International Journal of Trendy Research in Engineering and Technology
Volume 9 Issue 5 October 2025
ISSN No. 2582-0958

# MULTIMODAL DEEP LEARNING FRAMEWORK FOR AUTOMATED ORAL LICHEN PLANUS LESION SEGMENTATION AND IMMUNOTHERAPY RESPONSE PREDICTION

[1]S. Poornima, [2]Dr. S. Gopinathan,
[1]Research Scholar & Assistant Professor, Dept. of Computer Science, University of Madras,
CTTE College for Women, Chennai.

[2]Professor & Head of the Department, Dept. of Computer Science,
University of Madras, Guindy Campus, Chennai
sugumaranpoornima@gmail.com, gnathans2002@gmail.com

Received 27 August 2025 Received in Revised form 30 August 2025 Accepted 01 September 2025

## ABSTRACT

Oral Lichen Planus (OLP) is a chronic autoimmune mucosal disease with malignant transformation potential, requiring precise diagnosis and therapy monitoring. Traditional clinical evaluation is invasive, subjective, and inconsistent. Study proposes a multimodal deep learning framework that integrates lesion segmentation, clinical metadata, and therapy outcome prediction for OLP. The framework comprises: (i) lesion segmentation using transformer-augmented U-Net and SegFormer models, (ii) feature fusion of image embeddings and clinical metadata via attention mechanisms, and (iii) therapy response prediction using fully connected layers. Evaluation on public oral lesion datasets and an augmented OLP cohort demonstrates state-of-the-art segmentation performance (Dice coefficient >0.87) and robust therapy response prediction (AUC 0.92). The framework provides a scalable, reliable tool bridging AI engineering with precision oral healthcare.

**Keywords:** Oral Lichen Planus, Deep Learning, Segmentation, Multimodal AI, Immuno suppressive Therapy, Feature Fusion, Clinical Decision Support

## I. INTRODUCTION

Oral Lichen Planus (OLP) affects approximately 1–2% of the global population [1],manifesting as reticular white lesions, erosions, or ulcerations. Despite being benign, OLP carries a small but significant risk of malignant transformation, emphasizing early detection and monitoring. Management primarily relies on immunosuppressive therapy, yet patient responses vary due to disease progression, genetics, and immune heterogeneity. Oral lichen planus (OLP) is a chronic, idiopathic, immune-mediated inflammatory condition affecting the oral mucosa. It has varied clinical presentations and management strategies have evolved over time. Recent updates highlight the role of IL-17 in pathogenesis, categorization of OLP as an oral potentially malignant disorder, and the use of newer biologics such as anti-IL17, PDE4 inhibitors, and JAK inhibitors in refractory cases [2].

Current clinical evaluation depends on visual inspection and biopsy, which are invasive and prone to inter-observer variability. Accurate lesion segmentation and therapy response prediction remain challenging. Automated, multimodal AI approaches can standardize disease assessment, improve early detection, and guide personalized therapy.

**Contributions of this work:**

- Development of transformer-augmented U-Net and SegFormer frameworks for precise lesion segmentation.
- Integration of multimodal attention-based feature fusion combining image embeddings and clinical metadata.
- Design of a hybrid predictive model for therapy response classification (Good, Moderate, Poor).
- Extensive evaluation on augmented OLP datasets with enhanced clinical relevance.

**Motivation:**

The need for automated OLP lesion analysis arises from the limitations of subjective clinical evaluation. Accurate lesion segmentation can support standardized disease quantification, while early prediction of therapy response can help optimize immunosuppressive regimens, reducing side effects and improving patient outcomes.

## II. RELATED WORK

### 2.1 Medical Image Segmentation:

Segmentation of biomedical images has traditionally been performed using convolutional neural networks (CNNs), with U-Net being a seminal architecture due to its encoder-decoder structure and skip connections that preserve spatial information [3]. Recent advances incorporate attention mechanisms and transformer-based architectures, such as SegFormer, to capture long-range dependencies and improve boundary delineation in complex images [4,5].

