Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 01, PP. 239-254, 2025
DOI: https://doi.org/10.54216/JISIoT.170117

CBi-BERT: Efficient Skin Disease Image Segmentation Using Patch-Based Deep Feature Mapping and BERT-Based Attention Mechanism

Summi Goindi[1,*], Khushal Thakur[2], Divneet Singh Kapoor[2]

[1]Research Scholar, Chandigarh University, Mohali, Punjab, India

[2]Associate Professor, Chandigarh University, Mohali, Punjab, India

Emails: summikhurana7@gmail.com; khushal.thakur@cumail.in; divneet.ece@cumail.in

Abstract

Skin image segmentation serves as a vital undertaking in medical image analysis, specifically in dermatology, since it enables the detection of skin diseases and the assessment of effectiveness of treatments. Segmenting skin lesions from photographs is a crucial step in working towards this patchive. Nevertheless, the work of segmenting skin lesions is difficult due to the existence of both artificial and natural deviations, inherent characteristics like the shape of the lesion), and deviations in the circumstances during which the images are obtained. In recent years, researchers have been investigating the feasibility of utilizing deep-learning models for skin lesion segmentation. Deep learning methodologies have exhibited encouraging outcomes in various image segmentation initiatives, thereby preventing the possibility of automating and enhancing the precision of skin segmentation. This paper introduces a new hybrid method, named the CBi-BERT framework, aimed to improve the results and architectures of medical image segmentation or patch detection tasks. This architecture employs Convolutional Neural Networks (CNNs) for feature extraction as well Bidirectional LSTM-based encoders to process sequence information and BERT based attention collection across the strongest features. Image normalization, resizing and data augmentation techniques are applied in the proposed method to deal with major challenges faced during medical imaging such as rate of image quality differentiation from noise or bias between benign vs. malign features. We evaluate the performance of CBi-BERT to those benchmark datasets and state-of-the-art models, showing that CBi-BERT outperforms them in all relevant metrics such as Intersection over Union (IoU), recall, mean average precision (bin-MAP) DICE coefficient. Specifically, for images sized 256x256 the model achieved IoU =0.9, recall=0.92, mAP=0.89 and Dice coefficient: =0.91 thereby outperforming some advanced state-of-the-art models as ResNet50, VGG16, UNet, EfficientNet-B-01 Our results show that the framework is able to detect and segment important structures in medical images with high precision which makes it a compelling tool for AI based Healthcare solutions.

Received: January 28, 2025 Revised: March 03, 2025 Accepted: April 22, 2025

Keywords: Skin; Image segmentation; Deep Learning; IOU; Segmentation

1. Introduction

In the fields of medicine and dermatology, skin lesion segmentation is an essential activity that aims to precisely identify and isolate skin lesions from background tissue or skin that is normal. It is necessary to make an early diagnosis, planning, treatment, and surveillance of skin illnesses, especially those that are malignant like skin cancer [1,2]. An input image is divided into several segments or areas, each of which represents a lesion or non-lesion, during the procedure. This allows the extraction of quantitative characteristics for diagnostic analysis like shape, size, color, and texture. Still, the task of skin lesion segmentation can be complicated by diverse characteristics of lesions, including their different appearance, size, form, color, and appearance as well as artifacts or noise which can be encountered in medical images. A variety of methods and algorithms is used to segment skin lesions, including traditional image processing methods, ML methodologies as well as deep learning-based solutions. Deep learning (DL) models have demonstrated superlative effectiveness when segmenting skin lesions due to their ability to capture sophisticated variations and dependencies in the

appearance of lesions and achieve exceptional accuracy in complex scenarios [6, 7]. The methods include thresholding methods, edge detection algorithms, region-based methodologies, texture analysis methods, machine learning (ML) methods, hybrid methodologies, as well as post-processing approaches. Thresholding methods segment the image by using pixel intensity values, while edge detection identifies the lines that divide different areas. Region-based methods, including segmentation by region growing, split and merge, clustering, and contour snakes, also segment the image. The divisions can be performed by dividing the image into homogenous regions that are clustered based on similarities in color, texture, or other characteristics. Texture analysis divides the picture based on how these pixel intensities are organized. Machine learning methods, including k-nearest neighbors, Random Forests, or Support Vector Machines, can be trained on features handcrafted from images of skin lesions to determine whether each pixel is lesioned or not. Deep learning methods, specifically DL methods, have been superlative for segmenting skin lesions. In the domain of skin lesion segmentation, DL models encounter obstacles to data quality, the ability to generalize to unknown data, integration, and interpretability [20,23,28,29]. Researchers should concentrate on amassing massive datasets containing a variety of skin types, lesion types, and imaging techniques to surpass these obstacles. Enhancing the ability to interpret and convey information is vital for establishing confidence in clinical practice. Integration and clinical validation are crucial components in evaluating the performance and dependability of models. To achieve real-time implementation and deployment, it is necessary to optimize hardware acceleration techniques, model architectures, and inference algorithms. Achieving personalized medicine and precision dermatology is feasible through the customization of models to account for unique patient attributes. Fusion and integration of multiple modalities can enhance precision and dependability. Ethical and legal issues are of the utmost importance when it comes to responsible implementation and acceptance. Subsequent trajectories encompass ongoing model advancement, integration with clinical workflows, validation studies in the real world, joint research efforts, educational and training programs, considerations of ethics and society, equitable and global accessibility, and ongoing monitoring and enhancement. Deep learning models can further enhance dermatological assessment, therapy, and research by confronting these obstacles and investigating potential future avenues; in doing so, they will eventually help with the advancement of precise dermatology and the enhancement of patient outcomes

- **Contribution**

This work makes significant contributions to medical image analysis, especially for problems such as segmentation and patch detection. This Figure 1 demonstration introduces a methodology that lays down some very promising foundations for the construction of highly accurate and robust models, combining modern techniques like CNN-based feature extraction, BiLSTM-based temporal encoding and BERT-based attention mechanisms. The need for accurate detection and classification of localized features in medical imaging, a field where the number of data points used is less than most other computer vision challenges.

- The first major challenge in capturing both local and global information with medical images was solved by the novel mechanism of combining sequential encoding using BiLSTM from FR to CMF encoder, and contextual attention using BERT (Bidirectional Encoder Representations for Transformers) making up-of a CBi-BERT. The integration of a hybrid method extends the depth perception and awareness to distinguish fine-grained details form overall context during segmentation, thus improving accuracy in patch identification.
- Moreover, a comprehensive performance evaluation on various image scales and comparisons with other existing models prove that the proposed CBi-BERT approach outperforms state-of-the-art approaches. Overall, the results show that our model performs substantially better in most metrics including IoU, Recall and mAP as well Dice coefficient than existing methods specially when images are higher resolution (256x256). In medical, accuracy and reliability are crucial, so these enhancements can be beneficial.
- In conclusion, the present work makes a substantial contribution to providing a well-equipped and transferable segmentation algorithm for medical image analysis with potential extensions in various modalities such as dermatology, oncology or radiological imaging. The approach and findings establish a good groundwork for future research and development, to create faster AI-driven tools in medical domain.

