Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3455~3470
e-ISSN: 2548-6861

Benchmarking Pseudo-Mask Generation Methods for ResNet34-U-Net Acne Segmentation

Nikita Amelia Valencia [1], Puguh Hiskiawan [2]*
Data Science, Faculty of Technology and Design, Bunda Mulia University
s36230005@student.ubm.ac.id [1], phiskiawan@bundamulia.ac.id [2]

Article Info

Article history:

Received 2026-07-04
Revised 2026-07-25
Accepted 2026-08-08

Keyword:

Acne Lesion Segmentation,
Pseudo-Mask Generation,
U-Net,
Superpixel,
Weakly Supervised Segmentation.

ABSTRACT

Acne lesion segmentation is essential for automated dermatological image analysis. However, developing accurate deep learning segmentation models requires pixel-level annotations, which are costly and time-consuming to obtain. To address this limitation, pseudo-mask generation methods can be utilized to produce surrogate labels for training segmentation networks without manual annotation. This study presents a comparative evaluation of five pseudo-mask generation methods for acne lesion segmentation, namely Contour-Based, Superpixel (SLIC), K-Means, Weakly Supervised Semantic Segmentation (WSSS) based on Otsu Thresholding and Morphological Operations, and Pseudo-Mask-Based Generation. The generated pseudo masks were employed as supervisory labels to train a ResNet34-U-Net segmentation model under identical experimental settings. To improve the robustness of the training process, data augmentation was applied exclusively to the training dataset. Segmentation performance was quantitatively evaluated using Dice Score, Intersection over Union (IoU), Precision, Recall and Validation Loss, and qualitatively assessed through visual comparisons of the generated pseudo masks and predicted segmentation results. The experimental results demonstrate that pseudo-mask quality has a substantial impact on segmentation performance. Among the evaluated methods, Superpixel (SLIC) achieved the highest performance with a Dice Score of 0.898, an IoU of 0.815, a Precision of 0.889, a Recall of 0.908, and the lowest validation loss of 0.268 indicating superior lesion boundary preservation and region consistency. WSSS (Otsu + Morphology) also produced competitive results, whereas Pseudo-Mask-Based and Contour-Based methods yielded comparatively lower performance. These findings demonstrate that high-quality pseudo masks can provide effective supervision for ResNet34-U-Net training, offering an annotation-efficient approach for acne lesion segmentation and providing practical insights into selecting suitable pseudo-mask generation methods for dermatological image analysis.

I. Introduction

Acne vulgaris is one of the most common chronic inflammatory skin disorders affecting adolescents and young adults worldwide. Recent epidemiological research shows that the worldwide prevalence of acne has risen over the last few decades, especially in those aged 10-24 years, establishing it as an ongoing public health issue [1]. Acne lesions, such as blackheads, whiteheads, papules, pustules, and cysts, frequently occur on facial skin and can differ in severity across individuals. Although acne is usually not classified as a serious medical issue, its visible effects can result in lasting scars, diminished self-worth, and multiple psychosocial difficulties that adversely influence a person's quality of life [2]. Consequently, accurate identification and assessment of acne lesions are essential for supporting clinical evaluation, treatment planning, and long-term skin condition monitoring.

Recent advances in computer vision and artificial intelligence have accelerated the development of automated dermatological image analysis systems. Deep learning methods have shown exceptional effectiveness in numerous medical imaging applications, such as skin lesion detection, disease classification, object localization, and semantic segmentation [3], [4], [5]. Moreover, deep learning has demonstrated remarkable capability across various pattern recognition tasks, highlighting its ability to learn discriminative feature representations from complex data in different application domains [6]. These technologies enable computers to automatically identify significant visual characteristics from medical images, thus facilitating more effective and unbiased clinical evalutation. In acne analysis, image-based systems have been increasingly utilized to identify acne lesion types, assess severity levels, and assist in treatment monitoring using facial skin images [7]. However, the efficacy of these systems greatly relies on the quality and relevance of the visual data obtained from the image, especially when differentiating lesion areas from adjacent healthy skin and various background artifacts [4].

One of the major challenges in acne image analysis is the complexity of background details, such as normal skin texture, pores, facial hair, variations in lighting, and shadows [8]. These factors may interfere with the model's ability to focus on the actual lesion region, leading to reduced analytical performance. As a result, lesion segmentation has become an important preprocessing step in dermatological image analysis [9]. Segmentation techniques can enhance the interpretability of future image analysis processes by isolating acne lesions from adjacent skin areas, thereby emphasizing significant visual features [10].

Various pseudo-mask generation techniques have been proposed to facilitate image segmentation tasks by reducing the dependence on expensive pixel-level annotations [11]. Traditional image processing techniques such as contour detection, clustering, thresholding, and superpixel generation have been widely adopted to produce lesion masks due to their simplicity and low computational cost. Contour-based methods identify lesion boundaries through edge information, while clustering approaches such as K-Means group pixels based on color similarity. Superpixel segmentation further improves image representation by aggregating neighboring pixels into perceptually meaningful regions [12]. Although these methods are computationally efficient, their performance may be influenced by lighting conditions, skin color variations, and the diverse appearance of acne lesions.

However, existing pseudo-mask generation approaches still face several limitations. The quality of automatically generated masks is highly dependent on the characteristics of the employed generation strategy, where different approaches may produce varying levels of boundary accuracy, background noise, and lesion localization capability. Previous weakly supervised segmentation studies have also highlighted that pseudo-label quality remains a critical factor affecting segmentation performance, as inaccurate pseudo-masks may introduce noise during model training [13], [14]. Nevertheless, a systematic comparison of different pseudo-mask generation strategies under identical segmentation architectures and experimental settings remains limited, particularly for acne lesion segmentation.

Previous studies have primarily focused on acne classification or the development of individual segmentation models, while comparative investigations of pseudo-mask generation methods remain relatively limited [15]. Furthermore, most existing studies evaluate only a small number of mask generation techniques, making it difficult to determine which approach provides the most suitable pseudo labels for acne lesion segmentation. As a result, the relative strengths and limitations of different pseudo-mask generation strategies for acne lesion analysis are not yet fully understood.

To address this research gap, this study presents a comparative evaluation of five pseudo-mask generation methods for U-Net-based acne lesion segmentation, namely Contour-Based Segmentation, Superpixel Segmentation (SLIC), K-Means Segmentation, WSSS based on Otsu Thresholding and Morphological Operations, and Pseudo Mask Generation. The generated pseudo masks are subsequently used as supervisory labels for training a ResNet34-U-Net segmentation model under identical experimental settings. The main contribution of this study is the establishment of a unified benchmarking framework that systematically evaluates five pseudo-mask generation methods using the same dataset, ResNet34-U-Net architecture, training configuration, and evaluation metrics, enabling an objective and fair comparison of pseudo-mask quality and its impact on acne lesion segmentation performance. The U-Net architecture was selected because of its encoder–decoder structure with skip connections, which effectively preserves fine spatial information while enabling accurate localization of lesion boundaries. In addition, the ResNet34 encoder provides rich hierarchical feature representations through residual learning, facilitating the extraction of both low-level texture information and high-level semantic features from acne images without introducing excessive computational complexity [16]. This combination is therefore well suited for acne lesion segmentation, where lesions exhibit considerable variations in size, shape, color, and boundary characteristics. The performance of each pseudo-mask generation method is evaluated using Intersection over Union (IoU), Dice Score, and Validation Loss to assess segmentation accuracy and model convergence. Through both quantitative and qualitative analyses, this study aims to identify the most effective pseudo-mask generation strategy for acne lesion segmentation and provide insights into the development of annotation-efficient dermatological image analysis systems.

II. METHOD

A. Dataset

Acne lesion images have been widely utilized for the development and evaluation of image segmentation techniques in dermatological image analysis [17]. The dataset used in this study was obtained from the Acne Dataset Image available on the Kaggle platform. The original dataset consists of 4,617 facial acne images organized into predefined training, validation, and testing subsets. It includes five acne lesion categories, namely Blackheads (1240 images), Whiteheads (299 images), Papules (1032 images), Pustules (1006 images), and Cysts (1040 images). The images exhibit variations in lesion appearance, skin texture, image resolution, facial regions, and illumination conditions, providing diverse visual characteristics for evaluating the robustness of acne lesion segmentation methods.

From the original dataset, 50 representative images were selected for this study, consisting of 10 images from each lesion category to maintain a balanced class distribution. The image selection process was conducted manually by reviewing image quality, lesion visibility, and category relevance. Images with severe blur, excessive occlusion, or unclear lesion boundaries were excluded to ensure consistency throughout the experimental process.

The selected images exhibit variations in lesion size, shape, color intensity, and background characteristics, providing diverse visual patterns for segmentation evaluation. All images were stored in RGB format and subsequently processed using the same preprocessing and segmentation pipeline to ensure fair comparison among the investigated methods.

The Acne Dataset provides image-level acne category labels only and does not contain manually annotated pixel-level segmentation masks as ground truth. Therefore, pseudo masks were automatically generated using the evaluated pseudo-mask generation methods and subsequently utilized as surrogate supervisory labels for training the ResNet34-U-Net model. This experimental design enables a comparative evaluation of different pseudo-mask generation approaches without requiring manually annotated segmentation references.

