Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Instance Segmentation of Oral Cancer Images with Fusion of Swin Transformer and Mask RCNN

Kavyashree C.[1], Vimala H S.[2]

Department of Computer Science & Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, India.

E-mail: [1]kavyashree.csd@gmail.com, [2]vimalahsuvce25@gmail.com

Abstract

Oral cancer is the most preventable cancer if it is diagnosed at an early stage. Artificial intelligence can be a great help in cancer detection. Deep learning architectures are predominantly useful in medical image analysis by identifying patterns and the ability to predict the insights. The study proposes a deep learning methodology using Mask RCNN (Region Based Convolutional Neural Network) for the precise detection and segmentation of oral lesions in photographic images. With the swin transformer as a backbone, it aids the model in extracting features more effectively, thus supporting precise detection. Its ability to identify relationships among different parts of an image is particularly useful in locating the smallest lesions. The precise annotation has helped generate the segmentation mask accurately. The model attains a mean average precision (mAP) of 99.5%, a precision of 92.7% and a recall of 96.6%. This exceptional performance of the model is useful for the medical community to use it as a tool for the early detection of oral cancer.

Keywords: Oral Cancer, Mask RCNN, Swin Transformer; Object Detection, Instance Segmentation.

1. Introduction

Cancer is the uncontrollable growth of cell invasion that damages neighboring tissue. It is an abnormal condition or sore in the oral cavity including the cheek, lips, tongue, and mouth and it can spread to the oropharynx [1] [2]. Cancer is the second most common cause of mortality worldwide. Oral cancer can be attributed to various risk factors, such as tobacco consumption, including smokeless tobacco, chewing betel quid, extreme alcohol consumption, lack of oral hygiene, a diet lacking essential nutrients, and viral infections like human papillomavirus (HPV). Oral squamous cell carcinoma (OSCC) significantly accounts for oral cancer cases. According to the American Cancer Society's estimate for 2024, there have been 116,900 cases of oral cancer diagnosed in America, with 24,460 of those cases ending in death [3]. The statistics also show that it is significantly higher in males compared to females. It is evident from the study that oral cancer is curable if diagnosed at an early stage. Oral cancer diagnosis is done using various methods such as physical examination, biopsy, imaging techniques, spectroscopy, and biomarker detection.

DOI: https://doi.org/10.36548/jiip.2025.3.007
Received: 06.07.2025, received in revised form: 05.08.2025, accepted: 19.08.2025, published: 29.08.2025

The advancement of Artificial Intelligence (AI) has significantly improved the speed and affordability of cancer diagnosis [4]. The data gathered from diverse techniques is employed for model training and can be utilized for real-time diagnosis. Data can be gathered from imaging techniques like MRI scans, CT scans, PET scans and even photographic images. Advanced deep learning methods are employed to categorize and identify objects within an image. They help to automate the process of extracting features from images to detect objects and classify images. Pretrained models like DenseNet, Resnet and so on are used for feature extraction. Several object detection algorithms based on deep learning help in automating cancer detection. Feature Pyramid Networks (FPN) are utilized to acquire additional information or valuable features at a reduced resolution. Region-based Convolutional Neural Networks (R-CNN) were developed for the purpose of feature extraction and generating region proposals [5]. These region proposals are used to generate the Region of Interest (RoI) by implementing selective search algorithms. Furthermore, they can be classified and used to locate the objects in an image. R-CNN has evolved over time with enhanced models like Fast R-CNN, Faster R-CNN and Mask R-CNN. Fast R-CNN accelerates the model training process by utilizing shared convolutional features across various regions. The introduction of the Region Proposal Network (RPN) in Faster R-CNN allowed for the seamless training of region proposals and the efficient handling of object detection tasks [6]. Faster R-CNN is primarily designed for object detection and offers only bounding box-level localization, which is insufficient for medical image analysis where pixel-level precision is essential. Its lack of segmentation capabilities limits its effectiveness, particularly in identifying small or irregularly shaped regions common in medical imaging. While Mask R-CNN extends Faster R-CNN by enabling instance segmentation, it still relies heavily on local feature extraction, restricting its ability to model long-range dependencies and global context. Conversely, transformer-based architectures like the Swin Transformer excel at capturing hierarchical and global features but fall short in providing the detailed localization accuracy required for precise segmentation. To address these complementary limitations, this research proposes a fusion of the Swin Transformer with Mask R-CNN. The objective is to improve the accuracy and robustness of instance segmentation in oral cancer images by integrating global attention mechanisms with precise region-based predictions. This hybrid approach is intended to overcome the shortcomings of existing methods and push the boundaries of current segmentation performance in the domain of medical imaging.