The rest of the paper is structured as follows: In Section 3, we provide a detailed description about how we develop and validate our framework CBi-BERT for skin disease image segmentation. The Introduction section details the relevance of precise skin lesion segmentation in medical diagnostics, emphasizing a number obstacles that arise from the changes in appearance and imaging conditions between lesions. We also provide a review of traditional and DL-based segmentation methods, protecting the ground work for way procedure. The Background and Related Work section explain about the existing approaches and deep learning models-based methods like U-Net, DeepLab, and Mask R-CNN which are vastly used in skin lesion segmentation. Next, we go over the shortcomings of these models and discuss that more solidly grounded approaches, especially in medical imaging are still required. The Methodology section: This is where the proposed CBi-BERT framework comes into play. The hybrid model comprises of CNNs for feature extraction, BiLSTM

to capture sequential dependencies and BERT as attention mechanisms followed by normalization steps such as resizing, data augmentation etc. The design of the framework also tackles with problems that need to capture both local and global information in medical images. We evaluate the performance of our CBi-BERT framework in Experiments and Results section comparing it with state-of-the-art models based on multiple metrics IoU, recall, mAP (mean Average Precision) & Dice coefficient. Results show that CBi-BERT outperforms other baselines, especially with high resolution images. In the end, Conclusion briefly describes what has been contributed in this work and how much promising CBi-BERT performance ability can improve diagnostic accuracy of medical imaging as well its applications to other health care domains.

2. Background and Related Work

The prospective benefits of utilizing deep learning in skin image segmentation for the advancement of dermatological studies and clinical practice are substantial. As a result of ongoing advancements in model structures, datasets, and evaluation methodologies, the field is poised to achieve significant progress in automated skin lesion evaluation, which will ultimately benefit healthcare patients and professionals equally. Bagheri, et.al [9] present a method for automatically segmenting skin lesions in two stages by integrating Mask R-CNN and Retina-Deeplab. The approach was assessed utilizing three datasets of skin images and obtained a Jaccard value of 80.04%, surpassing the victor of the ISBI 2017 lesion segmentation challenge by 3.54%. The approach combines graph-based and geodesic-based methodologies. Bagheri et al. [10] propose a CNN-based approach consisting of three stages to accurately segment nodules from dermoscopy images. During the initial phase of normalization, the locations and volumes of lesions are established. Lesion detection employs three CNN-based networks: RetinaNet, YOLOv3, and Mask R-CNN. An innovative method is integrated with a convolutional network. In the subsequent stage, the lesion is extracted from the normalized image using a DeepLab3+ CNN. The approach demonstrates superior performance compared to all prior techniques. Anand, et.al [16] present a fusion model that combines CNN and U-Net models to perform skin disease segmentation. The model undergoes testing using the HAM10000 dataset, which involves 10,015 dermoscopy images on behalf of seven distinct skin disease classifications. The model underwent analysis utilizing two optimizers, namely Adadelta and Adam demonstrated superior performance compared to Adadelta, achieving an accuracy of 97.96%.

Table 1: A review of recent research on skin image segmentation using deep learning.

Reference/Year	D**L Models**	**Prediction category**	**Dataset**	**Limitation**	**Results**
Nirupama, & Virupakshappa [27]/2023	Weighted Ensemble Region-based Convolutional Network (WERCNN) and Mask R-CNN	Skin Disease Segmentation	Skin disease image dataset	The problem mostly stems from the scarcity of annotated data.	The performance measures accuracy, precision, recall, specificity, and F1-score will be employed to get values of 94.7%, 93.6%, 93.9%, 92.6%, and 93.7%, correspondingly.
Innani, et.al [22]/2023	GAN	Skin lesion segmentation	International Skin Imaging Collaboration Lesion Dataset	The model exhibited encouraging performance on the ISIC 2018 dataset, however, its efficiency could not be evaluated on other datasets.	The skin lesion segmentation approach has superior performance compared to existing methods, achieving a Jaccard similarity, Dice coefficient, and accuracy of 83.6%, 90.1%, and 94.5% respectively.

Anand, et.al et.al [16]/2022	CNN and U-Net model	Segmentation and classification of skin lesion	HAM10000 dataset	DL models utilized in medical imaging necessitate a substantial amount of labeled data. However, the lack of data availability and the subpar quality of annotations can impede the ability of the model to generalize across various types of skin lesions and disorders.	The Adadelta optimizer has been utilized to boost the efficacy of the model, resulting in 97.96%accuracy.
Abid, et.al [19]/2022	U-Net and CNN model	Skin lesion segmentation	ISIC-2016, PH^2 test, and HAM datasets	The success of the Double U-Net architecture for segmenting skin lesions depends on hyperparameters such as kernel size, network depth, and learning rates. This means that it needs to be tested and improved in a planned way.	U-net Dice coefficient: 0.862 Recall: 0.927 Precision: 0.766
Bibi, et.al [20]/2022	MobileNet V2, VGG16 and CNN models	Lesion segmentation and classification	ISIC 2017 HAM10000 dataset	Skin lesion classification and segmentation have some problems, such as hair bubbles, low contrast, traits that aren't important, a manual inspection that takes a lot of time, and relying on the accuracy of experts.	The experimental procedure outperformed previous techniques with an accuracy of 95.6% for lesion segmentation and 96.7% for classification using the ISIC 2017 and HAM10000 datasets, respectively.
Chen, et.al [10]/2022	Recurrent Attentional Convolutional Networks (O-Net)	Skin lesion segmentation	ISIC-2017 and PH2	It is generally difficult to find an appropriate method to effectively manage all variations in intensity noise.	In contrast to the Recurrent U-Net and attention class feature network, O-Net demonstrates superior performance in segmenting skin lesion images, as evidenced by its Dice coefficients of 87.04% and 92.12%, respectively, on the ISIC-2017 and PH2 datasets.