This study is designed as a preliminary benchmarking study to evaluate the feasibility and comparative performance of different pseudo-mask generation methods for acne lesion segmentation using a ResNet34-U-Net architecture. Therefore, a balanced subset of 50 representative images was intentionally selected to establish a controlled experimental setting in which all methods were evaluated under identical conditions. This controlled subset enables a fair comparison of pseudo-mask quality and its impact on segmentation performance while minimizing the influence of class imbalance and data variability. The primary objective of this study is to benchmark pseudo-mask generation methods rather than develop a large-scale clinical segmentation system. Consequently, the complete dataset will be utilized in future work for more comprehensive model development and validation.

B. Image Preprocessing

Dermatological images often exhibit variations in image resolution, illumination conditions, skin texture, lesion appearance, and background characteristics. These variations may affect the performance of segmentation algorithms by introducing inconsistencies in lesion representation [18]. Therefore, image preprocessing was performed to improve image quality, enhance lesion visibility, and standardize image inputs before the segmentation stage [17].

Several preprocessing operations were applied to the selected acne images. First, image resizing was conducted to adjust image dimensions according to the requirements of each segmentation method while reducing computational complexity [19]. RGB images were then maintained for segmentation approaches that utilize color information, whereas grayscale conversion was applied for methods that rely on intensity-based feature extraction [20]. Thresholding techniques were subsequently employed to separate potential lesion regions from surrounding skin areas. In particular, Otsu Thresholding was utilized to automatically determine optimal threshold values based on image intensity distributions [21].

Additional preprocessing steps were conducted to enhance the extracted lesion areas. Filtering and masking techniques were applied to reduce image noise and suppress irrelevant artifacts that may interfere with lesion detection [22]. Furthermore, morphological processing using a 3 × 3 kernel was employed to refine lesion masks and improve the representation of acne lesion regions generated during thresholding and clustering processes [23]. To improve the robustness and generalization of the segmentation model, data augmentation was applied exclusively to the training images using random horizontal and vertical flipping, random rotation within ±15°, and random brightness and contrast adjustment. Subsequently, all training and validation images were normalized and converted into tensor format before being used for model training and evaluation. These processes contribute to generating clearer and more accurate lesion masks for later examination.

Different pseudo-mask generation methods employed distinct preprocessing configurations according to their underlying methodological principles. Contour-Based Segmentation and Traditional Pseudo-Mask Segmentation utilized grayscale conversion, thresholding, filtering, and masking operations to emphasize acne lesion boundaries and generate surrogate lesion masks. Weakly Supervised Semantic Segmentation (WSSS) generated pseudo masks through a combination of thresholding and morphological refinement without relying on pixel-level annotations, leveraging image intensity information to approximate lesion regions. Meanwhile, Superpixel Segmentation and K-Means Segmentation utilized color and intensity characteristics to partition images into coherent regions corresponding to acne lesions. The generated pseudo masks were subsequently used as supervisory signals for training the U-Net model, which produced the final acne lesion segmentation masks.

C. Research Framework

The proposed framework was developed to evaluate the effectiveness of different pseudo-mask generation methods for U-Net-based acne lesion segmentation. The overall

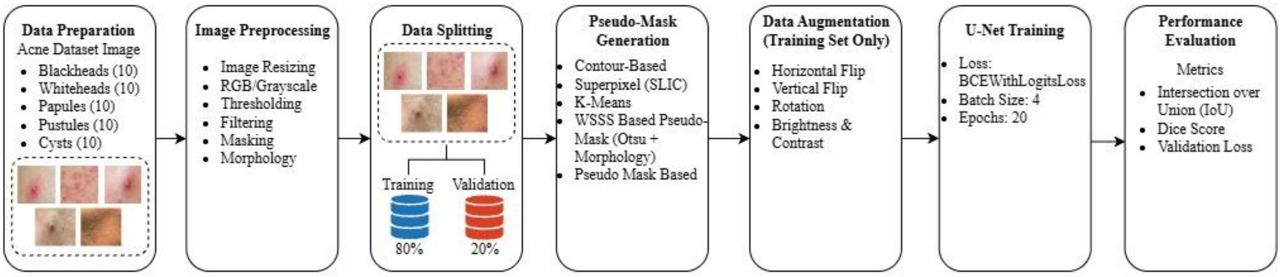


Figure 1. Framework of Pseudo-Mask Generation Methods for ResNet34-U-Net Acne Segmentation

workflow consists of dataset preparation, image preprocessing, dataset splitting, pseudo-mask generation, U-Net training, and performance evaluation. Each stage contributes to transforming raw acne images into segmentation masks that can be utilized as supervisory labels for training the segmentation model.

Acne images obtained from the Acne Dataset Image were first processed through an image preprocessing stage to improve image consistency and enhance lesion visibility. The preprocessing operations included image resizing, color space conversion, thresholding, filtering, masking, and morphological refinement. The resulting preprocessed images were subsequently divided into training and validation subsets using an 80:20 ratio, consisting of 40 training images and 10 validation images.

Five pseudo-mask generation methods were investigated in this study, namely Contour-Based Segmentation, K-Means Segmentation, Superpixel Segmentation, Traditional Pseudo-Mask Segmentation, and Weakly Supervised Semantic Segmentation (WSSS)-based Pseudo-Mask Generation using Otsu thresholding and morphological refinement. Each method generated lesion masks that were subsequently employed as training labels for the U-Net segmentation model. Before the U-Net training process, data augmentation was applied to the training dataset to increase data diversity and improve the model's generalization capability, while the validation dataset remained unchanged for performance evaluation. To ensure a fair comparison, the same U-Net architecture and training configuration were applied across all experiments.

After training, the U-Net model produced segmentation masks representing acne lesion regions from the input images. The segmentation results were then evaluated using Intersection over Union (IoU), Dice Score, and Validation Loss. These evaluation metrics were employed to measure segmentation quality and assess the effectiveness of each pseudo-mask generation method. The experimental results were further analyzed to determine which pseudo-mask generation strategy provides the most effective supervision for U-Net-based acne lesion segmentation. Figure 1 illustrates the overall framework of the proposed segmentation system.

D. Pseudo-Mask Generation Methods

This study investigates and compares five pseudo-mask generation methods for U-Net-based acne lesion segmentation: Contour-Based Segmentation, Superpixel Segmentation, K-Means Segmentation, Weakly Supervised Semantic Segmentation (WSSS), and Pseudo Mask-Based Segmentation. These methods were selected because they represent the major categories of automatic pseudo-mask generation techniques, including edge-based, region-based, clustering-based, weakly supervised, and multi-stage image processing approaches. Together, they provide representative coverage of widely used pseudo-mask generation strategies that do not require manually annotated pixel-level masks, enabling a fair and consistent benchmarking under identical experimental conditions. These methods represent different strategies for generating surrogate lesion annotations, ranging from classical image processing techniques to weakly supervised learning. The generated pseudo-masks are then used to train a U-Net model under identical training settings to ensure a fair comparison. Each method generates lesion masks based on different image characteristics, including edge information, color similarity, region homogeneity, and activation-based localization.

Contour-Based Segmentation was performed by identifying lesion boundaries through image intensity variations. Initially, the input image was converted into grayscale and processed using thresholding techniques to separate foreground and background regions. Contours corresponding to potential lesion areas were subsequently extracted and transformed into binary segmentation masks. This approach relies on edge information and object shape characteristics to delineate acne lesion boundaries [17].

Superpixel Segmentation was implemented to divide images into perceptually meaningful regions by grouping neighboring pixels with similar visual characteristics. The generated superpixels preserve local image structures and reduce image complexity while maintaining important lesion boundaries. Lesion masks were subsequently generated by selecting superpixel regions corresponding to acne lesions. This approach provides a more compact image representation compared with pixel-wise processing and has been widely adopted in medical image analysis for region-based segmentation tasks [17].

K-Means Segmentation was utilized as a clustering-based approach for lesion extraction. In this method, image pixels were grouped into several clusters according to their color similarity. The cluster representing acne lesion regions was subsequently selected to generate a binary segmentation mask. By partitioning pixels into homogeneous groups, K-Means enables automatic separation of lesion regions from surrounding skin areas without requiring manually annotated labels [24].

A weakly supervised segmentation approach was employed to generate pseudo lesion masks without requiring pixel-level annotations [25]. Unlike conventional segmentation methods that depend on manually delineated masks, this approach approximates lesion regions directly from the input images using grayscale conversion followed by adaptive Otsu thresholding. The resulting binary masks were subsequently refined using morphological opening and closing operations to suppress small artifacts and improve regional continuity. Since the pseudo labels were generated using only intensity-based thresholding and simple morphological refinement, the produced masks provide coarse lesion localization rather than precise boundary delineation. These pseudo masks were then used to supervise the training of the U-Net model, allowing the network to learn lesion localization while substantially reducing annotation effort.

