International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 176-188
https://doi.org/10.70917/ijcisim-2026-5012
Article

Towards Clinically Interpretable Skin Lesion Analysis: A Unified Deep Learning Framework for Classification and Localization

Sameer Tembhurney[1], Rucha Rajiv Shastrakar[2], Anjali Gondane[3], Pratiksha Yogesh Ramteke[4], Bhushan Gedam[5], Sneha Patle[6], Priyanka Mangalkar[7], Ashvini D. Tarar[8]

[1,2,3,4,5,6,7] Department of Computer Science and Engineering, Ramdeobaba University, Nagpur, Maharashtra, India.
[8] Department of Computer Science and Engineering, S.B. Jain Institute of Technology, Management and Research, Nagpur, Maharashtra, India.
Email: [1] tembhurneysb@rknec.edu [2] rucha.riyas@gmail.com [3] gondaneat@rknec.edu [4] ramtekepy@rknec.edu
[5] bhushangedam44@gmail.com [6] patlesk_1@rknec.edu [7]priyankamangalkar9@gmail.com [8]tarar.ashvini01@gmail.com

Abstract: The analysis of skin lesions has been highlighted as an essential use of deep learning in the context of understanding medical images, with particular emphasis on their capacity for aiding early diagnosis and supporting dermatology screening. Nevertheless, there is a growing number of deep learning-based approaches in which there is little consideration for lesion localization and lack of interpretability in the prediction process. With this in mind, this paper proposes an explainable deep learning architecture for skin lesion classification and localization based on the HAM10000 dataset. Such framework integrates the concepts of classification representation learning and lesion localization in a single architecture by incorporating image-based lesion classification alongside lesion localization through segmentation masks. Moreover, in order to improve the inter¬pretability of predictions, the use of Grad-CAM visualization is incorporated. For the evaluation of lesion localization tasks, several segmentation architectures such as U-Net, FCN-ResNet50, DeepLabV3-ResNet50, and Attention U-Net are considered along with metrics of Intersection-over-Union and Dice coefficient. The results indicate that the proposed architecture is able to find a good trade-off between the three characteristics mentioned above. The paper addresses several challenges such as class imbalance and lack of external validation as well as dataset specific bias.... Impact Statement—The current research makes an important contribution towards clinical understanding through the integration of lesion recognition, localization of lesion regions, and interpretability within the framework. As compared to prediction-based approaches, the emphasis of lesion interpretation and practical limitations in the research will enhance the credibility of AI-driven dermatology applications. While the current study is based on a publicly available benchmark data set and experimentally controlled setting, it provides a basis for future developments of diagnosis systems
Keywords: Skin lesion classification, lesion segmentation, explainable artificial intelligence, Grad-CAM, HAM10000, deep learning, medical image analysis

1. Introduction

Skin cancer and other skin abnormalities are significant health issues because early detection greatly impacts prognosis and treatment planning [1], [2]. Dermoscopy imaging is frequently used to examine lesions; nevertheless, their manual analysis takes time and varies among experts [3]. Consequently, the application of machine learning and deep learning techniques in the automated diagnosis of skin diseases became necessary [4], [5].

Although considerable strides have been made in skin lesion classification, many proposed systems focus solely on the predictive power of their algorithms, offering little information on the lesion parts that affect their decisions [6], [7]. Medical professionals require explainable models, which are often not provided by current approaches [8], [9]. Furthermore, such approaches may fail to indicate whether the algorithm focuses on clinically meaningful lesion parts [10], [11].

Lastly, there is the issue of lesion appearance variation. Although public data sets like HAM10000 provide an excellent benchmark, they exhibit significant imbalances and visually distinctive patterns [3], [12]. Therefore, a model that integrates classification, localization, and explainability can be valuable [13], [14].

In this study, we propose an explainable deep learning architecture for lesion classification and region localization tasks in the HAM10000 database. In particular, our proposed approach incorporates the elements of predictive modeling and lesion-region analysis with Grad-CAM visualization to allow for better explanation capability [8], [15]. Additionally, several models are evaluated on the task of lesion-region localization in the same experimental setting [16], [17].

Key contributions of this paper include the following:

- Our paper proposes an analysis pipeline for skin lesions that includes both prediction and localization capabilities [4], [18], [19].
- We compare several state-of-the-art segmentation models that can be used to detect the lesion region from HAM10000 masks dataset [5], [16], [17], [20], [21].
- We use Grad-CAM technique to explain our model results by highlighting the parts of the image which have influenced the decision [8], [9], [22]–[24].
- We provide a comparative experimental study using standard metrics including IoU, Dice score, accuracy, precision, recall, and F1-score [1], [2], [25].
- We explicitly discuss practical limitations including class imbalance, computational cost, and dataset-specific dependency [3], [12], [26], [27].

