



OPEN Screening of oral potentially malignant disorders and oral cancer using deep learning models

Karishma Madhusudan Desai^{1,2}, Pragya Singh²✉, Mahima Smriti², Vivek Talwar³, Manav Chaudhary², George Paul², Subhas Chandra Kolli², Parisa Sai Raghava², Golla Vamshi Krishna², C. V. Jawahar³, P. K. Vinod^{4,5}, Varma Konala^{2,6} & Ramanathan Sethuraman⁶

Oral cancer though preventable, shows high mortality and affect the overall quality of life when detected in late stages. Screening techniques that enable early diagnosis are the need of the hour. The present work aims to evaluate the effectiveness of AI screening tools in the diagnosis of OPMDs and Oral cancers via native or web-application (cloud) using smart phone devices. We trained and tested two deep learning models namely DenseNet201 and FixCaps using 518 images of the oral cavity. While DenseNet201 is a pre-trained model, we modified the FixCaps model from capsule network and trained it ground up. Standardized protocols were used to annotate and classify the lesions (suspicious vs. non-suspicious). In terms of model performance, DenseNet201 achieved an F1 score of 87.50% and AUC of 0.97; while FixCaps exhibited F1 score of 82.8% and AUC of 0.93. DenseNet201 model (20 M) serves as a robust screening model (accuracy 88.6%) that can be hosted on a web-application in the cloud servers; while the adapted FixCaps model with its low parameter size of 0.83 M exhibits comparable accuracy (83.8%) allowing easy transitioning into a native phone-based screening application.

Keywords Screening, Oral cancer, Oral potentially malignant disorders, Deep learning, Web application

Oral and lip cancers, though preventable, are associated with 1,77,757 deaths annually^{1,2}. The low- and middle-income countries have witnessed snowballing incidence of oral cancers and their precursors lesions 'oral potentially malignant disorders (OPMDs)'. Screening via clinical oral examination (COE) is considered as the most effective method for early detection of OPMDs and oral cancer¹. Additionally, as visual screening is cost-effective, researchers and governing bodies globally have laid emphasis to promote and implement conventional oral screening programs^{3–5}. Though early detection of OPMDs and early-stage cancers can reduce patient mortality⁶, there are several roadblocks such as heterogeneity in examination techniques, diverse clinical presentations and lack of health literacy. Apart from the known grave health consequences and morbidity, treatment of advanced carcinomas adds enormous economic and psychological burden on the affected individual⁷.

In several developing countries, besides screening and diagnosis, the access to healthcare services remains a pertinent challenge. This is further aggravated by urban–rural divide in low- and middle-income countries (LMIC) like India, wherein two-thirds of the population reside in rural areas which have limited accessibility and affordability to health services⁸. Recently, community health workers in India such as Accredited social health activists (ASHA) have helped to address this challenge. ASHA's have been trained to perform screening of the oral cavity and capture images of oral cavity using smartphones. These images when uploaded to AI applications can guide the health workers to refer high-risk patients to specialists for timely management^{9–11}. AI models have shown value in early detection of OPMDs and oral cancers^{11–13}.

Advanced computer vision algorithms based on deep learning help recognize, process and present desirable output based on the quality of images, annotations and data reconstruction techniques. Clinicians globally have put together one such effort via the Mobile Mouth Screening Anywhere (MeMoSA) project and deep learning based convolutional neural networks (CNN)¹⁴. They employed the box annotation approach with residual

¹Tokyo Dental College, 2 Chome-9-18 Misakicho, Chiyoda City, Tokyo 101-0061, Japan. ²INAI, International Institute of Information Technology, Hyderabad 500032, India. ³CVIT, International Institute of Information Technology, Hyderabad 500032, India. ⁴iHUB-Data, International Institute of Information Technology, Hyderabad 500032, India. ⁵CCNSB, International Institute of Information Technology, Hyderabad 500032, India. ⁶Intel Technology India Private Limited, Bengaluru 560103, India. ✉email: chief.pragya009@gmail.com

networks (ResNet-101) for classification of images (lesion Vs no lesion) and Faster region-based CNN (Faster R-CNN) for object detection (referral based on risk), exhibiting an F1 score of 87.07% for images with lesions. Similarly, researchers have used web and mobile applications for classification of images into normal, OPMD and oral cancer^{10–12,15}. Emerging studies are showing promising results with oral cancer screening using deep learning approaches.

