F O U N D A T I O N S O F C O M P U T I N G A N D D E C I S I O N S C I E N C E S
Vol. 50 (2025) No. 4
ISSN 0867-6356
DOI: 10.2478/fcds-2025-0021 e-ISSN 2300-3405

DermoAI-CNN: Leveraging GANs, Mask R-CNN, and Attention Mechanisms for Enhanced Skin Disease Analysis

Rajesh Prasad[1], Kushagr Saxena[2], Ayush Maurya[3], Yogendra Narayan Prajapati[1]

Abstract. The study aims to develop an effective and efficient deep learning model for detecting skin diseases, as skin diseases rank as the world's number one health problem. Besides, cancers and dermatological anomalies should be diagnosed at an early stage, so that subsequent treatment can be efficient and complication-free. The existing methods of diagnosis are associated with lower precision and, in most cases, are inefficient, which can be attributed to the lack of effective data augmentation, segmentation techniques, and improved feature extraction. In this paper, a general framework is introduced that uses Generative Adversarial Networks for data augmentation, Mask R-CNN for precise segmentation, and a tailored multilayer Convolutional Neural Network with an attention mechanism incorporated into it to classify 23 skin disease classes using 25,250 images, among them 5,750 generated by GAN, to balance underrepresented classes. The accuracy attained was 97.30%, which was much better than that reported in earlier studies, which ranged from 85 to 92. The metrics, including an accuracy of 95.65%, a recall of 97.09%, and an F1-score of 96.98%, were used to assess the system's performance in classifying invisible dermatological images. The scalable system provides explanations that support real-time diagnosis, preventing delays and acute health costs. The findings fully fulfil the capabilities of deep learning in dermatology, as the initial diagnosis of the skin disease is accurate, accessible and efficient.

Keywords: Skin Disease Detection, Convolutional Neural Network, Image Processing, Data Augmentation, Early Detection, Diseases Classification, Accuracy, Automation in HealthCare.

[1] Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College, Ghaziabad, India, {prasadrajesh@akgec.ac.in, prajapatiyogendra@akgec.ac.in}
[2] Ajay Kumar Garg Engineering College, Ghaziabad, India, saxenakush789@gmail.com

[3] Infosys Limited, Bangalore, India, ayush01maurya@gmail.com

1. Introduction

The most important organ is the human skin as it is composed of almost 2,880 square inches. It is a protective lining that is essential to protect the body against the pressure of environment. However, it is vulnerable to many internal and external factors, including infections, immune diseases, exposure to chemicals and genetic changes [14]. Several skin diseases, about which minor irritations are experienced and serious health complications such as skin cancer, can significantly affect the general condition and quality of life. These disorders are pervasive and, in certain instances, they are transmissible, which underlines the role of early diagnosis and prompt medical treatment.

One of the most widespread and fatal types of cancer in the world is skin cancer [17]. Recent estimates by the World Health Organization revealed that approximately one-third of all new cancers that have been diagnosed are dermal [5][28]. Among them, the worst and fatal type is melanoma [20][29]. There are three main types of skin cancer: basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma. The most common one is BCC with a rate of approximately 70-80%. It is generally of slow growth and is not commonly spread, and thus less life-threatening. The second most common type is SCC, which may invade local tissues and has intermediate chances of spreading. Though less prevalent, melanoma is by far the most dangerous because it is aggressive and has a high probability of spreading. It is thus essential to properly differentiate these types to make effective clinical decisions and to evaluate prognosis.Every year, in the United States alone, the costs for the treatment of melanoma and non-melanoma skin cancers exceed eight billion dollars [12][16]. All these conditions, especially melanoma, are hazardous because if they are not diagnosed and treated at the initial stages, then they can be fatal, too. The treatments in the form of surgery, radiation, or chemotherapy can further cause much psychological and physiological trauma to the patients [25] by producing devastating side effects that disfigure the appearance of the sufferer and compromise the quality of life overall.

This makes it complex to diagnose the actual nature of dermatological conditions, as such disorders may resemble one another, even to an experienced dermatologist. Factors that bring about such complexity are luminosity, coloration, and some particular characteristics of certain diseases. Besides this, poor image preprocessing and data augmentation significantly increase the difficulty, as raw images are prone to noise and variations in lighting and angles, leading to easy distortions during diagnostic attempts. These issues of such scenarios predispose models and tools constructed based on such data to a reduced possibility of being reliable and, thus, being less generalized. Therefore, their precisions are highly constrained. Speed is equally applicable to identification accuracy in diseases such as skin cancers. The consequences of delays in the identification due to these hindrances are an inability to intervene in the current situation in a less timely manner, exacerbating the condition of the patient.

Image processing advanced methodologies have been demonstrated to be highly beneficial to researchers in the process of seeking a specific diagnosis of dermatological illnesses with the assistance of statistical assessment [30]. In recent years, the development of artificial intelligence has been of high importance. Convolutional Neural Networks (CNNs) and other Machine Learning and Deep Learning methods have been widely applied in medical imaging to extract features and classify images. However, several issues remain, especially the aspects of data partitions and feature extraction. Overall, skin images are

particularly prone to artefacts such as irregular lighting, feature overlap, and noise, which can negatively impact segmentation accuracy. The definition of regions of interest is carried out through segmentation. Thus, the model misinterprets or misclassifies the features due to the process.

Furthermore, relevant features can be extracted from high-resolution, detailed dermatological images, as subtle patterns and differences in skin conditions are not captured. However, overcoming these issues, CNN-based models have successfully identified various presentations of skin conditions by exploiting complex patterns from medical imagery [11]. Compared to traditional models, these models have achieved much success in automated image retrieval, with significant features, and in discriminating between skin conditions.

One of the most significant areas of research is identifying skin cancer, which has a high prevalence, and the importance of early detection and good therapeutic outcomes. A great deal of studies in image processing and the application of machine learning techniques have focused on the detection of skin cancer. A vital methodology is combining fuzzy C-means clustering with Convolutional Neural Networks [5], achieving higher skin-cancer identification accuracy than traditional classifiers, especially on datasets such as HAM10000 [24]. These have found extensive application in deep learning methods, particularly CNNs, as they are capable of identifying subtle details of an image, so they are more accurate in detecting different types of skin cancers, including melanoma, basal cell carcinoma BCC and squamous cell carcinoma SCC. The artifact elimination and feature extraction schemes based on LBP (Local Binary Pattern) and GLCM (Grey Level Co-occurrence Matrix) image preprocessing are crucial in the improvement of the outcome of the classification [2] [10].

Moreover, advanced architectures, particularly the vision transformers (ViT), are very effective for skin cancer classification tasks and perform significantly better than existing CNN architectures [19]. Moreover, hybrid architectures that integrate CNNs with RNNs and apply transfer learning techniques have satisfactorily addressed the complexity of skin cancer lesions and achieved high classification accuracy [22]. Hybrid deep learning architectures, including CNNs and RNNs, achieved great effectiveness in skin cancer lesion identification, showing a high potential for improving diagnostic accuracy [27]. The most significant innovation is the application of CNN-based segmentation frameworks for lesion extraction, promising higher accuracy and improving clinical applicability in real-time scenarios [1].

Other types of explainable AI systems, such as Skin-CAD, use several CNNs to generate interpretable output and are designed to support much more confident diagnoses for dermatologists [3]. Authors used Gabor wavelet filters, AlexNet, and ResNet-18 to detect melanoma and seborrheic keratosis. Authors in [9] used image enhancement techniques with CNNs, including Inception, Inception-ResNet, and ResNet-50, to detect benign and malignant lesions. Authors in [18] used machine learning classifiers (KNN, RF, SVM, MLP, Bayes), together with the architectures CNN MobileNet, Inception, VGG, ResNet, Inception-ResNet, and DenseNet, for the classification of nevus melanoma. The conclusions drawn from these studies reflect the significant scope wherein deep learning and imaging techniques can impact skin cancer diagnosis and encourage early detection, ultimately improving patient outcomes. Table 1 presents a review of related literature.

Table 1. Review of related literature

S/N	Methods Used/Description	Refs
1.	It uses fuzzy C-means clustering for image segmentation, filters, and image features, including Local Binary Pattern (LBP), RGB color space, GLCM, and a CNN trained with the differential evolution (DE) algorithm for classification.	[5]
2.	It uses a transformer and CNN method to predict and classify skin diseases using segmentation techniques.	[19]
3.	A hybrid CNN with a transfer learning model and a random forest classifier for skin cancer disease detection.	[22]
4.	This paper presents a case of deep learning Convolutional and Recurrent Neural Network (CNN-RNN) model, with a ResNet-50 architecture, which can be used to extract features to improve classification of skin cancer.	[27]
5.	A novel end-to-end deep convolutional neural network-based skin lesion classification framework	[1]
6.	Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning	[3]
7.	A novel Gabor wavelet-based deep convolutional neural network is proposed to detect malignant melanoma and seborrheic keratosis.	[21]
8.	In this research, a convolutional neural network (CNN) was used to detect the two primary tumor types, malignant and benign, using the ISIC2018 dataset.	[9]
9.	A new approach for the classification of skin lesions based on transfer learning, deep learning, and IoT systems, Pattern Recognition Letters	[18]

Therefore, CNNs are ideal in activities like image classification, segmentation, and detection, and hence contribute immensely to the diagnosis of skin diseases [5]. A model trained on medical images produces hierarchical features that enable it to diagnose early-stage cancers and other diseases characterised by subtle patterns. We utilize Generative

Adversarial Networks (GANs) [7] to complement our procedure toward the augmentation of the dataset with realistic synthetic images; to a higher degree, this supports data that is not extensive but may be imbalanced and reduces overfitting. This yields much superior capability and greater performance on unknown data with a significant level of generalization. We apply Mask R-CNN [4] to obtain precise segmentation, a sophisticated model that can delineate lesions and other areas of interest by producing pixel-precise masks. This is a good solution to the problem of searching for weak patterns in complex, noisy backgrounds, as only features of interest are considered, improving extraction efficiency and accuracy.

