A Novel Feature Extraction Approach Using Deformable Adaptive Instance-Based U-Net Architecture for Segmentation and Classification of Oral Mucosal Lesion

S. M. Sagari
DSCE, Bengaluru, India
sagari-cs@dayanandasagar.edu (corresponding author)

Vindhya P. Malagi
DSCE, Bengaluru, India
vindhya-cs@dayanandasagar.edu

B. Chandrahas
DSCE, Bengaluru, India
chandrahas@gmail.com

Received: 3 April 2025 | Revised: 16 May 2025 and 7 June 2025 | Accepted: 9 June 2025

 DOI: https://doi.org/10.48084/etasr.11273

ABSTRACT

Oral cancer is one of the six cancer types having high morbidity and mortality rates, especially among socioeconomically deprived groups of people due to their lack of knowledge about oral hygiene. This study aimed to detect oral lesions in different areas of the oral cavity based on visual features of suspicious regions. The localization, detection, and classification of regions of interest in digital images stemming from diverse resolution cameras presents a formidable challenge due to the variation in illumination, image size, and varied noise. The proposed method employs image pre- and post-processing approaches to locate the regions effectively. A dataset of 2050 oral cavity images was used, having 1000 malignant, 700 benign, and 350 premalignant cases. The proposed method uses deformable convolution and instance normalization in the U-Net architecture to segment the region of interest by preprocessing the images using canny and local binary pattern feature extractors. These segmented regions are classified by combining the Bresenham circle and flood fill algorithms. The experimental analysis of the proposed approach showed precision, recall, and F1 scores of 93.85%, 97.37%, and 95.58% for noised malignant images and 96.20%, 96.82%, and 96.51% for denoised malignant images. Similarly, precision, recall, and F1 scores were 98.67%, 94.94%, and 96.77% for benign lesion noise images, and 96.95%, 96.36%, and 96.66% for benign denoised lesion images.

Keywords-oral cancer; mucosal lesion; U-Net

I. INTRODUCTION

The oral region of the human body is referred to as the oral cavity composed of oral mucosal or buccal cells. These cells act as a protection barrier in the biochemical processes of the oral region. The detriment of these cells is due to several factors, including genetics, exposure to harmful environmental customs, such as tobacco use, HPV infections, and DNA damage, leading to chronic irritation due to the presence of carcinogens that contribute to the development of oral cancers. The various patterns, structures, and appearances manifest different types of oral cancer, including squamous cell carcinoma, melanoma, sarcomas, and mucoepidermoid carcinoma. Oral cancer can occur in any part of the oral cavity in the lingual, gingiva, labium, palatum durum, palatum molle, and dentes [1]. Detecting and localizing these regions of interest and classifying them plays a vital role [2]. Oral cancer detection techniques have progressed significantly, and deep learning techniques use different modalities, including histopathology, to classify cells as malignant or benign [3-5]. Advances in image processing, computer vision, and machine learning in the field of healthcare have led to various modern approaches that achieve good accuracy in predicting different diseases by detecting and classifying regions of interest [5, 6]. Early detection of oral cancer can reduce morbidity and mortality rates [7-11].

The proposed method involves examining photographic images of oral cavities that include suspicious regions. The oral cancer image dataset was collected from Kaggle [12] and Mendeley [13], which comprise RGB images of oral cavity lesions. These images were captured using portable mobile and intraoral cameras. The images of both datasets were considered for analysis to locate the region of possible oral malignancies. These images were taken by medical professionals from various sources, including hospitals and medical institutions in Karnataka, India. The dataset in [13] comprises original and augmented data using rotation, resizing, and flipping. The datasets varied in size and illumination and entailed noise. The proposed work involves an effective approach to pre- and post-processing the dataset and suggests a novel approach to classifying the region as cancer or noncancer.

II. METHODOLOGY

The proposed approach performs a segmentation-based classification for a noisy oral cancer dataset containing many variations. The entire masking-based classification pipeline has multiple components that work in tandem.

```
Algorithm1: Proposed  method workflow
Input: Oral cancer labeled images
Output: Benign or Malignant
Procedure:
Step 1: Data collection and augmentation
Step 2: Preprocess images to remove noise
  and increase feature visibility
  combining Canny filter and Local Binary
  Pattern (LBP) methods
Step 3: Use the preprocessed images to
  train U-Net for segmentation
Step 4: Replace batch normalization with
  instance normalization
Step 5: The convolutional layer is
  replaced with a deformable convolution
  layer
Step 6: Postprocessing using guided
  filtering and Gaussian blur to smoothen
  the sharp edges after step 3
Step 7: Generate vectors to extract the
  textures as RoIs having consistently
  similar patterns using the Bresenham
  circle and flood fill algorithms
Step 8: Classification of texture
  extracted data using a CNN with global
  average pooling
```