Despite the promising prospects, two major hindrances^{3,16} in the global adoption and regularization of AI include—(a) the lack of understanding pertaining to functioning of deep learning algorithms in accurately distinguishing noise from data and (b) model fit for broad high and low-resource constrained device. To facilitate remote screening via mobile app in LMICs, an ideal solution would be to deploy a low-parameter size model that practically functions efficiently on high as well as resource-constrained devices. Our previous work¹¹ explored and analyzed performance of CNN architectures such as VGG19, Inception ResNet-V2, MobileNet-V2, DenseNet-121, DenseNet-169 and DenseNet-201 for screening of suspicious oral lesions; identified DenseNet-201 as the best performing model with moderate levels of compute complexity. In the present work, we aimed to evaluate the effectiveness of AI screening tools (moderate vs low compute complexity) in the diagnosis of OPMDs and Oral cancers via native or web-application using a smart phone device.

Methods

Ethical clearance and application flow

Following Institutional ethical clearance (SBS-2020-2) and after obtaining informed consent, all methods were performed in accordance with relevant guidelines; oral images were captured from the patients using smartphone devices. These images were used for training and testing the deep learning (DL) models explained in this work. The DL model-based disease screening approach presented in this work uses a smartphone device to run a native-application or a web-application. The captured smartphone image(s) are input to the DL models that provides a binary classification output (suspicious and non-suspicious). This helps to deliver OPMD and cancer screening even in locations with limited or no healthcare services. Illustrated in Fig. 1a is the end-to-end (E2E) solution for deploying such DL models for disease screening purposes.

Datasets

A total of 518 oral cavity images from Grace dataset¹¹ classified into 229 suspicious images and 289 non-suspicious images were included. Out of these, 46 suspicious and 59 non-suspicious images were used for the independent test set and the remaining were included for training. The data consisted of oral cavity images from both male (52%) and female (48%) patients of South Indian ethnicity. All patients between the age of 18 to 75 years, who reported to the outpatient department and Hyderabad district camp sites included under the screening effort by the Grace Cancer Foundation. A standardized protocol was used to annotate and classify the oral cavity images into suspicious and non-suspicious. Amongst the suspicious images, OPMDs included were classified based on Warnakulasuriya et al.¹⁷ definition and criteria for OPMDs. The common classes (Fig. 1b) encountered consisted of leukoplakia homogeneous (LKH) and leukoplakia nonhomogeneous (LKN), oral submucous fibrosis (OSF) and oral cancer (OCC). The images that presented with physiologic appearance of oral structures, benign growths, or with other oral conditions were included in the non-suspicious (NS) category. The data sets were screened and annotated by a general dentist with 8 years of experience, an oral pathologist with 10 years of pathology experience and an oncology specialist with 18 years of oncology experience. In case of discrepancies, a consensus was reached and considered. The Grace intraoral dataset was captured using smartphones (with 3-to-5-megapixel resolution camera sensors) by untrained frontline healthcare workers (FHWs) and the oral cavity images were used for training the models without any image enhancements. This process allowed model training that simulates a deployment scenario, wherein the captured oral cavity images without any image enhancements would be subject to the DL model inference.

Annotations

Though the model output is binary (suspicious and non-suspicious), the input data consisted of common subtypes of OPMDs prevalent in the Indian population. The region of interest in each image was annotated in the training set using CVAT¹⁸. A polygon tool was chosen for marking the lesion-level ROI in each image. The ROI was extended to include the lesion proper and a margin of surrounding tissue. The following parameters were considered while annotating the lesions—lesion color (red/white/mixed), surrounding tissue response (inflamed/noninflamed), topography (affected site and extent), and texture of the lesion (flat/verruciform/ulcerated/proliferative). For cases showing generalized involvement of the oral cavity, such as oral submucous fibrosis, a larger site demonstrating key parameters (blanching, pallor, and co-occurrence of other red–white lesions) was considered. This additional lesion-level characterization was done to enable objective outcome analysis for the DL models.

DL model architecture

From the previous DL model architecture exploration by Talwar et al.¹¹ of CNN and Vision Transformer (ViT) styles, DenseNet201 turned out to be the best performing model with lower model parameter size. Besides DenseNet201 model which is more suitable for cloud-based deployments (accessible through web-applications using smartphone and more suited for urban areas that have internet connectivity), we have explored DL architecture options in this work for standalone deployments (wherein smartphone-based native android applications that do not rely on internet connectivity and more suited for rural areas that have minimal or no internet connectivity). One such DL architecture used in the present work is a modification of the Capsule network¹⁹.

