Proceedings of the 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA-2025)
DVD Part Number: CFP25J88-DVD; ISBN: 979-8-3315-9928-7

Emerging Semi-Supervised Skin Lesion Segmentation using Unlabelled Samples with GAN and Transformer Representations

M. Megala
Research Scholar
Department of Computer Science and Information Technology
Vels Institute of Science, Technology & Advanced Studies (VISTAS)
Chennai, Tamil Nadu, India
megalasenthil12@gmail.com

T.R. Nisha Dayana
Assistant Professor
Department of Computer Science and Information Technology
Vels Institute of Science, Technology & Advanced Studies (VISTAS)
Chennai, Tamil Nadu, India
nisha.dayana1984@gmail.com

Abstract: Early cancer detection and diagnosis depend on accurate and automated skin lesion segmentation helping dermatologists in their work. On the other hand, deep learning models find great challenge in the absence of large, annotated datasets in the medical field. These problems consist in overfitting and limited generalisation. Manual labelling calls for knowledge, time, and money; hence, depending on very large, annotated datasets is generally not practical for medical applications. We introduce a new semi-supervised method using unlabelled samples including generative adversarial networks (GANs). This approach is proposed to transcend the already mentioned limitations. The method makes advantage of a two-stage training structure. First phase of the supervised learning process uses annotated data to learn semantic segmentation maps. After that, the unsupervised phase uses Structured Prediction-based Deep Reinforcement Learning (SP-DRL) to improve generalisation, which lowering the dependence on annotations. Strong data-driven features have been proposed to be derived from a surrogate task using convolutional and Transformer-based representations, which reducing the demand for annotations. The proposed method was evaluated using three benchmark datasets: PH2, ISIC 2017, and ISIC 2018. The segmentation performance shows to be better with increasing Dice Similarity Coefficient (DSC) scores of 89.2% (ISIC 2018), 87.4% (ISIC 2017), and 91.1% (PH2) respectively. Further, the model shows better generalisation and resistance against domain changes, which surpassing conventional approaches under total supervision. This semi-supervised method closes the difference between the availability of annotated medical datasets and those without labels, which enabling the useful adoption of automated lesion segmentation systems.

Keywords: Skin lesion segmentation, Semi-supervised learning, GAN, Transformer representations, Medical image analysis

I INTRODUCTION

Melanoma is among the most aggressive forms of skin cancer among all the others; it also is one of the most often occurring kind of cancer. Since it significantly affects the outcome of treatment, increasing survival rates depends on early identification of skin lesions. Early skin lesion diagnosis is essential. Based on automated skin lesion segmentation, many different diagnostic systems ensure exact localisation of lesions in dermoscopic images. The capacity of these systems to correctly classify lesions and extract features is mostly due to the use of advanced deep learning approaches. Still, the limited availability of defined medical datasets limits the efficiency of deep learning models still. Not only time-consuming but also costly; the acquisition of annotations for large medical datasets calls for the participation of specialists and makes advantage of a significant volume of resources [1-3].

Usually depending on completely supervised learning and requiring the use of extensive annotated datasets, traditional used skin lesion segmentation techniques rely on This depending results in the following difficulties:

- Dermatologists must label images; thus, the operation is expensive and labour-intensive [4]. Few medical datasets with sufficient annotations thus exist.
- Models trained on small datasets with frequent annotations overfit, which limit their ability to generalise to data they have not before coming across [5].
- Factors including imaging conditions, skin types, and lesion appearances often show a significant degree of variability in skin lesion datasets, which complicate the process of generalising models [6].
- Without semi-supervised approaches, huge volumes of unlabelled medical data remain underused in the process of training deep models [7].

Deep learning models for skin lesion segmentation depend on annotated datasets, hence scalability and generalisation are seriously challenged. Conventional methods cannot effectively use unlabelled data, thus models with limited scope and performance follow from their result. Moreover, the lack of strong means to learn from unlabelled samples aggravates overfitting problems and lowers the model's applicability to the real-world conditions. Overcoming these limitations will be facilitated by the development of new approaches including unlabelled data into the training process maintaining a high degree of segmentation accuracy [8-12]. The objective of the proposed work involves the following:

- The aim is to efficiently build a semi-supervised framework for skin lesion segmentation using both annotated and unlabelled data.
- The goal is to increase the generalisation capacities of the model over a spectrum of datasets so improving the segmentation model and reducing the overfitting probability.

