applied sciences

MDPI

Article

Automated Segmentation of Psoriasis in Uncontrolled Environments Using a Three-Class Ensemble Architecture

Robertas Audinys [1], Vaiva Paskeviciute [1], Vidas Raudonis [1,2,*], Linas Eidimtas [3], Dominyka Stragyte [4] and Skaidra Valiukeviciene [4]

[1] Automation Department, Electrical and Electronics Engineering Faculty, Kaunas University of Technology, 51367 Kaunas, Lithuania; vaiva.paskeviciute@ktu.edu (V.P.)
[2] Artificial Intelligence Center, Kaunas University of Technology, 51367 Kaunas, Lithuania
[3] "Dts Solutions" UAB, Naujoji Str. 2, 60180 Raseiniai, Lithuania; linas.eidimtas@dts-solutions.lt
[4] Department of Skin and Venereal Diseases, Lithuanian University of Health Sciences, 50161 Kaunas, Lithuania; dominyka.stragyte@lsmu.lt (D.S.); skaidra.valiukeviciene@lsmu.lt (S.V.)
* Correspondence: vidas.raudonis@ktu.lt

Abstract

Psoriasis is a heterogeneous inflammatory skin disease requiring continuous monitoring to assess treatment efficacy. Automated lesion segmentation remains a significant computer vision challenge due to irregular plaque boundaries, variable skin tones, and uncontrolled lighting conditions in clinical photography. This study proposes a robust hybrid deep learning framework for the automated segmentation of psoriatic lesions in unconstrained environments. We constructed a unique dataset utilizing a hierarchical three-class labeling scheme (psoriatic plaque, healthy skin, and background) to mitigate the class imbalance and background noise often found in binary segmentation tasks. Following a systematic hyperparameter optimization using the Optuna framework, three distinct architectures—DeepLabV3+, UperNet, and SegFormer—were identified as optimal. A novel ensemble architecture was then developed to integrate the high sensitivity of DeepLabV3+, the precision of UperNet, and the contextual balance of SegFormer via a conflict-resolution voting algorithm. Experimental results demonstrate that the proposed hybrid model outperforms individual state-of-the-art architectures, achieving a Dice coefficient of 89.3% for lesion segmentation and an F1 score of 90.7% across skin classes. These findings confirm the system's adaptability to real-world imaging conditions, validating its potential as an objective decision-support tool for dermatological practice.

Keywords: psoriasis; semantic segmentation; ensemble learning; medical image analysis; deep learning; automated diagnosis

Academic Editor: Thomas Lindner

Received: 2 February 2026
Revised: 23 February 2026
Accepted: 26 February 2026
Published: 2 March 2026

1. Introduction

Psoriasis is a chronic autoimmune skin disease characterized by the rapid acceleration of skin cell growth, causing cells to build up quickly on the surface. This accumulation results in the formation of raised, red patches known as plaques, which are typically covered with distinct silvery-white scales [1]. Unlike the diffuse borders often seen in eczema, psoriatic lesions usually present with sharp, well-defined boundaries that clearly separate the affected tissue from healthy skin. The condition predominantly affects extensor surfaces such as the knees, elbows, and scalp, rather than the skin folds commonly affected by atopic dermatitis. Patients often experience significant discomfort, including itching and burning, which can fluctuate in severity due to environmental triggers or stress. Psoriasis

affects approximately 125 million people globally, which is about 2–3% of the world's population [2]. While there is no permanent cure, continuous monitoring and targeted therapies can effectively manage the inflammation and reduce the visibility of lesions.

Monitoring psoriasis-affected skin areas is an essential part of disease management, as it allows clinicians to assess healing progress and evaluate the suitability of prescribed treatments. In clinical practice, this assessment is mainly performed through visual inspection by experienced dermatology specialists. Such evaluations are time-consuming, require significant clinical expertise, and may vary between observers, especially when monitoring is performed over long periods or under uncontrolled conditions. As a result, objective and repeatable assessment of disease progression remains a challenge. Automated image-based analysis methods offer a promising solution to these limitations. By reducing reliance on manual visual assessment, such methods could decrease the workload of medical professionals and provide more consistent evaluation of psoriasis severity and progression. Recent advances in computer vision and machine learning have enabled increasingly accurate analysis of medical images. However, despite the rapid development of this field, relatively few studies focus specifically on the segmentation of psoriasis-affected skin regions. From a computer vision perspective, the segmentation of psoriasis-affected skin areas is particularly challenging. Lesions often exhibit irregular shapes, distinct borders, and large variations in color and texture due to the presence of thick, silvery scales overlaying red plaques compared to surrounding healthy skin (see Figure 1). In addition, differences in illumination, skin tone, and background complexity further complicate accurate delineation of affected regions. These characteristics make psoriasis segmentation significantly more difficult than the segmentation of other skin lesions with smoother surface structures, such as melanoma.

Figure 1. Sample images from the dataset.

In this study, different machine learning-based segmentation methods are investigated for identifying psoriasis-affected skin regions in images acquired under uncontrolled conditions. A dedicated dataset was prepared for this purpose, and multiple segmentation architectures were evaluated and compared. The aim of this work is to improve the accuracy and robustness of psoriasis segmentation and to propose a practical approach that could support dermatology specialists in long-term monitoring of affected skin areas.

2. Materials and Methods

2.1. Computer Vision Methods for Skin Image Segmentation

Image segmentation plays a particularly important role in medical applications, where it is used to detect and isolate tissue abnormalities, tumors, or other pathological changes. In dermatology, segmentation methods enable the identification of affected skin regions

and support quantitative analysis of disease extent. Modern segmentation algorithms are capable of processing images with high accuracy and speed, making them an essential component of advanced medical image analysis systems. Image segmentation is a computer vision task that partitions a digital image into distinct groups of pixels, referred to as segments. Each segment represents a region with similar visual properties and typically corresponds to a specific object or object part within the image. By decomposing complex images into structured regions, segmentation enables more effective image analysis and facilitates the extraction of relevant information. In medical diagnostics and clinical image analysis, segmentation techniques are used to identify and quantify pathological regions in medical images [3,4]. In robotic and machine vision systems, segmentation supports object recognition and scene understanding, enabling more reliable interaction with the environment [5,6]. Vision systems for autonomous vehicles rely on segmentation to distinguish road elements, obstacles, and dynamic objects, ensuring safe navigation [7,8]. Additionally, image segmentation is commonly used in satellite imagery analysis to identify and extract dominant objects and land-cover structures from large-scale remote sensing data [9–11].

Computer vision commonly distinguishes three main image segmentation types: semantic, instance, and panoptic. These approaches differ in the level of contextual detail provided. Semantic segmentation, the most basic form, assigns a class label to every pixel without distinguishing between individual objects of the same category. All pixels of a single class are grouped into one mask, which effectively separates classes but cannot resolve distinct instances located near each other. Instance segmentation extends this by identifying specific objects within a class, generating unique masks for each object. This allows the model to separate overlapping objects, making the task significantly more complex. Panoptic segmentation unifies these methods, assigning a class label to every pixel while distinguishing instances to provide a holistic scene understanding. However, integrating semantic and instance-level data remains computationally demanding [5]. Earlier techniques, including thresholding, edge-based, region-based, and clustering methods, are effective for simple and controlled scenarios but typically fail to generalize to complex real-world images. Advances in deep learning have led to the development of fast and highly accurate segmentation models capable of handling large variability in visual data. In particular, convolutional neural networks (CNNs) have become the dominant approach due to their ability to automatically learn hierarchical feature representations and achieve high segmentation accuracy.

Current research, therefore, focuses on the analysis and improvement of AI-based segmentation methods. These methods are commonly categorized into several groups, such as fully convolutional networks (FCNs), region proposal–based models such as R-CNN, convolutional autoencoders, convolutional models combined with graphical models, adversarial learning approaches, recurrent neural networks, and attention-based models [12].

Previous research indicates that simple single-stage segmentation models often perform poorly when lesions occupy small regions of the image or lack clear structural features. In contrast, two-stage architectures, which first detect potential lesion regions and then perform fine-grained segmentation, have demonstrated improved performance for well-structured lesions. One representative example is the R-CNN-based approach, where detected regions are cropped from high-resolution images and subsequently segmented [13,14]. When applied to melanoma segmentation using combinations of Mask R-CNN and DeepLab-based models, such methods achieved relatively high Jaccard index values. However, these approaches are not well-suited for psoriasis segmentation, as the disease does not exhibit a consistent shape or structure. As a result, the detection stage becomes unreliable, leading to inaccurate segmentation and reduced overall performance.

