



OPEN ACCESS

EDITED BY

Vinayakumar Ravi,
Prince Mohammad bin Fahd University,
Saudi Arabia

REVIEWED BY

Mahaboob S. Hussain,
Vishnu Institute of Technology (VITB), India
Salwan Tajjour,
Shoolini University, India
Mohamed Fouad,
Zagazig University, Egypt

*CORRESPONDENCE

Seong-Han Kim
✉ shkim8@sejong.ac.kr

†These authors have contributed equally to
this work

RECEIVED 07 November 2024

ACCEPTED 17 February 2025

PUBLISHED 10 March 2025

CITATION

Amin J, Azhar M, Arshad H, Zafar A and
Kim S-H (2025) Skin-lesion segmentation
using boundary-aware segmentation
network and classification based on
a mixture of convolutional and transformer
neural networks.
Front. Med. 12:1524146.
doi: 10.3389/fmed.2025.1524146

COPYRIGHT

© 2025 Amin, Azhar, Arshad, Zafar and Kim.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks

Javaria Amin^{1†}, Marium Azhar², Habiba Arshad², Amad Zafar^{3†}
and Seong-Han Kim^{3*}

¹Rawalpindi Woman University, Rawalpindi, Pakistan, ²Department of Computer Science, University of Wah, Wah Cantt, Pakistan, ³Department of Artificial Intelligence and Robotics, Sejong University, Seoul, Republic of Korea

Background: Skin cancer is one of the most prevalent cancers worldwide. In the clinical domain, skin lesions such as melanoma detection are still a challenge due to occlusions, poor contrast, poor image quality, and similarities between skin lesions. Deep-/machine-learning methods are used for the early, accurate, and efficient detection of skin lesions. Therefore, we propose a boundary-aware segmentation network (BASNet) model comprising prediction and residual refinement modules.

Materials and methods: The prediction module works like a U-Net and is densely supervised by an encoder and decoder. A hybrid loss function is used, which has the potential to help in the clinical domain of dermatology. BASNet handles these challenges by providing robust outcomes, even in suboptimal imaging environments. This leads to accurate early diagnosis, improved treatment outcomes, and efficient clinical workflows. We further propose a compact convolutional transformer model (CCTM) based on convolution and transformers for classification. This was designed on a selected number of layers and hyperparameters having two convolutions, two transformers, 64 projection dimensions, tokenizer, position embedding, sequence pooling, MLP, 64 batch size, two heads, 0.1 stochastic depth, 0.001 learning rate, 0.0001 weight decay, and 100 epochs.

Results: The CCTM model was evaluated on six skin-lesion datasets, namely MED-NODE, PH2, ISIC-2019, ISIC-2020, HAM10000, and DermNet datasets, achieving over 98% accuracy.

Conclusion: The proposed model holds significant potential in the clinical domain. Its ability to combine local feature extraction and global context understanding makes it ideal for tasks like medical image analysis and disease diagnosis.

KEYWORDS

skin lesion, compact convolution transformer, tokenizer, dermoscopy, hybrid loss, ResNet-34

1 Introduction

The skin is a major part of the human body and consists of the epidermis, dermis, lymphatic vessels, muscles, blood vessels, subcutaneous tissue, and nerves (1). Apart from protecting the whole body and various organs against external invasions such as chemical damage, the skin can also prevent adventitious viruses (2). Regardless of its protective and barrier functions, the skin is destructible; it tends to be affected by a diversity of genetic and external factors. Deterioration of lipids in the epidermis can be prevented by using liquids to improve the barrier features of the skin. Fungal development on the skin owing to allergic reactions, hidden bacteria, degradation of skin texture due to microbial reactions, and pigment creation can lead to various skin diseases (3). Skin diseases are considered chronic and may infrequently propagate into malicious tissues. Skin diseases must be treated promptly to restrict their growth and propagation (4). In recent years, imaging-technology-based research to identify the effects of various skin diseases has been in high demand. In most cases, different skin diseases may have similar appearances, making early detection of skin diseases difficult. Owing to the lack of contrast between adjacent tissues, predicting the type of skin lesion is also challenging. Other systems are unable to manage environmental- and texture-based variations in the input image, which is a significant concern because environmental and lighting conditions cannot always be controlled. The sheer number of skin diseases, combined with the difficulties caused by diverse environments and limited datasets, make skin disease classification a significant challenge (5). Computer-aided diagnostic techniques are preferred to classify skin diseases reliably and efficiently, assisting in medication prescription (6). Diseased growth propagation is assessed using a grey-level co-occurrence vector. To improve medication and minimize treatment costs, accurate diagnosis is essential to thoroughly assess abnormalities.

Owing to the limited and false distribution of experienced dermatologists, effective and efficient diagnosis using data-driven approaches is required as skin diseases are spreading in various shapes. The growing trend towards photonics-based and laser medical technology has made accurate and robust diagnosis of skin diseases feasible. However, these treatments are expensive and have limited applications. To address this issue, researchers are in search of more robust solutions, and convolutional neural networks (CNN) have been considered (7, 8). The major constraint in detecting skin diseases using CNNs is that they tend to learn and represent the bias inherent in the training data (9). For instance, the diagnostic accuracy of lesions on light skin is higher than on dark skin. This is because there may have been insufficient dark skin samples with the same lesion in the training set, or the image markers of protective factors and disease-affected regions may have an inherent correlation.

Approaches based on deep learning (DL) are considered more efficient than these techniques in classifying diseased parts from images in a dataset (6, 10–12). The growing need in healthcare diagnosis is to detect abnormalities precisely and to classify the category of disease from various types of biomedical images, such as magnetic resonance imaging, positron emission tomography, X-ray, and computed tomography scan data in the form of signals, that is, electroencephalogram (EEG), electrocardiogram (ECG), and electromyography (EMG) (13–19). Better treatment of patients according to the type of disease can be achieved by precisely identifying the disease category. Critical problems can be solved using DL models, enabling them to automatically detect input features. The inferred data can be obtained through deep-learning-based models that use unexposed data patterns to identify data features. Even DL models with low computational costs can result in optimal efficiency.

Owing to the challenges in previous studies, the key objectives of the proposed methodology are as follows:

- To develop an optimized deep-learning model for skin lesion segmentation by integrating prediction and refinement modules for enhanced accuracy.
- To utilize a hybrid loss function (SSIM, IoU, BCE) for improved segmentation performance by preserving structural details and optimizing overlap measures.
- To propose a compact convolutional transformer model (CCTM) with optimized layers and hyperparameters for efficient and accurate classification of skin lesions.
- To evaluate the performance of the proposed model against existing state-of-the-art methods in terms of segmentation accuracy, computational efficiency, and classification effectiveness.

To achieve the abovementioned objectives, a deep learning-based solution is proposed, whose major contributions are as follows:

- The proposed BASNet model comprises prediction and refinement modules for skin-lesion segmentation. It is trained from scratch on hybrid loss, which is a fusion of structural similarity, intersection over union, and binary cross-entropy.
- The prediction module works like a U-Net and is densely supervised by the encoder and the decoder. The encoder contains an input convolution layer and six stages. Four stages were adopted from ResNet-34 and retained from the basic residual block. The bridge and decoder use three convolution layers and side outputs, respectively. This module generates seven probability maps for segmentation, of which the last map is considered the final output.
- The module of residual refinement is used to further refine the map of the final output through residual learning among ground truth courses and maps.
- A compact convolutional transformer model (CCTM) is proposed for a selected number of layers and hyperparameters with two convolutional transformers, two transformers, 64 projection dimensions, 64 batch sizes, two heads, 0.1 stochastic depth, 0.001 learning rates, 0.0001 weight

Abbreviations: BASNet, Boundary-aware segmentation Net; CCTM, Compact Convolutional Transformer Model; WALNN, Wolf Ant Loin Neural Network; HorUNet, High; CFF, Contextual Feature Fusion; SSD-KD, Self-supervised diverse knowledge distillation; SC, Skin Cancer; SOO, Sum of Outputs; BCE, Binary Cross-Entropy; SSIM, Structural Similarity Index; IoU, Intersection over Union; ViT Model, Vision Transformers; CCT, Compact Convolutional Transformer; XAI, Explainable Artificial Intelligence; GSO, Golden Search Optimization.

TABLE 1 Existing methods for skin-lesion detection.

References	Methods	Findings	Limitations
Khan et al. (23)	Mobile health units for data collection; multiple modal data fusion approach	Effective segmentation and classification of skin lesions with improved results	Requires mobile data collection; complex feature selection and fusion
Wu et al. (24)	HorUNet with multilevel dimensional fusion mechanism	Improved segmentation accuracy for skin lesions using ISIC2017, ISIC2018, PH datasets	High computational requirements for large datasets
Rahman et al. (25)	Features fusion method	98.3% accuracy and 95.7% AUC score on ISIC-2020 dataset	Limited exploration of concurrent multi-image comparison methods beyond melanoma
Akilandasowmya et al. (26)	Deep hidden features, ensemble classifier	Improved early detection results using ISIC-2019 and Kaggle skin-lesion data	Harmony search method complexity for feature optimization
Vidhyalakshmi and Kanchana (48)	Binary butterfly optimization algorithm (BBOA) and DCNN for skin disease categorization	Automated categorization with high prognosis accuracy	Limited applicability beyond specific skin disease types
ul haq et al. (27)	Hybrid equilibrium Aquila optimization with random forests and ensemble support vector kernels	Achieved approximately 97.4% accuracy for skin disease categorization using HAM10000 dataset	Dataset quality dependent due to requirement for enhanced image resolution
Kalpana et al. (28)	Combination of CNN and U-Net models for lesion segmentation and classification	Effective and fast recognition of skin lesions	Optimization complexity with two distinct batch size optimizers
Anand et al. (29)	SSD-KD integrating knowledge distillation (KD) for skin disease classification	85% accuracy for eight skin illnesses with condensed MobileNetV2 using ISIC-2019	Limited generalizability of KD technique for diverse datasets

TABLE 2 Overview of datasets.

Dataset name	Images	Class names	Resolution of images
PH2	200	Atypical Nevi, Common Nevi, Melanoma	768 × 560
MED-NODE	170	Benign Nevus, Melanoma	Not specified
ISIC 2016	1,279	Background, Skin Lesion	576 × 786 to 2,848 × 4,288
ISIC 2017	2,750	Nevus, Seborrheic Keratosis, Melanoma	556 × 679 to 4,499 × 6,748
ISIC 2018	10,015	Skin Cancer, Pigment Network, Globule, Milia-like Cyst, Negative Network, Streaks	Not specified
DermNet	19,500	23 Categories (e.g., Eczema, Seborrheic Keratoses, Poison Ivy, Acne, etc.)	Varies, generally low resolution

decays, and 100 epochs that provide excellent outcomes for classification.

The remainder of the paper is organized as follows. Related work is described in Section 2. The proposed methodology, with its various steps for identifying and classifying skin diseases, is discussed in Section 3. In Section 4, the quantitative and qualitative results are discussed, and Section 5 provides a brief conclusion of the proposed research work and its future perspectives.

2 Related work

Recently, researchers have introduced extensive methods to detect and classify skin diseases in their early stages using computer vision, machine learning, pattern recognition, deep CNN models, and artificial intelligence (20, 21). Rajeswar et al. (22) used the wolf antlion neural network (WALNN) to classify skin melanomas using magnetic resonance imaging data. A hybrid algorithm was introduced for feature selection. WALNN was compared with established methodologies, such as Cuckoo search-based SVM, decision tree, and CNN on the ISIC skin-lesion dataset. It improved sensitivity, specificity, recall, accuracy, and

precision, and accurately identified skin melanoma. Khan et al. (23) presented a method that uses mobile health units to collect skin data using a multimodal data fusion for skin-lesion detection. This system uses a hybrid approach for lesion segmentation by combining two CNN models. The HAM10000 dataset was used to train the CNN model for lesion classification. A summation discriminant correlation testing approach was applied to combine features from the two connected layers. A feature selection method was introduced to prevent feature redundancy. An ultimate machine-learning classifier was applied to classify the selected features with remarkable outcomes, in contrast to those of the traditional methods. Renkai et al. (24) suggested that dermoscopy is highly useful in diagnosing skin diseases, particularly skin lesions. Automatic skin-lesion segmentation is crucial for accurate diagnoses. Although U-Net models are commonly used for segmentation tasks, they have limitations in terms of spatial dependence and remote interactions. Transformers are emerging as alternatives; however, they require large amounts of data and significant computational resources. To address these issues, HorUNet, along with a multilevel dimensional fusion mechanism, was introduced. Extensive experiments were performed using private and publicly available ISIC2017, ISIC2018, and PH datasets. Ordinary convolution, which fails to exhibit spatial dependence or