The proposed method combines Transformer-based surrogate tasks with Generative Adversarial Networks (GANs) to include unlabelled data into the training pipeline. Integrated into the approach is structured prediction-based Deep Reinforcement Learning (SP-DRL), which maximises segmentation without depending on extensive annotations unlike conventional fully supervised methods. New paradigm for the segmentation of medical images has been developed by means of the combination of supervised and unsupervised techniques discussed here. The contribution of the proposed work involves the following:

- Unsupervised learning aiming at improving model generalisation and supervised learning for semantic segmentation maps generates a hybrid training approach.
- By using Generative Adversarial Networks, one can use unlabelled samples, which improve the learning process of the model with a spectrum of data distributions reflecting the situation.
- A surrogate task is presented to extract strong, data-driven features in minimum reliance on annotations using Transformer and convolutional representations.

II RELATED WORKS

Xie et al. [13] a new semi-supervised learning framework based on a "mean learner" scheme. Their method enhances the learning process by projecting the confidence of the model on the actual class probability using a confidence module. Consistent targets the learner model discovers under this configuration come from this source. Consistent material inclusion into pseudo-labels enhances segmentation process. Their experiments on the ISIC skin lesion segmentation dataset revealed that this method exceeded other semi-supervised methods considered to be state-of- the-art.

Emphasising the creation of denoised pseudo-labels, Zhang et al. [14] suggested a dynamic prototypical feature representation learning framework. Their developed method efficiently removes doubtful components from the first pseudo-labels and offers a means for the learning of feature representations. Moreover shown were memory relations learning method to improve the global feature representation and a prototype-based confidence-aware contrastive learning method to strengthen the local feature structure. Based on long-running studies on skin lesion segmentation datasets, their method is superior than others SSL methods.

Alahmadi et al. [15] focused on the usage of unlabelled samples into the training process using a two-stage semi-supervised network. First stage addresses semantic segmentation, which is achieved by a supervised approach. On the other hand, the second stage is focused on unsupervised methods housed inside the encoder module. Their network learns data-driven features from images themselves by means of a surrogate task developed atop convolutional and Transformer representations. Their approach suggests a feasible path for skin lesion segmentation since it shows better Dice scores—for example, a 2% increase in the ISIC 2017 dataset. They reached this by including unsupervised samples into their method.

Based on the Simple Linear Iterative Clustering (SLIC) approach, Dos Santos et al. [16] proposed a semi-automatic segmentation method grounded on the building of superpixels. After that, they generated a descriptor able to independently express every superpixel by means of a grey-level co-occurrence matrix combined with Tamura texture features. Lesion from background areas is separated using a seeded fuzzy C-means clustering method with semi-supervised approach. Especially, their approach achieved a remarkable average segmentation accuracy of 96.78%, above the values of other methods mentioned in the literature.

To achieve skin lesion segmentation, the authors Shen et al. [17] built a cross-supervised network aware of confidence. Their method consists of two separate segmentation networks running in parallel; one network acts as a pseudo-label for the output of the other network. The networks filter low-confidence areas using confidence maps; a consistency loss is then used to optimise the unlabelled data contribution. One can reach this by means of cross-entropy between pseudo-labels and predictions. To exhibit state-of- the-art performance on the ISIC 2017 dataset, they proposed an attention module that helps to capture fine-grained features.

Nouboukpo et al. [18] presented the Multiscale Spatial Consistency Network (MSCNet), so optimising the limited labelled data that is at hand using both local and global spatial consistency. MSCNet guarantees that predictions match the related spatial context by means of a single encoder-decoder architecture augmented by a spatially limited mixture model (SCMM). Moreover exploited by the model to capture fine-grained lesion details are hierarchical superpixel structures. Especially in the segmentation of complex lesions, pseudo-labels—which enable the inclusion of unlabelled data—improve segmentation performance much more than other SSL methods can offer.

Aiming for semi-supervised skin lesion segmentation, Dzieniszewska et al. [19] presented a self-training approach anchored on a noisy student approach. Under their method, pseudo-labels are generated after a learner model is trained on labelled data. These pseudo-labels then support a student model's learning process. Good use of unlabelled data made feasible by this self-training approach improves segmentation computation performance. On the ISIC 2018 dataset, their method obtained a mean intersection-over-union (mIoU) of 88.0% by means of DeepLabV3 architecture. This demonstrates appreciable performance increases even from a small volume of labelled data.

Feng et al. [20], medical image segmentation was proposed for a GAN inversion-based semi-supervised learning framework known as InvSSL. The framework generates variation samples from labelled training images using GAN inversion. New generated data shows a distribution similar to the original data. This approach assures a great degree of consistency between pixels by means of a multi-level dense contrastive learning mechanism, so improving the accuracy of segmentation. Their method provides a viable path for the use of generative models in semi-supervised learning since it outperformed other SSL methods on lung and skin lesion segmentation datasets.