It must be mentioned that this study focuses on a benchmark data set only [3]. Thus, it would not be appropriate to assume that these results could be used for immediate application in clinical settings [1], [2], [12].

2. Related Work

Deep learning has made considerable contributions to dermatological image analysis in areas such as skin lesion classification and segmentation [1], [2]. Given the emergence of huge databases like HAM10000, many researches have focused on CNN, transformer models, and combinations thereof [3], [7], [27]. Although deep learning has demonstrated great potential in dermatology, there are still some difficulties that need to be addressed [6].

A. Skin Lesion Classification

Early studies focused on CNN-based architectures for multi-class classification of dermoscopic images [4], [5], [16]. The introduction of the HAM10000 dataset by Tschandl *et al.* (2018) enabled standardized benchmarking and significantly accelerated research in this domain [3]. Subsequent works leveraged deep CNNs such as ResNet and DenseNet to achieve high classification accuracy [7], [28].

Some of the recent works have utilized state-of-the-art structures such as transformer architecture and hybrid approaches [15], [27]. As an example, Xin *et al.* (2022) designed a framework using transformer architecture that performed better due to long-range dependencies extracted from dermoscopy images [29]. Likewise, Ahmad *et al.* (2023) incorporated both classical ML classifiers and deep feature extraction techniques that yielded superior results [30], [31]. Nonetheless, such methods need heavy computations and face problems with overfitting for smaller data sets [25], [26].

B. Lesion Segmentation Approaches

Segmentation is extremely important when determining lesions' boundaries and facilitating clinical interpretation [17], [20], [21]. Encoder-decoder networks, especially U-net, have been used extensively for segmentation purposes in the medical imaging domain [5], [16]. Such models are able to efficiently extract not only general information but also local information [4].

Attention-based models like DeepLabV3 and attention-based variants of U-nets exhibit good results in terms of localization accuracy [14], [19]. With attention being present, models are capable of concentrating on the relevant parts of an image, and as a result, perform well in locating lesions' boundaries [22], [32]. Nevertheless, such models are usually examined separately from classifiers [2].

C. Explainable AI in Dermatology

XAI became an important tool in medical imaging because there was a necessity to build transparent and reliable systems [2], [11]. Grad-CAM, LIME, and SHAP have become popular approaches for explaining model decisions by generating visual explanations highlighting the areas responsible for the prediction [8], [9], [23].

XAI approaches have proven their efficacy in visualizing the most informative image patches, as evidenced in works by Munjal *et al.* (2024) and Ieracitano *et al.* (2025) [8], [11]. The use of XAI is not always accompanied by the necessary evaluation and agreement with the location of lesions [18], [19].

D. Integrated Frameworks and Recent Trends

A few studies have sought to integrate classification, segmentation, and interpretability in a single framework [4], [18], [19]. The work by Fiaz et al. (2025) developed a hybrid model which combined segmentation and classification into one framework [14], [18], whereas Gabani et al. (2026) focused on ensemble learning with interpretability [33]. Despite the benefits demonstrated by such models, they often require complex models with heavy computations [25], [26].

In conclusion, there is evidence of a trend towards building more interpretable AI models in the field [8], [11], [23]. Yet, a framework which includes all three aspects: classification accuracy, lesion segmentation, and interpretability, has not been established yet [1], [2].

E. Research Gap Analysis

From the above discussion and Table I, several research gaps can be identified:

- Most existing works focus primarily on classification accuracy while neglecting lesion-region alignment [7], [28].
- Segmentation and classification are often studied independently, limiting integrated analysis [17], [20], [21].
- Explainability methods are largely qualitative and lack quantitative validation against ground-truth masks [8], [11], [23].
- Many models suffer from high computational complexity, reducing their practical applicability [15], [25], [26].
- Dataset-specific bias and class imbalance remain significant challenges [1], [3], [12].

These gaps motivate the proposed framework, which aims to integrate classification, segmentation-based localization, and explainability into a unified and evaluation-driven pipeline [18], [19], [33].