Pseudo Mask-Based Segmentation in this study refers to a handcrafted pseudo-mask generation approach that integrates multiple image processing techniques to generate surrogate labels for U-Net training. Unlike the other evaluated methods, which rely primarily on a single segmentation principle such as contour detection, superpixel partitioning, K-Means clustering, or threshold-based weak supervision, this approach combines a sequence of preprocessing and post-processing operations to progressively refine the generated pseudo masks before they are used as supervisory labels. The input images were first converted to grayscale and enhanced using Gaussian filtering to suppress image noise and Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast. Otsu thresholding was subsequently applied to separate lesion regions from the surrounding skin, followed by morphological opening and closing operations to refine the binary masks by removing small artifacts and improving region continuity. Connected component analysis was then employed to eliminate isolated regions and retain only meaningful lesion candidates, producing cleaner and more consistent pseudo labels. Finally, mask inversion was performed to ensure a consistent foreground representation of acne lesions. The resulting pseudo masks were subsequently used as surrogate labels for training the U-Net segmentation model, thereby reducing the need for manually annotated pixel-level masks while still enabling effective lesion segmentation [26], [27]. Compared with the WSSS approach, which primarily relies on adaptive thresholding and morphological refinement, Pseudo Mask-Based Segmentation incorporates additional contrast enhancement, noise reduction, connected component filtering, and foreground adjustment. These additional processing stages are intended to suppress background artifacts, improve lesion boundary continuity, and generate more refined pseudo masks before model training, thereby distinguishing this approach from the other pseudo-mask generation methods evaluated in this study.

The parameters of each pseudo-mask generation method were determined based on preliminary observations of the generated masks. The selected configurations were chosen to achieve a balance between lesion localization, boundary preservation, and noise reduction while maintaining consistent experimental conditions. In particular, the SLIC parameters, including the number of superpixels (n_segments) and compactness, were selected to preserve lesion boundaries while avoiding excessive fragmentation of image regions. All selected parameters were fixed and consistently applied across the dataset. The detailed parameter configurations are presented in Table 1.

TABLE 1
PARAMETER CONFIGURATION OF PSEUDO-MASK GENERATION METHODS

Pseudo-Mask Generation Method	Parameter	Configuration
Contour-Based	CLAHE clip limit	2.0
	CLAHE grid size	8 x 8
	Gaussian blur kernel	5 x 5
	Thresholding	Otsu + Binary Inverse
	Morphological kernel	3 x 3
	Morphological iteration	2
	Contour selection	Largest contour (>100 pixels)
Superpixel	Number of segments	250
	Compactness	10
	Sigma	1
	Color space	CIELAB
	Lesion criterion	Mean A-channel >12
	Morphological kernel	Elliptical 3 x 3
	Minimum component area	50 pixels
K-Means	Color space	CIELAB
	Feature space	L, A, B channels
	Number of clusters (K)	2
	Initialization	K-Means++
	Maximum iteration	10
	Cluster selection	Lowest L-channel cluster
	Morphological refinement	Opening and closing (3x3 kernel)
WSSS (Otsu + Morphology)	Color conversion	Grayscale
	Thresholding method	Otsu thresholding
	Morphological refinement	Opening and closing

	Kernel size	3 x 3
	Iteration	1
Pseudo Mask Based	Gaussian filtering	Kernel 5x5
	Contrast enhancement	CLAHE (clip limit=2.0, grid size=8x8)
	Thresholding	Otsu thresholding
	Morphological refinement	Opening (1 iteration) and Closing (2 iterations)
	Small object removal	Connected component filtering (>50 pixels)
	Foreground adjustment	Mask inversion

E. U-Net Architecture

The generated pseudo-masks were subsequently used as training labels for a U-Net-based segmentation model with a ResNet34 encoder. This architecture was selected for its ability to capture both high-level semantic information and fine-grained spatial details, which are essential for accurate acne lesion segmentation. U-Net-based architectures have also demonstrated excellent performance in medical image segmentation by effectively integrating encoder–decoder feature representations for accurate anatomical structure delineation [28]. The encoder employs a pre-trained ResNet34 backbone consisting of an initial convolutional layer followed by four residual stages with 3, 4, 6, and 3 residual blocks, respectively. Through successive convolution and downsampling operations, the encoder progressively extracts hierarchical features while reducing the spatial resolution of the input image [16], [29], [30].

The decoder follows the standard U-Net architecture and consists of five upsampling stages with channel dimensions of 256, 128, 64, 32, and 16. At each stage, the upsampled feature maps are concatenated with the corresponding encoder features via skip connections, enabling the recovery of fine spatial details lost during downsampling and improving lesion boundary localization. Each decoder block includes convolution, batch normalization, and ReLU activation layers to refine the reconstructed feature maps.

In this study, the network takes RGB acne images with a resolution of 256 × 256 pixels as input and produces a binary lesion segmentation mask of the same resolution. The final segmentation head uses a 3 × 3 convolution layer to generate a single-channel prediction mask. The overall U-Net. architecture with a ResNet34 encoder is illustrated in Figure 2.

To ensure a fair comparison among the investigated pseudo-mask generation methods, all experiments were conducted using the same network architecture and identical training settings. As summarized in Table 2, the model was trained for 20 epochs with a batch size of 4 using the Adam optimizer and a learning rate of 1×10^{-4}. The BCEWithLogitsLoss function was employed as the optimization objective. By maintaining consistent training parameters across all experiments, the observed performance differences can be primarily attributed to the pseudo-mask generation methods rather than variations in the segmentation model configuration.

TABLE 2
U-NET TRAINING CONFIGURATION

Parameter	**Value**
Architecture	U-Net with ResNet34 Encoder
Batch Size	4
Number of Epochs	20
Learning Rate	1×10^{-4}
Optimizer	Adam
Loss Function	BCEWithLogitsLoss

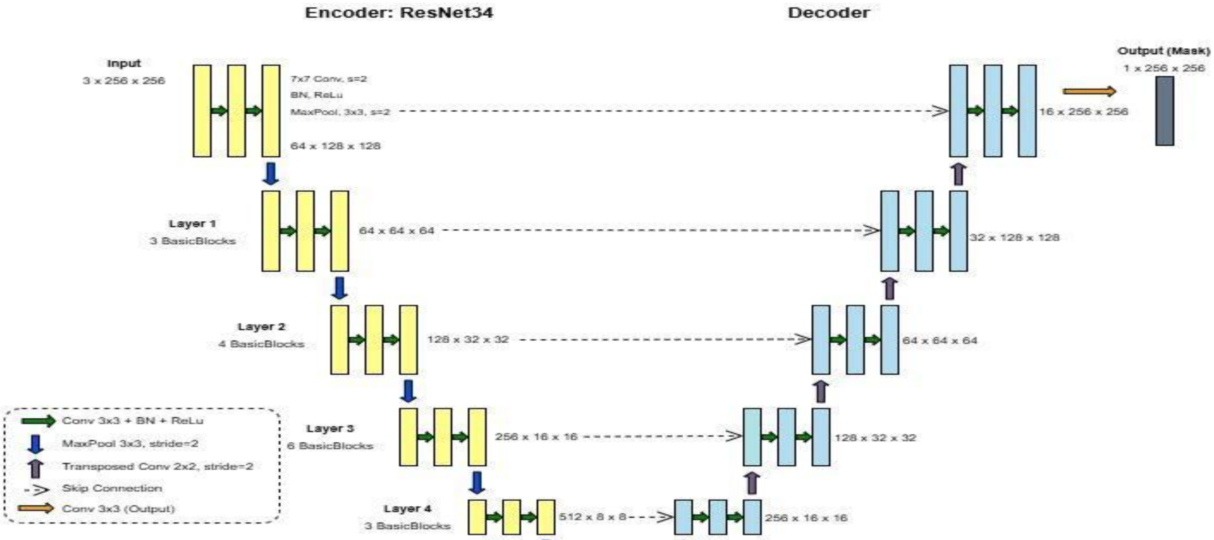


Figure 2. U-Net Architecture with ResNet34 Encoder

F. Loss Function

In this study, the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss) was used as the optimization objective for training the U-Net model across all experiments. The loss function was applied between the model predictions and the corresponding pseudo masks generated by each segmentation approach, including Contour-Based Segmentation, Superpixel Segmentation, K-Means Segmentation, Weakly Supervised Semantic Segmentation (WSSS), and Pseudo Mask-Based Segmentation. By using a single and consistent loss function, the study ensures a fair comparison across all methods, where performance differences are primarily influenced by the quality of the generated pseudo masks rather than variations in the training objective.

The BCEWithLogitsLoss combines a sigmoid activation function and binary cross-entropy loss into a single numerically stable formulation, which is well-suited for binary segmentation tasks. It is defined as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\sigma(x_i)) + (1-y_i)\log(1-\sigma(x_i))] \quad (1)$$

where y_i denotes the pseudo-label for pixel i, x_i represents the predicted logit output from the U-Net model, and $\sigma(x)$is the sigmoid activation function. Lower loss values indicate better agreement between the predicted segmentation output and the pseudo ground truth masks [31].