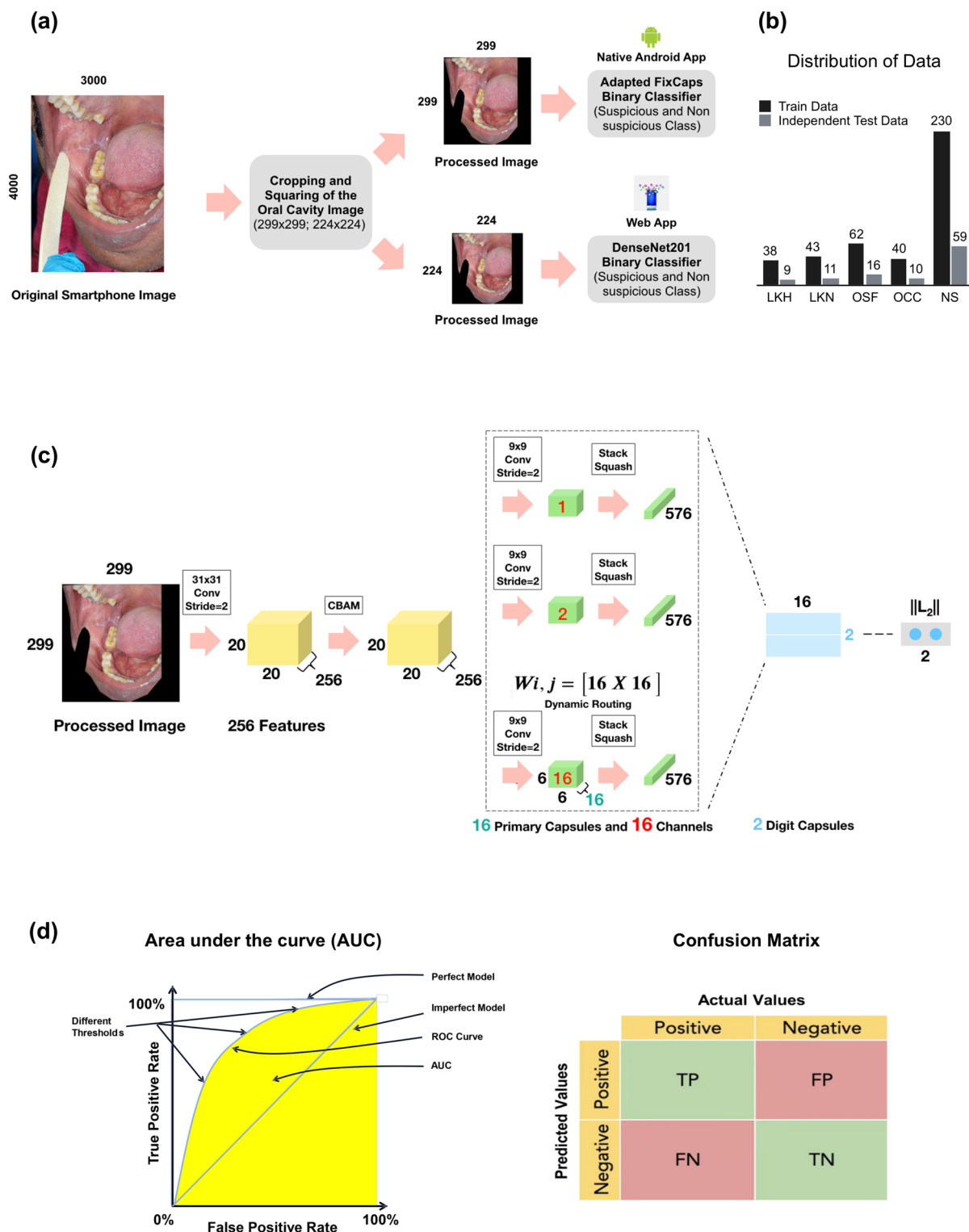


Fig. 1. (a) End-2-End (E2E) solution for OPMD screening. The inference pipeline uses smartphone image(s) as input and delivers binary classification as output. There are two pathways for OPMD screening. One enabled by the DL model running on the smartphone device as a native android app. The second enabled by the DL model running on the cloud accessed through the web application on a smartphone device. (b) Graph showing distribution of training and test data sets. (c) Adapted FixCaps Network19 architecture used for OPMD screening. (d) Schematic representation of area under the curve and confusion matrix.

A capsule represents the probability of existence of an image part defined by its dimensions, color, texture apart from its position and orientation. A capsule network made of multiple layers of capsules, aim to build the knowledge of the spatial relationships between parts of a whole object in an image. This is done using dynamic routing to create stronger connections between spatially related object parts while training the multi-layer capsule network. A variant of the Capsule network, called FixCaps²⁰, was applied for diagnosis of skin cancer achieving reasonable screening quality. FixCaps distinguishes itself from the Capsule network through its larger receptive field (due to the larger convolution kernel of 31×31) apart from the convolutional block attention module (CBAM) that was introduced to mitigate loss of spatial information due to convolution and pooling.