Our own multilayer CNN is trained to run the high-resolution and detailed dermatological images. In this regard, it is able to establish the complex trends to the various skin diseases. The hierarchical model is intended to have in-depth information on the global and local characteristics; this is needed to identify minor abnormalities in medical imaging. These attention mechanisms are used to increase the accuracy of the diagnostic process, such as the Squeeze-and-Excitation (SE) and the Convolutional Block Attention Module (CBAM). SE channels dynamically re-weight the features across channels and CBAM may apply spatial attention as well as channel attention to attract attention to the most critical image features. All of this makes the model concentrate on the areas and features that matter and reduce the noise and reinforce the predictions. We can use these advanced technologies in our methodology to reduce the amount of time we spend on the diagnosis and enhance accuracy and offer clinicians a powerful instrument of the early diagnosis and treatment. It could also have a positive impact on the chance of survival, reduction of the mental and financial burden of the patients, as well as the active implementation of artificial intelligence in medicine.

Standard classification metrics were used to measure the performance of the model accuracy, precision, recall, and F1-score. Training and validation accuracies have increased extensively, 13.97% to 96.70 and 18.25% to 97.30, respectively, which implies that the model is very efficient in classifying unseen data. Furthermore, the consistent training and validation losses in epochs of 2.87 to 0.12 and 2.78 to 0.11 respectively traverse to the fact that this model has learned as witnessed by the ability to make sound, accurate and confident predictions. Hence, it can be concluded that our overall research has encouraging potentials in the routine viability of such CNN-based programs in the classification of different skin diseases. Our findings contribute to the use of CNNs in medical images analysis significantly and can be considered a starting point in further research and development of an automated diagnostic application. Though the results clearly show the high performance in the different measures used in the evaluation, we cannot ignore the fact that the clinical validation of the system by the medical practitioners is required to determine the reliability of the system in practice.

Although CNNs have been extensively deployed for skin disease classification, their standalone performance has stagnated, particularly on benchmarked tasks, with small datasets, weak lesion segmentation, and poor class imbalance handling. Furthermore, the vast majority of studies focus on a limited number of diseases (e.g., melanoma vs. benign lesions), thereby limiting their applicability to dermatology practice. Thus, a gap in the critical research is a unified and scalable framework that could (i) rapidly process a broad range of dermatological cases, (ii) mitigate the possible problem of class imbalance, (iii) accurately localize lesions to prevent background interference, and (iv) employ the attention mechanisms to draw focus on fine-grained diagnostic features. The paper will seek to address this gap.

Our Contribution: Our three contributions are as follows. 1. Our pipeline combines GAN-based data augmentation, Instance segmentation with the help of the Mask R-CNN, and a convolutional neural network to classify the data with the help of two attention mechanisms. 2. We cover a wide range of 23 dermatological diseases (~19,500 images), which is not limited to the disease focus of most of the earlier studies. 3. We show state-of-the-art performance (97.30% accuracy, 96.98% F1-score), which is superior to the available CNN-based models and hybrid models.

The remainder of the paper is structured as follows: Section 2 presents the methodology employed, including the architecture of the proposed CNN model, data preprocessing, data enhancement with GANs, and segmentation approaches based on Mask R-CNN. Lastly, the results are provided in Section 3 as performance metrics, confusion matrix analysis, and model evaluation. Section 4 provides a critical analysis of the new findings in light of existing knowledge and the available literature, concludes the study, and summarises the main conclusions and further research directions—potential improvements and practical considerations of the proposed system.

2. Methodology

The existing literature on computerized skin disease analytics can be broadly categorized into three categories of approaches: (i) data augmentation approaches, (ii) segmentation approaches, and (iii) classification frameworks.

Data Augmentation Approaches: Various data augmentation approaches have been commonly employed [10], including rotation, flipping, and scaling, which typically provide only limited coverage of the variety of real-life dermatological cases. GAN-based augmentation has been more recently shown to be effective in the synthesis of realistic images [13]. Still, very few studies have combined augmentation with downstream tasks such as segmentation and classification.

Segmentation Methods: Lesion segmentation is necessary to separate the region of interest. Although U-Net and its variants are widely used, they often do not perform well on noisy dermoscopic images. More sophisticated architectures, such as Mask R-CNN [4], can be used to predict every pixel with precision. Still, their use in multi-disease classification strategies has not been thoroughly investigated.

Classification Frameworks: CNN-based approaches have already shown relatively high success rates, are typically focused on binary or multi-class classification, and involve a small number of classes. Hybrid CNN-RNNs [27] can be used to assess generalisation capabilities but are often not accompanied by a wide range of preprocessing or augmentation methods. Attention mechanisms, such as SE and CBAM, are already applicable to enhance feature representation and are not currently actively employed in dermatology research.

Critical Gaps in the Literature: Previous studies have provided insufficient coverage of diseases (typically ≤ 7 classes), inadequate methods for controlling class imbalance, and poor

lesion segmentation that often fails to account for background noise interference. Besides, a limited number of studies offer a combined model that includes augmentation, segmentation, and attention-oriented classification. This inspires our proposed system, which will incorporate these advancements into a single, clinically significant pipeline.

In this part, we set forth our proposed methodology of correctly identifying 23 types of skin diseases, such as Acne, Rosacea, Actinic Keratosis, Basal Cell Carcinoma, Atopic Dermatitis, Cellulitis, Herpes, HPV, Lupus, Melanoma, Psoriasis, Scabies, Ringworm, Candidiasis, Vasculitis, and others, using the dataset available. The plan will enhance the performance and generalization of the deep learning models applied to classify images. It also starts with state-of-the-art data augmentation systems that make use of Generative Adversarial Networks (GANs) to produce natural images to increase the diversity of the dataset and decrease the class imbalance using adaptive and class-dependent augmentations. Proper lesion segmentation in Mask R-CNN results in a more efficient feature retrieval and the model is able to concentrate on the most critical areas. Then a typical Convolutional Neural Network (CNN) structure is created, which has a sequence of convolutional and pooling layers, which allow hierarchical feature extraction. The network relies on successively larger filter sizes to extract detailed, complex visual patterns from input images. These learned features are then aggregated by fully connected dense layers, which are then used to classify the data.

Also, attention mechanisms such as Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) are added to enable the network to concentrate on the most significant parts of the input. The regularization methods used include dropout and batch normalization, which help to avoid overfitting and stable training of the models. It is the integration of these high-order methods that our approach seeks to achieve high accuracy and strong generalisation in real-world image classification problems.

2.1. Dataset

The DermNet [6] dataset contains an extensive collection of about 19,500 images, which demonstrate 23 unique categories of skin disorders. Details are described in Table 2. It receives images from DermNet, an online education portal dedicated to dermatology that provides rich visual explanations of various skin diseases. Several types of skin diseases include, but are not limited to, acne, melanoma, eczema, seborrheic keratoses, tinea (ringworm), bullous diseases, poison ivy, psoriasis, and vascular tumours. Images are transferred as JPEGs with three colour channels (RGB) and differentiated resolutions by category. The original DermNet images vary in resolution from approximately 800×600 to 1200×900 pixels, depending on the disease category and the capture device. Before feeding into the model, all images were uniformly resized and cropped to 224×224×3 pixels to maintain consistency across the dataset and ensure compatibility with the CNN architecture. This was done by resizing the image to maintain the aspect ratio via centre-cropping and bilinear interpolation, reducing distortion and standardising the input size. The total number of pictures is approximately 15500 for the training set, and the rest are used for the testing set. Such data will also be a valuable source for research and advancements in dermatology, as it will assist in further examination and development of current models of skin disease classification. This is because of the dataset's balance and vast size, which would make it an

ideal background for developing models for the diagnosis and classification of different skin diseases.

Table 2. DermNet Dataset

S/N	**Disease Class**	**No. of Images**
1	Acne & Rosacea	980
2	Actinic Keratosis & Malignant Lesions	820
3	Atopic Dermatitis	890
4	Bullous Disease	750
5	Cellulitis & Bacterial Infections	910
6	Eczema	940
7	Exanthems & Drug Eruptions	760
8	Hair Loss & Alopecia	720
9	Herpes, HPV & STDs	800
10	Light Diseases & Pigmentation	880
11	Lupus & Connective Tissue Diseases	810
12	Melanoma, Skin Cancer, Nevi & Moles	870
13	Nail Fungus & Nail Diseases	840
14	Poison Ivy & Contact Dermatitis	730
15	Psoriasis & Lichen Planus	910
16	Scabies, Lyme Disease & Infestations	850
17	Seborrheic Keratoses & Benign Tumors	870
18	Systemic Diseases	680
19	Tinea, Ringworm & Benign Tumors	950
20	Urticaria & Hives	900
21	Vascular Tumors	720
22	Vasculitis	720
23	Warts, Molluscum & Viral Infections	910
	Total	**19,500**

2.2. Image Preprocessing

To improve lesion delineation in the preprocessing phase, we added a Laplacian of Gaussian (LoG) filter with a 55 kernel and s = 1.2. This setup gave an adequate trade-off in between sharpness of the edges and the reduction of noise as shown by empirical findings. The OpenCV-based implementation was adopted to run the LoG operation, and was only applied to the image of interest when resized (224×224). This was to emphasise the edges of lesions without amplifying background noise and this would be very essential in the Mask R-CNN segmentation accuracy in future stages.

