

RESEARCH

Open Access



# Development of an oral cancer detection system through deep learning

Liangbo Li<sup>1,2†</sup>, Cheng Pu<sup>3,4†</sup>, Jingqiao Tao<sup>1,5†</sup>, Liang Zhu<sup>1,2</sup>, Suixin Hu<sup>2</sup>, Bo Qiao<sup>2</sup>, Lejun Xing<sup>2</sup>, Bo Wei<sup>2</sup>, Chuyan Shi<sup>2</sup>, Peng Chen<sup>2\*</sup>  and Haizhong Zhang<sup>2\*</sup> 

## Abstract

**Objective** We aimed to develop an AI-based model that uses a portable electronic oral endoscope to capture intraoral images of patients for the detection of oral cancer.

**Subjects and methods** From September 2019 to October 2023, 205 high-quality annotated images of oral cancer were collected using a portable oral electronic endoscope at the Chinese PLA General Hospital for this study. The U-Net and ResNet-34 deep learning models were employed for oral cancer detection. The performance of these models was evaluated using several metrics: Dice coefficient, Intersection over Union (IoU), Loss, Precision, Recall, and F1 Score.

**Results** During the algorithm model training phase, the Dice values were approximately 0.8, the Loss values were close to 0, and the IoU values were around 0.7. In the validation phase, the highest Dice values ranged between 0.4 and 0.5, while the Loss values increased, and the training loss began to decrease gradually. In the test phase, the model achieved a maximum Precision of 0.96 with a confidence threshold of 0.990. Additionally, with a confidence threshold of 0.010, the highest F1 score reached was 0.58.

**Conclusion** This study provides an initial demonstration of the potential of deep learning models in identifying oral cancer.

**Keywords** Artificial intelligence, Deep learning, Oral cancer

<sup>†</sup>Liangbo Li, Cheng Pu and Jingqiao Tao Contributed equally.

\*Correspondence:

Peng Chen

chenpeng@301hospital.com.cn

Haizhong Zhang

zhanghaizhong@301hospital.com.cn

<sup>1</sup>Medical School of Chinese PLA, Beijing, China

<sup>2</sup>Department of Stomatology, Chinese PLA General Hospital, 28 Fuxing road, Haidian District, Beijing 100853, China

<sup>3</sup>Key Laboratory of Animal Disease and Human Health of Sichuan Province, Beijing, China

<sup>4</sup>College of Veterinary Medicine, Sichuan Agricultural University, Sichuan, China

<sup>5</sup>Department of stomatology, Southern Medical Branch of PLA General Hospital, Beijing 100842, China

## Introduction

Oral cancer ranks 18th among 36 major cancers worldwide, with an estimated 377,731 new diagnoses and over 177,757 deaths in 2020 [1]. According to the latest epidemiological survey data released by the American Cancer Society in 2023, there are 54,540 new cases of oral cancer and 7,400 deaths annually in the United States [2]. The 5-year survival rate for U.S. patients diagnosed with localized disease is 83%, compared to only 32% for patients with metastatic cancer [3]. More than two-thirds of oral cancers are detected after distant metastasis, marking the highest incidence and mortality rates among major cancers [4]. Therefore, early detection is essential to ensure optimal outcomes.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Early detection of oral cancer is the most effective method to reduce incidence and improve patient survival rates [5]. However, the detection rates of traditional clinical screening methods under visible light are relatively low [6]. Research by Andrade et al. highlights the similarity between the signs and symptoms of benign and potentially malignant conditions and those of oral cancer [7]. Additionally, in resource-limited settings, the challenges of early screening are further exacerbated.

In recent years, artificial intelligence has been increasingly applied in the field of stomatology [8–10]. The implementation of artificial intelligence involves the integration of deep learning and machine learning, and several deep learning models have been successfully applied in the field of computer vision. Medical image processing in computer vision typically includes three components: image classification, object detection, and semantic segmentation. Numerous studies have demonstrated that using convolutional neural network models to detect caries in X-rays or through near-ray retransmission imaging can achieve high accuracy [11, 12]. Additionally, dental caries detection has been successfully conducted using mobile phones and cameras to capture images of the mouth [13]. Furthermore, Warin K et al. have utilized novel deep convolutional neural networks for the early detection of oral cancer by analyzing photographs taken with a camera [14, 15].

The advancement of artificial intelligence technology has introduced new opportunities in the field of stomatology. However, AI is often regarded as a “black box,” necessitating the development of specialized tools or equipment to make it practically useful. While mobile phones and cameras are potential options for capturing oral images, the limited space within the mouth makes it challenging to achieve comprehensive imaging with these devices, increasing the risk of missed detections. To address this limitation, we designed and developed a portable oral electronic endoscope. Its “toothbrush-like” appearance allows for easy maneuvering within the mouth, enabling clear and thorough imaging of each area, thus enhancing the effectiveness of oral cancer screening.

To our knowledge, this is the world’s first artificial intelligence model designed to detect oral cancer by capturing RGB images with a portable oral electronic endoscope. Our goal is to develop a tool that can assist in the early detection of oral lesions by enabling patients to use the oral electronic endoscope to capture images, which can then be reviewed by professional dentists for further evaluation and diagnosis. Additionally, we aim to explore the feasibility of using this technology as an auxiliary tool to help identify suspicious lesions, particularly in high-risk populations or resource-limited settings.