G. Evaluation Metrics

The performance of each segmentation method was evaluated using Intersection over Union (IoU) and Dice Score. IoU and Dice Score are among the most widely used evaluation metrics in medical image segmentation because they measure the overlap between predicted masks and reference masks.

Intersection over Union (IoU) measures the ratio between the overlapping area of the predicted mask and the reference mask to the total area covered by both masks. Higher IoU values indicate better agreement between the predicted segmentation result and the target lesion region [32].

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

where TP, FP, and FN denote true positive, false positive, and false negative pixels, respectively. IoU values range from 0 to 1, where values closer to 1 indicate higher segmentation accuracy.

Dice Score evaluates the similarity between predicted and reference segmentation masks by measuring the degree of overlap between the two regions. This metric is particularly suitable for medical image segmentation because it is sensitive to lesion boundary agreement and small segmented regions [33].

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (3)$$

where TP, FP, and FN denote true positive, false positive, and false negative pixels, respectively. Similar to IoU, Dice Score ranges from 0 to 1, with higher values indicating greater similarity between predicted and reference masks.

Precision and Recall were additionally used to provide a more detailed evaluation of segmentation performance, particularly in measuring the accuracy of detected lesion regions and the ability of the model to capture actual acne lesions. Precision measures the proportion of correctly predicted lesion pixels among all pixels predicted as lesion pixels. A high precision value indicates that the segmentation model produces fewer false positive pixels, meaning that most detected regions correspond to actual lesion areas [34].

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

where TP and FP denote true positive and false positive pixels, respectively. Precision values range from 0 to 1, where higher values indicate fewer incorrectly predicted lesion regions.

Recall measures the proportion of actual lesion pixels that are successfully identified by the model. This metric reflects the ability of the segmentation model to capture complete lesion regions and is particularly important when missing lesion areas (false negatives) should be minimized [34].

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

where TP and FN denote true positive and false negative pixels, respectively. Higher recall values indicate that the model is able to detect a larger proportion of the actual lesion regions.

III. RESULTS AND DISCUSSION

A. Performance Evaluation

The performance of the proposed methods was evaluated using the Dice coefficient, Intersection over Union (IoU), and validation loss. Since this study focuses on benchmarking pseudo-mask generation methods, the segmentation masks predicted by the ResNet34-U-Net model were evaluated against the corresponding pseudo-mask labels generated by each method rather than manually annotated ground truth masks. Accordingly, the Dice and IoU metrics quantify the agreement between the predicted segmentation masks and the generated pseudo-mask labels, while the validation loss reflects the optimization performance of the model during training. This evaluation protocol enables a fair comparison of the effectiveness of different pseudo-mask generation methods under identical experimental settings. The results are presented in Table 3.

TABLE 3
QUANTITATIVE PERFORMANCE EVALUATION

Method	Dice	IoU	Precision	Recall	Loss
Contour-Based	0.695	0.545	0.633	0.822	0.385
Superpixel (SLIC)	0.898	0.815	0.889	0.908	0.268
K-Means	0.755	0.608	0.747	0.780	0.390
WSSS	0.871	0.771	0.836	0.915	0.381
Pseudo-Mask Based	0.808	0.679	0.779	0.856	0.345

The results show that the Superpixel (SLIC) method achieves the best overall segmentation performance, obtaining a Dice score of 0.898, an IoU of 0.815, a Precision of 0.889, a Recall of 0.908, and the lowest validation loss of 0.268. These results indicate that SLIC generates high-quality pseudo-masks with strong spatial consistency and well-preserved lesion boundaries, providing reliable supervisory labels for U-Net training. The WSSS method, based on Otsu thresholding and morphological operations, also demonstrates competitive performance, achieving a Dice score of 0.871, an IoU of 0.771, a Precision of 0.836, and the highest Recall of 0.915. The high Recall suggests that WSSS successfully identifies most lesion pixels; however, its relatively lower Precision and higher loss indicate the presence of additional false-positive regions compared with the Superpixel approach.

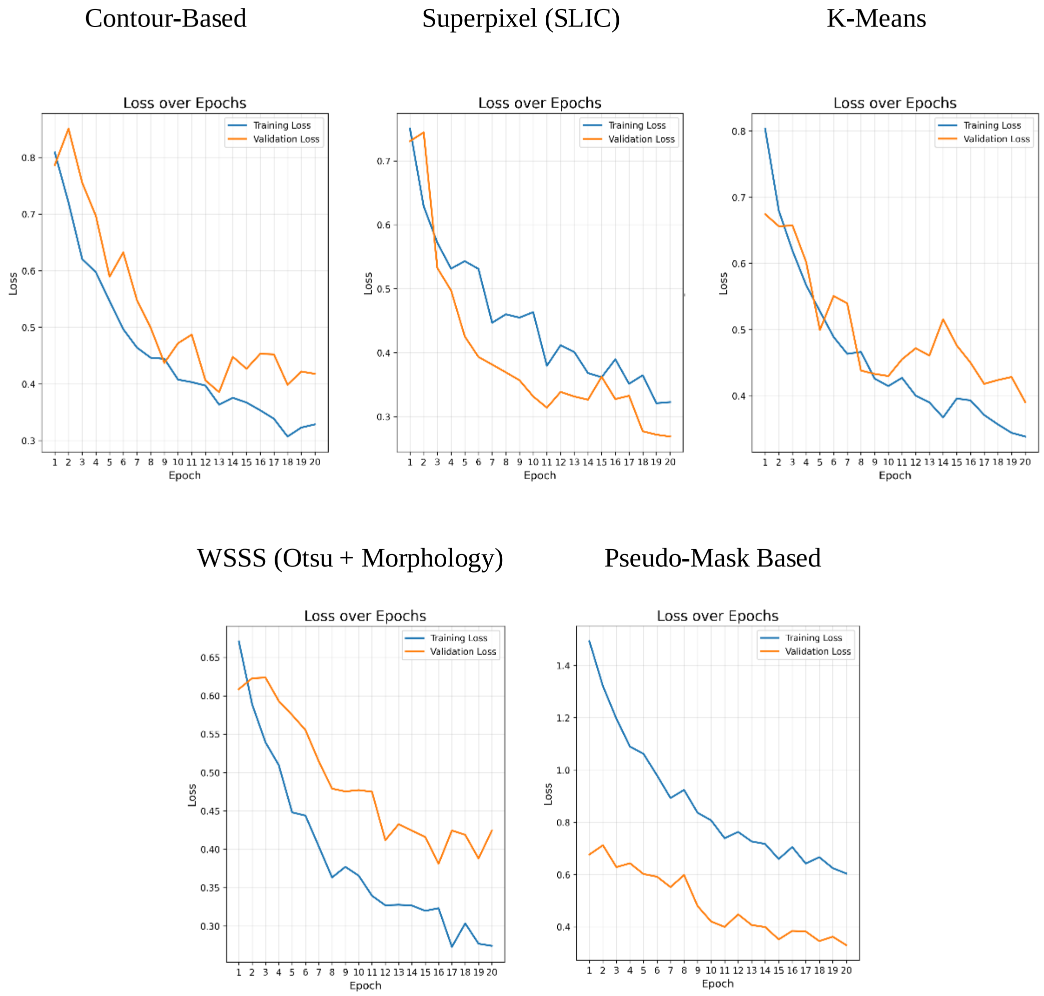
Among the remaining methods, the Pseudo-Mask Based method achieves a Dice score of 0.808, an IoU of 0.679, a Precision of 0.779, a Recall of 0.856, and a loss value of 0.345, demonstrating a balanced segmentation performance through its multi-stage preprocessing pipeline. The K-Means method follows with a Dice score of 0.755, an IoU of 0.608, a Precision of 0.747, a Recall of 0.780, and a loss value of 0.390, indicating that clustering-based pseudo-mask generation can effectively identify lesion regions based on color similarity, although its performance is limited by variations in lesion appearance and skin color. In contrast, the Contour-Based method produces the lowest overall performance, with a Dice score of 0.695, an IoU of 0.545, a Precision of 0.633, a Recall of 0.822, and a loss value of 0.385. This suggests that edge-based pseudo-mask generation is more susceptible to weak or discontinuous lesion boundaries, resulting in less accurate supervisory masks. Overall, the results demonstrate that pseudo-mask generation methods capable of preserving spatial consistency and region homogeneity, particularly the Superpixel (SLIC) approach, provide more reliable supervisory labels and consequently achieve superior U-Net segmentation performance. While these quantitative results indicate clear performance differences among the evaluated methods, it is necessary to determine whether the observed differences are statistically significant rather than arising from random variation. Therefore, statistical analyses were conducted using the Friedman test, followed by pairwise Wilcoxon signed-rank tests with Holm correction and rank-biserial correlation (RBC) effect size analysis.

The statistical significance of the segmentation performance differences among the five pseudo-mask generation methods was evaluated using the Friedman test based on the Dice scores obtained from the same ten test images. The Friedman test yielded a statistic of $x^2 = 10.80$ with a p-value of 0.0289, indicating a statistically significant overall difference among the evaluated methods. These results suggest that the choice of pseudo-mask generation method has an overall influence on the segmentation performance of the U-Net model.