Several studies have attempted to address the challenges of psoriasis segmentation using encoder–decoder architectures such as U-Net [15,16]. Experimental results indicate that model performance can vary significantly depending on the input representation. In particular, experiments using individual color channels demonstrated that different channels emphasize different lesion characteristics. While acceptable segmentation accuracy was reported in these studies, the evaluated images typically contained lesions with clear contrast and well-defined boundaries. Such conditions do not fully represent real-world clinical scenarios, where psoriasis lesions often exhibit gradual transitions between affected and healthy skin, making segmentation considerably more difficult.

Another critical challenge identified in the literature is the quality and consistency of annotated datasets. High-performance segmentation models require accurate expert annotations. However, in medical imaging, boundary definitions often vary between specialists due to subjective interpretation [17]. This inconsistency negatively affects model training and evaluation. In the clinical context, both precision and sensitivity are essential. False positive predictions may incorrectly label healthy skin as diseased, while false negatives may lead to missed lesions, potentially affecting clinical decisions and patient outcomes. For this reason, several studies suggest that combining multiple models into a hybrid or ensemble system can improve robustness and balance between these metrics [18].

Hybrid segmentation approaches have been successfully applied in skin lesion analysis, particularly for melanoma. For example, Goyal et al. proposed a hybrid architecture combining Mask R-CNN and DeepLabV3+, where segmentation outputs from both models were merged and refined using morphological post-processing techniques [19]. This approach improved segmentation accuracy and Jaccard index values compared to individual models. Similar conclusions were reported by Wang et al., who combined VGG-based U-Net architectures with DeepLabV3+ in a parallel configuration, averaging the outputs of both models to obtain the final segmentation result [20]. More complex ensemble systems, such as DivergentNets, further demonstrate that combining multiple architectures can significantly enhance segmentation accuracy by leveraging complementary feature representations [21–24].

2.2. Data Preparation and Tools

The experimental workflow for psoriasis segmentation was designed to ensure reliable and reproducible results. The main steps included the preparation, processing, and augmentation of input images, followed by the implementation and training of selected deep learning models, and finally the evaluation of segmentation performance. The dataset consisted of images of psoriasis-affected skin acquired under uncontrolled conditions, reflecting real-world variability in lesion appearance, skin tone, and lighting conditions. Preprocessing steps involved resizing, normalization, and optional selection of specific color channels to enhance model performance. All programming tasks were performed using Python 3.11. Deep learning models were implemented and trained using the PyTorch (ver. 2.1.1) framework, with additional libraries, such as PyTorch (ver. 2.7.1+cu118), Torchvision (ver. 0.22.1+cu118), NumPy (ver. 2.2.1), OpenCV (ver. 4.10.0.84), Pandas (ver. 2.2.3), Matplotlib (ver. 3.8.2), and Optuna (ver. V5) for hyperparameter optimization. Selected models were trained using CUDA on an NVIDIA GPU for accelerated training and large-scale computations. The hardware configuration for the experiments included an Intel® Core™ i5 processor (13th generation or higher, 2.5 GHz), 32 GB of RAM, and an NVIDIA GeForce RTX GPU (40 series or higher, 2460 MHz, 8 GB VRAM) running Windows 11 Home, with at least 80 GB of free storage.

2.3. Evaluation Metrics and Loss Functions

In this research, different segmentation models and their variations were evaluated to assess performance and enable comparisons between algorithms. While accuracy is a common metric in deep learning applications, it is not sufficient for segmentation tasks with imbalanced classes, such as medical images. Therefore, the evaluation focused on more informative metrics, such as F1-score, Intersection over Union (IoU), and Dice coefficient. The proposed framework performs a three-class segmentation (psoriasis, healthy skin, and background), the evaluation metrics are computed on a per-class basis using the class index c. The class-specific F1-score represents the harmonic mean of precision and recall for that specific class, calculated as:

$$\mathrm{F1}_c = 2(\mathrm{Precision}_c \times \mathrm{Recall}_c)/(\mathrm{Precision}_c + \mathrm{Recall}_c), \quad (1)$$

where $\mathrm{Precision}_c$ is the proportion of correctly predicted positive pixels for class c, and Recall_c is the proportion of true positive pixels identified by the model for class c. F1-score is particularly useful when class distribution is imbalanced, as it accounts for both the correctness and completeness of predictions. Evaluating the overlap between predicted masks and ground truth masks is essential in the segmentation task. The class-specific intersection over union (IoU), also known as the Jaccard index, is defined as:

$$\mathrm{IoU}_c = |\mathrm{P}_c \cap \mathrm{T}_c|/|\mathrm{P}_c \cup \mathrm{T}_c|, \quad (2)$$

where P_c represents the number of pixels for class c, T_c is the number of pixels in the ground truth for class c. IoU provides a strict measure of mask similarity, reflecting the proportion of correctly segmented regions. The Dice coefficient is widely used in semantic segmentation, especially for small or irregular objects such as psoriasis lesions. The Dice score for class c is calculated as:

$$\mathrm{Dice}_c = 2 \times |\mathrm{P}_c \cap \mathrm{T}_c|/(|\mathrm{P}_c| + |\mathrm{T}_c|). \quad (3)$$

Dice emphasizes the overlap between predicted and actual regions and is less sensitive to small mismatches than IoU, making it suitable for fragmented or small lesions. For the final system evaluation, these metrics are reported specifically for the target pathology (psoriasis class), as well as utilizing weighted averages across classes to reflect clinical priorities.

The cross-entropy loss is commonly employed for pixel-wise classification and was utilized in this research to calculate the difference between predicted probabilities and true labels for every pixel. While this method provides granular supervision at the pixel level, it is known to be sensitive to class imbalance, which is a frequent challenge in medical image segmentation. To mitigate this limitation, region-level loss functions are introduced to focus on the overall accuracy of the predicted mask rather than individual points. A prominent example is the Dice loss, derived from the Dice coefficient, which assesses the overlap between predicted regions and ground truth labels while incorporating class weights to better handle imbalanced data. A combined loss function was implemented to adopt the strengths of both methodologies. This hybrid strategy integrates cross-entropy loss with Dice loss, using a specific parameter to balance the relative contribution of each component. This combination enables the model to produce detailed masks while effectively addressing class imbalance, ultimately ensuring stable gradients and delivering improved segmentation performance [25,26].

2.4. Dataset

Research on psoriasis lesion segmentation is limited compared to other skin conditions, such as melanoma. While large datasets and challenge platforms exist for melanoma (e.g., ISIC—International Skin Imaging Collaboration), no publicly available, standardized dataset exists for psoriasis suitable for deep learning applications. Therefore, this study utilizes a unique dataset of anonymized patient images collected by dermatologists at the Lithuanian University of Health Sciences. The dataset contains images with lesions of varying severity, from small, scattered plaques to highly visible lesions, covering multiple body areas, including head, neck, limbs, and feet. This variability poses a challenge for lesion localization, particularly for high-resolution images, which standard segmentation models may not process in a single pass. Figure 1 shows sample images from the dataset.

The study included digital images from 258 patients. The demographic distribution consisted of 163 females (63.2%) and 95 males (36.8%). The median age of the cohort was 57 years (interquartile range [IQR]: 44–66), with an overall age range spanning from 15 to 94 years. All patients included in the study presented with Fitzpatrick skin types I–II. The dataset encompassed various clinical forms of the disease and corresponding body regions, such as common psoriasis (L40.0) predominantly affecting extensor surfaces such as elbows and knees, as well as the scalp (n = 238), palmar-plantar psoriasis (L40.3) specifically affecting the palms and soles (n = 11), and other psoriasis forms (L40) (n = 9). Image acquisition was conducted following a standardized clinical photo-documentation protocol under automatic shooting modes. The setup variability was controlled by using two primary imaging devices, such as a specialized Fotofinder system (FotoFinder Systems GmbH, located in the state of Bavaria, Germany) (n = 110) and a DSLR Nikon 5300 (Nikon Corporation, Minato City, Tokyo, Japan) camera (n = 148). The multi-tier annotation protocol was employed to ensure the highest quality of ground truth generation. The images were initially evaluated and annotated by specially trained medical students with a focused interest in dermatology. Instead of relying on a standalone statistical agreement measure, the accuracy of these initial annotations was rigorously double-checked and validated by a consensus panel consisting of two senior dermatology residents and one experienced dermatologist.

The dataset comprises 589 images accompanied by manually annotated masks indicating lesion regions (see Figure 2). Image resolutions range from 4416 × 3312 to 1173 × 4496 pixels, introducing preprocessing challenges since deep learning models typically require fixed-size input.