## Materials and methods

### Data source

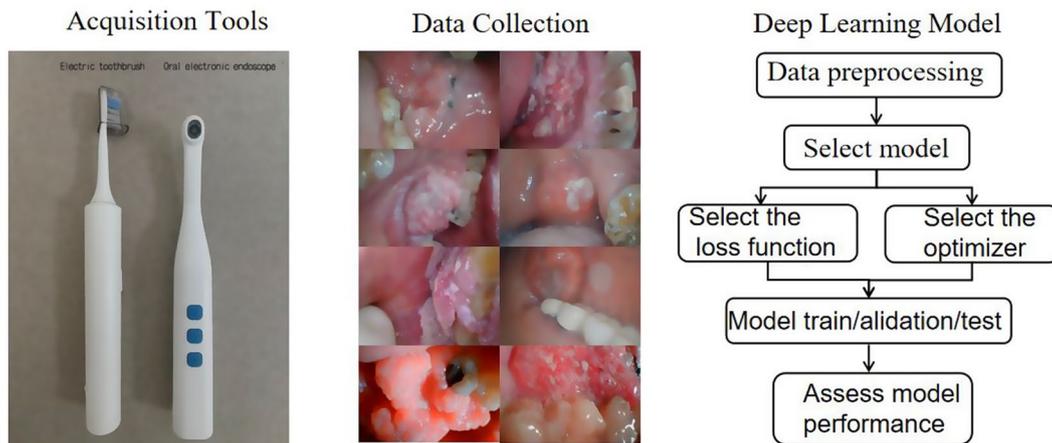
This study was approved by the Ethics Committee of the PLA General Hospital (Ethics Number: S201901602) and was conducted in accordance with the principles of the Declaration of Helsinki. The 205 clinical oral photographs analyzed in this study were collected at the Department of Stomatology, First Medical Center of the PLA General Hospital using a portable oral electronic endoscope. The collection period spanned from September 2019 to December 2023. The photographic images were captured at four different resolutions: 2190×1080 pixels, 1280×720 pixels, 320×240 pixels, and 320×180 pixels. All images were normalized to a uniform resolution of 256×256 pixels, and this modification was conducted using Python’s OpenCV library. We conducted a validation study to assess whether reducing image resolution introduced any bias. A representative subset of images was selected and processed at both their original high resolution and at 256×256 pixels using Python’s OpenCV library. The model was then tested on both datasets, and key performance metrics such as Dice coefficient, IoU, Precision, Recall, and F1 Score were compared. Statistical analysis showed no significant differences in the metrics, indicating that the reduction in resolution did not adversely affect model performance. The entire study process is illustrated in Fig. 1.

### Data acquisition and annotation process

All photographs were collected by students from the Department of Stomatology at the First Medical Center of the PLA General Hospital. The students were trained by a senior professional doctor, Dr. Haizhong Zhang, in the use of electronic oral endoscopy and the image acquisition process. The inclusion criteria for oral cancer are as follows: (1) Oral masses confirmed as malignant via pathology (2) Pathologically confirmed diagnoses of oral squamous cell carcinoma (T1–T4 stages).

All images were annotated using Labelme 4.5.6 software. Initial annotations were made by four junior doctors, Liangbo Li, Nenghao Jin, Liang Zhu, and Suixin Hu. These annotations were then reviewed by two intermediate-level doctors, Bo Qiao and Lejun Xing, who checked for any omissions, mislabeling, or boundary errors. In cases where there were disagreements between the initial annotations and the review, the final decision was made by a senior doctor, Dr. Haizhong Zhang. The annotated images were saved in JSON format.

To label the images, each oral cancer area in a tooth image is identified and marked separately. For each identified oral cancer area, a mask of size  $h * w$  is created. In this mask, pixels with a value of 1 represent the presence of oral cancer (white area), while pixels with a value of 0 represent non-cancerous areas (black area). If there are  $n$



**Fig. 1** Workflow for the development of an oral cancer detection system using deep learning. **Acquisition Tools:** The portable oral electronic endoscope (right) used in this study is shown alongside an electric toothbrush (left) for scale comparison. The endoscope is designed for easy, non-invasive imaging of the oral cavity. **Data Collection:** Representative images of oral lesions captured using the portable oral electronic endoscope. These images form the dataset used for training and validating the deep learning model in this study. **Deep Learning Model:** The diagram outlines the steps involved in the deep learning model's development process, including data preprocessing, model selection, choice of loss function and optimizer, model training/validation/testing, and performance assessment

distinct oral cancer areas in the image,  $n$  separate masks are generated, each corresponding to one cancerous area. These  $n$  masks are then combined to form a complete annotation with dimensions  $n * h * w$  (Fig. 2).

### Model selection

In this study, we utilized a deep learning model based on the U-Net architecture, which demonstrated excellent performance in image segmentation tasks by integrating a pre-trained ResNet-34 as the encoder. ResNet-34, known for its robust feature extraction capabilities, has achieved high accuracy in various computer vision tasks, as reported in several publications [16–18]. The ResNet-34 model used in this study was provided through the PyTorch deep learning framework, which offers pre-trained models that have been optimized on the ImageNet dataset. U-Net features a symmetric structure comprising two main components: the encoder and the decoder. The encoder captures the contextual information of the image, while the decoder precisely locates the target region of interest. Our network employs continuous convolutional layers and pooling layers to progressively reduce the spatial dimensions of the feature map, while upsampling and feature concatenation are used to restore image details, achieving high-precision segmentation.

As the core of the encoder, ResNet-34 is a deep residual network that introduces residual learning to address the degradation problem in deep networks. Its structure comprises multiple residual blocks, enabling the network to be effectively trained even at substantial depths. ResNet-34 has demonstrated powerful feature extraction

capabilities in various computer vision tasks, particularly on the ImageNet dataset.

To leverage the network parameters already trained on large datasets, we selected the ResNet-34 model with weights pre-trained on the ImageNet dataset. These pre-trained weights initialize the encoder of our model, significantly enhancing the convergence rate and improving the model's generalization ability.

