I.J. Image, Graphics and Signal Processing, **2026, 4, 82-105**
Published Online on August 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2026.04.05

A Novel Deep Learning Framework for Multi-Class Skin Cancer Classification Using Modified Stacked Capsule Network

Mudiyam Sivani*
Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, India
Email: mudiyamsivani-rs-cse@dsu.edu.in
ORCID iD: https://orcid.org/0009-0003-7656-4742
***Corresponding author

M. Shahina Parveen
Department of Computer Science & Technology, School of Engineering Dayananda Sagar University, Bangalore, India
Email: shahina-cst@dsu.edu.in
ORCID iD: https://orcid.org/0000-0002-5914-4127

Received: July 18, 2025; Revised: January 9, 2026; Accepted: February 2, 2026; Published: August 8, 2026

Abstract: Skin cancer is one of the deadliest kinds of cancer because of it's due to its visible onset and the potential for rapid progression and spread. Oncologists employ several techniques, including imaging and biopsies, to determine whether skin cancer is present, but these are labor-intensive and time-consuming. Developing an automated and accurate framework is essential for early skin cancer detection, which greatly increases the probability of a successful treatment and recovery. Therefore, to address the aforementioned problems, this study aim to propose a novel Deep Learning (DL)-based multi-class skin cancer classification (MC-SC2) model for automatically diagnosing skin cancers using dermoscopic images. Initially, the dermoscopic images are preprocessed using a Gaussian Filter (GF) to effectively remove noise. To address the class imbalance issue, we applied advanced augmentation techniques to oversample minority classes, ensuring a uniform class distribution and enhancing the model's ability to generalize. To extract the complex and dominant features, we propose a novel DL-based model called BAM-EfficientNet (Bottleneck Attention Module with EfficientNetB7). In BAM-EfficientNet, we replace each Squeeze-and-Excitation (SE) attention module with a BAM in the traditional EfficientNetB7; this modification enables the network to concentrate on the most relevant regions in the images. The extracted features fed into the proposed Modified Stacked Capsule Network (MSCNet) to classify a skin lesion image as AKIEC, BCC, BKL, DF, MEL, NV, and VASC. In the proposed MSCNet employs the Disperse Dynamic Routing (DDR) algorithm to improve the capsule network's performance, and network's initial weights and biases are fine-tuned using Pelican Optimization Algorithm (POA) to further enhance performance. The HAM10000 dataset is used to assess the suggested model. The findings indicate that our approach beats current methods with an accuracy of 99.21% in classifying the seven different types of skin cancer, yielding substantial outcomes. These results demonstrate the potential of the suggested model as a quick, precise, and useful tool for early skin cancer diagnosis, providing important assistance in diagnosing skin cancer for medical professionals.

Index Terms: Skin cancer, Multi-class classification, Dermoscopic images, EfficientNetB7, Bottleneck Attention Module, Modified Stacked Capsule Network, Disperse Dynamic Routing, and Pelican Optimization Algorithm.

1. Introduction

The human body's biggest essential part is the skin. It shields the body's internal organs and external environment and performs several key tasks [1]. The initial stage of skin cancer development is uncontrollable skin cell growth. It is caused by UV radiation from the sun and artificial sources, including tanning beds, which increased the number of skin cells and caused malignant tumors to form. The most common types of skin cancer that affect people are dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), benign keratosis-like lesions (BKL), basal cell carcinoma (BCC), actinic keratoses and intraepithelial carcinoma (AKIEC), and vascular lesions (VASC) [2]. The most critical kind of skin cancer is melanoma. Melanoma frequently spreads to other organs, such as the brain, liver, spleen,

or lungs [3]. The most prevalent kind of skin cancer is BCC. The lowest layer of the epidermis gives rise to these spherical cells, which typically grow slowly before migrating to other parts of the body.

A pigmented mole that can vary in skin tone is called melanocytotic nevi. It generally happens during youth and the first few years of adulthood. Early detection of skin cancer can lead to effective treatment [4]. Histopathological analysis is a conventional way for diagnosing skin cancer. Skin cancer is detected by visual inspection using general diagnostic techniques. Medical imaging techniques such MRIs, CT (SPECT), PET scans, and X-rays are used to scan the damaged organs [5]. These imaging techniques can be used to evaluate the biological structure and physiological activities of different organs. In histology, biopsy is a painful procedure used to find cancer [6]. It causes discomfort, and other human organs are harmed by sophisticated microwave imaging techniques. Nevertheless, it is challenging to reliably detect skin cancer from images of skin lesions [7]. Conventional techniques for diagnosis cannot reliably detect skin lesions. Automated techniques are required for precise and prompt detection due to the rise in skin cancer. The primary use of dermatological ultrasonography is to diagnose skin lesions.

Researchers employed DL and automated machine learning (ML) techniques to address this. Traditional machine learning (ML) techniques including support vector machines (SVM), decision trees (DT), and the Naïve Bayes (NB) classifiers are employed. However, because the moles differ in size, shape, and color, these techniques do not produce significant results. Furthermore, the features were manually obtained, a time-consuming procedure requiring more human effort [8]. In order to identify skin cancer automatically, DL-based methods were developed. The most relevant features of the affected skin region can be extracted using DL approaches, taking into consideration variations in illumination and intensity. It has trouble with images of skin diseases that are low contrast and inaccurate [9]. Furthermore, the majority of recent studies focus on the binary classification of skin lesions (benign vs. malignant). However, there hasn't been much research done on the multi-class classification of certain skin cancer types. In order to classify the various kinds of skin lesions, an efficient method is required [10]. To address these issues, we present a novel DL-based framework for multi-class classification of skin cancer with dermoscopic images. The following are the main contributions of this work:

- A novel DL-based model called MSCNet is proposed for the multi-class classification of skin cancer. From dermoscopic images, the suggested MSCNet effectively detects seven different forms of skin cancer, including AKIEC, BCC, BKL, DF, MEL, NV, and VASC.
- Skin lesion images are collected from the HAM10000 dataset. Initially, the collected skin lesion images are preprocessed using a GF to effectively remove noise and improve overall image quality.
- To address the class imbalance issue in HAM10000 dataset, we applied advanced augmentation techniques to oversample minority classes, ensuring a uniform class distribution and enhancing the model's ability to generalize.
- We propose a new DL-based model named BAM-EfficientNet to extract the intricate and prominent features from the preprocessed images.
- In BAM-EfficientNet, we enhance the traditional EfficientNetB7 architecture by replacing each SE attention module with a BAM. This modification enables the network to concentrate on the most significant features while suppressing redundant information, thereby enhancing the accuracy of classification.
- Next, the extracted features are fed into the proposed MSCNet to classify different types of skin cancer.
- The proposed MSCNet employs the DDR algorithm instead of Dynamic Routing (DR) to improve the capsule network’s performance. Moreover, the network's initial weights and biases are optimized by the POA to improve initial performance and speed up training convergence.
- The effectiveness of the proposed method is analyzed using various metrics including accuracy, recall, precision, sensitivity, specificity, and f-score and it is compared against existing techniques.

The structure of the paper is as follows. Section 2 reviews relevant research on the detection of skin lesions. The proposed multi-class skin cancer detection system is presented in detail in Section 3. The simulation results of the proposed method are discussed in Section 5. The paper is concluded in Section 5.

2. Literature Survey

To detect skin cancer, the researchers have developed various methods, some of the methods are discussed below. Saeed, A., et al. [11] introduced an Efficient Global Vision Attention-Net (EG-VAN) for early detection of skin cancer. The dual-path architecture was a unique combination of a modified ResNet50 enhanced with a Spatial-Context Group Attention (SCGA) module and Non-Local Block, and EfficientNetV2S, which processed multi-scale features with exceptional efficiency. A Multi-Scale Feature Fusion (MFF) module was used that combined hidden properties across scales in order to enhance generalization across skin cancer types. The colours in the images were then modified employing a new colour balancing approach that was based on Grey World and Retinex theories. This improved the input quality of the dual-path network and made the significant regions more visible. The technique increased the skin

cancer categorisation system's accuracy to an impressive level. However, the method struggled to identify all forms of variation within the same class and was computationally complex.

A LEVit framework developed by Sakib, A. H., et al. [12] aids in the automatic classification of skin lesions and the early detection of several types of skin cancer. For improved skin lesion image analysis, they integrated LEVit with Vision Transformer (ViT) and Convolutional Neural Network (CNN). They used two sizable image datasets to test LEVit: ISIC 2019, which had 25,331 images in 8 classes, and HAM10000, which had 10,015 images in 7 classes. They employed augmentation techniques, produced a balanced number of images for each class and more images for the less prevalent classes. Finally, skin lesions were accurately classified by the LEVit. The LEVit model produced exceptional outcomes with better accuracy. However, during the augmentation, the method encountered complexity, which damaged the model by producing an unrealistic image.

Haque, M. Z., and Morzina, M. S. [13] presented a DL based CNN method to detect skin cancer. They identified and detected multi-class skin cancer using ISIC 2018 and 2019 datasets, using a CNN for classification and an enhanced canny edge for detection. An enhanced canny edge detector was used in the pre-processing stage to identify the impacted region from the image. They utilized the ebola optimization search algorithm (EOSA) to find the best solution to the challenges by striking a balance among exploration and exploitation. This was the most effective way to accurately diagnose the skin lesions. However, CNN had trouble accurately identifying dermatological images.

Using the HAM10000 dataset, Haque, M. I., et al. [14] created a skin lesions analysis model using CNN. More than 10,000 labelled dermoscopic images from the collection showed different classes included benign keratosis, basal cell carcinoma, and melanoma. The approach used pre-processing techniques like scaling, normalization, and augmentation to alleviate class imbalance and increase data diversity. The technique detected skin cancer and recognized the issue of class imbalance. However, the model had trouble in accurately classifying classes with limited representation.

Houssein, E. H., et al. [15] presented a deep CNN (DCNN) with two unbalanced datasets, such as HAM10000 and ISIC-2019, which helped to detect skin cancer in the early stage. Oversampling techniques were used to deal with the unbalanced datasets. The DCNN method was contrasted with several transfer learning approaches, such as MobileNetV2, DenseNet-121, DenseNet-201, VGG-16, and VGG-19. Its performance was evaluated using four commonly used metrics: F1-score, accuracy, recall, and precision. Both datasets yielded the best accuracy for the DCNN model. However, an imbalanced issue with the chosen datasets affected the model.

Alruwaili, M., and Mohamed, M. [16] developed a fusion DL model to classify different skin diseases. The model integrated three CNN-based algorithms, such as ResNet50, EfficientNet-B2, and EfficientNet-B0, which functioned independently under separate branches. Each model extracted the characteristics from the images in its own way. Following the extraction, a fusion technique was used to integrate the results. The kaggle skin illnesses image dataset, which includes 27,153 images of skin conditions and ten distinct types of skin illnesses, served as their training dataset. The technique showed great promise for automating the diagnosis of skin conditions. However, the EfficientNet faced computational complexity, increased resource requirements, and showed inadequate performance.

A DL-based skin lesion classification model was given by Noaman, A., et al. [17] utilizing AlexNet's transfer learning technique. The strategy used an AlexNet-based transfer learning algorithm to classify eight different skin diseases. The model was trained, validated, and tested on the ISIC 2019 dataset, which has an anomalous class distribution. The image was shrunk using pre-processing, and the data augmentation helped to reduce the class size variance. The image's features and color details were preserved during pre-processing. The model obtained a better F1-score, accuracy, precision, and sensitivity. However, the model encountered difficulties when using techniques like segmentation and masking for enhancement or extraction.