Since the Friedman test indicated a significant overall difference, pairwise comparisons were subsequently performed using the Wilcoxon signed-rank test with Holm correction to control for multiple comparisons. Although several method pairs showed relatively small unadjusted p-values, none remained statistically significant after Holm correction (adjusted $p > 0.05$). This finding indicates that, while the methods differ collectively, no individual pair of methods demonstrated a statistically significant difference under the applied multiple-comparison criterion.

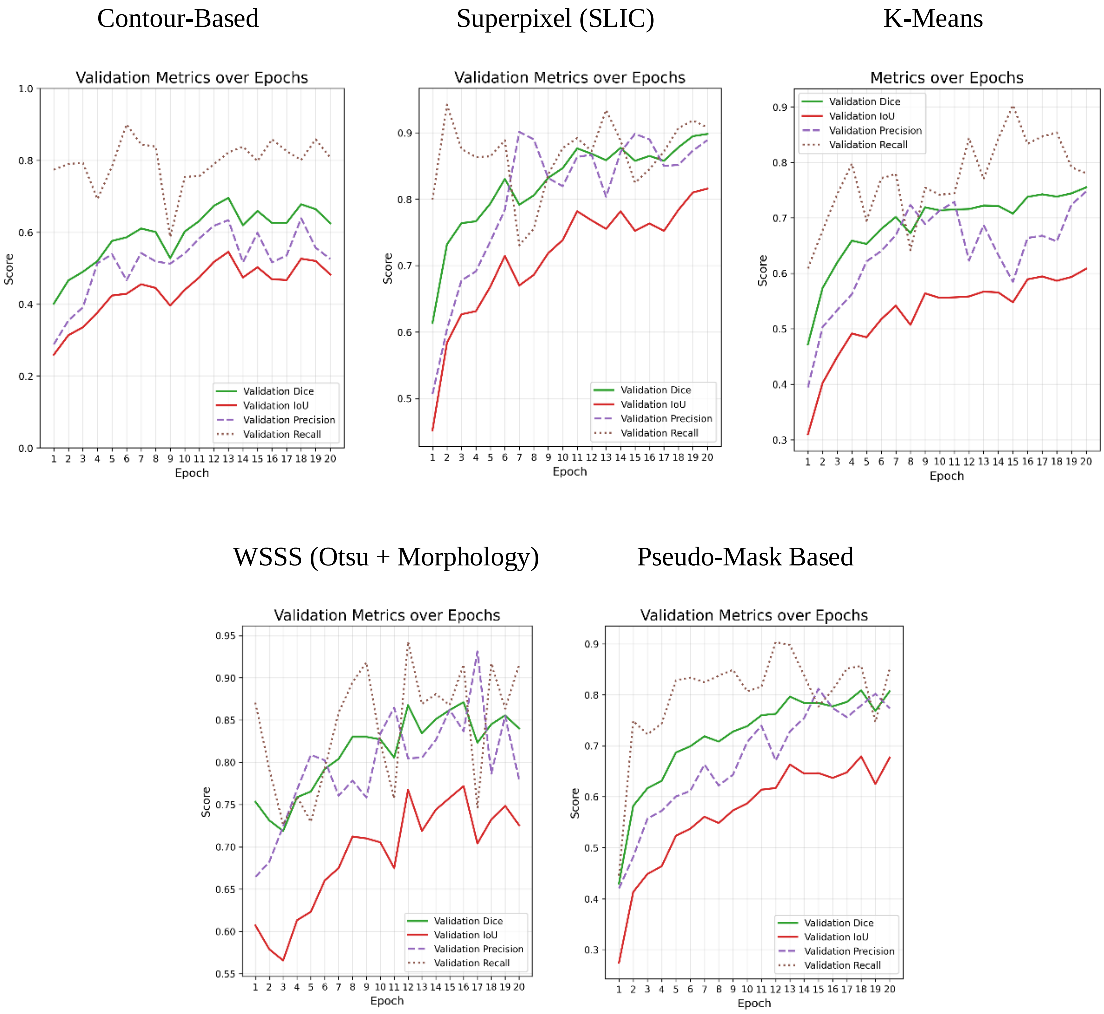
To complement the hypothesis testing, the rank-biserial correlation (RBC) was computed to estimate the magnitude of the pairwise differences. Several comparisons exhibited moderate to large effect sizes, including Contour-Based versus Pseudo-Mask ($|r| = 0.82$), Contour-Based versus WSSS ($|r| = 0.78$), K-Means versus Pseudo-Mask ($|r| = 0.71$), and Contour-Based versus Superpixel ($|r| = 0.64$). These effect sizes indicate that substantial practical differences exist among several pseudo-mask generation methods despite the absence of statistically significant pairwise differences after Holm correction. This outcome is likely influenced by the relatively small number of paired test images ($n = 10$), which limits the statistical power of pairwise non-parametric tests while still allowing the Friedman test to detect an overall difference among methods.

FIGURE 3
LOSS OVER EPOCHS VISUALIZATION



The training process of the U-Net model was further analyzed through the loss curves shown in Figure 3. Overall, all methods exhibit a decreasing trend in training loss, indicating that the U-Net model progressively learns meaningful representations from the generated pseudo masks. Among the evaluated methods, the Superpixel (SLIC) approach demonstrates the most stable convergence, with both training and validation losses decreasing consistently and reaching the lowest validation loss throughout training. The Contour-Based method also shows a continuous reduction in training loss; however, its validation loss fluctuates noticeably during the later epochs, suggesting less consistent generalization. The Pseudo-Mask Based approach exhibits a steady decrease in training loss accompanied by moderate fluctuations in validation loss, indicating reasonably stable learning despite minor variations. In contrast, WSSS (Otsu + Morphology) presents a relatively large gap between the training and validation loss curves after the initial epochs, suggesting a tendency toward overfitting. Meanwhile, the K-Means method demonstrates a gradual and relatively stable decrease in both training and validation losses, although its convergence remains slower than that of the Superpixel (SLIC) approach.

FIGURE 4
VALIDATION METRICS OVER EPOCHS VISUALIZATION



To evaluate segmentation performance throughout training, the progression of the validation Dice Score and IoU is presented in Figure 4. Among the evaluated methods, Superpixel (SLIC) consistently achieves the highest validation performance, maintaining the highest Dice and IoU values with stable convergence across epochs. WSSS (Otsu + Morphology) demonstrates the second-best performance, showing rapid improvement during the early training stages followed by relatively stable validation metrics throughout the remaining epochs. The K-Means approach exhibits steady improvements in both validation Dice and IoU, achieving intermediate segmentation performance with relatively stable convergence. Similarly, the Pseudo-Mask Based method shows continuous improvement in both metrics despite moderate fluctuations during training, although its overall validation performance remains lower than that of the Superpixel, WSSS, and K-Means approaches. In contrast, the Contour-Based method exhibits the lowest validation Dice and IoU values throughout most training epochs, indicating limited segmentation capability. Overall, the validation metric trends are consistent with the quantitative results presented in Table 3, where Superpixel (SLIC) achieves the best overall performance, followed by WSSS (Otsu + Morphology), K-Means, Pseudo-Mask Based, and Contour-Based segmentation.

Although segmentation accuracy is the primary criterion for evaluating pseudo-mask quality, computational efficiency is also an important consideration, particularly for large-scale datasets and practical clinical applications. Therefore, in addition to quantitative segmentation performance, this study further evaluates the computational efficiency of each pseudo-mask generation method by analyzing the pseudo-mask generation time and the estimated U-Net training time under identical experimental settings. The corresponding results are presented in Table 4.

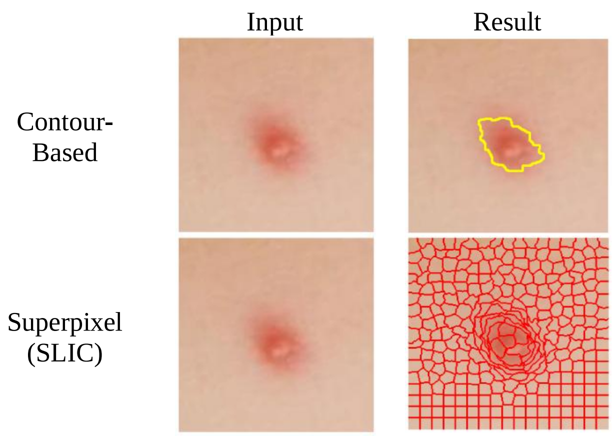
TABLE 4
QUANTITATIVE PERFORMANCE EVALUATION

Method	Pseudo Mask Generation Time (s/image)	Estimated U-Net Training Time (min)
Contour-Based	0.0598	19 min 29 s
Superpixel (SLIC)	0.2812	17 min 28 s
K-Means	0.0420	17 min 12 s
WSSS	0.0031	16 min 53 s
Pseudo-Mask Based	0.0041	11 min 1 s

Table 4 presents the pseudo-mask generation time and estimated U-Net training time for each method. All experiments were performed using Google Colab with a CPU runtime under the same training configuration. WSSS showed the fastest pseudo-mask generation time, while Superpixel (SLIC) required the longest processing time. These results indicate that computational efficiency should be considered together with segmentation accuracy when selecting a pseudo-mask generation method.