### Input data preprocessing

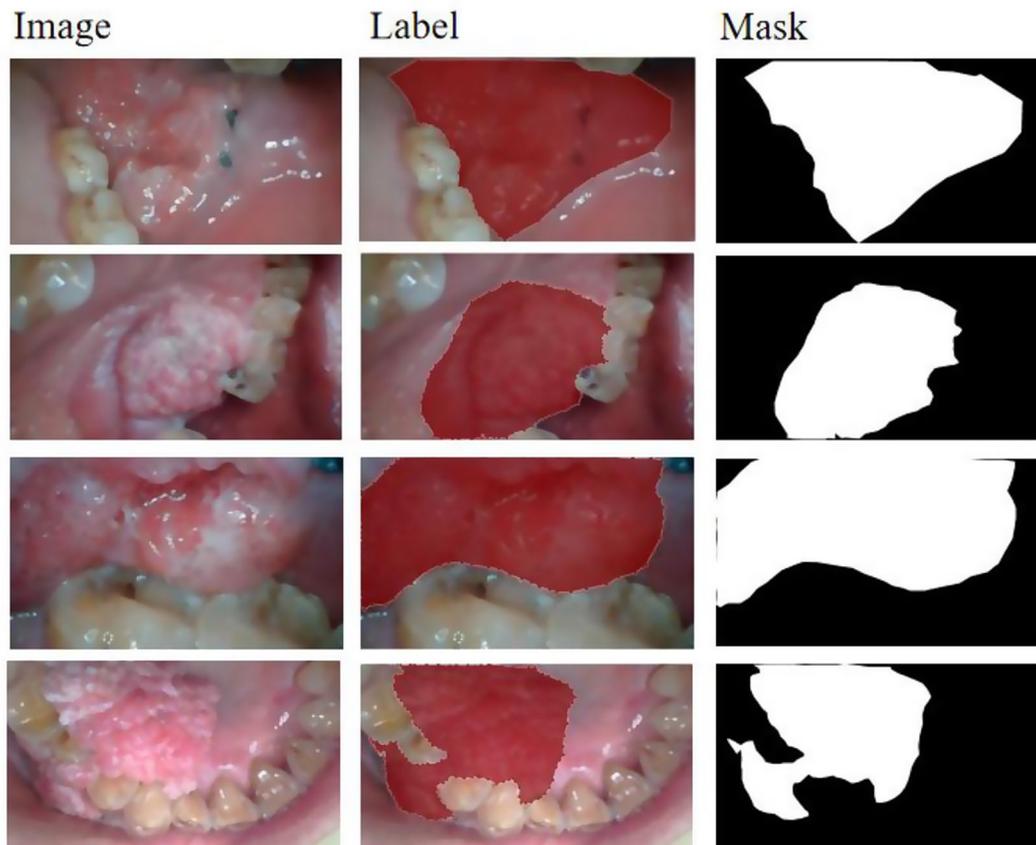
To ensure compatibility with the pre-trained ResNet-34 model, we applied a series of preprocessing steps to the image data.

#### Adjust the data size and pixels to fit the ResNet-34

We used the 'get\_preprocessing' function provided by the 'segmentation\_models' Python library to adjust the image size and normalize pixel values, ensuring compatibility with the pre-trained ResNet-34 model. The 'get\_preprocessing' function applies a series of transformations, including resizing the input images to the model's required dimensions and normalizing the pixel values based on the mean and standard deviation of the ImageNet dataset. This ensures that the input data is consistent with the data used during the ResNet-34 model's initial training.

### Data augmentation

To enhance the model's ability to handle deformation, rotation, and the identification of objects at different scales, we applied a range of data augmentation strategies to the training data. These strategies included: (1) random rotation of up to 1.90 degrees, (2) random



**Fig. 2** Example of oral cancer image labeling and mask generation. **Left Column (Image)**: Original intraoral images captured using the portable oral electronic endoscope. **Middle Column (Label)**: The corresponding labeled regions of interest (ROI) where oral cancer is present. These labels were manually annotated by experts to create ground truth data for training the deep learning model. **Right Column (Mask)**: Binary masks generated from the labeled images, where the white regions correspond to the annotated areas of oral cancer, and the black regions represent the background. These masks are used by the deep learning model to learn and predict cancerous regions in new images

translation up to 30% of the width and height, (3) shear transformation up to 50%, (4) random zoom up to 30%, and (5) random horizontal and vertical flips. These data augmentation techniques were implemented using the 'ImageDataGenerator' and utilized reflection mode for pixel filling.

#### **Binarization processing**

For the segmentation mask, we applied binarization post-processing after data augmentation to ensure that the mask retained its binary nature, with each pixel value set to either 0 (representing the background) or 1 (representing the foreground).

#### **Evaluation index**

The evaluation metrics in this study included the Dice coefficient, IoU, and Loss values. Both the Dice coefficient and IoU are statistics that measure the similarity between two sets, with IoU additionally accounting for the diversity between sample sets. Higher values for the Dice coefficient and IoU indicate better model

performance. The Loss value measures the gap between the predicted results and the actual outcomes; a smaller Loss value indicates that the model's predictions are closer to the actual results, reflecting better model performance. In addition, Precision, Recall, and F1 Score are also included. Precision is the correct proportion of all predicted samples; recall is the proportion of correctly identified positive samples relative to the total number of actual positive samples. The F1 Score is a metric used to evaluate the performance of a binary classification model by combining precision and recall. It is the harmonic mean of precision and recall, providing a balanced measure of accuracy. F1 scores range from 0 to 1, with values closer to 1 indicating better classifier performance and values closer to 0 indicating worse performance. The performance metrics of the model in oral cancer detection are detailed in Table 1.

