https://doi.org/10.31436/iiumej.v26i3.3680

BOUNDARY AWARE SALIENCY-BASED LEVEL SET WITH MOMENTUM CONTRAST METAFORMER FOR SKIN CANCER SEGMENTATION AND CLASSIFICATION

ANJANI GUPTA[1*], SNEHLATA SHEORAN[2], SURABHI RASTOGI[3], ANSHULA GUPTA[4], KAJAL SALUJA[5], SABNAM KUMARI[6]

[1,3,4]Department of CSE, School of Engineering & Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU-Block (Outer Ring Road) Pitampura, Delhi - 110034, India
[2]Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India
[5]Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India
[6]Department of Information Technology, Bhagwan Parshuram Institute of Technology, New Delhi, India

**Corresponding author: anjani.gupta@vips.edu*

(Received: 5 April 2025; Accepted: 1 August 2025; Published online: 9 September 2025)

ABSTRACT: Skin cancer is considered one of the most widespread and life-threatening cancers and remains a challenging task for dermatologists. This challenge arises due to the small boundaries and regions of affected tissue. Existing methods yield poor classification accuracy due to inefficient classifier performance in recognizing complex patterns. Therefore, an effective skin cancer segmentation and classification method called Boundary-Aware Saliency-based Level Set (BASLS) with Momentum Contrast Metaformer (MC-Metaformer) is proposed in this research. BASLS enables improved lesion segmentation by identifying significant structural components along lesion edges. Using residual connections, MC-Metaformer provides a better gradient path, addressing gradient fading issues during deep feature extraction. Preprocessing is performed on the HAM10000, ISIC-2019, and ISIC-2020 datasets to improve and standardize the data. ResNet50 is used to extract relevant features for classification. Experimental results demonstrate that the proposed MC-Metaformer outperforms the Deep Convolutional Neural Network (DCNN), achieving classification accuracies of 99.58%, 99.32%, and 98.62% on the HAM10000, ISIC-2019, and ISIC-2020 datasets, respectively. These results confirm the robustness and efficiency of the model in accurate skin cancer segmentation and classification.

ABSTRAK: Kanser kulit merupakan salah satu jenis kanser paling meluas dan mengancam nyawa serta masih menjadi cabaran utama kepada pakar dermatologi. Cabaran ini berpunca daripada sempadan dan kawasan kecil tisu yang terjejas. Kaedah sedia ada mempunyai ketepatan klasifikasi yang rendah kerana prestasi pengelasan yang kurang berkesan dalam mengenal pasti corak kompleks. Oleh itu, kajian ini mencadangkan kaedah segmentasi dan klasifikasi kanser kulit yang berkesan dikenali sebagai *Boundary-Aware Saliency-based Level Set* (BASLS) bersama *Momentum Contrast Metaformer* (MC-Metaformer). BASLS membolehkan segmentasi lesi yang lebih baik dengan mengenal pasti komponen struktur penting di sepanjang tepi lesi. Melalui penggunaan sambungan residu, MC-Metaformer menyediakan laluan kecerunan yang lebih baik bagi menangani isu kehilangan kecerunan semasa pengekstrakan ciri mendalam. Pra-pemprosesan dilakukan ke atas set data HAM10000, ISIC-2019, dan ISIC-2020 bagi meningkatkan serta menyeragam data. ResNet50 digunakan bagi mengekstrak ciri relevan untuk klasifikasi. Dapatan eksperimen menunjukkan bahawa MC-Metaformer yang dicadangkan mengatasi Rangkaian Neural Konvolusi Mendalam (DCNN) dengan ketepatan klasifikasi sebanyak 99.58%, 99.32%, dan

98.62% masing-masing pada set data HAM10000, ISIC-2019, dan ISIC-2020. Dapatan ini mengesahkan keteguhan dan kecekapan model dalam segmentasi dan klasifikasi kanser kulit secara tepat.

KEYWORDS: *Lesion Edges, Momentum Contrast, Relevant Features, Saliency-based Level Set, Skin Cancer Segmentation and Classification.*

1. INTRODUCTION

The term 'cancer' refers to a broad group of diseases that affect various parts of the human body. Abnormal growth of cancer cells invades neighboring tissues before metastasizing to other organs [1]. Uncontrolled cellular proliferation throughout the body leads to cell death, making cancer the second leading cause of death worldwide. Annual statistics indicate that cancer causes 9.6 million deaths, accounting for one-sixth of all fatalities [2]. Skin cancer is one of the most dangerous types of cancer and frequently affects individuals. It becomes life-threatening when not diagnosed and treated at an early stage [3]. Delayed detection increases mortality rates, as treatment becomes less effective in advanced stages. The two primary categories of skin cancer are melanoma and non-melanoma [4]. Skin pigment cells, known as melanocytes, produce melanomas within the skin tissue, often spreading to other organs. According to global statistics from 2020, non-melanoma skin cancer is the most prevalent subtype, accounting for 6.2% of all annual cancer diagnoses, although it generally results in lower mortality [5][6]. In contrast, the aggressive behavior of melanoma leads to higher fatality rates, despite its lower occurrence compared to non-melanoma [7]. Early detection of melanoma is critical due to its rapid progression. It typically develops on the back, arms, legs, and facial regions [8]. The primary risk factors for developing skin cancer include prolonged exposure to ultraviolet (UV) radiation, the presence of numerous moles, fair skin with sun sensitivity, and a family history of skin cancer [9].