The contributions of this study are:

- Implementation of modified Mask R-CNN along with swin transformer to detect the cancer or non-cancerous part and generate a mask for it.
- Enhanced feature extraction with use of swin transformer and reduced complexity.
- Comparative evaluation of the proposed model in comparison to other cutting-edge models.

The organization of research paper is follows. Section 2 provides the literature study, followed by data collection and processing in section 3. The methodology is discussed in section 4, with experimental set up and performance analysis is discussed in section 5 and 6. Section 7 provides the conclusion of the proposed work.

2. Related Work

Convolutional Neural Networks (CNN) are deep neural networks used to analyze images. They are not only used for image recognition but also for segmentation tasks, which involve labeling each pixel in an image. CNNs often struggle with detecting objects of various sizes and locations in an image. To overcome this problem, RCNNs use a selective search algorithm [7]. It generates region proposals that could possibly contain objects. Mask RCNN is part of the RCNN family, focused on performing instance segmentation. This is used to generate masks for the detected objects. Zhao et al. [8] used it to identify irregularities in dental X-rays. The model attained a high accuracy of 95.41%, thereby demonstrating its efficiency. Guo et al. [9] applied Mask RCNN to dental X-rays, significantly improving the inference time to 0.12 seconds to segment non-normal teeth in an image. This helped in early detection and reducing patient costs. Brahmi et al. [10] used Mask RCNN for object instance segmentation, achieving an acceptable result with 90% mAP and 96% precision. Even though it is a small dataset, it consists of real data obtained from hospitals and annotated with the help of experts. This accurate annotation contributed to obtaining good results. Fatima et al. [11] proposed a model with the objective of automating dental disease detection from X-ray images using Mask RCNN with MobileNet-v2 as a backbone. The major advantage is the reduced complexity and enhanced results, with an accuracy of 94% and mAP of 85%, making it suitable for real-time deployment. This model has also been tested with a limited dataset and hence needs validation. Zhang et al. [12] worked on the detection of breast cancer using MRI images with Mask RCNN. They used two datasets to verify the detection of lesions. Both provide good specificity and are able to classify normal and lesion regions. The major outcome of this research is to significantly reduce false positives. Yuan et al. [13] applied a deep learning model to improve diagnosis performance in thyroid carcinoma patients. Ultrasound images were used for segmentation and further classification. The model was trained and tested with 1,000 images, and results were validated with those of a sonographer. This exhibits the power of deep learning models in medical image analysis.

Table 1. Summary of Literature study

Author	Precision	Recall	F1-score	mAP
Zhao et al. [8]	0.965	0.96	0.965	
Guo et al. [9]	0.795	0.956		
Brahmi et al. [10]	0.96		0.63	
Fatima et al. [11]	0.86	0.89	0.89	0.85
Zhang et al. [12]	0.96	0.91		
Yuan et al. [13]	0.886	0.856		88.8
Soltani et al. [14]	0.96	0.99		
Kang et al. [15]	0.667	0.7	0.683	
Shetty et al. [16]				0.911

Soltani et al. [14] used transfer learning to segment and classify the lesions with Mask R-CNN. The transfer learning helped in achieving better results, with a high accuracy of

99.44% using the DenseNet121 model. The use of images that contain lesions in the INbreast dataset reduces complexity during the training of the deep learning model. Kang et al. [15] focused on the class imbalance and bias problem in diagnosing oral diseases. The research focused on identifying objects at different scales, emphasizing both local and global attention in a patch. The data augmentation technique helped in effectively addressing the class imbalance problem. The model performed exceptionally well, improving performance by over 40% in recall and F1 scores. DenseNet121 was used to reduce the model capacity and encode patches. Shetty et al. [16] tested Mask R-CNN with various ResNet and MobileNet-based architectures as backbones. The popular image processing technique, named contrast enhancement, was used to improve image quality and thereby enhance detection. Table 1 shows the results of the studies considered in the review. The related work demonstrates the use of Mask R-CNN with various backbone architectures for the segmentation and classification of different types of abnormalities in several disease diagnoses. The limitation identified is the availability of datasets in large volumes, which limits the validation of the model to its full potential. The proposed research addresses these issues and proposes a solution for the accurate and efficient segmentation and classification of oral cancer photographic images.