Toprak, A. N. and Aruk, I.[18] developed a DeepLabV3+ powerful image segmentation model to accurately isolate the skin lesion areas from dermoscopic images. Balanced performance and reliable feature learning were guaranteed by employing three distinct pre-trained DL models, such as MobileNetV2, EfficientNetB0, and DenseNet201. To choose the most pertinent features, the ReliefF algorithm was used with the extracted features. DF, squamous cell carcinoma, MEL, NV, BKL, BCC, AKIEC, and VASC were the eight categories into which the acquired features were categorised using the kNN algorithm. The method identified skin cancer at an early stage with high accuracy and generalizability. However, the process encountered numerous difficulties when creating reliable and useful categorisation systems for skin cancer.

2.1. Research Gap

Even though substantial progress has been made in skin lesion detection, existing research still faces various limitations. Traditional ML-based approaches suffer from inefficient feature extraction, high misclassification rates, and limited generalization across different lesion types. While using DL techniques improves classification accuracy through automatic feature extraction, they still face challenges such as high computational complexity, overfitting, difficulty in handling intra-class variations and the lack of advanced optimization methods. These techniques might not identify skin malignancies early and could not yield significant results when tested on unseen data. Furthermore, a lot of current studies classify a skin lesions image as benign or malignant. However, there hasn't been much research done on the classify a different forms of skin lesion diseases. A more effective DL-based approach is required to identify the

various forms of skin lesion diseases. To address these shortcomings, this study presents a novel DL-based system for enhancing detection accuracy of the multi-class skin lesion diseases using dermoscopic images.

3. Proposed Methodology

Skin cancer develops when UV radiation damages skin cells, altering their DNA and impairing their normal growth. Researchers commonly employ dermoscopic images to detect skin cancer. If skin cancer is detected early, physicians have a better chance of halting the disease's progression and starting treatment on time. To improve the detection accuracy of skin cancer, DL algorithms are employed. Therefore, a novel DL-based automatic method for multi-class skin cancer classification utilizing dermoscopic images is proposed. The suggested technique is intended to efficiently identify the seven different forms of skin cancer, including AKIEC, BCC, BKL, DF, MEL, NV, and VASC.

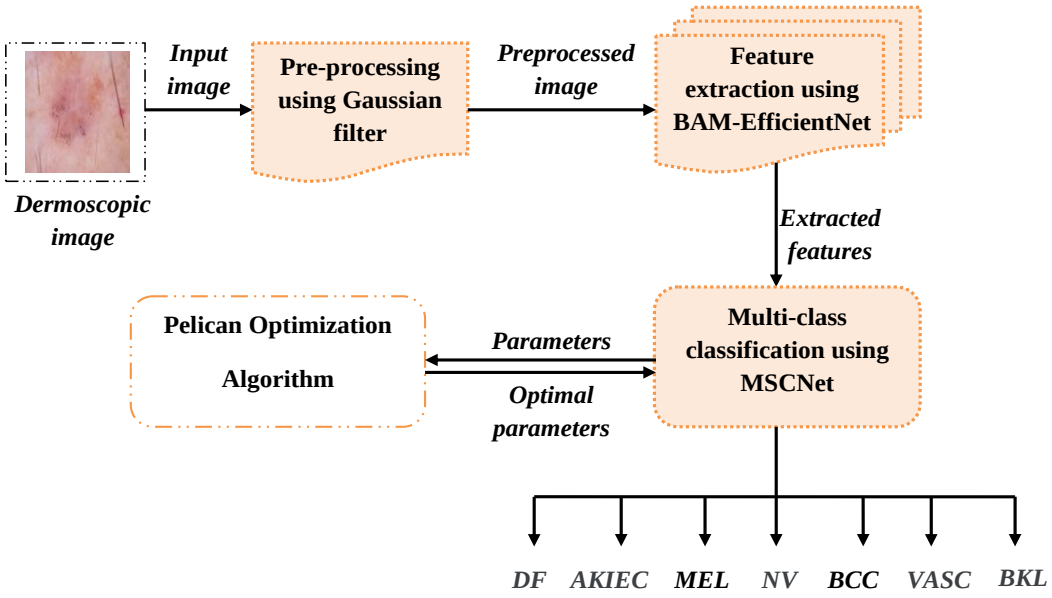


Fig. 1. Overview diagram of ProposedMC-SC2.

To effectively remove noise and enhance the image quality, the dermoscopic images are first preprocessed using a GF. Next, dataset balancing is performed via augmentation and oversampling, ensuring a uniform class distribution and enhancing the model's ability to generalize. Next, the complex and prominent features are extracted using the proposed BAM-EfficientNet model. In BAM-EfficientNet, a BAM is used in place of each SE attention module in the EfficientNetB7 architecture. This modification enables the network to capture important features and suppressing superfluous ones, improving the accuracy of classifying multiple types of skin cancer. In order to detect different types of skin cancer, the extracted features were then sent into the proposed MSCNet. To improve the performance of the capsule network, the MSCNet uses the DDR algorithm. Additionally, to further improve performance, the network's initial weights and biases are optimized using the POA. Figure 1 depicts a schematic design of the suggested MC-SC2 system. The following section offers a thorough explanation of the suggested MS-SC2 framework.

3.1. Dataset Collection

Skin lesion images are collected from the HAM10000 dataset. In the HAM10000 database, 10,015 images are included. Seven different data types that identify skin lesions are included in this collection. Among them include DF, MEL, NV, BKL, BCC, AKIEC, and VASC. The original class distribution is highly imbalanced, with **N**V representing approximately 67% of the dataset**,** while minority classes such as DF and VASC account for less than 1.5% each**.** Specifically, the dataset includes 514 images from the BCC class, 1113 images from the MEL class, 327 images from the AKIEC class, 1099 images from the BKL class, 115 images from the DF class, 142 images from the VASC class, and 6705 images from the NV class in this dataset. The images were acquired using a dermatoscopy device. The device is a set of lenses that functions as a magnification and is used to take images of skin lesions. The dataset provides a rich ground truth for supervised learning models by labeling each image with one of four diagnostic categories.

3.2. Preprocessing using Gaussian Filters

Dermoscopic images are often impacted by various artifacts and distortions arising from multiple sources. Thus, it is crucial to carry out preprocessing operations on images in order to improve their quality and attain more precise results. The raw images are transformed into a suitable format for further processing during this pre-processing stage.

Two-dimensional GFs are employed to enhance the images and eliminate the noise. The 2D GF is a widely used low-pass filter in image processing that plays a vital role in removing noise from images. This 2D GF uses smoothing and blurring operations to enhance the image quality and eliminate noise. In this preprocessing step, Gaussian operators act as convolutional filters, where smoothing is achieved through convolution operations. The following is the one-dimensional Gaussian operator:

$$GF^{(y)}{}_{1d} = \frac{1}{\sqrt{2\pi\sigma}} e^{\left(\frac{y^2}{2\sigma^2}\right)} \tag{1}$$

In the frequency and spatial domains, the crucial filter for image smoothing undergoes localization, and the uncertainty relation is executed as follows:

$$\Delta y \Delta \varpi \geq \frac{1}{2} \tag{2}$$

The Gaussian filter's two-dimensional operator is provided below:

$$GF_{2d}(y,z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{y^2+z^2}{2\sigma^2}\right)} \tag{3}$$

In this case, sigma (σ) represents a Gaussian function's standard deviation. The smoothing will be highest when the value is maximum, while y and z display the image's Cartesian coordinates, which indicate the window dimension.

3.3. Data Balancing

To address the issue of class imbalance in the HAM10000, we employed an oversampling technique. Minority classes, such as DF and VASC, were underrepresented compared to the dominant classes like NV, which posed a risk of biased learning. We synthetically increased the sample count of the minority classes by augmenting their images using the methods such as rotation, horizontal/vertical flipping, random cropping and translation, brightness and contrast adjustment, shear and zoom transformations, and sigmoid-based intensity correction. This balancing strategy ensured uniform representation across all class types, reducing model bias and enhancing feature learning for the less frequent classes. Once the datasets were balanced, the dataset was officially split into a training, validation, and testing subsets using an 80:10:10 ratio on the patient level. All images belonging to the same patient were strictly assigned to only one subset, preventing data leakage and ensuring unbiased evaluation.

Table 1. Balanced data distribution for dataset via oversampling techniques.

HAM10000 dataset balancing					
Class	**Actual**	**After Augmentation**	**Training**	**Validation**	**Testing**
AKIEC	327	6,705	5,364	335	1,006
BCC	514	6,705	5,364	335	1,006
BKL	1,099	6,705	5,364	335	1,006
DF	115	6,705	5,364	335	1,006
MEL	1,113	6,705	5,364	335	1,006
NV	6,705	6,705	5,364	335	1,006
VASC	142	6,705	5,364	335	1,006

3.4. Feature Extraction using BAM-EfficientNet

After preprocessing, a feature extraction stage is carried out, which plays a crucial role in capturing meaningful features from the images. In this stage, the proposed BAM-EfficientNet is utilized to capture complex and significant features from the preprocessed dermoscopic images. BAM-EfficientNet is an improved version of traditional EfficientNet-B7, in which each SE attention module is replaced with a BAM.

The BAM-EfficientNet model is developed based on EfficientNetB7, which has an effective network structure. EfficientNet-B7 is the most advanced model in the EfficientNet family, which equally scales depth, width, and resolution for increased accuracy and efficiency using a compound scaling technique. Simple and incredibly effective compound scaling techniques are used to build the EfficientNet-B7 model. For compound scaling, EfficientNet-B7 depth, width, and resolution parameters, such as dep, wid, and res, are provided in Equations (4), (5), and (6).

$$dep = \alpha^{\nu} \tag{4}$$

$$wid = \beta^{v} \tag{5}$$

$$res = \gamma^{v} \tag{6}$$

Where, $\alpha\beta^2\gamma^2 \approx 2$ and $\alpha \geq 1$, $\beta \geq 1$, $\gamma \geq 1$. In this case, v stands for the compound coefficient, and α, β, and γ are the scaling coefficients. The EfficientNet-B7's dep, wid, and res are all determined by the number of channels, layers, and related parameters.

Multiple Mobile Inverted Bottleneck Convolution (MBConv) modules are stacked to create the EfficientNet-B7 model. These MBConv blocks are arranged hierarchically so that deeper blocks can extract higher-level, more abstract features at various scales. The MBConv block consists of a Conv layer with a kernel size of 1 x 1, depth-separable conv, the SE attention block, and a dropout layer. However, the SE attention block can only focus on the significance of distinct feature channels; it cannot determine the significance of features at various spatial locations. This limitation can reduce the model classification accuracy by failing to capture spatial attention information of images. To address this, we suggest BAM-EfficientNet, an enhanced EfficientNet-B7 model, where the BAM block is used in place of SE block. BAM enhances network capacity to extract important features by accurately capturing spatial and channel attention information of images. Through this modification, the network is able to solve the problem of uneven spatial feature distribution in images, enhances network capacity to capture pertinent features, and thereby improve accuracy.

The architecture of the suggested BAM-EfficientNet is depicted in Figure 2. BAM-EfficientNet incorporates the BAM-MBConv module, designed to provide improved feature representation capabilities. Figure 3 (a) and (b) illustrate the structure of the BAM-MBConv1 and BAM-MBConv6 modules. The BAM-EfficientNet can capture both the accurate spatial information and the long-term relationships across channels, resulting in more robust and informative feature extraction.