Skin cancer diagnosis is challenging due to poor image quality, combined with factors such as physician experience, visual limitations, and fatigue. These elements contribute to inconsistent diagnostic outcomes among dermatologists [10]. Medical experts have developed computer systems for automatic diagnosis to assist dermatologists in analyzing dermoscopic images [11]. Automated image analysis provides diagnostic support for clinical decision-making; however, healthcare professionals face difficulties distinguishing between different skin cancer types, as their visual presentations often share overlapping characteristics [12]. The diagnostic accuracy of dermatologists in identifying skin cancer ranges from 62% to 80%, depending on their experience level [13]. The dermoscopy technique requires comprehensive training, as new dermatologists need considerable time to achieve precision in their diagnoses [14]. Artificial Intelligence (AI) and Machine Learning (ML) algorithms have driven significant progress in both medical image analysis and disease diagnosis [15]. Enhancements in medical diagnostic systems using Deep Learning (DL) algorithms have led to their widespread adoption across medical research fields [16]. The accuracy of high-performance deep neural networks is improved by expanding Convolutional Neural Networks (CNNs) through increased depth, width, and resolution parameters [17]. However, current skin cancer detection models face challenges in handling small and indistinct lesion boundaries that often blend into surrounding tissue, making accurate segmentation difficult. Traditional models fail to capture fine-grained structural details, especially in noisy or low-contrast images. Additionally, many classifiers underperform due to limited feature extraction and shallow architectures, resulting in poor recognition of complex patterns. This leads to misclassifying visually similar lesions, especially when trained on imbalanced or low-quality datasets, thereby

limiting their generalizability. To address these limitations, this research develops an efficient DL-based approach.

The key findings of this research are listed as follows:

- A Boundary-Aware Saliency-based Level Set (BASLS) with a Momentum Contrast Metaformer (MC-Metaformer) is proposed for effective skin cancer segmentation and classification.
- The BASLS enables improved lesion segmentation by identifying essential structural components along lesion edges. Furthermore, it utilizes adaptive level-set functions to preserve accurate lesion boundaries while maintaining optimal regional uniformity.
- The MC-Metaformer provides an improved gradient path through residual connections during feature extraction. Moreover, the combination of token mixing and depth-wise convolution enhances feature representation, thereby improving classification accuracy.
- In preprocessing, a dull razor removes hair artifacts and improves lesion boundary clarity, while Contrast-Limited Adaptive Histogram Equalization (CLAHE) increases contrast for better lesion visibility. Normalization is applied to scale pixel values, thereby improving model performance.
- ResNet50 extracts deep hierarchical features from skin lesion images using skip connections, which help prevent vanishing gradient issues and enhance model stability. Additionally, it captures complex patterns and textures, thereby enhancing classification performance.

The remainder of this research is organized as follows: Section 2 presents the literature review, Section 3 describes the proposed method, Section 4 provides the experimental results with discussion, and Section 5 concludes the study.

2. LITERATURE REVIEW

The existing research models introduced for skin cancer segmentation and classification are analyzed in this section.

Houssein et al. [18] suggested an efficient multiclass skin cancer classification using a Deep Convolutional Neural Network (DCNN). The DCNN addressed the class imbalance problem and efficiently handled complex visual inputs, rendering it appropriate for image classification. Furthermore, the preprocessing steps were refined to enhance the model's predictive capability with optimal classification accuracy. However, the model's efficiency was limited to smaller datasets, with a deteriorated success rate when tested on larger data quantities.

Ahmad et al. [19] developed a DeepLabv3+ and Vision-based Transformer (ViT) for segmenting and classifying skin lesions. ViT processed information across various image locations in the first layer due to the self-attention mechanism. Furthermore, DeepLabv3+ and ViT were trained from the beginning using selected hyperparameters. However, this method did not account for how different features related to one another, which impaired its accuracy in detecting skin cancers.

Pandey et al. [20] introduced a sparse dictionary-based CNN for skin cancer classification. The Non-Local Means (NLM) denoising and CNN were backed by a sparse dictionary, which utilized a multi-stage approach to improve input data quality, extract meaningful features, and enhance overall classification performance. The combined approach addressed challenges such

as feature extraction, noise robustness, and classification accuracy, making it more suitable for intricate image analysis. However, training this model was difficult due to gradient value decay.

Khan et al. [21] presented an early-stage melanoma diagnosis method from dermoscopic images using a DenseNet and CNN-based approach. The developed model included CNN, DenseNet, batch normalization, max pooling, and a Rectified Linear Unit (ReLU) layer to address the overfitting issue. Moreover, numerous test samples were included, and data augmentation was applied to address class imbalance. The Adam optimizer managed large datasets, effectively handling small gradients. Despite its strengths, this method produced lower-quality results when processing small lesions.

Pandimurugan et al. [22] proposed a DCNN-based model for the early identification and classification of melanoma skin cancer. Segmentation techniques were used to delineate cancer lesions, and DL in computer vision assisted in classifying these lesions. However, automatic detection of melanoma from dermoscopic samples to determine lesion classes was challenging. The model exhibited lower performance due to the images displaying inconsistent shapes and border characteristics.

From the overall analysis, each method proved ineffective in feature selection, as they chose poor-quality features. Furthermore, existing models produced low-quality results when evaluating smaller lesions. This research introduces MC-Metaformer technology to identify different types of skin cancer. ResNet50 captures both low-level texture details and high-level patterns using residual learning, thereby enabling the extraction of discriminative and high-level features. Furthermore, the Boundary-Aware Saliency-based Level Set (BASLS) segmentation enhances boundary localization by preserving complex structural components, particularly in small and indistinct lesions. This segmentation and feature representation improvement enables the model to perform superiorly on small lesion regions, as validated by higher accuracy across the HAM10000, ISIC-2019, and ISIC-2020 datasets.