#### **Model and endoscopic parameters**

The model is based on the Tesla V100 chip, with a single inference speed of 29.14 ms. The training iteration

**Table 1** Performance Metrics of the AI model in oral Cancer detection

Metric	Training Phase	Validation Phase	Testing Phase
Dice Coefficient	0.80	0.40–0.50	-
IoU	0.70	0.30	-
Loss	0.00	Increased	-
Precision	-	-	0.96
Recall	-	-	1.00
F1 Score	-	-	0.58

IoU:Intersection over Union

reached 10 epochs, the batch size was set to 8, and the maximum number of detections was 50. The optimizer used was SGD(Stochastic Gradient Descent), with an

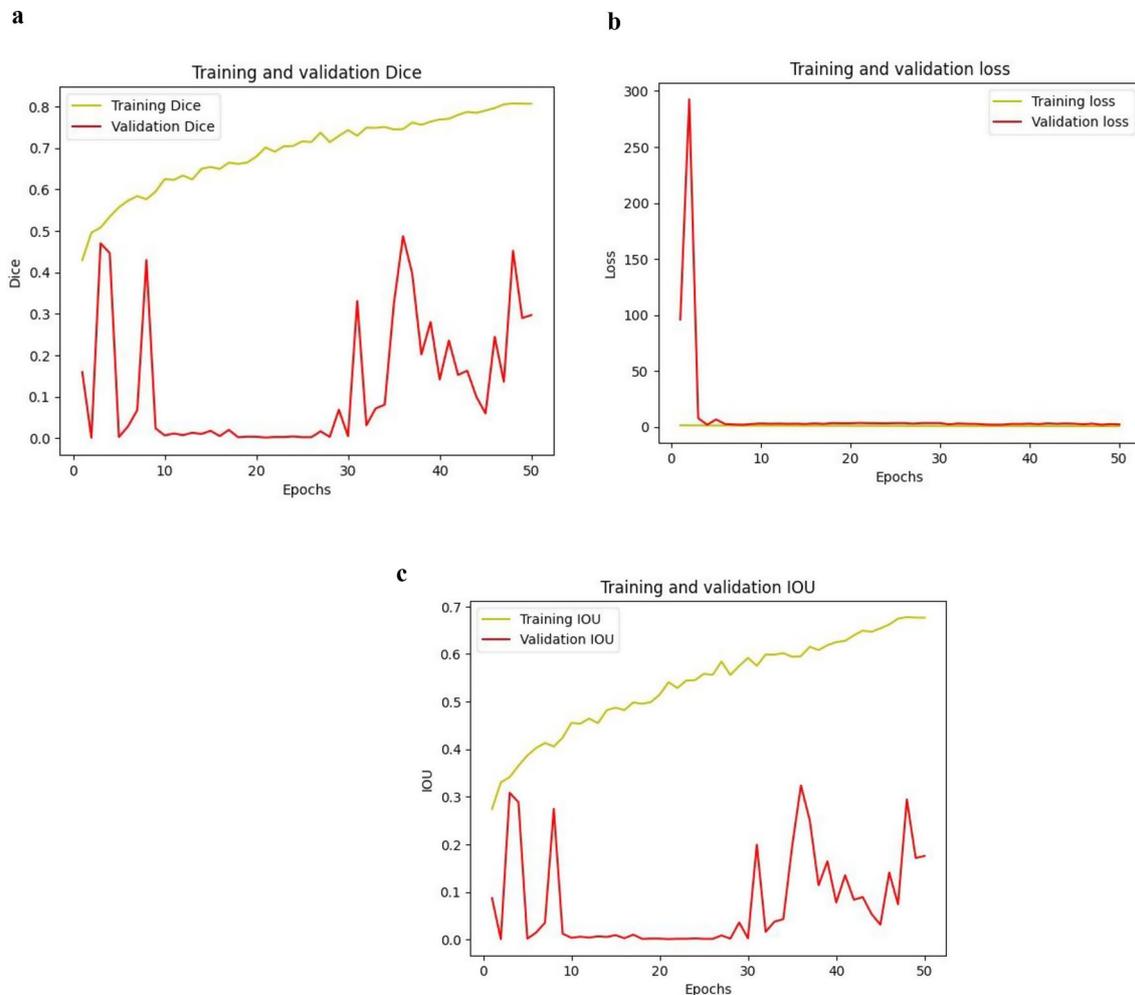
initial learning rate of 0.01 and a warm-up period of 3 epochs.

The lens of the portable electronic oral endoscope is a customized Zeiss lens, featuring five million pixels and surrounded by six white LED lights that can adjust the brightness across five levels. The focus range is from 10 mm to 50 mm.

**Results**

**Model training and validation**

In Fig. 3a, the abscissa “Epochs” represents the number of iterations during model training, and the ordinate “Dice” value represents the degree of overlap between the predicted results and the true labels—the closer the Dice value is to 1, the higher the similarity. The Dice



**Fig. 3** Training and validation metrics of the model. **3a:**The Dice coefficient, which assesses the similarity between predicted and actual segmentation, is plotted for both the training (yellow line) and validation (red line) datasets. The training Dice improves consistently, while the validation Dice is unstable and lower, reflecting challenges in achieving consistent performance on the validation set, potentially due to the small dataset size or overfitting.**3b:**This graph shows the loss values for both the training (yellow line) and validation (red line) datasets over 50 epochs. The training loss decreases rapidly and stabilizes near zero, indicating effective learning from the training data. However, the validation loss spikes initially and then stabilizes at a higher level.**3c:**This graph depicts the IoU values, a metric that measures the overlap between predicted and actual segmentation regions. The training IoU (yellow line) increases steadily, showing improvement in model accuracy during training. In contrast, the validation IoU (red line) fluctuates significantly and remains lower

coefficient is calculated using the formula:  $\text{Dice coefficient} = 2 * TP / (2 * TP + FP + FN)$ , where TP represents the number of true positive samples, FP represents the number of false positive samples, and FN represents the number of false negative samples. This formula quantifies the similarity between the predicted results and the true labels. The yellow line represents the “Training Dice.” As the number of training iterations increases, the Dice value on the training dataset gradually improves, indicating that the model’s ability to fit the training data is being enhanced. The red line represents the “Validation Dice,” which shows the Dice value changes on the validation dataset. Unlike the training dataset, the scores on the validation dataset remained relatively stable and did not show significant improvement with additional training. This pattern suggests that our model may be overfitting, becoming too sensitive to the training data and less predictive for unseen data (validation datasets). This issue could be related to the smaller size of our dataset.

In Fig. 3b, the ordinate “Loss” represents the loss value during each iteration. The yellow line indicates the training loss, showing how the loss varies on the training dataset. The red line represents the validation loss, illustrating how the model performs on the validation dataset. At the beginning of training, the validation loss initially shows an upward trend, likely because the model’s parameters have not yet been optimized, leading to a significant gap between the predicted results and the actual outcomes. As the number of iterations increases, the training loss gradually decreases, indicating that the model’s performance on the training dataset is improving. Simultaneously, the validation loss remains relatively low and exhibits minimal fluctuation, suggesting that the model demonstrates good generalization ability on the validation dataset.