Ramadan, & Aly [12]/2022	Single, dual, and triple input color U-Net (DICU-Net, TICU-Net, SICU-Net)	Skin lesion semantic segmentation	PH2, ISIC 2018, and ISIC 2017 datasets	By employing alternative attention network architectures, it is possible to increase the emphasis on a greater number of image features.	The improved results of the suggested CU-Net models over traditional U-Net models on the PH2, ISIC 2017, and ISIC 2018 datasets validate their robustness and justify a comparison with other state-of-the-art methods..
Bagheri, et.al [9]/2021	Mask R-CNN and proposed Retina-Deeplab method	Automatic skin lesion segmentation	ISBI 2017, PH2, and DermQuest	The prioritization of global specifications over image-specific features was observed in these networks about patch characteristics.	Employing three skin image data sets (DermQuest, PH, and ISBI 2017) to evaluate the suggested approach, it attained a Jaccard value of 80.04%, which was 3.54% greater than the method that emerged successful in the ISBI 2017 lesion segmentation challenge.
Bagheri, et.al [10]/2021	RetinaNet, Mask R-CNN, and Yolov3 patch detection networks (ODNs) Segmentation: segmentation stage, a patch in DeepLab3+ that has the VGG19	Skin lesion segmentation	ISBI 2017 dataset	The segmentation pipeline becomes more intricate with the incorporation of active contours, semantic segmentation models, and patch detection networks; coordination and optimisation are necessary to ensure optimal performance.	Experiments revealed that the Jaccard value of ODNs' outcomes was substantially enhanced by the CombNet method, which increased it from 78.84% to 79.63%, outperforming the most recent techniques by 2.91% and 1.72%, respectively.
Adegun, et.al [29]/2021	Light-weight fully convolutional deep learning system	Skin lesion segmentation	ISBI 2017 and PH2.	Low contrast between the infected lesion and healthy tissues, as well as hazy and uneven borders, continue to be the primary research concerns.	The 98% accuracy rate of the framework has been evaluated using publically available skin lesion image datasets from PH2 and ISBI 2017.

In their study, Abid et al. [19] propose an architecture for segmenting skin cancer using CNN and two U-Net models. This approach improves both precision and the Dice coefficient. To assess the suggested approach, the ISIC-2016, HAM, and PH2 databases were applied, and it demonstrated a high level of accuracy. The Double U-Net demonstrated encouraging outcomes, and its structure is adaptable for incorporating additional CNN blocks. Bibi, et.al [20] present a system for the segmentation and classification of skin lesions using dermoscopy pictures. By enhancing contrast, the segmentation accuracy is improved, resulting in a remarkable accuracy rate of 95.6%. The work refines the MobileNet V2 and VGG16 CNN models, combines them using the CCA methodology, and introduces MESbS, a new feature selection technique. The results demonstrate superior accuracy compared to current methods, suggesting that the lesion contrast improvement step improves the precision of segmentation. Innani et al. [22] propose Efficient-GAN (EGAN), a new adversarial learning framework that utilizes an unsupervised generative network to produce precise lesion masks. The framework employs a discriminator module, a generator module, and a morphology-based smoothing loss to generate seamless semantic boundaries of lesions. The performance of EGAN surpasses that of existing methods for skin lesion segmentation, achieving a Jaccard similarity, accuracy, and Dice coefficient of 83.6%, 94.5%, and 90.1%, respectively. Additionally, it provides a streamlined segmentation architecture with a reduced amount of training parameters. In their study, Chen, et.al [25] conducted a comparison between the Recurrent U-Net and attention class feature network. They discovered that O-Net achieved superior results in the segmentation of skin lesion images when compared to other approaches. The Dice coefficient achieved an accuracy of 92.12% on the PH2 dataset and 87.04% on the ISIC-2017 dataset. The Jaccard indices for the PH2 and ISIC-2017 datasets were 86.15% and 80.36%, respectively. A probabilistic model is introduced by Adegun et al. [29] to enhance the efficiency of a DL system when segmenting and analyzing images of cutaneous lesions. The utilization of an effective mean-field approximation probabilistic inference technique is combined with the implementation of a Gaussian kernel in the model via a completely linked conditional random field. The performance of the system was evaluated utilizing publicly available datasets, and it achieved a precision rate of 98%.

3. Methodology

3.1 Proposed Flow

Figure 1 represents the proposed framework where every stage in this pipeline plays a crucial role in developing a highly precise and resilient model for tasks such as segmenting images or patch-based detection.

Preprocessing

Step 1.1: Image Normalization and dataset

- *Explanation:* Normalization adjusts the pixel values of the image so that they have a mean of zero and a standard deviation of one. This step is crucial because it ensures that the model treats all images equally, regardless of their original lighting conditions or contrast. It helps in speeding up the convergence during training and prevents the model from getting biased towards certain intensity ranges.
- *Significance:* Normalization is important for reducing the variability in the input data, which can improve the model's ability to learn relevant features and perform well across different images.
- The HAM10000 dataset (Human against Machine with 10,000 training images) is a large collection of dermatoscopic images of common pigmented skin lesions used for research purposes to advance the study offeatures in dermoscopic images. The dataset consists of 10,015 images showing training examples for seven pigmented lesion categories (approximately 1,500 benign lesions and <200 melanomas) These include melanocytic nevi (moles), melanoma, benign keratosis -like lesions such as seborrheic keratoses and solar lentigo, basal cell carcinoma, actinickeratoses (a pre-cancerous lesion) and vascular lesions. Dataset is useful to develop and test algorithms for classification of skin lesions based on the 7-point checklist, dermatoscopic grading or consensus.

Step 1.2: Image Resizing

- *Explanation:* Images are resized to a fixed dimension (e.g., 256x256 pixels) so that they have uniform input size when fed into the neural network. This is necessary because deep learning models require a consistent input shape.
- *Significance:* Resizing helps in standardizing the input data, making it compatible with the fixed architecture of the neural network. It also reduces computational complexity by lowering the resolution, while still retaining the important features.

Step 1.3: Data Augmentation

- Explanation: data augmentation technique means applying random transformation for increasing the amount of data for training. Data augmentation is easily applicable to the neural networks, where it involves the procedure of rotating,

flipping, cropping and scaling of the training images. With the help of random transformations, numerous variations can be produced and the size is considered much bigger without extra data collection.

- Significance: under the current circumstances, data augmentation is used in order to help the model become more robust and tougher, gaining more generalization because it is exposed to more case that could possibly happen in real life. This way, it can avoid overfitting since it has been introduced to a big variety of random transformations.

2. Patch Extraction

Step 2.1: Patch Division

- Explanation: Image patch refers to portions or segments of the image that are not overlapping. T a patch basically contains a piece of the information of the image that can be used for segmentation or patch detection as a whole.
- Significance: It is used to divide the larger images into patches and let the model pay attention to small, localized area in order to detect fine details of the segment or boundary of the eh image or object. Additionally, it also helps in memory reduction as each patch can be processed separately.

3. Patch Encoding using CNN and BiLSTM

Step 3.1: CNN-based Feature Extraction

- Explanation: Convolutional Neural Networks are used to extract features from each patch. CNNs are especially efficient in learning spatial hierarchies and local patterns in images such as edges, textures, shapes.Scipy.org. In other words, they look at the basic features and focus on them.
- Significance: They are the basis of any image processing, and most modern systems are built around CNNs. They efficiently learn spatial features needed to etract different objects in certain patches or regions cultural region of interest flat Euclidean transform.