3. PROPOSED METHODOLOGY

Identifying different types of skin cancer is challenging due to complex patterns and indistinct feature lines. This research introduces a novel DL method for skin cancer segmentation and classification. The proposed model involves five key steps: data collection, preprocessing, segmentation, feature extraction, and classification. The overall procedure for skin cancer segmentation and classification is illustrated in Fig. 1.

https://doi.org/10.31436/iiumej.v26i3.3680

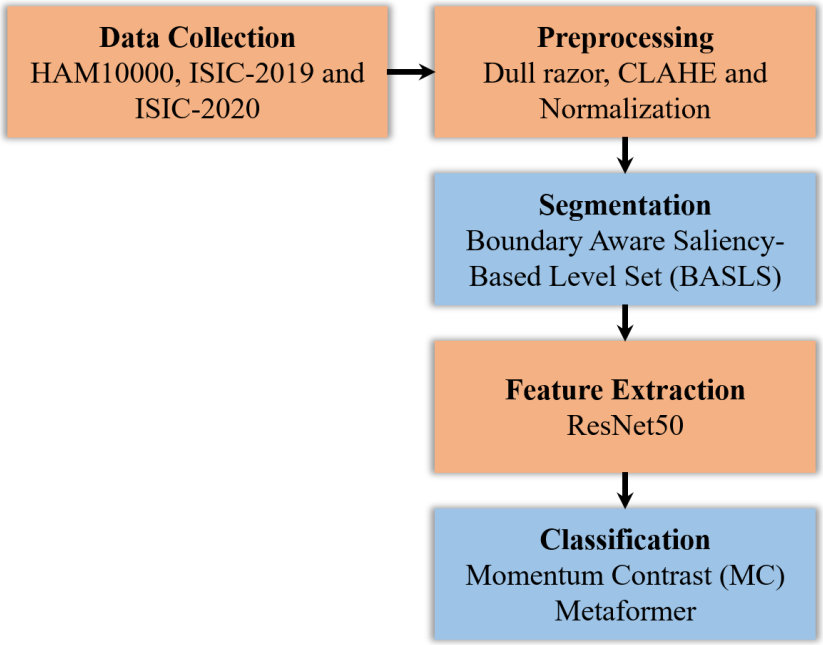

Figure 1. Process for skin cancer segmentation and classification

3.1. Data Collection

This research considers the HAM10000 [23], ISIC-2019 [24], and ISIC-2020 [25] datasets for skin cancer segmentation and classification. The following section provides a detailed description of these datasets.

3.1.1. HAM10000 dataset

The HAM10000 is a publicly available and widely employed dataset for skin cancer classification. It contains 10,105 images divided into seven classes: melanoma (MEL), Melanocytic Nevus (MN), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Benign Keratosis Lesions (BKL), Dermatofibroma (DF), and vascular lesions (VL). The class distributions and sample images are listed in Table 1.

Table 1. Class distributions of the HAM10000 dataset

Classes	MEL	MN	BCC	AK	BKL	DF	VL
No. of samples	1113	6705	514	327	1099	115	142

3.1.2. ISIC-2019 dataset

ISIC-2019 is a widely used dataset published by the International Skin Imaging Collaboration (ISIC). It comprises 25,331 images divided into eight classes: Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Dermatofibroma (DF), Squamous Cell Carcinoma (SCC), Vascular Lesion (VASC), Melanoma (MEL), and Melanocytic Nevus (NV). The class distribution and number of samples are listed in Table 2.

Table 2. Class distributions of the ISIC-2019 dataset

Classes	AK	BCC	BKL	DF	SCC	VASC	MEL	NV
No. of samples	867	3323	2624	239	628	253	4522	12875

3.1.3. ISIC-2020 dataset

ISIC-2020 is a publicly available and widely used dataset published by ISIC. In this dataset, histopathology is applied to confirm malignant diagnoses, while continuous monitoring and specialist consultation are used to verify benign cases. It includes 33,126 images categorized into seven classes: MEL, Seborrheic Keratosis (SK), Solar Lentigo (SL), Lentigo NOS (LNOS), Lichenoid Keratosis (LK), Nevus, and Unknown. The class distributions and sample images are presented in Table 3.

Table 3. Class distributions of the ISIC-2020 dataset

Classes	MEL	Nevus	LNOS	LK	SL	SK	Unknown
No. of samples	584	5193	44	37	7	135	27126

3.2. Preprocessing

The three datasets are preprocessed using a dull razor, CLAHE, and normalization, as explained in the following subsection.

Dull razor: This method removes hair from skin images to enhance their effectiveness for further analysis. Dull Razor operates according to the following process. 1) A standardized technique identifies dark hair areas across various image backgrounds. 2) The model utilizes bilinear interpolation to replace detected hair by accounting for its thin and elongated characteristics. 3) An adaptive median filter is applied to the replaced pixels, enhancing image clarity.

CLAHE: This method improves local contrast enhancement using the basic Histogram Equalization model. CLAHE does not operate on individual image pixels but instead adjusts contrast within specific regions. Bilinear interpolation handles transitions between these regions and reduces visual distortions. However, CLAHE may introduce unnatural visual effects by boosting contrast, adding unwanted artifacts to the original scenes.

Normalization: Scaling pixels across all areas prevents any feature from becoming more visible than the others [19]. This technique makes training easier by making models less unstable and helping them achieve consistent learning progress. The model becomes more efficient at handling various medical images by reducing its dependence on exact brightness levels.