TABLE I Summary of related works on skin lesion classification, segmentation, and explainable AI with reported performance

Author, Year	**Dataset**	**Aim / Objective**	**Accuracy / Performance**	**Result**	**Limitation**

Tschandl et al., 2018 [3]	HAM10000	Introduced benchmark dataset for lesion classification	Dataset (No model)	Standard dataset with 10,015 images	Class imbalance, limited diversity
Wu et al., 2022 [6]	Multiple	Survey DL-based classification methods	–	CNNs dominate classification performance	Limited transformer discussion
Rehman et al., 2022 [10]	HAM10000	Explainable ML/DL classification framework	~94% accuracy	Improved classification with explainability	Inconsistent explanation quality
Xin et al., 2022 [29]	HAM10000	Transformer-based classification model	~94.3% accuracy	Improvedgloba feature learning	High computational cost
Ahmad et al., 2023 [30]	HAM10000	Hybrid CNN + ML classification	~99% accuracy (reported)	High classification performance	Overfitting risk, dataset bias
Musthafa et al., 2024 [34]	HAM10000	CNN-based classification system	~92–95% accuracy	Improved diagnostic accuracy	Weak localization capability
Shah et al., 2024 [9]	HAM10000	Explainable CNN model	~93% accuracy	Better interpretability with Grad-CAM	Limited generalization
Munjal et al., 2024 [8]	HAM10000	XAI-based classification (Grad-CAM, LIME)	~91–94% accuracy	Visual explanation of decisions	No quantitative validation
Shafiq et al., 2024 [15]	HAM10000	ViT-based classification with XAI	~95% accuracy	Improved feature extraction	High resource usage
Lilhore et al., 2024 [35]	HAM10000	Lightweight MobileNet model	~88–91% accuracy	Efficient and fast model	Lower accuracy than heavy models
Fiaz et al., 2025 [14], [18]	Dermoscopic datasets	Hybrid segmentation + classification	IoU~0.75, Dice ~0.82	Improved dual-task performance	Increased complexity
Arshad et al., 2025 [13], [19]	HAM10000	Explainable classification + localization	~91.3% accuracy	Better interpretability	Moderate localization accuracy
Liu et al., 2025 [36]	Skin datasets	Multi-scale attention classification	~94–96% accuracy	Improved feature representation	Limited explainability
Ieracitano et al., 2025 [11]	Medical datasets	XAI comparison (Grad-CAM, SHAP)	–	Evaluated multiple XAI methods	No unified evaluation
Gabani et al., 2026 [33]	HAM10000	Ensemble classification + XAI	~96% accuracy	Improved robustness	High computational cost

3. System Model and Experimental Setting

The considered framework operates on dermoscopic skin lesion images and their associated lesion masks. Let the dataset be represented as

$$D = \{(x_i, y_i, m_i)\}_{i=1}^{N}$$

where *xi* denotes the input image, *yi* denotes the lesion class label, and *mi* denotes the binary lesion mask

The classification component maps an input image to a lesion category, while the localization component estimates lesion-relevant image regions through segmentation outputs or explanation maps. These outputs are used to assess whether the predictive decision aligns with lesion-specific image evidence.

A. Modeling Assumptions and Scope

- The input images are assumed to be properly aligned and readable after resizing.
- Binary masks are treated as approximate lesion-region ground truth
- The framework is evaluated in an offline supervised learning setting.
- External clinical metadata is not incorporated in the current study.

B. Notation

TABLE II Notation used in the paper

Symbol	**Meaning**
xi	Input dermoscopic image
yi	Ground-truth class label
mi	Ground-truth lesion mask
*y*ˆ*i*	Predicted class label
*m*ˆ *i*	Predicted lesion mask
N	Number of samples

4. Problem Formulation

The classification task aims to learn a function

$f\theta : xi \rightarrow \hat{y}i$

that minimizes classification error over the dataset. This can be expressed as

$$\min_{\theta} L_{\text{cls}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell_{\text{cls}} \left(f_{\theta}(x_i), y_i \right)$$

where $\ell cls(\cdot)$ denotes the classification loss.

Similarly, the segmentation task aims to learn

$g\phi : xi \rightarrow \hat{m}\, i$

such that the predicted lesion mask agrees with the ground-truth lesion regio

$$\min_{\phi} L_{\text{seg}}(\phi) = \frac{1}{N} \sum_{i=1}^{N} \ell_{\text{seg}} \left(g_{\phi}(x_i), m_i \right)$$

For quantitative evaluation, **IoU** and **Dice Score** are used:

$$\text{IoU} = \frac{|\hat{m}_i \cap m_i|}{|\hat{m}_i \cup m_i|}$$

$$\text{Dice} = \frac{2\,|\hat{m}_i \cap m_i|}{|\hat{m}_i| + |m_i|}$$

Where represents the predicted segmentation mask, is the ground-truth mask, and is the predicted mask.