Even with the great progress made by the semi-supervised skin lesion segmentation methods now in use, several research gaps still exist. While pseudo-labeling and self-training have advanced techniques' capacity to precisely predict outcomes, label noise and data imbalance remain challenges to most approaches' capacity to depend mostly on a small set of labelled data. Many approaches also concentrate on enhancing the segmentation of individual lesions; but, for some reason they might not be able to generally apply to highly varied or complicated forms of lesions. Further, the implementation of innovative techniques such as domain adaptation or few-shot learning may improve the model even more, particularly in clinical settings where either broad annotated dataset are either rare or inaccessible.

III PROPOSED METHOD

Transformer-based surrogate tasks and GAN are used in proposed semi-supervised framework for skin lesion segmentation. This framework includes paradigms of both supervised and unsupervised learning. One can classify the operation into two main groups. Semantic segmentation maps of a CNN are taught using an annotated set of images in the supervised stage of the evaluation process. By means of Structured Prediction-based Deep Reinforcement Learning (SP-DRL), unlabelled samples are used in the unsupervised stage to improve generalisation and resilience. High-dimensional, data-driven features are extracted from transformer representations by means of a surrogate task. This helps the model to identify minute characteristics in skin lesions. Through effective augmentation of the training set, GANs let one generate reasonable lesion patterns from unlabelled data. These components in figure 1 help to create a strong model that can efficiently segment lesions over a large spectrum of datasets and requires just a minimum of annotation resources.

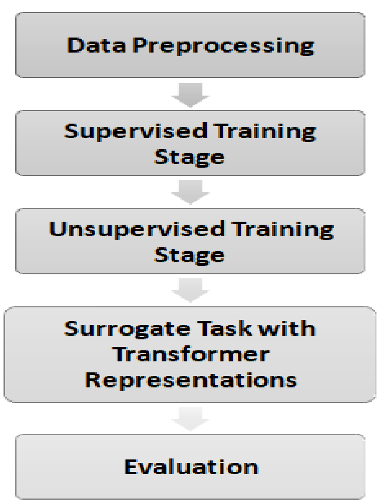

Figure 1: Proposed Process

A. Preprocessing

Preprocessing is a critical first step in preparing the dataset for the proposed semi-supervised skin lesion segmentation method. It ensures that the input images are suitable for training deep learning models, enabling efficient learning and better performance.

Normalization : Normalization is applied to scale the pixel values of images to a standard range, usually [0, 1], or to a range with zero mean and unit variance. This helps in reducing training time and allows the model to converge more quickly. It is essential because neural networks often perform better when input features are on similar scales.

For an image I with pixel values in the range [0, 255], normalization can be performed using:

$$I_{norm} = \frac{I - \mu}{\sigma}$$

where, I_{norm} is the normalized image, I is the original pixel value, μ is the mean pixel value of the image, σ is the standard deviation of the pixel values in the image. Alternatively, for simpler normalization to the range [0, 1], the formula would be:

$$I_{norm} = \frac{I}{255}$$

This ensures all pixel values are within a consistent and manageable range for the network.

B. Data Augmentation

Data augmentation is applied to artificially expand the training dataset by generating modified versions of the original images. This step is especially useful when there are limited annotated data, as it helps prevent overfitting and increases the model’s robustness. Typical augmentation techniques include rotation, flipping, scaling, cropping, and color jittering.

Let I be the input image. For a rotation augmentation, the image is rotated by an angle θ\thetaθ, and the new image I_{aug} is:

$$I_{aug} = R_{\theta}(I)$$

Where R_{θ} represents the rotation operation by angle θ.

Similarly, flipping and scaling operations can be represented as:

$$I_{flip} = F(I), \quad I_{scale} = S(I, \lambda)$$

Where F is the flipping operation, and S is the scaling operation with a scaling factor λ.

Image Resizing: Since deep learning models typically require input images of a consistent size, resizing is an essential step. Resizing involves transforming the images to a fixed resolution, say H×W, which is suitable for the network architecture. If the input image I has dimensions h×w, resizing to H×W is performed as:

$$I' = r\left(I, H, W\right)$$

This ensures all images fed into the model are of the same dimensions, avoiding errors during training and inference.

C. Edge Detection

For some segmentation tasks, applying edge detection techniques can enhance the boundary of skin lesions, improving the model’s ability to distinguish between the lesion and surrounding tissue. One common method is the Sobel operator, which computes the gradient magnitude of an image to highlight edges:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I,$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I$$

Where G_x and G_y are the gradients in the x and y directions, and ∗*∗ represents the convolution operation. The magnitude of the gradient at each pixel can be calculated as:

$$G = \sqrt{G_x^2 + G_y^2}$$

This highlights the edges of lesions, making them more distinguishable for the segmentation task.