3.3. Segmentation

The preprocessed images are given as input to BASLS for skin cancer segmentation. The traditional saliency-based segmentation methods segment images to produce maps based on their texture and color. However, the recognition of accurate cancer region boundaries by saliency-based segmentation is affected by low intensity. The level set detects boundaries without color and texture data. However, segmentation using a level set is affected by illumination problems. Since the level set requires data on the cancer forms, and saliency requires boundary indicator functions, they are integrated into this study as BASLS. Furthermore, the traditional level set utilizes an edge detector, but fails to detect accurate boundaries due to lower intensity and illumination issues. Hence, a boundary indicator function is applied to predict accurate skin cancer boundaries. Furthermore, the contour is incorporated as the zero level set of functions. Let φ be the level set determined by the domain Ω, Ω_0 is determined as the zero level set, while Ω_{in} is the region inside Ω_0 and the region outside Ω_0 is defined as Ω_{out}. The energy function $E(\varphi)$ is given by Eq. (1) Where, ε_{img} is the external

https://doi.org/10.31436/iiumej.v26i3.3680

energy determined by the image attribute, ε_{reg} is the internal energy applied as a level set growth constraint, and g_p is an edge detector, which is given in Eq. (2).

$$E(\varphi) = \varepsilon_{img}(\varphi, g_p) + \varepsilon_{reg}(\varphi, g_p) \tag{1}$$

$$g_p = \frac{1}{1+\frac{1}{2}(1-|\nabla I_\sigma|^2\rho^2)(|\nabla I_\sigma|^2\rho^2)} \tag{2}$$

where, I_σ denotes the image smoothed by a Gaussian filter with a standard deviation, and ρ is a threshold boundary function obtained from the standard deviation of the image to regulate the adaptive stopping speed. The boundary threshold function is mathematically expressed in Eq. (3).

$$\rho(I) = \frac{1+\sqrt{S(I_\sigma)}}{3} \tag{3}$$

The derived edge detector, such as the boundary indicator function, is applied in the gradient descent expression, as shown in Eq. (4), comprising three parts. The first part determines the distance-regularized term, which is applied to remove re-initialization. The second part provides long-term driving for the zero-level set model of target boundaries. The final term is an area created through a region-growing matrix and the boundary indicator function. This third part increases the region energy among neighboring targets and enhances the evolution rate.

$$\frac{\partial\varphi}{\partial t} = udiv\left(d_p(|\nabla_\varphi|)\nabla_\varphi\right) + \lambda\delta_\varepsilon(\varphi)div\left(g_\rho\frac{\nabla_\varphi}{|\nabla_\varphi|}\right) + (\alpha g_p + mY)\delta_\varepsilon(\varphi) \tag{4}$$

where, u, λ, α and m are weight coefficients which manage every parameter effect, $div(.)$ is a parameter applied to estimate the deviation, ∇ is a gradient of the level set received through the gradient operator, and δ_ε is the Dirac function, as mathematically expressed in Eq. (5).

$$\delta_\varepsilon(\varphi) = \begin{cases} \frac{1}{2\varepsilon}\left(1+cos\left(\frac{\pi\varphi}{\varepsilon}\right)\right) & |\varphi| \le \varepsilon \\ 0 & |\varphi| > \varepsilon \end{cases} \tag{5}$$

where, δ_ε is obtained from the Heavisine function. The Dirac function is helpful when deployed in conjunction with the integral. The boundary indicator line is integral to an active contour, which is estimated by including the Dirac function, as it achieves boundary adjustment to a zero level set. Subsequently, external energy is attained with the contour acting as a part of the growth stimulus, which is affected by the parameter ε applied to the Dirac function. A large ε enhances the contour-capturing range; however, a higher value reduces contour detection accuracy. Therefore, ε is usually set to 0.5.

3.4. Feature Extraction

The segmented images are passed into the feature extraction process to extract DL features. ResNet50 enhances classification accuracy by capturing significant details from image edges, shapes, textures, and distinctive patterns. Its residual blocks enable the network to learn deeper representations while maintaining stability for better model performance. ResNet50 uses skip connections to address the vanishing gradient problem, allowing the model to learn more effective features. As data progresses through the network, it first detects simple visual elements before identifying more complex object parts in deeper layers.

https://doi.org/10.31436/iiumej.v26i3.3680

3.5. Classification

The extracted features are passed into the skin cancer classification process. The MC algorithm enables unsupervised learning by implementing dynamic dictionary systems for contrastive learning and an adapted Noise-Contrastive Estimation (NCE) loss, which optimizes data representations. The loss function described in subsequent sections serves as the base loss function for this research. The algorithm employs a temperature parameter τ to distinguish a query q from k_+ among $k+1$ options, comparing q to one positive key k_+ and K negative keys. The learning model facilitates information retrieval from both similar and dissimilar image pairs using Eq. (6).

$$L_q = -log \frac{exp(q.k_+/T)}{\sum_{i=0}^{K} exp\ (q.k_i/T)} \tag{6}$$

MC allows for better model expansion by adding fresh mini-batches to the buffer while removing older ones. This technique decouples mini-batch parameters from dictionary size, enabling continuous updates with new mini-batch elements. To stabilize the learning process, the model uses momentum-based updates that gradually modify key parameters θ_k, while multiple query batches are updated with model parameters from the query model θ_q. These optimized parameters follow the gradient descent steps as shown in Eq. (7).

$$\theta_K \leftarrow m\theta_k + (1-m)\theta_{q'} \qquad m \in [0,1) \tag{7}$$

This updated model allows gradual changes, thereby preserving functionality and rendering the model more stable. After the Transformer, the Metaformer offers a universal design with an input embedder with two residual blocks that combine token mixers and a channel MLP. $Input\_Embedding$ converts the input feature map I into N tokens, each with a feature dimension of C as shown in Eq. (8).