In Fig. 3c, the ordinate “IoU” (Intersection over Union) represents the overlap between the predicted and the actual bounding boxes. The closer the IoU value is to 1, the higher the overlap between the predicted and real boxes. The yellow line shows the IoU changes during the training phase, and around 50 epochs, the model finds a good optimization path, causing the IoU values to peak between 0.6 and 0.7. The red line represents the IoU changes during the validation stage. As the number of epochs increases, the IoU value hovers around 0.3, indicating that the model’s generalization ability on the validation dataset is weaker than on the training dataset. This discrepancy may be related to the relatively small size of our dataset.

### Model testing

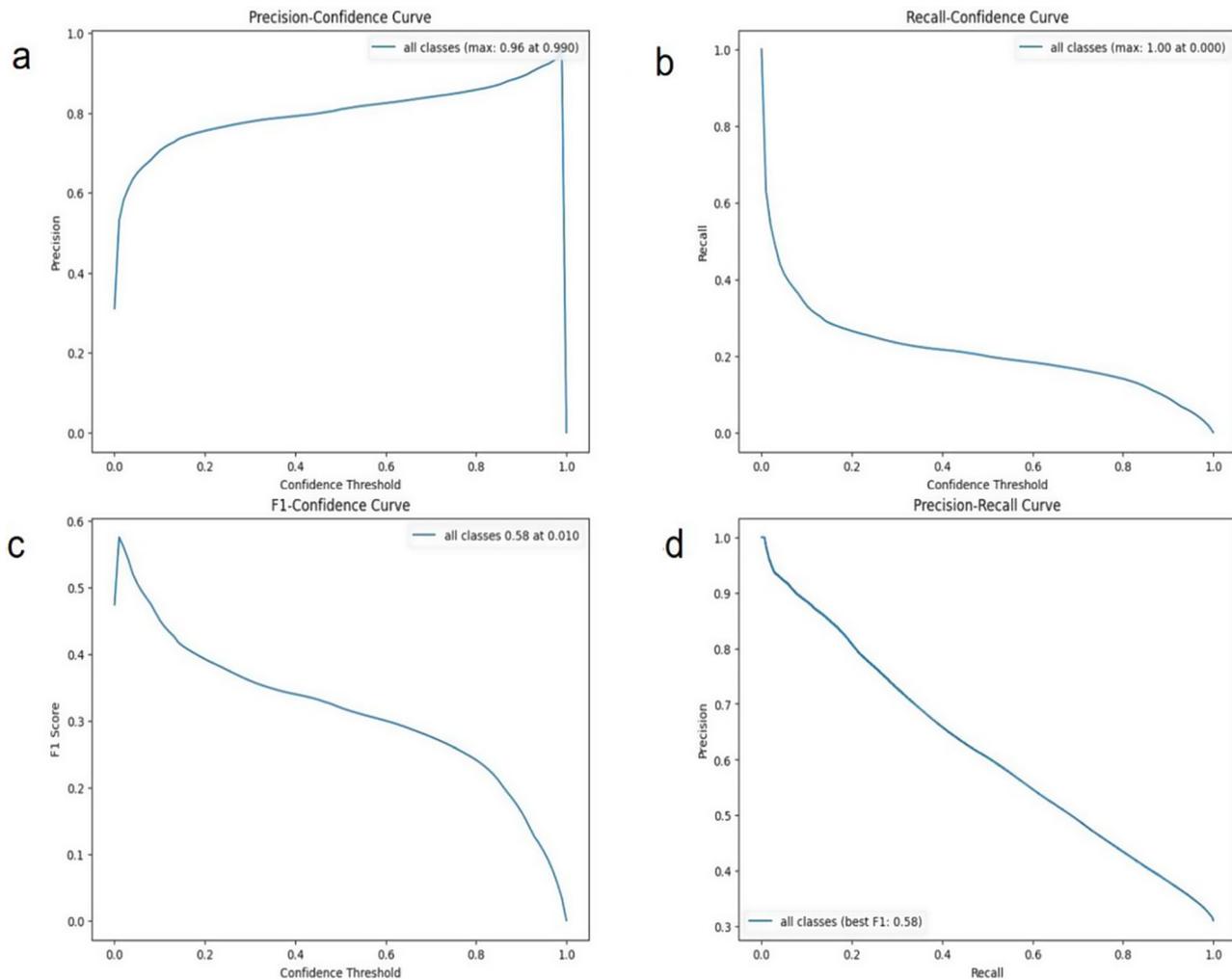
In Fig. 4a, Precision represents the proportion of samples that the model correctly predicts as positive for oral cancer. The confidence threshold is the criterion used to

determine whether a prediction should be considered a positive result. When the model’s prediction confidence exceeds this threshold, the sample is classified as positive. Precision measures the accuracy of the model in identifying true positive cases of oral cancer. A high precision rate indicates that the model is accurate in identifying true cases of oral cancer, while a low precision rate suggests a higher likelihood of misclassifying healthy oral mucosa as cancerous. At a confidence threshold of 0.990, the model achieves a maximum precision of 0.96, meaning that when the model is highly confident in its predictions (close to 1), its accuracy in correctly identifying positive samples is also very high, reaching 0.96.

In Fig. 4b, recall refers to the proportion of all positive samples that the model correctly identifies as positive. An ideal recall value is 1, indicating that the model successfully identifies all true positive samples without missing any. A higher recall rate reflects the model’s improved ability to detect positive cases. As confidence decreases, the recall rate reaches 1.00 at a confidence threshold of 0.000 (the lowest confidence), suggesting that the model can identify nearly all positive samples at very low confidence levels. However, this may also result in a higher number of false positives. As an early screening model for oral cancer, our goal is to minimize the risk of missed diagnoses by improving the recall rate.