A. Dataset and Augmentation

The dataset comprises a total of 2050 images with diverse instances of oral cancer occurrences. These instances cover various regions within the oral cavity, such as the tongue, inner cheeks, chin, and gums, as shown in Figure 1, among others. Images were annotated using Label Studio and subsequently reviewed by a certified oncologist. This dataset presents a challenge in terms of object detection due to the diverse shapes of the lesion regions. In addition, the dataset contains significant amounts of noise resulting from changes in background lighting during image capture. Notably, the dataset includes various regions of the oral cavity, leading to significant diversity for analysis. Therefore, this study aimed to identify a robust architecture that is more resilient to noise while also capturing all the variations present in the dataset.

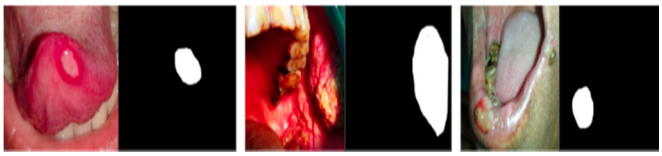

Fig. 1. Images and respective lesion regions on the tongue, inner cheek, and lips, respectively.

The training dataset comprises augmented images that incorporate five main variations: random rotation, image inversion, left-right flipping, and horizontal and vertical translation. During the training pipeline, these augmentations are combined as part of a permutation combination approach, decided in random order. In this way, the training dataset was populated with approximately 1000 augmented images per epoch. The nonaugmented images were reserved for validation. A separate subset of 40 images without augmentation was kept and used for measuring model generalizability, resulting in a train-test-validation split of 70:20:10.

B. Preprocessing

Various preprocessing techniques were employed to mitigate the influence of noise and improve feature visibility. Through numerous experiments, it was found that combining the Canny filter with Gaussian blur and the image generated using a variant of the LBP texture operator [14] yielded superior results. The optimal thresholds for the Canny filter were identified using the Otsu algorithm [15]. Superior results were observed when Canny and LBP were fused and the image was broadcast across all channels and added element-wise to the input image, as shown in Figures 2-7.

C. Segmentation Architecture Details

The baseline architecture adopts a U-Net architecture [16] with a depth of 4, featuring an encoder-decoder structure with identical blocks. Each block consists of two identical convolutions, followed by max pooling and batch normalization. The second convolution layer is concatenated from the encoder to the decoder block. The architecture concludes with a single-channel segmentation layer.

This model was improvised due to the variation in the dataset and the sensitivity of batch normalization to the batch size. Adaptive instance normalization [17] was adapted as a replacement for batch normalization to adapt to the different forms present in the images. The introduction of instance normalization in the model's architecture improves robustness and adaptability to variations in individual images, significantly increasing performance. The following equation implements the instance normalization, which has two learnable parameters (namely γ and β) for every feature map x present in the input.

$$IN\ (x; s) = \gamma^s\ (\frac{x - \mu(x)}{\sigma(x)}) + \beta^s \quad (1)$$

To further enhance the segmentation results, the penultimate convolutional layer is replaced with a deformable convolution [18] layer to adapt to varying shapes, providing better learning capabilities and adaptable receptive fields for the segmentation task, as shown in Figure 1.

D. Postprocessing

The use of deformable convolutions and instance normalization produces sharp and geometrically unpleasant results. To smooth out the edges after the segmentation results, a variant of guided filtering [19] is implemented to reduce structural inconsistencies in the segmentation results. The preprocessed input image is given as the guidance image with the target image being the postprocessed segmentation results.

E. Region Of Interest (RoI) Extraction

To extract the segmented region, binary masking was implemented using the Intersection over Union (IoU) metric. All the segmented regions were cropped and resized to a constant shape.