3. Data Collection and Processing

The research utilizes an oral cancer photographic image dataset obtained from the Roboflow universe [17]. A subset of the images (30) was also obtained from Suraksha Specialty Dental Care, Hosakote, India. A total of 1,500 images were collected from both sources. The dataset includes not just the cancerous portions in an image but also some of the cavity portions, which have been disregarded in the study. The images are transformed and resized to 640x640 to support the architecture for further processing. This high resolution allows Swin Transformer's window attention to focus on localized and small lesions. The images are annotated with labels "Cancer" and "Non-cancer" using the VGG Image Annotator (VIA) [18] [19]. This uses polygon-based annotation that ensures accurate coverage of the region. The annotations are exported in JSON file format for further training. The dataset was further split into 80% training, 10% validation, and 10% testing.

4. Methodology

The study proposes the development of deep learning based oral cancer detection using Mask R-CNN by systematically integrating it with the Swin transformer. Figure 1 shows the proposed methodology. Figure 2 shows the proposed architecture with the fusion of the Swin transformer and Mask RCNN.

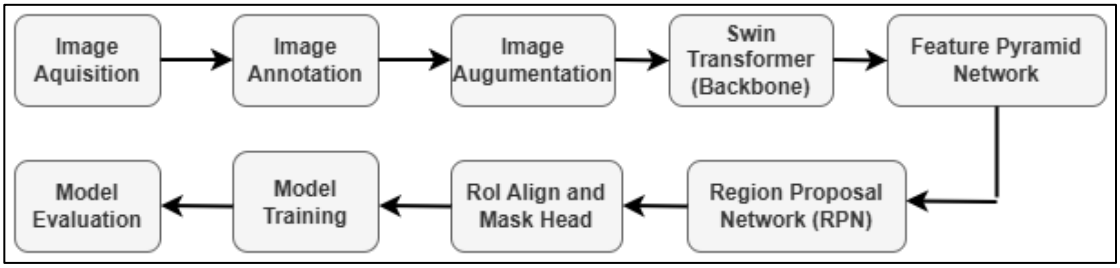

Figure 1. Flow of the Proposed Work

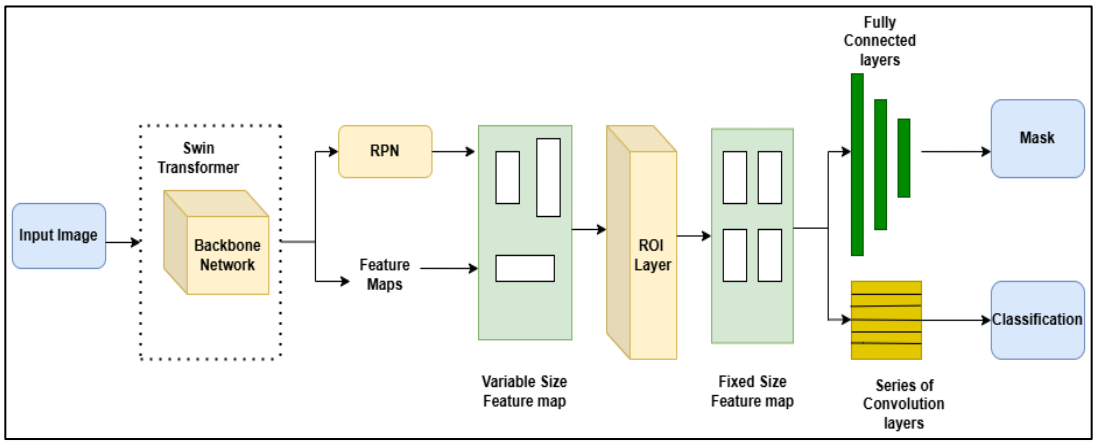

Figure 2. Proposed Architecture for Oral Cancer Detection

4.1 Swin Transformer

Transformers represent a deep learning architecture primarily utilized in natural language processing (NLP) tasks and have recently been investigated for their potential in computer vision applications. The Swin (Shifted window) transformer partition the image into non-overlapping windows rather than applying self-attention to the entire image [20]. Attention is computed within these individual windows. By adjusting the positions of the windows in subsequent layers, this method effectively captures the relationships between different windows, thereby enhancing the understanding of global context. This has moved from quadratic complexity to linear, thereby reducing computational complexity. The hierarchical design generates feature maps at various scales. This approach is advantageous for identifying objects of different sizes within images. It helps to capture complex patterns and relationships in images that are crucial for medical image analysis. The Swin transformer has different versions like tiny, small, base, large and huge. The study uses Swin-B (base) because of its ability to provide high accuracy in object detection and segmentation tasks.