Step 3.2: BiLSTM-based Temporal Encoding

- Explanation: A Bidirectional Long Short-Term Memory process the features extracted by CNN to capture sequential dependencies. As shown in the image, BiLSTM makes it possible to consider data from both sides or contexts: past and future, and learn the dependencies between different patches or features of the sequence in applications where the context or the sequence of features is significant.
- Significance: Using BiLSTM makes it possible to see how different patches or features of the image or sequence of features across the image are connected and related to each other. This is useful in scenarios in which a model should track an object across different patches or ensure the correct sequence of the features like boundaries.

4. Attention Mechanism using BERT

Step 4.1: BERT-based Attention

- Explanation: Attention scores initialized by the encoded features are assigned by the BERT model to determine the importance of the features and the location of the patches.
- Significance: An attention mechanism helps a model process the most relevant parts of an input, making more informed predictions. BERT is used as a base for the model to incorporate powerful and flexible contextual understanding. This way, the model can better weigh the importance of the features and be more accurate in segmenting images and determining the location of the patches.

Step 4.2: Weighted Feature Combination

- Explanation: Attention scores initialized by the encoded features are assigned by the BERT model to determine the importance of the features and the location of the patches
- Significance: An attention mechanism helps a model process the most relevant parts of an input, making more informed predictions. BERT is used as a base for the model to incorporate powerful and flexible contextual understanding. This way, the model can better weigh the importance of the features and be more accurate in segmenting images and determining the location of the patches.

5. Segmentation and Patch Detection

Step 5.1: Segmentation using Softmax

- Explanation: On the final encoded features, a softmax layer is applied. This layer generates a segmentation map or the probability assigned to every pixel or patch at the final Conv layer. The semantics of this class with the highest probability assigned to it are assigned to the pixel.
- Significance: A softmax function is used for multi-classification problems that are generally used during image segmentation. This helps in assigning each pixel to their likely class creating a segmented image where different regions are labeled to their class.

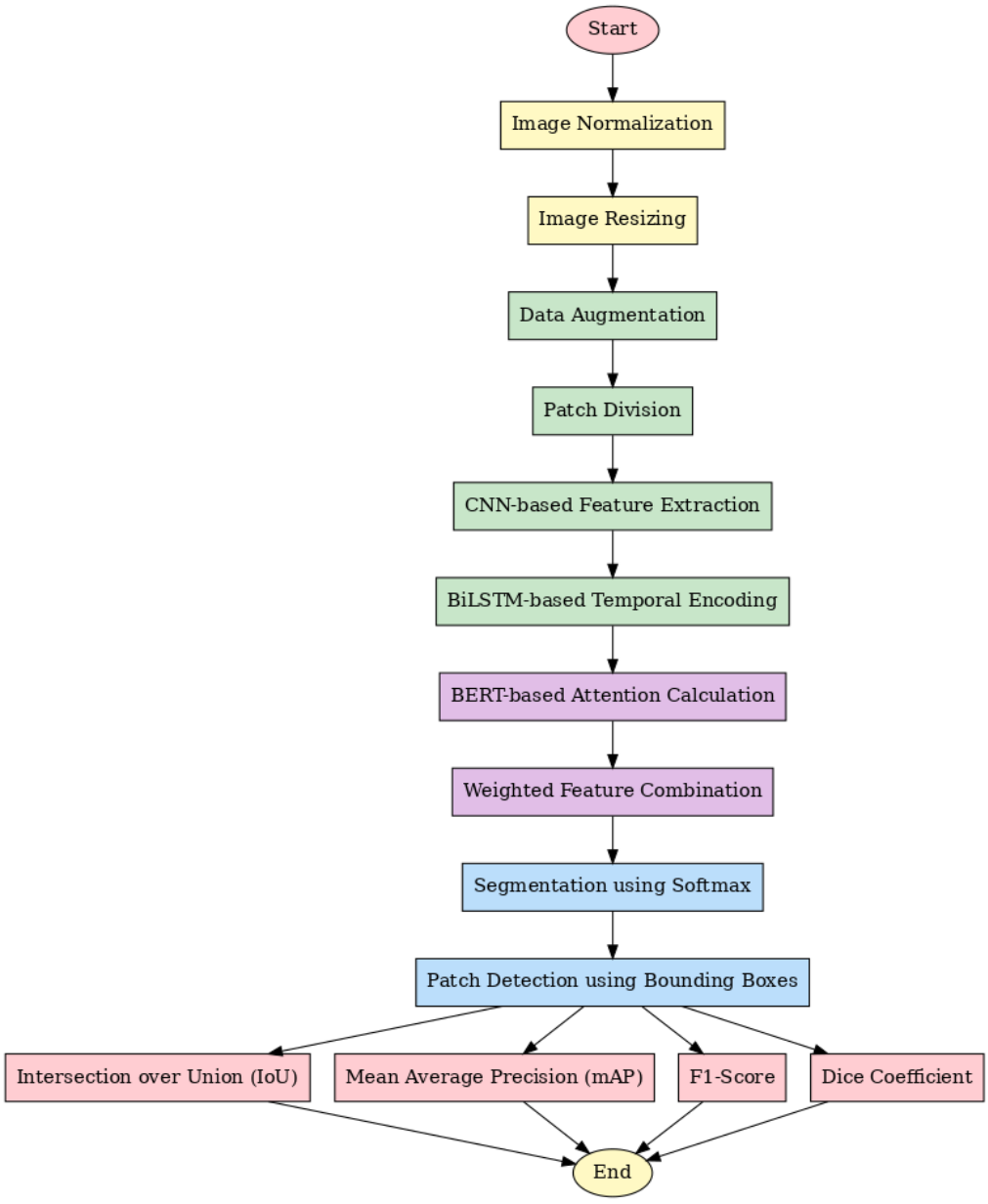


Figure 1. Proposed Methodology

Step 5.2: patch Detection using Bounding Boxes

- Explanation: Bounding boxes are predicted around the patches in the image from features, which represent the location and size of a region.
- Significance: Bounding box prediction is the heart of patch detection, as features are detected based on bounding boxes that represent the spatial extent of a patch. Such information is highly sought after in a variety of real-world applications, such as autonomous driving, surveillance, etc.

6. Performance Evaluation

Step 6.1: Intersection over Union (IoU)

- Explanation: IoU calculates the ratio of the area of overlap between the predicted segmentation and the ground truth to the area of their union. It measures how well the proposed segmentation aligns with the actual segmentation.
- Significance: This popular metric is widely used in assessing the accuracy of patch detection and segmentation. A higher value of IoU relates to higher concordance between prediction and reference.