F. Bresenham Circle with Flood Fill

The Bresenham circle algorithm along with the scan floodfill algorithm [20] was utilized to identify the most important features that play a role in the classification task.

```
Algorithm 2: Proposed approach using
  Bresenham Circle Algorithm and flood
  fill for ROI extraction
Input:  Segmented Images
Output: Region of Interest (RoI)
Procedure:
Step 1: Identify the center of an image
  in both X and Y directions: center_x,
  center_y.
Step 2: Find the radius based on the
  center(X, Y) point of the image
Step 3: Locate the next pixel point by
  measuring the decision parameter d
  when x = 0 and y = radius
    d = 3 - 2 * radius
Step 4: Add points  to draw center_x and
  center_y if x <= y
Step 5: Draw a circle with radius as in
  step 2
  if d<0
    d= d+4*x+6
  else
    d=d+4*(x-y)+10
    y=y-1
    x=x+1
Step 6: Draw a circle based on identified
  points to extract the region of interest
Step 7: Apply the flood fill algorithm to
  fill the region of interest with the
  pixel color starting from the center(X,
  Y). Iterate over this filled region
  until reaching the boundary defined by
  the Bresenham circle algorithm.
```

This technique generates circular vectors that capture nuanced spatial relationships. The Bresenham circle algorithm combined with flood fill enables 2D convolution-based classification, preserving geometric properties and enhancing information extraction within circular regions. This approach extracts textures that have consistently similar patterns in the image.

G. Classification of the Texture-Extracted Data

A simple convolutional neural network is employed with global average pooling [21] rather than utilizing traditional fully connected layers. The final layer consists of 3 neurons, each representing the three classes.

III. RESULTS

Extensive experiments were performed that included multiple preprocessing techniques, architectural changes, and post-processing techniques, including the following:

- Preprocessing: Canny (with and without OTSU), Sobel, Prewitt, and LBP, and holistically-nested edge detection [22].
- Architectures: Baseline U-Net, U-Net with instance normalization, deformable U-Nets with batch/instance normalization, transfer learning with MobileNet and EfficientNetB0.
- Postprocessing: Guided filtering and Gaussian blur (for smoothing).

These techniques were utilized during the experimentation. The U-Net architecture consisted of 4 blocks, each containing two convolutions separated by a normalization layer. All the experiments were performed using the Tesla T4 (12 GB) GPU provided by Google Colab. All models were trained for 100 epochs with a batch size of 16, a linear learning rate decay scheduler that decreases the learning rate from 0.003 to 0.0003, and the Adam optimizer. For the loss function, Binary Cross Entropy (BCE) was initially used but had to be replaced because the dataset had regions of different sizes and the model was extremely sensitive to the class weights.

$$L_{BCE}(y,\bar{y}) = -\frac{1}{n}\sum_{i=1}^{n} y_i \cdot log(\bar{y}_i) + (1-y_i)\cdot log(1-\bar{y}_i) \quad (2)$$

A combination of the WBCE and the Dice Loss was chosen as the loss function for all experiments [23]. Using instance normalization with a small batch size risks overfitting, hence L2 norm was added to the loss function. Additionally, the loss components were controlled with a constant through the hyperparameters α (0.9) and β (0.001), respectively.

$$L_{Dice}(y,\bar{y}) = 1 - \frac{2\cdot|y \cap \bar{y}|}{|y| + |\bar{y}|} \quad (3)$$

$$L_{tot} = L_{BCE} + \lambda \cdot L_{Dice} + \beta \cdot |W|^2 \quad (4)$$

where $|W|^2$ is L2 regularization, implemented to reduce overfitting and smoother initial training iterations. The transfer learning techniques failed to provide good results as the domains were completely different with minimal knowledge transfer capabilities.

For further analysis, the segmented region is postprocessed with guided filtering to smooth out the edges. Once the complete masked image is generated, it is used to mask the input image to extract the RoI. This RoI is cropped, and the Bresenham circle and the flood fill algorithms are applied to extract the relevant features from the image. These features are then passed through a simple classification model containing three convolutional layers followed by a global pooling layer and a Multi-Layered Perceptron (MLP) containing two neurons, each for one class: Benign and Malignant.

IV. DISCUSSION AND COMPARISON

The performance of each experiment was mainly evaluated based on the IoU score with a threshold of 50%. The models were deemed successful or failed based on whether they exceeded or fell short of this IoU threshold, respectively. As mentioned in previous sections, the testing was performed on original nonaugmented images. The results were benchmarked using the IoU and F1 scores.

Table I compares and benchmarks all the experiments performed on the test data. As expected, the existing approaches did not yield optimal results due to the number of variations and noise present during the training. The figures below depict the results of the models on the dataset images.