D. Supervised Training

Phase of supervised training aims to train the model using annotated data, images with ground truth segmentation labels that correspond one another. Aim is to learn the mapping between the individual input images and the several segmentation masks. Usually training a Convolutional Neural Network (CNN), able to directly learn semantic segmentation maps from labelled data during this phase

Loss Function: A loss function measuring the variation between the ground truth segmentation as projected and actual output forms a guiding concept for the training process. In skin lesion segmentation, the loss function typically motivates the network to forecast accurate pixel-level segmentation by combining cross-entropy loss with dice loss. This is done thus to generate the desired results. Applied widely in the field of medical image segmentation, the Dice Similarity Coefficient (DSC) is sensitive to even the smallest lesions. It calculated like this:

$$D(P,G) = \frac{2 \cdot |P \cap G|}{|P| + |G|}$$

Where: P is the predicted segmentation mask, G is the ground truth mask. Dice Loss is the complement of the Dice coefficient and is defined as:

$$\mathrm{L}_{Dice} = 1 - \frac{2 \cdot \sum_i P_i \cdot G_i}{\sum_i P_i + \sum_i G_i}$$

where, P_i and G_i represent the predicted and ground truth values at pixel *I*, respectively. The dice loss reduces the overlap error between expected and actual lesions rather significantly.

Pixel-wise classification is supported alongside Dice Loss by cross-entropy loss. This helps especially to ensure that the model not only pays attention to lesion areas but also acquires proper awareness of the background. Here is defined the Cross-Entropy Loss for a pixel x with class label *y*:

$$\mathrm{L}_{CE} = -\sum_c y_c \cdot \log(p_c(x))$$

Where: y_c is the true label for class ccc at pixel x, $p_c(x)$ is the predicted probability of class ccc for pixel x. The weighted sum of the Dice Loss and the Cross-Entropy Loss allows one to determine the overall loss for supervised training for the following purposes:

$$\mathrm{L}_s = \alpha \cdot \mathrm{L}_D + \beta \cdot \mathrm{L}_{CE}$$

The weighting elements α and β help to guarantee the homogeneous distribution of every loss term. Training to lower this total loss, the model learns from the annotated data an ideal segmentation strategy for skin lesions. One can accomplish this by using the statistics as a learning tool.

E. Unsupervised Training

Under the framework of the unsupervised training phase, using unlabelled data will help to increase the generalisation capacity of the model. Since obtaining large-scale annotated data is often impractical in medical domains, unlabelled data provides valuable additional training samples to improve the model's robustness and prevent overfitting. In this method, the model is trained to learn from both the labelled and unlabelled datasets using techniques such as Generative Adversarial Networks (GANs) and Structured Prediction-based Deep Reinforcement Learning (SP-DRL).

Generative Adversarial Networks (GANs**) :** GANs consist of two networks: a Generator that creates synthetic images from random noise, and a Discriminator that distinguishes between real and generated images. The unsupervised phase leverages unlabelled images to train the GAN, where the Generator learns to create plausible lesion patterns, while the Discriminator evaluates the realism of the generated lesions.

The Generator GGG tries to minimize the adversarial loss, while the Discriminator DDD tries to maximize it. The adversarial loss is typically defined as:

$$\mathrm{L}_{GAN} = \mathrm{E}[\log D(x)] + \mathrm{E}[\log(1 - D(G(z)))]$$

Where: D(x) is the probability that image *x* is real, G(z) is the synthetic image generated from noise zzz, E denotes the expected value over the training set. Through adversarial training, the Generator creates increasingly realistic skin lesion images, which can then be used to augment the training set, aiding the segmentation model during subsequent training steps.

F. Structured Prediction-Deep Reinforcement Learning (SP-DRL)

In SP-DRL, the goal is to optimize the segmentation model's prediction strategies. Here, the network learns to make segmentation decisions by exploring unlabelled data in a reinforcement learning framework. The model is treated as an agent that interacts with an environment (the input image), where it takes actions (segmentation predictions) based on states (image features). The reward function evaluates the quality of the segmentation, guiding the agent to improve over time. The agent's objective is to maximize the expected cumulative reward:

$$R_{total} = E\left[\sum_{t=0}^{T} \gamma^t R_t\right]$$

Where: E is the expected reward over all actions, γ is the discount factor (determining how future rewards are weighted), R_t is the reward at time step t, T is the total number of steps in an episode. The reward R_t at each step is typically computed based on the segmentation quality (e.g., based on Dice similarity or pixel-level classification accuracy). The agent is trained to improve its segmentation strategy, leveraging both the synthetic data generated by the GAN and unlabelled data.

G. Surrogate Task with Transformer Representation

The surrogate task with transformer representation is a key component of the proposed method to enhance skin lesion segmentation by leveraging unlabelled data. Unlike traditional approaches that rely solely on pixel-level annotations, the proposed method introduces a surrogate task to extract high-level features from the input images. This task aids in learning robust representations without requiring extensive labeled datasets. By integrating a Transformer-based model with convolutional networks, the method enables the model to learn complex relationships within image data at different scales and improve its generalization capabilities.