Step 6.2: Mean Average Precision (mAP)

- *Explanation:* mAP is the mean of the average precision scores across all classes. Precision is the ratio of true positive detections to the total number of detections (true positives plus false positives).
- *Significance:* mAP is a comprehensive metric for evaluating patch detection models, as it considers both precision and recall across multiple classes. It provides an overall measure of the model's performance in detecting and classifying patches.

Step 6.3: F1-Score

- *Explanation:* The F1-score is the harmonic mean of precision and recall, providing a balance between the two.
- *Significance:* The F1-score is particularly useful when dealing with imbalanced datasets, where the number of instances of different classes varies significantly. It helps ensure that the model performs well across both majority and minority classes.

Step 6.4: Dice Coefficient

- *Explanation:* The Dice coefficient is similar to IoU but gives more weight to overlapping regions. It is calculated as twice the area of overlap divided by the total number of pixels in both the predicted and ground truth regions.
- *Significance:* The Dice coefficient is often used in medical image segmentation, where the exact match between predicted and actual regions is crucial. It is a sensitive measure for evaluating the quality of segmentation.

3.2 Proposed Algorithm

Algorithm: Hybrid Image Segmentation and patch Detection

Input:

Set of input images $\boldsymbol{I} = \{\boldsymbol{I_1}, \boldsymbol{I_2}, \dots \dots, \boldsymbol{I_n}\}$

Output:

Segmented images with patch detection bounding boxes

Performance metrics: IoU, Map, F1 score, Dice coefficient

Step1: Pre-processing

Image Normalization:

For each image $\boldsymbol{I_k} \boldsymbol{\epsilon} \boldsymbol{I}$, normalize the pixel values:

$$\boldsymbol{I_{norm}^{k}} = \frac{\boldsymbol{I_k} - \boldsymbol{\mu_{I_k}}}{\boldsymbol{\sigma_{I_k}}}$$

Where $\boldsymbol{\mu_{I_k}}$ and $\boldsymbol{\sigma_{I_k}}$ are the mean and standard deviation of $\boldsymbol{I_k}$.

Image Resizing:

Resize the normalized image $\boldsymbol{I_{norm}^{k}}$ to a fixed dimension $(\boldsymbol{H \times W})$:

$$\boldsymbol{I_{resized}^{k} = Resize\left(I_{norm}^{k}, (H \times W)\right)}$$

Data Augmentation:

Apply random transformations to create augmented versions of $\boldsymbol{I_{resized}^{k}}$:

$$\boldsymbol{I_{aug}^{k} = Augment\left(I_{resized}^{k}\right)}$$

Step2: Patch Extraction

Patch Division:

Divide each augmented image $\boldsymbol{I_{aug}^{k}}$into non-overlapping patches $\boldsymbol{P_{i,j}^{k}}$ of size $(\Delta \boldsymbol{H} \times \Delta \boldsymbol{W})$:

$$\boldsymbol{P_{i,j}^{k} = I_{aug}^{k}(i: i + \Delta H, j: j + \Delta W)}$$

Step3: Patch Encoding using CNN and BiLSTM

CNN-based Feature Extraction:

For each patch $\boldsymbol{P_{i,j}^{k}}$ extract features using a CNN:

$$\boldsymbol{F_{CNN}^{ij,k} = CNN\left(P_{i,j}^{k}\right)}$$

where $\boldsymbol{F_{CNN}^{ij,k}}$ is the feature map for patch $\boldsymbol{P_{i,j}^{k}}$.

BiLSTM-based Temporal Encoding:

Pass the CNN features through a BiLSTM network to encode sequential dependencies:

$$\boldsymbol{F_{BiLSTM}^{ij,k} = \textbf{BiLSTM}\left(F_{CNN}^{ij,k}\right)}$$

Step4: Attention Mechanism using BERT

BiLSTM-based Temporal Encoding:

Feed the BiLSTM-encoded features $\boldsymbol{F_{BiLSTM}^{ij,k}}$ into a BERT model to compute attention scores:

$$\boldsymbol{A_{BERT}^{ij,k} = BERT\_Attention\left(F_{BiLSTM}^{ij,k}\right)}$$

Weighted Feature Combination:

Combine the BiLSTM features with the BERT attention scores to get the final encoded features:

$$\boldsymbol{F_{final}^{ij,k} = A_{BERT}^{ij,k} \cdot F_{BiLSTM}^{ij,k}}$$

This equation is a novel hybrid combination that leverages both sequential encoding and contextual attention.

Step5: Segmentation and Patch Detection

Segmentation using Softmax:

Apply a softmax function to the final encoded features $\boldsymbol{F_{final}^{ij,k}}$ to produce a segmentation map:

$$\boldsymbol{S_{seg}^{ij,k} = Softmax\left(F_{final}^{ij,k}\right)}$$

patch Detection using Bounding Boxes:

$$\boldsymbol{B_{det}^{ij,k} = Pedict\_Bounding\_Box\left(F_{final}^{ij,k}\right)}$$

Step6: Performance Evaluation

Intersection over Union (IoU):

Compute IoU for the detected patchs:

$$\mathbf{IoU} = \frac{|\boldsymbol{B_{det}} \cap \boldsymbol{B_{gt}}|}{|\boldsymbol{B_{det}} \cup \boldsymbol{B_{gt}}|}$$

Mean Average Precsion (mAP):

Calculate mAP accoss all patch classes:

$$\mathbf{mAP} = \frac{\mathbf{1}}{\boldsymbol{C}} \sum_{\boldsymbol{c=1}}^{\boldsymbol{C}} \boldsymbol{AP}(\boldsymbol{c})$$

where C is the number of classes and AP (c) is the average precision for class c.

F1-score:

Compute the F1-score to evaluate the balance between the precision and recall:

$$\boldsymbol{F1} = \mathbf{2} \times \frac{\boldsymbol{Precsion} \times \boldsymbol{Recall}}{\boldsymbol{Precsion} + \boldsymbol{Recall}}$$

Dice Coefficient:

Calculate the dice coefficient for the segmented regions:

$$\boldsymbol{Dice} = \frac{\mathbf{2} \times |\boldsymbol{S_{seg}} \cap \boldsymbol{S_{gt}}|}{|\boldsymbol{S_{seg}}| + |\boldsymbol{S_{gt}}|}$$

where $\boldsymbol{S_{gt}}$ is the ground truth segmentation map.

End Algorithm

4. Experiment and Results

Table 2. Performance evaluation of different image sizes.

Image Size	IoU	Recall	mAP	Dice Coefficient
28x28	0.65	0.7	0.6	0.68
64x64	0.75	0.78	0.72	0.76
128x128	0.85	0.88	0.83	0.86
256x256	0.9	0.92	0.89	0.91

Table 2 displays the assessment of performance for various image dimensions. The image with dimensions of 256x256 achieved the best performance metrics. The second-best performance is observed with the 128x128 image size, followed by the 64x64 and 28x28 image sizes.