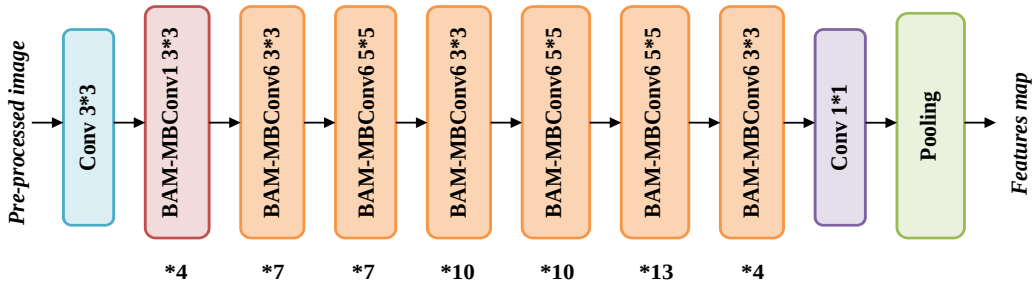


Fig. 2. BAM-EfficientNet Architecture.

The BAM-EfficientNet model is constructed by stacking Multiple BAM-MBConv modules. The BAM-MBConv module comprises a 1 × 1 Conv layer, depth-separable conv, BAM and a dropout layer. Following convolutional layers, the BAM-MBConv module uses the Swish activation function and Batch Normalization (BN). BN helps to standardize data and accelerates network convergence, and the Swish function adds non-linearity and helps prevent overfitting. The integration of BAM into the BAM-MBConv block enhances network feature extraction ability by obtaining spatial and channel attention information from input feature map. This enhances overall feature representation and assists in classifying skin lesions more accurately.

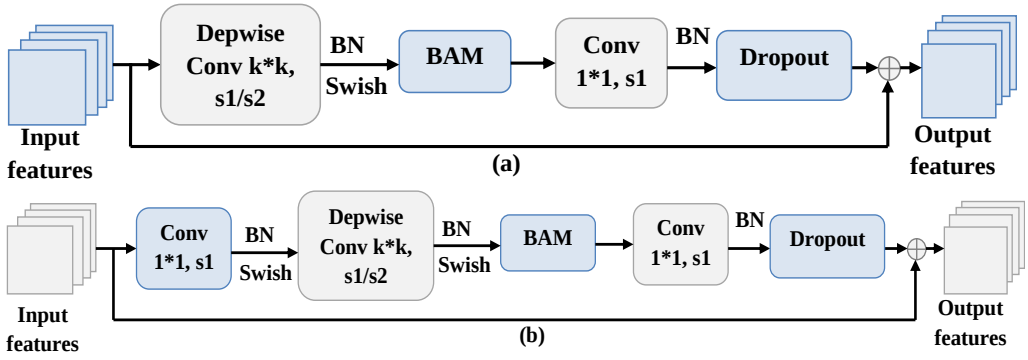


Fig. 3. (a) Structure of BAM-MBConv1 (b) BAM-MBConv6 Module.

The following is the feature extraction procedure with the BAM-EfficientNet model: The pre-processed images serve as the BAM-EfficientNet model's input. To accomplish space squeeze and channel expansion, the first layer first downsamples the input image using a Conv function with a kernel size of 3 × 3. After, the seven BAM-MBConv layers that include the BAM to further obtain high-dimensional features from the images. This stage aids in extracting intricate and multifaceted features from images of skin lesions. Following the last BAM-MBConv module, 1 × 1 Conv layer is employed to refine extracted features and followed by pooling layer that minimizes the spatial dimensions, producing a compact and semantically rich feature vector. This final feature map is then fedinto the proposed MSCNet for detection, facilitating accurate identification of multiple skin cancer types.

3.4.1. Bottleneck Attention Module

The BAM is incorporated into the EfficientNetB7, which allows the network to accurately extract the spatial and channel attention information of images, enhancing its ability to capture significant features.BAM uses both channel and spatial pathways to infer an attention map. To construct an effective module, the channel attention module (CAM) and the spatial attention module (SAM) are calculated in two separate branches. The BAM structure is shown in Figure 4.

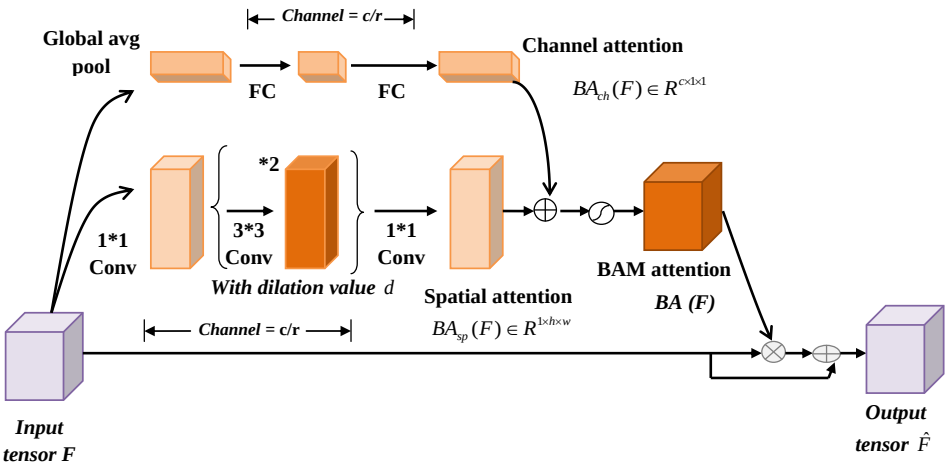


Fig. 4. Structure of BAM.

3.4.1.1. Channel Attention Module

Since every channel has a unique feature response, the channel branch takes advantage of the inter-channel interaction. Global average pooling (GAP) is utilized to aggregate feature maps F of every channel, producing a channel vector $F_{ch} \in R^{c\times1\times1}$. This vector subtly captures global information within every channel. The channel vector F_{ch} is used to measure attention across channels utilizing a multi-layer perceptron (MLP) with one hidden layer. The size of hidden layer is adjusted to $R^{c/r\times1\times1}$ in order to avoid a parameter overhead, where r denotes the reduction ratio, in order to avoid a parameter overhead. A BN layer is added after the MLP to adjust the scale with the output of spatial branch. The following formula is used to determine the CAM:

$$\begin{aligned} BA_{ch}(F) &= BN(MLP(avgpool(F))) \\ &= BN(w_1(w_0 avgpool(F) + b_0) + b_1 \end{aligned} \tag{7}$$

Here, $w_0 \in R^{c/r\times c}, b_0 \in R^{c/r}, w_1 \in R^{c\times c/r}, b_1 = R^c$.

3.4.1.2. Spatial Attention Module

In order to highlight or suppress features at various spatial locations, the SAM is responsible for producing a spatial attention map $BA_{sp}(F) \in R^{h\times w}$. It is necessary to use contextual information to identify which spatial regions that demand attention. A broad receptive field is required to use contextual information effectively. The receptive fields are efficiently expanded by the dilated Conv [19]. The dilated Conv makes it simpler to produce an efficient spatial map than the typical conv. Specifically, 1×1 Conv is used to fuse and compress the feature map along the channel dimension, projecting it $F \in R^{c\times h\times w}$ into a smaller dimension $R^{c/r\times h\times w}$. The channel branch uses the same reduction ratio r for

consistency. To effectively exploit contextual information, two 3x3 dilated convolution layers are employed after the reduction. Finally, 1×1 conv is used to compress the features to a $R^{1\times h\times w}$ spatial attention map. To adjust the scale, a BN layer is applied at the end of the spatial branch. The following formula is used to determine the SAM:

$$BA_{sp}(F) = BN(f_3^{1\times1}(f_2^{3\times3}(f_1^{3\times3}(f_0^{1\times1}(F))))) \tag{8}$$

In this case, f indicates a conv function; the superscripts indicate the size of the conv filter.

3.4.1.3. BAM

BAM derives the 3D attention map $BA(F) \in R^{c\times h\times w}$ for the input feature map $F \in R^{c\times h\times w}$ by integrating the CAM $BA_{ch}(F)$ and SAM that were obtained $BA_{sp}(F)$ from two attention branches. Then, the attention map $BA(F)$ should be calculated as follows:

$$BA(F) = \sigma(BA_{ch}(F) + BA_{sp}(F)) \tag{9}$$

Where, σ denotes a sigmoid function. Prior to addition, both branch outputs are scaled to $R^{c\times h\times w}$ because their shapes differ. The element-wise summing is chosen for an effective gradient flow. The final 3D attention map $BA(F)$ is created by applying a sigmoid function after summing, and it ranges from 0-1.

The 3D attention map $BA(F)$ is element-wise multiplied by the input feature map F , and then added to the original input feature map to produce the refined feature map $\widehat{F}$ .To make the gradient flow easier, a residual learning technique is integrated with the attention mechanism. The following is the calculation for the revised feature map $\hat{F}$:

$$\hat{F} = F + F \otimes BA(F) \tag{10}$$

Here, $\otimes$ stands for element-wise multiplication.

The BAM effectively captures the both channel and spatial attention information of images. By integrating BAM into the EfficientNetB7 architecture, the proposed BAM-EfficientNet significantly increase the network capacity to concentrate on most significant regions while suppressing irrelevant features. Hence, the proposed BAM-EfficientNet extracted the most relevant features from the preprocesssed image and sent into MSCNet for multi-class skin cancer classification. These extracted high-dimensional feature representations serve as a robust foundation for accurately classifying different skin cancer types in the subsequent classification stage.

3.5. Multi-Class Skin Cancer Classification Using Modified Stacked Capsule Network

The extracted features are then inputted into the proposed MSCNet to classify a multiple types of skin cancer diseases. The proposed MSCNet uses the DDR algorithm in place of the DR algorithm to improve the capsule network performance. Additionally, the POA is employed to fine-tune the network's initial weights and biases to further enhance performance.

The architecture of proposed MSCNet is shown in figure 5. The proposed MSCNet consists of the two primary capsule networks (CapsNet) stacked sequentially, with the output of the initial CapsNet acting as the input to the second CapsNet. This stacked model enables for progressive refinement of features, improving the network capacity to extract intricate features. In this architecture, the first CapsNet concentrates on extracting basic relationships and patterns, establishing the framework for the next network to extract more intricate and subtle features related to skin cancer. By iteratively refining features through this stacked architecture, MSCNet enables a deep understanding of the image, ultimately supporting for robust and accurate classification of multiple kinds of skin lesion diseases. Figure 4 shows architecture of the proposed MSCNet.

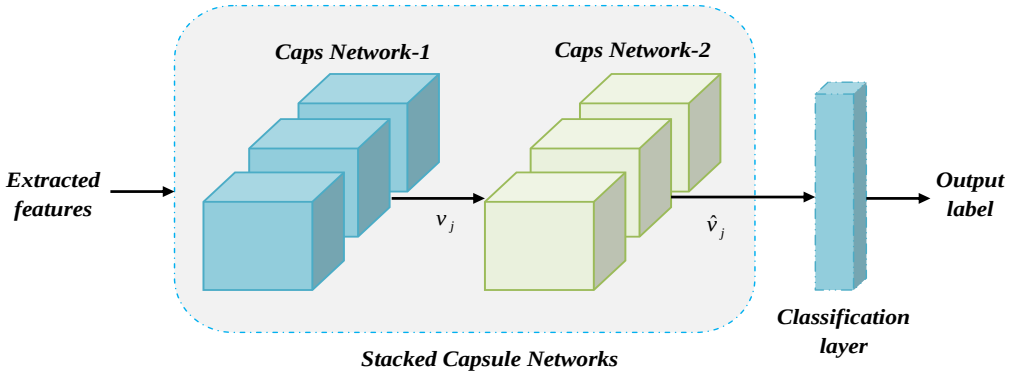


Fig. 5. Architecture of Proposed MSCNet.