4.2 Mask RCNN

Mask RCNN is a well-known framework for instance segmentation that is constructed on the foundation of faster RCNN [21]. It produces a top-notch segmentation mask for the identified object. The process consists of two phases: the region proposal network (RPN) is used to locate the objects, while the network head is responsible for classifying the objects and creating the mask. The image is scanned by the RPN using a sliding window technique to search for regions that may contain objects. It generates anchor boxes that are used for analysis in the subsequent stages. As the study does not use the bounding boxes, the RPN is not used. The model excludes the bounding box component in Mask RCNN, to improve the segmentation of small and irregular objects in the cancer images. This approach also reduces computational complexity. The generated proposals are fed to the RoI (Region of Interest) Align stage along with the feature maps generated from the swin transformer. RoI align utilizes bilinear interpolation to estimate the value of a function at any given point within the grid created by the collected data points. Thus the extracted features align with objects, ensuring accurate segmentation and classification. The head part takes care of generating a mask for the identified RoI.

The decision-making process relies on the features extracted through fully connected layers. It was also used for the purpose of categorizing between cancer and non-cancer cases. The study primarily concentrated on creating the mask and conducting classification with the confidence score, therefore the bounding box was not taken into consideration.

The methodology uses the Swin-B version as a backbone network to extract the feature maps. The hierarchical feature maps are fed to the feature pyramid network (FPN) for further refinement. This has been passed through RoI Align and is adjusted to be compatible with the feature maps generated from the swin transformer.

4.3 Loss Function

The loss function guarantees that the model is able to accurately segment objects, thus providing precise instance segmentation in images [22].

The loss function of the proposed Mask RCNN is given in equation (1)

$$\text{Lall} = \text{Lcls} + \text{Lmask} \tag{1}$$

Lall represents the overall loss function of the model. L specifies the classification loss, while Lmask is the map loss that specifies the average binary cross-entropy loss, exclusively defined for the ground truth boxes. The loss function aids in separating the class prediction from the mask generation process, making it more convenient to train and capable of yielding superior results. Lbox is ignored in our study to minimize computational cost and focus more on generating precise mask.

$$\text{Lcls} = -\frac{1}{A}\sum_{i=1}^{A}(\text{gi} * \log(\text{pi})) + ((1 - \text{gi}) * \log(1 - \text{pi})) \tag{2}$$

Where A specifies the number of anchor boxes used for classification, gi represents the ground truth label and pi is the probability of the predicted class.

$$\text{Lmask} = -\frac{1}{N}\sum_{i=1}^{N}(\text{gi} * \log(\text{pi})) + ((1 - \text{gi}) * \log(1 - \text{pi})) \tag{3}$$

N represents the total number of mask pixels, gi is the ground truth value of the ith pixel and pi refers to the predicted probability of pixel I, considering the sigmoid activation function.

The complete Mask R-CNN model undergoes end-to-end training through backpropagation in order to reduce the overall loss. The loss function significantly contributes to improving object detection and segmentation accuracy.

The proposed methodology uses the Swin transformer as a backbone and is used as the feature extractor. The fusion between the Swin Transformer and Mask R-CNN primarily occurs at the feature level, where the output of the Swin transformer feeds into the downstream components of Mask R-CNN.

5. Experimental Setup

The proposed Mask R-CNN with the Swin Transformer as the backbone model is trained using a photographic image dataset. Table 2 shows the optimal parameters identified to

enhance the performance of the model. The model is trained to differentiate between the cancer and non-cancer classes after iterating for 200 epochs. The optimizer and learning rate help the model to learn and perform better. The various thresholds used in the table help to identify objects at various scales and generate a pixel-level mask on them.