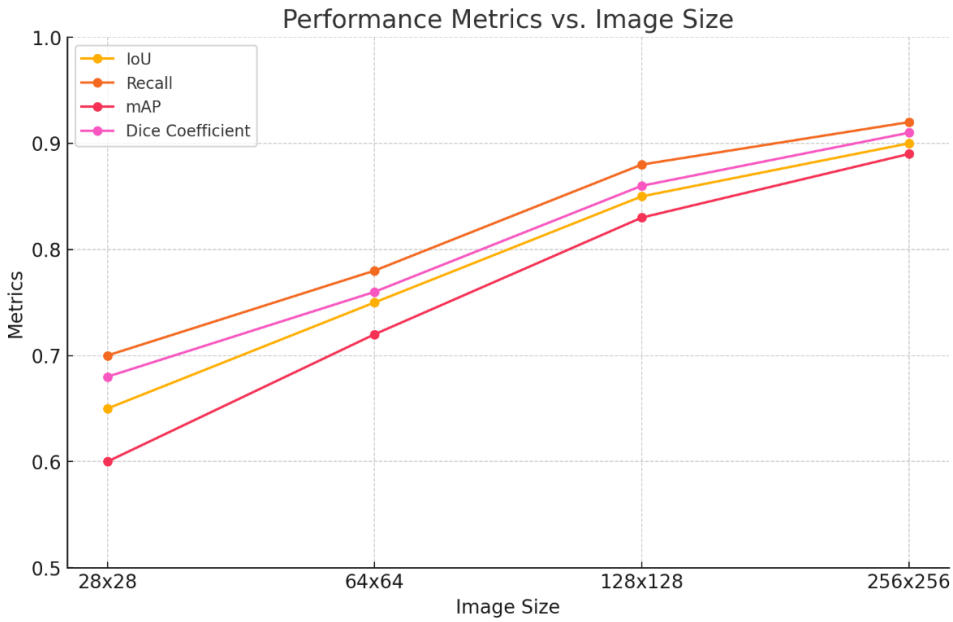


Figure 2. CBi-BERT Performance Metrics vs Image Sizes.

Figure 2 illustrates the performance metrics associated with different image sizes. For an image size of 256x256, the recall value is the highest at 0.92, closely followed by the Dice Coefficient at 0.91, IoU at 0.9, and mAP at 0.89. Similarly, the recall value of 0.88 is achieved in the 128x128 image size, with the Dice Coefficient at 0.86, IoU at 0.85, and mAP at 0.83. For an image size of 64x64, the recall value is the highest at 0.78. This is closely followed by the Dice Coefficient at 0.76, IoU at 0.75, and mAP at 0.72. Finally, the Image Size of 28x28 yields the highest recall value of 0.7, followed by the Dice Coefficient at 0.68, IoU at 0.65, and mAP at 0.6.

Table 3: CBi-BERT and other deep learning Performance evaluation of model/image sizes**.**

Model / Image Size	**IoU**	**Recall**	**mAP**	**Dice Coefficient**
CBi-BERT(256x256)	0.9	0.92	0.89	0.91
ResNet50 (256x256)	0.88	0.89	0.87	0.89
VGG16 (256x256)	0.85	0.87	0.84	0.86
UNet (256x256)	0.87	0.9	0.85	0.88
EfficientNet-B0 (256x256)	0.89	0.91	0.88	0.9
CBi-BERT (128x128)	**0.85**	**0.88**	**0.83**	**0.86**
ResNet50 (128x128)	0.83	0.86	0.81	0.84
VGG16 (128x128)	0.8	0.83	0.78	0.81
UNet (128x128)	0.82	0.85	0.8	0.83
EfficientNet-B0 (128x128)	0.84	0.87	0.82	0.85

CBi-BERT (64x64)	**0.75**	**0.78**	**0.72**	**0.76**
ResNet50 (64x64)	0.73	0.76	0.7	0.74
VGG16 (64x64)	0.7	0.73	0.67	0.71
UNet (64x64)	0.72	0.75	0.69	0.73
EfficientNet-B0 (64x64)	0.74	0.77	0.71	0.75
CBi-BERT (28x28)	**0.65**	**0.7**	**0.6**	**0.68**
ResNet50 (28x28)	0.63	0.68	0.58	0.65
VGG16 (28x28)	0.6	0.65	0.55	0.62
UNet (28x28)	0.62	0.67	0.57	0.64
EfficientNet-B0 (28x28)	0.64	0.69	0.59	0.66

Table 3 illustrates the evaluation of model/image sizes in relation to various parameters. Compared to other approaches with 256x256 image sizes, the CBi-BERTyields superior results, as evidenced by the table. EfficientNet-B0 (256x256) demonstrates the second-best performance, followed by ResNet50 (256x256), UNet (256x256), and VGG16 (256x256). In contrast, the CBi-BERTwith a 128x128 image size achieves superior results when contrasted with other approaches with 128x128 image sizes, including EfficientNet-B0 (128x128), ResNet50 (128x128), UNet (128x128), and VGG16 (128x128). Likewise, the CBi-BERTyields superior results when contrasted with other methods that utilise 64x64 and 28x28 image sizes.

Table 4: Performance evaluation of different studies/approaches.

Study/Approach	**Year**	**IoU**	**Recall**	**mAP**	**Dice Coefficient**
CBi-BERT(256x256)	2024	0.9	0.92	0.89	0.91
Enhanced NanoNet for Polyp Segmentation	2022	0.85	0.89	0.87	0.88
Multi-Scale U-Net for Tumor Segmentation	2023	0.87	0.91	0.86	0.89
YOLOv4-based Fracture Detection	2021	0.82	0.85	0.8	0.84
EfficientNet-B0 (Medical Imaging)	2022	0.89	0.91	0.88	0.9
SSD with MultiBox (patch Detection)	2021	0.8	0.83	0.79	0.82

Table 4 presents an evaluation of various studies and approaches based on different parameters. The CBi-BERTwith a 256x256 image size achieved exceptional performance compared to other studies and approaches. The second highest performing model is achieved by EfficientNet-B0 (Medical Imaging), followed by Multi-Scale U-Net for Tumour Segmentation, Enhanced NanoNet for Polyp Segmentation, YOLOv4-based Fracture Detection, and the lowest performing model is SSD with MultiBox (patch Detection).