In this work on building DL model architecture suited for OPMD screening using mobile phone images of the oral cavity, we use 256 features and 16 primary capsules (Fig. 1c) instead of 128 features and 8 primary capsules as used in FixCaps for dermatoscopy images. Since the OPMD screening using mobile phone images are to be used by the FHWs, the DL model classifies the images into two classes (namely, suspicious and non-suspicious). The adapted FixCaps DL model used in this work is 0.83 million (M) parameters (see Table 1), which makes it a suitable option as a standalone solution for deployment in smartphones.

Model training and independent testing

DenseNet201 was a pretrained model from ImageNet dataset and was fine-tuned for Oral cavity images. The oral cavity images of up to 5-megapixel resolution from the smartphone camera were first cropped for the mouth region and the resulting cropped image is resized to 224×224. The retraining process uses resized input images with standard data augmentation (center cropping, horizontal flipping, vertical flipping, colour jitter and random brightness contrast transformations), Adam optimizer, learning rate of 10⁻⁴, retraining epochs of 100, batch size of 32 and cross-entropy loss function.

Adapted FixCaps model was trained from scratch. The input images of up to 5-megapixel resolution from the smartphone camera were first cropped for the mouth region and the resulting cropped image is resized to 299×299. The resized image is subject to ImageNet normalization scheme²¹ and image augmentation (horizontal flipping, vertical flipping and normalize transformations; 10.9X augmentation of 413 train images) before feeding as an input to the adapted FixCaps network²⁰ for training. The training settings include Adam optimizer with a learning rate of 0.001, 350–500 epochs of training, StepLR scheduler with step size of 3 and gamma of 0.97. All the experiments in this paper were implemented in PyTorch (version 2.0) on a NVidia GeForce GTX Titan X system.

Both the DenseNet201 and adapted FixCaps models have been trained using the Grace train dataset (413 images) with ratio split of 80:20 (train: test) and 64:16:20 (train: validate: test) respectively. Subsequently, the quality of the resulting models was assessed using the Grace independent dataset (105 images). The training data (64%) and validation data (16%) were merged and randomly partitioned into five-folds, maintaining class balance.

Quality assessment metrics

The standard quality assessment metrics are used, namely, Accuracy, Precision, Recall (sensitivity or TPR), FPR, Specificity, F1-score, CI, AUC measure under ROC curve and Confusion matrix. All results presented in this work will use these quality metrics to assess the performance of the deep learning model for the specific binary classification application of OPMD screening. Primary metrics that are used for quality assessment are true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Rest of the quality metrics are defined based on these primary metrics.

Accuracy = (TP + TN) / T; T = TP + FP + TN + FN (total number of samples) (1)

Precision = TP / (TP + FP) (2)

Recall (or Sensitivity) = True Positive Rate (TPR) = TP / (TP + FN) (3)

False Positive Rate (FPR) = FP / (FP + FN) (4)

Specificity = TN / (TN + FP) (5)

DL models	Parameters (M: Million)
Adapted FixCaps (This work)	0.83
MobileNet ¹²	4.2
DenseNet201 ¹¹	20
Swin (base) ¹¹	88
VGG + BDL ¹²	143.7

Table 1. Model parameter size of DL models used in OPMD screening applications.

$$F1 - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (6)$$

$$Confidence\ Interval\ (95\% \ CI) = 1.96 \times \sqrt{\frac{Accuracy \times (1 - Accuracy)}{T}} \quad (7)$$

Similar CIs can be computed for all other quality metrics stated above. AUC is an area measure under the receiver operator characteristic (ROC) curve, which is a plot of TPR versus FPR. An example ROC curve is shown in Fig. 1d where the AUC lies between 0.5 and 1. This implies, the classifier can distinguish the positive class from the negative ones. Figure 1d also shows a representative confusion matrix that highlights the performance of the classifier in terms of predicted values to actual values.

Results

For evaluating the relative performance, pretrained DenseNet201¹¹ and the adapted FixCaps network were tested for screening of OPMDs and oral cancers using 105 oral cavity images (Fig. 2a,b). The DenseNet201 model with a parameter size of 20 M presented with 88.6% accuracy, 84% precision, 91.3% sensitivity (recall), 86.4% specificity, F1 score of 87.5% and AUC of 0.97. Whereas, the adapted FixCaps model with a parameter size of 0.83M showed reliable recognition of lesions and exhibited accuracy of 83.8%, specificity of 79.7% and F1 score of 82.8%. The other key parameters of precision, recall and area under the curve were 77.4%, 89.1% and 0.93, respectively. A five-fold cross-validation was executed for different train and validation splits in order to study the generalizability of the models as illustrated in Supplementary Fig. 1 (S1). The performance metrics across folds is reported on the Grace independent test set with a 95% confidence interval (CI) in Supplementary Fig. 2 (S2).