Table 2. Hyperparameters for Fine-Tuning the Model

Hyper parameters	Value
Backbone	Swin Transformer
Image scaling	640, 640
Learning Rate	0.001
Batch Size	10
No. of epochs	200
No. of classes	2+1
Optimizer	Adam
Object threshold	0.7
NMS threshold	0.3
Weight decay	0.001

6. Performance Analysis

This section analyses the performance of the proposed model using the metrics precision, recall and mAP. Precision mainly focuses on the positive predictions of the model. Recall concentrates on the correct predictions of positive instances. mAP is an important performance metric that focuses on the average precision with across number of classes. IoU (Intersection over Union) is another important metric for the generation of segmentation masks and classification.

$$Precision = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \tag{4}$$

$$Recall = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{5}$$

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{6}$$

$$mAP = \frac{1}{\text{N}} \sum_{i=1}^{N} \text{APi} \tag{7}$$

Table 3 shows the results of the proposed model with the mAP going up to 99.5. The mAP computed for the training data is shown in Figure 3 for the threshold and the mean of all the values from 0.5 to 0.95. After the model was trained, it was tested with test dataset that consists of 150 images. A high confidence interval of 95% is considered in the study and the model could able to identify the cancerous and non-cancerous regions effectively. Figure 4 shows the classification loss of the model during training and indicates reduced errors and better classification performance. Figure 5 shows the object loss of the model, thereby reducing the

misclassifications. These greatly reduce the error and minimizes false positives and negatives improving the performance of oral cancer detection. Table 3 shows the comparison of the proposed model with other models, and Figure 6 shows the graphical representation of it. It is compared with Mask RCNN and other CNN based backbone networks, as well as with U-Net models. The proposed model performs better with the good precision and recall values compared to other models.

Table 3. Result Analysis of the Proposed Model and Comparison with the Other Models

Model	Precision	Recall	mAP
Proposed Model	92.17	96.6	99.5
U-Net [24]	70	52	
U-Net++ [25]	96.55	96.59	
Mask R-CNN+ ResNet101	90.2	93.4	97.6
Mask R-CNN + ResNet152	91.1	92.6	92.5
Mask R-CNN +DenseNet121	90.2	92.1	91.2