$$X = Input\_Embedding(I) \tag{8}$$

The processor passes the embedded tokens through multiple Metaformer blocks to extract image features, as shown in Fig. 2. In the network design, each Metaformer block consists of two residual modules with normalization. A token mixer supports interactions between tokens in the first residual module. This mixer functions as an identity mapping or a simple MLP, as shown in Eq. (9).

$$Y_1 = Token\_Mixer(Norm(X)) + X \tag{9}$$

A normalization operation, known as $Norm$ appears before the residual modules. The second residual module includes an MLP with two layers and an activation function that follows them. Channel MLP improves feature selection by increasing the number of channels from C to rC. The channel dimension is extended from C to rC using text before returning to C. The operation is presented in Eq. (10).

$$Y_2 = \sigma(Norm(Y_1W_1)W_2) + Y_1 \tag{10}$$

where, $\sigma(.)$ is a function that performs the activation process, and W_1 and W_2 are residual modules builds its output from two learned linear transformation layers.

https://doi.org/10.31436/iiumej.v26i3.3680

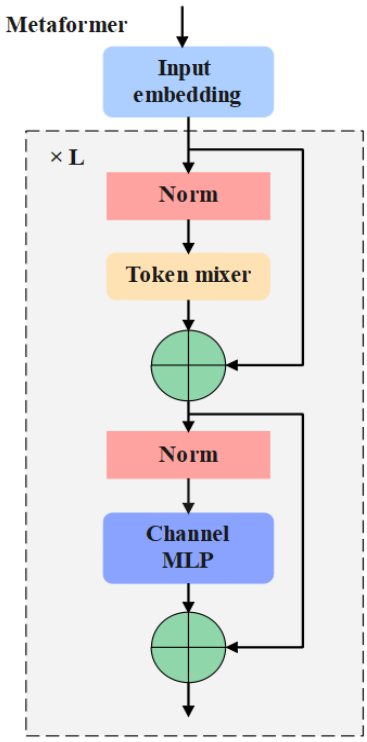

Figure 2. Structure of Metaformer

The proposed novel network design employs a 3×3 depth-wise convolution for token mixing to gather the spatial feature data. This method conserves more computer processing power than self-attention. The operation is defined in Eq. (11).

$$F_S = \delta\left(BN\left(DW(X)\right)\right) \tag{11}$$

where, $\delta(.)$ denotes the LeakyReLU activation element with a depth-wise convolution block. In the Metaformer architecture, the main task of the multilayer perceptron (MLP) is to extract channel-based features, requiring fewer computations than self-attention. The channel MLP component of the token mixer accounts for most of the network's computations when depth-wise convolutions or pooling operate within the token mixer. Experiments with DenseNet and CSPNet demonstrate that adding channel features through concatenation improves network accuracy more than simple element-wise addition. Channel expansion is important in Metaformer to capture more complex and diverse feature information. Metaformer employs a Channel Mixer (CM) module to connect its base blocks to the residual path. The new CM design achieves high overall performance with low computational demands, as shown in Fig. 2.

When the CM module is used, the first linear projection layer transforms the number of channels into $(r-1)C$. The produced channel diversity feature F_C combines the generated channels with the incoming features. The lower-layer feature joins with the spatial result F_S to form a mixed feature O_1 combines the texture details of the upper layers with different types of information. A linear projection layer compresses the features in the channel dimension and selects O_2 based on relevance, as shown in Eqs. (12) - (14).

$$F_C = \delta\left(BN\left(Conv_1(F_S)\right)\right) \tag{12}$$

$$O_1 = Concat\left(F_C, F_S\right) \tag{13}$$

https://doi.org/10.31436/iiumej.v26i3.3680

$$O_2 = \delta\left(BN\left(Conv_2(O_1)\right)\right) \tag{14}$$

where, $Conv_1$ and $Conv_2$ are 1×1 convolutional layers, respectively. CM normalization employs BN as its normalization technique, which runs faster than Layer Normalization (LN) and is a typical CNN component. This model uses a ratio of 2 between the input and output sizes to achieve better results.

4. RESULT ANALYSIS

This section describes the experiments conducted to evaluate the proposed method, including model construction and performance measurement based on specific standards.

4.1. Experimental setup and evaluation metrics

The proposed MC-Metaformer runs on an 8GB RAM and an Intel Core i5 processor that uses the Windows 10 operating system. The developed model measures its performance based on accuracy, F1-score, recall, precision, Dice coefficient, and Jaccard index metrics. Table 4 lists the evaluation criteria and corresponding formulas to determine the model performance. Where, TP, TN, FP and FN denote the True Positive, True Negative, False Positive, and False Negative.

Table 4. Evaluation metrics

Performance metrics	Formula
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
F1-score	$2 \times \frac{precision \times recall}{precision + recall}$
Recall	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
Dice coefficient	$\frac{TP \times 2}{TP \times 2 + FP + FN}$
Jaccard	$\frac{TP}{TP + FP + FN}$

4.2. Segmentation Performance Analysis

This section evaluates different segmentation methods through state-of-the-art techniques to verify their outcomes. The BASLS segmentation model is tested against the GAC, DRLS, and SLS models to determine its effectiveness in detecting skin lesion boundaries. Table 5 presents the segmentation performance of the proposed approach. The proposed BASLS method shows superior segmentation performance compared with other methods, as shown in Table 5. The method achieves 99.58% accurate segmentation of skin lesion areas on HAM10000, producing better results than GAC (93.23%), DRLS (95.37%), and SLS (97.13%). For ISIC-2019 and ISIC-2020, BASLS produces 99.32% accuracy, outperforming GAC, DRLS, and SLS. When segmenting ISIC-2020, BASLS achieves 98.62% accuracy, outperforming all other methods, including GAC (92.65%), DRLS (94.67%), and SLS (96.53%). BASLS generates better lesion segmentation by identifying and retaining important structural components when detecting lesion edges. This approach uses adaptive level-set functions to maintain accurate lesion boundaries with better regional uniformity. The optimized energy function helps BASLS avoid excessive and incomplete segmentations, leading to

improved scores in the Dice and Jaccard coefficient measurements. Understanding lesion boundaries during processing helps the model detect more lesions in different image sets