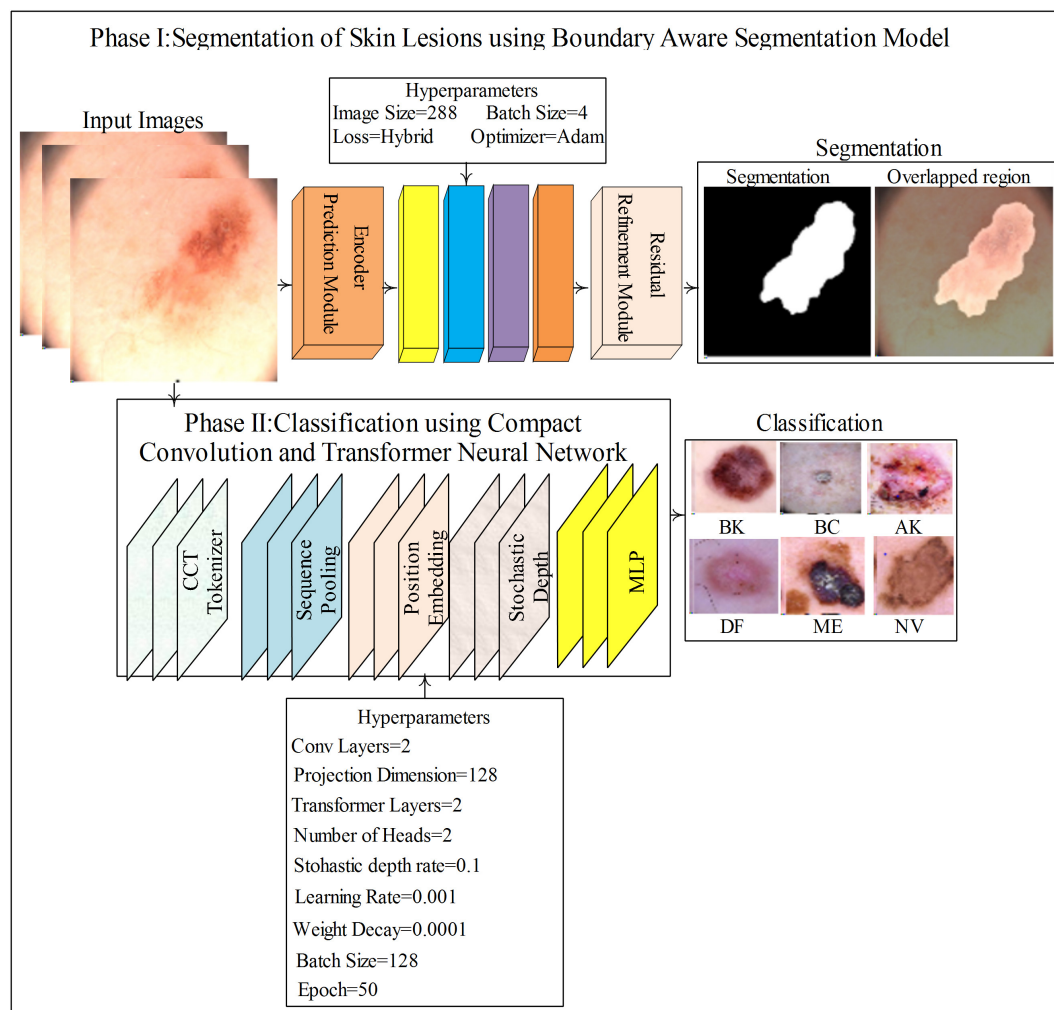


FIGURE 1
Proposed method for skin lesions detection.

remote interaction, was used in the U-Net model. Transformers are becoming a popular alternative. They require large amounts of data and huge computational resources, making them less practical for clinical medical problems. Rahman et al. (25) suggested that existing deep-learning-based schemes do not explore concurrent multi-image comparative methods. To improve the diagnosis of melanoma, a feature fusion method that integrates patient-related information was proposed. The introduced multiple-kernel self-attention segment provides an optimal overview of the extracted features, which are combined using a contextual feature fusion approach from various images into a distinct feature matrix. The introduced contextual-learning scheme significantly improved performance. Akilandasowmya et al. (26) split the deep hidden features to ensure accurate predictions and, applied a harmony search technique to optimize the features (25) and reduce data size. They also used ensemble classifiers for early disease diagnosis. Results on ISIC-2019 and Kaggle skin-lesion datasets showed considerable improvement compared with traditional methods. ul haq et al. (27) introduced a hybrid-equilibrium Aquila optimization method using random forests and ensemble support vector kernels. The HAM10000 dataset, with enhanced image resolution after

removing intrusions and noise, was used to test the model. The framework subcategorized the segmented images into five classes based on the feature properties. This approach achieved a high performance rate with an accuracy of approximately 97.4%. Kalpana et al. (28) suggested a combination model combining CNN and U-Net models to create an automated system capable of recognizing skin lesions in biomedical dermoscopic images. CNN was used to classify segmented images into multiple classes. The system was designed to handle biomedical image data, ensuring accurate and fast recognition of skin lesions to enhance the effectiveness of the DL-based approach in treating various illnesses. Two optimizers with distinct batch sizes were used to optimize the proposed scheme. Anand et al. (29) presented a cutting-edge technique called SSD-KD that integrates disparate information into a general knowledge distillation technique for skin disease classification. This methodology combined the current knowledge distillation research by developing intra-instances to represent relational features. The apprentice model captured more information than the instructor model using weighted softened outputs. The condensed MobileNetV2 classified eight distinct skin illnesses with an accuracy of up to 85%. This is the

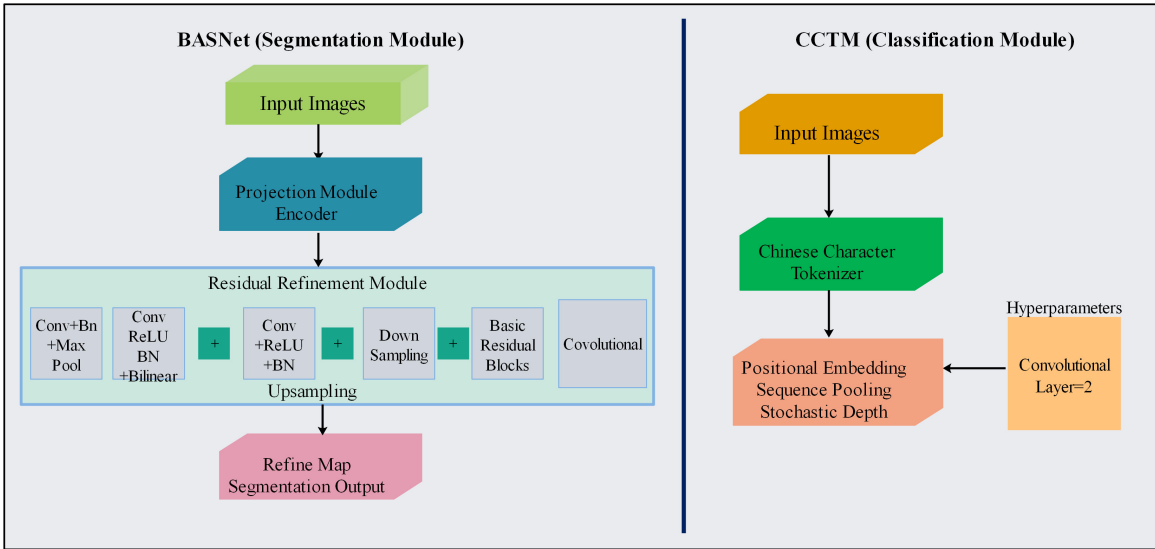


FIGURE 2
Flowchart of proposed models for segmenting and classifying skin lesions

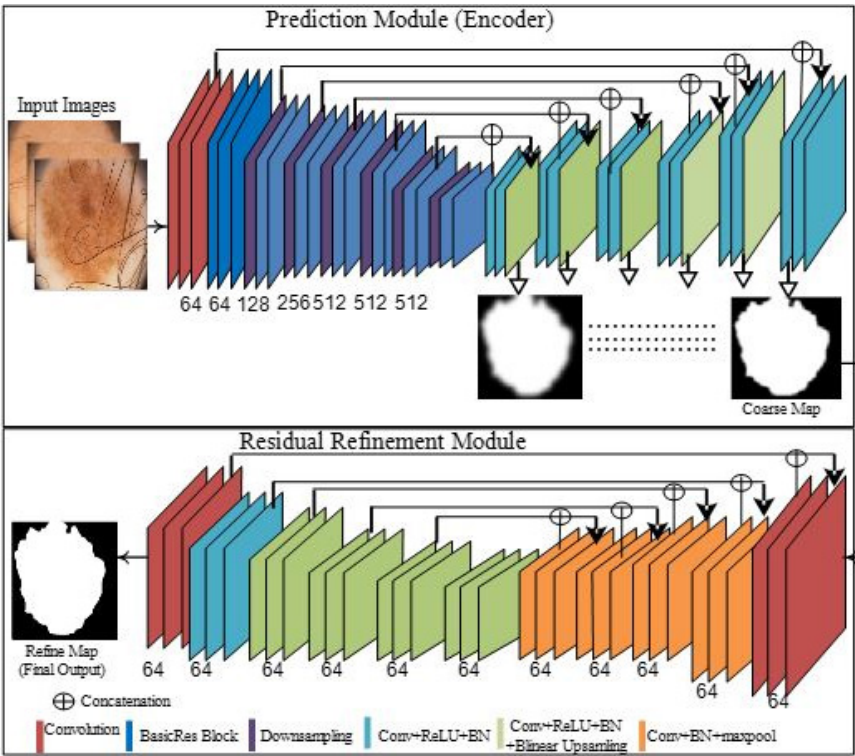


FIGURE 3
Proposed segmentation model of the skin lesions.

first application of deep knowledge distillation to a large-scale dermoscopy database for multidisease categorization. The SSD-KD method for skin disease classification, although effective, is computationally intensive and may not be suitable for all portable devices. However, its performance depends on the training data quality, which can limit its generalizability. The complexity of the dual relational knowledge distillation architecture adds to

implementation challenges. Table 1 provides a summary of existing techniques.

Table 2 summarizes the datasets used in the proposed method. To overcome the existing challenges, the proposed BASNet effectively generalizes features learned in a multiscale and dense supervision module, allowing the capture of fine-grained and global contexts regarding the details of the boundary.

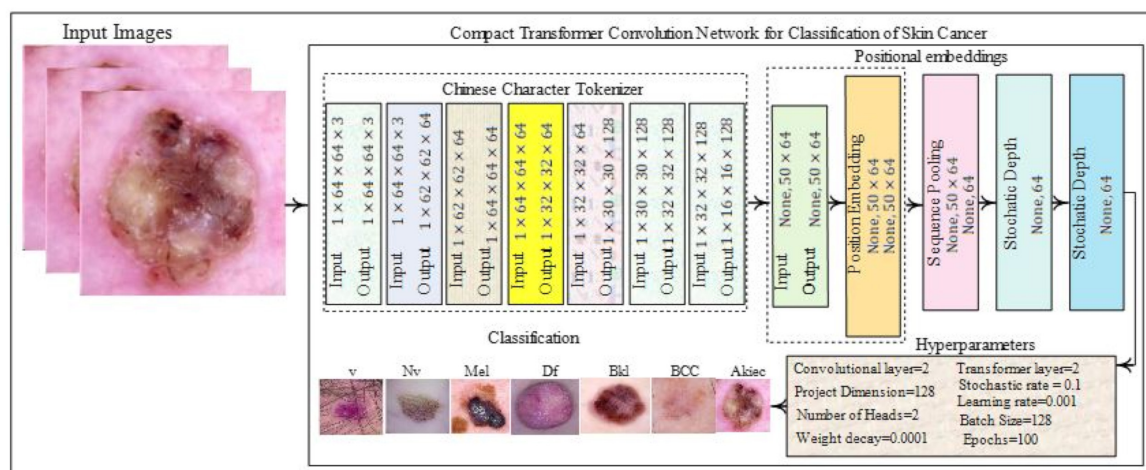


FIGURE 4
Proposed classification model.

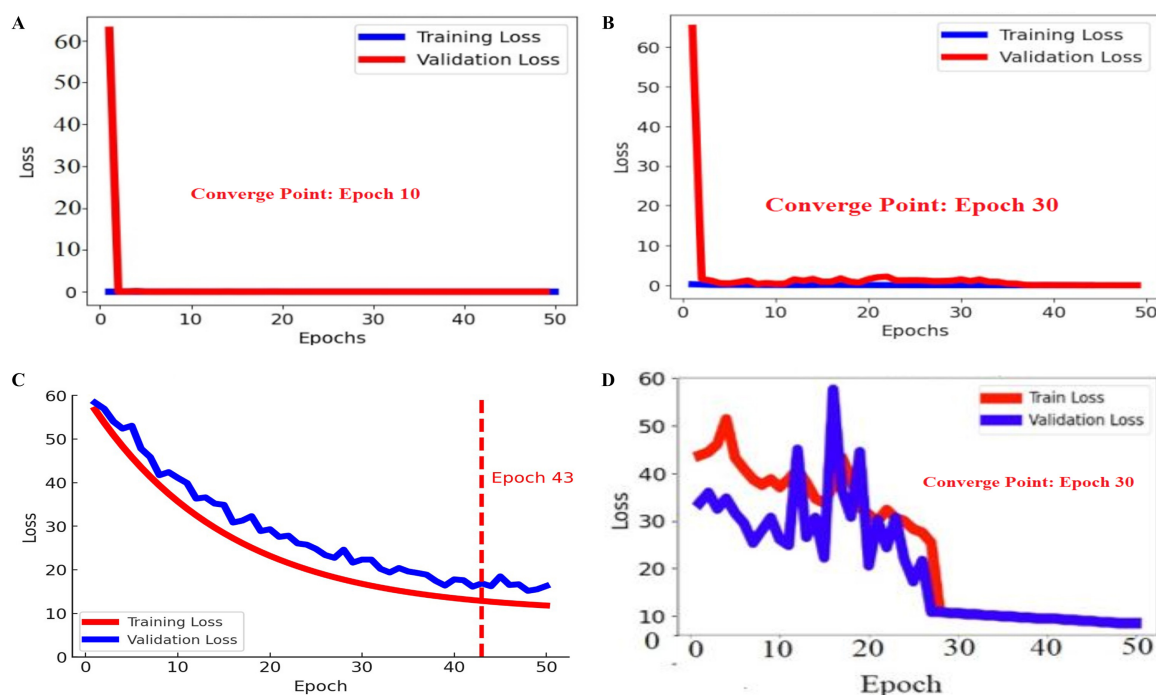


FIGURE 5
Training and validation losses of the proposed BASNet: (A) PH2, (B) ISIC 2016, (C) ISIC 2017, and (D) ISIC 2018.