3.5.1. CapsNet-1

The CapsNet-1 serves as the core component of the MSCNet, utilizing the unique concept of capsules-groups of neurons that respond to various attributes of a specific entity type-while maintaining both spatial relationships and feature hierarchies. Capsules preserve multidimensional information reflecting the orientation and other attributes of features. Because of its design, CapsNet-1 is able to capture and preserve complex features, laying a strong foundation for higher-order feature analysis in the subsequent capsule network. The Capsule Network structure is shown in Figure 6. The CapsNet-1 is made up of the following layers: input layer, Conv layer, primary capsule (PrimaryCaps) layer, DDR algorithm, digital capsule (DigitCaps) layer, and output layer.

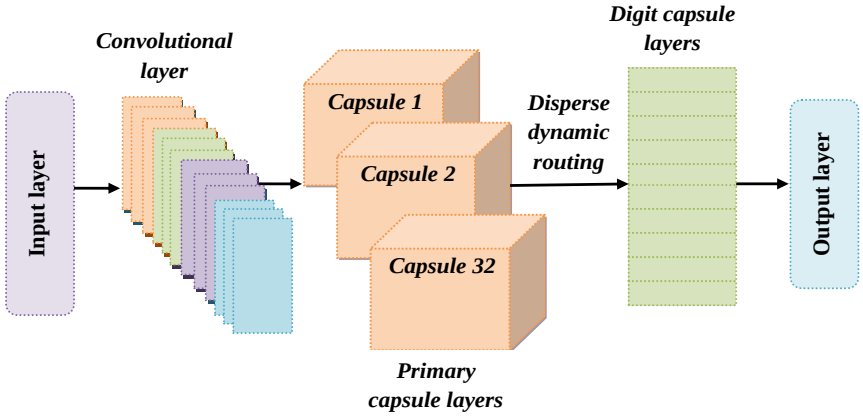


Fig. 6. Structure of CapsNet.

The conv layer uses a conv filter to obtain low-level features from input data. The capsules are used by the PrimaryCaps layer to capture spatial correlation within the features. The PrimaryCaps layer encodes the low-level features, which are then sent to the DigitCaps layer, where the DDR strategy is used to make predictions. The DDR algorithm adjusts the coupling coefficient (CC) values based on the degree of similarity among the low-layer capsule and the corresponding high-layer capsule; the more similar they are, the higher the CC between them. The DigitCaps' layer contains 16 capsules per digit classes, each capsule obtains input from each capsule in the preceding layer and is responsible for performing classification. The output vector from DigitCaps layer encapsulates richer information-such as the entity's pose, orientation, and other properties. The probability that the related class will exist is indicated by the output vector's magnitude. The longer the vector, the higher the likelihood of that class being detected. These layers are explained as follows:

Input Layer: The output of the BAM-EfficientNet is a feature vector, denoted by the symbol u_i, which is sent to the CapsNet-1 input layer. The crucial features that are essential for recognizing the various forms of skin cancer are encoded and highlighted in this layer.

Conv Layer: A convolutional layer is used following the input layer. These layers apply a series of learnable filters to the input data in order to capture local feature patterns such as edges or texture. This phase is carried out prior to the feature vector being passed into the PrimaryCaps.

Primary Caps Layer: In CapsNet-1's capsule hierarchy, this is the first layer. Every PrimaryCaps is a small cluster of neurons intended to identify specific characteristics in the feature maps obtained from the conv layer. The spatial relationships between the features are captured by this Primary Caps layer. This layer contains 32 capsules, each

of which generates an activity vector that encodes spatial information through instantiation parameters. These capsules provide a set of prediction vectors, which are lower-dimensional representations of the input data, by capturing the likelihood of feature existence and their spatial orientations. Prediction vectors are computed for every capsule i in the layer l using transformation matrices w_{ij} :

$$\hat{u}_{j|i} = squash(w_{ij}, u_i) \tag{11}$$

Where, u_i and w_{ij} represents prediction vector of the lower layer capsule i and weight matrix, respectively, and $\hat{u}_{j|i}$ denotes prediction vector of the upper-level j -th capsule. Squash is a non-linear function that decreases the length of a vector to a value between 0 and 1, which can be expressed as follows:

$$v_j = squash(sh_j) = \frac{\| sh_j \|^2}{1+\| sh_j \|^2} \frac{sh_j}{\| sh_j \|} \tag{12}$$

Here, v_j denotes its vector output of capsule j , and sh_j denotes its entire input.

Each PrimaryCaps effectively represents the pose of features as vectors, preserving both their spatial and orientational information within a specific receptive field.

Disperse Dynamic Routing: DDR mechanism exists between the PrimaryCaps and the higher-level capsules to overcome the limitations of conventional DR in CapsNet. In standard CapsNet routing, coupling coefficients are computed using the Softmax function, which enforces a probability distribution across output capsules. Although effective in simple classification tasks, Softmax-based routing often leads to tightly clustered coefficient distributions, causing weak discrimination among prediction vectors and reducing inter-class separability in complex multi-class problems. Prior studies have shown that Softmax routing yields coupling coefficients concentrated within a narrow interval, resulting in similar vector magnitudes across DigitCaps and increasing the risk of misclassification. To address this limitation, MSCNet uses the DDR algorithm instead of the conventional DR to further improve efficiency of CapsNet. The DDR algorithm makes it possible for capsules to communicate with one another and provide information to higher-level capsules. It establishes the links between the capsules. It ensures that the network concentrates on the spatial hierarchies in the data and calculates the CC (c_{ij}) among capsules by using the current input data. Unlike traditional DR that uses a Softmax function to calculate coupling coefficients, DDR uses the Sigmoid function to determine the CC, as expressed in Equation (13).

$$c_{ij} = \frac{1}{1+\exp(b_{ij})} \tag{13}$$

Where, c_{ij} no longer indicates allocation probabilities toward the final capsules, but the relationship strength among the PrimaryCaps and the DigitCaps. The discrepancy between the minimum and maximum CC is significantly greater. The essential prediction vectors are multiplied with greater CC for relevant features, and lower ones for irrelevant features, in order to make the relevant features more decisive. In addition, it extends the differentiation among the vector lengths in the DigitCaps layer. At that point, the length of the correct DigitCaps is greater than the length of the other DigitCaps. Sigmoid routing performance is better than the Softmax routing. Algorithm 1 displays the pseudocode of the Disperse dynamic algorithm.

Algorithm 1. Pseudocode of the DDR algorithm

1: Input to routing procedure: $(\hat{u}_i), r, l)$

2: for every Caps i in layer l and Caps j in layer $(l+1)$: $b_{ij} \leftarrow 0$

3: for r iterations

4: for every Caps i in layer l :

$c_i \leftarrow sigmoid(b_{ij})$

5: for every Caps j in layer $(l+1)$:

$sh_j \leftarrow \sum_i c_{ij} \hat{u}_{j|i}$

6: for every Caps j in layer $(l+1)$:
$v_j \leftarrow squash(sh_j)$
7: for every Caps i in layer l and Caps j in layer $(l+1)$: $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i}.v_j$
8: Return v_j

DigitCaps Layer: Following the PrimaryCaps is a layer of DigitCaps. The DigitCaps layer is responsible for transforming the initial capsules into more abstract output capsules using DDR process. This layer captures more intricate combinations of the features determined by the PrimaryCaps. In the multi-class skin lesion classification, this could entail looking for anomalies in image patterns to identify higher level patterns, including structural irregularities in lesion images that are essential for accurate classification.

Output Layer: A set of activity vectors from each capsule in the DigitCaps layer makes up CapsNet-1's output. These vectors include the instantiation parameters of the features found in the image data. The length of each vector indicates the likelihood that a certain feature would be present in the input data, and the feature's orientation in the vector space reflects its specific features. The output of CapsNet-1 is then utilized as input for CapsNet-2, which performs feature abstraction and representation at an even higher level.

3.5.2. CapsNet-2

CapsNet-2 acts as the next stage in the stacked MSCNet, focusing on processing and interpreting the intricate features transmitted by CapsNet-1. It is designed to further transform and process the output high-level feature vectors (v_j) from CapsNet-1 using secondary caps layer and DDR.

Input Layer: The output vectors v_j from CapsNet-1 make up the input to CapsNet-2. These vectors encode high-level feature information and the probability that they exist, which is determined by the DDR of the CapsNet-1.

Secondary Capsule Layer: Here, the DDR algorithm is utilized again. However, the outputs from are used to build the prediction vectors in CapsNet-2 $\hat{u}_{j|i}$, meaning that they are predicated on feature data that has already been processed and interpreted. This layer creates prediction vectors for every higher-level capsule using transformation matrices similar to those in CapsNet-1. The DDR algorithm then refines the data representation by iteratively updating the coupling coefficients between the capsules based on the "agreement" between their predictions. The DDR algorithm then subsequently improves the data representation by iteratively changing the CC among the capsules according to the "agreement" between their predictions.

Dynamic Routing in CN-2: The goal of CapsNet-2 DDR algorithm is to iteratively improve CC among the secondary caps in order to detect and highlight intricate, high-level features that are most relevant for skin lesion detection.

Output Layer: The secondary caps' final activity vectors are included in the CapsNet-2 output, represented by $\hat{v}_j$. These vectors reflect the network's comprehensive understanding of the input, capturing complex patterns and relationships essential for identifying multiple types of skin cancer.

Each output capsule in this final layer corresponds to a specific class and is represented as a vector. The length of this vector denotes the probability that the corresponding class is present in the input image.

3.5.3. Classification stage

The final stage of the CapsNet is classification, where the various types of skin cancer are detected. This stage classifies the lesion as multi-type skin cancer, including AKIEC, BCC, BKL, DF, MEL, NV, and VASC, using the output vectors of the DigitCaps lengths and orientations. These vectors' geometric characteristics provide a robust basis for classifying between different skin cancer types. The classification is performed by measuring each output vector's length, with the longest vector indicating the predicted skin lesion type.

3.5.4. Margin Loss Function

To direct the network toward accurate classification, the margin loss function is essential. It is created to provide a balance between sensitivity and specificity while penalizing inaccurate classifications. This balance is particularly essential in classification tasks, where the cost of false positives and negatives can be high. Margin loss can be expressed mathematically as follows:

$$Loss_c = t_c \max(0, mr^+ - \| v_c \|)^2 + \lambda(1-t_c)\max(0, \| v_c \| - mr^-)^2 \quad (14)$$

Here, λ denotes down-weighting term for the absence of a class, t_c is 1 in the case of a class present c and 0 otherwise, and mr^+ and mr^- are the margins for accurate and inaccurate classifications, respectively.

To further enhance classification performance, the POA is employed to optimize the network's initial weights and biases.

3.5.5. Parameter optimization using POA

The POA is employed to fine-tune the initial weights and biases of the MSCNet, thereby improving its overall performance. The POA is a stochastic optimization method that draws inspiration from pelicans' innate hunting tactics [20]. POA is well-known for its strong search space exploration and exploitation capabilities, effectively pursuing the global optimum. Pelicans often hunt in groups and employ a multi-step approach. They make a coordinated descent and then stretch their wings after determining the location of their prey. The pelicans can easily capture their prey by forcing them to surface and move into shallower waters. The following are the main stages of the POA:

Step 1: Initialization: Every pelican in the group is taken into consideration as a potential solution in the POA. To begin the optimization process, every member of the population is initialized at random. The weight and bias values of the MSCNet are taken into account in the study's solutions. Thus, the initial solution is determined as follows:

$$Y = \{y_1, y_2, ..., y_N\} \tag{15}$$

Where, y_N denotes quantity of members in the population or the N-th solution, it is denoted using (16).

$$y_N = \{w_1, w_2, ..., w_k; b_k\} \tag{16}$$

Step 2: Fitness calculation: After initialization, the fitness of each solution is determined. The fitness of every solution is determined by classification accuracy. The best accuracy indicates the highest level of fitness. The fitness value of each iteration should to be determined. The following formula can be used to calculate the fitness function.

$$F(Y) = Maximum(Accuracy) \tag{17}$$

Step 3: Updation using POA: To update potential solutions, the POA mimics the tactics and movements of pelicans during attack and prey-seeking. There are two steps to this hunting strategy: (i) exploration and (ii) exploitation. Below is a mathematical representation for each of these updating phases.