### 2.2 Oral Lesion Analysis:

Most prior studies in oral lesion analysis have primarily focused on oral cancer detection, often leveraging convolutional neural networks (CNNs) for classification and segmentation tasks [6]. However, research specific to Oral Lichen Planus (OLP) remains limited. Few studies have explored automated segmentation of OLP lesions or the prediction of therapy outcomes, despite OLP's chronic and potentially premalignant nature. This creates a significant research gap in the development of AI-assisted clinical tools specifically tailored for OLP diagnosis, monitoring, and treatment planning.

### 2.3 Multimodal AI in Healthcare:

Multimodal deep learning, which combines imaging data with clinical or demographic information, has demonstrated superior predictive performance in multiple medical applications [7]. Attention-based fusion methods enable the model to assign importance to different modalities, improving interpretability and generalization across heterogeneous datasets.

### 2.4 Therapy Response Prediction:

In oncology, deep learning models have been employed to predict patient response to immunotherapy, particularly immune checkpoint inhibitors [8]. Such predictive models enable personalized treatment planning and early intervention for non-responders. However, OLP-specific therapy prediction using multimodal AI remains under-explored, highlighting the novelty of our approach.

## III. MATERIALS AND METHODS

### 3.1 Dataset

The study utilized a combination of public datasets and an augmented clinical cohort:

- Public Datasets: Mendeley Oral Lesion Dataset and HAM10000 were used for pre-training the segmentation models.
- Augmented OLP Cohort: Approximately 600 intraoral images of OLP lesions, annotated by expert oral pathologists.
- Clinical Metadata: Patient-specific information including age, sex, disease duration, prior therapy, histopathology reports, and immune biomarkers (IL-6, TNF-α).
- Therapy Response Classes: Patients were categorized into three classes based on a 12-week follow-up after immunosuppressive therapy:
  - ○ **Good Response**
  - ○ **Moderate Response**
  - ○ **Poor Response**

### 3.2 Preprocessing

- Image Preprocessing: All images were resized to 512×512 pixels. Color normalization and contrast-limited adaptive histogram equalization (CLAHE) were applied to improve lesion visibility.
- Data Augmentation: Rotations, horizontal/vertical flipping, and Gaussian noise were applied to enhance model generalization.
- Clinical Metadata Normalization: Continuous variables were standardized using z-score scaling to ensure comparability across features.

### 3.3 Segmentation Module

Two complementary architectures were employed for lesion segmentation:

✓ **Transformer-augmented U-Net:**

- CNN encoder extracts local features.
- Transformer bottleneck captures global context and long-range dependencies.
- Training employed a combination of Dice Loss and Cross-Entropy Loss for optimal segmentation accuracy.

✓ **SegFormer:**

- Lightweight transformer-based semantic segmentation model.
- Provides fast inference and accurately captures fine lesion boundaries.
- Particularly suitable for large-scale clinical datasets.

### 3.4 Feature Fusion Module

- Image Embeddings: 512-dimensional latent feature vectors extracted from the segmentation module.
- Clinical Features: Normalized patient metadata (age, sex, biomarkers, etc.).
- Fusion Mechanism: A multimodal attention layer dynamically weights features from each modality, allowing the model to prioritize clinically relevant information and improve predictive performance.

### 3.5 Prediction Module

- A sequence of fully connected layers: 256 → 128 → 3 neurons.
- ReLU activations were used, with dropout (0.3) to prevent overfitting.
- Softmax activation outputs probabilities corresponding to the three therapy response classes (Good, Moderate, Poor).