Table 5. Segmentation performance for all three datasets

Dataset	Method	Accuracy (%)	Recall (%)	Dice coefficient (%)	Jaccard (%)
HAM10000	GAC	93.23	92.81	88.53	81.63
	DRLS	95.37	94.82	91.36	85.27
	SLS	97.13	96.85	94.24	89.41
	BASLS	99.58	99.19	97.83	95.12
ISIC-2019	GAC	92.51	92.13	87.24	80.42
	DRLS	94.94	94.56	90.72	84.38
	SLS	96.84	96.43	93.53	88.07
	BASLS	99.32	98.87	97.65	94.82
ISIC-2020	GAC	92.65	91.74	86.89	79.92
	DRLS	94.67	94.18	90.23	83.57
	SLS	96.53	96.26	93.08	87.54
	BASLS	98.62	98.12	97.31	94.03

4.3. Feature Extraction Performance Analysis

This section presents the results of various feature extraction techniques compared to current state-of-the-art approaches. It compares the ResNet50's performance against that of the VGG16, DenseNet121, and ResNet18 baseline models, evaluating the efficiency of each structure in utilizing the preprocessed output to discover meaningful features. ResNet50 proves to be an efficient feature extractor, as shown in Table 6. The findings of this research indicate that ResNet50 works best in identifying useful image information. ResNet50 achieves 99.58% compared to 91.45% for VGG16, 94.72% for DenseNet121, and 96.81% for ResNet18 operating on the HAM10000 dataset. Under ISIC-2019 testing, ResNet50 delivered an accuracy of 99.32% above VGG16 at 90.88%, Densenet121 at 93.92%, and ResNet18 at 95.94%. ResNet50 achieved 98.62% accuracy on the ISIC-2020 dataset, surpassing VGG16 (89.77%), DenseNet121 (92.35%), and ResNet18 (94.57%). ResNet50 distinguishes itself from other networks by capturing surface-level visual details and complex lesion structures. Its deep structure contains several convolutional layers, which effectively help learn the visual characteristics at different levels.

Table 6. Feature extraction performance for all three datasets

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	VGG16	91.45	91.12	90.98	91.07
	DenseNet121	94.72	94.39	94.15	94.22
	ResNet18	96.81	96.52	96.37	96.41
	ResNet50	99.58	99.33	99.19	99.24
ISIC-2019	VGG16	90.88	90.54	90.31	90.47
	DenseNet121	93.92	93.51	93.28	93.42
	ResNet18	95.94	95.68	95.33	95.47
	ResNet50	99.32	98.87	98.53	98.68
ISIC-2020	VGG16	89.77	89.34	89.12	89.24
	DenseNet121	92.35	91.98	91.72	91.86
	ResNet18	94.57	94.31	94.05	94.19
	ResNet50	98.62	98.36	98.12	98.24

4.4. Classification performance analysis

This section describes how the proposed classification method outperforms existing techniques. The MC-Metaformer's performance is evaluated by processing skin cancer types and comparing the results against those of CNN and Metaformer-based approaches for precise classification analysis. Table 7 presents the classification performance of the proposed MC-Metaformer across three datasets. The MC-Metaformer classifier outperforms conventional methods based on the test results. It achieves 99.58% accuracy, surpassing CNN at 93.53%, MC at 95.54%, and Metaformer at 97.56% on the HAM10000 dataset. During the ISIC-2019 evaluation, MC-Metaformer attains an accuracy of 99.32%, exceeding CNN (93.29%), MC (95.31%), and Metaformer (97.35%). For both ISIC-2020 and ISIC-2019 datasets, MC-Metaformer scores over 98% accuracy, outperforming CNN, MC, and Metaformer. The model delivers superior performance by employing enhanced feature extraction and data representation techniques. While processing data, MC-Metaformer effectively detects connections between spatial elements and within channel spaces. It also reduces processing time without losing essential information through its Channel Mixer (CM) setup. The model enhances gradient flow by integrating residual connections, thereby addressing the vanishing gradient problem that occurs during deep feature extraction. The combined use of token mixing and depth-wise convolution architecture further enhances feature understanding, leading to better results across all datasets.

Table 7. Classification performance for all three datasets

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	CNN	93.53	93.27	93.18	93.22
	MC	95.54	95.28	95.15	95.20
	Metaformer	97.56	97.29	97.16	97.23
	MC-Metaformer	99.58	99.33	99.19	99.24
ISIC-2019	CNN	93.29	92.84	92.52	92.68
	MC	95.31	94.86	94.50	94.67
	Metaformer	97.35	96.88	96.49	96.66
	MC-Metaformer	99.32	98.87	98.53	98.68
ISIC-2020	CNN	92.63	92.38	92.11	92.24
	MC	94.68	94.39	94.14	94.26
	Metaformer	96.66	96.39	96.17	96.28
	MC-Metaformer	98.62	98.36	98.12	98.24