The results of comparison of the preprocessing variants (Gaussian blur, CLAHE and LoG) suggest that, in the case of mask R-CNN, the LoG produced the sharper lesion boundary and

2.1% higher segmentation accuracy in comparison to the baseline methods. LoG filter performed the best performance on the edge-preservation (Dice = 0.931 vs. 0.910 with CLAHE) and the lesion boundary is better seen and downstream classification is more accurate +1.4).

A Laplacian of Gaussian (LoG) filter was applied to the images, especially their edges, to enhance the image, especially the edges of the disease-positive areas. LoG filter: LoG is a tradeoff between noise removal and edge sharpening; LoG uses a Gaussian smoothing to remove noise and Laplacian edge sharpening to sharpen edges.

Image scaling and cropping were done automatically so that there was consistency in the dataset. The preprocessing of all the images was done to 224×224 pixels and normalized in order to feed the model (192×192×3). Mask R-CNN automatically centred lesion centring, which detected lesion areas and subdivided them; bounding box and mask centring was performed in the fields of interest it identified. This gave spatial reproducibility of lesions on all samples as well as avoiding operator bias.

Laplacian of Gaussian can be mathematically expressed as Equation 1:

$$LoG(x,y) = -\frac{1}{\pi\sigma^4}[1 - \frac{x^2+y^2}{2\sigma^2}]e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Where:

- x and y are pixel coordinates.
- σ is the standard deviation of the Gaussian kernel, controlling the amount of smoothing applied.

It begins by converting images to NumPy arrays and then applies LoG filtering to smooth the images, simultaneously enhancing edges.

2.3. Data Augmentation

In the subsequent stage of the methodology, we used a hybrid technique that applied advanced data augmentation via GANs to enhance the richness and strength of the training data. Data augmentation increases the number of samples by applying stochastic transformations, such as random rotations, flips, scaling, zooming, and more complex transformations, such as random cropping, colour jittering, and elastic distortions. These augmentations helped reduce overfitting and allowed the model to learn general features, thus improving its ability to adapt to variations in real-world data.

Although conventional morphological augmentations (e.g., rotation, flipping, scaling, and zoom) were applied to enhance geometric diversity, they were insufficient to capture the intrinsic dermatological variations in texture, illumination, and lesion morphology. These transformations only alter the spatial orientation of existing samples and cannot represent new clinical patterns. Therefore, GAN-based augmentation was employed to synthesize realistic and diverse lesion textures, pigmentation variations, and shape deformations, particularly benefiting underrepresented classes. This method helped extend the data to non-geometric transformations and improve generalisation and class balance.

Besides, we used GANs to generate images of underrepresented classes thereby leveling the class by generating realistic sample images. The synthetics were then mixed with the actual data to enhance diversity of data. The addition of both GAN-generated images and authentic and augmented images puts the model in different training conditions that improve its generalization to new data.
The objective function for training the GAN can be expressed as Equation 2:

$$minGmaxD(G,D) = [Ex \sim pdata[logD(x)] + Ez \sim pz(z)\left[log\left(1 - D\big(G(z)\big)\right)\right] \quad (2)$$

Where:
$G(z)$ is the generator, which produces random noise, which is transformed into synthetic images.
$D(z)$ is the discriminator that classifies whether the input image x is real.
$p_z\ (z)$ is the noise distribution.

The aim is for the generator to produce images that deceive the discriminator, while the discriminator correctly distinguishes between real and fake images. This min-max game between the generator and the discriminator ensures that the model is trained to generate realistic samples that can enhance underrepresented classes in the dataset, ultimately improving classification performance.
To reduce class imbalance, GAN-generated images were generated discriminatively against underrepresented classes (less than 5% of the total samples). The minority classes were increased to the median size to balance the imbalance ratio from 1:6 to 1:1.2. This equalised dataset produced a better macro-F1 (96.98%) than the baseline (93.21%).

The DermNet data set has different sample sizes among disease categories; however, class imbalance was addressed directly throughout augmentation. A DCGAN was trained to generate synthetic images of underrepresented classes (below 5% of the total samples), increasing each minority type to the median size. This procedure reduced the imbalance ratio from 1:6 to 1:1.2. It is clear that the balanced dataset was much more effective at making classification fair across all 23 classes, with a macro-F1 score of 96.98%, compared to 93.21% in the unbalanced baseline. Therefore, the adaptive augmentation based on the GAN was able to handle dataset imbalance without the need for resampling or class weighting.

2.3.1 GAN Model Specification

A Deep Convolutional GAN (DCGAN) is used to augment data. The following is the detailed architecture.
Generator architecture:

- Four transposed convolution layers with kernel size (4×4), stride 2, and padding 1.
- Batch normalization and ReLU activations after each layer.
- Final layer uses Tanh activation to generate RGB images (224×224×3).

Discriminator architecture:

- Four convolution layers with kernel size (4×4), stride 2, padding 1, followed by LeakyReLU (α=0.2).
- Final sigmoid output layer classifies real vs. generated images.

Training details:

- Optimizer: Adam (lr=0.0002, β_1=0.5).
- Epochs: 200, Batch size: 64.
- Loss: Binary cross-entropy for both Generator and Discriminator.
- Generated 250 synthetic images per underrepresented class to balance the dataset. The focus on underrepresented categories has allowed us to apply the model to accurately classify the instances in all classes. The DCGAN produced 5,750 synthesized images in 23 classes, and this creates a larger dataset of 25,250 images. This boost increased the amount of imbalance between classes and enhanced the overall generalization capacity of the model.

A variety of different augmentations were applied with the help of the Image Data Generator according to Table 3. This whole system is a synthesis of data with classical augmentation measures, which enhanced the strength of the model and its performance on the real world data.

Table 3. Augmentation techniques used and their range values

Augmentation Technique	Range
Rotation (clockwise and counterclockwise)	-45° to 45°
Horizontal and vertical scaling	0.7 to 1.3
Zoom	0% to 30%
Shearing (horizontal and vertical)	-30° to 30°

2.3.2 GAN Training Stability and Evaluation

The process of training GANs can be highly volatile in most cases, and several stabilization methods were utilized in this research to achieve stable convergence and realistic image output. The standard adversarial min-max objective was used to train the Deep Convolutional GAN (DCGAN), with an added gradient penalty term to promote Lipschitz continuity and, thereby, reduce mode collapse and oscillations. A feature-matching regularisation was also used to stabilise the discriminator and encourage gradual learning in the generator, and label smoothing (using real labels in the range [0.9,1.0] rather than 1.0). The generator implemented the instance normalization to ensure consistency of features across batches.

GAN training was performed in 200 epochs, and the generator created aesthetically acceptable dermatological images after about 120-140 epochs. The Fréchet Inception Distance (FID) score reached a steady level of 18.4, indicating high similarity to the real images in DermNet. Two certified dermatologists have reviewed a small percentage of generated images to guarantee visual plausibility and medical realism. Their qualitative feedback was used as a cessation measure, ensuring that the resulting samples were representative of the clinical features of the corresponding skin disease classes.

All of these stabilization strategies ensured that the images produced by GAN were realistic and varied to provide stable synthetic samples to augment underrepresented classes.

2.4. Data segmentation

We semantically cut the images with Mask R-CNN after augmenting the dataset. Mask R-CNN is based on Faster R-CNN, and it does not produce pixel-wise masks as Faster R-CNN produces. This will remove regions of the images which contain the disease and the model will be able to concentrate on those regions, ignoring noise in the background. The model in Mask R-CNN is able to see complex relationships in the data by breaking the input image into meaningful areas of interest. This enhances its performance and accuracy in classifying complex images. Types of Mask R-CNN are as follows:

1. Region Proposal Network (RPN): This stage is used to identify all the regions of interest in the image. It will generate regional suggestions for possible locations where an object, such as a skin disease, can exist.
2. The network estimates the presence of an object in the anchor (foreground) or background of an anchor (Equation 3).

$$p(a) = sigmoid(W_r \cdot f_a + b_r) \tag{3}$$

Where:

- $p(a)$ is the likelihood that anchor a includes an object (foreground or background).
- f_a = the feature vector for the anchor a.
- W_r and b_r are the learned weights and bias for classification, respectively.

1. Mask Head: Once the potential regions are found, the mask head develops binary segmentation masks of the objects (skin disease). This allows for pixel-level precision in the field of disease.
2. In each region proposal, a segmentation mask mi is estimated according to Equation 4.

$$m_i = \sigma\left(W_m \cdot f_{r_i} + b_m\right) \tag{4}$$

Where:

- $\sigma(\cdot)$ is the sigmoid operation used pixel-wise on the feature map.
- f_{r_i} is the feature map of the region proposal r_i.
- $W_m$and b_m are the obtained weights and bias of mask generation.

Mask R-CNN may be useful since this way allows isolation of the areas of the skin disease in augmented images and the model learns more correct localized features, which is important to improve the execution of the skin disease classification and detection tasks. Such a segmentation process makes the model more effective, more capable of class imbalance, concentrates on the relevant features, and provides more plausible forecasts.