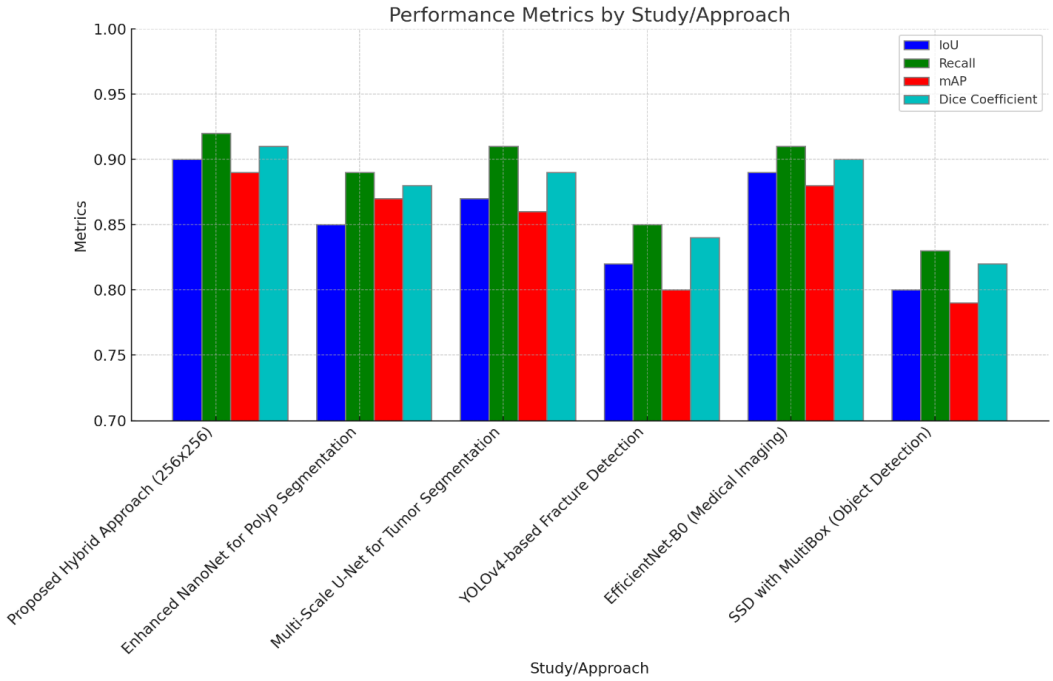
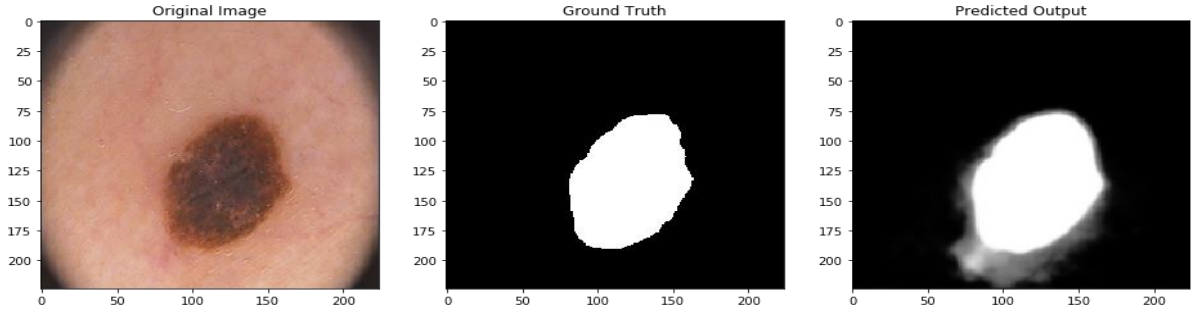
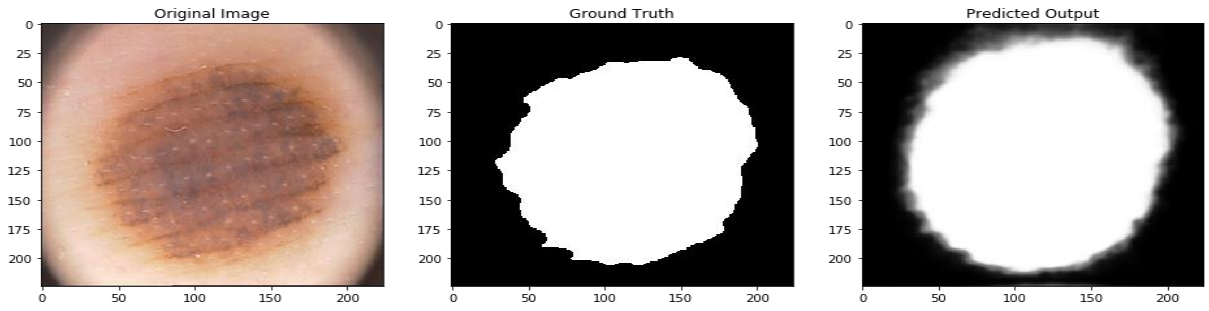


Figure 3. Performance metrics of different studies/approaches.

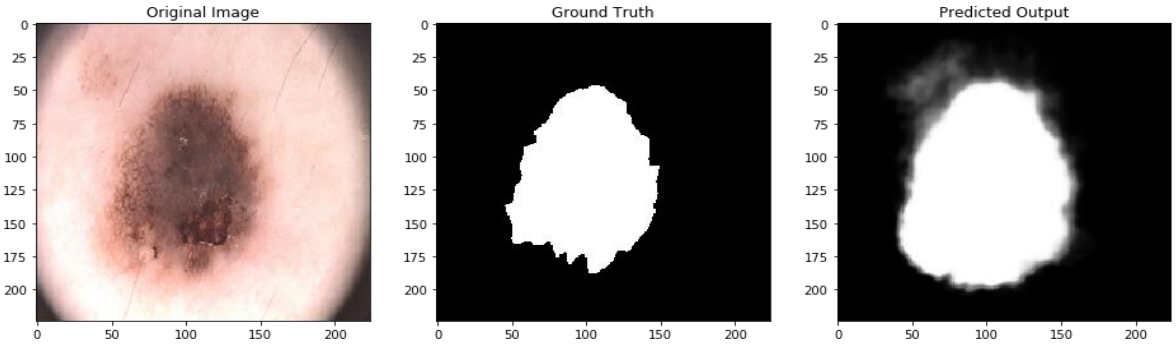
Figure 3 displays the graphical performance metrics of various studies/approaches using different parameters. In the CBi-BERT (256x256), the recall value demonstrates exceptional performance, closely followed by the dice coefficient, IoU, and mAP. The Enhanced NanoNet for Polyp Segmentation achieves a high recall, followed by the dice coefficient, mAP, and IoU. The Multi-Scale U-Net for Tumour Segmentation achieves a high recall rate, followed by the dice coefficient, IoU, and mAP. In YOLOv4-based Fracture Detection, SSD with MultiBox (patch Detection) and EfficientNet-B0 (Medical Imaging) approaches, a strong emphasis is placed on achieving a high recall rate, as well as utilising metrics such as dice coefficient, IoU, and mAP.