TABLE I. COMPARISON OF SEGMENTATION RESULTS WITH VARIOUS STATE-OF-THE-ART MODELS

Experiments	IoU	F1	Passed	Failed	Accuracy
UNet	0.177	0.27	82	58	10.37
UNet+Sobel	0.255	0.165	79	61	14.39
UNet+Canny+LBP	0.34	0.235	100	40	24
UNet+Canny+Instance normalization	0.47	0.342	122	18	40.96
Deformable UNet+ Instance+Canny+LBP	0.47	0.332	118	22	39.62
Deformable UNet+ Instance+Canny+LBP+ Guided filtering (proposed)	**0.48**	**0.54**	**128**	**12**	**43.90**

Once the segmented regions are identified, the appropriate regions are masked from the image, and the Bresenham circle algorithm is applied to it along with the flood fill algorithm. The circle algorithm ensures that no point outside the circle is considered and the flood fill algorithm considers all points that have similar patterns as the masked regions, as shown in Figure 8. Here, it can be seen that the classification model was able to identify all the regions of the Benign class correctly, based on the masked RoI using the segmentation. However, there are some flaws in predicting the other two classes. The overall accuracy of the classification model is around 96.43%. Table II shows a comparative analysis when considering noisy and denoised Malignant and Benign lesion region images. The results indicate the reliability of the proposed approach across different image conditions, highlighting its relevance in imaging applications where noise reduction is essential without sacrificing accuracy.

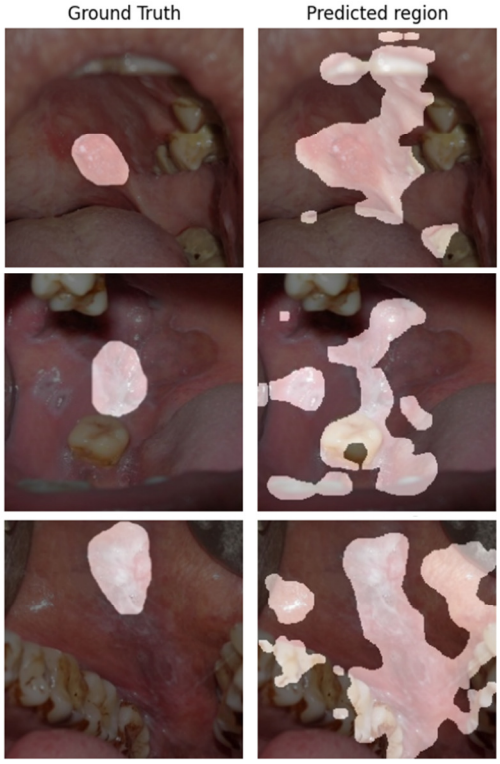

Fig. 2. Results of the baseline UNet model.

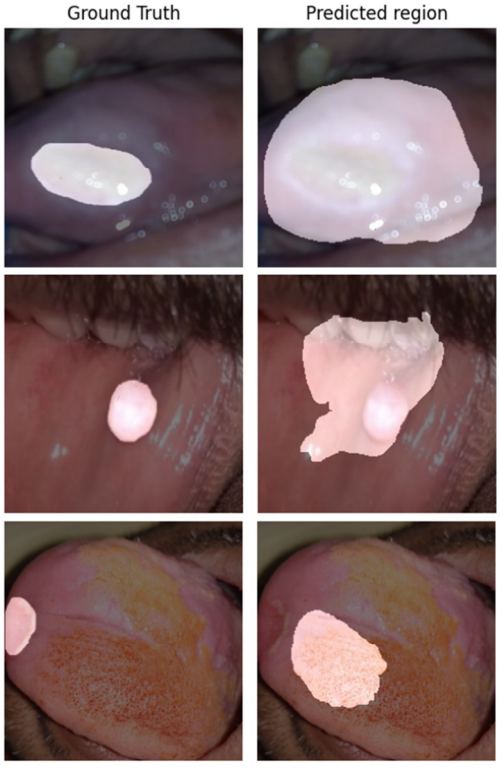

Fig. 3. Results from the Gaussian blur and Sobel with UNet.