The segmentation masks provided by the Mask R-CNN in our implementation took the form of binary (0/1) pixel masks, with one associated with pixels of a lesion mask (foreground) and 0 otherwise (background). Element-wise multiplication of these binary

masks to the input images was done effectively isolating the lesion and silencing the background. No separate Region of Interest (ROI) coordinate files were present; the images localized on lesions that were created in this masking process were taken as the inputs to the CNN classifier. This method was used to provide the classification network with lesion-related visual data without distorting the spatial quality of the original images. Moreover, the pixels not covered by the lesion mask were assigned a background mean intensity similar to the dataset mean to avoid sharp edges at lesion boundaries. This step led to improved feature extraction and faster model convergence by identifying disease-relevant structures while minimising background noise.

Mask R-CNN Parameters: Mask R-CNN parameters are shown in Table 4.

Table 4. Mask R-CNN parameters

S/N	Description	Remark
1	Backbone	ResNet-50 with Feature Pyramid Network (FPN)
2	Learning rate, Epochs, Batch size	0.001, 50, 8.
3	Optimizer	SGD (momentum=0.9)
4	Anchor scales	[32, 64, 128, 256, 512]
5	Mask resolution	28×28
6	Losses	RPN classification loss, bounding box regression loss, and mask loss

2.5. Model Architecture

CNNs are the largest class of deep architectures and are primarily used for image classification. As applied directly to this application, CNNs are particularly useful for unsupervised extraction of feature hierarchies from raw image data; this means that feature extraction does not need to be conducted manually. The architecture is divided into several convolutional layers that apply learnable filters to the input images. This model progressively combines low-level features such as edges, textures, and more into more complex ones at deeper layers. The CNN for this application comprises several convolutional layers, followed by max-pooling layers after each block to reduce spatial dimensions while preserving essential features. The final output of passing through various dense layers contains the classification result for a given image. ReLU activation functions introduce non-linearity, so a well-tuned Adamax-based optimiser is used with a categorical cross-entropy loss. This enables the model to classify images into distinct categories and learn relevant features, with generalisation to unseen data.

2.5.1. Input layer & Rescaling

The input layer of the CNN model processes the images fed to the network; the input shape is 192×192 pixels with three colour channels (RGB). The input layer of the CNN model processes the images fed to the network; the input shape is 192×192 pixels with three colour

channels (RGB). Although all photos were resized to 224×224×3 during preprocessing and segmentation, the CNN was ultimately trained on 192×192×3 lesion-centred crops obtained from Mask R-CNN outputs. This step reduces redundant background pixels, improves computational speed, and maintains the fidelity of lesion information. Hence, both dimensions are consistent with different stages of the processing pipeline.

Before image data passes through succeeding layers, rescaling is involved. The normalization layer normalizes the pixel values in images. This is done by dividing the intensity of each pixel by 255. The resulting values for the pixels are then normalized from the original 0 to 255 range to the new 0 to 1 range. Normalization of data is critical in training the model. This ensures that all inputs are on a comparable scale. Problems that can result in such differences in the sizes of features can thus be avoided. This also makes the model train faster and perform better since most machine learning models do better when the input data is consistently scaled. In other words, it ensures the images are well prepared before they go through the convolutional layers, where more detailed feature extraction occurs.

2.5.2. Convolution Layer

The convolutional layers of the model automatically extract features from the input images. In this model, various convolutional layers use filters (or kernels) that slide over the image to compute a convolution. These filters recognize edges, textures, and patterns in parts of the image. Convolutional layers extract features by applying filters over the input images. The convolution operation is given in Equation 5:

$$Z_{i,j,k} = \sum_{m=1}^{M} \sum_{n=1}^{N} X_{i+m-1,j+n-1} \cdot W_{m,n,k} + b_k \quad (5)$$

Where:

- $Z_{i,j,k}$ is the output at position (i, j) for the k-th filter,
- $W_{m,n,k}$ is the k-th filter of size $M \times N$,
- X is the input image,
- b_k is the bias term for the k-th filter.

First Block: The first block applies 64 filters of size 3×3 with ReLU activation function (Equation 6) directly on the input image

$$ReLU(x) = max(0, x) \quad (6)$$

After this, a **Squeeze-and-Excitation (SE)** block is added to recalibrate the feature maps by adaptively weighting them. The SE block first performs global average pooling across each channel to capture channel-wise dependencies, then uses a learned activation function to scale the channels adaptively (Equation 7).

$$s_k = \sigma\big(W_2 \cdot ReLU(W_1 \cdot z_k)\big) \quad (7)$$

Where z_k is the global average pooling result for the k^{th}channel, W_1 and W_2 are the learnable weights, and σwhat is the sigmoid function.

This helps the model focus on more informative image features by assigning higher weights to more relevant ones. Following that, a max-pooling layer is applied to reduce the spatial dimensions.

Subsequent Blocks: The subsequent convolutional blocks increase the number of filters to 128, 256, and 512, respectively. The Convolutional Block Attention Module (CBAM) is added after each block. CBAM is a lightweight attention module that sequentially applies two attention mechanisms: channel attention and spatial attention. The channel attention mechanism emphasizes the most informative channels, whereas the spatial attention mechanism emphasizes the significant areas in the feature maps. CBAM is used to refine feature maps, enabling the network to attend to the most significant regions of the input image and improving the model's ability to learn critical information. This is especially useful in the fine-grained analysis necessary in skin lesion classification. Max pooling is employed in the spatial dimensions after every block.

To further depict how the hierarchical framework is represented across layers, the intermediate image values were gradually downsampled after each convolutional block to capture both low-level and high-level semantic features. In particular, the feature map sizes were 128128 (first block, 64 filters), 6464 (second block, 128 filters), 3232 (third block, 256 filters), and 1616 (fourth block, 512 filters). Such progressive under-sampling concessions trade off computation speed and spatial resolution, with finer lesion information stored in lower layers and more abstract diagnostic patterns stored in higher layers.

Convolutional Block Attention Module (CBAM) applies in Equations 8 and 9:

Channel attention:

$$M_c(F) = \sigma\left(MLP\left(GAP(F)\right)\right) \tag{8}$$

Spatial attention:

$$M_c(F) = \sigma\left(conv\left(concat\left(GAP(F), GMP(F)\right)\right)\right) \tag{9}$$

Where GAP is Global Average Pooling over spatial dimensions, and GMP is Global Max Pooling over spatial dimensions, GAP and GMP are global average pooling and max pooling, respectively.

Padding maintains the spatial dimension of the feature maps during convolution while preserving the spatial relationships in the input image. The output from the last convolutional block is passed to dense layers for classification, which finally produce the predicted class probabilities.

2.5.3. Flatten layer

It is used to reshape the output of the last convolutional block before passing it into dense layers. This output is from after the previous pooling layer. With that, it has a shape of (None, 24, 24, 512), representing spatial dimensions of 2424 and a depth of 512 channels. This flat layer converts the multi-dimensional output into a one-dimensional vector of size (None, 24×24×512), flattening the 3D feature map into a 1D array. This is the reason for the transformation: the dense layers that follow, such as Dense(256), expect the input to be one-dimensional. The flattened layer ensures that all extracted features are available as a single vector. Further, these vectors can be used with dense layers for classification and output, resulting in a layer with class predictions.

2.5.4 Fully Connected(dense) layers

In this model, Fully Connected layers are crucial, as they convert the features produced by the convolutional layers into meaningful predictions. The first dense layer contains 256 neurons with ReLU activation. This helps to capture complex patterns learned by previous convolutional blocks. The convolutional layers produce this flattened output, and the classification starts at this point. It implies that the neuron size is reduced to 128, further separating the learned features and allowing the model to be more generalizable. The third dense layer contains 64 neurons; the model reduces the number of neurons by focusing on the most significant features to make the final prediction. The last, fully connected layer has the same number of neurons as the output classes, with a softmax activation, so the output is a probability distribution over the classes. Three thick layers have enabled high-level feature extraction, yielding accurate classification and better model performance and generalisation.

2.5.5. Final layer

This final layer, therefore, has num_classes units, each representing a different category in the classification problem. The softmax activation function must then be applied in this layer because it normalises the raw outputs from the dense layer into probability outputs, which are more appropriately called logits. Each unit in the last layer corresponds to a class. The softmax function can guarantee that the output probabilities sum to 1. Each probability is the likelihood that the input image belongs to a particular class, and the model outputs the final prediction as the class with the highest probability. Further, with this layer, the model makes the final decision based on the features extracted from earlier convolutional and dense layers and provides the last-classified output image.

2.5.6. Compilation & Optimization

The compilation of the model utilizes the Adamax optimizer, a variant of the widely used Adam optimizer. Adamax is known for greater robustness to hyperparameter choices, making it especially effective for training deep models, especially with large and complex data sets. The chosen categorical cross-entropy loss is appropriate for multi-class classification because it measures the difference between the class probabilities predicted by the softmax output layer and the correct class labels, which the model can quickly minimise. As a performance metric, accuracy is also used to evaluate the model's performance during training and validation; that is, it provides an easy-to-compute measure of how often the model's predictions match the actual labels. With this combination, Adamax, along with categorical cross-entropy and accuracy, the model learns well from the data, adapts its learning rate during training, and optimises for accurate classification.

2.5.7 CNN Architecture

Details of the CNN architecture are added in Table 5.