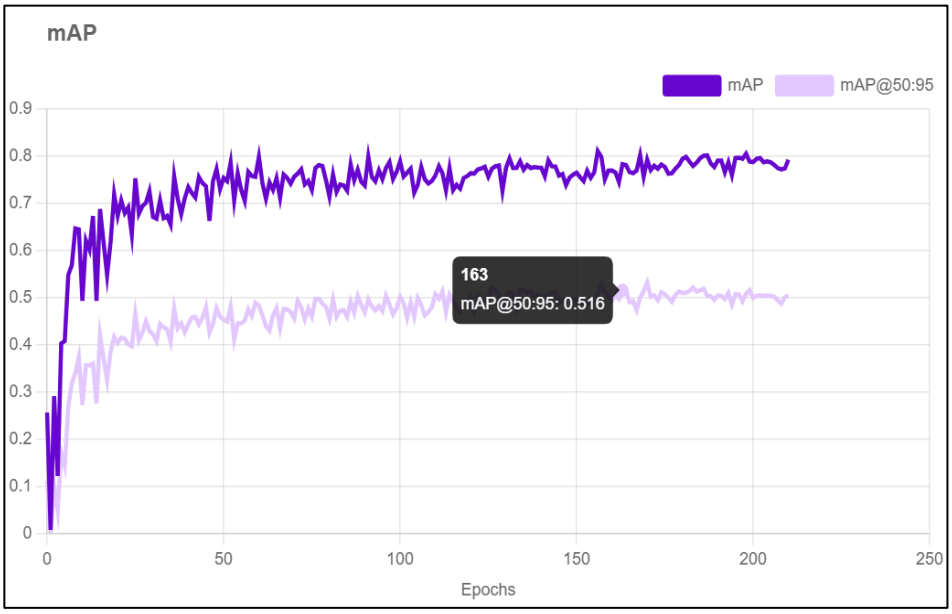

Figure 3. mAP of the Proposed Model

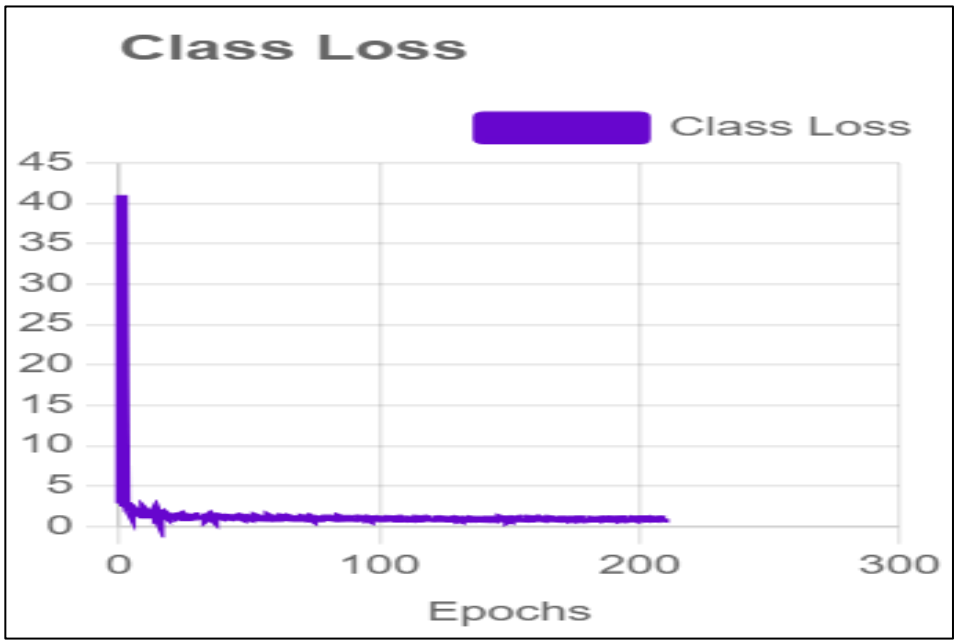

Figure 4. Classification Loss of the Proposed Model

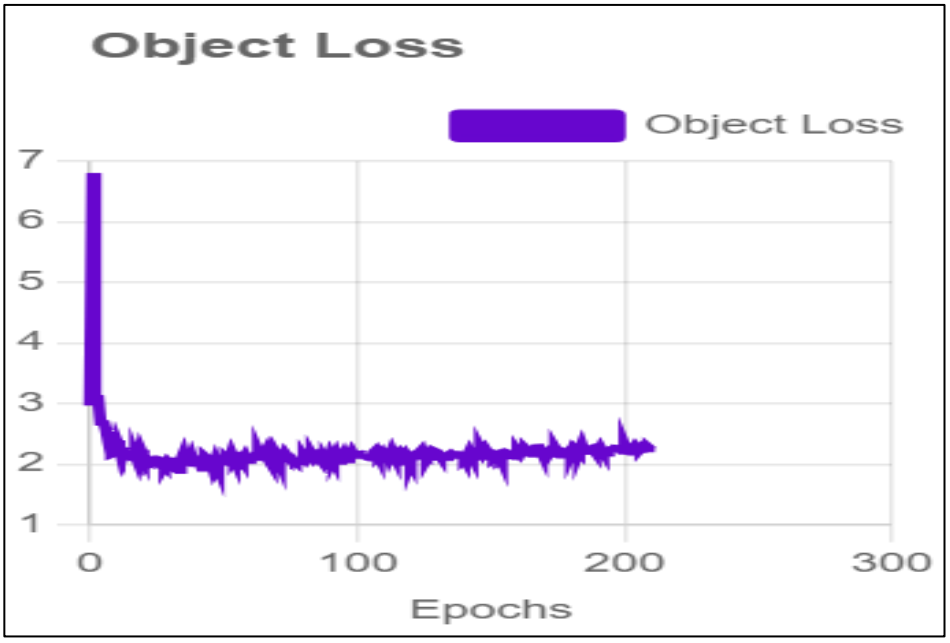

Figure 5. Object Loss of the Proposed Model

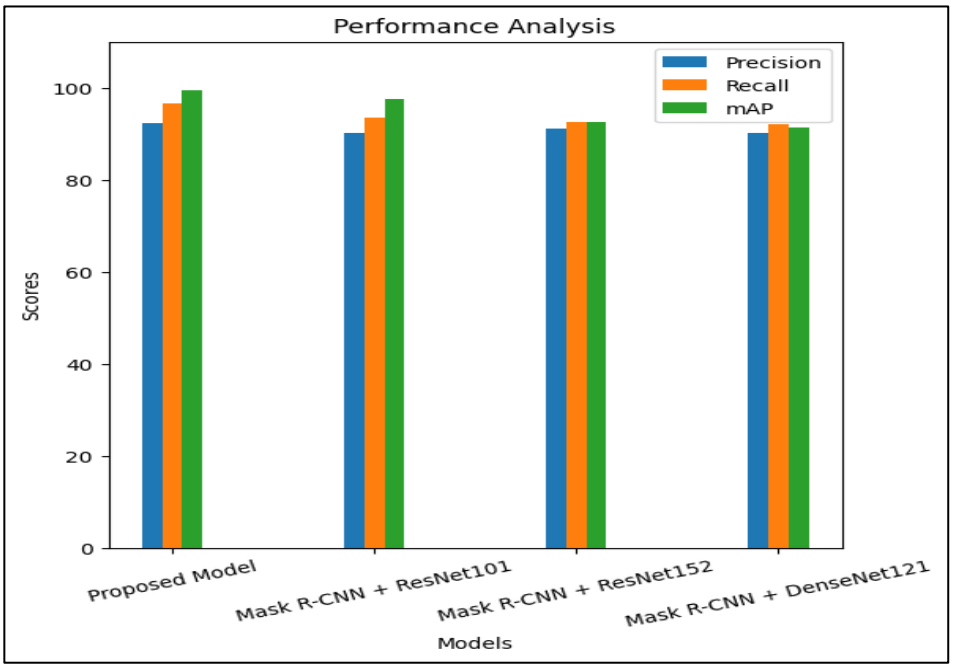

Figure 6. Comparison of the Proposed Model with the Other Models

The use of high-resolution images combined with Swin and FPN architectures improves the detection of small objects or lesions, reducing false negatives and making the approach well-suited for cancer detection. The study focuses mainly on the cancerous parts that could create a class imbalance, leading to misclassification. High-resolution images used in the research help preserve fine details with improved segmentation but increase the computational overhead. This also reduces the inference time, making it unsuitable in clinical or real-time settings. Low-resolution images reduce the computational load, making them suitable for mobile or edge deployment or low-resource settings, but at the cost of segmentation degradation. This problem can be alleviated by training the model using a varied range of image resolutions to guarantee equitable learning across different scales.

7. Conclusion

This study presents a highly efficient, accurate, and scalable method for the detection of oral cancer. The integration of Mask R-CNN with Swin Transformer provides a robust and precise approach for detecting oral cancer in medical imaging. This method utilizes the robust instance segmentation features of Mask RCNN alongside the efficient and context-sensitive feature extraction of the Swin Transformer, resulting in accurate detection and segmentation. It provides a high precision of 92.7 and a recall of 96.6, maximizing the true positives and true negatives. The mAP score of 99.5 signifies the accuracy of the detection capabilities provided by this model. There are obstacles related to computational expenses and data necessities when dealing with large medical images. False positives and false negatives are reduced with proper identification of the objects. The results indicate that this