Surrogate Task : A **surrogate task** refers to an auxiliary task that is designed to help the model learn useful features that are indirectly related to the main task (skin lesion segmentation in this case). The main idea behind this surrogate task is to guide the model to learn semantic features from the input images, even when the true segmentation labels are not available for all data. The surrogate task operates by predicting certain high-level representations or abstractions of the data (such as image context or texture) that can inform the segmentation model about the key features of lesions.

One common surrogate task for skin lesion segmentation is image reconstruction or patch prediction, where the network learns to predict certain parts of the input image, based on other parts. This will inspire the model to acquire more global and semantic elements regarding the skin lesions.

Transformer Representation: Transformers are a fantastic tool for capturing in images long-range dependencies and contextual relationships. A Transformer model can ignore pointless areas of the image and focus on important ones by means of attention mechanisms. This is especially helpful for tasks involving segmentation since it enables the model to concentrate on the most important areas of skin lesions considering their context inside the whole image.

Feature extraction in the proposed framework makes use of Transformer encoder-decoder architecture. Following division of the input image into patches, a learnable embedding based on Transformer encoder processing projects each patch into a high-dimensional space. At last the space shows the patches. The model can then compute pairwise interactions between patches, which allowing it to capture global dependencies. The attention mechanism makes this practical.

Self-attention mechanism of the Transformer model is its most important feature. Every individual patch in the image has a computed attention score using this mechanism. Should the self-attention mechanism receive an input patch representation x_i for patch I, it will compute a weighted sum of all the patch representations, where the attention scores will determine the weights.

$$V(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)$$

Where: Q, K, and V are the Query, Key, and Value matrices, derived from the input features, d is the dimension of the feature space,

$\frac{QK^T}{\sqrt{d}}$ calculates the pairwise similarity between patches,

The softmax function normalizes the attention scores, ensuring they sum to 1. The attention weights allow the model to focus on relevant parts of the image (e.g., lesion areas) and ignore less relevant regions (e.g., background).

Learning Surrogate Task with Transformer Representations: Once the Transformer extracts contextual information from the image, the next step is to apply this knowledge to the surrogate task. In this case, the surrogate task could involve reconstructing the image patches or predicting high-level semantic features (e.g., lesion boundaries or shapes) from the Transformer's learned representations. The idea is that by training the model to perform this auxiliary task, the model can capture important features that are useful for the final skin lesion segmentation.

To formalize this surrogate task, let I be the input image, and PPP be the set of patches obtained from the image. The Transformer network computes the patch embeddings and attends to the context of each patch:

$$H_i = \vartheta(P_i)$$

Where: P_i represents the patch I of the input image I, H_i is the output feature representation of patch I. For the reconstruction task, the goal is to predict the original image from these patch representations. The model is trained to minimize the reconstruction error, which can be expressed using a loss function such as Mean Squared Error (MSE):

$$L_{recon} = \frac{1}{N}\sum_{i=1}^{N} \| I_i - \hat{I}_i \|^2$$

Where: I is the original image patch, $\hat{I}_i$ is the predicted patch, N is the number of patches in the image. By minimizing the reconstruction loss, the model learns to extract useful features from the input image, such as the shape, texture, and context of the skin lesions.

Integrating Surrogate Task into Segmentation Model: The main skin lesion segmentation model includes the last output of the Transformer-based surrogate task. Regarding the segmentation process of decision-making, the output of the

Transformer shows great value as rich representations. Combining the conventional pixel-level loss, that of Dice loss and Cross-Entropy loss, forms the total segmentation loss from the surrogate task loss. This combination ensures the semantic features obtained from the surrogate task as well as the optimisation of the pixel-level segmentation details jointly:

$$L_{total} = L_{seg} + \lambda \cdot L_{recon}$$

Where: L_{seg} is the segmentation loss (e.g., Dice loss), L_{recon} is the reconstruction loss from the surrogate task, λ is a hyperparameter that controls the balance between the segmentation and surrogate task losses. Through this surrogate activity, the model is encouraged to develop global context and semantic features from unlabelled data. This helps the model to generalise to hitherto unheard-of types of lesions and improves its segmentation performance.

IV RESULTS AND DISCUSSION

The proposed segmentation scheme for skin lesions was evaluated using extensive simulations in a virtual environment. Mostly used as the simulation tool for model development, training, and evaluation was Python. Training and evaluation make use of well-known deep learning libraries including PyTorch and TensorFlow. The computational setup for training and model testing consisted on high-performance systems supplied with NVIDIA RTX 3090 graphics processing units (GPUs). These GPUs could handle the heavy computations the deep learning algorithms required. The machines provided the necessary computational capacity with Intel i9 CPUs and 32 gigabytes of random access memory (RAM), which enabling efficient training and evaluation of large models.