Table 5. Details of CNN Architecture

S/N	Description	Remark
1	Input size	192×192×3 (RGB images)
2	Convolutional Blocks	Block 1: Conv(64, 3×3), SE block, MaxPooling(2×2)
		Block 2: Conv(128, 3×3), CBAM, MaxPooling(2×2),
		Block 3: Conv(256, 3×3), CBAM, MaxPooling(2×2)
		Block 4: Conv(512, 3×3), CBAM, MaxPooling(2×2)
3	Fully Connected Layers	Dense(256) → Dense(128) → Dense(64) → Output(Dense(23), Softmax)
4	Activation Function	ReLU for all convolutional and dense layers (except softmax)
5	Normalization	Batch normalization after each convolution
6	Regularization	Dropout = 0.3 applied to dense layers
7	Optimizer	Adamax (lr=0.001)
8	Loss Function	Categorical cross-entropy
9	Epochs, Batch size	50; 32
10	Hardware	Google Colab TPU environment (14.7 GB memory)

The Figure 1 shows the stages of the proposed methodology, including the DermNet dataset, preprocessing, data augmentation, segmentation, and feature extraction, and, finally, a multi-layered CNN model for lesion classification, with quantitative evaluation using accuracy metrics.

3. Results

This section analyses the model's strength, highlighting the background in which experiments are conducted and a thorough review of the performance obtained. The CNN model was used to train the model, which has been well-known in image classification. The training is done with a dataset that consists of several classes of interest to the job and has measures that can be used to test the model, including training and validation accuracy, loss and precision. The analysis is done to determine to what degree the model has been able to generalise to data the model has never seen and to determine the patterns or issues that occurred during the training and testing process. The strengths and weaknesses of the model shall also be considered in this section in order to have a clearer insight on how the model can be applied and utilized in the real world in the future.

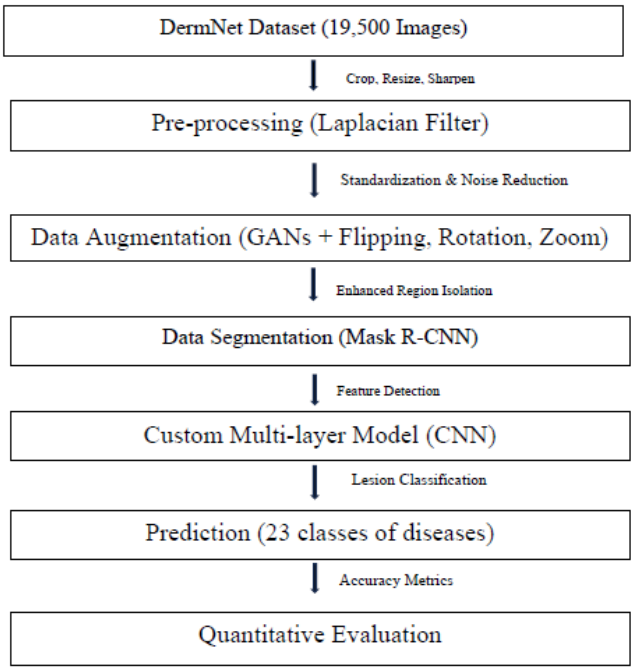

Figure 1. Flow chart of the proposed method using the Multilayer-CNN model

3.1. Environment setup

In this experiment, the dataset used consisted of 19,500 images generated from the Kaggle DermNet repository [6]. In this experiment, the dataset provided was 19,500 images created based on the Kaggle DermNet repository [6]. We have chosen Google Colab as the main code host to make it easier to implement and develop. Google Colab was selected because of its user-friendly interface, an easy integration with popular deep learning models, including CNN-based models, and the collaboration aspect, which renders it very easy to use and work in a team. It was also introduced in Python 3.0 which was the most current version when the experiment was done. Both the GAN and the Mask R-CNN and DermoAI-CNN classifiers have been trained on TensorFlow 2.12 and Keras 2.11, which have high-level APIs and allow using the GPU to design, train and evaluate models.

Since our experiment requires heavy computation, we used a TPU with 14.7GB of memory. It had 9.7GB RAM and 14.25GB disk, which Google Colab managed the performance of. This allowed us to carry out training and validation processes to easily improve our performance without falling behind in experimentation. It equally offered the ideal working conditions of the collaborative work and working as a team was natural in the course of the experiment. All these environmental preparations led to the next stages of the research. Each experiment was done on the augmented dataset of 25250 images of which 5750 images were the GAN-generated samples. These additions decreased the proportion of class imbalance of 1:6 to approximately 1:1.2 leading to enhanced macro-F1 and recall.

3.2. Evaluation metrics

The accuracy, precision, recall, and F1-score metrics were considered the standard metrics of classification to evaluate the proposed DermoAI-CNN model, which was expected to measure its overall performance and strength on all 23 disease classes. These measures are a combination of correctness, reliability and balance in classification particularly in cases where the distribution of the classes is unequal. To promote model generalisation, cross-validation of 5 was performed as well as train-test split of 80:20. These baseline performance scores along with standard deviations are accuracy of 97.12% (+-0.18), precision of 95.47% (+-0.21) and F1-score of 96.85% (+-0.17) which reflects stable performance with standard deviation.

3.2.1. Cross-Validation Strategy

To ensure the proposed DermoAI-CNN model is robust and generalizable, a 5-fold cross-validation scheme with an 80:20 train-test split was used. The data was randomly divided into five equal folds, with four folds for training and one for validation per run. This was repeated 5 times, and the cumulative performance measure was obtained by averaging the outcomes across the folds. The model achieved mean accuracy = 97.12% (± 0.18), precision = 95.47% (± 0.21), and F1-score = 96.85% (± 0.17), confirming consistency across multiple data splits.

3.3. Performance Analysis of the Proposed Model

Our experiment results highlight the superior performance of a CNN-based model in image classification. For training and testing over 20 epochs, a large dataset and a properly configured computing environment were used. This led to higher accuracies, precisions, recalls, and F1 scores. All frameworks used: TensorFlow 2.12, Keras 2.11, Python 3.10, Random seed fixed (42) for deterministic results, Public dataset [6] with train-test split (80:20), Detailed preprocessing pipeline and augmentation parameters. In the succeeding sections, the results are detailed, and the visual aids and statistical analyses are used.

Table 6 presents the disease-wise performance of the DermoAI-CNN model. In the original DermNet dataset, the label “Melanoma, Skin Cancer, Nevi & Moles” is a composite category encompassing malignant melanoma, other non-melanoma skin cancers (such as basal cell carcinoma and squamous cell carcinoma), and benign pigmented lesions (such as nevi and moles). To maintain consistency with the dataset’s taxonomy, this composite label was retained in Table 6.

Table 6. Disease-wise performance of DermoAI-CNN model on Accuracy, Precision, F1-score

Class	Precision(%)	Recall(%)	F1-score(%)
Acne & Rosacea	98.43	98.43	98.43
Actinic Keratosis & Malignant Lesions	99.34	98.05	98.69
Atopic Dermatitis	100.00	100.00	100.00
Bullous Disease	100.00	97.06	98.51
Cellulitis & Bacterial Infections	100.00	97.44	98.70
Eczema	97.99	98.98	98.48
Exanthems & Drug Eruptions	95.83	100.00	97.87
Hair Loss & Alopecia	87.50	89.74	88.61
Herpes, HPV & STDs	86.21	87.72	89.96
Light Diseases & Pigmentation	97.83	98.90	98.36
Lupus & Connective Tissue Diseases	98.46	100.00	99.22
Melanoma, Skin Cancer, Nevi & Moles	98.46	100.00	99.22
Nail Fungus & Nail Diseases	93.15	88.31	90.67
Poison Ivy & Contact Dermatitis	100.00	100.00	100.00
Psoriasis & Lichen Planus	95.42	96.62	96.02
Scabies, Lyme Disease & Infestations	98.59	100.00	99.29
Seborrheic Keratoses and Benign Tumors	99.04	98.10	98.57
Systemic Disease	98.82	97.67	98.25
Tinea, Ringworm, Benign Tumors	99.49	99.49	99.49
Urticaria & Hives	100.00	100.00	100.00
Vascular Tumors	100.00	100.00	100.00
Vasculitis	94.74	98.18	96.43
Warts, Molluscum & Viral Infections	94.58	95.15	94.86

3.3.1. Training and Validation Metrics

The training process showed steady improvement in the model's performance. Figure 2(a): Training and Validation Accuracy graph; Accuracy iteratively improved across epochs. The accuracy of the model began at 13.97 and reached 96.70. In the same way, the accuracy of validation increased to 18.25 to 97.30 which shows the generalization property of the model.

Figure 2(b) demonstrates Training and Validation Loss, which shows the decrease of the loss during epochs. Losses in training and validation have decreased by 2.87 and 2.78 to 0.12 and 0.11 respectively. These values of loss are decreasing steadily denoting better learning and reduced uncertainty of prediction with more training.

3.3.2. Prediction Results

The model's classification performance can be assessed from class-level predictions. Some sample predictions are displayed in Figure 3, which exhibits correctly classified images and their predicted labels. Figure 4 presents a graphical comparison of classification performance with DermoAI-CNN. In this regard, the output provides a qualitative assessment of the model's ability to classify different classes correctly.

Figure 4 compares the performance of the proposed DermoAI-CNN model with both *standard CNN architectures* (such as MobileNet-V2, DenseNet201, and EfficientNet-B7) and *state-of-the-art models reported in literature*, including Vision Transformer (ViT)–based approaches (YoTransViT, Swin-Transformer) and hybrid CNN–RNN frameworks. The comparison was performed using the same dataset (DermNet) and standardised training conditions, as detailed in Table 6. This ensures that the graphical results in Figure 4 correspond directly to the quantitative benchmark results summarized later in Section 4.1 ("Comparison with Previous Approaches"). Accordingly, Figure 4 gives a visual comparison of the performance of the literature benchmark and the other CNN-based baselines that we use in our experiments.