In Fig. 4c, the F1 Score is a metric used to evaluate the performance of a binary classification model by considering both precision and recall. It is the harmonic mean of precision and recall, providing a balanced measure of accuracy. F1 scores range from 0 to 1, with values closer to 1 indicating better classifier performance, and values closer to 0 indicating worse performance. The curve starts at the top left corner, with the F1 score initially rising and then decreasing as the confidence threshold increases. At a particular threshold value (0.58), the F1 score drops sharply, indicating poor model performance at this threshold. This may occur because the model becomes too conservative when the threshold is set too high, leading to many samples that should be predicted as positive being incorrectly classified as negative. This reduction in recall consequently lowers the F1 score.

In Fig. 4d, precision consistently declines as the recall rate increases. When selecting the optimal threshold, the goal is to achieve high values for both precision and recall. However, due to the inherent trade-off between these two metrics, it is impossible to maximize both simultaneously. The F1 score, being the harmonic mean of precision and recall, provides a balanced assessment of model performance. The model’s performance is considered optimal when the F1 score reaches its peak value of 0.58.



**Fig. 4** Performance metrics of the model in oral cancer detection. **4a:** This curve shows the relationship between the confidence threshold and the precision of the model. The model achieved a maximum precision of 0.96 at a confidence threshold of 0.990, indicating that at this threshold, 96% of the positive predictions were correct. **4b:** This curve illustrates the relationship between the confidence threshold and the recall of the model. The recall reaches 1.00 at a confidence threshold of 0.000, meaning the model correctly identified all positive samples at this threshold, but with a potential increase in false positives. **4c:** This curve displays the F1 score, which is the harmonic mean of precision and recall, across different confidence thresholds. The highest F1 score of 0.58 was achieved at a confidence threshold of 0.010, indicating the optimal balance between precision and recall at this threshold. **4d:** This curve shows the trade-off between precision and recall. As recall increases, precision decreases, highlighting the challenge of maintaining both high precision and high recall simultaneously. The model's best F1 score (0.58) represents the optimal point on this curve

### Model predictions

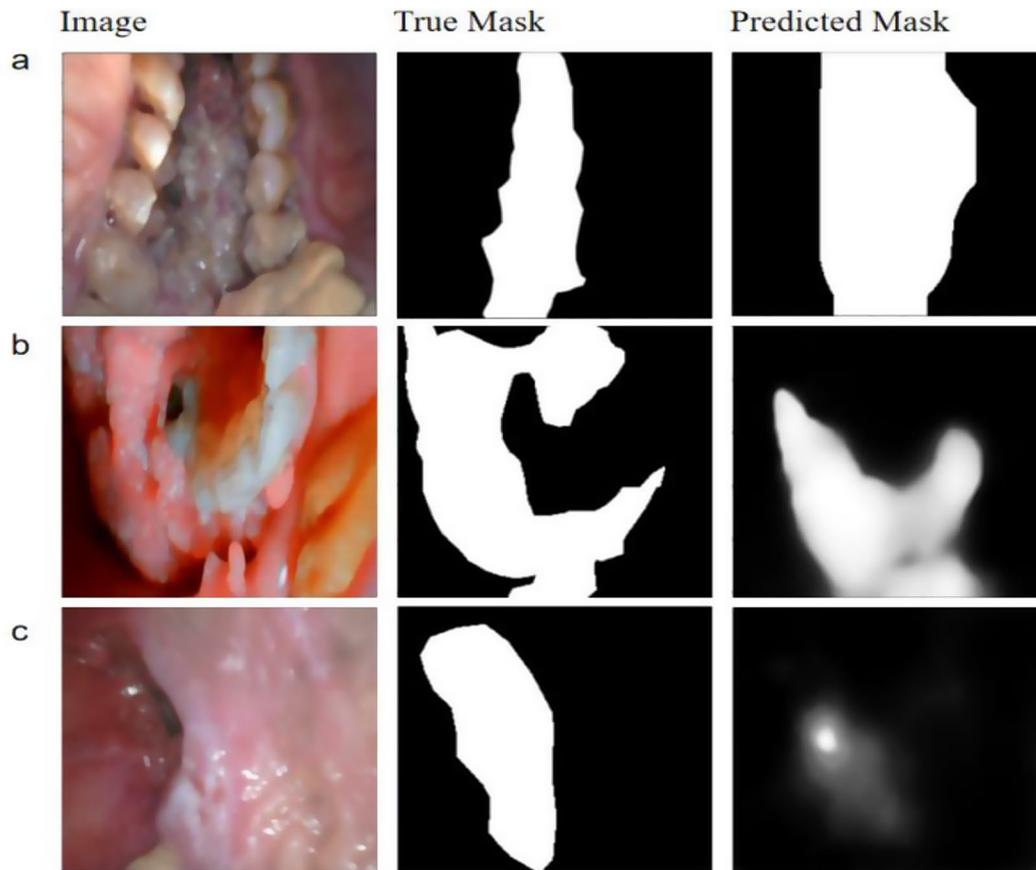
Figure 5 shows the results of the model predictions.

### Discussion

Early screening of oral cancer is of great significance. In a 15-year randomized trial involving 87,655 participants in India, mortality in the screened population decreased by 81%, and the incidence of oral cancer was reduced by 38% [19]. A study conducted in Taiwan on oral cancer screening showed a 26% reduction in mortality in the screening group and a reduction of 57.5 per 100,000 in the non-screening group [20]. These studies demonstrate that screening asymptomatic individuals through visual examination is feasible and that early screening

enables the early detection of oral cancer, reducing both morbidity and mortality [21]. However, visual inspection for oral cancer may be more applicable to specialist dentists. Some studies have shown that general practitioners (GPs) often lack knowledge and awareness in diagnosing oral cancer, particularly in recognizing its early symptoms, which is a significant factor in delaying treatment [22]. The 5-year survival rate for oral cancer is 83% when detected early, compared to just 32% for patients with late-stage diagnosis. Therefore, early screening and diagnosis are crucial in reducing the incidence of oral cancer and improving patient survival rates.