The proposed method was evaluated using six skin lesion segmentation techniques considered as modern standards. Using the ISIC 2018, ISIC 2017, and PH2 datasets, each of which included skin lesion images with varying degrees of annotation quality and lesion complexity, the performance of the techniques was compared among confidence-aware semi-active learning (CASSL), convolutional network transformer (CNT), SLIC, MSCNet, DeepLabV3-ST, and Inv SSL. The comparison focused on generalising the model to non-traditional lesion types, segmentation accuracy, and model strength to unlabelled data.

Table 2: Experimental Setup

Parameter	Value
Learning Rate	0.0001
Batch Size	16
Epochs	50
Optimizer	Adam
Weight Decay	1e-5
Dropout Rate	0.3
Loss Function	Cross-Entropy Loss, Dice Loss
Transformer Encoder Layers	4
Transformer Decoder Layers	4
Patch Size	16x16
Patch Overlap	50% (8x8)
Surrogate Task Weight (λ)	0.5

A. Performance Metrics

These metrics assess various aspects of model performance, such as accuracy, precision, and generalization ability:

Dice Similarity Coefficient (DSC): The Dice Similarity Coefficient is a measure of overlap between the predicted segmentation mask and the ground truth mask. It is calculated as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Where A is the predicted segmentation area, and B is the ground truth. A higher DSC indicates better overlap.

Intersection over Union (IoU): **IoU** quantifies the similarity between the predicted segmentation mask and the ground truth mask by measuring the ratio of the intersection to the union of the two areas: A higher IoU value indicates better segmentation quality.

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

Precision: **Precision** measures the percentage of true positive pixels (correctly segmented lesion pixels) relative to all pixels predicted as positive:

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity): **Recall** measures the percentage of true positive pixels relative to all the actual positive pixels (lesion pixels in the ground truth):

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The **F1-Score** is the harmonic mean of precision and recall, providing a single metric that balances both:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

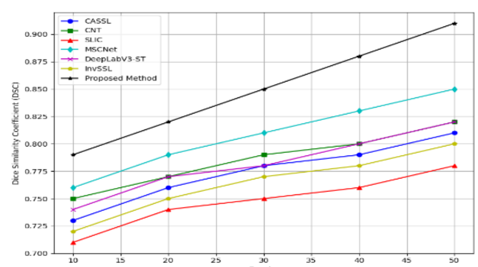

Figure 2: DSC

In figure 2, consistent increasing from 0.79 at epoch 10 to 0.91 at epoch 50, the DSC shows better performance over all epochs. This implies that the approach is always improving with time. While CNT gets a DSC of 0.82 at the same period, at epoch 50 CASSL reaches a DSC of 0.81. This is a significant progress over the methods now applied. Since it promotes generalisation and refinement during the method is under development, the proposed approach gains from the integration of unlabelled data and the surrogate task. In terms of segmentation accuracy, the proposed approach continuously shows better performance than other ones.

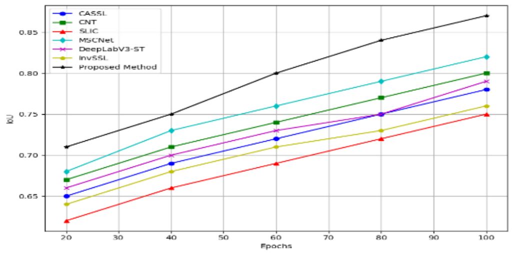

Figure 3: IoU

The IoU results in figure 3 clearly show that the proposed method is superior than the ones now used over all periods. The proposed approach beats CASSL (0.78), CNT (0.80), and SLIC (0.75), by a margin of much higher value with an IoU of 0.87 at epoch 100. The significant rise in IoU obtained shows the effectiveness of the proposed semi-supervised method, which employs unlabelled data via means of the surrogate task and Transformer representations. This improved performance makes one conclude that the model generalises more effectively, which finally results in lower overfitting in comparison to conventional techniques and better segmentation quality.