All the pictures shown in Figure 3 are parts of the original DermNet images that were automatically cropped using the Mask R-CNN segmentation step. These cut areas are centred, leaving the entire lesion intact, while the background is removed to create visual emphasis on the diseased area. Preprocessing was conducted to ensure consistency across samples, as cropping and resizing were performed consistently (224×224×3 pixels).

3.3.3. Confusion Matrix Analysis

Accuracy of the classification results is quantitatively studied by the help of a confusion matrix. Figure 5 indicates the total of true positive, false positive, true negative, and false negative. Most importantly, the diagonal values of most classes are high. On the other hand, the off-diagonal values show the misclassifications and are small, which means that there are few misclassifications.

The confusion matrix also indicates potential improvement points like certain classes that have a minor higher misclassification rate, and it would be interesting to explore the feature

extraction technique or data augmentation to represent these classes in a more appropriate way.

In order to evaluate the generalization prowess of the proposed DermoAI-CNN model, the external validation experiment was performed in relation to the ISIC 2019 dermoscopic images database, which did not rely on the training one, the DermNet database. This external evaluation is associated with the confusion matrix in Figure 5 and it indicates that the model is strong in datasets of different imaging modes. The level of diagonal dominance of the matrix is high which proves the fact that the model can be used on the unseen dermoscopic data as well which proves that the model is flexible to the variations of clinics.

3.3.4. Precision-Recall Curve

The precision-recall curve in Figure 6 is an analysis of a classification model, where precision is plotted against recall at different thresholds, and is particularly applicable to imbalanced data. In this study, micro-average, macro-average, and weighted-average PR curves were considered. The micro-average PR curve computes measures across all classes, giving more weight to performance on more frequent instances. By comparison, the macro-average PR curve is fairer because it treats each class equally. Moreover, the weighted-average PR curve is class-imbalanced, since each class's contribution is weighted by its size, thereby better reflecting the overall performance. Overall, these curves indicate the model's strength and its ability to categorise the various class distributions correctly.

3.3.5. Training Dynamics and Model Stability

This is because the gradual overlap between the accuracy and loss curves indicates a well-calibrated training process. The early periods have shown slight variations in validation measures, which are naturally due to the model's adaptation to complex data structures. The metrics, however, stabilised as training progressed, indicating the strength and stability of the CNN-based architecture.

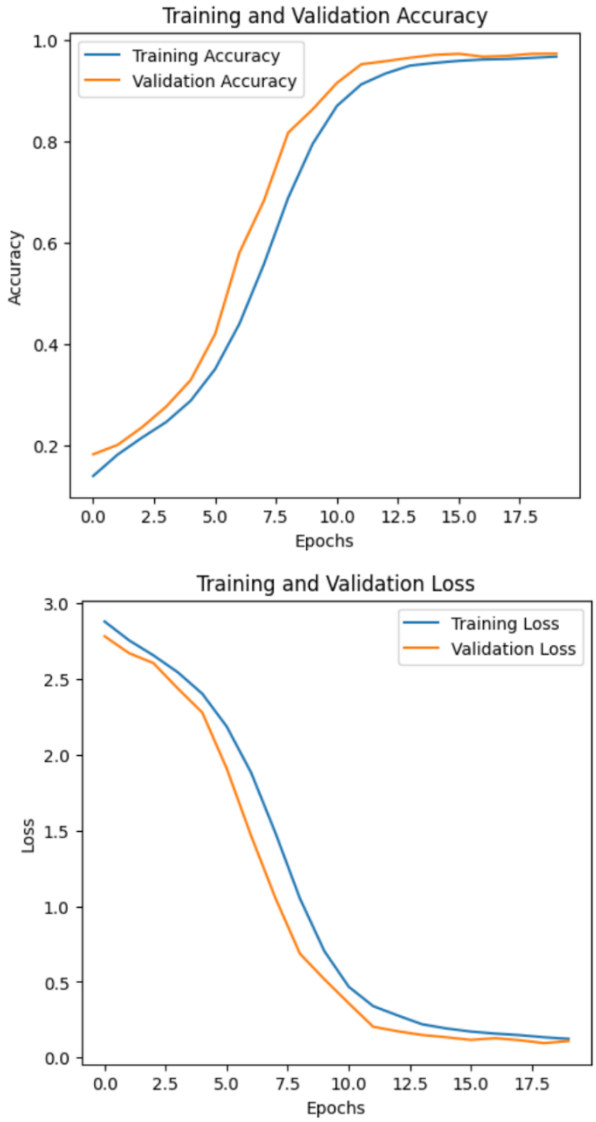

Figure 2. (a) Training and Validation Accuracy (b) Training and Validation loss

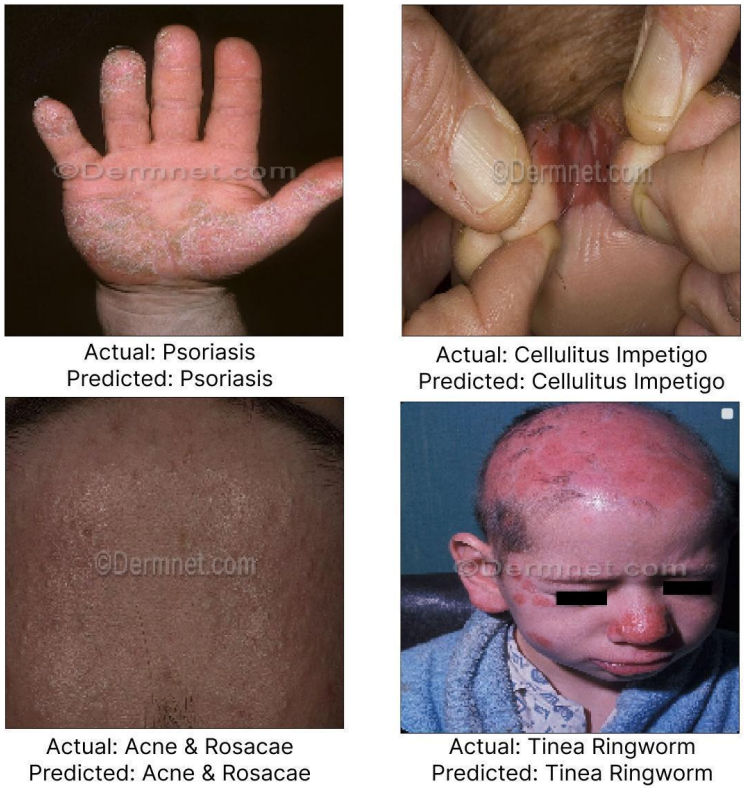

Figure 3. Prediction sample results on skin disease images using the DermoAI-CNN model

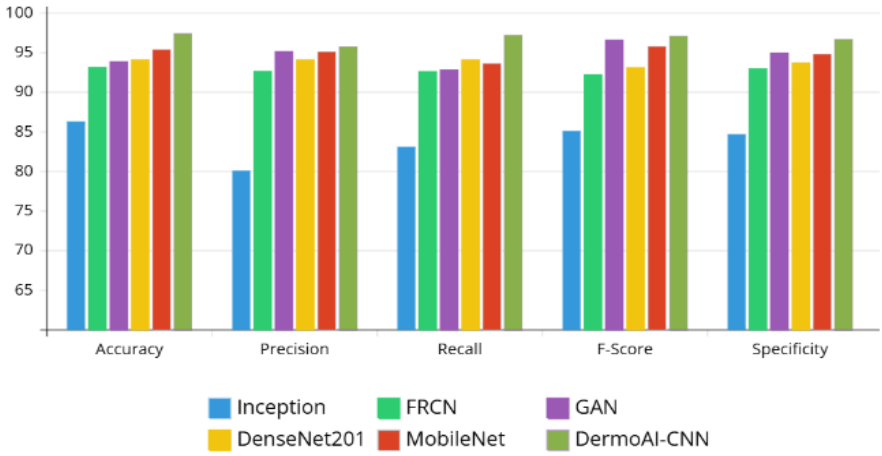

Figure 4. Graphical representation of classification approaches' performance compared to DermoAI-CNN

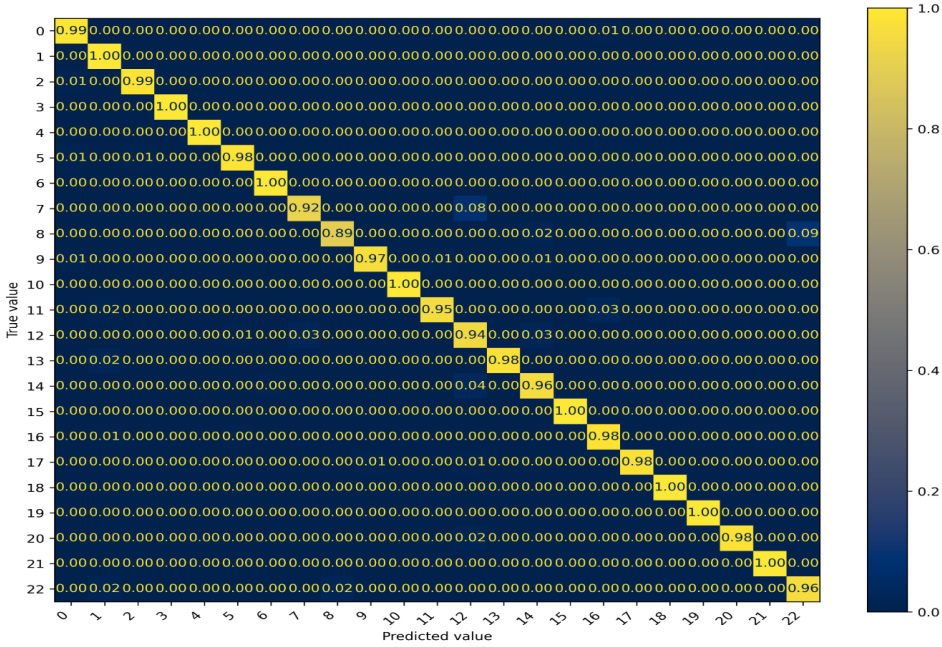

Figure 5. Confusion matrix on the ISIC 2019 dataset of right and wrong predictions for DermoAI-CNN