B. Comparative Analysis of Pseudo-Mask Generation Methods

The following subsection presents a comparative analysis of several pseudo-mask generation methods used in the segmentation process, including Contour-Based, Superpixel (SLIC), K-Means, WSSS (Otsu + Morphology), and Pseudo-Mask Based approaches. The analysis focuses on the ability of each method to capture lesion regions, preserve boundary details, and handle variations in skin appearance. Through qualitative visual assessment, the strengths and limitations of each approach are examined to provide a better understanding of their effectiveness in generating reliable pseudo-masks for subsequent U-Net segmentation.



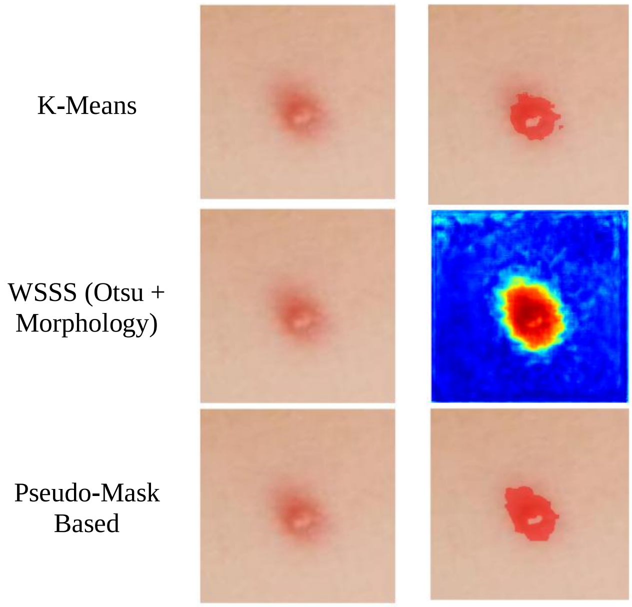


FIGURE 5
PSEUDO-MASK GENERATION VISUALIZATION

As shown in Figure 5, each pseudo-mask generation method produces distinct segmentation characteristics. The Contour-Based method generates coarse lesion boundaries that often fail to capture the complete lesion region, particularly in low-contrast areas. In contrast, the Superpixel (SLIC) method preserves lesion boundaries and local structures more effectively by grouping pixels with similar color and texture characteristics. The K-Means method produces relatively compact lesion masks but may include adjacent healthy skin due to its reliance on color similarity. Meanwhile, the WSSS (Otsu + Morphology) method generates smoother and more continuous lesion masks by combining adaptive thresholding with morphological refinement, whereas the Pseudo-Mask Based approach provides relatively good lesion coverage with minor boundary inaccuracies. Overall, the qualitative comparison shows that the Superpixel (SLIC) and WSSS (Otsu + Morphology) methods produce more accurate lesion localization, better boundary preservation, and fewer background artifacts, which is consistent with the quantitative results presented in the following section.

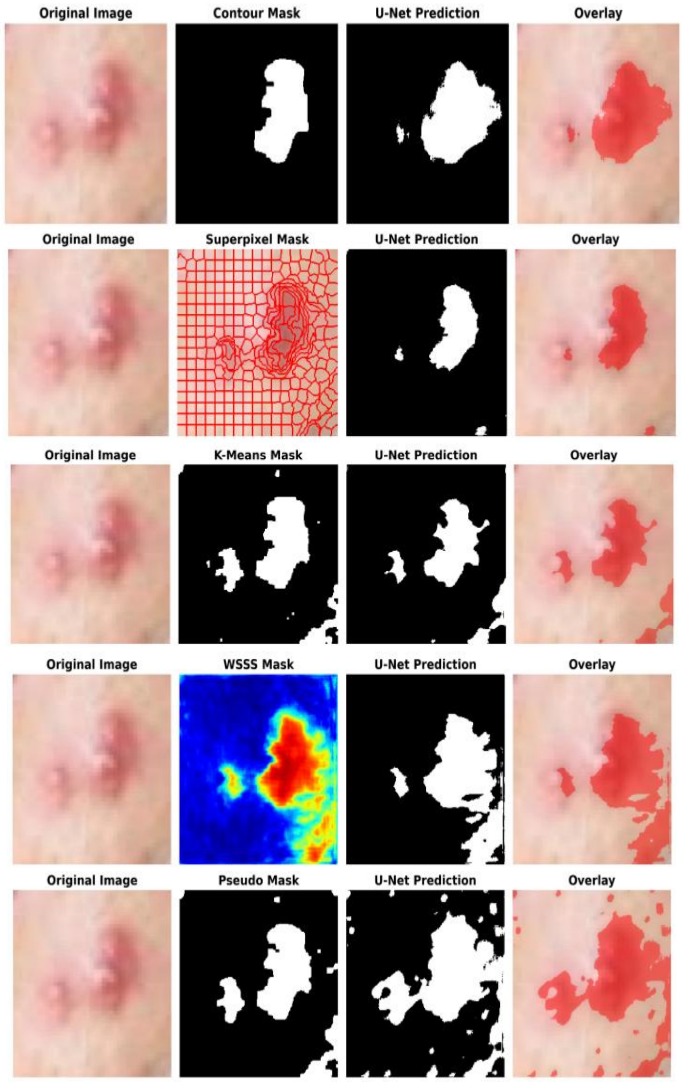


FIGURE 6
QUALITATIVE COMPARISON OF PSEUDO-MASK GENERATION METHODS AND THE CORRESPONDING RESNET34-U-NET SEGMENTATION RESULTS

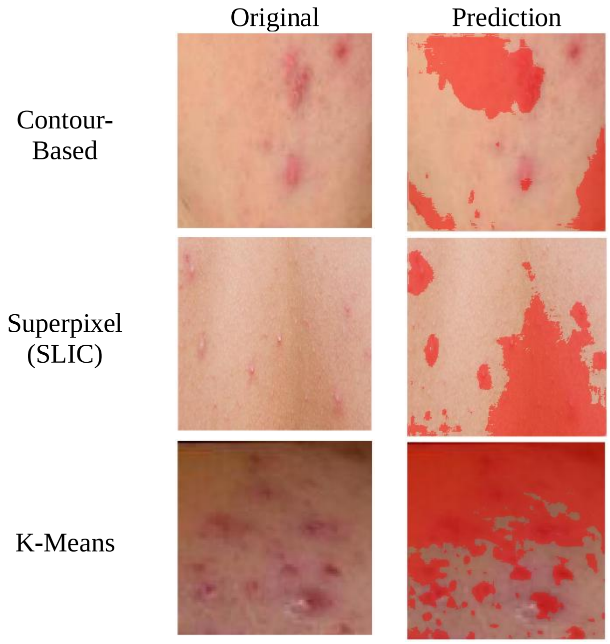
To further illustrate the differences among the pseudo-mask generation methods and their influence on the final segmentation performance, Figure 6 presents qualitative examples consisting of the original acne images, the pseudo masks generated by each method, and the corresponding segmentation results produced by the ResNet34-U-Net model. This visual comparison complements the quantitative evaluation by demonstrating how the quality of the generated pseudo masks affects lesion localization, boundary delineation, and the overall segmentation performance.

The Contour-Based method generates relatively simple binary masks that capture the primary lesion region but fail to preserve complete lesion boundaries, resulting in predictions that underestimate the lesion extent. In contrast, the Superpixel (SLIC) method preserves local image structures and lesion boundaries more effectively, enabling the U-Net model to produce well-localized segmentation results with fewer background artifacts. The K-Means approach successfully identifies the main lesion region; however, its reliance on color similarity introduces several non-lesion regions into the pseudo mask, leading to minor false-positive predictions in the final segmentation. The WSSS (Otsu + Morphology) method produces coarse activation-based pseudo masks that provide reliable lesion localization, allowing the U-Net model to capture the principal lesion area while maintaining relatively smooth and continuous boundaries. Nevertheless, slight over-segmentation is observed around the lesion margins. Finally, the Pseudo-Mask Based approach generates broader pseudo masks containing additional background regions despite the applied refinement procedures. Consequently, the corresponding U-

Net predictions also exhibit more extensive segmented areas and several false-positive regions surrounding the lesion.

The superior performance of the Superpixel (SLIC) method can be attributed to its ability to simultaneously consider color similarity and spatial proximity when grouping pixels into homogeneous regions. Unlike the Contour-Based method, which relies primarily on edge information, or K-Means clustering, which groups pixels mainly based on color similarity without enforcing spatial continuity, SLIC generates compact superpixels that closely adhere to natural lesion boundaries. This characteristic enables the generated pseudo-masks to preserve lesion shape and boundary information more consistently while reducing fragmented regions and image noise. As a result, the supervisory labels provided to the U-Net model more accurately represent the actual lesion morphology, allowing the network to learn more discriminative lesion features and ultimately achieve higher segmentation accuracy.