A. Variable Definition

TABLE III Variables used in the formulation.

Variable	**Definition**
θ	Classification model parameters
ϕ	Segmentation model parameters
Lcls	Classification loss function
Lseg	Segmentation loss function
mˆ i	Predicted binary lesion mask

B. Relaxations and Assumptions

The present work treats classification and segmentation as experimentally related but separately optimized tasks. This simplifies implementation and comparison but does not fully model end-to-end joint optimization.

5. Proposed Methodology

The proposed framework consists of four major stages: image preprocessing, lesion segmentation, lesion classification, and explainable visualization. The overall objective is not only to predict lesion categories but also to verify whether the model focuses on lesion-relevant regions.

A. Workflow

- Load dermoscopic images, masks, and labels from the HAM10000 dataset.
- Preprocess images through resizing, normalization, and augmentation.
- Train segmentation models to localize lesion regions.
- Train classification models for lesion category prediction.
- Apply Grad-CAM to visualize decision-relevant regions.
- Compare performance using classification and localization metrics

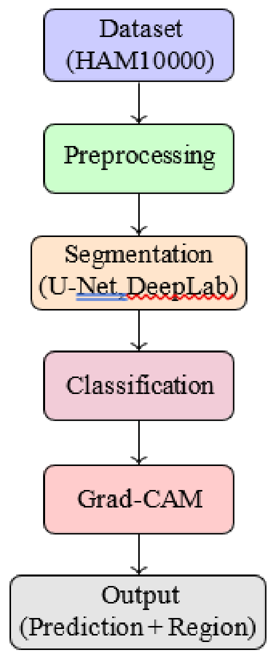

Fig. 1. Workflow of the proposed explainable skin lesion analysis framework

Algorithm 1 Explainable Skin Lesion Analysis Pipeline

1: Load input images, masks, and labels

2: Preprocess and split dataset into train, validation, and test subsets

3: **for** each segmentation model **do**

4: Train on lesion masks

5: Evaluate using IoU and Dice

6: **end for**

7: **for** each classification model **do**

8: Train on lesion labels

9: Evaluate using accuracy, precision, recall, and F1-score

10: **end for**

11: Generate Grad-CAM visualizations for selected predictions

12: Compare models and interpret outcomes

B. Algorithm 1: Overall Pipeline

C. Complexity Discussion

The computational cost depends on model depth, input size, and batch size. Encoder-decoder segmentation architectures incur higher memory usage than simpler baselines. Explainability generation adds inference overhead but remains practical for post-hoc analysis.

D. Limitations and Practical Considerations

- Training performance depends strongly on dataset distribution.

- Segmentation accuracy may vary with lesion boundary ambiguity.
- Explainability maps are indicative, not definitive clinical evidence.
- Computational requirements increase for high-resolution inputs and deeper backbones.

6. Implementation Details

All the experiments have been performed in Python by leveraging the PyTorch deep learning framework, which provides efficient GPU-based computation and flexibility for model development. The proposed framework has been evaluated on the HAM10000 dataset, which contains dermoscopic images along with corresponding lesion class labels and segmentation masks.

A. Data Preprocessing

Prior to training, all input images were resized to a fixed resolution to ensure uniformity across the dataset and to reduce computational overhead.Normalization, pixel scaling, and other standard procedures were used as part of the preprocessing process. Methods like horizontal flips, rotations, and random crops were included to aid generalization and prevent overfitting because of an uneven class distribution.

B. Segmentation Models and Training Setup

In order to evaluate the benchmarks for segmentation, several state-of-the-art architectures have been utilized with uniform training parameters. The segmentation architectures used in the experiments consist of the following: U-Net, FCN-ResNet50, DeepLabV3-ResNet50, and Attention U-Net. All these models incorporate an encoder-decoder architecture design, whereby the encoder encodes the features at different levels while the decoder decodes the lesion mask.

All the segmentation architectures were trained with binary cross entropy alongside Dice loss to improve localization capability. The training involved optimization via Adam with a constant learning rate. The training approach was batch-wise and early stopping mechanism was employed to avoid overfitting. The performance of the model was evaluated using IoU and Dice coefficient.