4.5. Comparative analysis

The MC-Metaformer demonstrates superior capabilities against the DCNN [18], ViT [19], Sparse Dictionary-Based CNN [20], Densenet-CNN [21], and DCNN [22]. The hybrid approach is tested against other techniques on HAM10000, ISIC-2019, and ISIC-2020 datasets, and the results are presented in Table 8. When a specific value was unavailable, this research used the symbol NA to indicate it. The results demonstrate that the MC-Metaformer surpasses all state-of-the-art methods across performance metrics. DCNN [18], Sparse Dictionary-Based CNN [20], and DenseNet-CNN [21] deliver lower accuracy compared to the MC-Metaformer, which achieves 99.58% accuracy on the HAM10000 dataset. The MC-Metaformer attains 99.32% accuracy on the ISIC-2019 dataset, outperforming DCNN [18], ViT [19], Sparse Dictionary-Based CNN [20], DenseNet-CNN [21], and DCNN [22]. When processing the ISIC-2020 dataset, MC-Metaformer achieves 98.62% accuracy, surpassing existing models such as ViT and DCNN in labeling precision. This method confirms that MC-Metaformer

produces superior results by effectively capturing more discriminative image features compared to other skin cancer classification approaches.

Table 8. Comparative analysis for proposed MC-Metaformer

Dataset	Method	Accuracy (%)	F1-score (%)	Recall (%)	Precision (%)
HAM10000	DCNN [18]	98.5	98.48	98.51	98.56
	Sparse dictionary-based CNN [20]	85.61	NA	NA	NA
	DenseNet-CNN [21]	93.24	69.68	70.59	68.80
	MC-Metaformer	99.58	99.33	99.19	99.24
ISIC-2019	DCNN [18]	97.11	97.08	97.12	97.09
	ViT [19]	97.73	NA	NA	NA
	Sparse dictionary-based CNN [20]	81.23	NA	NA	NA
	DenseNet-CNN [21]	95.70	80.34	80.93	79.76
	DCNN [22]	88.82	NA	NA	NA
	MC-Metaformer	99.32	98.87	98.53	98.68
ISIC-2020	ViT [19]	96.97	NA	NA	NA
	DCNN [22]	93.45	NA	NA	NA
	MC-Metaformer	98.62	98.36	98.12	98.24

4.6. Discussion

The proposed model achieves strong results and demonstrates clear benefits, as analyzed using HAM10000, ISIC-2019, and ISIC-2020 datasets. The MC-Metaformer model outperforms others in multiclass classification and feature extraction capabilities across all three datasets. The ResNet50 architecture effectively identifies features that yield the highest accuracy in disease classification. On the HAM10000 dataset, ResNet50 achieves 99.58% accuracy, surpassing state-of-the-art methods such as VGG16 (91.45%), DenseNet121 (94.72%), and ResNet18 (96.81%). For both ISIC-2019 and ISIC-2020 datasets, the MC-Metaformer outperforms CNN, Metaformer, and MC models. Metaformer alone also shows better performance than previous approaches, with accuracies of 99.32% on ISIC-2019 and 98.62% on ISIC-2020, outperforming CNN (92.63%), MC (94.68%), and Metaformer (96.66%) on ISIC-2020. The MC-Metaformer's superior performance stems from DL techniques that efficiently extract and classify complex skin lesion features. The classifier detects diverse lesion patterns more effectively by leveraging depth-wise convolution and the Channel Mixer (CM). Furthermore, the proposed method surpasses DCNN, ViT, and Sparse Dictionary-Based CNN in skin cancer detection accuracy. Integrating the BASLS segmentation method enhances lesion boundary detection, further improving medical image classification. Overall, the contrastive learning approach with dynamic dictionaries delivers robust results, establishing the proposed methodology as the leading technique for skin cancer diagnosis.

5. CONCLUSION

The MC-Metaformer significantly enhances the accuracy of skin cancer classification, aiding dermatologists in precise early detection. ResNet50 is employed for feature extraction, effectively capturing small and complex image patterns necessary for multiclass classification. The BASLS segmentation method enhances lesion edge delineation, thereby improving the proposed approach's effectiveness in identifying various skin lesion types. This method addresses two key limitations of previous techniques: inaccurate boundary localization and

poor generalization of deep features. The model introduces a novel framework that achieves superior classification performance by integrating BASLS for precise lesion segmentation with MC-Metaformer for robust representation learning. The MC-Metaformer demonstrates excellent performance and outperforms the existing classification models. The results demonstrate that the proposed method outperforms other methods, including CNN, MC, and Metaformer, with accuracies of 99.58%, 99.32%, and 98.62% on the HAM10000, ISIC-2019, and ISIC-2020 datasets, respectively. The consistent performance of CNN, MC, and Metaformer highlights the contribution of this work in optimizing segmentation boundaries and enhancing multiclass discrimination using contrastive learning and depthwise convolution. The proposed model improves both efficiency and accuracy across datasets. Future research will explore other medical datasets with novel feature selection methods for enhanced performance. Incorporating interpretability tools, the model can be improved to make it accessible and valuable for dermatologists, thereby aiding in early skin cancer detection.