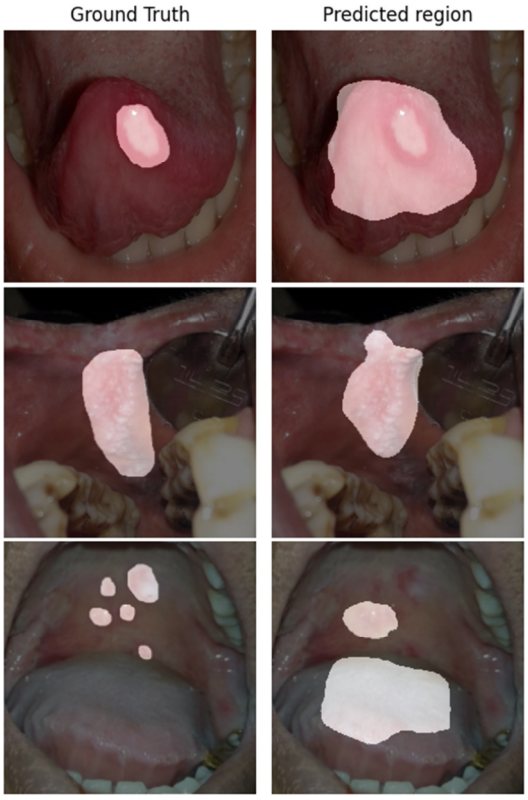

Fig. 4. Results from Canny Edge and LBP with U-Net.

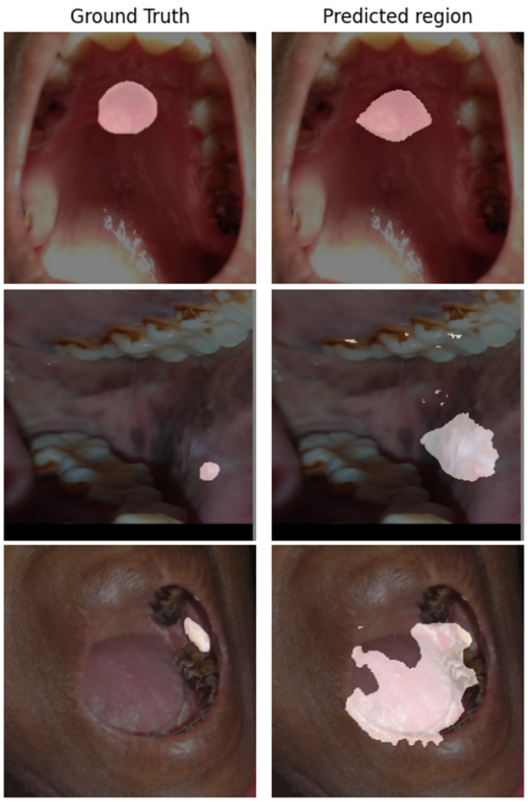

Fig. 5. Results from the Canny + LBP with instance normalized U-Net.

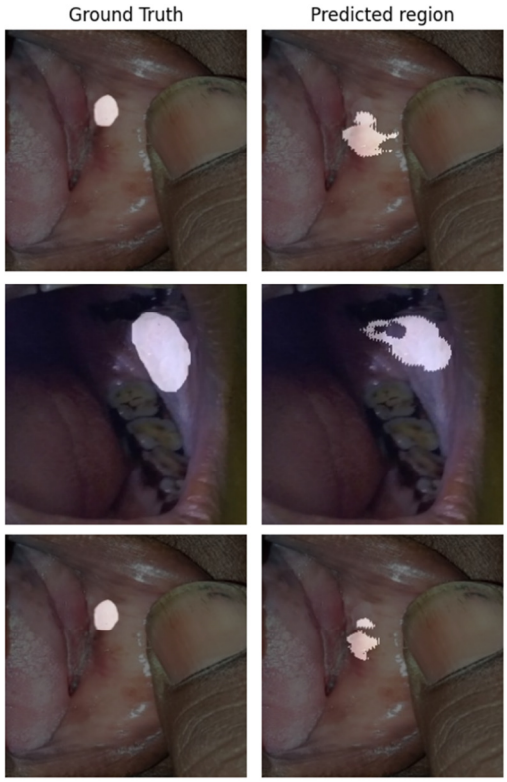

Fig. 6. Results from Canny+LBP with deformable U-Net.