Phase 1: Approaching prey (Exploration): In the first stage, the pelicans locate the victim and then move in on it. The POA can more efficiently scan the search space and investigate a variety of areas inside it by adopting the pelican's approach. The POA's capacity to generate the prey's location at random, which enhances its exploration abilities, is one of its noteworthy characteristics. A mathematical simulation of the pelican's approach to the target position is given by equation (18).

$$Y_{i,j}^{L_1} = \begin{cases} Y_{i,j} + rand.(L_j - I.Y_{i,j}), & F_l < F_i; \\ Y_{i,j} + rand.(Y_{i,j} - L_j), & else; \end{cases} \tag{18}$$

Here, $Y_{i,j}^{L_1}$ denotes i-th pelican's current status in j-th dimension, I denotes a random number that can be either 1 or 2, L_j denotes prey's position the j-th dimension, and F_l denotes value of its fitness function. The parameter I is chosen at random for every iteration and every member. When this parameter's value is 2, a member is displaced further, potentially bringing them to more recent regions of the search space. Consequently, the parameter influences the POA's exploration capability I to precisely search the space.

A pelican's current position is accepted if it improves the fitness value. It keeps the algorithm from reaching less-than-ideal locations. This procedure is described in equation (19).

$$Y_i = \begin{cases} Y_i^{L_1}, & F_i^{L_1} < F_i; \\ Y_i, & else; \end{cases} \tag{19}$$

Where, $F_i^{L_1}$ denotes range of this fitness value, and $Y_i^{L_1}$ denotes i-th pelican's most recent position in accordance with Phase 1.

Phase 2: Winging on the water surface (Exploration): In the second stage, the pelicans use the tactic of spreading their wings to catch fish higher after they have reached the water's surface. The pelicans' throat pouch then collects fish that have been forced to the water's surface. More fish are caught in the targeted area as a result of this strategy. Its convergence towards more advantageous locations within the hunting region is improved by including this pelican behavior into the POA, which increases its capacity for local searches and exploitation efficiency. Mathematically speaking, the method needs to look at the points close to the pelican's location in order to converge to a better option. The behavior of pelicans during hunting is represented mathematically as follows:

$$Y_{i,j}^{L_2} = Y_{i,j} + R.\left(1 - \frac{1}{T}\right).(2.rand - 1).Y_{i,j} \tag{20}$$

Here, $Y_{i,j}^{L_2}$ denotes most recent location of the i-th pelican in j-th dimension, t denotes current iteration, T denotes maximum number of iterations, R denotes constant set at 0.2, and $R.(1-t/T)$ denotes neighborhood radius of $Y_{i,j}$. The coefficient $R.(1-t/T)$ denotes the neighborhood radius surrounding each population member, enabling local search activities close to each member and increasing the likelihood of finding a better solution. With the use of this parameter, the POA exploitation power can get closer to the optimal global solution. The current pelican location, which has also been approved or rejected at this point, is represented by equation (21).

$$Y_i = \begin{cases} Y_i^{L_2}, & F_i^{L_2} < F_i; \\ Y_i, & else; \end{cases} \tag{21}$$

Where, $F_i^{L_2}$ denotes this fitness value, and $Y_i^{P_2}$ denotes i-th pelican's most recent position in accordance with Phase 2.

Step 4: Termination: Once all population members have been updated through the first and second stages, the best solution thus far is updated based on the updated population and the fitness function. As the algorithm progresses to the next iteration, the POA updation phases using equations (18)–(21) are carried out repeatedly until the termination condition is met. Finally, the algorithm produces the optimal solution discovered during the iterations. The POA's pseudo-code is displayed in Algorithm 2.

Algorithm 2. Pseudo-code of POA

Input: Parameters
Output: Optimal Parameters
Start POA
1. Input information of the optimization problem.
2. Determine the number of iterations (T) and the POA population size (N).
3. Initialize the pelican location and compute the fitness function.
4. for $t = 1:T$
5.Set the victim location at random.
6. for $I = 1:N$
7. **Phase 1**
8. for $j = 1:m$
9. Determine j-th dimension's current status by applying equation (18).
10. end
11. Update i-th population member utilizing equation (19).
12. **Phase 2**
13. for $j = 1:m$
14. Determine the current status of j-th dimension utilizing equation (20).
15. end
16. Update i-th population member utilizing equation (21).
17. end
18. Update optimal solution.
19. End
20. POA provided the optimal solution (optimal parameter).
End POA

By employing the POA, the parameters of the MSCNet are optimized to ensure optimal classification performance. Hence, the dermoscopic image is effectively classified as AKIEC, BCC, BKL, DF, MEL, NV, and VASC classes using the proposed MSCNet. The integration of the stacked CapsNet structure enables a hierarchical feature abstraction, and the DDR algorithm improves CapsNet performance by enhancing feature routing and representation. Moreover, optimizing the initial weights and bias using the POA ensures faster convergence and improved generalization. These contributions significantly improve the accuracy of the proposed multi-class skin cancer detection framework.

4. Results And Discussion

This section presents the results of the proposed multi-class skin cancer detection framework. The proposed method uses an Intel Core i5 processor running on Windows 10 and 6 GB of RAM. Skin lesion images are collected from the HAM10000 dataset. The proposed method is implemented using Python. The proposed method performance is analyzed using several metrics and it is compared against existing methods. Table 2 shows the parameter settings of the proposed method.

Table 2. Simulation Settings.

Model	Parameter	Range
BAM-EfficientNet	Optimizer	Adam
	Learning Rate	0.0001
	Attention Module	BAM
	Batch Size	16
	Number of Epochs	50
	Activation Function	Swish
	Dropout Rate	0.3
	Random Seed	42
	Normalization	BN
MSCNet	Number of CapsNet Layers	2
	Optimizer	Adam
	Loss	Margin loss
	Batch Size	32
	Epochs	60
	Activation	Squash
	Learning Rate	0.0001
	Dropout Rate	0.4
	No. of channel in primary capsule layer	32
	Dimension of capsule in primary capsule layer	8
	Dimension of capsule in digit capsule layer	16
	No. of routing iterations	2
	Loss function	Margin loss
	random seed	42
	Weight/Bias Initialization	POA
	Routing Algorithm	DDR

4.1. Ablation Study

To validate the individual contributions of each component in the proposed framework, an ablation study was conducted on the HAM10000 dataset. The study systematically evaluates the impact of (i) BAM-EfficientNet-based feature extraction, (ii) stacked capsule architecture, (iii) DDR, and (iv) POA by incrementally incorporating these components into the baseline models. Performance is assessed using accuracy, precision, recall, and F1-score.

Table 3 presents the comparison between the baseline EfficientNetB7 and the proposed BAM-EfficientNet. Replacing the conventional SE blocks with BAM significantly enhances performance. The proposed BAM-EfficientNetB7 improves accuracy from 96.84% to 98.82% and F1-score from 96.75% to 98.66%. This improvement demonstrates that BAM enables the network to better focus on discriminative lesion regions by jointly exploiting spatial and channel attention, leading to more informative feature representations.

Table 3. Ablation results for BAM-EfficientNet.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)
EfficientNetB7 +SE	96.84	96.79	96.71	96.75
Proposed BAM-EfficientNetB7	98.82	98.89	98.76	98.66

Table 4 summarizes the ablation analysis of the proposed MSCNet model. The baseline CapsNet with DR achieves an accuracy of 94.62%, a stacked CapsNet with DR increases accuracy to 95.78%, indicating improved hierarchical feature modeling. When DDR is incorporated, the performance further improves, achieving 97.26% accuracy for the stacked CapsNet. This confirms that DDR enhances the routing process by improving capsule agreement and feature

coupling. Similarly, the integration of POA for optimizing network weights and biases contributes to additional gains by improving convergence and avoiding local minima. The stacked CapsNet with DR and POA achieves 97.01% accuracy. Finally, combining stacked capsules, DDR, and POA in the proposed MSCNet yields the highest performance, with an accuracy of 99.21%, precision of 99.03%, recall of 99.23%, and an F1-score of 99.20%.

Table 4. Ablation results for MSCNet.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)
CapsNet + DR	94.62	94.55	94.48	94.51
Stacked CapsNet + DR	95.78	95.69	95.63	95.66
CapsNet + DDR	96.41	96.35	96.29	96.32
Stacked CapsNet + DDR	97.26	97.18	97.12	97.15
CapsNet + DR + POA	96.03	95.96	95.90	95.93
Stacked CapsNet + DR + POA	97.01	96.94	96.88	96.91
Proposed MSCNet (Stacked + DDR + POA)	99.21	99.03	99.23	99.20

Table 5. Impact of POA on MSCNet Performance.

Initialization Method	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)
MSCNet with Xavier Initialization	97.68	97.55	97.61	97.58
MSCNet with He Initialization	97.92	97.84	97.79	97.81
MSCNet with Random Initialization	96.84	96.71	96.68	96.70
Proposed MSCNet with POA-based Initialization	99.21	99.03	99.23	99.20

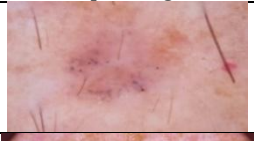
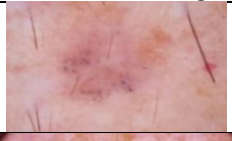
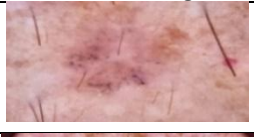
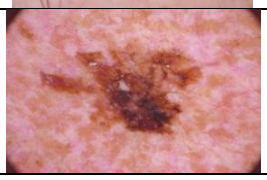
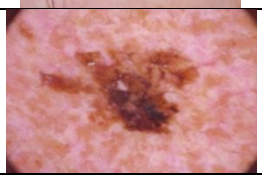
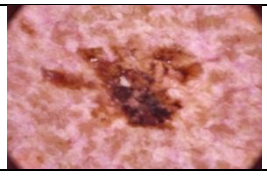
Table 5 presents a detailed comparison between standard initialization strategies (Xavier, He, and random initialization) and the proposed POA-based initialization within the MSCNet framework. Results show that MSCNet with random initialization achieves the lowest performance, with an accuracy of 96.84% and an F1-score of 96.70%, indicating suboptimal convergence due to poor initial parameter distribution. Xavier and He initialization strategies provide noticeable improvements, achieving accuracies of 97.68% and 97.92%, respectively, and F1-scores above 97.5%, confirming their effectiveness in stabilizing training and accelerating convergence. In contrast, the proposed POA-based initialization clearly outperforms all standard methods, achieving an accuracy of 99.21%, precision of 99.03%, recall of 99.23%, and an F1-score of 99.20%. This improvement demonstrates that POA is able to identify a more optimal initial search space for network parameters, enabling faster convergence to better local minima and improving overall generalization performance.

The ablation results clearly demonstrate that each component of the proposed framework significantly contributes to the overall performance. BAM-EfficientNet significantly improves feature discrimination, stacked capsule architecture enhances hierarchical representation learning, DDR strengthens capsule routing efficiency, and POA optimizes network parameters for better generalization. The integration of all components results in the proposed MSCNet achieving the best performance across all evaluation metrics.