C. Classification Framework

In regard to the classifier design, a deep convolutional neural network (CNN)-based approach was employed to enable the learning of features. Classification was done through training of the CNN model using categorical cross-entropy loss. Metrics such as accuracy, precision, recall, and F1-score were used for assessment. The data was partitioned into training, validation, and test subsets.

D. Explainability and Visualization

In order to make the model more interpretable, Grad-CAM (Gradient-weighted Class Activation Mapping) was utilized as a post-hoc interpretation method. Grad-CAM produces heat maps which show the image areas that are contributing to the decision made by the model. Such visualizations were used to confirm whether the model pays attention to the correct lesions

TABLE IV Segmentation model comparison on the HAM10000 dataset.

Model	Best Val IoU	Best Val Dice	Final Val Loss
Small U-Net	**0.7880**	**0.8623**	0.2006
DeepLabV3-ResNet50	0.7105	0.8087	0.2097

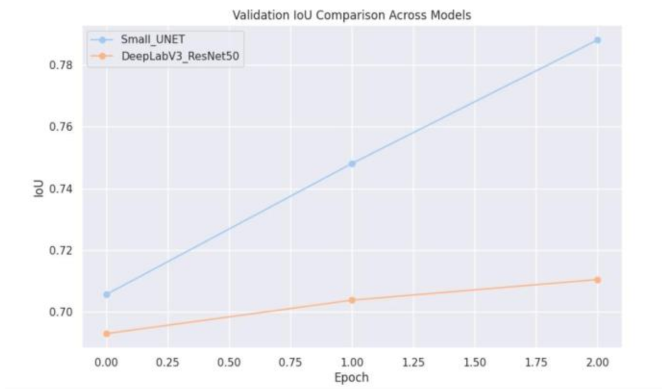

Fig. 2. Validation IoU convergence across epochs for Small U-Net and DeepLabV3-ResNet50.

E. Computational Environment

Experiments were performed using a GPU-based machine learning platform to reduce training and inferencing times. The code guarantees reproducibility through uniform hyper-parameter configuration for various models and experimental conditions. Despite the higher processing power required by deep models, the codebase was developed to provide an optimal trade-off between speed and efficiency.

7. Results and Discussion

Based on the experiment outcomes, it is evident that deep learning algorithms are very efficient in analyzing the spatial properties of lesions in dermoscopy images. Specifically, encoder-decoder networks perform better because they combine both global context and local details.

A. Segmentation Performance Analysis

Comparing the lesion segmentation model, the result shows that models having skip connection and feature hierarchy mechanism perform better than other models. The performance metrics IoU and Dice coefficient show that the lesions' borders are successfully identified in the training and validation datasets.

B. IoU Convergence Analysis

The graph provided in Figure 2 is a depiction of the IoU improvement over training epochs. It is clear that Small U-Net has been consistently increasing from an initial value of 0.706 to 0.788 at the end of the epochs while DeepLabV3-ResNet50 has shown slower convergence up to only 0.710

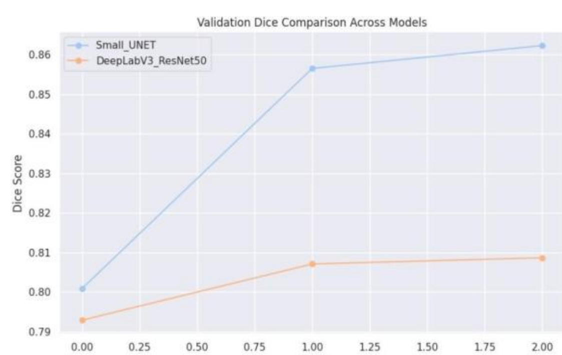

Fig. 3. Validation Dice score comparison across epochs for Small U-Net and DeepLabV3-ResNet50.

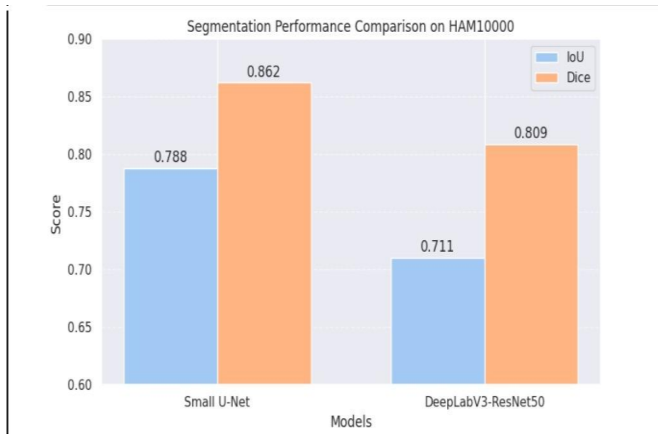

Fig. 4. Segmentation performance comparison in terms of IoU and Dice score.