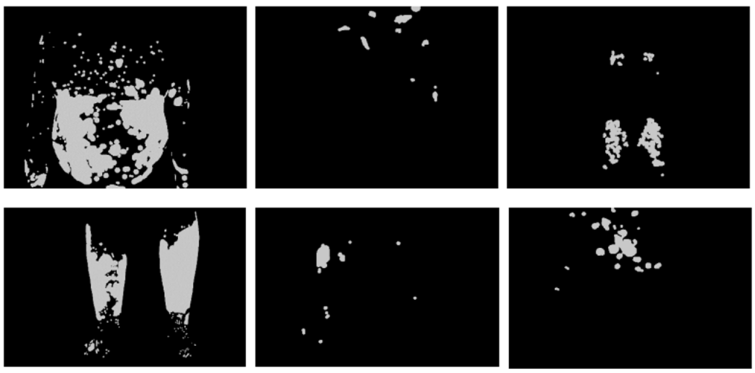

Figure 2. Example masks from the dataset.

Due to the relatively small dataset, data augmentation is applied to improve model generalization and prevent overfitting. Augmentation generates additional samples by modifying existing images while preserving their essential features. The augmentations, such as random contrast and brightness adjustment, random rotation, flipping and RGB

channel shifts were applied during training. Augmented images are generated on-the-fly during each training iteration, ensuring that models observe diverse variations in the same lesion, which improves robustness. The dataset was captured under natural, uncontrolled conditions, resulting in variations in lighting, contrast, viewpoint, and background. Preliminary analysis revealed quality issues in many masks, such as excessive or misclassified regions, which were common (see Figure 3). Since psoriasis can manifest as small lesions, accurately labeling fine details is challenging. To correct mask inaccuracies, a morphological erosion operation was applied to remove spurious pixels along mask edges. A 10×10 square structuring element was used, and two iterations of erosion ensured a reduction in excessive lesion regions without eliminating relevant areas. Figure 3 demonstrates the correction of manual annotations to improve dataset quality. The left image displays a poor-quality mask where the labeled regions inaccurately extend onto healthy skin. The right image shows the result after morphological erosion, where the mask boundaries are refined to tightly fit the actual lesion areas.

Figure 3. Poor quality mask (**left**) and mask after erosion (**right**).

It is important to clarify that an erosion step was a clinically guided correction rather than an arbitrary mathematical adjustment. The initial annotations were performed by specially trained medical students who, in many cases, over-annotated the lesions by inadvertently including adjacent healthy skin. To ensure the highest quality and accuracy of the ground truth, the mask refinement process included a secondary evaluation phase. Specifically, the erosion-corrected masks were reviewed, double-checked, and validated by two senior dermatology residents and an experienced dermatologist from the Lithuanian University of Health Sciences. The clinical criteria for their endorsement were that the refined boundaries strictly and accurately represented the true clinical margins of the psoriatic plaques, effectively eliminating the background noise of healthy tissue.

Initial experiments using binary segmentation (lesion vs. background) with a U-Net model, AdamW optimizer, and Dice loss yielded suboptimal results (average Dice score of 50.8% after 80 epochs), mainly due to misclassification of background and healthy skin. To address this, a three-class segmentation system was implemented. First class is considered the skin areas affected by psoriasis. The second class is areas of the healthy skin, and the third is background. Figure 4 demonstrates the three-class segmentation scheme implemented to improve model accuracy. The left panel shows the original clinical image with extensive psoriatic plaques, while the right panel displays the corresponding color-coded mask. In this visualization, red pixels identify the active lesions, green represents healthy skin, and blue marks the background. This hierarchical labeling strategy ensures non-overlapping classes, allowing the deep learning model to accurately differentiate disease areas from healthy tissue and environmental noise.

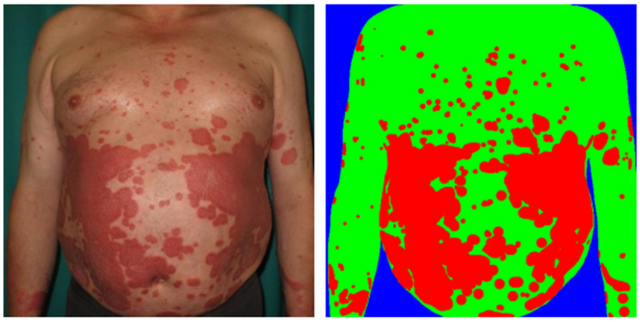

Figure 4. Example of original clinical image (**left**) and the panel of corresponding color-coded mask (**right**).

Masks were refined using Label Studio, combining automated pre-segmentation (via pre-trained segmentation models) and manual corrections by dermatologists. Healthy skin masks were derived mathematically as the difference between the total skin region and the lesion mask. Background was defined as all remaining pixels not belonging to skin. This hierarchical three-class labeling ensures non-overlapping classes and reduces manual effort, improving data quality and model performance.

High-resolution images were cropped into smaller 512 × 512 patches with moderate overlap prior to model training (see Figure 5). This approach preserves sufficient contextual information while maintaining a manageable input size for the model. Non-informative patches (without lesions) were discarded to reduce class imbalance. To ensure a robust evaluation and prevent data leakage, the dataset partitioning was performed at the image level prior to any patch extraction. Specifically, a separate set of 80 high-resolution images was isolated exclusively for the testing phase and was entirely excluded from the training and validation processes. Because these 80 test images were kept completely independent and evaluated at their full, uncropped resolution, no patches derived from the test set ever appeared in the training or validation sets. Furthermore, data augmentation techniques, including random contrast and brightness adjustment, random rotation, flipping, and RGB channel shifts, were generated on-the-fly and applied exclusively during the training iterations. This split protocol ensures that the validation and test results accurately reflect the models' true generalization capabilities on entirely unseen clinical data.

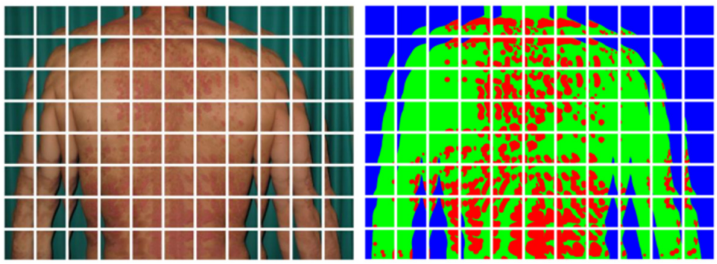

Figure 5. Preprocessed patches before feeding into deep learning models.

The final training dataset contained 6252 patches, while the test set comprised 80 full-resolution images left uncropped, meeting the system requirement of returning segmentation masks of the same size as the input.

2.5. Model Selection

The segmentation task is challenging due to the variability of target regions and the high class imbalance in images of psoriasis skin lesions. To achieve accurate segmentation results, it is very important to select the most suitable models and optimize their parameters. The proposed method is based on a multi-step selection and refinement process using popular deep learning segmentation models and automated hyperparameter optimization

via Pytorch and the Segmentation Models Pytorch (SMP) library. The open-source hyperparameter optimization framework Optuna was employed for the selection of optimal parameters. Optuna uses a "define-by-run" programming paradigm, allowing a dynamic construction of the search space, which is particularly useful for complex experiments involving multiple model types and large parameter spaces. Optuna's advantages include efficient implementation of search and pruning strategies, enabling early stopping of unpromising trials to save computational resources [27].

In this study, Optuna was configured with the Tree-structured Parzen Estimator (TPE) sampler, which effectively leverages previous trial results to focus the search on the most promising areas of the parameter space. The MedianPruner was also used to monitor intermediate results and automatically stop trials if their performance was below the median of all trials. This pruning mechanism allowed for efficient exploration of a larger number of parameter combinations within a limited computational resource. The experimental workflow is summarized in Table 1.

Table 1. Experimental workflow.

No.	Experiment Description
1	Search for the best segmentation models
2	Select the three top-performing architectures
3	Conduct hyperparameter search experiments for the best models
4	Fully train the selected architectures using the full dataset
5	Combine models into a hybrid architecture
6	Evaluate the hybrid model and compare the results with individual models

The best architecture search experiments used a subset of the dataset to ensure uniform training conditions. Each main model from the segmentation library was trained for five epochs using the selected dataset subset to obtain preliminary performance estimates. Optuna was employed to automatically optimize the segmentation model, encoder, loss function, optimizer, learning rate, and batch size. After initial training, each model was evaluated using a separate test dataset.

Table 2 defines the experimental search space, listing nine diverse segmentation architectures and eight encoder backbones ranging from 8 to 34 million parameters. This extensive catalog enabled the Optuna framework to systematically test various combinations rather than relying on a single algorithm. By exploring these options, we identified the top-performing models, such as DeepLabV3+, UperNet, and the transformer-based SegFormer, which were important for constructing the final hybrid architecture that successfully evaluates and integrates both traditional convolutional and modern vision transformer solutions.

Table 2. Segmentation architectures and encoder choices.