4.2. Experimental Results

This section analyzes the findings from the proposed multi-class skin cancer detection framework. The accuracy-epoch, loss-epoch, confusion matrix, ROC, and AUC graph of the proposed method against existing methods are analyzed.

Table 6. Visual Representation of Dermoscopic Image Preprocessing Stage with Corresponding Class Labels.

Input Image	Noise Removed Image	Enhanced Image	Label
			BCC
			MEL

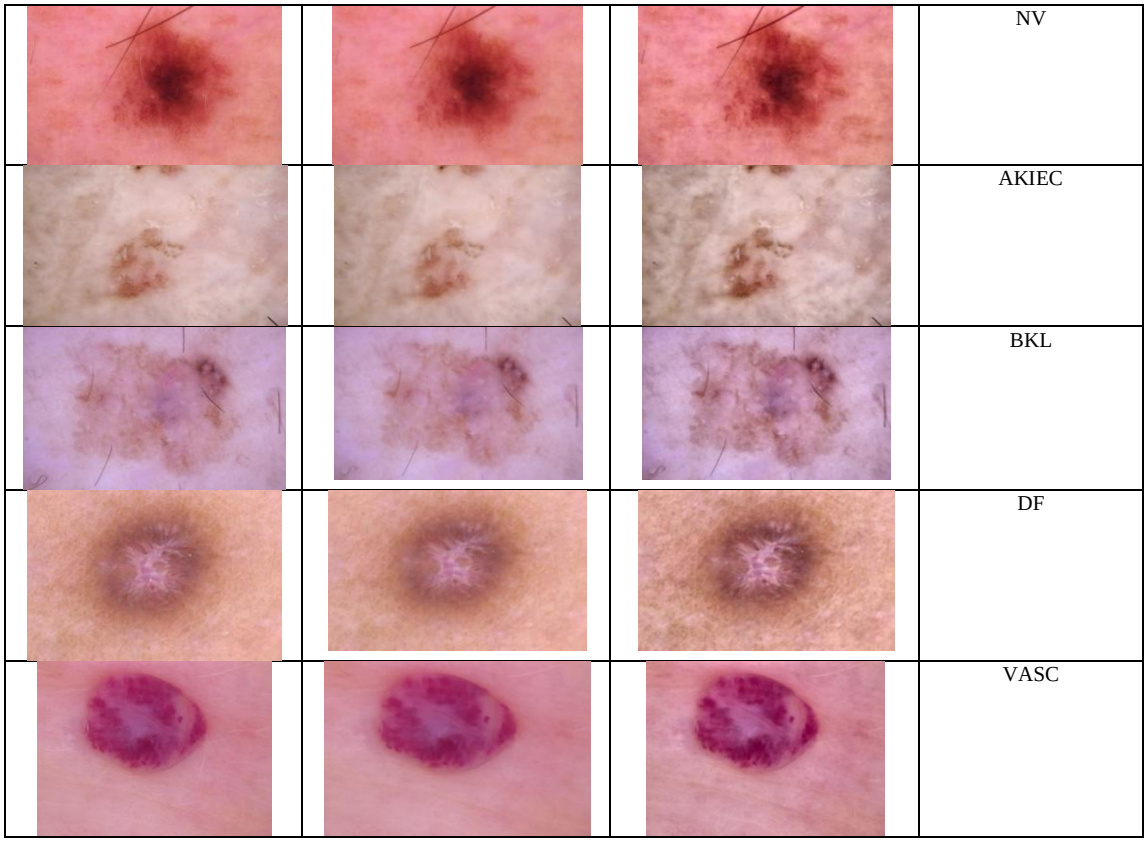


Table 6 shows the outputs of the dermoscopic image preprocessing stage with corresponding class labels. It includes original input image, noise removed image, enhanced image, and final associated class label. The Preprocessing stage is critical for preparing the data for accurate feature extraction and classification by the proposed MSCNet framework.

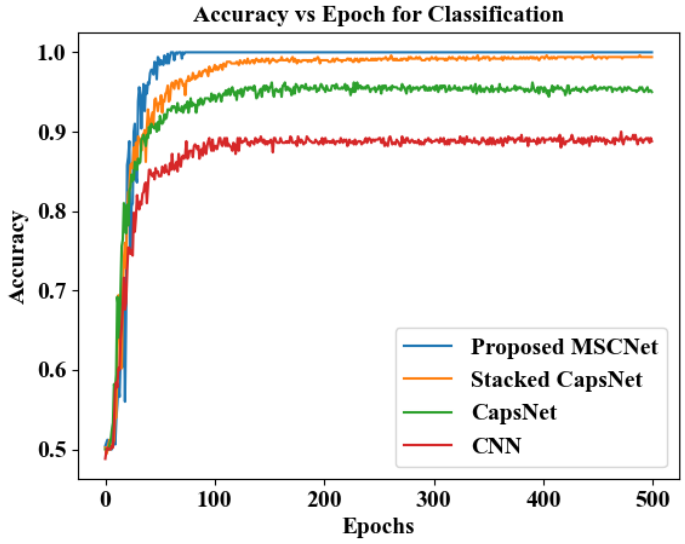


Fig. 7. Accuracy-Epoch graph.

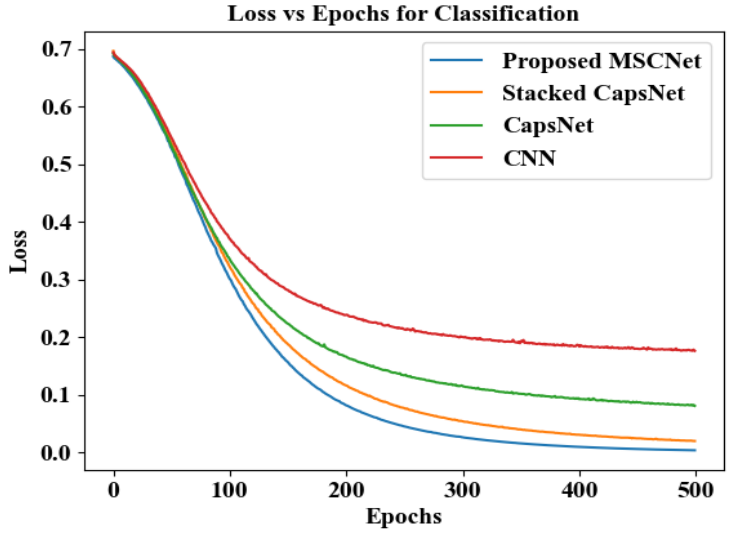


Fig. 8. Loss-Epoch graph.

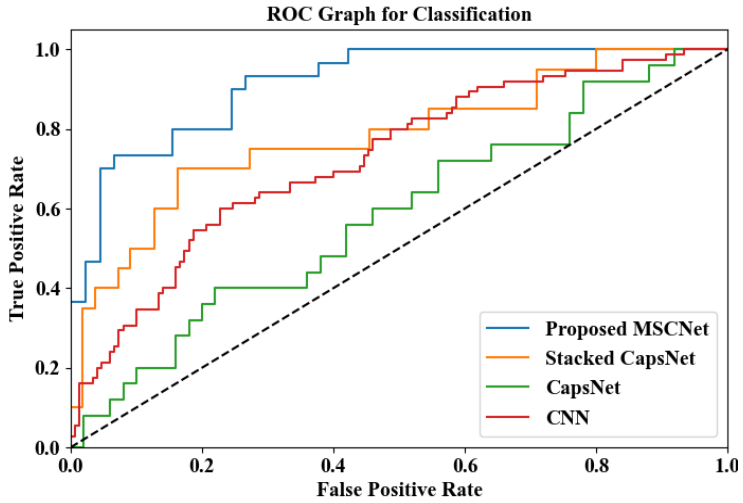


Fig. 9. ROC graph.

Figure 7 shows accuracy versus epoch graph for the proposed MSCNet method and existing methods such as stacked CapsNet, CapsNet and CNN over 500 training epochs. From the results, the accuracy of the suggested method gradually increases as number of epoch increases. Comparatively, the stacked CapsNet and CNN shows slightly lower accuracy than the proposed method but still perform better than CNN. The CNN model shows comparatively lower accuracy than the other existing techniques. The suggested MSCNet framework significantly surpasses the existing methods.

Figure 8 displays a loss versus epoch graph for the proposed MSCNet method compared with existing methods over 500 training epochs. From the graph, the suggested MSCNet consistently achieves the lowest loss and performs better than the existing methods such as stacked CapsNet, CapsNet and CNN. Comparatively, the stacked CapsNet and CapsNet show relatively better performance than CNN, but both still show higher loss compared to the proposed method. The CNN model shows a higher loss among all methods. Overall, the loss curve of the proposed method drops rapidly during the initial epochs and continues to gradually decrease as the number of epoch increases, indicating effective learning and convergence.

The ROC curve for the suggested MSCNet technique compared to other techniques is displayed in Figure 9. The trade-off between the TPR and the FPR at various threshold values is displayed by the ROC curve. It is clear from the graph that the suggested MSCNet approach performs better in classification, achieving a higher TPR with a lower FPR. The ROC curve for the suggested approach is closer to the upper-left corner, suggesting that it has a strong capacity for discrimination. In comparison, stacked CapsNet performs moderately well, followed by CapsNet, with CNN showing lower performance. The results show that the suggested MSCNet model performs better than the current techniques, including CNN, stacked CapsNet, and CapsNet.

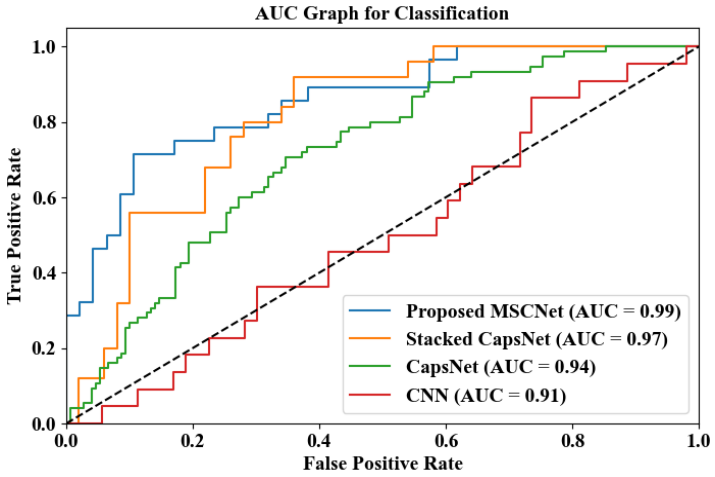


Fig. 10. AUC graph.

Figure 10 shows the AUC graph for the suggested MSCNet approach comparison with different approaches, including CNN, combined CapsNet, and CapsNet. The FPR and TPR are displayed on the x and y axes of the AUC graph, respectively. According to the graph, the suggested MSCNet approach achieves a higher AUC value of 99%, indicating superior predictive accuracy and classification ability. Following closely, the stacked CapsNet and CapsNet attain AUC values of 97% and 94%, respectively. Comparatively, the CNN model shows a lower AUC value of 91%, indicating relatively lower performance compared to all other methods. According to the results, the suggested MSCNet model performs noticeably better than the current models and is more successful in correctly classifying multiple types of skin cancer.

The confusion matrix for the suggested MSCNet approach is displayed in Figure 11. The proposed MSCNet method's classification performance across seven distinct skin lesion classes-AKIEC, BCC, BKL, DF, MEL, NV, and VASC-is displayed in the confusion matrix. The suggested multi-class classification approach yields the lowest possible rates of false positives and false negatives when evaluating confusion matrices. The results indicate that the suggested MSCNet approach provides high accuracy.