A residual network was combined to extract semantic high-level features with boundary refinement to improve the accuracy of boundary predictions. Hybrid loss includes structural similarity and boundary-aware losses, which aid the model in focusing on fine boundaries and shapes to improve performance on unseen and diverse data. By combining the advantages of transformers with those of convolutional layers, the proposed CCTM performed well in terms of generalization. Convolutional layers are prone to significant inductive biases when learning local patterns, such as edges and textures. This improves the generalization of the model on small datasets or sparse data. The model can

comprehend complicated spatial linkages successfully with small amounts of data, which enhances the performance of various applications and datasets.

3 Materials and methods

In this study, two novel models are proposed for the segmentation and classification of skin lesions. BASNet is fine-tuned to segment skin lesions in poorly contrasted, illuminated, and hair dermoscopy images of the skin. The compact transformer

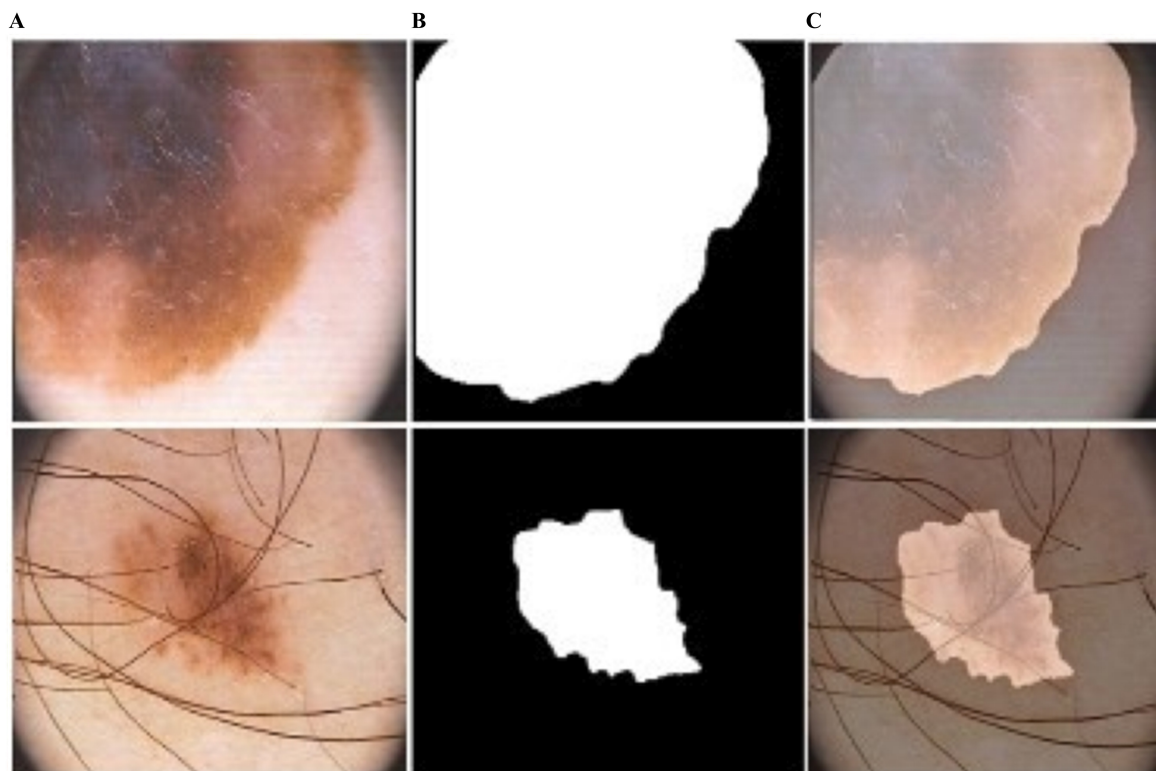


FIGURE 6

Predicted segmentation outcomes from ISIC-2017: (A) input (B) predicted masks, and (C) overlapped predicted output.

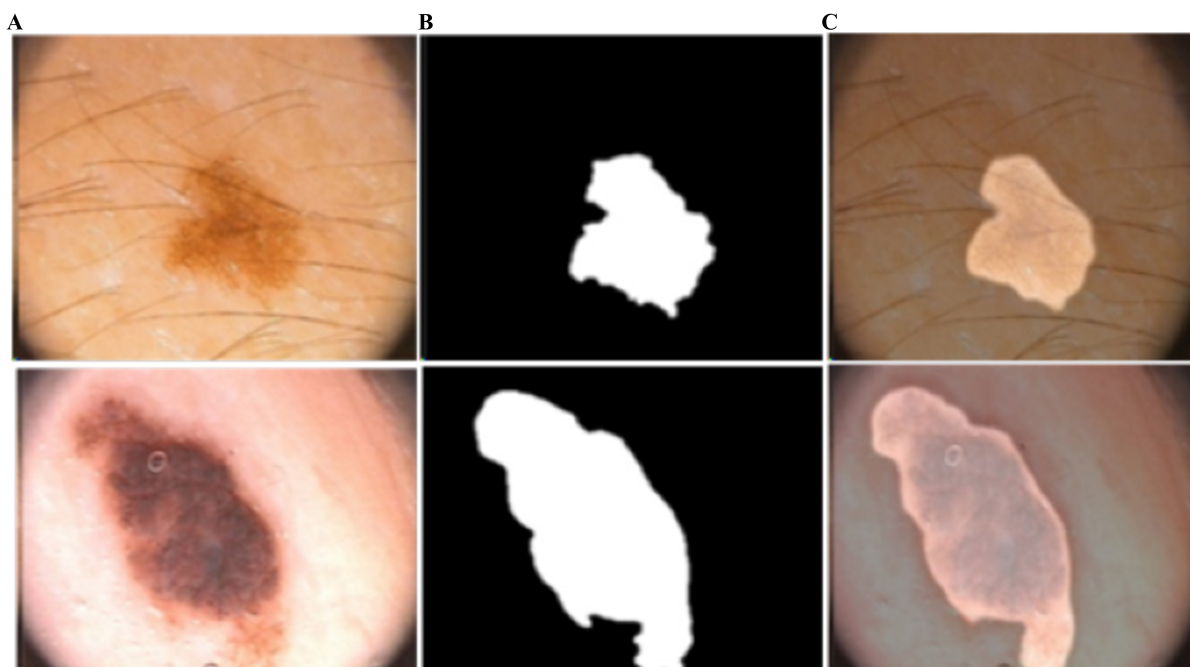


FIGURE 7

Predicted segmentation outcomes from PH2: (A) input, (B) predicted masks, and (C) overlapped predicted output.

model, which is a mixture of convolutional and transformer models, is proposed for classification. The proposed method steps are illustrated in Figure 1.

The classification model comprises compact convolutional and transformer models (Figure 1), consisting of a tokenizer, position embedding, sequence pooling, stochastic depth, and MLP to classify

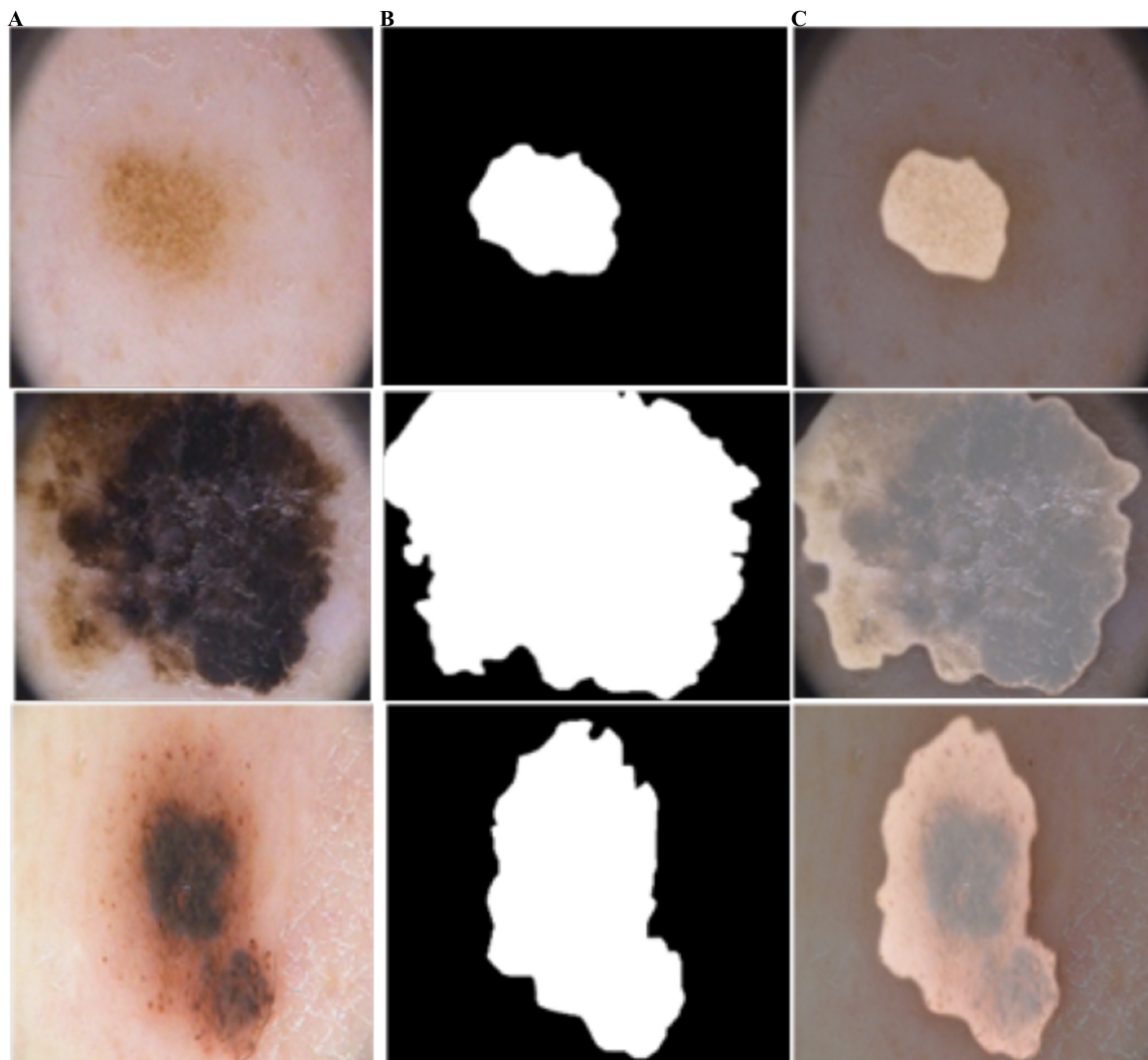


FIGURE 8

Predicted segmentation outcomes from ISIC-2018: (A) input, (B) predicted masks, and (C) overlapped predicted output.

SL. Figure 2 shows the segmentation and classification model architectures in detail.

3.1 Proposed boundary-aware segmentation model

BASNet contains prediction and residual multiscale refinement modules and a refined network for prediction and hybrid losses (30). The refined prediction model comprises a dense encoder/decoder supervised model. The prediction module consists of an encoder and decoder similar to that of U-Net. The encoder section contains an input layer of convolutional layers and six stages, where the first four are taken from the ResNet-34 and the remaining are from the primary residual blocks. The first convolution layer and pool of ResNet-34 are skipped, and four blocks are extracted. The decoder and bridge use three convolutional layers with side outputs. This module generated seven probability maps for segmentation during training, the last of

which is the output layer. The objective of the refinement module—based on a residual block—is to refine noisy and blurry boundaries on the segmentation maps produced through prediction. This model comprises four stages, each consisting of a convolutional block. Finally, the coarse and residual maps produce a more refined output map, as shown in Figure 3.

Hybrid loss helps the network learn in a hierarchy, such as the patch, pixel, and map levels, as defined in Equation 1 (31).

$$\text{Sum of output (SOO)} = \sum_{n=1}^N \alpha_n^{(n)} \quad (1)$$

where SOO is the sum of the outputs, N is the total output, n is a single instance of outputs, $l^{(n)}$ is a loss of n outputs, and α_n is the weight of each loss. This model has eight outputs: seven related to the prediction model and one for refinement.