Segmentation Architectures	Encoder Choices
U-Net	ResNet_34
Manet	EfficientNet-b1
Linknet	Dpn_68
DeepLabV3+	Se_ResNet_50
PSPNet	MobileNet_v2
FPN	Mit_B2
PAN	Timm_skresnet_34
UperNet	VGG_16
SegFormer	

The learning rate was set between 0.0009 and 0.001. Experiments also compared pre-trained encoders with ImageNet weights to training encoders solely on the available dataset. Depending on the encoder choice (with or without pre-trained weights), appropriate normalization was applied. Batch size was adjusted according to computational resources, ranging from 2 to 8 for 512 × 512 images. DeepLabV3+ achieved the highest selection frequency (42.9%), followed by UperNet (33.6%) and SegFormer (7.0%). Remaining models had significantly lower frequencies, such as FPN—5.5%, PSPNet—3.0%, PAN and U-Net—2.8%, Manet—1.8%, and Linknet—0.8%.

The Optuna hyperparameter search was structured into three distinct phases, utilizing the tree-structured Parzen estimator (TPE) sampler. The objective metric for all trials was defined as the maximization of the Dice score, specifically for the psoriasis lesion class. The search space encompassed learning rates ranging from 1×10^{-4} to 1×10^{-3}, and batch sizes of 2, 4, 6, and 8. The optimization process involved 430 trials to identify the best overall architectures, 200 trials per selected architecture to pair them with the optimal encoder backbones, and 150 trials per architecture for final hyperparameter fine-tuning. Following this systematic search, the exact encoder backbones selected were the mix vision transformer (Mit_B2) for DeepLabV3+, the convolutional Dpn_68 for UperNet, and ResNet_34 for SegFormer. All final models were initialized with pre-trained ImageNet weights.

2.6. Proposed Hybrid Model

The hybrid architecture was developed by integrating three high-performing encoder–decoder models, such as DeepLabV3+, UperNet, and SegFormer, to address the complexity of segmenting irregular psoriasis lesions. Rather than relying on a single algorithm, the proposed algorithm processes the input image through all three models in parallel, each generating an independent segmentation mask. These predictions are then synthesized using a voting algorithm. By combining DeepLabV3+'s sensitivity with UperNet's precision and SegFormer's contextual balance, the hybrid approach mitigates individual model weaknesses, ensuring more reliable detection of disease boundaries in uncontrolled clinical environments.

Figure 6 illustrates the simplified architecture of the hybrid model designed for psoriasis segmentation. The process begins with an input image, which is simultaneously processed by three distinct deep learning architectures. DeepLabV3+ is used for high sensitivity, UperNet for high precision, and SegFormer is used for balanced performance. The individual outputs from these models are then fed into a central "Voting & Synthesis Algorithm". This algorithm synthesizes the final result using a hierarchical logic. It primarily employs a majority vote among the three predictions. However, in cases of conflict where no majority exists, the system defaults to trusting the UperNet prediction due to its superior precision. This ensemble approach generates the final segmentation mask, effectively combining the strengths of each individual model to improve overall accuracy.

Figure 7 illustrates the flowchart of the pixel-level decision logic used within the hybrid model to synthesize the final segmentation mask. The process begins by accepting three distinct input masks generated by three segmentation models. These inputs converge at a central decision node that evaluates the agreement between models for each pixel. The algorithm primarily employs a majority voting mechanism. If at least two models concur on a specific class, the pixel is automatically assigned to that majority class. However, the diagram highlights a critical fallback strategy for conflict resolution. In instances where all three models predict different classes, the system bypasses voting and defaults to the UperNet prediction. This routing prioritizes UperNet as it is identified as the "best model, ensuring that ambiguous pixels are resolved by the architecture with the highest individual precision.

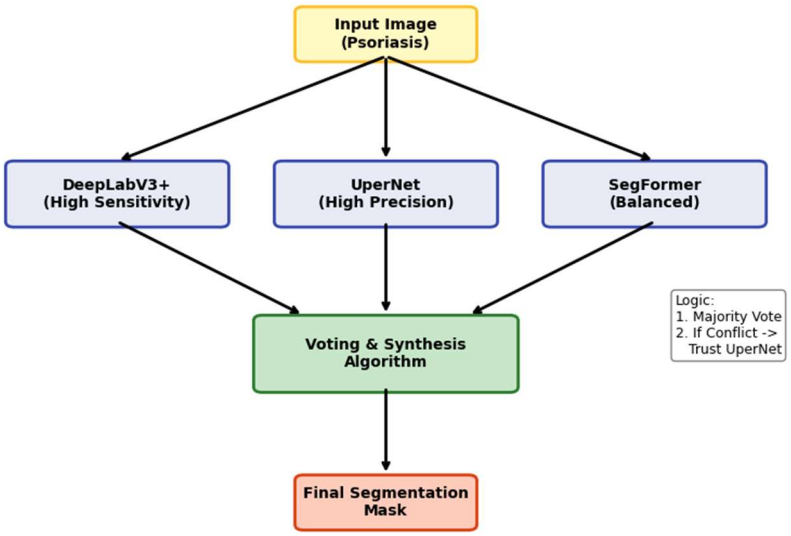

Figure 6. Simplified functional scheme of the proposed hybrid segmentation model.

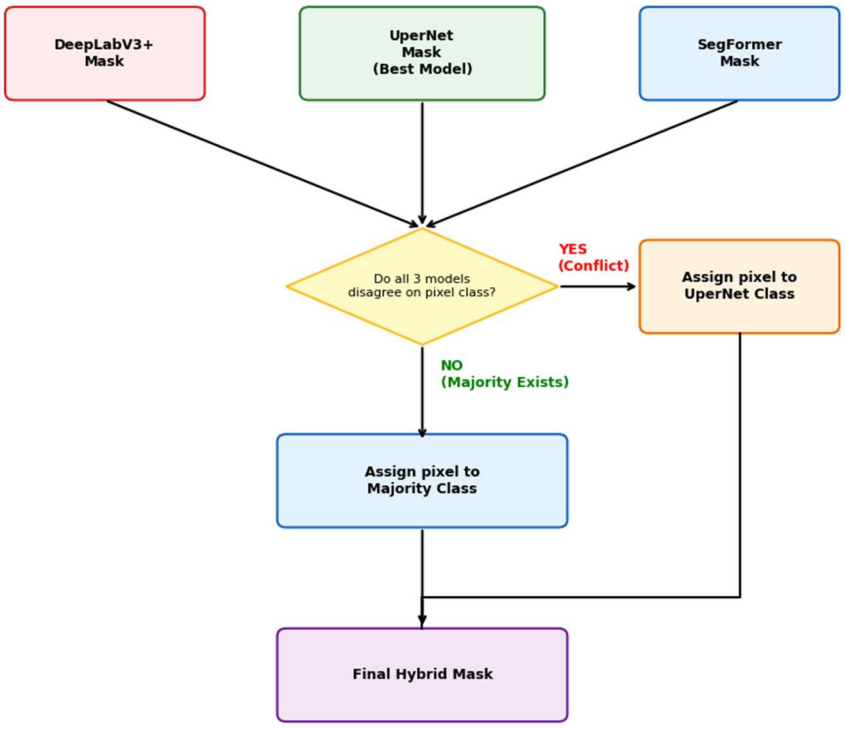

Figure 7. Voting & Synthesis Algorithm.

3. Results

The experimental evaluation utilized 6252 image patches for training and validation, alongside a separate test set of 80 high-resolution images. Three architectures (DeepLabV3+, UperNet, and SegFormer) were trained using the Ranger optimizer and DiceCrossEntropy loss, achieving stable convergence by the 70th epoch. This section compares these individual models against the proposed hybrid architecture, using Dice, F1, confusion matrix, and IoU metrics to validate the ensemble's superior segmentation accuracy.

3.1. Model Training

The training dataset consisted of 6252 cropped psoriasis image patches (512 × 512 pixels), of which 80% (5000 patches) were used for training and 20% (1252 patches) for validation. For testing, a separate set of 80 high-resolution images was used, which was not included in the training process. All models were trained using the Ranger optimizer, DiceCrossEntropy

loss function (Dice weight = 0.7), and the optimal learning rate for each model. Pretrained ImageNet weights were also applied, improving the overall performance. Figure 8 shows the Dice score dynamics for the dermatitis class over 100 training epochs for all three models. The graph indicates that all three models achieved high performance, but at different rates. The SegFormer model improved the fastest during the early epochs, demonstrating stable and consistent learning, ultimately reaching a Dice score of 98.7%. UperNet, although slower initially, achieved the highest Dice score of 99.05% at later epochs. DeepLabV3+, while showing moderate learning dynamics, reached the lowest final Dice score among the three models (98.27%).