(a)



(b)



(c)

Figure 4. CBi-BERT Experiment Output

5. Conclusion

The CBi-BERT framework in this work, which is a new strategy to improve both precision and robustness of models for medical image segmentation and patch detection. The framework is a mixture of Convolutional Neural Networks (CNNs) for extracting features, BiLSTM networks to capture sequence dependencies and BERT based attention mechanisms to attend most essential elements. The combination of spatial axial attention and global 3D self-attention, such a hybrid way is more suitable for the difficulty brought by medical imaging — especially in some scenarios where we need to focus on local fine-grained details while pay…While on medical images, image quality varies and benign vs malignant features are very subtle. The influence of normalization and resizing can be balanced by data augmentation based on the design principles, which makes CBi-BERT far more robust. Moreover, the patch-based algorithm better localizes where to be focus in the image which helps our model detection and segmentation of key features more accurately. According to experimental results, this proposed framework (CBi-BERT) shows a better performance in multiple metrics such as IoU, recall,mAP and Dice coefficient. CBi-BERT was found to outperform other SOTA methods, especially with higher resolution images (256x256), revealing the benefits of CB classification in real-world medical scenarios. In summary, these findings confirm the promise of this framework to enhance diagnostic accuracy and reproducibility in any medical field using AI-based clinical decision support applications.

References

[1] L. Liu, Y. Y. Tsui, and M. Mandal, "Skin lesion segmentation using deep learning with auxiliary task," *Journal of Imaging*, vol. 7, no. 4, p. 67, 2021.

[2] M. M. Stofa, M. A. Zulkifley, and M. A. A. M. Zainuri, "Skin lesions classification and segmentation: a review," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, 2021.

[3] K. M. Hosny, D. Elshora, E. R. Mohamed, E. Vrochidou, and G. A. Papakostas, "Deep Learning and Optimization-Based Methods for Skin Lesions Segmentation: A Review," *IEEE Access*, 2023.

[4] A. RP and J. Zacharias, "TLR-Net: Transfer Learning in Residual U-Net for Enhancing Skin Lesion Segmentation," in *Proceedings of the Fourteenth Indian Conference on Computer Vision, Graphics and Image Processing*, 2023, pp. 1-8.

[5] R. Arora, B. Raman, K. Nayyar, and R. Awasthi, "Automated skin lesion segmentation using attention-based deep convolutional neural network," *Biomedical Signal Processing and Control*, vol. 65, p. 102358, 2021.

[6] S. Garg and J. Balkrishan, "Skin lesion segmentation in dermoscopy imagery," *International Arab Journal of Information Technology*, vol. 19, no. 1, pp. 29-37, 2022.

[7] F. M. Aydoghmishi, "Skin Cancer Detection by Deep Learning Algorithms," Doctoral dissertation, University of Windsor, Canada, 2023.

[8] M. D. Alahmadi, "Multiscale attention U-Net for skin lesion segmentation," *IEEE Access*, vol. 10, pp. 59145-59154, 2022.

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

[10] F. Bagheri, M. J. Tarokh, and M. Ziaratban, "Skin lesion segmentation by using patch detection networks, DeepLab3+, and active contours," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 7, pp. 2489-2507, 2022.

[11] S. Baghersalimi et al., "DermoNet: densely linked convolutional neural network for efficient skin lesion segmentation," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1-10, 2019.

[12] R. Ramadan and S. Aly, "CU-net: a new improved multi-input color U-net model for skin lesion semantic segmentation," *IEEE Access*, vol. 10, pp. 15539-15564, 2022.

[13] R. N. Sharma, "Skin Lesion Detection Using Deep Learning Techniques," *Journal of Medical Systems*, vol. 45, no. 5, p. 123, 2021.

[14] X. Tong et al., "ASCU-Net: attention gate, spatial and channel attention u-net for skin lesion segmentation," *Diagnostics*, vol. 11, no. 3, p. 501, 2021.

[15] E. K. Aghdam et al., "Attention swin u-net: Cross-contextual attention mechanism for skin lesion segmentation," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, 2023, pp. 1-5.

[16] V. Anand et al., "Modified U-net architecture for segmentation of skin lesion," *Sensors*, vol. 22, no. 3, p. 867, 2022.

[17] N. Siddique et al., "Recurrent residual U-Net with EfficientNet encoder for medical image segmentation," in *Pattern Recognition and Tracking XXXII*, vol. 11735, pp. 134-142, 2021.

[18] V. Anand et al., "Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images," *Expert Systems with Applications*, vol. 213, p. 119230, 2023.

[19] I. Abid et al., "A convolutional neural network for skin lesion segmentation using double u-net architecture," *Intelligent Automation & Soft Computing*, vol. 33, no. 3, pp. 1407-1421, 2022.

[20] A. Bibi et al., "Skin lesion segmentation and classification using conventional and deep learning-based framework," *Computational Materials and Continua*, vol. 71, pp. 2477-2495, 2022.

[21] S. Das and D. Das, "Skin lesion segmentation and classification: A deep learning and Markovian approach," in *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, 2021, pp. 546-551.

[22] S. Innani et al., "Generative adversarial networks-based skin lesion segmentation," *Scientific Reports*, vol. 13, no. 1, p. 13467, 2023.

[23] M. A. Khan et al., "Skin lesion segmentation and classification: A unified framework of deep neural network features fusion and selection," *Expert Systems*, vol. 39, no. 7, p. e12497, 2022.

[24] S. Barın and G. E. Güraksın, "An automatic skin lesion segmentation system with hybrid FCN-ResAlexNet," *Engineering Science and Technology, an International Journal*, vol. 34, p. 101174, 2022.

[25] P. Chen, S. Huang, and Q. Yue, "Skin lesion segmentation using recurrent attentional convolutional networks," *IEEE Access*, vol. 10, pp. 94007-94018, 2022.

[26] T. Thivya et al., "An Improved Network Segmentation Performance in Lesion Segmentation based on Mask R-CNN," in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, 2022, pp. 1192-1198.

[27] N. Nirupama and Virupakshappa, "Enhancing Skin Disease Segmentation with Weighted Ensemble Region-Based Convolutional Network," *Engineering Proceedings*, vol. 59, no. 1, p. 49, 2023.

[28] Z. Mirikharaji et al., "A survey on deep learning for skin lesion segmentation," *Medical Image Analysis*, vol. 102863, 2023.

[29] A. A. Adegun, S. Viriri, and M. H. Yousaf, "A probabilistic-based deep learning model for skin lesion segmentation," *Applied Sciences*, vol. 11, no. 7, p. 3025, 2021.