Clearer and higher-quality segmentation boundaries were obtained using hybrid loss, expressed as Equation 2.

$$l^{(n)} = l^{(n)}_{\text{bce}} + l^{(n)}_{\text{ssim}} + l^{(n)}_{\text{iou}} \quad (2)$$

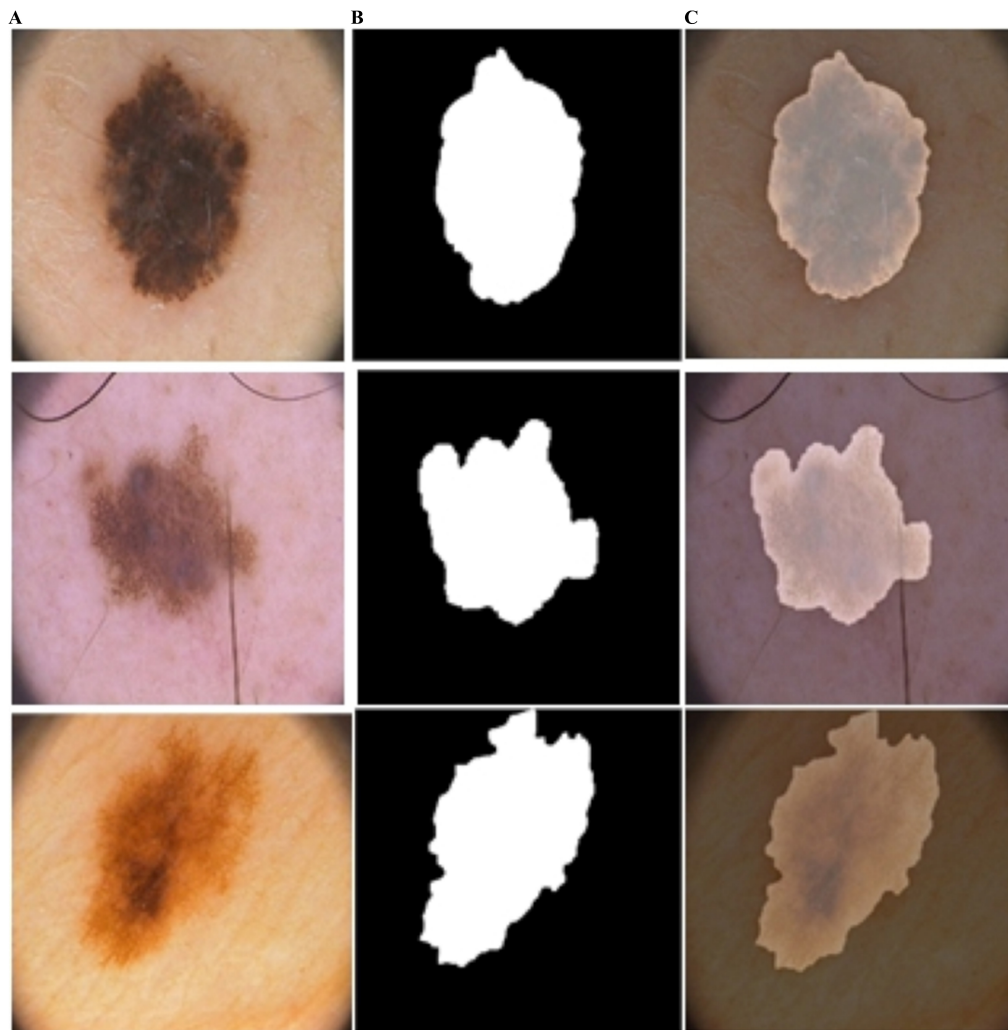


FIGURE 9

Predicted segmentation results from ISIC-2016: (A) input, (B) predicted masks, and (C) overlapped predicted output.

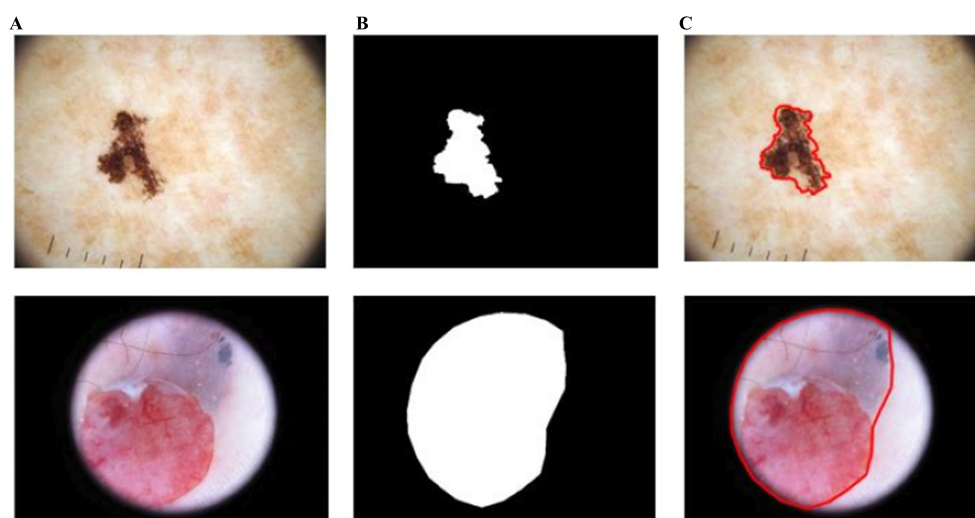


FIGURE 10

Segmentation results of BASNet: (A) input, (B) segmentation, and (C) skin-lesion boundaries.

TABLE 3 Segmentation outcomes of the proposed BASNet model.

Segmentation datasets	Testing	Validation	Training	Average (%)		Confidence interval
				IoU	Dice score	
PH2	✓			0.96	0.98	[0.9383–0.9817]
		✓		0.96	0.98	[0.9383–0.9817]
			✓	0.95	0.98	[0.9283–0.9717]
ISIC 2016	✓			0.89	0.96	[0.8683–0.9117]
		✓		0.87	0.95	[0.8483–0.8917]
			✓	0.88	0.97	[0.8583–0.9017]
ISIC 2018	✓			0.94	0.97	[0.9183–0.9617]
		✓		0.93	0.96	[0.9083–0.9517]
			✓	0.96	0.98	[0.9383–0.9817]
ISIC 2017	✓			0.97	0.98	[0.9483–0.9917]
		✓		0.95	0.97	[0.9283–0.9717]
			✓	0.94	0.96	[0.9183–0.9617]

TABLE 4 Comparison of training time on HAM-10000 dataset.

Model	Total epochs	Time of each epoch
BASNet model	50	4 min
Yaseliiani et al. (49)		17.46 s

where $l_{bce}^{(n)}$, $l_{ssim}^{(n)}$, $l_{iou}^{(n)}$ are the BCE, SSIM, and IoU losses, respectively. Mathematically, the BCE is defined as Equation 3.

$$l_{bce} = -\sum_{(r,c)} [M(r,c)\log(P(r,c)) + (1-M(r,c))\log(1-P(r,c))] \quad (3)$$

where $M(r, c) \in \{0, 1\}$ is the pixel (r, c) of the annotated mask, and $P(r, c)$ is the probability of the predicted pixel. SSIM loss captures the structural information of the image. Let $x = x_j; j = 1, \dots, N^2$ and $y = y_j; j = 1, \dots, N^2$ represent the pixel values of the patch size in $N = N$ region cropped from p probability map and m mask, respectively. The SSIM of x and y are expressed as Equation 4.

$$l_{ssim} = 1 - \frac{(2m_x m_y + Cov_1)(2\sigma_{xy} + Cov_2)}{(m_x^2 + m_y^2 + Cov_1)(\sigma_x^2 + \sigma_y^2 + Cov_2)} \quad (4)$$

where m_x , m_y , σ_x , and σ_y represent the mean and standard deviation of x and y , respectively. σ_{xy} denotes covariance. The values of the Cov_1 and Cov_2 (0.01^2 and 0.03^2) were used to prevent division by 0.

The intersection over union (IoU) matrix was used to compute the similarity using Equation 5.

$$l_{iou} = 1 - \frac{\sum_{r=1}^H \sum_{c=1}^W P(r, c) M(r, c)}{\sum_{r=1}^H \sum_{c=1}^W [P(r, c) + M(r, c) - P(r, c) M(r, c)]} \quad (5)$$

where $M(r, c) \in \{0, 1\}$ is pixel (r, c) and $p(r, c)$ is predicted pixel probability.

The BCE loss was used to compute pixel-level values. The neighborhood of the labels was not considered, and foreground and background pixels were equally weighted, which helps with convergence points on pixels and guarantees the best local optima. The SSIM loss was used for patch-level information around the

TABLE 5 Results of proposed BASNet model compared to the existing approaches.

References	Year	Datasets	Results (%)
Giotis et al. (38)	2024	PH2	0.96
Hong et al. (42)	2023		0.96
Srikanteswara and Ramachandra (41)	2023		0.82
Nampalle et al. (40)	2022		0.92
Mustafa et al. (50)	2025		0.93
Proposed model			0.98
Giotis et al. (38)	2024	ISIC-2018	0.96
Nampalle et al. (40)	2022		0.89
Venugopal et al. (51)	2023		0.97
Proposed model			0.97
Hong et al. (42)	2023	ISIC-2017	0.94
Srikanteswara and Ramachandra (41)	2023		0.89
Rajendran and Shanmugam (52)	2024		0.97
Proposed model			0.98

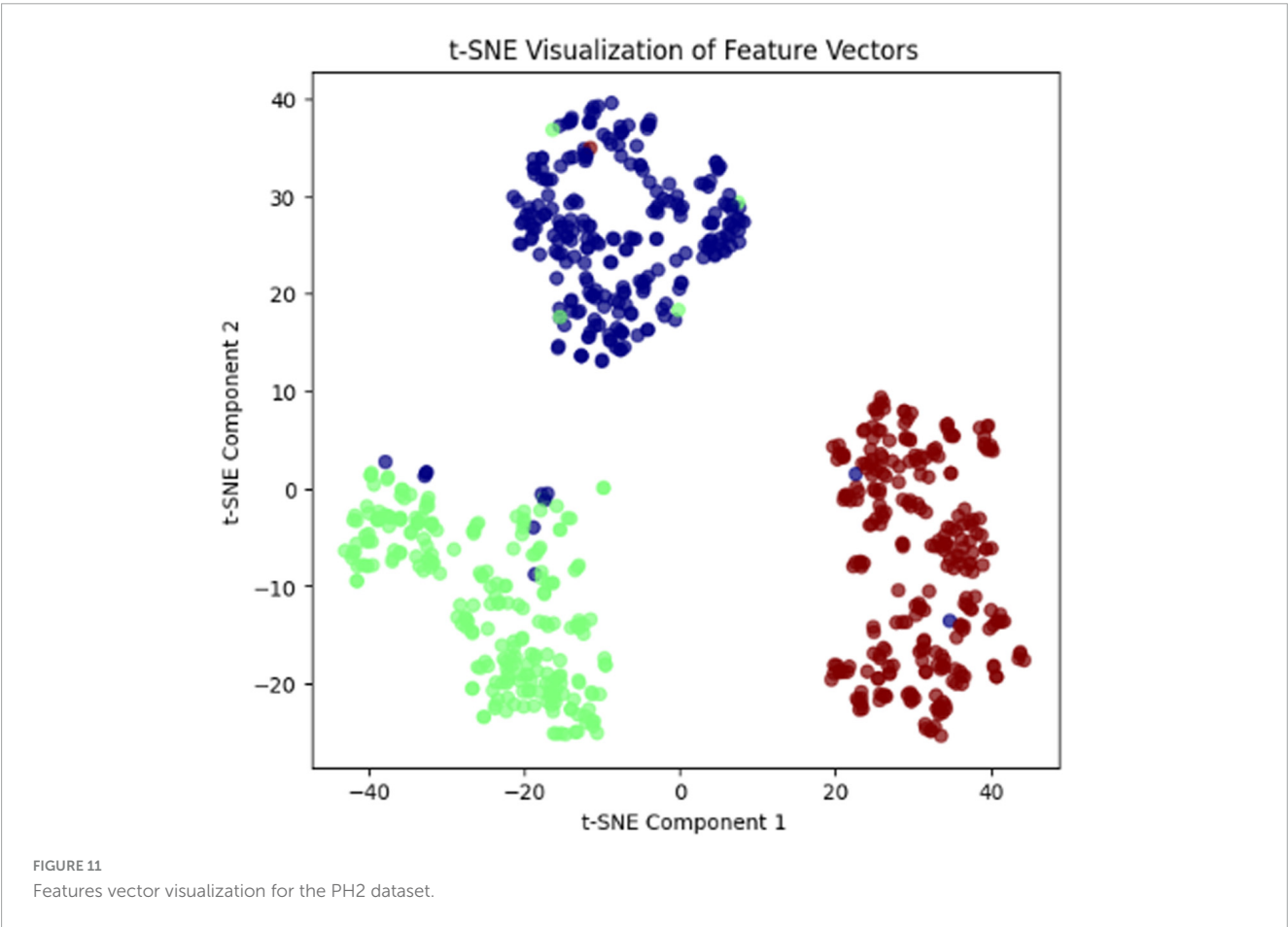
pixels of the local neighborhood. Higher weight values were assigned to the pixels located in the region of the transitional buffer between the foreground and background pixels that are similar or higher than the region of the foreground. The background loss was not applied in training until the background pixels were closer to the mask, where the loss dropped rapidly to 1–0.

The values of m_x , σ_{xy} , $m_x m_y$, and σ_y^2 in $SSIM_{loss}$ in Equation 4 were 0 in the region of the background. Therefore, the SSIM value was computed using Equation 6.

$$l_{ssim}^{background} = 1 - \frac{Cov_1 Cov_2}{(m_x^2 + Cov_1)(\sigma_x^2 + Cov_2)} \quad (6)$$

TABLE 6 Proposed classification results on the HAM-10000 dataset.