The test set presented with 12 FP and 5 FN in the adapted FixCaps model, while there were 8 FPs and 4 FNs with the DenseNet201 model. A detailed clinical analysis of these FPs and FNs revealed that the adapted FixCaps model FNs included 2 cases of OSF and LKN each, and 1 case of LKH; while the DenseNet201 showed failures with 1 case of OSF, 2 cases of LKH and 1 case of LKN. Specific characterization of lesions, helped in identification of overlapping failures between the models (Fig. 2ci, ii). OSF cases with generalized involvement of mucosa (Fig. 2ciii) failed and were classified as nonsuspicious. Amongst FPs, most images showing vascular sites such as labial mucosa (Fig. 2civ) and ventrolateral tongue were misdiagnosed. Clinically, FPs included non-suspicious images of linea alba buccalis with overlapping traumatic lesion (Fig. 2ci), non-specific redness (inflammation) of faucial pillars and periodontal disease.

Discussion

In healthcare settings, AI-assisted clinical and histological diagnosis can save time and person effort. AI models have shown promising results in screening of skin cancers²², ocular²³ and cardiovascular diseases²⁴. For oral cancers and OPMDs, it is recommended that early detection and prompt referral to specialists (within two weeks) can help curb the malignant progression³. Considering the reported malignant transformation rate of commonly encountered OPMDs such as leukoplakia (7.9%)²⁵ and OSMF (6%)²⁶, there is a need to employ active efforts in promoting screening and quick referrals via existing and new diagnostic modalities. Screening techniques ranging from basic clinical examination using white light/fluorescence, vital staining, cytology to several others have been developed to meet this objective²⁷. Recently, smart phone-based screening of oral lesions by AI models have been used for field screening programs, enabling quick referrals to tertiary care centers^{10,12}. Existing CNNs namely Densely connected convolution networks (DenseNet), ResNet, SqueezeNet and Swin Transformers (Swin) are widely being explored for classification of images^{28,29}. Combination of CNN models with object detection algorithms (Faster R-CNN, YOLOv5, RetinaNet and CenterNet2) presented with better performance compared to general practitioners²⁹. Developing functional, easy-to-use solutions that can enable quick diagnosis and referrals by general practitioners, allied healthcare personnel and accredited health activists is the need of the hour. In the present work, we explain one such solution using our AI model for the screening of OPMDs and oral cancers.

Given the earlier exploration of DL model architectures of CNN and Vision Transformer (ViT) styles¹¹ that demand more data for training the model apart from being computationally expensive, in this paper we focus on an architecture that needs less amount of training data while being computationally inexpensive for deployment in mobile phones. Our DenseNet201¹¹ demonstrated a relatively higher accuracy of 88.6% compared to the adapted FixCaps model (83.8%). However, the adapted FixCaps DL model used in this work is 0.83 M parameter and is much smaller compared to any of the other DL models used in OPMD screening applications (Table 1). Amongst models using phone-based images for screening of OPMDs, Birur et al.¹² reported similar screening quality while using a relatively higher parameter size model (VGG + BDL with 143.7 M; MobileNet with 4.2 M). Though their data sets were heterogenous, we understand that using a compliant smaller parameter model such as adapted FixCaps could facilitate better transition to smartphone-enabled oral screening.

Our models exhibited a few limitations, clinical appearance of anatomical/physiological variations superseded by trauma-induced ulcer or keratosis were misinterpreted as suspicious. A possible solution for which could include addition of supporting history and training for identification of trauma-induced lesions. Additionally, as we used an unbalanced data set, the site distribution discrepancy led to failures with identification of non-suspicious images of non-keratinized and vascular mucosal sites such as lower labial mucosa, floor of mouth and ventrolateral surface of tongue. In contrast, Rabinovici-Cohen et al.³⁰ who also used smart phone-based images for detection of oral cancer, stratified the data based on lesion location and obtained high accuracy for diagnosis