REFERENCES

[1] Alabdulkreem E, Elmannai H, Saad A, Kamil IS, Elaraby A. (2024) Deep Learning-Based Classification of Melanoma and Non-Melanoma Skin Cancer. Traitement du Signal, 41(1): 213-223. https://doi.org/10.18280/ts.410117

[2] Mavaddati S. (2025) Skin cancer classification based on a hybrid deep model and long short-term memory. Biomedical Signal Processing and Control, 100A: 107109. https://doi.org/10.1016/j.bspc.2024.107109

[3] Behera N, Singh AP, Rout JK, Balabantaray BK. (2024) Melanoma skin cancer detection using deep learning-based lesion segmentation. International Journal of Information Technology, 16(6):3729-3744. https://doi.org/10.1007/s41870-024-02004-8

[4] Saha S, Hemal MM, Eidmum MZA, Mridha MF. (2025) Knowledge distillation approach for skin cancer classification on lightweight deep learning model. Healthcare Technology Letters, 12(1):e12120. https://doi.org/10.1049/htl2.12120

[5] Teodoro AAM, Silva DH, Rosa RL, Saadi M, Wuttisittikulkij L, Mumtaz RA, Rodriguez DZ. (2023) A skin cancer classification approach using gan and RoI-based attention mechanism. Journal of Signal Processing Systems, 95(2):211-224. https://doi.org/10.1007/s11265-022-01757-4

[6] Toprak AN, 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. https://doi.org/10.1002/ima.23180

[7] Mushtaq S, Singh O. (2024) A deep learning based architecture for multi-class skin cancer classification. Multimedia Tools and Applications, 83(39):87105-87127. https://doi.org/10.1007/s11042-024-19817-1

[8] Sabir R, Mehmood T. (2024) Classification of melanoma skin Cancer based on Image Data Set using different neural networks. Scientific Reports, 14(1):29704. https://doi.org/10.1038/s41598-024-75143-4

[9] Mui-zzud-din, Ahmed KT, Rustam F, Mehmood A, Ashraf I, Choi GS. (2024) Predicting skin cancer melanoma using stacked convolutional neural networks model. Multimedia Tools and Applications, 83(4):9503-9522. https://doi.org/10.1007/s11042-023-15488-6

[10] Alhudhaif A, Almaslukh B, Aseeri AO, Guler O, Polat K. (2023) A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks. Chaos, Solitons & Fractals, 170:113409. https://doi.org/10.1016/j.chaos.2023.113409

[11] Al Mahmud A, Azam S, Khan IU, Montaha S, Karim A, Haque A, Hasan ZM, Brady M, Biswas R, Jonkman M. (2024) SkinNet-14: a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. Neural

Computing and Applications, 36(30):18935-18959. https://doi.org/10.1007/s00521-024-10225-y

[12] Subhashini G, Chandrasekar A. (2024) Hybrid deep learning technique for optimal segmentation and classification of multi-class skin cancer. The Imaging Science Journal, 72(8):1043-1064. https://doi.org/10.1080/13682199.2023.2241794

[13] Jaisakthi SM, Mirunalini P, Aravindan C, Appavu R. (2023) Classification of skin cancer from dermoscopic images using deep neural network architectures. Multimedia Tools and Applications, 82(10):15763-15778. https://doi.org/10.1007/s11042-022-13847-3

[14] Verma N, Ranvijay, Yadav DK. (2024) Hybrid of Deep Feature Extraction and Machine Learning Ensembles for Imbalanced Skin Cancer Datasets. Experimental Dermatology, 33(12):e70020. https://doi.org/10.1111/exd.70020

[15] Ahmed N, Tan X, Ma L. (2023) A new method proposed to Melanoma-skin cancer lesion detection and segmentation based on hybrid convolutional neural network. Multimedia Tools and Applications, 82(8):11873-11896. https://doi.org/10.1007/s11042-022-13618-0

[16] Himel GMS, Islam MM, Al-Aff KA, Karim SI, Sikder MKU. (2024) Skin cancer segmentation and classification using vision transformer for automatic analysis in dermatoscopy-based noninvasive digital system. International Journal of Biomedical Imaging, 3022192. https://doi.org/10.1155/2024/3022192

[17] Khan MA, Alam S, Ahmed W. (2025) Enhanced Skin Cancer Diagnosis via Deep Convolutional Neural Networks with Ensemble Learning. SN Computer Science, 6:124. https://doi.org/10.1007/s42979-024-03581-y

[18] Houssein EH, Abdelkareem DA, Hu G, Hameed MA, Ibrahim IA, Younan M. (2024) An effective multiclass skin cancer classification approach based on deep convolutional neural network. Cluster Computing, 27:12799-12819. https://doi.org/10.1007/s10586-024-04540-1

[19] Ahmad I, Amin J, Lali MI, Abbas F, Sharif MI. (2024) A novel Deeplabv3+ and vision-based transformer model for segmentation and classification of skin lesions. Biomedical Signal Processing and Control, 92:106084. https://doi.org/10.1016/j.bspc.2024.106084

[20] Pandey A, Teja MS, Sahare P, Kamble V, Parate M, Hashmi MF. (2024) Skin cancer classification using non-local means denoising and sparse dictionary learning based CNN. Journal of Electrical Systems and Information Technology, 11:36. https://doi.org/10.1186/s43067-024-00162-0

[21] Khan AR, Mujahid M, Alamri FS, Saba T, Ayesha N. (2025) Early-Stage Melanoma Cancer Diagnosis Framework for Imbalanced Data From Dermoscopic Images. Microscopy Research and Technique, 88(3): 797-809. https://doi.org/10.1002/jemt.24736

[22] Pandimurugan V, Ahmad S, Prabu AV, Rahmani MKI, Abdeljaber HAM, Eswaran M, Nazeer J. (2024) CNN-Based Deep Learning Model for Early Identification and Categorization of Melanoma Skin Cancer Using Medical Imaging. SN Computer Science, 5:911. https://doi.org/10.1007/s42979-024-03270-w

[23] HAM10000 dataset link: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000 (Accessed on March 2025).

[24] ISIC-2019 dataset link: https://www.kaggle.com/datasets/andrewmvd/isic-2019 (Accessed on March 2025).

[25] ISIC-2020 dataset link: https://challenge2020.isic-archive.com/ (Accessed on March 2025).