Akiec	Bcc	Bkl	Df	Mel	Nv	Vl	Precision (P)	Recall (R)	F1-score (F)
✓							0.96	0.96	0.96
	✓						0.96	0.97	0.96
		✓					0.95	0.95	0.95
			✓				0.98	1.00	0.99
				✓			0.96	0.93	0.95
					✓		0.95	0.96	0.96
						✓	1.00	1.00	1.00
✓	✓	✓	✓	✓	✓	✓	0.97 Accuracy (A)		
✓	✓	✓	✓	✓	✓	✓	0.97	0.97	0.97
✓	✓	✓	✓	✓	✓	✓	0.97	0.97	0.97



where $Cov_1 = 0.01^2$ and $Cov_2 = 0.03^2$, when the prediction of x was close to zero.

IoU was used to compute map-level information. If a large region is included in the IoU, the model trained on the IoU loss emphasizes a large region of the foreground and generates homogeneous and whiter probabilities in these areas. However, this model produced a false-negative region in finer structures. To obtain the advantages of the three losses—BCE, IoU, and SSIM—the hybrid loss was formulated. BCE was used to maintain the gradient among all pixel values, and IoU focused on the foreground pixels.

3.2 Classification of the skin lesions based on the proposed mixture of convolutional and transformer mode

The proposed CCTM model for skin-lesion classification is shown in Figure 4. The model uses a tokenizer to process input images. The vision-transformer (ViT) model was applied to organize all images into a uniform non-overlapping patch, removing information related to the boundaries between distinct patches, which is vital for effectively exploiting the

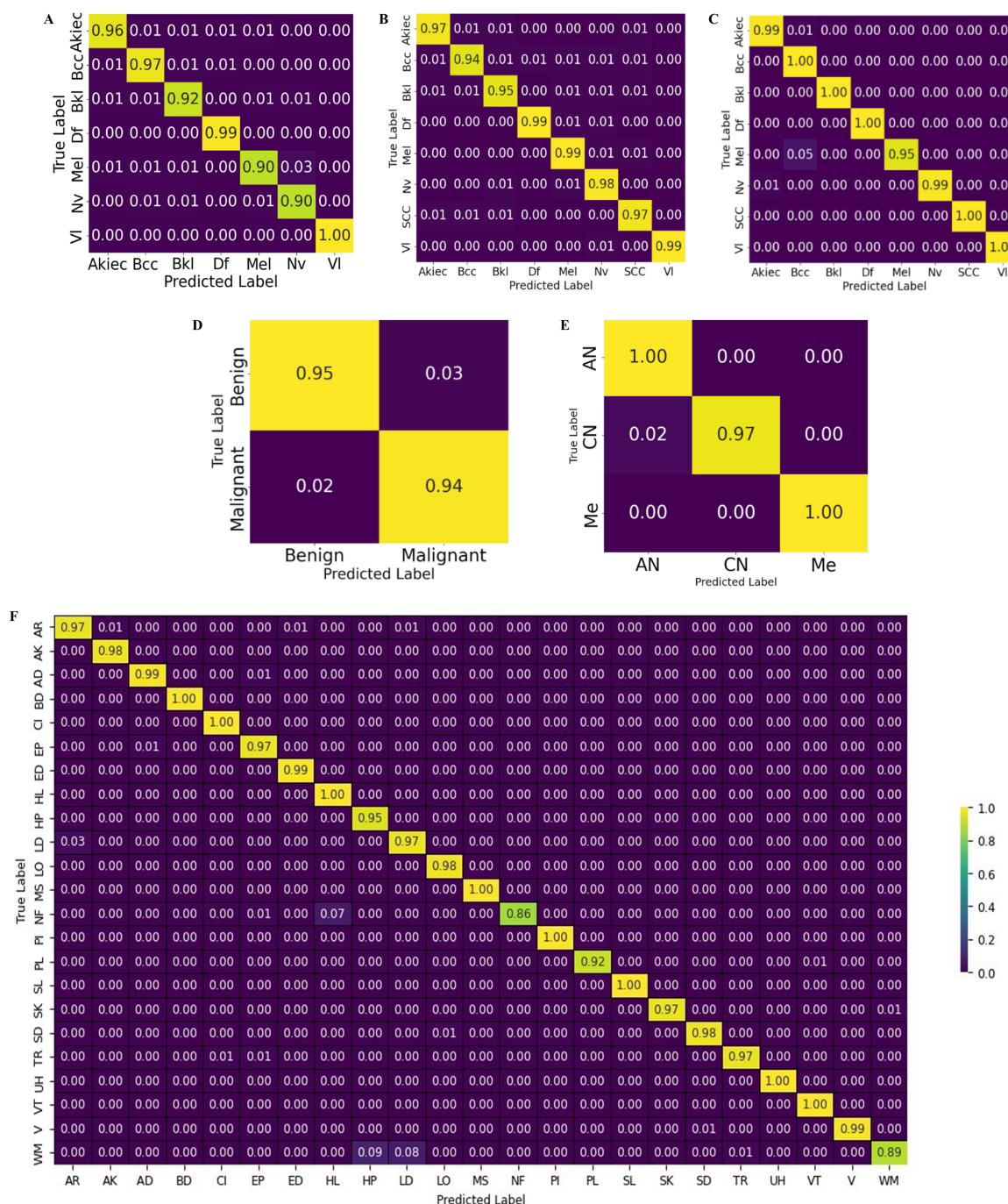
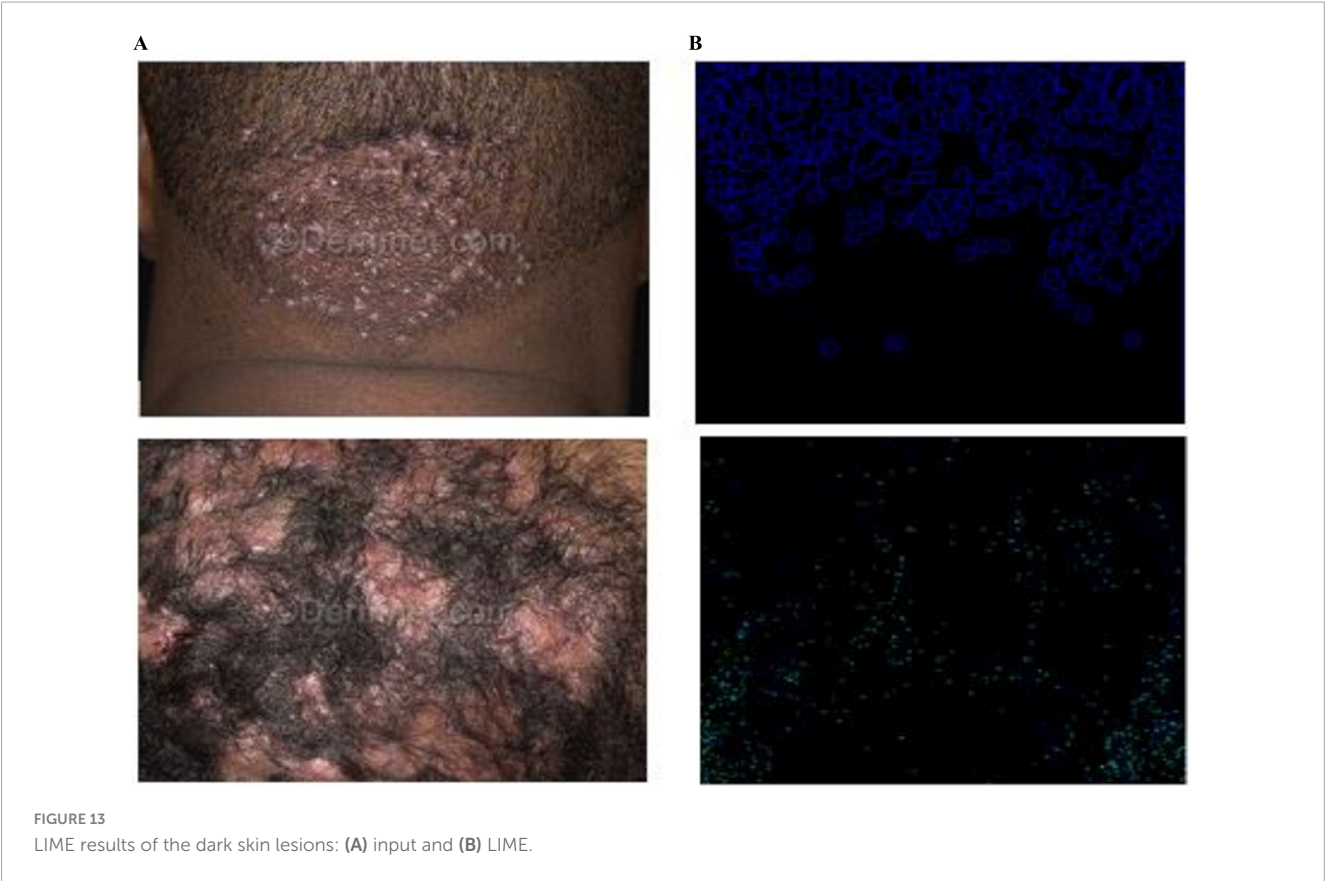


FIGURE 12

Confusion chart of the proposed CCTM for skin-lesion classification: (A) HAM-10000, (B) ISIC-2019, (C) ISIC-2020, (D) Med-node (E) PH2, and (F) DermNet.

local information. Convolution works well in extracting local information. The input images were processed using a tokenizer. The ViT model divided images into non-overlapping uniform patches to reduce the information at the boundaries between different types of patches. This is vital for neural networks to effectively exploit local information. The convolution kernel effectively exploits the local information; thus, the convolution in the mini-model generates patches. Positional embedding in CCT is optional, and sequence or attention pooling was added. In the

ViT model, the mapping of features related to the token of the class is pooled and used subsequently for skin-lesion classification. Stochastic depth is a regularization method of randomly dropping a set of layers. This is similar to dropout but operates on a block of layers as compared to separate nodes that are included in the layer. Stochastic depth is used before the residual blocks of the transformer section of the encoder. Finally, the encoder section of the transformer is weighted and fed to the final specific layer for skin-lesion classification.



4 Experimental results and discussion

To evaluate the effectiveness of the segmentation method, four publicly available datasets were used: PH2 (32), ISIC 2016 (32), 2017 (33), and 2018 (34). Five publicly available datasets were used to evaluate the performance of the classification method: PH2 (32), HAM10000 (35), ISIC 2019 (36), MED-NODE (37), and DermNet (accessed on 9 November 2022).¹ The entire datasets were divided into 0.4 hold-out validation, in which 60% of the data was used for training, 20% for validation, and 20% for testing. The proposed method was implemented on a Desktop-T20JD6R, 12th Generation, Core i7-12700K, processor 3.60 GHz, RAM 32.0 GB, graphics card NVIDIA RTX A4000 with a Windows-11 operating system.

4.1 Segmentation and classification datasets

In this section, the datasets used for segmentation and classification are discussed. The details of the segmentation datasets are as follows.

PH2: The PH2 dataset contained 200 images of melanocytic lesions categorized as atypical nevi, common nevi, and melanomas. Each image was verified by expert dermatologists

TABLE 7 Elapsed time on benchmark datasets.

Datasets	Execution time
HAM-10000	28 min 4 s
MED-NODE	15 min 2 s
ISIC-2019	23 min 1 s
ISIC-2020	21 min 0 s
PH2	16 min 0 s
DermNet	29 min 3 s

TABLE 8 Execution time comparison of the proposed CCTM model with the existing method.

Ref	Dataset	Running/execution time
Araújo et al. (43)	HAM-10000	1 h 30 min
Proposed Model	HAM-10000	28 min 4 s

who manually segmented skin lesions for clinical analysis. This database is valuable for evaluating and validating computer-based segmentation and classification algorithms for melanoma diagnosis.

ISIC 2016: This dermoscopy dataset is categorized into two classes named “background” and “skin lesion.” The anatomical region “full body” skin is considered.

ISIC 2017: The size of the training and validation data was 2000/150 whereas the test data size was approximately 600.

¹ <https://www.kaggle.com/datasets/shubhamgoel27/dermnet>

TABLE 9 CCTM classification outcomes using ISIC-2019 dataset.

Akiec	Bcc	Bkl	Df	Mel	Nv	Scs	Vl	P	R	F
✓								0.97	0.97	0.97
	✓							0.96	0.94	0.95
		✓						0.95	0.95	0.95
			✓					0.99	0.99	0.99
				✓				0.97	0.98	0.97
					✓			0.96	0.98	0.97
						✓		0.95	0.97	0.96
							✓	1.00	0.99	0.99
✓	✓	✓	✓	✓	✓		✓	0.97 A		
✓	✓	✓	✓	✓	✓		✓	0.97	0.97	0.97
✓	✓	✓	✓	✓	✓		✓	0.97	0.97	0.97

TABLE 10 CCTM classification outcomes using ISIC-2020 dataset.

Akiec	Bcc	Bkl	Df	Mel	Nv	Scs	Vl	P	R	F1
✓								0.99	0.99	0.99
	✓							0.95	1.00	0.97
		✓						1.00	1.00	1.00
			✓					1.00	1.00	1.00
				✓				1.00	0.95	0.97
					✓			1.00	0.99	1.00
						✓		1.00	1.00	1.00
							✓	1.00	1.00	1.00
✓	✓	✓	✓	✓	✓		✓	0.99 A		
✓	✓	✓	✓	✓	✓		✓	0.99	0.99	0.99
✓	✓	✓	✓	✓	✓		✓	0.99	0.99	0.99

ISIC 2018: A large-scale dataset published by ISIC containing 10,015 dermoscopic images with skin lesions annotated with seven classes of skin diseases such as skin cancer, pigment network, globule, and others: milia-like cysts, negative networks, and streaks. These classes were used to detect various skin diseases. The dataset was used for instances and semantic segmentation.