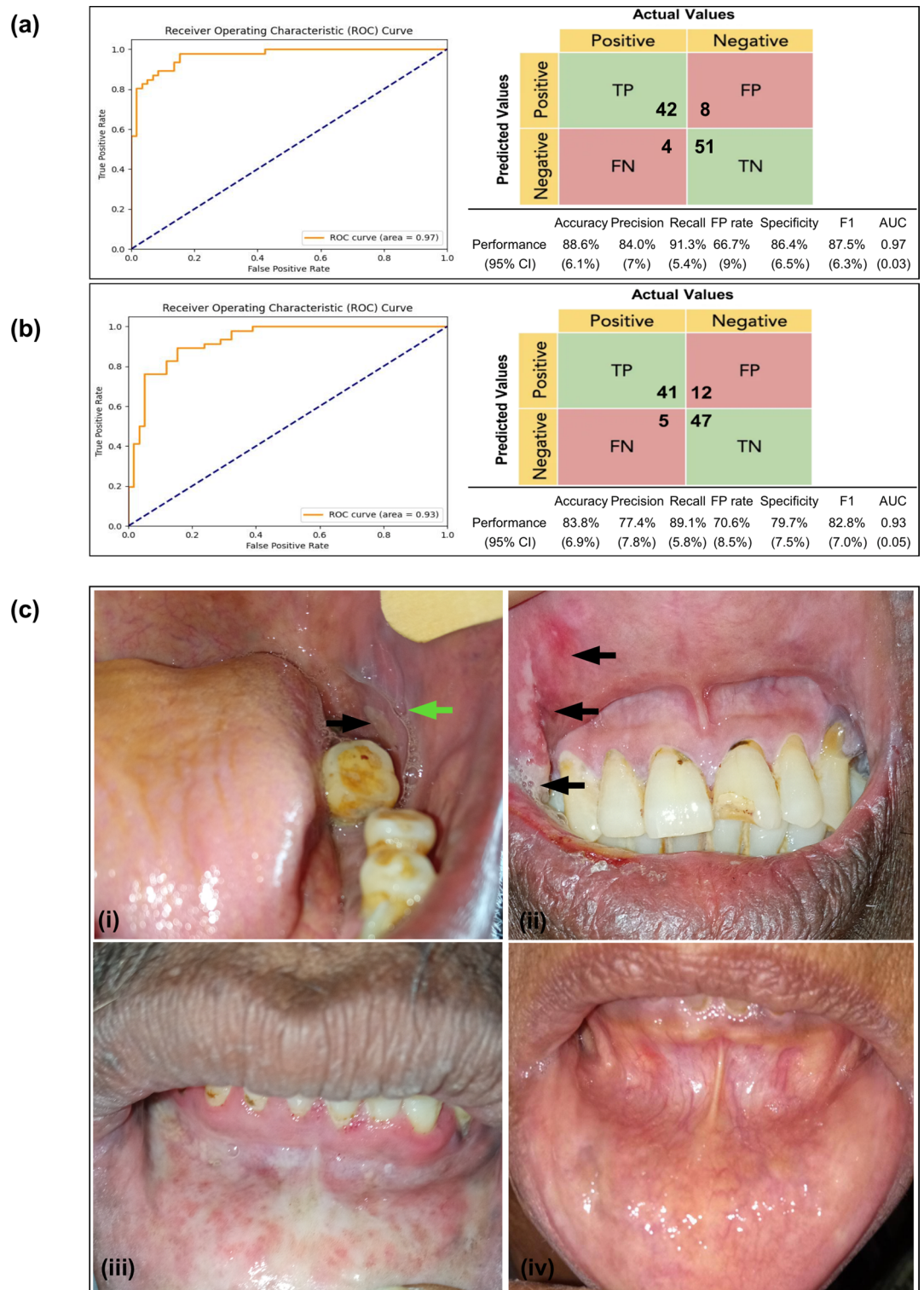


Fig. 2. (a) Performance of DenseNet201 showing area under the curve and confusion matrix. (b) Performance of Adapted FixCaps model showing area under the curve and confusion matrix. (c) False positive and false negative failures. (i) Linea alba (green arrow) with traumatic ulcer (black arrow) diagnosed as suspicious by both models. (ii) Red–white lesion (indicated by black arrows) diagnosed as non-suspicious by both models. (iii) False negative case of oral sub mucous fibrosis showing generalized involvement of the lower labial mucosa. (iv) False positive vascular site of lower labial mucosa.

of lesions involving lingual mucosa, floor of mouth and posterior tongue. This stratification approach combined with ROI-based annotations can direct future DL model architectures towards better clinical outcomes.