Overall, these findings indicate that preserving lesion boundaries and spatial continuity is a key factor in generating high-quality pseudo-masks for weakly supervised segmentation. Compared with the other evaluated methods, the Superpixel (SLIC) approach produces pseudo-masks that more closely resemble the actual lesion morphology, enabling the ResNet34-U-Net model to achieve more accurate lesion localization, better boundary delineation, and fewer false-positive regions. This observation is consistent with the quantitative evaluation, where the Superpixel (SLIC) method achieved the highest Dice Score, IoU, Precision, and the lowest validation loss among all evaluated pseudo-mask generation methods.



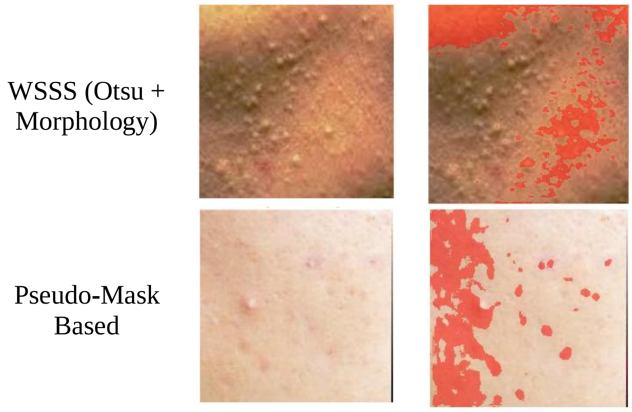


FIGURE 7
REPRESENTATIVE FAILURE CASES OF THE RESNET34-U-NET SEGMENTATION RESULTS USING DIFFERENT PSEUDO-MASK GENERATION METHODS

To complement the quantitative evaluation, Figure 7 presents representative failure cases observed for each pseudo-mask generation method. The results indicate that each method encounters different challenges depending on the visual characteristics of the acne lesions and surrounding skin. The Contour-Based method frequently fails when lesion boundaries exhibit low contrast or weak edge information, causing incomplete lesion localization and under-segmentation. As illustrated in the first example, only a small portion of the lesion is correctly identified, while several surrounding lesion regions are missed due to discontinuous edge detection.

The K-Means method demonstrates limitations when lesion and healthy skin exhibit similar color distributions. Because clustering relies primarily on pixel color similarity, several surrounding skin regions are incorrectly grouped as lesion areas, resulting in noticeable false-positive predictions. This limitation becomes more apparent in images containing uneven illumination or multiple acne lesions with varying color intensities, where the generated segmentation masks cover substantially larger regions than the actual lesions.

The WSSS (Otsu + Morphology) and Pseudo-Mask-Based methods generally succeed in identifying most lesion regions; however, both methods tend to produce over-segmentation under challenging lighting conditions. Since these approaches rely on intensity-based thresholding, neighboring skin regions with similar intensity values are occasionally included in the generated masks. Consequently, the predicted segmentation extends beyond the actual lesion boundaries, leading to additional false-positive regions despite achieving relatively high recall.

Although Superpixel (SLIC) achieved the highest overall segmentation performance, challenging cases remain. Small acne lesions with subtle color differences from the surrounding skin may not form independent superpixel regions, causing slight under-segmentation or fragmented lesion boundaries. Nevertheless, compared with the other evaluated methods, Superpixel better preserves local spatial continuity and lesion morphology, resulting in fewer false-positive regions and more accurate boundary delineation.

Overall, these failure cases indicate that lesion size, boundary contrast, illumination variation, and color similarity between lesions and healthy skin remain the primary challenges for automatic pseudo-mask generation. Future work may improve robustness by incorporating adaptive pseudo-label refinement or learning-based pseudo-mask generation techniques capable of better handling these complex dermatological image characteristics.

C. Implications for Acne Lesion Analysis

The findings of this study highlight the importance of pseudo-mask quality in acne lesion analysis. Methods that preserve spatial consistency and lesion boundary information, particularly Superpixel (SLIC), provide more reliable supervision for U-Net training and lead to more accurate lesion segmentation. Accurate segmentation is essential for acne analysis because it enables the model to focus on lesion regions while minimizing interference from surrounding skin texture, pores, illumination variations, and other background artifacts.

The results also demonstrate that high-quality pseudo-labels can reduce the dependence on manually annotated masks, which are often costly and time-consuming to obtain in dermatological applications. The strong performance achieved by the Superpixel (SLIC) and WSSS approaches suggests that automatically generated pseudo-masks can serve as effective alternatives for developing acne lesion segmentation systems. Therefore, these methods have the potential to support computer-aided dermatological analysis, treatment monitoring, and future acne assessment systems while reducing annotation effort and improving scalability.

In practical clinical settings, the proposed Superpixel (SLIC)-based pseudo-mask generation approach can be integrated into computer-aided dermatology systems as an automatic pre-annotation tool. Instead of manually delineating lesion boundaries for every image, dermatologists can review and refine automatically generated pseudo-masks, substantially reducing annotation time while maintaining clinically meaningful lesion localization. This semi-automatic workflow can accelerate the development of annotated dermatological datasets, facilitate the training of deep learning models, and improve the scalability of acne analysis systems. Furthermore, accurate lesion segmentation can serve as a preprocessing stage for downstream tasks such as acne severity assessment, lesion counting, treatment response monitoring, and automated clinical decision support.

IV. CONCLUSION

This study presented a comparative evaluation of five pseudo-mask generation methods for ResNet34-U-Net-based acne lesion segmentation, namely Contour-Based Segmentation, Superpixel (SLIC), K-Means, WSSS (Otsu + Morphology), and Pseudo-Mask-Based Segmentation. The experimental results demonstrated that pseudo-mask quality has a significant impact on segmentation performance. Among the evaluated methods, Superpixel (SLIC) achieved the best overall performance, obtaining the highest Dice Score (0.9273) and IoU (0.8657), indicating superior lesion boundary preservation and spatial consistency. WSSS (Otsu + Morphology) also produced competitive segmentation results, whereas Pseudo-Mask-Based Segmentation showed satisfactory performance despite relying on automatically generated labels. In contrast, K-Means and Contour-Based Segmentation yielded lower segmentation accuracy due to their limited ability to capture the complex characteristics of acne lesions.

Overall, the findings demonstrate that pseudo-mask generation methods capable of preserving spatial coherence and lesion boundary information provide more effective supervision for ResNet34-U-Net training. This study confirms that high-quality pseudo masks can serve as reliable surrogate annotations, reducing the dependence on manual pixel-level labeling while maintaining accurate segmentation performance. Future work may focus on validating the proposed approach using larger and more diverse dermatological datasets, investigating more advanced weakly supervised learning strategies, and exploring recent segmentation architectures to further improve acne lesion segmentation performance.

IV. LIMITATIONS AND FUTURE WORK

This study has several limitations that should be considered. First, an ablation study was not conducted to evaluate the individual contribution of each preprocessing component and data augmentation strategy. This study focuses on benchmarking and comparing different pseudo-mask generation methods rather than optimizing a single segmentation pipeline. Therefore, the analysis of the contribution of individual preprocessing components and the influence of augmentation strategies are beyond the scope of this research. Future studies may perform systematic ablation experiments to investigate the effect of each component on segmentation performance.

Second, this study does not include comparisons with fully supervised segmentation approaches due to the absence of manually annotated pixel-level ground truth masks in the utilized dataset. Since the generated pseudo-masks were used as surrogate supervisory labels, direct evaluation against models trained using expert annotations could not be performed. Future work will involve comparing pseudo-mask-based segmentation approaches with fully supervised methods using datasets containing manually annotated lesion masks.

Third, all pseudo-mask generation methods in this study were evaluated using a ResNet34-U-Net architecture as the segmentation backbone. Therefore, the consistency of the benchmarking results when using other segmentation architectures, such as DeepLabV3+, U-Net++, or SegFormer, remains unexplored. Future research should investigate whether the observed performance differences among pseudo-mask generation methods are maintained across various segmentation architectures.

Fourth, this study benchmarked representative conventional pseudo-mask generation methods and did not include more advanced approaches, such as GrabCut, Conditional Random Fields (CRF) refinement, Segment Anything Model (SAM), or CAM-based pseudo-labeling. Consequently, the comparative performance of these recent pseudo-mask generation techniques remains unexplored. Future studies may extend the proposed benchmarking framework by incorporating these advanced methods to provide a more comprehensive evaluation of pseudo-mask generation strategies.

Finally, this study has not fully investigated the limitations of pseudo-mask generation under diverse clinical conditions, including variations in illumination, skin color, lesion size, and complex dermatological appearances. These factors may affect the accuracy and consistency of automatically generated pseudo-masks. Future studies should evaluate the robustness of pseudo-mask generation methods using more diverse datasets representing broader clinical variations.

REFERENCES

[1] Z. Zhu *et al.*, "Global, regional and national burdens of acne vulgaris in adolescents and young adults aged 10–24 years from 1990 to 2021: a trend analysis," *Br. J. Dermatol.*, vol. 192, no. 2, pp. 228–237, Jan. 2025, doi: 10.1093/bjd/ljae352.