The details of classification datasets are as follows:

HAM10000: The HAM10000 dataset comprised a considerable group of multisource dermoscopic images containing colored skin diseases. Diverse populations and acquisition methods were used to collect 10,015 images. The dataset included seven generic classes of pigmented lesions, selected for simplicity and practical clinical relevance. The dataset was meticulously cleaned and standardized to ensure high quality using manual screening to exclude specific attributes and ensure appropriate color reproduction.

ISIC 2019: consisted of 19,424 dermoscopic images captured over 16 years using high-resolution cameras. These images comprised approximately 11 diagnostic groups. The dataset may become more disabled in medical practice by relating each captured image to the age and sex of the patient and the position of the lesion.

TABLE 11 CCTM classification outcomes using PH2 dataset.

Benign	Malignant	P	R	F
✓		0.97	0.96	0.97
	✓	0.96	0.97	0.97
✓	✓	0.97 A		
✓	✓	0.97	0.97	0.97
✓	✓	0.97	0.97	0.97

MED-NODE: This is a non-dermoscopic skin-lesion dataset consisting of two classes: benign nevi with 100 images and melanoma lesions with 70 images.

DermNet: Consisted of 19,500 images with three RGB channels and 23 distinct categories of skin diseases, such as eczema, borrelic keratoses, poison ivy, acne, vascular tumors, tinea ringworm, psoriasis, melanoma, and bullous disease. However, the images were mostly of low resolution.

The datasets had limitations such as data imbalance, low resolution, and poor contrast. The data were augmented through horizontal and vertical flipping to increase the number of images. Hybrid loss and residual refinements were applied to BASNet to handle occlusion and poor contrast images.

TABLE 12 CCTM classification outcomes using MED-NODE dataset.

AN	CN	Me	P	R	F1
✓			0.98	1.00	0.99
	✓		1.00	0.97	0.98
		✓	1.00	1.00	1.00
✓	✓	✓	0.99 A		
✓	✓	✓	0.99	0.99	0.99
✓	✓	✓	0.99	0.99	0.99

TABLE 13 CCTM classification outcomes using DermNet dataset.

Classes	P	R	F1
AR	0.96	0.97	0.97
AK	0.99	0.98	0.99
AD	0.99	0.99	0.99
BD	0.99	1.00	1.00
CI	0.99	1.00	0.99
EP	0.97	0.97	0.97
ED	0.97	0.99	0.98
HL	0.93	1.00	0.96
HP	0.88	0.95	0.91
LD	0.98	0.97	0.98
LO	0.99	0.98	0.99
MS	0.99	1.00	1.00
NF	0.91	0.86	0.88
PI	1.00	1.00	1.00
PL	0.95	0.92	0.93
SL	1.00	1.00	1.00
SK	0.99	0.97	0.98
SD	0.98	0.98	0.98
TR	0.99	0.97	0.98
UH	1.00	1.00	1.00
VT	0.99	1.00	0.99
V	0.99	0.99	0.99
WM	0.93	0.89	0.91
Total Classes = 23	Accuracy = 0.97		
	0.97	0.97	0.97
	0.97	0.97	0.97

4.2 Experiment 1: segmentation of the skin lesions using the proposed BASNet model

In this experiment, the BASNet Model was trained on the optimal hyperparameters that are shown in Figure 5.

In Figure 4, the blue and red lines represent the training and validation losses, respectively. On the PH2 dataset, the training/validation loss was stable at the initial epochs, whereas on

TABLE 14 Classification results of CCTM compared with those of existing methods.

References	Year	Datasets	Accuracy (%)
Rokhsati et al. (44)	2024	MED-NODE	0.96
Georgiadis et al. (53)	2025		0.66
Rasel et al. (54)	2024		0.75 F1
Proposed model			0.99
Rokhsati et al. (44)	2024	PH2	0.96
Reis et al. (39)	2024		0.88
Mustafa et al. (50)	2025		0.94
Georgiadis et al. (53)	2025		0.50
Proposed model			0.97
Singh et al. (45)	2024	ISIC-2020	0.73
Georgiadis et al. (53)	2025		
Proposed model			0.99
Reis et al. (39)	2024	ISIC-2019	0.96
Mustafa et al. (50)	2025		0.93
Georgiadis et al. (53)	2025		0.48
Proposed model			0.97
Singh et al. (45)	2024	HAM-10000	0.96
Mustafa et al. (50)	2025		0.92
Georgiadis et al. (53)	2025		0.73
Proposed Model			0.97
Medhat et al. (46)	2024	DermNet	91.92 ± 1.74
Mui-zzud-din et al. (47)	2024		0.91
Hanum et al. (55)	2025		0.94
Proposed model			0.97

the ISIC [2016, 2017, and 2018] datasets, the training/validation curves were stable after 30 epochs. The segmentation model results are shown in Figures 6–9.

As shown in Figure 10, BASNet segments the skin-lesion boundaries more accurately, even with illumination and lighting effects. The computed segmentation results are given in Table 3.

Table 3 shows the evaluation of the segmentation model using four dermoscopy datasets, PH2, ISIC 2016, ISIC 2018, and ISIC 2017. During training, the average IoU and Dice scores were 0.95 and 0.98 on PH2; 0.88 and 0.97 on ISIC 2016; 0.96 and 0.98 on ISIC 2018, and 0.94 and 0.96 on ISIC 2017 datasets, respectively. Similarly, in the validation stage, the results in terms of IoU and Dice scores were 0.96 and 0.98 on PH2; 0.87 and 0.95 on ISIC 2016; 0.93 and 0.96 on ISIC 2018, and 0.95 and 0.97 on ISIC 2017 datasets, respectively. Finally, in the testing stage, the results were 0.95 and 0.98 on PH2; 0.88 and 0.97 on ISIC 2016; 0.96 and 0.98 on ISIC 2018, and 0.94 and 0.96 on ISIC 2017 datasets, respectively.

The confidence intervals demonstrate high consistency and reliability across datasets. Most intervals fall within narrow ranges, indicating stable performance. For instance, several scores show

TABLE 15 Segmentation results on the variant of ablation using ISIC-2018 dataset.

Variant of ablation	Testing	Validation	Training	Average (%)	
				IoU	Dice score
Without using residual refinement	✓			0.78	0.79
		✓		0.75	0.76
			✓	0.75	0.78
Without using hybrid loss (structured similarity and boundary aware) Only binary cross-entropy	✓			0.79	0.76
		✓		0.77	0.75
			✓	0.78	0.77

TABLE 16 Experiment for the selection of hyperparameters.

Batch size	Optimizer	Loss	Learning rate	Error rate
2	Sgdm	Tversky	1e ⁻⁵	0.093
4	Adam	Hybrid	1e⁻⁴	0.049
8	RMSprops	Focal	1e ⁻³	0.100

The bold numbers represent the values selected from the experiment for further analysis and experimentation.

a confidence interval of [0.9383, 0.9817], reflecting a robust and precise estimation. Similarly, intervals such as [0.9283, 0.9717] and [0.9183, 0.9617] suggest minimal variability, while slightly wider intervals like [0.8683, 0.9117] and [0.8483, 0.8917] indicate reduced but still consistent performance. Notably, the highest confidence interval [0.9483, 0.9917] showcases the peak reliability within the dataset. Overall, the reported intervals confirm a high level of accuracy and confidence. BASNet was trained for 50 epochs; the training time of each epoch is presented in Table 4.

Table 4 shows the training times of BASNet and the existing method on the HAM-10000 dataset. The input image size is 288 × 288. This model takes four minutes on each epoch. However, the existing method is trained for 50 epochs with 240 × 240 input image size, taking 17.46 sec on each epoch. The difference in time is due to the image size. The outcomes are shown in Table 5.

Table 5 presents the results of the proposed model and those of the existing approaches, such as (38–41). The U-Net model was designed using skip paths to the encoder to reduce the semantic gap between concatenated maps of features for skin-lesion segmentation. The method was evaluated using PH2 and ISIC-18 datasets providing accuracies of 96.18 and 96.09%, respectively (38). The adaptive contour was applied for segmentation on the ISIC-17 and PH2 datasets, with accuracies of 0.94% and 0.96%, respectively (42). The LinkNet and U-Net models were combined to transfer skin-lesion segmentation learning on ISIC-18 and PH2 datasets with 0.89% and 0.92% accuracy (40). A two-stage method was designed based on a modified CNN classifier to segment the skin lesions. This method was evaluated on the PH2 and ISIC-17 datasets with accuracies of 0.82% and 0.89%, respectively.

4.3 Experiment 2: classification of the skin lesions using the proposed compact convolutional transformer model

The proposed CCTM was trained for 100 epochs, and the results are shown in Table 6. The features vectors visualization for PH2 dataset is shown in Figure 11.

In Figure 11, t-SNE visualization of feature vectors illustrates how well the model has learned to distinguish between different classes. Each dot represents a sample in the dataset, and the color coding corresponds to different class labels. The clear separation of clusters indicates that the model has successfully extracted meaningful features, with samples of the same class grouping together while maintaining distinct boundaries between different classes. Some overlap might suggest areas where the model could improve, possibly due to similarities between certain categories. This visualization helps understand the model's representation learning and provides insights into feature space organization.

The test results are shown in terms of the confusion matrix in Figure 12.

Table 6 presents the classification results on the HAM-10000 dataset. Precision, recall, and F1-score were computed for the individual classes, as well as the overall accuracy, micro average, and weighted average. The values of precision, recall and F1-score were 0.96, 0.96, and 0.96 for Akiec; 0.96, 0.97, and 0.96 for BCC; 0.95, 0.95, and 0.95 for Bkl; 0.98, 1.00, and 0.99 for Df; 0.96, 0.93, and 0.95 for Mel; 0.95, 0.96, and 0.96 for Nv; and 1.00, 1.00, and 1.00 for V1. The overall attained accuracy of the seven classes was 0.97, the macro average of the precision, recall, and F1 scores was 0.97, and the weighted average was 0.97.

The proposed CCTM classified all skin lesions, including light and dark lesions. Furthermore, to authenticate the model performance, explainable AI (XAI) was applied using LIME to highlight the important features of the model as shown in Figure 13.

The study addresses AI interpretability by applying explainable AI (XAI) techniques, specifically LIME, to highlight important features contributing to the model's decisions. Figure 13 visually demonstrates how the proposed CCTM model classifies both light and dark skin lesions with transparency. This ensures clinical trust by making the model's decision-making process more interpretable and justifiable.

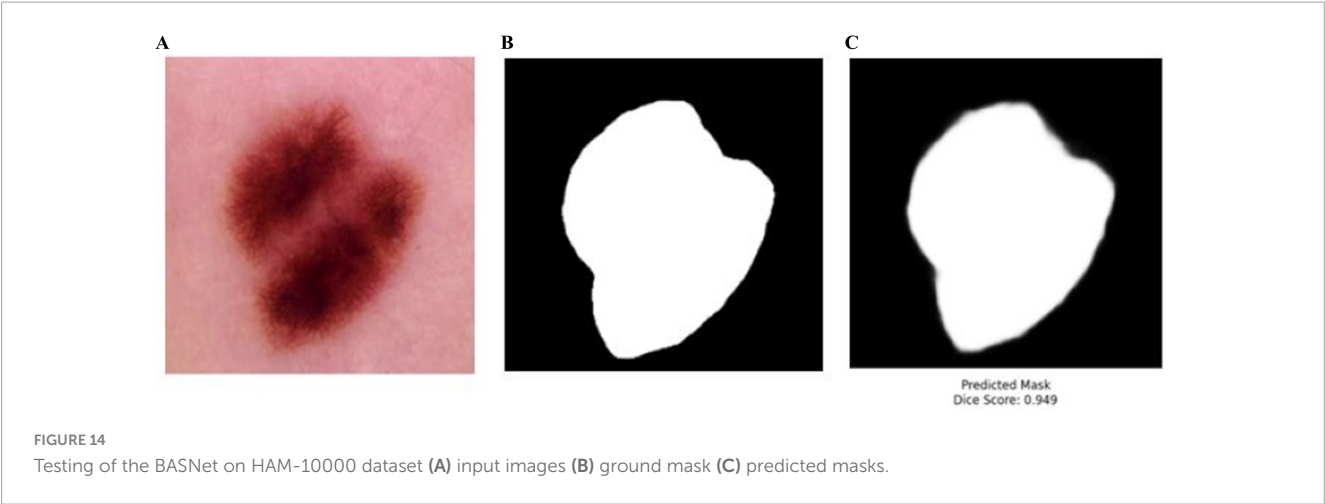


TABLE 17 Comparison of training time.

Model	Time of each epoch
BASNet model	4 min
Öztürk and Özkaya (56)	432.3 sec

The time elapsed was computed for both training and testing on the benchmark datasets, as shown in Table 7.

The running/execution time of the proposed model was also compared with that of the existing method (43), as listed in Table 8.

Table 8 shows the elapsed time for the same benchmark dataset, where the execution time of the existing method is 1 h 30 min, whereas that of the proposed model is 28 min 4 s. Thus, the proposed model is computationally efficient. The classification outcomes are given in Table 9.