While FP errors can be resolved by training and addition of mouth segmentation strategy, a defined approach is needed to bypass the FNs. In our work, we encountered challenges with OSF cases, especially when the key presenting feature was generalized blanching of the oral mucosa. In a clinical setting, these features are often evaluated by the clinician using visual inspection coupled with manual palpation for the region of interest. Evaluating the degree of underlying submucosal fibrosis and making an informed decision regarding the same, presents as a major limitation with the use of AI for screening of (early) OSF. Identification of clinical changes in color, texture and contour of affected mucosa may allow AI models to overcome this pertinent limitation. Shah P et al.³¹ developed an algorithm for automatic detection of oral cancer using Red–Green–Blue (RGB) channels for images and a MATLAB image processing. Guided via the threshold value and histogram for red channel, their model could distinguish oral images into normal (no lesion) and abnormal (leukoplakia Vs erythroplakia). They further designed a two-step approach using a YCbCr (segmented Y and Cb–Cr planes) processing path with gray-level co-occurrence matrix (GLCM) for evaluating details of homogeneity, entropy, pixel contrast etc. Similarly, Goswami et al.³² extended this application and used five color formats with gradient boosting (light gradient boosting) machine learning approach for evaluating features like lesion color and texture. They found remarkable results with an accuracy of 99.25%, precision 99.18% and F1 score of 99.24%. These approaches reinforce that incorporation of image segmentation and color formats can aid in recognition of critical lesion patterns. Additionally, cases such as OSF, the diagnosis can be better obtained by addition of areca nut consumption history and clinical information such as burning sensation, interincisal distance, uvula changes and tongue protrusion changes. Model training to differentiate normal uvula shape from the shrunken and deformed “hockey-stick” uvula, interpretations of interincisal and tongue protrusion measurements can help further to classify the stage of OSF. As compared to some of the existing high-end imaging technologies, deep learning-based applications using smart phone images may serve as a cost-effective option for the screening of oral lesions.

Conclusion

DenseNet201 is a robust model which on progressive training can facilitate improved screening of common OPMDs and oral cancer, especially while using a cloud-supported application. Conversely, the adapted FixCaps model with its low model parameter can serve as an effective tool for transition to an easy-to-use native smart phone-based screening application. In the future, to increase the clinical applicability and to increase the model specificity, there is a need to expand the dataset and include other OPMDs or lesions that present with similar clinical features such as lichen planus, actinic keratosis, amongst several others. We also plan to delineate the clinical relevance and applications of color formats, mouth segmentation and reconstruction techniques for the screening of OPMDs, oral cancers and other general pathologies.

Data availability statement

The datasets generated during and/or analyzed during the current study are not publicly available due to specific institutional requirements governing privacy protection; however, they are available from the authors on reasonable request within the terms of the data use agreement and compliance with ethical and legal requirements. Models for the detection of OPMDs are provided in the Github repository: https://github.com/mahimasmriti/OPMD_240625/

Received: 16 August 2024; Accepted: 15 May 2025

Published online: 23 May 2025

References

- Louredo, B. V. R. et al. Reported physical examination methods for screening of oral cancer and oral potentially malignant disorders: A systematic review. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **137**, 136–152 (2024).
- Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Vigarios, E., Warnakulasuriya, S., Barres, B. H. & Maret, D. Difficulty diagnosing oral cancer: seeking an early specialist opinion is key. *Lancet* **402**, 2018 (2023).
- Dissanayaka, D. W. V. N. et al. A preliminary study on early detection of oral cancer with opportunistic screening: Insights from dental surgeons in Sri Lanka. *Cancers (Basel)* **15**, 5511 (2023).
- Subramanian, S. et al. Cost-effectiveness of oral cancer screening: results from a cluster randomized controlled trial in India. *Bull World Health Organ.* **87**, 200–206 (2009).
- Rajaraman, P. et al. Recommendations for screening and early detection of common cancers in India. *Lancet Oncol.* **16**, e352–e361 (2015).
- Singh, A. G. et al. A prospective study to determine the cost of illness for oral cancer in India. *Ecancermedicalscience* **15**, 1252 (2021).
- Rural population (% of total population) – India. *The World Bank*. <https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=IN> (2018).
- Sankaranarayanan, R. et al. Long term effect of visual screening on oral cancer incidence and mortality in a randomized trial in Kerala, India. *Oral Oncol.* **49**, 314–321 (2013).
- Birur, N. P. et al. Role of community health worker in a mobile health program for early detection of oral cancer. *Indian J. Cancer* **56**, 107–113 (2019).
- Talwar, V. et al. AI-assisted screening of oral potentially malignant disorders using smartphone-based photographic images. *Cancers (Basel)* **15**, 4120 (2023).
- Birur, N. P. et al. Field validation of deep learning-based Point-of-Care device for early detection of oral malignant and potentially malignant disorders. *Sci. Rep.* **12**, 14283 (2022).