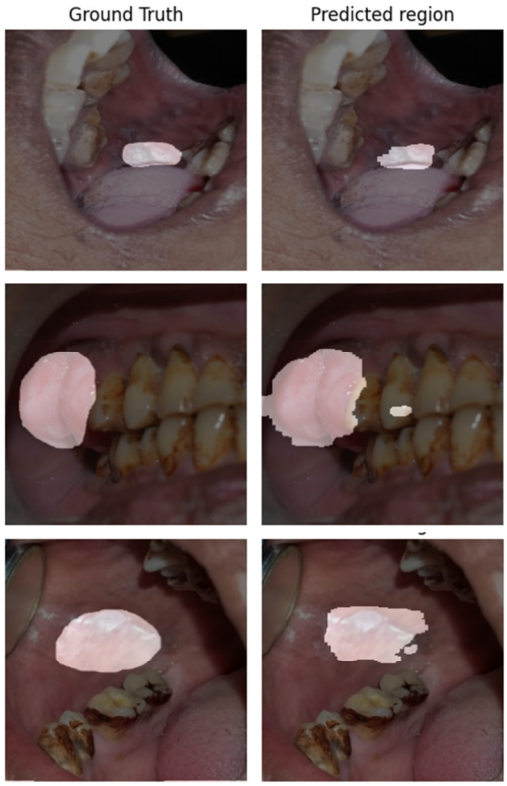

Fig. 7. Results from Canny+LBP with deformable U-Net and guided filtering.

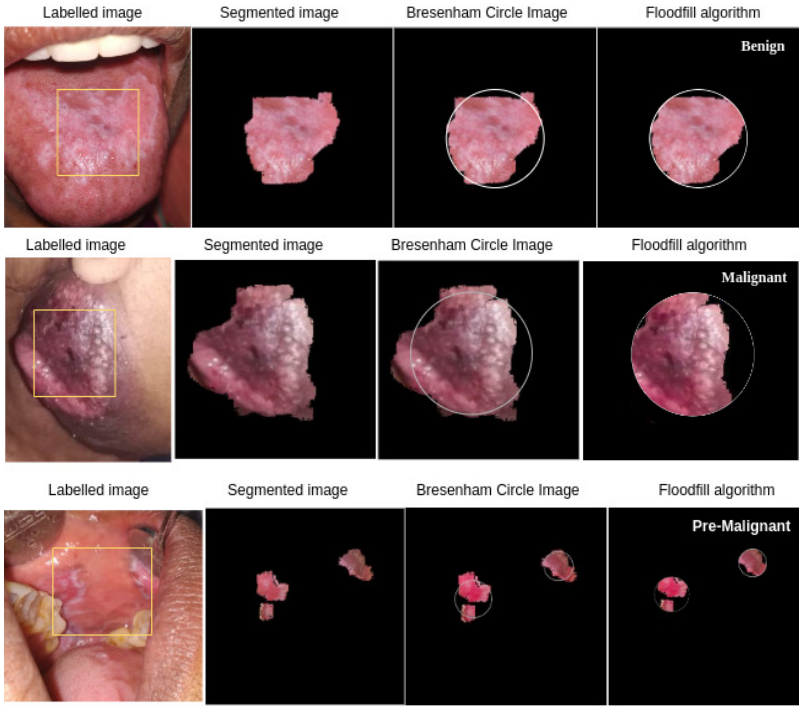

Fig. 8. Results from the Bresenham circle + flood fill.

TABLE II. COMPARATIVE ANALYSIS ON NOISY AND DENOISED IMAGE DATASETS

Class label	Noisy images (140 images)			Denoised images (322 images)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Malignant	93.85	97.37	95.58	96.20	96.82	96.51
Benign	98.67	94.94	96.77	96.95	96.36	96.66

Figures 9-11 depict ROC, a confusion matrix, and images that were misclassified along with failed cases in locating the RoI. Future work will focus on adding more augmentation techniques to reduce the misclassification of the region to improve accuracy.

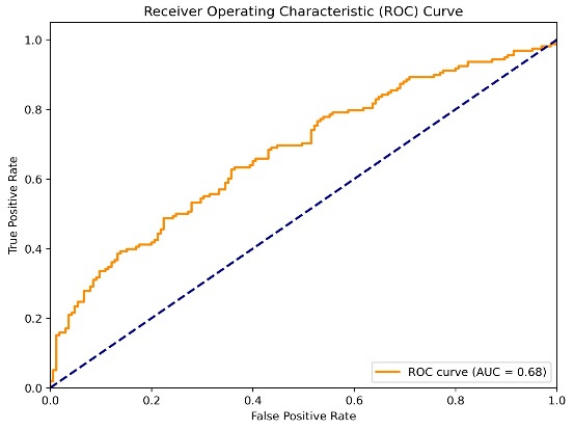

Fig. 9. ROC curve on the model performance.