model holds significant potential for future clinical applications, thereby facilitating the development of enhanced diagnostic tools in oncology. Moreover, the computational complexity can be minimized by creating lightweight variants for real-time deployment. Additionally, multi modal data contributes to enhancing the model's effectiveness and versatility, thereby improving the diagnosis of oral cancer.

References

[1] Borse, Vivek, Aditya Narayan Konwar, and Pronamika Buragohain. "Oral cancer diagnosis and perspectives in India." Sensors International 1 (2020): 100046.

[2] Thiruvengadam, Rekha, and Jin Hee Kim. "Therapeutic strategy for oncovirus-mediated oral cancer: A comprehensive review." Biomedicine \& Pharmacotherapy 165 (2023): 115035.

[3] Siegel, Rebecca L., Angela N. Giaquinto, and Ahmedin Jemal. "Cancer statistics, 2024." CA: a cancer journal for clinicians 74, no. 1 (2024): 12-49.

[4] Kavyashree, C., H. S. Vimala, and J. Shreyas. "A systematic review of artificial intelligence techniques for oral cancer detection." Healthcare Analytics 5 (2024): 100304.

[5] Hmidani, Oussama, and EM Ismaili Alaoui. "A comprehensive survey of the R-CNN family for object detection." In 2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet), IEEE, 2022, 1-6.

[6] Jiang, Du, Gongfa Li, Chong Tan, Li Huang, Ying Sun, and Jianyi Kong. "Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model." Future Generation Computer Systems 123 (2021): 94-104.

[7] Ren, Junsong, and Yi Wang. "Overview of object detection algorithms using convolutional neural networks." Journal of Computer and Communications 10, no. 1 (2022): 115-132.

[8] Zhao, X., Xu, T., Peng, L., Li, S., Zhao, Y., Liu, H., He, J. and Liang, S., 2023. Recognition and segmentation of teeth and mandibular nerve canals in panoramic dental x-rays by mask RCNN. Displays, 78, 102447.

[9] Guo, Yanbin, Jing Guo, Yong Li, Peng Zhang, Yuan-Di Zhao, Yundi Qiao, Benyuan Liu, and Guoping Wang. "Rapid detection of non-normal teeth on dental X-ray images using improved Mask R-CNN with attention mechanism." International Journal of Computer Assisted Radiology and Surgery 19, no. 4 (2024): 779-790.