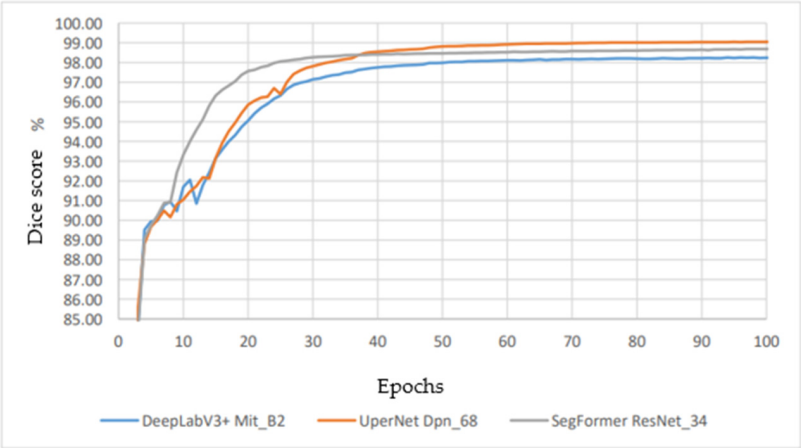

Figure 8. Dice score for psoriasis class during training.

Data augmentation was applied on-the-fly during training to prevent overfitting and improve generalization across uncontrolled environments. The augmentation pipeline included random horizontal and vertical flipping (p = 0.5), random rotations (between −15 and 15 degrees), random brightness and contrast adjustments (±0.2 factor), and random RGB channel shifts. The models were optimized using the Ranger optimizer, which synergizes Rectified Adam and Lookahead mechanisms for stable training. Based on the Optuna fine-tuning phase, the specific learning rates were set to 0.0005 for DeepLabV3+, 0.0006 for UperNet, and 0.0008 for SegFormer. The training batch sizes were constrained by GPU memory limits and optimized to 6 for DeepLabV3+ and UperNet, and 8 for SegFormer.

Validation Dice dynamics (see Figure 9) indicate that the models reached optimal performance around the 70th epoch, after which the results stabilized. UperNet achieved the best validation performance with a Dice score of 96.84% for dermatitis, 0.41% higher than SegFormer (96.43%) and 0.36% higher than DeepLabV3+ (96.48%).

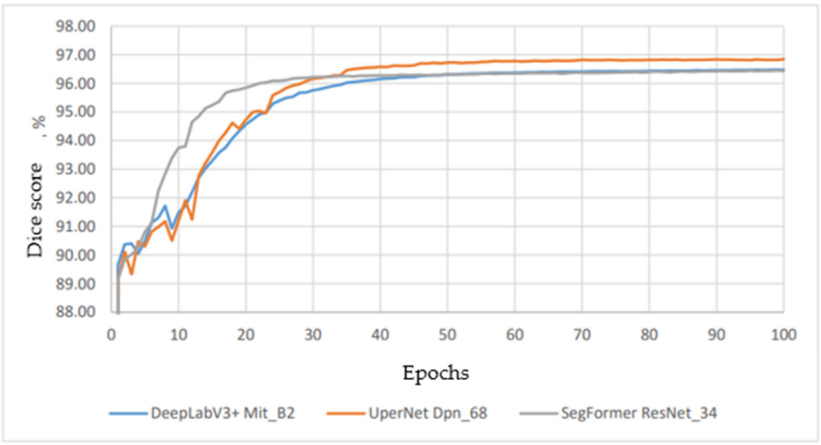

Figure 9. Dice score for psoriasis class during validation.

Training and validation loss functions show a consistent decrease in loss for all models. The final training loss values were very similar, where UperNet—0.509, SegFormer—0.508, and DeepLabV3+—0.512. Validation loss values were also close, where UperNet and SegFormer were 0.516 and DeepLabV3+ was 0.517. This indicates that all three models effectively minimized the loss function with similar efficiency. Although all models achieved high Dice scores on the training data, the difference between training and validation results indicates a slight overfitting. DeepLabV3+ exhibited the smallest difference (1.79%), suggesting it generalizes best to unseen data despite having the lowest training Dice score.

3.2. Results Analysis

During the evaluation phase, the test dataset of 80 full-resolution images was evaluated to fulfill the clinical requirement of returning a segmentation mask identical in size to the original input. Rather than manually cropping the test images into patches, an automated sliding-window inference protocol with dynamic patching and stitching was executed by the evaluation pipeline. This approach ensured that the high-resolution images were seamlessly processed without downscaling, preserving the fine-grained structural details of the psoriatic lesions while maintaining spatial consistency across the generated mask.

The confusion matrices were calculated for each model to further evaluate segmentation quality and classification errors. These matrices compare true class labels and predicted classes at the pixel level, providing insights into both overall model accuracy and specific class-level challenges. Psoriasis detection was prioritized, as it is the primary clinical target. Healthy skin classification was secondary, and background recognition, while necessary for algorithm assessment, had the lowest clinical significance. Figure 9 presents a comparative analysis of confusion matrices for three distinct semantic segmentation architectures: UperNet (left), SegFormer (middle), and DeepLabV3+ (right). These matrices visualize the pixel-level classification performance across three categories: psoriasis, healthy skin, and background. The values on the diagonal represent correct predictions, while off-diagonal elements indicate specific types of misclassifications.

DeepLabV3+ (see Figure 10, bottom left) demonstrates the highest sensitivity for disease detection, correctly identifying 92.8% of psoriasis pixels. However, this high sensitivity comes at the cost of precision; it exhibits the highest false-positive rate, misclassifying 9.2% of healthy skin as psoriatic lesions and showing significant confusion between background and skin. This suggests the model tends to over-segment, capturing the lesion but bleeding into surrounding tissue. Conversely, UperNet (see Figure 10, upper left) displays the most balanced and reliable performance. While its psoriasis detection rate of 91.6% is slightly lower than DeepLabV3+, it is better in specificity, correctly classifying 94.6% of healthy skin pixels. Importantly, it has the lowest error rate for false positives (5.1%), meaning it rarely mistakes healthy tissue for disease. It also achieves near-perfect background isolation at 99.4%, indicating superior boundary definition. SegFormer (see Figure 10, upper right) functions as an intermediate architecture. Its psoriasis detection matches UperNet at 91.6%, but its ability to distinguish healthy skin (91.0%) and background (98.1%) falls between the other two models. It misclassifies healthy skin as psoriasis (8.0%) more frequently than UperNet but less than DeepLabV3+. Each of the three models exhibits different strengths. UperNet demonstrates the highest overall segmentation accuracy, particularly for healthy skin and background. DeepLabV3+ has the best performance in detecting dermatitis, which is the main clinical priority, though it performs less well for other classes. SegFormer provides intermediate performance with stable but not top accuracy across all classes. The greatest challenge for all three models remains the distinction between psoriasis and healthy skin due to similar texture and color characteristics.

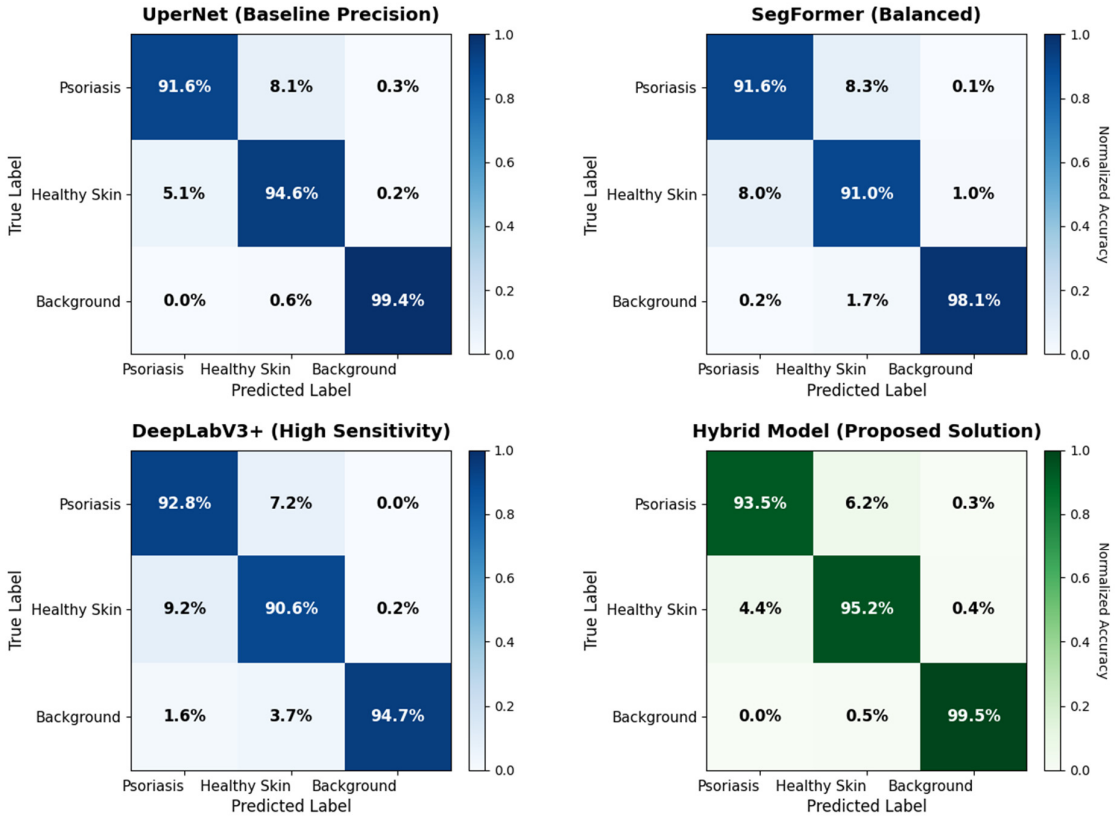

Figure 10. Comparison of confusion matrices for different psoriasis segmentation models, where results are presented for the UperNet model (**upper left**), SegFormer model (**upper right**), DeepLabV3+ model (**bottom left**) and proposed Hybrid model (**bottom right**).