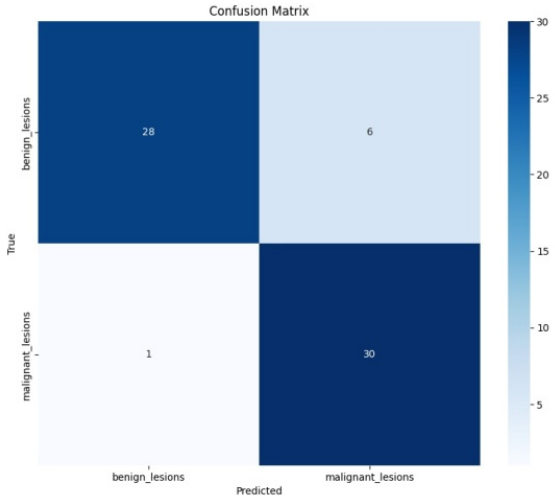

Fig. 10. Confusion matrix for the RoI-based classification.

V. CONCLUSION

Oral lesions require medical practitioners to identify suspicious regions at their early stages for early treatment plans. This study used deep-learning approaches to segment and classify the regions into Benign and Malignant. Extraction of features in RGB images plays a pivotal step in enhancing the detection and recognition of oral cancer regions. The deformable convolution layer provides the ability to adapt to varying shapes, providing better learning capabilities and adaptable receptive fields for the segmentation task. Then, binary masking was implemented using the IoU metric. All the segmented regions were cropped out and resized to a constant shape, thus avoiding unrelated surrounding region information.

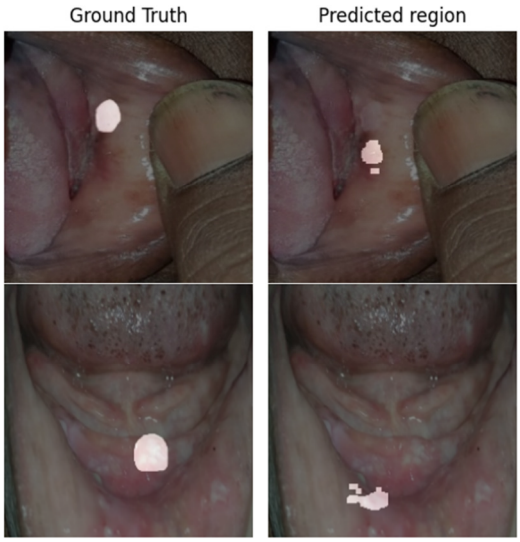

Fig. 11. Results of failure cases and misclassified images.

The novelty of the proposed method lies in the combination of the Bresenham circle algorithm with flood fill, which enables 2D convolution-based classification, preserving geometric properties and enhancing information extraction within circular regions. This approach extracts the textures having consistently similar patterns in the image. The study in [24] focused on the classification of malignant transformations but could not identify nuanced details based on effective similar pattern matching. This study focused on similar pattern matching by identifying the spread of similar pixels within the region. The comparative analysis showed that the proposed model achieved high F1 scores in noisy and denoised images. Future work could focus on improving accuracy using a deep architecture, such as DeepLab, attention, and transformers.

ACKNOWLEDGMENT

The authors thank Dr. Hamsa Nandini, Consultant Head and Neck surgeon, Kerudi Cancer Hospital, Bagalkot, for validating the results and for her constant support and input.

REFERENCES

[1] P. H. Montero and S. G. Patel, "Cancer of the Oral Cavity," *Surgical Oncology Clinics of North America*, vol. 24, no. 3, pp. 491–508, Jul. 2015, https://doi.org/10.1016/j.soc.2015.03.006.

[2] K. W. Aschheim, Esthetic Dentistry: A Clinical Approach to Techniques and Materia, 3rd ed. Mosby, 2015.

[3] A. Chaurasia, S. I. Alam, and N. Singh, "Oral cancer diagnostics: An overview," *National Journal of Maxillofacial Surgery*, vol. 12, no. 3, pp. 324–332, Sep. 2021, https://doi.org/10.4103/njms.NJMS_130_20.

[4] A. Rahman *et al.*, "Histopathologic Oral Cancer Prediction Using Oral Squamous Cell Carcinoma Biopsy Empowered with Transfer Learning," *Sensors*, vol. 22, no. 10, May 2022, Art. no. 3833, https://doi.org/10.3390/s22103833.

[5] M. S. Rahman *et al.*, "Evaluation of a low-cost, portable imaging system for early detection of oral cancer," *Head & Neck Oncology*, vol. 2, no. 1, Apr. 2010, Art. no. 10, https://doi.org/10.1186/1758-3284-2-10.