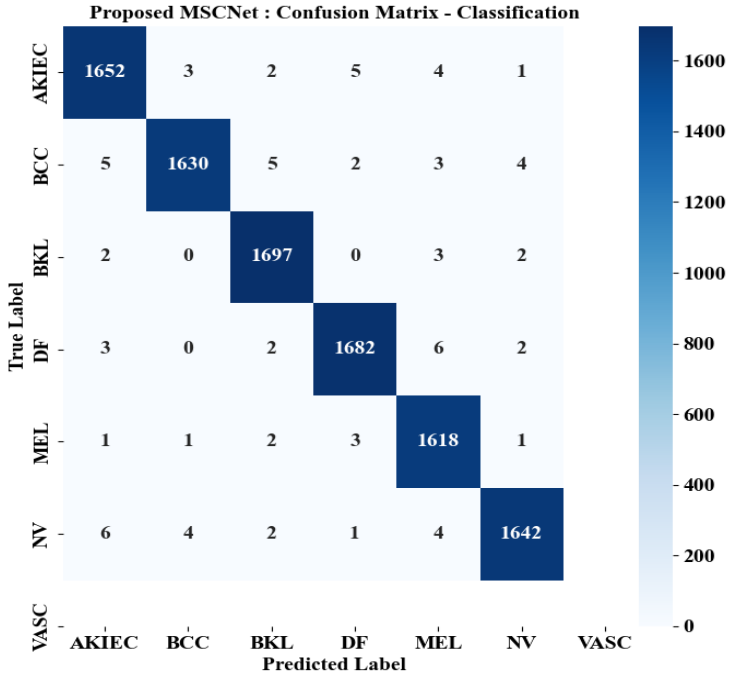


Fig. 11. Confusion matrix.

4.3. Results of comparative analysis

To demonstrate the effectiveness of the suggested approach, we contrast it with various approaches. Feature extraction and classification are the two main phases of the suggested MC-SC2 model. This section compares the

performance of both phases using different metrics, including f-score, accuracy, precision, recall, sensitivity, and specificity.

Figure 12 shows a performance comparison of feature extraction methods. The suggested BAM-EfficientNet-based feature extraction technique is compared with existing techniques, such as EfficientNetB7, ResNet50, and VGG16, using various metrics including, accuracy, precision, recall, sensitivity, specificity, and f-score. The presented BAM-EfficientNet achieves a highest accuracy of 98.83%, which is 2.81% better than EfficientNetB7, 5.97% better than ResNet50, and 9.37% better than VGG16. Additionally, the proposed method achieves a highest precision and recall of 98.90% and 98.76%, which is 2.77% and 2.46% higher than EfficientNetB7, 6.06% and 5.89%higher than ResNet50, and 9.70% and 9.33% higher than VGG16. Moreover, the proposed BAM-EfficientNet achieves a maximum sensitivity, and specificity of 98.67% and 98.72%, outperforming EfficientNetB7 by 2.51% and 2.32%, outperforming ResNet50 by 5.73% and 5.78%, and outperforming VGG16 by 9.33%and 9.38%, respectively. The suggested approach provides the greatest F-score of 98.66%, which is 9.30% higher than VGG16, 5.79% higher than ResNet50, and 2.34% higher than EfficientNetB7. From the results, the proposed BAM-EfficientNet model significantly outperforms the existing methods across all metrics. This significant improvement can be attributed to the integration of BAM into the EfficientNetB7 architecture, which improves the network performance by focusing on the most informative regions, thereby enhancing classification performance.

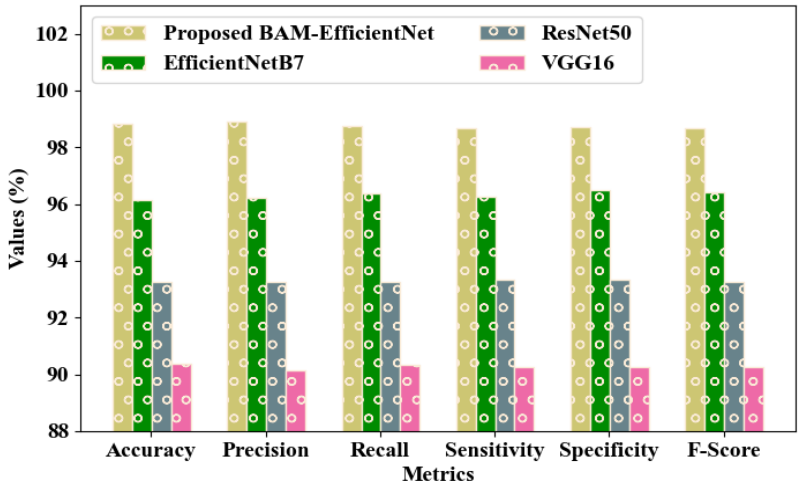


Fig. 12. Comparison based on feature extraction methods.

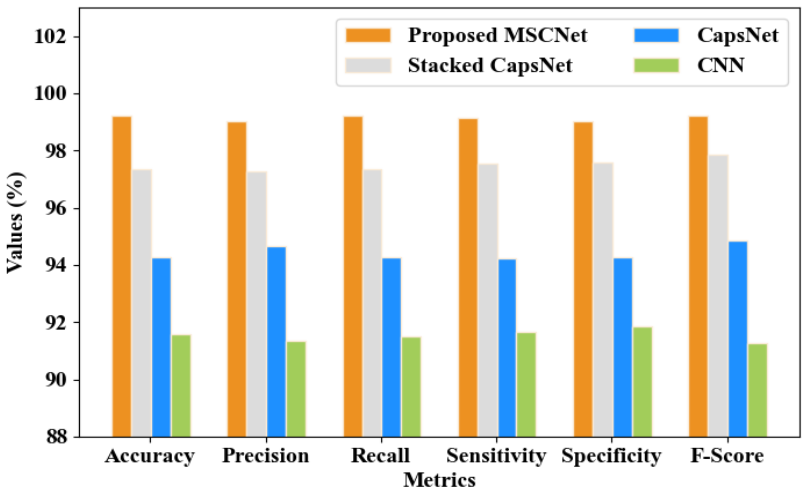


Fig. 13. Comparison based on classification methods.

Figure 13 compares the proposed MSCNet's classification results with the other approaches, including CNN, stacked CapsNet, and CapsNet, utilizing f-score, accuracy, precision, recall, sensitivity, and specificity. In every metric, the suggested MSCNet performs noticeably better than the current approaches. With a maximum accuracy of 99.21%, the suggested approach outperforms CNN by 8.33%, stacked CapsNet by 1.91%, and CapsNet by 5.26%. Additionally, the proposed method achieves a highest precision and recall of 99.04% and 99.24%, which is 1.83% and 1.92% higher than stacked CapsNet, 4.62% and 5.28% higher than CapsNet, and 8.39% and 8.42% higher than CNN. Moreover, the

proposed method achieves a maximum sensitivity, and specificity of 99.13% and 99.01%, outperforming stacked CapsNet by 1.62% and 1.46%, outperforming CapsNet by 5.22% and 5.05%, and outperforming CNN by 8.14% and 7.77%, respectively. A maximum F-score of 99.20% is attained by the suggested approach, outperforming stacked CapsNet by 1.37%, CapsNet by 4.59%, and CNN by 8.71%. This outcome clearly shows the proposed MSCNet's superior classification ability, which is due to the integration of BAM-EfficientNet for enhanced feature extraction, the use of disperse dynamic routing for enhanced capsule performance and POA for optimal parameter initialization. Because of these contributions, the model is now able to extract more relevant and discriminative features, which increases its efficiency for classifying skin cancer into several types.

Figure 14 compares the performance of the proposed MC-SC2 framework across the seven classes such as AKIEC, BCC, BKL, DF, MEL, NV, and VASC in terms of accuracy, precision, recall, sensitivity, specificity, and F1-score. From the results, the model achieved exceptionally high performance across all classes. NV, the majority class, attained the highest accuracy of 99.5%, with precision of 99.6%, recall of 99.5%, sensitivity of 99.5%, specificity of 99.8%, and F1-score of 99.5%. For the minority classes DF and VASC, the model achieved accuracies of 97.5% and 97.8%, respectively. Precision, recall, sensitivity, specificity, and F1-score for DF were 97.4%, 97.6%, 97.6%, 98.5%, and 97.5%, respectively, while for VASC they were 97.9%, 97.7%, 97.7%, 98.7%, and 97.8%, respectively. These results indicate that the data balancing and augmentation strategies effectively mitigated class imbalance, allowing reliable classification of underrepresented classes. Intermediate classes, including AKIEC, BCC, BKL, and MEL, also exhibited strong performance. AKIEC achieved 98.5% accuracy, 98.6% precision, 98.4% recall, 98.4% sensitivity, 99.0% specificity, and 98.5% F1-score. BCC achieved 99.0% accuracy, 98.9% precision, 99.1% recall, 99.1% sensitivity, 99.2% specificity, and 99.0% F1-score. BKL achieved 98.8% accuracy, 98.7% precision, 98.9% recall, 98.9% sensitivity, 99.1% specificity, and 98.8% F1-score. MEL achieved 98.9% accuracy, 99.0% precision, 98.8% recall, 98.8% sensitivity, 99.0% specificity, and 98.9% F1-score. The consistently high values of accuracy, precision, recall, sensitivity, specificity, and F1-score across all classes indicate that the model not only correctly identifies positive cases but also effectively minimizes false positives. Overall, these results confirm the effectiveness of the proposed MC-SC2 framework in multi-class skin cancer classification, demonstrating its potential as a reliable decision support tool for dermatologists, aiding in early detection and treatment planning.

Fig. 14. Comparison of the proposed framework across seven skin lesion classes.

4.4. Computational complexity analysis

This section presents the computational complexity of the proposed MSCNet. Table 7 compares the computational performance of the proposed model with different classifiers, evaluated in terms of the number of parameters, inference time, and floating-point operations (FLOPs).

Table 7. Computational performance of the proposed model.

Model	Parameters (Millions)	Inference Time (ms)	FLOPs (G)
CNN	3.9	2.6	0.42
CapsNet	11.6	34.8	2.9
Stacked CapsNet	22.4	61.5	5.1
Proposed MSCNet	24.0	24.1	4.6

As shown in Table 7, the conventional CNN classifier exhibits the lowest computational cost due to its simple architecture. In contrast, CapsNet shows increased inference time as a result of the DR mechanism, which is further amplified in the stacked CapsNet owing to multiple routing layers. Although the proposed MSCNet has a slightly higher parameter count than the stacked CapsNet, it achieves substantially lower inference time of 24.1 ms. This efficiency gain is mainly attributed to the DDR strategy, which improves routing efficiency by enhancing coupling coefficient distribution and reducing redundant computations. Overall, MSCNet provides a favorable trade-off between computational cost and classification accuracy, making it well suited for clinical skin cancer diagnosis, where diagnostic accuracy is prioritized over computational efficiency. Thus, while computational intensity is a consideration, our focus centers on achieving the highest levels of diagnostic accuracy.

4.5. Comparative analysis of the proposed work with other published studies

This section presents the comparative analysis of the presented MC-SC2 model with previously published works [11-18] to show the efficiency of the presented model. These previously published studies are reviewed in detail in the literature survey section. The findings of these published studies are compared with the suggested approach. A comparison between the suggested model and the most recent state-of-the-art models is presented in Table 8.

Table 8. Comparative analysis of the proposed work vs. published works.