On ISIC-2019, the values of P, R, and F1 on the classes of Akiec were 0.97, 0.97, 0.97, whereas those of Bcc were 0.96, 0.94, 0.95, Bkl were 0.95, 0.95, 0.95, Df were 0.99, 0.99, 0.99, Mel were 0.97, 0.98, 0.97, Nv were 0.96, 0.98, 0.97, Scc were 0.95, 0.97, 0.96, and of VI were 1.00, 0.99, 0.99 respectively. The outcomes achieved for the ISIC-2020 dataset are listed in Table 10.

Tables 10, 11 present the values of precision, recall, and F1-score on eight classes of ISIC-2020 dataset.

Table 11 presents results for benign and malignant classes. In the benign class, the results for P, R, and F1 were 0.97, 0.96, and 0.97, respectively, whereas those for the malignant class were 0.96, 0.97, and 0.97, respectively. Similarly, the accuracy was 0.97, and macro average and weighted rates were 0.97, 0.97, and 0.97 for precision, recall, and F1-scores, respectively. The classification results on MED-NODE dataset are listed in Table 12.

The classification results on the MED-NODE dataset were computed in three classes: AN, CN, and Me. The results of P, R, and F1 were 0.98, 1.00, and 0.99 for AN; 1.00, 0.97, and 0.98 for CN; and 1.00, 1.00, and 1.00 for the Me class, respectively. The accuracy was 0.99 and the macro-average and macro-weighted results were 0.99, 0.99, and 0.99, respectively. The results for the DermNet dataset are listed in Table 13.

Table 13 provides results on twenty-three skin-lesion classes. On the classes the results of AR were 0.96, 0.97, 0.97, on AK were 0.99, 0.98, 0.99, whereas 0.99, 0.99, 0.99 on AD, 0.99, 1.00, 1.00 on

BD, 0.99, 1.00, 0.99 on CI, 0.97, 0.97, 0.97 on EP, 0.97, 0.99, 0.98 on ED, 0.93, 1.00, 0.96 on HL, 0.88, 0.95, 0.91 on HP, 0.98, 0.97, 0.98 on LD, 0.99, 0.98, 0.99 on LO, 0.99, 1.00, 1.00 on Ms, 0.91, 0.86, 0.88 on Nf, 1.00, 1.00, 1.00 on PI, 0.95, 0.92, 0.93 on PL, 1.00, 1.00, 1.00 on SL, 0.99, 0.97, 0.98 on SK, 0.98, 0.98, 0.98 on SD, 0.99, 0.97, 0.98 on TR, 1.00, 1.00, 1.00 on UH, 0.99, 1.00, 0.99 on VT, 0.99, 0.99, 0.99 on V and 0.93, 0.89, 0.91 on WM. The accuracy of all the classes is 0.97.

The CCTM outcomes validated using the ISIC-2019, ISIC-2020, PH2, and DermNet datasets are shown in Tables 5–9. The proposed model has an accuracy of 0.97 on ISIC-2019, 0.99 on ISIC-2020, 0.97 on PH2, 0.99 on MED-NODE, and 0.97 on DermNet datasets.

The CCTM model provides the highest accuracy of 0.99 on ISIC-2020 and MED-NODE compared with the other datasets. The CCTM results were compared with those of the existing methods, as presented in Table 14.

Iterative magnitude pruning was used with AlexNet for skin-lesion classification on PH2 and MED-NODE datasets with an accuracy of 0.96 (44). The DT uses Bayesian learning and fuzzy ID3 values for skin-lesion classification. The results on PH2 and ISIC-19 datasets were 88% and 96%, respectively (39). A stacked CNN model was used for skin-lesion classification. This method was evaluated on ISIC-20 and HAM-10000 with accuracies of 0.73% and 0.96%, respectively (45). The features were optimized using GSO and the skin lesion was classified based on a random forest classifier (46). The ensemble model was created using a combination of pretrained models, such as ResNet50V2, ResNet152V2, and ResNet101V2, which were used for feature extraction to classify skin lesions (47).

In Table 14, on the MED-NODE dataset, where the proposed model achieved 99% accuracy, the misclassification rate is 1%. Similarly, for the PH2 dataset, achieving 97% accuracy, the misclassification rate is 3%. On the ISIC-2020 dataset, the proposed model outperformed previous methods with 99% accuracy, leading to a 1% misclassification rate. Likewise, for ISIC-2019, the model achieved 97% accuracy, corresponding to a 3% misclassification rate. On the HAM-10000 dataset, the proposed model attained 97% accuracy, resulting in a 3% misclassification rate. Finally, for the DermNet dataset, where the model achieved 97% accuracy, the misclassification rate remains 3%. These results indicate that the

TABLE 18 Training hyperparameters of CCTM.

Conv layers	Project dimension	Number of heads	Transformer layers	Stochastic rate	Learning rate	Batch size	Error rate
2	128	2	2	0.10	0.001	128	0.021
4	64	4	4	0.01	0.01	64	0.054

The bold numbers represent the values selected from the experiment for further analysis and experimentation.

TABLE 19 Results of classification on variant of ablation using HAM-10000 dataset.

Variant of ablation	Akiec	Bcc	Bkl	Df	Mel	Nv	Vl	P	R	F1
Without convolution layers K	✓							0.82	0.82	0.80
		✓						0.82	0.81	0.81
			✓					0.80	0.79	0.85
				✓				0.81	0.80	0.78
					✓			0.82	0.83	0.74
						✓		0.78	0.79	0.74
							✓	0.80	0.81	0.79
✓	✓	✓	✓	✓	✓	✓	✓	0.80 Accuracy (A)		
✓	✓	✓	✓	✓	✓	✓	✓	0.80	0.80	0.80
✓	✓	✓	✓	✓	✓	✓	✓	0.80	0.80	0.80
Kernel size of convolution = 4	✓							0.90	0.91	0.90
		✓						0.91	0.90	0.92
			✓					0.90	0.89	0.90
				✓				0.91	0.92	0.91
					✓			0.92	0.93	0.90
						✓		0.92	0.92	0.92
							✓	0.91	0.91	0.91
✓	✓	✓	✓	✓	✓	✓	✓	0.90 A		
✓	✓	✓	✓	✓	✓	✓	✓	0.91	0.91	0.91
✓	✓	✓	✓	✓	✓	✓	✓	0.91	0.91	0.91
Without patch embedding	✓							0.81	0.80	0.89
		✓						0.80	0.81	0.82
			✓					0.80	0.81	0.82
				✓				0.82	0.82	0.82
					✓			0.81	0.80	0.80
						✓		0.80	0.81	0.80
							✓	0.81	0.81	0.81
✓	✓	✓	✓	✓	✓	✓	✓	0.82 A		
✓	✓	✓	✓	✓	✓	✓	✓	0.82	0.82	0.82
✓	✓	✓	✓	✓	✓	✓	✓	0.82	0.82	0.82

proposed model significantly reduces errors compared to previous studies while maintaining robust classification performance.

4.4 Ablation study

An ablation study was performed using both segmentation and classification models.

BASNet used ResNet-34 as a backbone because it gives a balance between performance and efficiency. It is a lightweight, less computationally expensive, and faster model. The residual connections and hierarchical features extraction abilities of this model help capture the contextual and fine-detail information. The pretrained weights of ResNet-34 enable faster training and better generalization. Compared with ResNet-50 and ResNet-101, deeper

models exist, with higher computation and memory requirements and long training time.

Dermoscopy images of skin cancer often suffer from poor quality due to factors like lighting variations, hair occlusion, and low contrast, which can obscure critical features needed for accurate diagnosis. The hybrid loss and residual refinement alleviate the challenges of occlusion and poor contrast. Hybrid loss combines pixel-wise accuracy (e.g., cross-entropy) with structural sensitivity (e.g., dice loss), ensuring the model captures both fine details like lesion borders and broader patterns like texture irregularities. Residual refinement further improves predictions by iteratively correcting errors from earlier outputs, focusing on subtle but diagnostically significant features. This combination makes the system more robust and accurate, enabling it to handle the variability and imperfections common in dermoscopy images, ultimately supporting better skin cancer detection in real-world clinical applications. The hyperparameters of BASNet were finalized after the experimentation as shown in Table 15.

Table 15 presents different types of losses and batch sizes. Learning rates and loss function were used for model training. In this experiment, an error rate of 0.049 was obtained on the combination of hyperparameters such as Hybrid loss, Adam optimizer, and 10^{-4} , which is less than those of others.

In BASNet model, weight (0.8) for pixel-wise (L2) loss and perceptual loss (0.2) were applied. This helped the model prioritize high-level preservation of features. The high weight on pixel-level loss, helps reconstruct the image and preserve the overall structure, particularly when fine details are less distinguishable due to occlusion or low contrast. This combination strikes a balance between fidelity and original ability to retain the vital features.

The segmentation BASNet model was authenticated by performing different experiments on the ISIC-2018 dataset, as listed in Table 16.

Table 16 presents the ablation variant without using the residual refinement module. The IoU was 0.78 and Dice score was 0.79 in the testing stage, the IoU was 0.75 and Dice score was 0.76 in the validation stage, and the IoU was 0.75 and Dice score was 0.78 in the training stage. Similarly, without using hybrid loss in the testing stage, IoU and Dice scores were 0.79 and 0.76, respectively. In the validation stage, IoU and Dice scores were 0.77 and 0.75, respectively. In the training stage, IoU and Dice scores were 0.78 and 0.77, respectively. It was observed that the residual refinement module and hybrid loss played vital roles in the segmentation of skin lesions. The results of segmentation drastically decrease when used without these parameters.

The proposed BASNet model was trained and tested on four benchmark datasets: PH2, ISIC 2016, ISIC 2017 and ISIC 2018. To authenticate the performance of BASNet, some images of the HAM-10000 dataset were passed to the trained weights on ISIC 2018 datasets; the segmentation results of the predicted masks with DSC scores are shown in Figure 14.

In Figure 14, BASNet was not trained on the BASNet model. The trained weights of BASNet on ISIC-2018 dataset were used to test some images of the HAM-10000 dataset. BASNet achieved 0.949 DSC, which shows the generalizability and reliability of the model. The training time of BASNet was compared with that of the existing method (Table 17).

The training time of each epoch of BASNet is 4 minutes while that of the existing method, the iFCN model, is 432.3 seconds (Table 17). Before training CCTM, hyperparameters were selected after experimentation as shown in Table 18.

The hyperparameters which provide less error rates compared to others are highlighted in bold and italics, in Table 18. The results of the classification model were evaluated using a variant of ablation on the HAM-10000 dataset as shown in Table 19.

The results in Table 19 were computed based on different parameters, such as without convolution layers, by varying convolution kernel size, and without a patch embedding layer. Without convolutional layers, the accuracy was 0.80, whereas with the four kernel sizes of the convolutional layers, an accuracy of 0.90 was achieved. Similarly, without the patch embedding layer, an accuracy of 0.82 was achieved. The results can be drastically changed by reducing or changing the number of parameters.

5 Conclusion

Several studies have been conducted on the detection of skin lesions; however, accurate segmentation and classification of skin lesions remain great challenges. To overcome these challenges, we proposed a method, which is based on two novel models. To address the challenges of skin-lesion segmentation, a boundary-aware segmentation model was proposed based on hybrid loss and selected hyperparameters for more accurate skin-lesion segmentation. The model was assessed using four challenging dermoscopic datasets: PH2, ISIC-2016, ISIC-2017, and ISIC-2018. The average IoU and Dice scores were 0.96 and 0.98 for PH2; 0.89 and 0.96 for ISIC 2016; 0.94 and 0.97 for ISIC 2018; and 0.97 and 0.98 for ISIC 2017 datasets, respectively.

Skin-lesion classification remains a challenge owing to the similar shape, color, and size of skin lesions. Therefore, a CCTM was proposed and trained on optimal hyperparameters, achieving accurate skin-lesion classification at the testing stage. CCTM was evaluated on the ISIC Challenge and DermNet datasets with different types of skin lesions. The accuracy obtained was 0.99 on MED-NODE, 0.97 on PH2, 0.97 on ISIC-2019; 0.99 on ISIC-2020; 0.97 on HAM-10000, and 0.97 on DermNet datasets, respectively.

6. Limitations and future scope

BASNet is appropriate for poorly contrasted, illuminated, and hair dermoscopic images, it is computationally intensive. It focuses on both global and fine-grained details by employing a deep-learning model that undergoes several rounds of feature extraction, refinement, and fusion. This leads to significant processing and memory requirements, particularly when handling high-resolution dermoscopic images. Furthermore, the results demonstrated the superiority of CCTM. This is a great contribution to this domain; in the future, this model will be implemented in hospitals to evaluate its performance on real dermoscopic images. However, there remain obstacles to its incorporation into clinical workflows, including the need for strong regulatory approvals to guarantee safety, huge computing resources required

for real-time inference, and the need for clinician training to properly understand AI results. For smooth adoption and for AI to support human knowledge in managing skin cancer rather than replace it, these obstacles must be overcome and cooperation between AI developers and healthcare practitioners must be encouraged.

In the future, a method using quantum machine/DL may be proposed to achieve accurate and efficient outcomes. The proposed method may also be validated on ISIC Challenge-2024, which was not used in this study.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. The code associated with this article can be found at: <https://www.javeriaamin.site/2025/02/skin-lesion-segmentation-using-boundary.html>. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

JA: Conceptualization, Methodology, Software, Writing – original draft. MA: Formal Analysis, Investigation, Writing – review & editing. HA: Formal Analysis, Investigation, Writing – review & editing. AZ: Conceptualization, Formal Analysis, Methodology, Writing – review & editing. S-HK: Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing.