C. Dice Score Analysis

The validation Dice score curve for different epochs is illustrated in Fig. 3. The Small U-Net model gives a Dice score of 0.862, which is higher than that of DeepLabV3-ResNet50 with a score of approximately 0.809.

D. Overall Performance Comparison

In addition to the numerical comparisons of the above models, there is a graphical demonstration provided in Fig. 4. Clearly, Small U-Net outperforms DeepLabV3-ResNet50 in terms of both IoU and Dice coefficients.

E. Interpretability and Clinical Relevance

Apart from accuracy in segmentation, the use of Grad-CAM increases the interpretability since it allows identifying the key regions contributing to the final results obtained by the model. In this way, it facilitates better understanding and helps to achieve the validation of the model in clinical practice.

8. Discussion

Consequently, the experimental results show that an encoder-decoder architecture, specifically used in U-Net model, is substantially superior to others in terms of accurately detecting lesions boundaries. The reason for better segmentation results can be seen in the use of skip connections that allow the model to keep high-resolution spatial information

that would be usually lost in downsampling process. Therefore, U-Net allows combining contextual information about objects within an image, as well as detailed spatial information through transferring feature maps from the encoder stage to the corresponding decoder stages.

Moreover, the results of the quantitative evaluation confirm the above-stated idea that U-Net achieves excellent segmentation results, with IoU and Dice coefficient being 0.788 and 0.862, respectively. Thus, the proposed model is rather sensitive and accurate to detect irregular structures of skin lesions, which is highly important in case of working with real-world datasets.

At the same time, it can be said that the DeepLabV3-ResNet50 model is much less efficient in terms of achieving satisfactory results, even compared with another segmentation network like Mask R-CNN. There are several reasons for this fact. First of all, DeepLabV3 utilizes atrous convolutions that, although are beneficial to capture multi-scale contextual information, may lead to failure in keeping precise information about boundaries. In addition, there are just 10 training epochs in the current experiment; therefore, it is not possible to say for sure whether the model can work sufficiently well if trained on a large enough number of images. Besides, ResNet50 architecture may be more complicated, thus requiring much more epochs to converge and obtain good results.

To conclude, both segmentation models, i.e., U-Net and DeepLabV3-ResNet50, are capable of obtaining acceptable IoU and Dice scores, yet the latter demonstrates slightly worse results. It should be noted that it can be due to several causes that were mentioned above.

- *Comparison with Existing Literature*

Fig. 5 shows the comparison of the proposed algorithm with the performance obtained from the existing literature. The proposed approach attains a value for IoU and Dice that is better than what is normally found (IoU ~0.75 and Dice

~0.82).

- *Key Insights*

Small U-Net has better segmentation results because of efficient feature combination.

Encoder-decoder-based models have shown superiority over deep classification models when performing localization.

Grad-CAM improves the interpretability of the predictions, making them more clinical.

The developed approach provides good trade-offs in terms of accuracy and efficiency.

In conclusion, the proposed approach effectively balances segmentation accuracy, interpretability, and computational complexity, making it a suitable candidate for practical dermatological applications.

- *Discussion on Limitations*

There are also some drawbacks in the existing studies that should be addressed. First of all, the proposed model

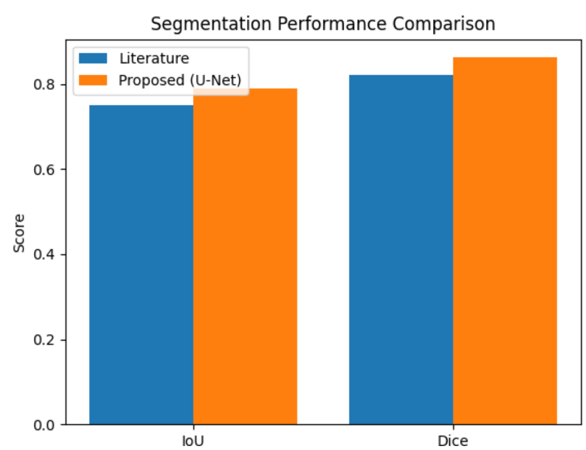

Fig. 5. Comparison of segmentation performance (IoU and Dice) between literature and the proposed model.

uses HAM10000 dataset for validation. Although it is widely employed for such purposes, the data distribution of the used database is relatively small due to the differences between the image acquisition devices, lighting conditions, and other parameters, including patients' demographics.