[2] A. Wianti, H. Heni, R. Setyowati, R. Nuraeni, and A. Setyawati, "Adolescents and Early Adults With Acne Vulgaris: a Scoping Review on Psychological and Psychosocial Impacts and Quality of Life," *J. Matern. Care Reprod. Heal.*, vol. 6, no. 3, pp. 194–208, 2024, doi: 10.36780/jmcrh.v6i3.12267.

[3] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3059968.

[4] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET Image Process.*, vol. 16, no. 5, pp. 1243–1267, 2022, doi: 10.1049/ipr2.12419.

[5] P. Hiskiawan, Y. M. Geasela, H. Heryanto, E. Stephanie, M. Ardianti, and F. A. Sukarno, "Trustworthy Data Science Framework for Non-Invasive Nutritional Screening Using Computer Vision," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Dec. 2025, pp. 1743–1748. doi: 10.1109/ICIMCIS68501.2025.11327268.

[6] P. Hiskiawan, T. P. Wijayanti, H. Heryanto, S. Everlin, S. A. Yasodhara, and D. Alexander, "Mel-Frequency Cepstral Coefficients and Neural Networks for Indonesian Traditional Music Recognition," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Dec. 2025, pp. 1707–1712. doi: 10.1109/ICIMCIS68501.2025.11327436.

[7] J. Li *et al.*, "Development and validation of an artificial intelligence-powered acne grading system incorporating lesion identification," *Front. Med.*, vol. 10, no. October, pp. 1–8, 2023, doi: 10.3389/fmed.2023.1255704.

[8] V. Venugopal, M. K. Nath, J. Joseph, and M. V. Das, "A deep learning-based illumination transform for devignetting photographs of dermatological lesions," *Image Vis. Comput.*, vol. 142, p. 104909, Feb. 2024, doi: 10.1016/j.imavis.2024.104909.

[9] M. Benčević, M. Habijan, I. Galić, D. Babin, and A. Pižurica, "Understanding skin color bias in deep learning-based skin lesion segmentation," *Comput. Methods Programs Biomed.*, vol. 245, no. November 2023, 2024, doi: 10.1016/j.cmpb.2024.108044.

[10] Z. He, X. Li, Y. Chen, N. Lv, and Y. Cai, "Attention-based dual-path feature fusion network for automatic skin lesion segmentation," *BioData Min.*, vol. 16, no. 1, pp. 1–22, 2023, doi: 10.1186/s13040-023-00345-x.

[11] L. Ran, Y. Li, G. Liang, and Y. Zhang, "Pseudo Labeling Methods for Semi-Supervised Semantic Segmentation: A Review and Future Perspectives," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 4, pp. 3054–3080, Apr. 2025, doi: 10.1109/TCSVT.2024.3508768.

[12] Y. Yu *et al.*, "Techniques and Challenges of Image Segmentation: A Review," *Electronics*, vol. 12, no. 5, p. 1199, Mar. 2023, doi: 10.3390/electronics12051199.

[13] M. Lee, S. Lee, J. Lee, and H. Shim, "Saliency as Pseudo-Pixel Supervision for Weakly and Semi-Supervised Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 12341–12357, Oct. 2023, doi: 10.1109/TPAMI.2023.3273592.

[14] J. Feng, X. Wang, T. Li, S. Ji, and W. Liu, "Weakly-supervised semantic segmentation via online pseudo-mask correcting," *Pattern Recognit. Lett.*, vol. 165, pp. 33–38, Jan. 2023, doi: 10.1016/j.patrec.2022.11.024.

[15] D. O. Traini, G. Palmisano, C. Guerriero, and K. Peris, "Artificial Intelligence in the Assessment and Grading of Acne Vulgaris: A Systematic Review," *J. Pers. Med.*, vol. 15, no. 6, pp. 1–13, 2025, doi: 10.3390/jpm15060238.

[16] K. Yojana and L. Thillai Rani, "OCT layer segmentation using U-NET semantic segmentation and RESNET34 encoder-decoder," *Meas. Sensors*, vol. 29, no. June, p. 100817, 2023, doi: 10.1016/j.measen.2023.100817.

[17] M. Moncho-Santonja, S. Aparisi-Navarro, B. Defez, and G. Peris-Fajarnés, "Segmentation of Acne Vulgaris Images Techniques: A Comparative and Technical Study," *Appl. Sci.*, vol. 13, no. 10, 2023, doi: 10.3390/app13106157.

[18] C. Yuan, D. Zhao, and S. S. Agaian, "MUCM-Net: a Mamba powered UCM-Net for skin lesion segmentation," *Explor. Med.*, vol. 5, pp. 694–708, 2024, doi: 10.37349/emed.2024.00250.

[19] L. Cai, K. Hou, and S. Zhou, "Intelligent skin lesion segmentation using deformable attention Transformer U-Net with bidirectional attention mechanism in skin cancer images," *Ski. Res. Technol.*, vol. 30, no. 8, pp. 1–13, 2024, doi: 10.1111/srt.13783.

[20] N. Lama, J. Hagerty, A. Nambisan, R. J. Stanley, and W. Van Stoecker, "Skin Lesion Segmentation in Dermoscopic Images with Noisy Data," *J. Digit. Imaging*, vol. 36, no. 4, pp. 1712–1722, Apr. 2023, doi: 10.1007/s10278-023-00819-8.

[21] M. Tamoor, A. Naseer, A. Khan, and K. Zafar, "Skin Lesion Segmentation Using an Ensemble of Different Image Processing Methods," *Diagnostics*, vol. 13, no. 16, pp. 1–14, 2023, doi: 10.3390/diagnostics13162684.

[22] A. K. Nugroho, Afiahayati, M. E. Wibowo, and H. Soebono, "The Effect of Preprocessing on Skin Lesion Segmentation," in *2024 8th International Conference on Information Technology (InCIT)*, IEEE, Nov. 2024, pp. 455–460. doi: 10.1109/InCIT63192.2024.10810559.

[23] S. Joseph and O. O. Olugbara, "Preprocessing Effects on Performance of Skin Lesion Saliency Segmentation," *Diagnostics*, vol. 12, no. 2, 2022, doi: 10.3390/diagnostics12020344.

[24] S. Jardim, J. António, and C. Mora, "Graphical Image Region Extraction with K-Means Clustering and Watershed," *J. Imaging*, vol. 8, no. 6, pp. 1–27, 2022, doi: 10.3390/jimaging8060163.

[25] K. Zhu, N. N. Xiong, and M. Lu, "A Survey of Weakly-supervised Semantic Segmentation," in *2023 IEEE 9th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, IEEE, May 2023, pp. 10–15. doi: 10.1109/BigDataSecurity-HPSC-IDS58521.2023.00013.

[26] J. Peng and Y. Wang, “Medical image segmentation with limited supervision: A review of deep network models,” *IEEE Access*, vol. 9, pp. 36827–36851, 2021, doi: 10.1109/ACCESS.2021.3062380.

[27] J. Su, Z. Luo, S. Lian, D. Lin, and S. Li, “Mutual learning with reliable pseudo label for semi-supervised medical image segmentation,” *Med. Image Anal.*, vol. 94, p. 103111, May 2024, doi: 10.1016/j.media.2024.103111.

[28] P. Hiskiawan, M. K. Sari, R. E. Siregar, N. A. Valencia, and T. P. Wijayanti, “A Deep Learning Data Fusion Approach for Prostate MRI Zonal Segmentation Using Dual-Channel U-Net with T2 and ADC Images,” in *2026 International Conference on Current Research in Artificial Intelligence and Data Science (ICCRAIDS)*, IEEE, Apr. 2026, pp. 1–7. doi: 10.1109/ICCRAIDS67816.2026.11519671.

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[30] P. Hiskiawan, C.-C. Chen, and Z.-K. Ye, “Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures,” *Arab. J. Geosci.*, vol. 16, no. 10, p. 581, Oct. 2023, doi: 10.1007/s12517-023-11690-w.

[31] H. Liu *et al.*, “MEA-Net: multilayer edge attention network for medical image segmentation,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–15, 2022, doi: 10.1038/s41598-022-11852-y.

[32] Z. Wang *et al.*, “Revisiting Evaluation Metrics for Semantic Segmentation: Optimization and Evaluation of Fine-grained Intersection over Union,” *Adv. Neural Inf. Process. Syst.*, vol. 36, no. NeurIPS, pp. 1–82, 2023.

[33] M. Yeung, L. Rundo, Y. Nan, E. Sala, C. B. Schönlieb, and G. Yang, “Calibrating the Dice Loss to Handle Neural Network Overconfidence for Biomedical Image Segmentation,” *J. Digit. Imaging*, vol. 36, no. 2, pp. 739–752, 2023, doi: 10.1007/s10278-022-00735-3.

[34] P. Fränti and R. Mariescu-Istodor, “Soft precision and recall,” *Pattern Recognit. Lett.*, vol. 167, pp. 115–121, 2023, doi: 10.1016/j.patrec.2023.02.005.