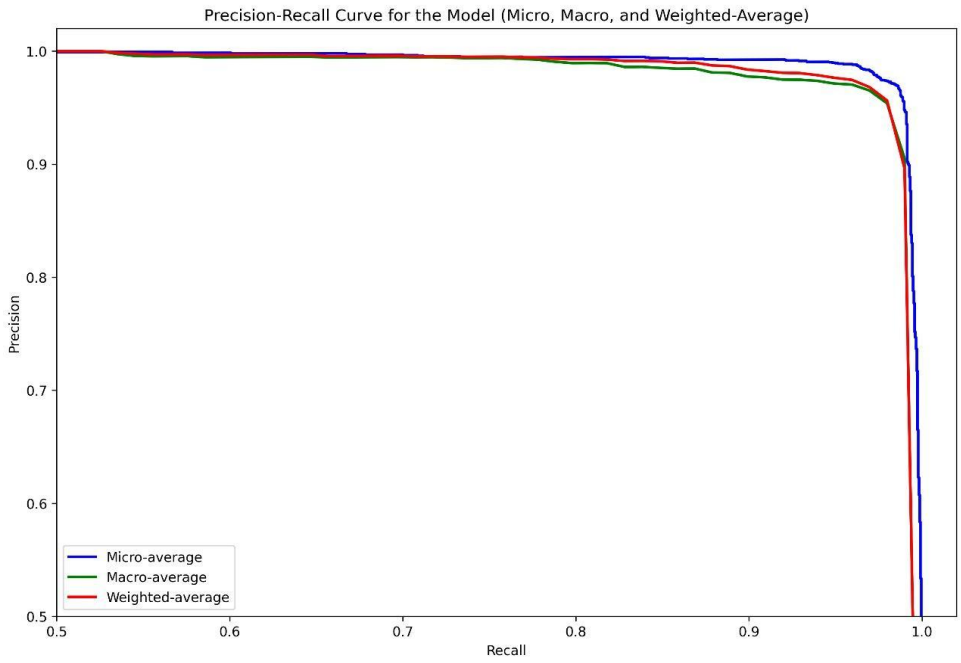

Figure 6. Precision-Recall Curve Showing Micro-Average, Macro-Average, and weighted average for Model Performance Evaluation.

3.4. Ablation study

We have provided a new ablation study (Table 7) that measures the contribution of each component. As this analysis shows, both SE and CBAM modules improve classification performance, whereas augmentation with a GAN further improves performance by +0.47% in accuracy.

Table 7. Ablation study

S/N	Model Variant	Components	Accuracy (%)	F1-score (%)
1	Baseline CNN	–	93.84	93.26
2	CNN + SE	SE only	95.16	94.89
3	CNN + CBAM	CBAM only	95.72	95.45
4	CNN + SE + CBAM	Dual Attention	96.83	96.61
5	Full Model + GAN + Mask R-CNN	Complete DermoAI-CNN	**97.30**	**96.98**

To determine the extent to which performance improvements were due to the SE/CBAM modules or network depth, we performed an ablation analysis of several architectural variants. First, we tested deeper CNNs (6 and 8 convolutional blocks, without SE/CBAM) to investigate whether deeper CNNs may be improved with the help of deeper depth. Although these deeper variants were slightly more accurate in training (~97.5%), validation accuracy was leveling off at 96.2% and overfitting was observed. The four block design was therefore chosen as a good compromise between performance and computing power.

The combination of Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) attention blocks showed significant improvement in performance due to the improvement of feature selectivity, but not just the depth. SE modules enhanced the separation of classes by dynamic weighting channels of features, whereas CBAM further narrowed the attention on the spatially relevant regions of lesions.

The incremental contribution of each of the processing stages is summarized in Table 7. The accuracy of the baseline CNN was 93.84% that rose to 95.16% and 95.72% during SE and CBAM respectively. The dual attention architecture (SE + CBAM) gave 96.83 per cent and the entire pipeline of GAN-based augmentation and segmentation with Mask R-CNN gave 97.30 per cent. This shows that the most critical improvement (~3%) was brought by attention mechanisms, and this fact confirms that they are highly important in improving feature representation and model generalization.

4. Discussion

We have examined in this paper how to use Generative Adversarial Networks (GANs), data augmentation, the Mask R-CNN, data segmentation, Convolutional Neural Networks (CNNs) and attention models such as the Squeeze-and-Excitation (SE) of Convolutional Block Attention Module (CBAM) classifier to classify dermatological images using the Kaggle data called DermNet NZ [6]. The model was able to provide such notable and outstanding results with the training and validation accuracy being not less than 97.30%. The loss presented in the training being minimal, that shows the strength and stability of the model in the question of training on the information it is given and the extent to which it will be able to generalize to the unknown cases. There were other assessment measures such as precision, recall, and F1-score, which were also applied and only enhanced and supported the high effectiveness of all classes included in the research. The results of the present study were thus measured by various visualizations, including confusion matrices, which can be used to see how well the model is apt at; precision recall curves, which can provide information on the interaction between precision and recall; and substantive plots of accuracy and loss that track the real-time learning behavior of the model. All these visual tools proved the validity and effectiveness of the model in the correct classification of complex dermatological images, which strengthens the validity of the findings.

4.1. Comparison with Previous Approaches

Our approach shows several obvious superiorities to the current works. The past models such as MobileNet [23] and DenseNet201 [8] had good performance, but they were limited to

particular disease types and not in image imbalance effective. GAN-based [13] augmentation was better, but none of lesion-focused segmentation. Background suppression was not considered crucial and hybrid CNN-RNN models [27] demonstrated better learning of features.

The ways our work facilitates the advancement of the field are: - The model becomes more clinically relevant as it is expanded to cover 23 dermatological conditions. Optimizing underrepresented CSSs (with GANs) and achieving much better balance and less overfitting. The use of Mask R-CNN to pixelly segment lesions to make sure that the classification is lesion-based and not background-based. SE and CBAM: dual attention mode was incorporated to achieve refined features representation, which allowed differentiating between similar conditions visually.

This triple statement is not so simple as the augmentation, segmentation, and attention mechanisms have to be balanced in the same training pipeline. The obtained structure has a better accuracy (97.30%) and stability in many metrics of evaluation, outperforming the previous SOTA solutions.

Our model will help mitigate these limitations in several ways:

1. **Expanded Dataset:** Unlike the previous models that focused on a specific set of circumstances, we have 23 categories, which is a more comprehensive and clinically useful solution.
2. **Improved Generalization**: GAN in data augmentation has significantly reduced overfitting and hence the model becomes resistant to unseen data. Additionally, resistance to variability of real-world data was enhanced by means of such techniques as rotation, scaling, and flipping.
3. **Robust Feature Extraction:** When the SE and CBAM realization of attention mechanism were used together, the model could be focused on informative features, which minimize noise and maximize the classification accuracy.:
4. **Expanded Dataset:** In contrast to the past models that were aimed at a particular set of conditions, our work has 23 categories, which provide a more holistic and clinically applicable solution.
5. **Improved Generalization**: The use of GAN in data augmentation has greatly minimized overfitting, and thus the model is more immune to unseen data. Moreover, such techniques as rotation, scaling, and flipping improved resistance to variability of real-world data.
6. **Robust Feature Extraction:** When the SE and CBAM implementation of the attention mechanism were combined, the model was able to focus on informative features, which reduce noise and maximize the classification accuracy.

Table 8 demonstrates the superiority of our model in terms of accuracy, precision, recall, and F1-score compared to state-of-the-art methods, confirming its high effectiveness. The current results of experiments are as follows:

i. Compared to YoTransViT and Swin Transformer, the current SOTA Vision Transformer architectures in dermatology.
ii. Included Skin-CAD and Hybrid CNN-RNN that are skin lesion specific.

iii. involves findings from previous studies using the Dermnet dataset or its subsets to establish a fair basis.
iv. The fact that the proposed DermoAI-CNN outperforms CNN-based as well as ViT-based baselines supports the validity and competitiveness of the experimental setup.

Table 8. Comparative Evaluation with State-of-the-Art Models for Skin Lesion Classification

S/N	MobileNet-V2	Lightweight CNN	Skin Cancer (ISIC)	95.27	96.00	93.52	95.68	[23]
1	DenseNet201	Deep CNN	ISIC	94.00	94.00	94.00	93.00	[8]
2	GAN-based CNN	CNN + GAN Augmentation	ISIC	93.78	95.06	92.70	96.50	[13]
3	SLDCNet (FRCN)	Full-Resolution CNN + Transfer Learning	ISIC	93.02	92.56	92.52	92.11	[26]
4	Hybrid CNN–RNN	CNN + RNN	HAM10000	96.84	95.12	96.60	95.86	[27]
5	Skin-CAD	Explainable CNN	Dermoscopic	96.50	95.30	96.10	95.70	[3]
6	YoTransViT	Vision Transformer + CNN	Dermnet	97.05	96.80	96.50	96.65	[19]
7	Swin-Transformer (ViT-based)	Pure Transformer	Dermnet (Subset)	96.90	95.90	96.40	96.10	[15]
8	EfficientNet-B7	Hybrid CNN	Dermnet	96.40	94.90	95.70	95.30	[22]
9	Proposed DermoAI-CNN	GAN + Mask R-CNN + Dual Attention CNN	Dermnet	97.30	95.65	97.09	96.98	This study

4.2. Implications

The conclusions drawn from this work have a direct connection to dermatological diagnosis and automated healthcare. This system can:

(i) Enhance the quality of diagnostics and decrease inaccuracies in clinical practice.
(ii) Reduce medical expenses through automation of medical diagnosis.
(iii) Facilitate easier access, especially in low-resource settings, by creating mobile-based diagnostic equipment.