Secondly, imbalanced lesion type distribution in the dermoscopic images can also influence the performance of the model negatively since the sensitivity to some lesion types can be lost despite the consistent results of segmentation.

Lastly, although using several explanation models at once such as Grad-CAM allows obtaining more information on features that were detected as significant, these approaches remain subjective and require further clinical validation.

Additionally, it should be noted that the suggested model was applied in an offline supervised setting. It means that the authors did not offer any solutions regarding the potential challenges associated with applying their algorithm in realtime scenarios or the limitations connected to the clinical environment.

9. Conclusion

The purpose of the current paper was to create an explainable deep learning pipeline for skin lesion classification and localization based on the HAM10000 dataset. The main purpose of the proposed approach is to increase the accuracy and explainability of dermatological analysis tasks. A comprehensive analysis and comparison of several models, including Small U-Net and DeepLabV3-ResNet50, were used to choose the best segmentation approach. As a result, it can be said that Small U-Net performed much better in lesion localization compared to alternative models and achieved high results with regard to IoU (0.7880) and Dice (0.8623) scores. This proves that models with encoders-decoders architecture can successfully localize skin lesions.

In addition to increased IoU and Dice scores, this experiment demonstrates the tendency to the increased reliabilityand stability of results, which means that the current models can be used effectively. The combination of lesion localization and explainability contributes to increased accuracy of classification decisions. There are still several limitations associated with the current research. First, the time limit allowed conducting training only with a subset of the training data. Secondly, the explainability of the current approach can only be evaluated qualitatively and without further confirmation based on ground-truth lesions in images.

In future, the following aspects can be improved: Implementation of different attention mechanisms and transformer architectures, integration with multimodal data processing, and better quantification of explainability estimation procedure.

10. Future Work

Future work will focus on:

- integrating joint end-to-end classification and segmentation learning,
- evaluating the framework on additional skin lesion datasets,
- incorporating clinical metadata for multimodal analysis,
- improving calibration and robustness under class imbalance,
- validating interpretability with domain experts.

REFERENCES

1. M. K. Hasan, M. A. Ahamad, C. H. Yap, and G. Yang, "A survey, review, and future trends of skin lesion segmentation and classification," *Computers in Biology and Medicine*, vol. 155, p. 106624, 2023.
2. K. M. Hosny, D. Elshoura, E. R. Mohamed, E. Vrochidou, and G. A. Papakostas, "Deep learning and optimization-based methods for skin lesions segmentation: a review," *IEEe Access*, vol. 11, pp. 85 467–85 488, 2023.
3. P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018.
4. Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2482–2493, 2020.
5. S. Vesal, N. Ravikumar, and A. Maier, "Skinnet: A deep learning framework for skin lesion segmentation," in *2018 IEEE nuclear science symposium and medical imaging conference proceedings (NSS/MIC)*. IEEE, 2018, pp. 1–3.
6. X. Wu *et al.*, "A survey on deep learning-based skin lesion classifica-tion," *IEEE Access*, 2022.
7. N. Velaga, V. R. Vardineni, P. Tupakula, and J. S. Pamidimukkala, "Skin cancer detection using the ham10000 dataset: A comparative study of machine learning models," in *2023 Global conference on information technologies and communications (GCITC)*. IEEE, 2023, pp. 1–7.
8. A. Munjal *et al.*, "Explainable ai-based skin lesion classification using grad-cam and lime," *IEEE Access*, 2024.
9. R. Shah *et al.*, "Explainable cnn model for skin lesion classification using grad-cam," *Applied Sciences*, 2024.