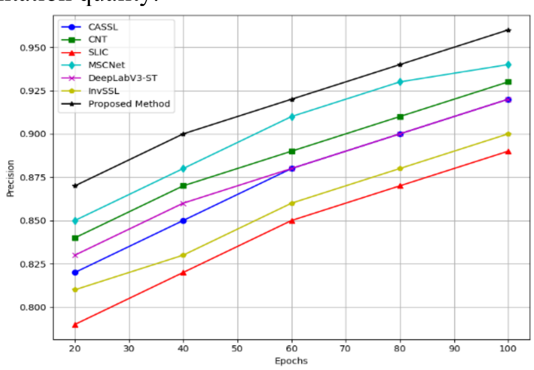

Figure 4: Precision

In figure 4, the Precision test shows a constant and appreciable improvement for the recommended approach. With a precision score of 0.96 by the time epoch 100 passes, the proposed method outperforms all other methods including CASSL (0.92), CNT (0.93), and SLIC (0.89). The proposed method reduces the number of false positives by including unlabelled samples and a surrogate task with Transformer representations, which generating a more accurate identification of true positive pixels. Since the constant increase in precision over the course of the epochs indicates, this shows that the proposed semi-supervised learning method is successful in improving segmentation accuracy.

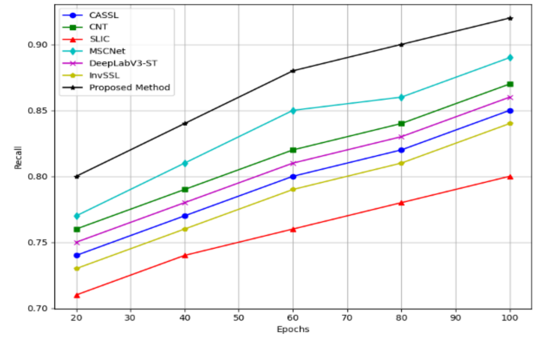

Figure 5: Recall

Figure 5 shows the results of the recall show how much more effectively the proposed approach detects true positive lesions than the methods now in figure 5. Higher than the recalls obtained by CASSL (0.85), CNT (0.87), and SLIC (0%), the proposed method achieves a recall of 0.92 by the time epoch 100 has passed. This improvement suggests that the semi-supervised training strategy, more especially, the use of unlabelled data and the surrogate task with Transformer representations, helps the model to better segment and capture all relevant skin lesions. The constant rise in the recall score indicates that the model can lower false negatives, which enhancing the general segmentation.

Table 3(a): DSC, IoU, and F1 Results for Training and Validation Sets

Metric	CASSL	CNT	SLIC	MSCNet	Proposed
	Train				
DSC	0.82	0.84	0.78	0.86	0.91
IoU	0.68	0.71	0.64	0.73	0.84
F1	0.79	0.81	0.74	0.82	0.94

Table 3(b): DSC, IoU, and F1 Results for Training and Validation Sets

Metric	CASSL	CNT	SLIC	MSCNet	Proposed
	Valid				
DSC	0.76	0.78	0.73	0.81	0.89
IoU	0.60	0.63	0.58	0.69	0.78
F1	0.71	0.73	0.68	0.78	0.87

Table 3(a) and (b) shows training and validation set DSC, IoU, F1 results. With regard to the DSC measure, the recommended approach earns a score of 0.91 in training and 0.89 in validation, above both other approaches and the CASSL methodology (0.82 in training and 0.76 in validation). Similar trends are shown by the IoU score; the recommended method gets a score of 0.84 in training and 0.78 in validation while the other approaches lag ever more behind. Compincreasing 0.94 in the training phase and 0.87 in the validation phase, the F1 score for the proposed approach is also the best. This points to rather good harmony between recall and accuracy. These results clearly show that the proposed semi-supervised method can effectively improve segmentation performance by using unlabelled data via means of the surrogate task and Transformer representations. Consistent improvement of the proposed method over the training and validation sets in comparison to more conventional approaches indicates its resilience and ability to generalise more effectively, which reducing the degree of underfitting that results.

V CONCLUSION

In this study, we proposed a novel semi-supervised approach for skin lesion segmentation, aiming to address the challenges of limited annotated data in medical image processing. The results presented a new semi-supervised method for skin lesion segmentation for this work. The unlabelled samples into the training process enabled us to effectively increase the generalisation and resilience of the model. The surrogate task using Transformer representations let one effectively extract features from skin lesion images, which reducing the demand for large-scale datasets with suitable annotations. Over a spectrum of performance criteria on the training set as well as the validation one, the experimental results revealed that the proposed approach outperforms current methods. Among these are DSC, IoU, Precision, Recall, and F1-score. These were consistent across several datasets including PH2, ISIC 2018, and ISIC 2017, containing skin lesions. In terms of reducing overfitting and false positives, common problems in medical image segmentation tasks, the proposed approach obtained the best segmentation performance and accuracy. The combination of the surrogate activity with Transformer-based feature learning offers a possible approach for segmentation of skin lesions depending on proposed semi-supervised learning framework. This method not only enhances segmentation performance but also reduces reliance on large, annotated datasets, which increasing their relevance in practical medical uses.

REFERENCES

[1] Du, F., Peng, B., Al-Huda, Z., & Yao, J. (2024). Semi-Supervised Skin Lesion Segmentation via Iterative Mask Optimization. International Journal of Image and Graphics, 24(02), 2450020.