In recent years, artificial intelligence has made significant strides in the medical field, with its applications in



**Fig. 5** Comparison of true and predicted masks in oral cancer detection. **Left Column (Image):** Original intraoral images showing various oral lesions. **Middle Column (True Mask):** Manually annotated masks representing the actual areas of oral cancer, used as ground truth for training the AI model. The white areas indicate the regions identified as cancerous, while the black areas represent the background. **Right Column (Predicted Mask):** Masks predicted by the AI model. The white areas indicate the regions predicted as cancerous by the model. Blurriness in the predicted masks may result from the model's interpolation process during segmentation, which attempts to smooth the boundaries to reduce noise but may also result in less precise edges. **a:** The predicted mask shows some differences compared to the true mask. **b:** The predicted mask shows some differences compared to the true mask, which may be due to complex or unclear boundaries within the lesion. **c:** Example of a more significant mismatch between the predicted and true masks, highlighting areas where the model's predictions need further refinement

deep learning and machine learning revolutionizing medical diagnosis and treatment. Utilizing deep learning algorithms, AI can automatically interpret and analyze vast amounts of medical imaging data, accurately identifying diseased areas and enhancing the accuracy and efficiency of diagnosis [23–25]. By training on large datasets of imaging data, AI models can detect small lesions that might be overlooked by doctors, thereby reducing the risk of missed diagnoses and misdiagnoses [26]. Several deep learning models have been successfully applied in the field of computer vision, which involves three key tasks: image classification, object detection, and semantic segmentation. Advances in computer vision and AI technologies have significantly improved visual detection capabilities, offering new tools to assist in visual diagnosis and clinical data interpretation, particularly for oral cancer screening systems [14].

Our results show that during the algorithm model training phase, the Dice coefficient was approximately

0.8, the Loss was near 0, and the IoU was around 0.7. In the model validation stage, the highest Dice value ranged between 0.4 and 0.5, which may indicate overfitting of our model. The Loss value increased while the training loss gradually decreased. The IoU value was around 0.3, suggesting that the model is less generalized on the validation dataset compared to the training dataset, which may be related to the limited size of our data. During the test phase, the model achieved a maximum Precision of 0.96 at a confidence threshold of 0.990. At the lowest confidence threshold of 0.000, the recall rate reached 1.00, indicating that the model could identify almost all positive samples, but this might also include a large number of false positives. As an early screening model for oral cancer, our aim is to minimize the risk of missed diagnoses by increasing the recall rate. At a confidence threshold of 0.010, the highest F1 score reached was 0.58.

This study offers an initial demonstration of the potential for applying artificial intelligence (AI) and deep

learning in the diagnosis of oral cancer, but several limitations must be acknowledged. Firstly, the AI system developed in this study is still in its early stages and requires further enhancement, particularly in its ability to identify occult lesions, determine optimal biopsy sites, and accurately define surgical margins. Additionally, the study primarily focused on more advanced and easily identifiable lesions (T1-T4 stages), which limits the generalizability of the findings to early-stage, occult lesions. Future research should aim to expand the dataset, improve the model's effectiveness in detecting early lesions, and validate the system's performance by comparing it with other diagnostic methods, such as biopsy and clinical examination. Moreover, this study did not address complex differential diagnoses, such as distinguishing between severe hairy tongue and verrucous carcinoma, or between erosive lichen planus and early squamous cell carcinoma. These complex diagnostic challenges are crucial in clinical practice, and we plan to refine the AI system to better handle such cases in future iterations. Additionally, the current technology is not designed to effectively define surgical margins or detect hidden lesions, both of which are critical challenges in oral cancer management. We will clearly state these limitations in the discussion. Despite these limitations, we believe that the results of this study, though preliminary, provide an important direction for exploring the use of AI in the early screening of oral cancer. With further development and refinement, this system has the potential to become a more reliable clinical tool, ultimately contributing to improved early diagnosis and treatment outcomes for patients with oral cancer.

In comparing our work with previous studies, such as Warin et al. [14], who utilized CNN-based models like DenseNet-169 and Faster R-CNN for OSCC and OPMD detection, we recognize key differences and similarities. Warin et al. achieved high AUCs (1.00 for OSCC and 0.98 for OPMDs), demonstrating the effectiveness of their models on larger datasets. In contrast, our study, using a different dataset and model architecture, achieved a maximum Precision of 0.96 during the test phase. These differences underscore the variability in model performance depending on the dataset size, type of architecture, and the specific challenges posed by the dataset used. Furthermore, our approach using U-Net combined with ResNet-34 is particularly advantageous for tasks requiring detailed segmentation and classification within smaller datasets, as often encountered in clinical settings. This approach, along with our specific preprocessing techniques and focus on achieving high precision in identifying oral lesions, is crucial for reducing false positives in clinical practice. This focus on precision, combined with the versatility of the U-Net architecture in

handling complex image features, distinguishes our study from others.

Despite these limitations, we believe that the results of this study, though preliminary, provide an important direction for exploring the use of AI in the early screening of oral cancer. With further development and refinement, this system has the potential to become a more reliable clinical tool, ultimately contributing to improved early diagnosis and treatment outcomes for patients with oral cancer.

## Conclusion

This study provides an initial demonstration of the potential of deep learning models in identifying oral cancer. By increasing the dataset and incorporating more types of diseases (such as oral potentially malignant disorders), it is expected to become a useful tool to assist clinicians in diagnosing oral cancer.

## Acknowledgements

We are thankful to the staff and faculty of the Department of Stomatology, Chinese PLA General Hospital.

## Author contributions

Liangbo Li, contributions to design, data acquisition, data annotation, analysis, and interpretation, drafted and critically revised the manuscript; Cheng Pu, Jingqiao Tao, contributions to data analysis and interpretation, critically revised the manuscript; Nenghao Jin, Liang Zhu, Suixin Hu, contributions to, data acquisition, data annotation, critically revised the manuscript; Bo Qiao, Lejun Xing, Bo Wei, Chuyan Shi, contributions to data acquisition; Peng Chen, Haizhong Zhang contributions to conception and design, analysis, and interpretation, critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

## Funding

None.