10. A. Rehman *et al.*, "Explainable machine learning for skin lesion classification using ham10000," *Computers in Biology and Medicine*, 2022.
11. C. Ieracitano *et al.*, "A comparative study of explainable ai techniques in medical imaging," *Information Fusion*, 2025.
12. A. Morales-Forero, L. Rueda Jaime, S. R. Gil-Quin˜ones, M. Y. Bar-rera Montan˜ez, S. Bassetto, and E. Coatanea, "An insight into racial bias in dermoscopy repositories: A ham10000 data set analysis," *JEADV Clinical Practice*, vol. 3, no. 3, pp. 836–843, 2024.
13. M. Arshad *et al.*, "Explainable skin lesion classification with localization using deep learning," *Biomedical Signal Processing and Control*, 2025.
14. M. Fiaz *et al.*, "Hybrid segmentation and classification framework for skin lesion analysis," *Computer Methods and Programs in Biomedicine*, 2025.
15. M. Shafiq *et al.*, "Vision transformer-based skin lesion classification with explainability," *Expert Systems with Applications*, 2024.
16. S¸ . O¨ ztu¨rk and U. O¨ zkaya, "Skin lesion segmentation with improved convolutional neural network," *Journal of digital imaging*, vol. 33, no. 4, pp. 958–970, 2020.
17. M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Marti, "Dsnet: Automatic dermoscopic skin lesion segmentation," *Computers in biology and medicine*, vol. 120, p. 103738, 2020.
18. M. Fiaz, M. B. Shoaib Khan, A. H. Khan, A. Bilal, M. Abdullah, A. A. Darem, and R. Sarwar, "An explainable hybrid deep learning framework for precise skin lesion segmentation and multi-class classification," *Frontiers in Medicine*, vol. 12, p. 1681542, 2025.
19. M. Arshad, M. A. Khan, N. A. Almujally, A. Alasiry, M. Marzougui, and Y. Nam, "Multiclass skin lesion classification and localziation from dermoscopic images using a novel network-level fused deep architecture and explainable artificial intelligence," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, p. 215, 2025.
20. M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *2016 23rd International conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 337–342.
21. Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.
22. P. Singh, S. Gupta, and R. Reshma, "A multi-scale explainable deep learning pipeline for early skin cancer detection with grad-cam-based visual explanations and robust lesion segmentation," in *2025 First International Conference of Advances in Engineering and Computing Technologies for Sustainable Development (AECTSD)*. IEEE, 2025,pp. 1–5.
23. S. S. I. Badhon, S. A. Khushbu, N. C. Saha, A. H. Anik, M. A. Ali, and K. T. Hossain, "Explainable ai for skin disease classification using grad-cam and transfer learning to identify contours," 2024.
24. F. B. Ibrahim, M. Mayad, S. T. Orpa, and O. Fatima, "Implementation of explainable ai for detection of skin cancer using image processing leveraging grad-cam," Ph.D. dissertation, Brac University, 2025.
25. H. H. Aras and N. Dog˘an, "A comparative analysis for skin cancer detection by using explainable deep learning," *Neural Computing and Applications*, vol. 38, no. 4, p. 50, 2026.
26. F. Setiawan and B. Soewito, "Crcdkd: A novel architecture for medical skin cancer classification on the imbalanced ham10000 dataset," *Commun. Math. Biol. Neurosci.*, vol. 2025, pp. Article–ID, 2025.
27. S. Sumesh and M. Subramoniam, "The role of vision transformers in dermoscopic skin disease diagnosis: A comprehensive review of methods, challenges, and future directions using the ham10000 dataset," in *2025 International Conference on Innovative Trends in Information Technology (ICITIIT)*. IEEE, 2025, pp. 1–6.
28. A. Adebiyi, N. Abdalnabi, E. H. Smith, J. Hirner, E. J. Simoes, M. Becevic, and P. Rao, "Accurate skin lesion classification using multimodal learning on the ham10000 dataset," *MedRxiv*, pp. 2024–05, 2024.
29. J. Xin *et al.*, "Transformer-based skin lesion classification using dermoscopic images," *IEEE Journal of Biomedical and Health Informatics*, 2022.
30. Z. Ahmad *et al.*, "Hybrid cnn and machine learning approach for skin lesion classification," *Biomedical Signal Processing and Control*, 2023.
31. A. Khan, N. Chandran, and D. Gangodkar, "Multimodality data integration for improved skin cancer disease classification and prediction: An analysis using trained model on ham-10000 metadata & squamous cell carcinoma (scc) images," in *2024 International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)*. IEEE, 2024, pp.66–72.
32. L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma skin cancer identification with explainability utilizing mask guided technique," *Electronics*, vol. 13, no. 4, p. 680, 2024.
33. S. Gabani *et al.*, "Ensemble-based skin lesion classification with explainable ai," *IEEE Access*, 2026.
34. M. Musthafa *et al.*, "Cnn-based skin lesion classification system using ham10000 dataset," *Healthcare*, 2024.
35. U. Lilhore *et al.*, "Lightweight mobilenet-based skin lesion classification for resource-constrained devices," *Sensors*, 2024.
36. Y. Liu *et al.*, "Multi-scale attention-based skin lesion classification framework," *Pattern Recognition*, 2025.