References

1. Tajjour S, Garg S, Chandel S, Sharma D. A novel hybrid artificial neural network technique for the early skin cancer diagnosis using color space conversions of original images. *Int J Imaging Syst Technol.* (2023). 33:276–86. doi: 10.1002/ima.22784
2. Asif S, Qurrat-ul-Ain S, Khan S, Amjad K, Awais M. SKINC-NET: An efficient lightweight deep learning model for multiclass skin lesion classification in dermoscopic images. *Multimedia Tool Appl.* (2024):1–27. doi: 10.1007/s11042-024-19489-x
3. Almeida M, Santos I. Classification models for skin tumor detection using texture analysis in medical images. *J Imaging.* (2020) 6:51. doi: 10.3390/jimaging6060051
4. Ki V, Rotstein C. Bacterial skin and soft tissue infections in adults: A review of their epidemiology, pathogenesis, diagnosis, treatment and site of care. *Can J Infect Dis Med Microbiol.* (2008) 19:173–84. doi: 10.1155/2008/846453
5. Lin Q, Guo X, Feng B, Guo J, Ni S, Dong H. A novel multi-task learning network for skin lesion classification based on multi-modal clues and label-level fusion. *Comput Biol Med.* (2024) 175:108549. doi: 10.1016/j.cmpbiomed.2024.108549
6. Srinivasu P, SivaSai J, Ijaz M, Bhoi A, Kim W, Kang J. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors (Basel).* (2021) 21:2852. doi: 10.3390/s21082852
7. Allugunti VR. A machine learning model for skin disease classification using convolution neural network. *Int J Comput Programm Database Manage.* (2022) 3:141–7. doi: 10.33545/27076636.2022.v3.i1b.53
8. Karthik R, Vaichole T, Kulkarni S, Yadav O, Khan F. Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification. *Biomed Signal Process Control.* (2022) 73:103406. doi: 10.1016/j.bspc.2021.103406
9. Hermann K, Chen T, Kornblith S. The origins and prevalence of texture bias in convolutional neural networks. *Adv Neural Inf Process Syst.* (2020) 33:19000–15.
10. Nawaz M, Nazir T, Masood M, Ali F, Khan M, Tariq U, et al. Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network. *Int J Imaging Syst Technol.* (2022) 32(6):2137–53. doi: 10.1002/ima.22750

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by the Ministry of Science and ICT (MSIT), Korea, under the ICT Challenge and Advanced Network of HRD (ICAN) program (IITP-2024-RS-2022-00156345) supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP). This work was partially supported by the National Research Foundation of Korea (NRF) (grant numbers RS-2023-00219051 and RS-2023-00209107) and the Unmanned Vehicles Core Technology Research and Development Program through the NRF and Unmanned Vehicle Advanced Research Center (UVARC), funded by the Ministry of Science and ICT, Republic of Korea (NRF-2023M3C1C1A01098408).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

11. Castillo D, Lakshminarayanan V, Rodríguez-Álvarez MJ. MR images, brain lesions, and deep learning. *Appl Sci.* (2021) 11:1675. doi: 10.3390/app11041675
12. SivaSai J, Srinivasu P, Sindhuri M, Rohitha K, Deepika S. An automated segmentation of brain MR image through fuzzy recurrent neural network. In: Bhoi AK, Mallick PK, Liu CM, Balas VE editors. *Bio-Inspired Neurocomputing*. Singapore: Springer (2021). p. 163–79. doi: 10.1007/978-981-15-5495-7_9
13. Civit-Masot J, Luna-Perejón F, Domínguez Morales M, Civit A. Deep learning system for COVID-19 diagnosis aid using X-ray pulmonary images. *Appl Sci.* (2020) 10:4640. doi: 10.3390/app10134640
14. Yamanakkanavar N, Choi J, Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: A survey. *Sensors (Basel).* (2020) 20:3243. doi: 10.3390/s20113243
15. Sato R, Iwamoto Y, Cho K, Kang D, Chen Y. Accurate BAPL score classification of brain PET images based on convolutional neural networks with a joint discriminative loss function. *Appl Sci.* (2020) 10:965. doi: 10.3390/app10030965
16. Avanzato R, Beritelli F. Automatic ECG diagnosis using convolutional neural network. *Electronics.* (2020) 9:951. doi: 10.3390/electronics9060951
17. Sridhar S, Manian V. Eeg and deep learning based brain cognitive function classification. *Computers.* (2020) 9:104. doi: 10.3390/computers9040104
18. Chen J, Bi S, Zhang G, Cao G. High-density surface EMG-based gesture recognition using a 3D convolutional neural network. *Sensors (Basel).* (2020) 20:1201. doi: 10.3390/s20041201
19. Da C, Zhang H, Sang Y. Brain CT image classification with deep neural networks. In: Handa H, Ishibuchi H, Ong YS, Tan KC editors. *Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems, 1*. Cham: Springer International Publishing (2015). p. 653–62. doi: 10.1007/978-3-319-13359-1_50
20. Hussain SM, Prasanthi B, Kandula N, Uppalapati J, Dasika S. Enhancing Melanoma skin cancer detection with machine learning and image processing techniques. In *International Conference on Advanced Computing, Machine Learning, Robotics and Internet Technologies*. Berlin: Springer (2023).
21. Aljohani K, Turki T. Automatic classification of melanoma skin cancer with deep convolutional neural networks. *AI.* (2022) 3:512–25.
22. Rajeswari R, Kalaiselvi K, Jayashri N, Lakshmi P, Muthusamy A. Meta-heuristic based melanoma skin disease detection and classification using wolf ant lion neural network (WALNN) model. *Int J Intell Syst Appl Eng.* (2023) 12:87–95.
23. Khan M, Muhammad K, Sharif M, Akram T, Kadry S. Intelligent fusion-assisted skin lesion localization and classification for smart healthcare. *Neural Comput Appl.* (2024) 36:37–52. doi: 10.1007/s00521-021-06490-w
24. Wu R, Liang P, Huang X, Shi L, Gu Y, Zhu H, et al. MHorUNet: High-order spatial interaction UNet for skin lesion segmentation. *Biomed Signal Process Control.* (2024) 88:105517. doi: 10.1016/j.bspc.2023.105517
25. Rahman M, Paul B, Mahmud T, Fattah SA. CIFF-Net: Contextual image feature fusion for Melanoma diagnosis. *Biomed Signal Process Control.* (2024) 88:105673. doi: 10.1016/j.bspc.2023.105673
26. Akilandasowmya G, Nirmaladevi G, Suganthi S, Aishwariya A. Skin cancer diagnosis: Leveraging deep hidden features and ensemble classifiers for early detection and classification. *Biomed Signal Process Control.* (2024) 88:105306. doi: 10.1016/j.bspc.2023.105306
27. ul haq I, Amin J, Sharif M, Anjum MA. "Skin lesion detection using recent machine learning approaches", in: Saba T, Rehman A, Roypp S editors. *Prognostic Models in Healthcare: AI and Statistical Approaches*. Singapore: Springer Nature Singapore (2022). p. 193–211.
28. Kalpana B, Reshmy A, Senthil Pandi S, Dhanasekaran S. OESV-KRF Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases. *Biomed Signal Process Control.* (2023) 85:104779. doi: 10.1016/j.bspc.2023.104779
29. Anand V, Gupta S, Koundal D, Singh K. Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. *Expert Syst Appl.* (2023) 213:119230. doi: 10.1016/j.eswa.2022.119230
30. Wang Y, Wang Y, Cai J, Lee T, Miao C, Wang ZJ. SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images. *Med Image Anal.* (2023) 84:102693. doi: 10.1016/j.media.2022.102693
31. Qin X, Fan D, Huang C, Diagne C, Zhang Z, Sant'Anna A, et al. Boundary-aware segmentation network for mobile and web applications. *arXiv [Preprint]* (2021):doi: 10.48550/arXiv.2101.04704
32. Mendonça T, Ferreira P, Marques J, Marcal A, Rozeira J. 2-A dermoscopic image database for research and benchmarking. *Annu Int Conf IEEE Eng Med Biol Soc.* (2013) 2013:5437–40.
33. Gutman D, Codella N, Celebi E, Helba B, Marchetti M, Mishra N, et al. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *arXiv [Preprint]* (2016):doi: 10.48550/arXiv.1605.01397
34. Codella N, Gutman D, Celebi M, Helba B, Marchetti M, Dusza S, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (ISIC). *Proceedings of the 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE (2018). p. 168–72. doi: 10.1109/ISBI.2018.8363547
35. Codella N, Rotemberg V, Tschandl P, Celebi M, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv [Preprint]* (2019):doi: 10.48550/arXiv.1902.03368
36. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Sci Data.* (2018) 5:180161. doi: 10.1038/sdata.2018.161
37. Combalia M, Codella N, Rotemberg V, Helba B, Vilaplana V, Reiter O, et al. Bcn20000: Dermoscopic lesions in the wild. *arXiv [Preprint]* (2019):doi: 10.48550/arXiv.1908.02288
38. Giotis I, Molders N, Land S, Biehl M, Jonkman M, Petkov N. MED-NODE A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Syst Appl.* (2015) 42:6578–85. doi: 10.1016/j.eswa.2015.04.034
39. Reis H, Turk V, Khoshelham K, Kaya S. InSiNet: A deep convolutional approach to skin cancer detection and segmentation. *Med Biol Eng Comput.* (2022) 60:643–62. doi: 10.1007/s11517-021-02473-0
40. Nampalle K, Pundhir A, Jupudi P, Raman B. Towards improved U-Net for efficient skin lesion segmentation. *Multimedia Tool Appl.* (2024) 83:71665–82.
41. Srikanteswara R, Ramachandra A. Analyzing color features to realize adaptive contour model for segmentation. *Indian J Sci Technol.* (2023) 16:4378–87. doi: 10.17485/IJST/v16i46.1215
42. Hong Y, Zhang G, Wei B, Cong J, Xu Y, Zhang K. Weakly supervised semantic segmentation for skin cancer via CNN superpixel region response. *Multimedia Tool Appl.* (2023) 82:6829–47. doi: 10.1007/s11042-022-13606-4
43. Araújo R, Araújo F, Silva R. Automatic segmentation of melanoma skin cancer using transfer learning and fine-tuning. *Multimedia Syst.* (2022) 28:1239–50. doi: 10.1007/s00530-021-00840-3
44. Rokhsati H, Rezaee K, Abbasi A, Belhaoui S, Shafi J, Liu Y, et al. An efficient computer-aided diagnosis model for classifying melanoma cancer using fuzzy-ID3-pvalue decision tree algorithm. *Multimedia Tool Appl.* (2024) 83:76731–51.
45. Singh J, Sandhu J, Kumar Y. An analysis of detection and diagnosis of different classes of skin diseases using artificial intelligence-based learning approaches with hyper parameters. *Arch Comp Methods Eng.* (2024) 31:1051–78. doi: 10.1007/s11831-023-10005-2
46. Medhat S, Abdel-Galil H, Aboutabl A, Saleh H. Iterative magnitude pruning-based light-version of AlexNet for skin cancer classification. *Neural Comput Appl.* (2024) 36:1413–28. doi: 10.1007/s00521-023-09111-w
47. Mui-zzud-din K, Ahmed K, Rustam F, Mehmood A, Ashraf I, Choi G. Predicting skin cancer melanoma using stacked convolutional neural networks model. *Multimedia Tool Appl.* (2024) 83:9503–22. doi: 10.1007/s11042-023-15488-6
48. Vidhyalakshmi A, Kanchana M. Skin cancer classification using improved transfer learning model-based random forest classifier and golden search optimization. *Int J Imaging Syst Technol.* (2024) 34:e22971. doi: 10.1002/ima.22971
49. Yaseliani M, Ijadi Maghsoodi A, Hassannayebi E, Aickelin U. Diagnostic clinical decision support based on deep learning and knowledge-based systems for psoriasis: From diagnosis to treatment options. *Comput Ind Eng.* (2024) 187:109754. doi: 10.1016/j.cie.2023.109754
50. Mustafa S, Jaffar A, Rashid M, Bhatti SM. Deep learning-based skin lesion analysis using hybrid ResUNet++ and modified AlexNet-Random Forest for enhanced segmentation and classification. *PLoS One.* (2025) 20:e0315120. doi: 10.1371/journal.pone.0315120
51. Venugopal V, Raj N, Nath M, Stephen N. A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. *Decision Anal. J.* (2023) 8:100278.
52. Rajendran V, Shanmugam S. Automated skin cancer detection and classification using cat swarm optimization with a deep learning model. *Eng Technol Appl Sci Res.* (2024) 14:12734–9.
53. Georgiadis P, Gkouvikos E, Vrochidou E, Kalampokas T, Papakostas G. Building better deep learning models through dataset fusion: A case study in skin cancer classification with hyperdatasets. *Diagnostics.* (2025) 15:352.
54. Rasel MA, Kareem A, Obaidallah U. Integrating color histogram analysis and convolutional neural networks for skin lesion classification. *Comput. Biol. Med.* (2024) 183:109250.
55. Hanum S, Dey A, Kabir M. An attention-guided deep learning approach for classifying 39 skin lesion Types. *arXiv [Preprint]* (2025). doi: 10.48550/arXiv.2501.05991.
56. Öztürk B, Özkeya U. Skin lesion segmentation with improved convolutional neural network. *J Digit Imaging.* (2020) 33:958–70. doi: 10.1007/s10278-020-00343-z