Ref. No	Purpose	Dataset	Pre-processing Methods	Dataset Balancing	Dataset splitting ratio	Methods	Performance Analysis
[11]	Multi-Class Skin Cancer Recognition	HAM-10000, skin cancer dataset	Hair Removal, Color Correction, Augmentation	Yes	-	Modified ResNet50, EfficientNetV2S	Accuracy-97.80%, Precision-97.18%, Recall-94.19%, F-Score-95.56%
[12]	Multi-class skin cancer diagnosis	ISIC 2019, HAM-10000	Resized, Normalization, Augmentation	Yes	80:5:15	ViT, CNN	Accuracy-96.62%, Specificity-95.40%, F-Score-98.11%
[13]	Multi-Class Skin Cancer Detection	HAM-10000	-	No	-	DenseNet201	Accuracy-97%, Precision-97%, Recall-97%, F-Score-97%
[14]	Multi-Class Skin Cancer Diagnosis	HAM-10000	Resizing, Normalization, Augmentation	Yes	-	CNN	Accuracy-98.17%, Precision-93%, Recall-92%, F-Score-92%
[15]	Multiclass skin cancer classification	HAM-10000, ISIC-2019	Normalization, Reshaping, Random Oversampling	Yes	80:20	DCNN	Accuracy-97.11%, Precision-97.09%, Recall-97.12%, Specificity-97.61%, F-Score-97.08%
[16]	Multi-Class Skin Disease Classification	Kaggle Skin Diseases Image Dataset	Data Augmentation	Yes	80:10:10	ResNet50, EfficientNet-B2, EfficientNet-B0	Accuracy-97.14%, Precision-96.14%, Recall-95.14%, F-Score-94.14%
[17]	Multi-class skin lesion classification	ISIC-2019	Augmentation	Yes	70:15:15	Alex Net, transfer learning	Accuracy-90.9%, Precision-90.8%, F-Score-90.8%
[18]	Classification of Multi-Class Skin Cancer	ISIC-2019, PH2	Data Augmentation	Yes	80:20	kNN	Accuracy-94.44%, Precision-94.13%, Specificity-96.68%, F-Score-94.43%
Ours	**Multi-Class Skin Cancer Classification**	**HAM10000**	**Gaussian Filter, Augmentation**	**Yes**	**80:10:10**	**BAM-EfficientNet, MSCNet**	**Accuracy-99.21%, Precision-99.03%, Recall-99.24%, Sensitivity-99.13%, Specificity-99.01%, F-Score-99.20%**

Table 8 compares the suggested model to the most recent state-of-the-art models based on several metrics, including F-score, accuracy, precision, recall, and specificity. Table 2 shows that the suggested MC-SC2 approach had the greatest accuracy of 99.21%.In comparison, the accuracy reported in [11-18] are 97.80%, 96.62%, 97%, 98.17%, 97.11%, 97.14%, 90.9%, and 94.44%, respectively. Additionally, the suggested model attained a greatest precision and recall were 99.03% and 99.24%, respectively. Furthermore, the proposed method achieved a highest sensitivity, specificity and F-score of 99.13%, 99.01% and 99.20, respectively. Our suggested approach performs noticeably better than the most recent state-of-the-art techniques. These enhancements are due to the BAM-EfficientNet feature extraction method, which allows the mode to learn more important and discriminative features. Additionally, the POA

is used for optimal parameter initialization, and DDR is used in MSCNet to enhance capsule network performance. These contributions are helping the model to correctly identify the various forms of skin cancer.

5. Conclusion

This study presented a novel DL-based automatic system for multi-class skin cancer classification using dermoscopic images. The dermoscopic images were first preprocessed using a Gaussian filter to efficiently eliminate noise. To address the issue of class imbalance, a data balancing strategy based on image augmentation was applied. Then, the complex and prominent features were extracted using the proposed BAM-EfficientNet model. In order to enable the network to focus on the most pertinent regions of the skin lesion, BAM-EfficientNet substituted a BAM for each SE attention module in the EfficientNetB7 architecture. The extracted features were then fed into the proposed MSCNet in order to classify a skin lesion image as VASC, NV, DF, MEL, BCC, BKL, or AKIEC. The POA was used to optimize the network's initial weights and biases, while the MSCNet employed the DDR method to improve the performance of capsule networks. We used the HAM10000 dataset for experimental analysis. The highest accuracy of 99.21%, precision and recall of 99.03% and 99.24%, sensitivity and specificity of 99.13% and 99.01%, and F-score of 99.20% were all attained by the suggested MC-SC2 framework. The suggested approach performed noticeably better than the state-of-the-art techniques. The model demonstrated its ability to accurately identify the various forms of skin cancer from dermoscopic images. The suggested method provides dermatologists with useful support in correctly identifying various forms of skin cancer early on, which can help with early diagnosis and treatment planning, ultimately leading to better patient outcomes and less strain on healthcare systems. In our future endeavors, the classification performance can be further enhanced by integrating dermoscopic image data with clinical metadata (e.g., age, gender, and patient history) using multimodal neural networks, enabling more precise detection.

All the Declarations and Statements

Author Contributions Statement

Mudiyam Sivani – Conceptualization, Methodology, and Supervision: Proposed research ideas, Constructed the overall framework, and supervised project execution.
M. Shahina Parveen – Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing, and implementing the research model.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

The authors declare that they have no competing interests and funding.

Data Availability Statement

Data sharing is not applicable to this article because of proprietary nature.

Ethical Declarations

Not applicable.

Acknowledgments

Mudiyam Sivani and M. Shahina Parveen. Finally, we confirm that all individuals mentioned in this section have given their permission to be acknowledged.

Declaration of Generative AI in Scholarly Writing

Not applicable.

Abbreviations

The following abbreviations are used in this manuscript:

DL - Deep Learning
GF - Gaussian Filter
BAM-EfficientNet - Bottleneck Attention Module integrated with EfficientNetB7

MSCNet - Modified Stacked Capsule Network
DDR - Disperse Dynamic Routing
DF - Dermatofibroma
MEL - Melanoma
BKL - Benign Keratosis-like Lesions
BCC - Basal Cell Carcinoma
AKIEC - Actinic Keratoses and Intraepithelial Carcinoma
SVM - Support Vector Machines
DT - Decision Trees
NB - Na ïve Bayes

Appendix

Not applicable.

References

[1] Ozdemir, B., &Pacal, I. (2025). An innovative Deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms. *Results in Engineering*, *25*, 103692.

[2] DorathiJayaseeli, J. D., Briskilal, J., Fancy, C., Vaitheeshwaran, V., Patibandla, R. L., Syed, K., & Swain, A. K. (2025). An intelligent framework for skin cancer detection and classification using fusion of Squeeze-Excitation-DenseNet with Metaheuristic-driven ensemble deep learning models. *Scientific Reports*, *15*(1), 7425.

[3] Ali, K., Shaikh, Z. A., Khan, A. A., & Laghari, A. A. (2022). Multiclass skin cancer classification using EfficientNets–a first step towards preventing skin cancer. *Neuroscience Informatics*, *2*(4), 100034.

[4] Purni, J. T., &Vedhapriyavadhana, R. (2024). Eosa-net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks. *Journal of King Saud University-Computer and Information Sciences*, *36*(3), 102007.

[5] Rosendahl, C., &Marozava, A. (2023). *Dermatoscopy and Skin Cancer, updated edition: A handbook for hunters of skin cancer and melanoma*. Scion Publishing Ltd.

[6] Balaha, H. M., & Hassan, A. E. S. (2023). Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Computing and Applications*, *35*(1), 815-853.

[7] Pathania, Y. S., Apalla, Z., Salerni, G., Patil, A., Grabbe, S., &Goldust, M. (2022). Non-invasive diagnostic techniques in pigmentary skin disorders and skin cancer. *Journal of cosmetic dermatology*, *21*(2), 444-450.

[8] Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbour, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071.

[9] Mampitiya, L. I., Rathnayake, N., & De Silva, S. (2023). Efficient and low-cost skin cancer detection system implementation with a comparative study between traditional and CNN-based models. *Journal of Computational and Cognitive Engineering*, *2*(3), 226-235.

[10] Jütte, L., González-Villà, S., Quintana, J., Steven, M., Garcia, R., & Roth, B. (2024). Integrating generative AI with ABCDE rule analysis for enhanced skin cancer diagnosis, dermatologist training and patient education. *Frontiers in Medicine*, *11*, 1445318.

[11] Saeed, A., Shehzad, K., Ahmed, S., & Azar, A. T. (2025). EG-VAN: A Global and Local Attention based Dual-Branch Ensemble Network with Advanced color balancing for Multi-Class Skin Cancer Recognition. *IEEE Access*.

[12] Sakib, A. H., Siddiqui, M. I. H., Akter, S., Al Sakib, A., & Mahmud, M. R. (2025). LEVit-Skin: A balanced and interpretable transformer-CNN model for multi-class skin cancer diagnosis. *International Journal of Science and Research Archive [Internet]*, 1860-73.

[13] Haque, M. Z., &Morzina, M. S. Multi-Class Skin Cancer Detection: Implementing Various CNN Architectures with Mobile App Integration. *International Journal of Computer Applications*, *975*, 8887.

[14] Haque, M. I., Khatun, K., Rohman, G., Al Rakib, A., &Billah, M. M. (2025). A Deep Learning Framework for Multi-Class Skin Cancer Diagnosis: Addressing Class Imbalance and Performance Optimization Through Convolutional Neural Network Architectures. *European Journal of Computer Science and Information Technology*, *13*(1), 63-71.

[15] Houssein, E. H., Abdelkareem, D. A., Hu, G., Hameed, M. A., Ibrahim, I. A., &Younan, M. (2024). An effective multiclass skin cancer classification approach based on deep convolutional neural network. *Cluster Computing*, *27*(9), 12799-12819.

[16] Alruwaili, M., & Mohamed, M. (2025). An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification. *Diagnostics*, *15*(5), 551.

[17] Noaman, A., Ahmad, R., Khan, M. F., Mohammed, A. S., Farooq, M., & Adnan, K. M. (2025). Beyond binary: multi-class skin lesion classification with AlexNet transfer learning-towards enhanced dermatological diagnosis. *Discover Applied Sciences*, *7*(1), 1-18.

[18] Toprak, A. N., &Aruk, I. (2024). A Hybrid Convolutional Neural Network Model for the Classification of Multi-Class Skin Cancer. *International Journal of Imaging Systems and Technology*, *34*(5), e23180.

[19] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. 2015.

[20] Trojovský, P. and Dehghani, M., 2022. Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. *Sensors*, *22*(3), p.855.

Authors' Profiles

Mudiyam Sivani is a research scholar in Computer Science & Engineering at Dayananda Sagar University, Bangalore, India. She received her B.Tech degree in Computer Science & Engineering in 2021 and holds a M.Tech degree in Computer Science & Engineering in 2023 from Jawaharlal Nehru Technological University, Anantapur. Her research interests are in S, Machine Learning, Artificial intelligence and Data Science.

M. Shahina Parveen graduated in Computer Science & Engineering from Jawaharlal Nehru Technological University (JNTU) in 2005, holds a Master's degree at Acharya Nagarjuna University, Andhra Pradesh and a Ph.D. in Computer Science and Engineering from Jawaharlal Nehru Technological University (2019). She is currently working as a professor and Chairperson of the Department of Computer Science & Technology, School of Engineering Dayananda Sagar University, Bangalore, India. She has authored several papers in International Conference Proceedings and referred journals. Her current research interests include machine learning, data mining, neural Networks, and big data analytics.

How to cite this paper:

Mudiyam Sivani, M. Shahina Parveen, "A Novel Deep Learning Framework for Multi-Class Skin Cancer Classification Using Modified Stacked Capsule Network ", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.18, No.4, pp. 82-105, 2026. DOI:10.5815/ijigsp.2026.04.05