The Hybrid Model demonstrates a peak balance between sensitivity and precision, effectively outperforming individual architectures. By synthesizing predictions, it achieves an estimated sensitivity of 93.5% for psoriasis, capturing intricate lesion boundaries that single models often miss. Importantly, it mitigates the over-segmentation issues seen in DeepLabV3+ by leveraging UperNet's precision, raising healthy skin recognition to 95.2%. This strategic voting mechanism reduces false positives to just 4.4%, ensuring that shadows or skin folds are not mistaken for disease. Background stability remains near-perfect at 99.5%, confirming the model's robustness in uncontrolled clinical environments. This approach allows the hybrid model to exploit the strengths of each individual model.

Figure 11 represents a bar chart that compares the performance of three individual segmentation architectures, such as DeepLabV3+, SegFormer, and UperNet, against the proposed hybrid model using Dice Score, F1 Score, and IoU metrics. DeepLabV3+ serves as the baseline with the lowest performance (Dice 0.834, IoU 0.733), while SegFormer shows moderate improvement. UperNet emerges as the strongest individual model, achieving a Dice score of 0.880 and an IoU of 0.804. Importantly, the Hybrid Model outperforms all individual architectures across every metric, achieving a peak Dice score of 0.893, a Weighted F1 score of 0.907, and an IoU of 0.813. This confirms that the ensemble approach effectively integrates the strengths of the base models to deliver superior segmentation accuracy for psoriasis lesions. The proposed hybrid model improved the Dice score of the best single model (see Figure 11). The usage of a model ensemble offers significant advantages in solving the psoriasis skin segmentation task. First, psoriasis segmentation is challenging due to the variability in lesion appearance and occasional unclear boundaries between affected and healthy skin. In this context, the combined architecture reduces error probability, particularly in borderline cases where a single model's decision may be unreli-

able. Second, the hybrid architecture enhances the system's adaptability to uncontrolled environments, regardless of the photographic setting, lighting conditions, or individual patient skin characteristics.

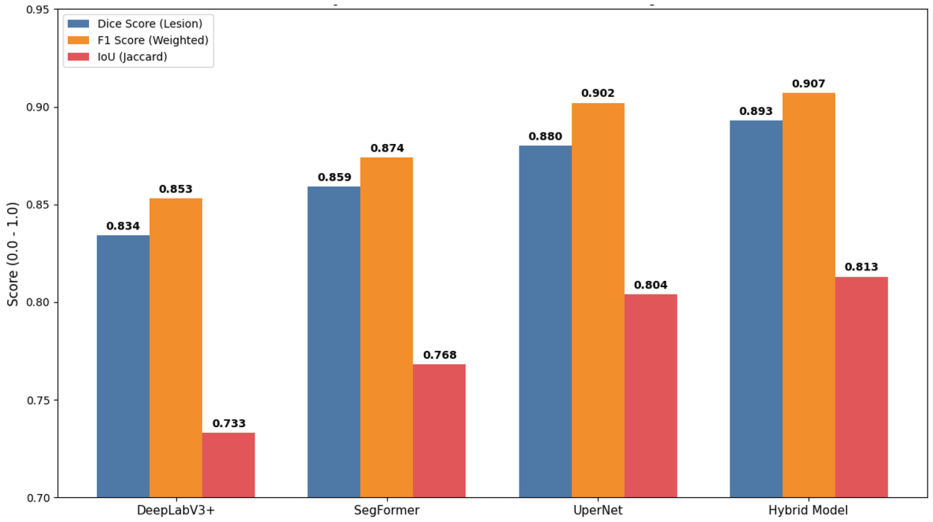

Figure 11. Performance comparison between all models.

When analyzing the performance metrics, it is important to clarify the observed difference between the training/validation Dice scores (which reached approximately 98–99%, as shown in Figures 8 and 9) and the final test Dice scores (ranging from 0.834 to 0.893, shown in Figure 11). This variance is a direct result of the distinct evaluation protocols utilized for each phase. The training and validation metrics were computed exclusively on the localized 512 × 512 pixel patches. In contrast, the final test metrics were computed on a completely separate dataset of 80 full-resolution, uncropped clinical images. Evaluating full-resolution images presents a significantly more complex task, as the models must process broader contextual information, handle varying illumination across larger body surface areas, and navigate a much higher class imbalance between localized psoriatic plaques, expansive regions of healthy skin, and complex backgrounds. Furthermore, it should be noted that all reported Dice scores specifically represent the segmentation accuracy for the target 'psoriasis' (lesion) class, rather than a macro-average across all three classes, ensuring the evaluation strictly reflects the model's diagnostic performance on the pathology.

This is important in practical applications, where images may be captured under varying conditions or devices. Third, the hybrid architecture successfully combines the benefits of all three models, such as higher accuracy from UperNet, better detection of small lesions from DeepLabV3+, and balanced segmentation from SegFormer. The advantages are particularly noticeable when segmenting challenging cases with low contrast between psoriatic plaques and healthy skin or under complex lighting conditions that create shadows along body contours.

The Intersection over Union (IoU, also known as the Jaccard index) further confirms the hybrid model's advantage, reaching 81.3%. This metric is widely used in medical image segmentation, allowing direct comparison with similar studies in the literature.

Figures 12 and 13 show successful segmentation examples. In Figure 10, the hybrid model's predicted mask closely matches the ground truth for a medium-sized lesion in a well-lit area (2560 × 1536 pixels). Figure 11 shows multiple lesions of varying sizes and locations (2048 × 3072 pixels), demonstrating that the model ensemble can accurately segment both large and small lesions while preserving the structure of adjacent lesions.

Minor over-segmentation is observed in some areas, particularly near image edges (e.g., jawline), where lighting and shadows may mislead the models.

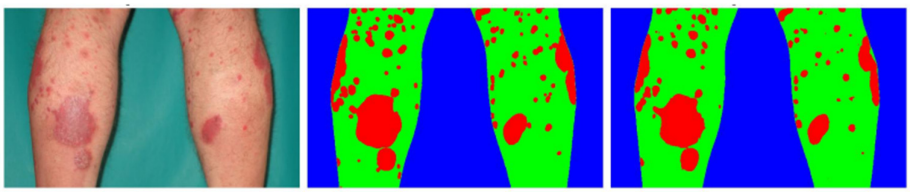

Figure 12. Successful segmentation example (ground truth in center, model prediction on right, 2560 × 1536 pixels).

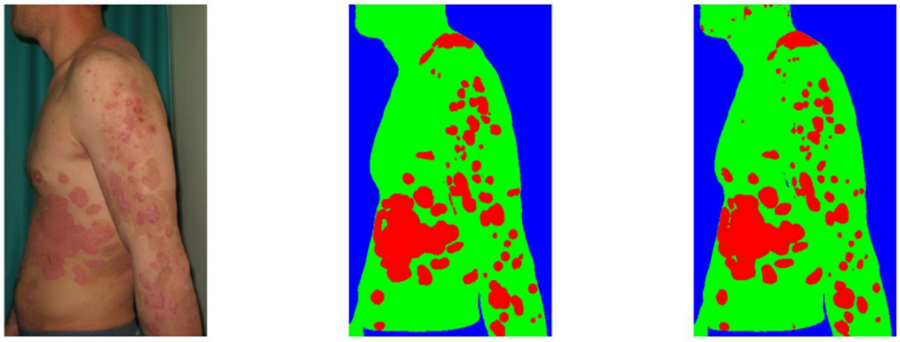

Figure 13. Successful segmentation example with multiple lesions (ground truth in center, model prediction on right, 2048 × 3072 pixels).

