:100036.
25. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging. 2016;36(4):994–1004.
26. Jain S, Singhania U, Tripathy B, Nasr EA, Aboudaif MK, Kamrani AK. Deep learning-based transfer learning for classification of skin cancer. Sensors. 2021;21(23):8142.
27. Al-Rasheed A, Ksibi A, Ayadi M, Alzahrani AI, Zakariah M, Ali Hakami N. An ensemble of transfer learning models for the prediction of skin cancers with conditional generative adversarial networks. Diagnostics. 2022;12(12):3145.
28. Fraiwan M, Faouri E. On the automatic detection and classification of skin cancer using deep transfer learning. Sensors. 2022;22(13):4963.
29. Balaha HM, Hassan AES. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. Neural Comput Appl. 2023;35(1):815–53.
30. DeVries T, Ramachandram D. Skin lesion classification using deep multi-scale convolutional neural networks. 2017. arXiv preprint arXiv:1703.01402.
31. Mahbod A, Schaefer G, Wang C, Ecker R, Ellinge I. Skin lesion classification using hybrid deep neural networks. In: ICASSP 2019—2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2019. p. 1229–33.
32. Mendes DB, da Silva NC. Skin lesions classification using convolutional neural networks in clinical images. 2018. arXiv preprint arXiv:1812.02316.
33. Dorj UO, Lee KK, Choi JY, Lee M. The skin cancer classification using deep convolutional neural network. Multimed Tools Appl. 2018;77:9909–24.
34. Kalouche S, Ng A, Duchi J. Vision-based classification of skin cancer using deep learning. In: 2015, conducted on Stanfords Machine Learning course (CS 229) taught. 2016.
35. Abbas Q, Daadaa Y, Rashid U, Ibrahim ME. Assist-dermo: a lightweight separable vision transformer model for multiclass skin lesion classification. Diagnostics. 2023;13(15):2531.
36. Xin C, Liu Z, Ma Y, Wang D, Zhang J, Li L, Zhou Q, Xu S, Zhang Y. Transformer guided self-adaptive network for multi-scale skin lesion image segmentation. Comput Biol Med. 2024;169:107846.
37. Aladhadh S, Alsanea M, Aloraini M, Khan T, Habib S, Islam M. An effective skin cancer classification mechanism via medical vision transformer. Sensors. 2022;22(11):4008.
38. Desale RP, Patil PS. An efficient multi-class classification of skin cancer using optimized vision transformer. Med Biol Eng Comput. 2024;62(3):773–89.
39. Rashid J, Ishfaq M, Ali G, Saeed MR, Hussain M, Alkhalifah T, Alturise F, Samand N. Skin cancer disease detection using transfer learning technique. Appl Sci. 2022;12(11):5714.
40. Ismail MA, Hameed N, Clos J. Deep learning-based algorithm for skin cancer classification. In: Proceedings of international conference on trends in computational and cognitive engineering: proceedings of TCCE 2020. Springer Singapore; 2021. p. 709–19.
41. Chaturvedi SS, Tembhurne JV, Diwan T. A multi-class skin cancer classification using deep convolutional neural networks. Multimed Tools Appl. 2020;79(39):28477–98.
42. Elshahawy M, Elnemr A, Oproescu M, Schiopu AG, Elgarayhi A, Elmogy MM, Sallah M. Early melanoma detection based on a hybrid YOLOv5 and ResNet technique. Diagnostics. 2023;13(17):2804.
43. Maniraj SP, Maran PS. A hybrid deep learning approach for skin cancer diagnosis using subband fusion of 3D wavelets. J Supercomput. 2022;78(10):12394–409.
44. Nagaraj P, Saijagadeeshkumar V, Kumar GP, Yerriswamyreddy K, Krishna KJ. Skin cancer detection and control techniques using hybrid deep learning techniques. In: 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN). IEEE; 2023. p. 442–6.
45. Skin Cancer MNIST: HAM10000. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000. Accessed 10 Jan 2024.
46. Li S, Zhao S, Zhang Y, Hong J, Chen W. Source-free unsupervised adaptive segmentation for knee joint MRI. Biomed Signal Process Control. 2024;92:106028.
47. Hong J, Zhang YD, Chen W. Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. Knowl-Based Syst. 2022;250:109155.
48. Hong J, Yu SCH, Chen W. Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. Appl Soft Comput. 2022;121:108729.
49. Alfano PD, Pastore VP, Rosasco L, Odone F. Top-tuning: A study on transfer learning for an efficient alternative to fine tuning for image classification with fast kernel methods. Image Vis Comput. 2024;142:104894.
50. Hu Y, Zhang X, Yang J, Fu S. A hybrid convolutional neural network model based on different evolution for medical image classification. Eng Lett. 2022;30(1):1.
51. Gökhan AKSU, Güzeller CO, Eser MT. The effect of the normalization method used in different sample sizes on the success of artificial neural network model. Int J Assess Tools Educ. 2019;6(2):170–92.
52. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. BMC Med Imaging. 2022;22(1):69.
53. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci data. 2018;5(1):1–9.
54. Garg R, Maheshwari S, Shukla A. Decision support system for detection and classification of skin cancer using CNN. In: Innovations in computational intelligence and computer vision: proceedings of ICICV 2020. Springer Singapore; 2021. p. 578–86.
55. Ramamoorthy M, Qamar S, Manikandan R, Jhanjhi NZ, Masud M, AlZain MA. Earlier detection of brain tumor by pre-processing based on histogram equalization with neural network. Healthcare. 2022;10(7):1218.
56. Saifullah S, Drężewski R. Modified histogram equalization for improved CNN medical image segmentation. Procedia Comput Sci. 2023;225:3021–30.
57. Daghrir J, Tlig L, Bouchouicha M, Sayadi M. Melanoma skin cancer detection using deep learning and classical machine learning techniques: a hybrid approach. In: 2020 5th international conference on advanced technologies for signal and image processing (ATSIP). IEEE; 2020. p. 1–5.
58. Li W, Raj ANJ, Tjahjadi T, Zhuang Z. Digital hair removal by deep learning for skin lesion segmentation. Pattern Recognit. 2021;117:107994.
59. Anand V, Gupta S, Altameem A, Nayak SR, Poonia RC, Saudagar AKJ. An enhanced transfer learning based classification for diagnosis of skin cancer. Diagnostics. 2022;12(7):1628.