[6] A. Sungheetha and R. Sharma R, "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network," *Journal of Trends in Computer Science and Smart Technology*, vol. 3, no. 2, pp. 81–94, Jul. 2021, https://doi.org/10.36548/jtcsst.2021.2.002.

[7] B. W. Neville and T. A. Day, "Oral Cancer and Precancerous Lesions," *CA: A Cancer Journal for Clinicians*, vol. 52, no. 4, pp. 195–215, 2002, https://doi.org/10.3322/canjclin.52.4.195.

[8] T. Baykul, H. Yilmaz, Ü. Aydin, M. Aydin, M. Aksoy, and D. Yildirim, "Early Diagnosis of Oral Cancer," *Journal of International Medical Research*, vol. 38, no. 3, pp. 737–749, Jun. 2010, https://doi.org/10.1177/147323001003800302.

[9] T. O. Bittar, L. R. Paranhos, D. H. Fornazari, and A. C. Pereira, "Epidemiological features of oral cancer: a world public health matter," *RFO UPF*, vol. 15, no. 1, pp. 87–93, Apr. 2010.

[10] D. P. Slaughter, H. W. Southwick, and W. Smejkal, "Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin," *Cancer*, vol. 6, no. 5, pp. 963–968, Sep. 1953, https://doi.org/10.1002/1097-0142(195309)6:5<963::aid-cncr2820060515>3.0.co;2-q.

[11] C. Scully, J. V. Bagan, C. Hopper, and J. B. Epstein, "Oral cancer: current and future diagnostic techniques," *American journal of dentistry*, vol. 21, no. 4, pp. 199–209, Aug. 2008.

[12] "Oral Cancer (Lips and Tongue) images." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images.

[13] B. R. Nanditha *et al.*, "Oral Images Dataset." Mendeley, Feb. 05, 2021, https://doi.org/10.17632/MHJYRN35P4.2.

[14] J. Lizé, V. Débordès, H. Lu, K. Kpalma, and J. Ronsin, "Local Binary Pattern and Its Variants: Application to Face Analysis," in *Advances in Smart Technologies Applications and Case Studies*, 2020, pp. 94–102, https://doi.org/10.1007/978-3-030-53187-4_11.

[15] Y. Wang, L. Shi, J. Lausanne, and D. Zhong, "Straight lane line detection based on the Otsu-Canny algorithm," in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China, Mar. 2022, pp. 27–30, https://doi.org/10.1109/ITOEC53115.2022.9734320.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

[17] X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Oct. 2017, pp. 1510–1519, https://doi.org/10.1109/ICCV.2017.167.

[18] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets V2: More Deformable, Better Results," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 9300–9308, https://doi.org/10.1109/CVPR.2019.00953.

[19] Z. Shi, Y. Chen, E. Gavves, P. Mettes, and C. G. M. Snoek, "Unsharp Mask Guided Filtering," *IEEE Transactions on Image Processing*, vol. 30, pp. 7472–7485, 2021, https://doi.org/10.1109/TIP.2021.3106812.

[20] Y. He, T. Hu, and D. Zeng, "Scan-Flood Fill(SCAFF): An Efficient Automatic Precise Region Filling Algorithm for Complicated Regions," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, Jun. 2019, pp. 761–769, https://doi.org/10.1109/CVPRW.2019.00104.

[21] M. Lin, Q. Chen, and S. Yan, "Network In Network." arXiv, Mar. 04, 2014, https://doi.org/10.48550/arXiv.1312.4400.

[22] S. Xie and Z. Tu, "Holistically-Nested Edge Detection," *International Journal of Computer Vision*, vol. 125, no. 1, pp. 3–18, Dec. 2017, https://doi.org/10.1007/s11263-017-1004-z.

[23] A. Galdran, G. Carneiro, and M. A. G. Ballester, "On the Optimal Combination of Cross-Entropy and Soft Dice Losses for Lesion Segmentation with Out-of-Distribution Robustness," in *Diabetic Foot Ulcers Grand Challenge*, 2023, pp. 40–51, https://doi.org/10.1007/978-3-031-26354-5_4.

[24] B. R. Nanditha, G. Kiran, and A. M. P. Sanathkumar, "Oral Cancer Detection using Machine Learning and Deep Learning Techniques," *International Journal of Current Research and Review*, vol. 14, no. 01, pp. 64–70, 2022, https://doi.org/10.31782/IJCRR.2021.14104.