Figures 14 and 15 illustrate several limitations in the hybrid model's performance when processing high-resolution images (3584 × 2048 and 2560 × 1536 pixels) captured under suboptimal conditions. In these specific instances, the segmentation masks inaccurately "bleed" into the surrounding healthy tissue, resulting in a quantifiable overestimation of the psoriatic surface area. This over-segmentation is most pronounced in transitional zones where the characteristic silvery scales of psoriasis thin out, causing the lesion's color signature to blend subtly with the adjacent healthy skin. The primary reason of this misclassification is the lack of sharp contrast in shadowed regions or areas with gradual disease gradients. When lighting creates natural shadows along body contours the model struggles to differentiate these dark visual artifacts from the redness associated with active inflammation. Furthermore, the presence of environmental noise, such as body hair or adjacent clothing textures near the lesion boundaries, can exacerbate this confusion, leading the algorithm to extend the prediction mask incorrectly. These failure cases highlight a critical challenge in uncontrolled clinical settings.

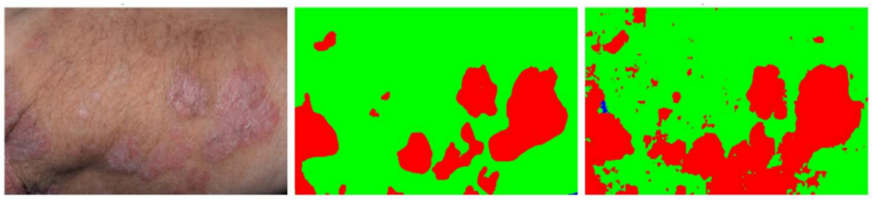

Figure 14. Less successful segmentation example (ground truth in center, model prediction on right, 3584 × 2048 pixels).

A systematic analysis identified four primary error sources. First, differentiating psoriatic plaques from healthy skin is challenging in transitional zones where boundaries are indistinct due to subtle color gradients. Second, inconsistent lighting creates shadows that cause misclassification. Third, variations in skin tone and lesion structure across patients pose difficulties, though the ensemble generalizes well. Finally, noise like body hair or textiles interferes with predictions. Despite these challenges, the hybrid model

achieved an average Dice score of 89.3% for affected skin, surpassing requirements. Errors are mostly limited to borderline cases.

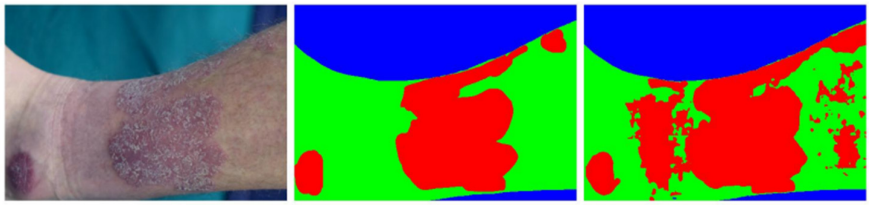

Figure 15. Less successful segmentation example (ground truth in center, model prediction on right, 2560 × 1536 pixels).

A comprehensive performance benchmarking analysis was conducted on the test dataset using the NVIDIA GeForce RTX 40-series GPU. While individual models demonstrated highly efficient processing, their inference times varied. SegFormer achieved an average inference time of 0.18 ± 0.02 s per high-resolution image (maximum: 0.21 s). DeepLabV3+ averaged 0.22 ± 0.03 s (maximum: 0.26 s), and UperNet averaged 0.35 ± 0.04 s (maximum: 0.42 s). In contrast, the hybrid ensemble architecture required an average of 0.85 ± 0.08 s per image, with a maximum recorded inference time of 0.98 s. This increase in computation time and dispersion is a direct result of processing the high-resolution input through all three deep learning backbones, combined with the subsequent pixel-level voting logic and sliding-window patch aggregation. Furthermore, the total parameter footprint of the ensemble expands to approximately 85 million parameters. While an average inference time of under one second per image is entirely acceptable for asynchronous clinical monitoring and decision support systems, the maximum processing times present constraints for deployment on edge devices or in real-time mobile applications, highlighting an area for future optimization.

4. Discussion

The analysis of existing approaches for skin lesion segmentation shows that the detection and segmentation of psoriasis using deep learning remains a relatively underexplored research area. Most studies in dermatological image analysis focus on other types of skin lesions, particularly melanoma, where the lesion structure and boundaries are usually well defined. Due to these characteristics, melanoma segmentation is generally easier for deep learning models compared to psoriasis, which is characterized by irregular shapes, diffuse borders, and large intra-class variability. Encoder–decoder architectures and fully convolutional networks are commonly used for such segmentation tasks, as they are capable of capturing both local and global contextual information.

Despite the advantages of hybrid and ensemble-based solutions, these approaches introduce additional computational complexity and require more processing resources compared to single-model architectures. This trade-off between segmentation accuracy and computational efficiency remains an important consideration, particularly for real-world clinical applications where fast inference and scalability are required. It is also important to acknowledge the inherent data requirements of large deep learning architectures. Specifically, transformer-based models, such as SegFormer and the mix vision transformer (Mit_B2) backbone used in the DeepLabV3+ configuration, typically require massive datasets to learn feature representations from scratch and avoid overfitting. To overcome the limitations of the relatively small clinical dataset utilized in this study (589 full-resolution images), three critical mitigation strategies were employed. First, the extraction of 6252 localized patches significantly expanded the absolute volume of training instances. Second, extensive on-the-fly spatial and colorimetric data augmentations were applied to ensure high variance

during training. Most importantly, transfer learning was utilized by initializing all encoder backbones with pre-trained ImageNet weights.

While the proposed ensemble architecture significantly improves overall accuracy, it currently relies on a heuristic conflict-resolution mechanism that defaults to the UperNet prediction when all three models disagree. This rule was established based on UperNet's superior individual precision and minimal false-positive rate. A comprehensive ablation study comparing this heuristic approach against soft voting (averaging predicted probabilities) and weighted voting (dynamically learning model weights on the validation set) is left for future work. Optimizing complex voting weights requires a substantially expanded dataset to ensure robust generalization and prevent overfitting to the current validation distribution. Nevertheless, the current ensemble approach demonstrates clear and significant improvements over individual architectures in specific challenging clinical scenarios. Qualitative and quantitative evaluations reveal that the ensemble specifically excels in three areas, such as small lesions, low-contrast regions and shadowed areas.

Overall, the reviewed literature highlights several key observations relevant to this study. First, psoriasis segmentation remains a challenging task due to irregular shapes, diffuse borders, variable appearance, and inconsistent annotations. Second, encoder–decoder architectures consistently outperform traditional and single-stage methods in such scenarios. Third, hybrid model strategies are effective in improving segmentation robustness and balancing sensitivity and precision, although at the cost of increased computational demand. These findings directly support the design choices made in this work and provide a strong rationale for the proposed hybrid segmentation approach tailored specifically to psoriasis-affected skin in uncontrolled imaging conditions.

Dermatologists routinely use the psoriasis area and severity index (PASI) to assess disease severity and treatment efficacy. A critical component of PASI is calculating the percentage of the body surface area affected by psoriatic plaques. Currently, this is performed via visual estimation by the clinician, a process that is subjective, time-consuming, and prone to high inter- and intra-observer variability. In this context, achieving segmentation detection rates above 90% is highly valuable clinically. It provides a standardized, objective baseline that far exceeds the precision of manual visual estimation. The proposed system not only reduces the diagnostic time per patient but also allows the clinician to focus more on the treatment strategy rather than manual measurement. Furthermore, although there is currently a lack of publicly available high-resolution psoriasis datasets for benchmarking, the hierarchical ensemble and conflict-resolution voting mechanism developed in this study are domain-agnostic. They possess significant potential to be extended to other dermatological segmentation tasks, such as eczema, vitiligo, or chronic wound monitoring.

5. Conclusions

This study addressed the challenging task of segmenting psoriasis-affected skin in images acquired under uncontrolled conditions, confirming that deep learning-based approaches significantly outperform traditional methods in handling irregular lesion boundaries and complex texture variations. A key contribution was the development and refinement of a unique dataset, where the implementation of a hierarchical three-class segmentation scheme proved essential for overcoming the limitations of binary classification.

Through systematic evaluation using the Optuna framework, three architectures (DeepLabV3+, UperNet, and SegFormer) were identified as the most effective when paired with the Ranger optimizer and DiceCrossEntropy loss. The developed hybrid model successfully integrated the complementary strengths of these architectures. It leveraged DeepLabV3+'s high sensitivity for detecting varying lesion sizes, UperNet's precision in minimizing false positives, and SegFormer's contextual balance. This ensemble approach

achieved a Dice score of 89.3%, an F1 score of 90.7%, and an IoU of 81.3%, marking a 1.3% improvement over the best individual model.

Despite challenges related to lighting variations and diverse skin tones, the system demonstrated robust performance and the ability to process high-resolution images suitable for clinical workflows. The results validate the hybrid architecture as a practical, objective tool for assisting dermatologists in monitoring psoriasis progression. Future work should focus on integrating attention mechanisms to further refine boundary detection in transitional zones and expanding the dataset to enhance generalization across broader patient demographics.