[10] Brahmi, Walid, and Imen Jdey. "Automatic tooth instance segmentation and identification from panoramic X-Ray images using deep CNN." Multimedia Tools and Applications 83, no. 18 (2024): 55565-55585.

[11] Fatima, Anum, Imran Shafi, Hammad Afzal, Khawar Mahmood, Isabel de la Torre Díez, Vivian Lipari, Julien Brito Ballester, and Imran Ashraf. "Deep learning-based multiclass instance segmentation for dental lesion detection." In Healthcare, vol. 11, no. 3, MDPI, 2023, 347.

[12] Zhang, Yang, Yan-Lin Liu, Ke Nie, Jiejie Zhou, Zhongwei Chen, Jeon-Hor Chen, Xiao Wang et al. "Deep learning-based automatic diagnosis of breast cancer on MRI using mask R-CNN for detection followed by ResNet50 for classification." Academic radiology 30 (2023): S161-S171.

[13] Yuan, Yuquan, Shaodong Hou, Xing Wu, Yuteng Wang, Yiceng Sun, Zeyu Yang, Supeng Yin, and Fan Zhang. "Application of deep-learning to the automatic segmentation and classification of lateral lymph nodes on ultrasound images of papillary thyroid carcinoma." Asian Journal of Surgery 47, no. 9 (2024): 3892-3898.

[14] Soltani, Hama, Mohamed Amroune, Issam Bendib, Mohamed-Yassine Haouam, Elhadj Benkhelifa, and Muhammad Moazam Fraz. "Breast lesions segmentation and classification in a two-stage process based on Mask-RCNN and Transfer Learning." Multimedia Tools and Applications 83, no. 12 (2024): 35763-35780.

[15] Kang, Junegyu, Van Nhat Thang Le, Dae-Woo Lee, and Sungchan Kim. "Diagnosing oral and maxillofacial diseases using deep learning." Scientific Reports 14, no. 1 (2024): 2497.

[16] Shetty, Shishir, Auwalu Saleh Mubarak, Leena R David, Mhd Omar Al Jouhari, Wael Talaat, Natheer Al-Rawi, Sausan AlKawas, Sunaina Shetty, and Dilber Uzun Ozsahin. "The application of mask Region-Based convolutional neural networks in the detection of nasal septal deviation using cone beam computed tomography images: Proof-of-Concept study." JMIR Formative Research 8 (2024): e57335.

[17] Freitas, Nuno R., Pedro M. Vieira, Catarina Tinoco, Sara Anacleto, Jorge F. Oliveira, A. Ismael F. Vaz, M. Pilar Laguna, Estêvão Lima, and Carlos S. Lima. "Multiple mask and boundary scoring R-CNN with cGAN data augmentation for bladder tumor segmentation in WLC videos." Artificial Intelligence in Medicine 147 (2024): 102723.

[18] Dataset accessed from https://universe.roboflow.com/srinivas-xujci/oral-cancer-new-1wgcw/dataset/2, 1st November 2024.

[19] Dutta, Abhishek, and Andrew Zisserman. "The VIA annotation software for images, audio and video." In Proceedings of the 27th ACM international conference on multimedia, pp. 2276-2279. 2019.

[20] Dutta, Abhishek, Ankush Gupta, and Andrew Zissermann. VGG image annotator (VIA). 2016.

[21] Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin transformer: Hierarchical vision transformer using shifted windows." In Proceedings of the IEEE/CVF international conference on computer vision, 10012-10022. 2021.

[22] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In Proceedings of the IEEE international conference on computer vision, 2961-2969. 2017.

[23] Tian, Yingjie, Duo Su, Stanislao Lauria, and Xiaohui Liu. "Recent advances on loss functions in deep learning for computer vision." Neurocomputing 497 (2022): 129-158.

[24] Damaceno-Araujo, A. L., E. Crespo, M. Cardoso-Moraes, M. Ajudarte-Lopes, P. A. Vargas, L. P. Kowalski, and A. R. Santos-Silva. "UNet-driven image segmentation for improved salivary gland tumor detection." Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology 139, no. 1 (2025): e34-e35.

[25] Priya, J., S. Kanaga Suba Raja, and S. Sudha. "An intellectual caries segmentation and classification using modified optimization-assisted transformer denseUnet++ and ViT-based multiscale residual denseNet with GRU." Signal, Image and Video Processing 18, no. 6 (2024): 5213-5227.