Even though the proposed model shows significant improvements in dermatological image classification, some remain.

4.3. Limitations and Future Directions

Following are the limitations that point to future work:

i. **Dataset Diversity**: The dataset that is utilized in the research is mainly composed of non-dermoscopic images, which is not necessarily reflective of real-life clinical situation. It is advised that future studies aim to expand the dataset to incorporate dermoscopic and other image data, adding strength and flexibility across diverse clinical environments.

ii. **Computational Complexity**: The combination of GANs, the Mask R-CNN, and the attention mechanism causes the model to become more complex regarding computational requirements and difficult to run on low-resource settings. Future work should focus on optimising the model for deployment on lightweight mobile and embedded systems to enable real-time diagnostics.

iii. **Exploration of Vision Transformers (ViTs)**: ViTs have already demonstrated the potential to capture long-range image representations, and may be useful in enhancing classification. Future studies could explore incorporating ViTs into the model, balancing their cost and performance.

iv. **Clinical Validation with Dermatologists:** Although the present study has been able to validate the model on a large scale with the help of benchmark datasets, the direct validation with practising dermatologists has not been done yet. The next step, clinical validation, is necessary, and an expert ground-truth comparison will be used to verify the system's applicability to real-world diagnostic procedures. We are also in the process of negotiating with clinical partners to carry out such validation in future projects.

5. Conclusion

This study emphasises the possibility of combining GANs, Mask R-CNN, and improved attention systems within a CNN-based architecture, resulting in dermatological image classification systems. In addition, the DermoAI-CNN model achieved an accuracy of 97.30% and high performance indicators. Despite the fact that there is room for improvement and this will be achieved at later research stages, the study provides a solid foundation for the further development and practical implementation of medical diagnostics.

References

[1] A Razia. S., Chamola, V., Hussain, Z., Albalwy, F., & Hussain, A, A novel end-to-end deep convolutional neural network based skin lesion classification framework, *Expert Systems with Applications*, *246*. 2024, 1-19.

[2] Ali, R., Manikandan, A., & Xu, J, A Novel framework of Adaptive fuzzy-GLCM Segmentation and Fuzzy with Capsules Network (F-CapsNet) Classification, *Neural Computing and Applications*, *35*, 2023, 22133–22149.

[3] Attallah, O, Skin-CAD: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level CNNs features and transfer learning, *Computers in Biology and Medicine*, *178*, 108798, 2024.

[4] Bagheri, F., Tarokh, M. J., & Ziaratban, M, Skin lesion segmentation from dermoscopic images by using Mask R-CNN, Retina-Deeplab, and graph-based methods, *Biomedical Signal Processing and Control*, *67*. 2021, 102533.

[5] Burada, S., Manjunathswamy, B. E., & Sunil Kumar, M, Early detection of melanoma skin cancer: A hybrid approach using fuzzy C-means clustering and differential evolution-based convolutional neural network, *Measurement: Sensors*, *33*, 2024, 101168, 1-9.

[6] DermNet NZ. *Dermnet*. https://www.kaggle.com/datasets/shubhamgoel27/dermnet, 2024

[7] Fan, S. K. S., & Chen, W. Y., A generative-adversarial-network-based temporal raw trace data augmentation framework for fault detection in semiconductor manufacturing, *Engineering Applications of Artificial Intelligence*, *139(part B)*, 2025, 109624.

[8] Farooq, M. A., Khatoon, A., Varkarakis, V., & Corcoran, P., Advanced deep learning methodologies for skin cancer classification in prodromal stages, *CEUR Workshop Proceedings*, *2019, 2563*.

[9] Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z., Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning, *Healthcare (Switzerland)*, *10*(7). 2022, 1183, 1-18.

[10] Gurunathan, A., & Krishnan, B., A Hybrid CNN-GLCM Classifier For Detection And Grade Classification of Brain Tumor, *Brain Imaging and Behavior*, *16*(3). 2022, 1410–1427.

[11] Jeong, H. K., Park, C., Henao, R., & Kheterpal, M., Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations, *JID Innovations, 3(1), 2023, 100150, 1-16.*

[12] Kao, S. Y. Z., Ekwueme, D. U., Holman, D. M., Rim, S. H., Thomas, C. C., & Saraiya, M., Economic burden of skin cancer treatment in the USA: an analysis of the Medical Expenditure Panel Survey Data, 2012–2018, *Cancer Causes and Control*, *34*(3), 2023, 205–212.

[13] Kumar, K. S., Suganthi, N., Muppidi, S., & Kumar, B. S., FSPBO-DQN: SeGAN based segmentation and Fractional Student Psychology Optimization enabled Deep Q Network for skin cancer detection in IoT applications, *Artificial Intelligence in Medicine*, *129*, 2022, 102299.

[14] McHale, C. M., Osborne, G., Morello-Frosch, R., Salmon, A. G., Sandy, M. S., Solomon, G., Zhang, L., Smith, M. T., & Zeise, L., Assessing health risks from multiple environmental stressors: Moving from G × E to I × E, *Mutation Research - Reviews in Mutation Research,* 775, 2018, 11-20.

[15] Mohan, J., Sivasubramanian, A., V., S., & Ravi, V., Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable

AI, *Computers in Biology and Medicine*, *190*, 2025, 110007.

[16] Oztel, I., Yolcu Oztel, G., & Sahin, V. H., Deep Learning-Based Skin Diseases Classification using Smartphones, *Advanced Intelligent Systems*, *5*(12), 2023, 2300211, 1-11.

[17] Ravi, V., Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification, *Cancers*, *14*(23), 2022, 1-26.

[18] Rodrigues, D. de A., Ivo, R. F., Satapathy, S. C., Wang, S., Hemanth, J., & Filho, P. P. R., A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system, *Pattern Recognition Letters*, *136*, 2020, 8–15.

[19] Saha, D. K., Joy, A. M., & Majumder, A., YoTransViT: A transformer and CNN method for predicting and classifying skin diseases using segmentation techniques, *Informatics in Medicine Unlocked*, *47*, 101495, 2024, 1-15.

[20] Schmitges, F. W., Radovani, E., Najafabadi, H. S., Barazandeh, M., Campitelli, L. F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T., Dai, W. F., Taipale, J., Emili, A., Greenblatt, J. F., Hughes, T. R., Zhou, X., Weeks, S. D., Ameloot, P., Callewaert, N., … Zammit, P. P. S. Melanoma: Risk factors, early detection, and treatment strategies-An updated review. *Skeletal Muscle*, *6*(1), 2016.

[21] Serte, S., & Demirel, H., Gabor wavelet-based deep learning for skin lesion classification, *Computers in Biology and Medicine*, *113*, 103423, 2019.

[22] Shukla M M et al., A hybrid CNN with transfer learning for skin cancer disease detection, *Medical & Biological Engineering & Computing*, *62*(10), 2024, 3057–3071.

[23] Srinivasu, P. N., Sivasai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J., Classification of skin disease using deep learning neural networks with mobilenet v2 and LSTM, *Sensors*, *21*(8), 2021, 1-27.

[24] Tschandl, P., Rosendahl, C., & Kittler, H., Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Scientific Data*, *5*, 2018, 180161.

[25] Urban, K., Mehrmal, S., Uppal, P., Giesey, R. L., & Delost, G. R., The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study, 1990–2017, *JAAD International*, *2*, 2021, 98-108.

[26] Varma, P. B. S., Paturu, S., Mishra, S., Rao, B. S., Kumar, P. M., & Krishna, N. V., SLDCNet: Skin lesion detection and classification using full resolution convolutional network-based deep learning CNN with transfer learning, *Expert Systems*, *39*(9), 2022.

[27] Zareen, S. S., Sun, G., Kundi, M., Qadri, S. F., & Qadri, S., Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach, *Computers, Materials and Continua*, *79*(1), 2024, 1497–1519.

[28] Zenghong Wu, Fangnan Xia, R. L., Global burden of cancer and associated risk factors in 204 countries and territories, 1980-2021: a systematic analysis for the GBD 2021, *J Hematol Oncol*, *17*(1: 119), 2024, 1-14.

[29] Zhang, B., Zhou, X., Luo, Y., Zhang, H., Yang, H., Ma, J., & Ma, L., Opportunities and

Challenges: Classification of Skin Disease Based on Deep Learning, *Chinese Journal of Mechanical Engineering (English Edition)*, 34(1), 2021.

[30] Zhang, J., Zhong, F., He, K., Ji, M., Li, S., & Li, C., Recent Advancements and Perspectives in the Diagnosis of Skin Diseases Using Machine Learning and Deep Learning: A Review, *Diagnostics*, 13(23), 2023, 1-30.

Received 15.02.2025, Accepted 14.11.2025