[2] Gu, Y., Zhou, T., Zhang, Y., Zhou, Y., He, K., Gong, C., & Fu, H. (2025). Dual-scale enhanced and cross-generative consistency learning for semi-supervised medical image segmentation. Pattern Recognition, 158, 110962.

[3] Ma, Y., Wang, J., Yang, J., & Wang, L. (2024). Model-Heterogeneous Semi-Supervised Federated Learning for Medical Image Segmentation. IEEE Transactions on Medical Imaging.

[4] Lu, Y., Li, W., Cui, Z., & Zhang, Y. (2025). Beyond low-dimensional features: Enhancing semi-supervised medical image semantic segmentation with advanced consistency learning techniques. Expert Systems with Applications, 261, 125456.

[5] Zhang, Y., Zhou, T., Wu, Y., Gu, P., & Wang, S. (2024, October). Combining Segment Anything Model with Domain-Specific Knowledge for Semi-Supervised Learning in Medical Image Segmentation. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (pp. 343-357). Singapore: Springer Nature Singapore.

[6] Yuan, W., Du, Z., & Han, S. (2024). Semi-supervised skin cancer diagnosis based on self-feedback threshold focal learning. *Discover Oncology*, *15*(1), 180.

[7] Yan, H., Wang, P., Jia, Y., Si, X., & Wei, B. (2024). Combining unsupervised constraints on weakly supervised semantic segmentation of skin cancer. *Biomedical Physics & Engineering Express*, *10*(5), 055024.

[8] Lei, T., Wang, Y., Wang, X., Wang, X., Hu, B., & Nandi, A. K. (2024). Unified Feature Consistency of Under-Performing Pixels and Valid Regions for Semi-Supervised Medical Image Segmentation. *IEEE Transactions on Radiation and Plasma Medical Sciences*.

[9] He, A., Li, T., Wu, Y., Zou, K., & Fu, H. (2024, October). FRCNet: Frequency and Region Consistency for Semi-supervised Medical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 305-315). Cham: Springer Nature Switzerland.

[10] Ye, P., Li, H., Zhang, S., & Wang, G. (2024, May). Semi-Supervised Medical Image Segmentation via Multi-Group Network with Consistency Among Heterogeneous Loss Supervision. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5). IEEE.

[11] Xia, Q., Zheng, H., Zou, H., Luo, D., Tang, H., Li, L., & Jiang, B. (2024). A Comprehensive Review of Deep Learning for Medical Image Segmentation. *Neurocomputing*, 128740.

[12] Li, B., Xu, Y., Wang, Y., & Li, X. (2024, May). Multi-Level Contrastive Student-Learner Structure for Semi-Supervised Medical Image Segmentation. In *2024 9th International Symposium on Computer and Information Processing Technology (ISCIPT)* (pp. 628-632). IEEE.

[13] Xie, Z., Tu, E., Zheng, H., Gu, Y., & Yang, J. (2021, June). Semi-supervised skin lesion segmentation with learning model confidence. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1135-1139). IEEE.

[14] Zhang, Z., Tian, C., Gao, X., Wang, C., Feng, X., Bai, H. X., & Jiao, Z. (2022). Dynamic prototypical feature representation learning framework for semi-supervised skin lesion segmentation. *Neurocomputing*, *507*, 369-382.

[15] Alahmadi, M. D., & Alghamdi, W. (2022). Semi-supervised skin lesion segmentation with coupling CNN and transformer features. *IEEE Access*, *10*, 122560-122569.

[16] Dos Santos, E. S., de MS Veras, R., Aires, K. R., Portela, H. M., Junior, G. B., Santos, J. D., & Tavares, J. M. R. (2022). Semi-automatic segmentation of skin lesions based on superpixels and hybrid texture information. *Medical Image Analysis*, *77*, 102363.

[17] Shen, X., Sun, Z., Sun, Y., & Chen, H. (2023). Confidence-aware cross-supervised model for semi-supervised skin lesion segmentation. *Journal of Electronic Imaging*, *32*(1), 013016-013016.

[18] Nouboukpo, A., Allaoui, M. L., & Allili, M. S. (2024). Multi-scale spatial consistency for deep semi-supervised skin lesion segmentation. *Engineering Applications of Artificial Intelligence*, *135*, 108681.

[19] Dzieniszewska, A., Garbat, P., & Piramidowicz, R. (2024). Improving skin lesion segmentation with self-training. *Cancers*, *16*(6), 1120.

[20] Feng, X., Lin, J., Feng, C. M., & Lu, G. (2024). GAN inversion-based semi-supervised learning for medical image segmentation. *Biomedical Signal Processing and Control*, *88*, 105536.