13. Camalan, S. et al. Convolutional neural network-based clinical predictors of oral dysplasia: Class activation map analysis of deep learning results. *Cancers (Basel)* **13**, 1291 (2021).
14. Welikala, R. A. et al. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access* **8**, 132677–132693 (2020).
15. Song, B. et al. Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning. *Biomed. Opt. Express* **9**, 5318–5329 (2018).
16. Sangers, T. E. et al. Position statement of the EADV artificial intelligence (AI) task force on AI-assisted smartphone apps and web-based services for skin disease. *J. Eur. Acad. Dermatol. Venereol.* **38**, 22–30 (2024).
17. Warnakulasuriya, S. et al. Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO collaborating centre for oral cancer. *Oral Dis.* **27**, 1862–1880 (2021).
18. Computer Vision Annotation Tool. <https://www.cvat.ai/> (2024).
19. Sabour, S., Nicholas, F. & Geoffrey, E. H. Dynamic routing between capsules. *Adv. Neural Inform. Process. Syst.* <https://doi.org/10.48550/arXiv.1710.09829> (2017).
20. Lan, Z., Cai, S., He, X. & Wen, X. FixCaps: An improved capsules network for diagnosis of skin cancer. *IEEE Access* **10**, 76261–76267 (2022).
21. Timothy. PyTorch. <https://discuss.pytorch.org/t/discussion-why-normalise-according-to-imagenet-mean-and-std-dev-for-transfer-learning/115670> (2021).
22. Sangers, T. E., Wakkee, M., Moolenburgh, F. J., Nijsten, T. & Lugtenberg, M. Towards successful implementation of artificial intelligence in skin cancer care: A qualitative study exploring the views of dermatologists and general practitioners. *Arch. Dermatol. Res.* **315**, 1187–1195 (2023).
23. El-Hag, N. A. et al. Classification of retinal images based on convolutional neural network. *Microsc. Res. Tech.* **84**, 394–414 (2021).
24. Samant, S. et al. Artificial intelligence, computational simulations, and extended reality in cardiovascular interventions. *JACC Cardiovasc. Interv.* **16**, 2479–2497 (2023).
25. Zhang, C. et al. The global prevalence of oral leukoplakia: A systematic review and meta-analysis from 1996 to 2022. *BMC Oral Health* **23**, 645 (2023).
26. Murthy, V. et al. Transformation rate of oral submucous fibrosis: A systematic review and meta-analysis. *J Clin Med.* **11**, 1793 (2022).
27. Lau, J. et al. Adjunctive aids for the detection of oral squamous cell carcinoma and oral potentially malignant disorders: A systematic review of systematic reviews. *Jpn. Dent. Sci. Rev.* **60**, 53–72 (2024).
28. Fu, Q. et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine* **27**, 100558 (2020).
29. Warin, K. et al. AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. *PLoS ONE* **17**, e0273508 (2022).
30. Rabinovici-Cohen, S. et al. From pixels to diagnosis: Algorithmic analysis of clinical oral photos for early detection of oral squamous cell carcinoma. *Cancers (Basel)* **16**, 1019 (2024).
31. Shah, P., Roy, N. & Dhandhukia, P. Algorithm mediated early detection of oral cancer from image analysis. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **133**, 70–79 (2022).
32. Goswami, B., Bhuyan, M. K., Sultan, A. & Mejdi, S. Classification of oral cancer into pre-cancerous stages from white light images using LightGBM algorithm. *IEEE Access* **12**, 31626–31639 (2024).

Acknowledgements

The authors thank Naren Akash R. J. for technical assistance with re-run of DenseNet201. The authors thank Dr. Chinnababu Sunkavalli and Grace Cancer Foundation for their support in data collection efforts.

Author contributions

KMD, PS, MS, RS—Manuscript preparation, conceptualization, interpretation and formal data analysis; PS, VT, MC, PKV, GP, SCK, PSR, GVK—architecture, data collection and management; CVJ, PKV, VK—supervision and review. All authors have read and approved the final manuscript.

Funding

This work was supported by iHUB, INAI-International Institute of Information Technology, Hyderabad, India. The funding body had no role in the study design and analysis.

Declarations

Competing interests

Authors declare no competing interests.

Institutional review board

The studies involving human participants were reviewed and approved by Ikon independent ethics committee (SBS-2020-2).

Informed consent

The patients/participants provided their informed consent to participate in this study.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02802-5>.

Correspondence and requests for materials should be addressed to P.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

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/>.

© The Author(s) 2025