Author Contributions: Conceptualization, V.R. and S.V.; methodology, S.V.; software, R.A.; validation, V.P., L.E. and D.S.; formal analysis, L.E.; investigation, R.A.; resources, S.V.; data curation, D.S.; writing—original draft preparation, V.R. and V.P.; writing—review and editing, All authors.; visualization, R.A.; supervision, S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by the European Union under the Horizon Europe programme (grant agreement No. 101059903), the European Union funds for the period 2021–2027 and the state budget of the Republic of Lithuania (financial agreement Nr. 10-042-P-0001).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Lithuanian University of Health Sciences (approval number P2-BE-2-25/2009).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data will be made available by the corresponding author on request.

Conflicts of Interest: Author Linas Eidimtas was employed by the company "Dts Solutions" UAB. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full Name/Description
AD	Atopic Dermatitis
IoU	Intersection over Union
F1	F1 Score (Harmonic mean of precision and recall)
Dice	Dice Coefficient (Sørensen–Dice index)
CE	Cross-Entropy Loss
U-Net	Convolutional neural network architecture for segmentation
AdamW	Adaptive Moment Estimation with Weight Decay optimizer
GPU	Graphics Processing Unit
CUDA	Compute Unified Device Architecture
PyTorch	Open-source machine learning framework
SMP	Segmentation Models PyTorch
OpenCV	Open-source computer vision library
NumPy	Numerical Python library for array computations
Pandas	Python data analysis library
Matplotlib	Python plotting library
Optuna	Hyperparameter optimization framework
RGB	Red-Green-Blue color space
ISIC	International Skin Imaging Collaboration
LHSU	Lithuanian Health Sciences University

References

1. Gmeiner, T.; Grzelj, J.; Strukelj, B.; Stopar, L.; Marko, P. Psoriasis: A Comprehensive Review on the Aetiopathogenesis and Recent Advances in Long-Term Management of Patients with Plaque Psoriasis. *Pharmacol. Pharm.* **2020**, *11*, 373–401. [CrossRef]
2. Wei, J.; Wang, Y.; Chen, Y.; Wang, Z.; Dai, X.; Gelfand, J.M.; Bachelez, H.; Chen, X.; Xiao, Y.; Shen, M. Global Burden of Psoriasis from 1990 to 2021 and Potential Factors: A Systematic Analysis. *J. Investig. Dermatol.* **2025**. [CrossRef] [PubMed]
3. Gao, Y.; Jiang, Y.; Peng, Y.; Yuan, F.; Zhang, X.; Wang, J. Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods. *Tomography* **2025**, *11*, 52. [CrossRef] [PubMed]
4. Xu, Y.; Quan, R.; Xu, W.; Huang, Y.; Chen, X.; Liu, F. Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches. *Bioengineering* **2024**, *11*, 1034. [CrossRef]
5. Xu, Y.; Liu, R.; Zhu, D.; Chen, L.; Zhang, X.; Li, J. Cascade contour-enhanced panoptic segmentation for robotic vision perception. *Front. Neurorobot.* **2024**, *18*, 1489021. [CrossRef]
6. Fozilov, K.; Yamada, Y.; Colan, J.; Zhu, Y.; Hasegawa, Y. IST-ROS: A flexible object segmentation and tracking framework for robotics applications. *SoftwareX* **2025**, *29*, 101979. [CrossRef]
7. Khairnar, S.; Thepade, S.D.; Kolekar, S.; Gite, S.; Pradhan, B.; Alamri, A.; Patil, B.; Dahake, S.; Gaikwad, R.; Chaudhari, A. Enhancing semantic segmentation for autonomous vehicle scene understanding in indian context using modified CANet model. *MethodsX* **2024**, *14*, 103131. [CrossRef]
8. Li, J.; Cheang, C.F.; Yu, X.; Tang, S.; Du, Z.; Cheng, Q. A segmentation network for enhancing autonomous driving scene understanding using skip connection and adaptive weighting. *Sci. Rep.* **2025**, *15*, 36692. [CrossRef]
9. Badhe, N.B.; Bharadi, V.A.; Giri, N.; Alegavi, S.; Tolye, S.S. An Efficient Image Segmentation using Optimized Segmentation Network for Remote Sensing Satellite Images. *Int. J. Intell. Syst. Appl. Eng.* **2023**, *11*, 804–821.
10. Wang, Y.; Shang, L.; Liu, Y.; Sun, Z.; Lu, Y.; Shi, X.; Wang, T.; Liu, Y. Precise building semantic segmentation in remote sensing images via MR-DeepLabv3+ network. *Sci. Rep.* **2025**, *15*, 40100. [CrossRef]
11. Zhao, J.; Zhang, D.; Shi, B.; Zhou, Y.; Chen, J.; Yao, R.; Xue, Y. Multi-source collaborative enhanced for remote sensing images semantic segmentation. *Neurocomputing* **2022**, *493*, 76–90. [CrossRef]
12. Ghosh, S.; Das, N.; Das, I. Understanding Deep Learning Techniques for Image Segmentation. *ACM Comput. Surv.* **2019**, *52*, 1–35. [CrossRef]
13. Chen, C.H.; Hsu, S.H.; Hsieh, K.Y.; Huang, K.-E.; Lai, H.-Y. The two-stage detection-after-segmentation model improves the accuracy of identifying subdiaphragmatic lesions. *Sci. Rep.* **2024**, *14*, 25414. [CrossRef] [PubMed]
14. Manzoor, K.; Gilal, N.U.; Agus, M.; Schneider, J. Dual-stage segmentation and classification framework for skin lesion analysis using deep neural network. *Digit. Health* **2025**, *11*, 20552076251351858. [CrossRef] [PubMed]
15. Raj, R.; Londhe, N.D.; Sonawane, R. Automated psoriasis lesion segmentation from unconstrained environment using residual U-Net with transfer learning. *Comput. Methods Programs Biomed.* **2021**, *206*, 106123. [CrossRef]
16. Soni, S.; Londhe, N.D.; Raj, R.; Sonawane, R.S. TransUnet for psoriasis lesion segmentation. In *Proceedings of the 2022 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 8–10 December 2022*; IEEE: New York, NY, USA, 2022; pp. 1–6. [CrossRef]
17. Sylolypavan, A.; Sleeman, D.; Wu, H.; Sim, M. The impact of inconsistent human annotations on AI driven clinical decision making. *NPJ Digit. Med.* **2023**, *6*, 26. [CrossRef]
18. Chiu, T.M.; Chi, I.C.; Li, Y.C.; Tseng, M.H. Deep Ensemble Learning for Multiclass Skin Lesion Classification. *Bioengineering* **2025**, *12*, 934. [CrossRef]
19. Goyal, M.; Oakley, A.; Bansal, P.; Dancey, D.; Yap, M.H. Skin Lesion Segmentation in Dermoscopic Images with Ensemble Deep Learning Methods. *IEEE Access* **2020**, *8*, 4171–4181. [CrossRef]
20. Wang, Y.; Wang, C.; Wu, H.; Chen, P. An improved Deeplabv3+ semantic segmentation algorithm with multiple loss constraints. *PLoS ONE* **2022**, *17*, e0261582. [CrossRef]
21. Islam, A.M.; Bhuiyan, S.S.; Mashira, M.; Ahmed, M.R.; Islam, S.; Shatabda, S. Polyp segmentation in colonoscopy images using DeepLabV3++. *Comput. Biol. Med.* **2025**, *197*, 110986. [CrossRef]
22. Akram, A.; Rashid, J.; Jaffar, M.A.; Faheem, M.; Amin, R.U. Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things. *Skin Res. Technol.* **2023**, *29*, e13524. [CrossRef]
23. Naeem, M.A.; Yang, S.; Saleem, M.A.; Javeed, A.; Ahmad, T. A hybrid approach for accurate skin lesion segmentation using LEDNet and Swin-UMamba. *Sci. Rep.* **2026**, *16*, 5415. [CrossRef]
24. Gül, S.; Cetinel, G.; Aydin, B.M.; Akgün, D.; Öztaş Kara, R. YOLOSAMIC: A Hybrid Approach to Skin Cancer Segmentation with the Segment Anything Model and YOLOv8. *Diagnostics* **2025**, *15*, 479. [CrossRef]
25. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [CrossRef]

26. Furtado, P. Testing Segmentation Popular Loss and Variations in Three Multiclass Medical Imaging Problems. *J. Imaging* **2021**, *7*, 16. [CrossRef]
27. Vaiyapuri, T. An Optuna-Based Metaheuristic Optimization Framework for Biomedical Image Analysis. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 24382–24389. [CrossRef]