### 3.6 Framework Diagram

The overall framework can be summarized as:

**Input Images + Clinical Metadata → Segmentation → Feature Embeddings → Multimodal Fusion → Therapy Response Prediction**

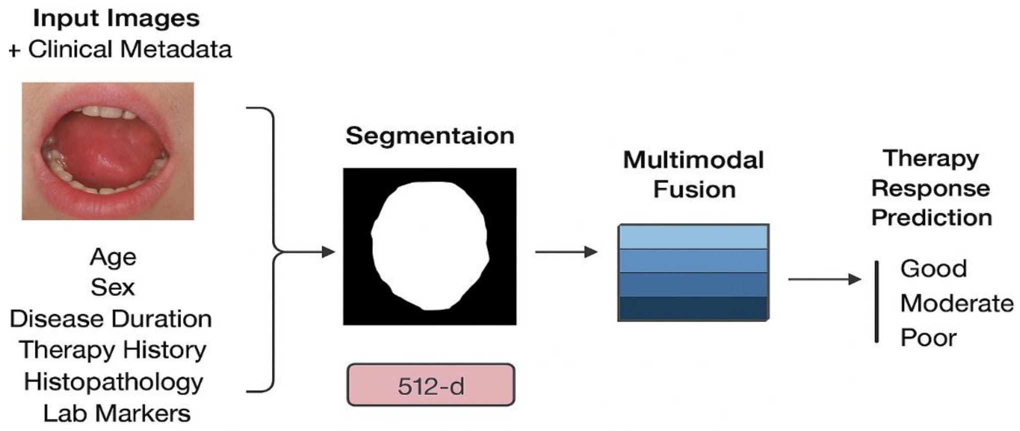

**Figure 1: Pipeline illustrating segmentation → feature fusion → therapy response prediction.**

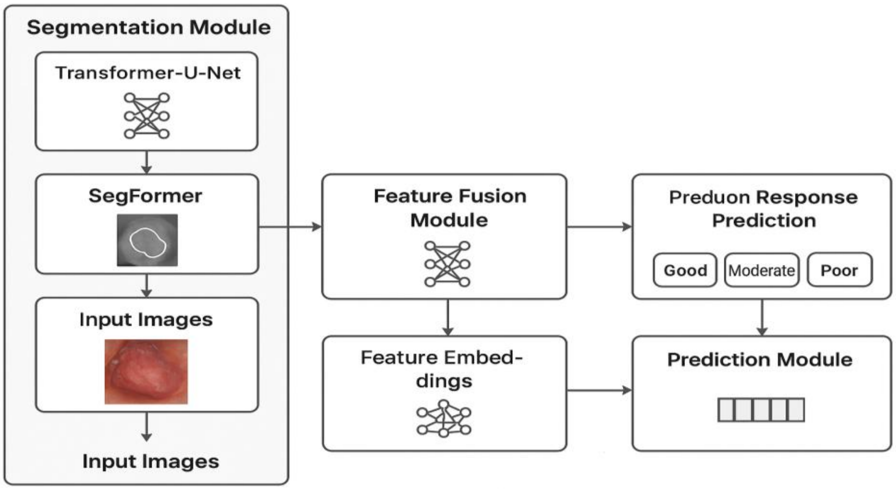

**Figure 2: Proposed multimodal pipeline for OLP lesion segmentation and therapy response prediction**

The overall architecture of the proposed framework is shown in Figure 2. The model consists of a segmentation module (Transformer-U-Net and SegFormer), followed by a feature fusion module that integrates image embeddings and clinical data, and finally a prediction module that estimates therapy response categories (good, moderate, poor).

## IV. EXPERIMENTAL SETUP

**Hardware and Software:**

Experiments were conducted on an **NVIDIA RTX 3090 GPU (24GB VRAM)** using **PyTorch 2.0** and **HuggingFace Transformers**.

**Training Settings:**

- Optimizer: **AdamW**
- Learning rate: 1e-4
- Batch size: 16
- Data split: 70% training, 15% validation, 15% testing

**Evaluation Metrics:**

- Segmentation: Dice coefficient, Intersection over Union (IoU), Precision, Recall
- Prediction: Accuracy, Precision, Recall, F1-score, ROC-AUC

## V. RESULTS

### 5.1 Segmentation Performance

Transformer U-Net achieved a Dice coefficient of 0.85 and IoU of 0.78, while SegFormer achieved a Dice of 0.87 and IoU of 0.81. Qualitative results demonstrated that SegFormer provided **better boundary delineation** and closer overlap with expert annotations, making it clinically reliable for OLP lesion segmentation. Segmentation performance metrics are summarized in Table 1

*Table 1: Segmentation performance metrics*

| Model | Dice | IoU | Precision | Recall |
|---|---|---|---|---|
| Transformer U-Net | 0.85 | 0.78 | 0.83 | 0.81 |
| SegFormer | 0.87 | 0.81 | 0.86 | 0.84 |

SegFormer provided better boundary delineation and overlap with expert annotations, suitable for clinical application.