## Data availability

Relevant data can be obtained by contacting the corresponding author.

## Declarations

### Ethics approval and consent to participate

The study received ethical approval from the Research Ethics Committee of the Chinese PLA General Hospital (S2019-016-02). Informed consent was obtained from all patients. All methods were conducted in accordance with the relevant guidelines and regulations.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 13 June 2024 / Accepted: 12 November 2024

Published online: 04 December 2024

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209–49.

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73:17–48.
3. Llewellyn CD, Johnson NW, Warnakulasuriya KA. Risk factors for squamous cell carcinoma of the oral cavity in young people—a comprehensive literature review. *Oral Oncol.* 2001;37:401–18.
4. Ilhan B, Lin K, Guneri P, Wilder-Smith P. Improving oral Cancer outcomes with imaging and Artificial Intelligence. *J Dent Res.* 2020;99:241–8.
5. Andrade SA, Pratavieira S, Ribeiro MM, Bagnato VS, de Pilla Varotti F. Oral cancer from the perspective of wide-field optical fluorescence: diagnosis, tumor evolution and post-treatment follow up. *Photodiagnosis Photodyn Ther.* 2017;19:239–42.
6. Andrade SA, Pratavieira S, Bagnato VS. Use of tissue natural fluorescence in clinical practice as a complementary tool for a differential diagnosis of multiple oral pathologies. *Photodiagnosis Photodyn Ther.* 2015;12:358.
7. Andrade SA, Ribeiro MM, Pratavieira S, Bagnato VS, Varotti F, De P. Hairy Tongue: Differential diagnosis by Use of Widefield Optical Fluorescence. *Braz Dent J.* 2019;30:191–6.
8. Elmakaty I, Elmarasi M, Amarah A, Abdo R, Malki MI. Accuracy of artificial intelligence-assisted detection of oral squamous cell carcinoma: a systematic review and meta-analysis. *Crit Rev Oncol Hematol.* 2022;178:103777.
9. Revilla-León M, Gómez-Polo M, Vyas S, Barmak BA, Galluci GO, Att W, et al. Artificial intelligence applications in implant dentistry: a systematic review. *J Prosthet Dent.* 2023;129:293–300.
10. Schwendicke F, Cejudo Grano de Oro J, Garcia Cantu A, Meyer-Lueckel H, Chaurasia A, Krois J. Artificial Intelligence for Caries Detection: value of data and information. *J Dent Res.* 2022;101:1350–6.
11. Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, et al. Caries Detection with Near-Infrared Transillumination using deep learning. *J Dent Res.* 2019;98:1227–33.
12. Li S, Liu J, Zhou Z, Zhou Z, Wu X, Li Y, et al. Artificial intelligence for caries and periapical periodontitis detection. *J Dent.* 2022;122:104107.
13. Kühnisch J, Meyer O, Heseniuss M, Hickel R, Gruhn V. Caries Detection on Intraoral images using Artificial Intelligence. *J Dent Res.* 2022;101:158–65.
14. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P, Vicharueang S. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. *PLoS ONE.* 2022;17:e0273508.
15. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J Oral Pathol Med.* 2021;50:911–8.
16. Freitas NR, Vieira PM, Cordeiro A, Tinoco C, Morais N, Torres J, et al. Detection of bladder cancer with feature fusion, transfer learning and CapsNets. *Artif Intell Med.* 2022;126:102275.
17. Quinsten AS, Umutlu L, Forsting M, Nassenstein K, Demircioğlu A. Determining the anatomical site in knee radiographs using deep learning. *Sci Rep.* 2022;12:3995.
18. Makhoulouf Y, Singh VK, Craig S, McArdle A, French D, Loughrey MB, et al. True-T - improving T-cell response quantification with holistic artificial intelligence based prediction in immunohistochemistry images. *Comput Struct Biotechnol J.* 2024;23:174–85.
19. Sankaranarayanan R, Ramadas K, Thara S, Muwonge R, Thomas G, Anju G, et al. Long term effect of visual screening on oral cancer incidence and mortality in a randomized trial in Kerala, India. *Oral Oncol.* 2013;49:314–21.
20. Chuang S-L, Su WW-Y, Chen SL-S, Yen AM-F, Wang C-P, Fann JC-Y, et al. Population-based screening program for reducing oral cancer mortality in 2,334,299 Taiwanese cigarette smokers and/or betel quid chewers. *Cancer.* 2017;123:1597–609.
21. Warnakulasuriya S, Kerr AR. Oral Cancer screening: past, Present, and Future. *J Dent Res.* 2021;100:1313–20.
22. Crossman T, Warburton F, Richards MA, Smith H, Ramirez A, Forbes LJJ. Role of general practice in the diagnosis of oral cancer. *Br J Oral Maxillofac Surg.* 2016;54:208–12.
23. Zhou J, Hu B, Feng W, Zhang Z, Fu X, Shao H, et al. An ensemble deep learning model for risk stratification of invasive lung adenocarcinoma using thin-slice CT. *NPJ Digit Med.* 2023;6:119.
24. Do HM, Spear LG, Nikpanah M, Mirmomen SM, Machado LB, Toscano AP, et al. Augmented Radiologist Workflow improves Report Value and saves time: a potential model for implementation of Artificial Intelligence. *Acad Radiol.* 2020;27:96–105.
25. Hassan C, Spadaccini M, Mori Y, Foroutan F, Facciorusso A, Gkolfakis P, et al. Real-time computer-aided detection of colorectal Neoplasia during Colonoscopy: a systematic review and Meta-analysis. *Ann Intern Med.* 2023;176:1209–20.
26. Wu L, He X, Liu M, Xie H, An P, Zhang J, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. *Endoscopy.* 2021;53:1199–207.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.