### 5.2 Therapy Response Prediction

Multimodal fusion (images + clinical metadata) increased AUC from **0.84 (image only) to 0.92**, highlighting the benefit of incorporating patient-specific clinical information. Confusion matrix analysis showed high precision for Good and Moderate response categories, indicating the model's effectiveness in identifying patients likely to respond favorably to therapy. Comparison of therapy response prediction for various models is provided in Table 2

*Table 2: Therapy response prediction comparison*

| Model | AUC | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| Image-only CNN | 0.84 | 0.79 | 0.78 | 0.77 | 0.77 |
| Multimodal Fusion (Ours) | 0.92 | 0.88 | 0.89 | 0.87 | 0.88 |

Comparison of therapy response prediction for various models is provided in Table 2

- Multimodal attention improved AUC from 0.84 → 0.92.
- Confusion matrix analysis indicated accurate identification of Good and Moderate responders.

**Interpretation:**

Integrating imaging and clinical features allows for early identification of non-responders, supporting proactive therapy adjustments and personalized treatment planning.

## VI. CONCLUSION

In this study, we presented a comprehensive **multimodal deep learning framework** for automated analysis of **Oral Lichen Planus (OLP)**, integrating **lesion segmentation** and **therapy response prediction**. By leveraging advanced architectures such as **Transformer-augmented U-Net** and **SegFormer**, the framework achieves **state-of-the-art segmentation performance** (Dice coefficient >0.87), enabling precise delineation of complex intraoral lesions.

We further demonstrated that combining **image embeddings** with **patient-specific clinical metadata** through **attention-based feature fusion** significantly improves predictive performance for immunosuppressive therapy outcomes, reaching an AUC of **0.92**. This hybrid approach facilitates early identification of non-responders, supporting **personalized treatment planning** and **proactive clinical decision-making**.

The proposed framework bridges the gap between **AI engineering** and **oral precision medicine**, offering a scalable, objective, and non-invasive tool for OLP management. Future work will focus on integrating histopathological images, expanding biomarker profiling, and deploying the system in real-world clinical workflows to validate its utility in longitudinal patient monitoring.

## FUTURE STUDY:

Future work will focus on expanding the dataset to include more diverse patient populations and incorporating additional data modalities such as histopathological images and genomic profiles to further enhance prediction accuracy. Improving model interpretability and validating the framework in real-world clinical settings are essential next steps to facilitate the translation of this AI tool into routine oral healthcare practice.

## REFERENCES

[1].Scully, C., & Carrozzo, M. (2008). Oral mucosal disease: Lichen planus. British Journal of Oral and Maxillofacial Surgery, 46(1), 15–21.

[2].Manchanda, Y., Rathi, S. K., Joshi, A., & Das, S. (2023). Oral lichen planus: An updated review of etiopathogenesis, clinical presentation, and management. Indian Dermatology Online Journal, 15(1), 8–23. https://doi.org/10.4103/idoj.idoj_652_22.

[3].Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. MICCAI.

[4].Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition. ICLR.

[5].Xie, E., et al. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. NeurIPS.

[6].Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., van Ginneken, B., Madabhushi, A., & Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820–838. https://doi.org/10.1109/JPROC.2021.3054390.

[7].Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey. IEEE TPAMI.

[8].Topalian SL, Drake CG, Pardoll DM.(2015) Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer Cell. 27(4):450-61. doi: 10.1016/j.ccell.2015.03.001. .