Research Paper

A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study

Qiuyun Fu[a,1], Yehansen Chen[e,1], Zhihang Li[e,1], Qianyan Jing[e,1], Chuanyu Hu[f,1], Han Liu[e], Jiahao Bao[e], Yuming Hong[e], Ting Shi[g], Kaixiong Li[a], Haixiao Zou[h], Yong Song[i], Hengkun Wang[j], Xiqian Wang[k], Yufan Wang[l], Jianying Liu[m], Hui Liu[n], Sulin Chen[o], Ruibin Chen[p], Man Zhang[d], Jingjing Zhao[q], Junbo Xiang[c], Bing Liu[a], Jun Jia[a], Hanjiang Wu[r], Yifang Zhao[a], Lin Wan[e,**], Xuepeng Xiong[a,b,*]

[a] Department of Oral and Maxillofacial Surgery, School and Hospital of Stomatology, Wuhan University, Wuhan, Hubei 430079, China
[b] The State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory of Oral Biomedicine Ministry of Education, Wuhan University, Wuhan, Hubei 430079, China
[c] Department of Periodontology, School and Hospital of Stomatology, Wuhan University, Wuhan, China
[d] Department of Orthodontics, Hubei-MOST KLOS and KLOBM, School and Hospital of Stomatology, Wuhan University, Wuhan, China
[e] School of Geography and Information Engineering, China University of Geosciences, Wuhan, China
[f] Center of Stomatology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[g] School of Information Engineering, Wuhan Huaxia University of Technology, Wuhan, China
[h] Department of Stomatology, the Second Affiliated Hospital of Nanchang University, Nanchang, China
[i] Department of Stomatology, Liuzhou People's Hospital, Liuzhou, China
[j] Department of Stomatology, Weihai Municipal Hospital, Weihai, China
[k] Oral Medical Center, Henan Provincial People's Hospital, School of Clinical Medicine, Henan University, Zhengzhou, China
[l] Department of Oral and Maxillofacial surgery, Peking University Shenzhen Hospital, Shenzhen, China
[m] Department of Stomatology, the People's Hospital of Zhengzhou, Zhengzhou, China
[n] Department of Oral and Maxillofacial Surgery, Shanghai Stomatological Hospital, Fudan University, Shanghai, China
[o] Department of Oral Implantology, School and Hospital of Stomatology, Fujian Medical University, Fuzhou, China
[p] Department of Oral Mucosal Diseases, Xiamen Key Laboratory of Stomatological Disease Diagnosis and Treatment, Stomatological Hospital of Xiamen Medical College, Xiamen, China
[q] Department of Oral Surgery, Jingmen No.2 People's Hospital, Jingmen, China
[r] Department of Oral and Maxillofacial Surgery, The Second Xiangya Hospital, Central South University, Changsha, China

ARTICLE INFO

ABSTRACT

Background: The overall prognosis of oral cancer remains poor because over half of patients are diagnosed at advanced-stages. Previously reported screening and earlier detection methods for oral cancer still largely rely on health workers' clinical experience and as yet there is no established method. We aimed to develop a rapid, non-invasive, cost-effective, and easy-to-use deep learning approach for identifying oral cavity squamous cell carcinoma (OCSCC) patients using photographic images.
Methods: We developed an automated deep learning algorithm using cascaded convolutional neural networks to detect OCSCC from photographic images. We included all biopsy-proven OCSCC photographs and normal controls of 44,409 clinical images collected from 11 hospitals around China between April 12, 2006, and Nov 25, 2019. We trained the algorithm on a randomly selected part of this dataset (development dataset) and used the rest for testing (internal validation dataset). Additionally, we curated an external validation dataset comprising clinical photographs from six representative journals in the field of dentistry and oral surgery. We also compared the performance of the algorithm with that of seven oral cancer specialists on a clinical validation dataset. We used the pathological reports as gold standard for OCSCC identification. We evaluated the algorithm performance on the internal, external, and clinical validation datasets by calculating the area under the receiver operating characteristic curves (AUCs), accuracy, sensitivity, and specificity with two-sided 95% CIs.
Findings: 1469 intraoral photographic images were used to validate our approach. The deep learning algorithm achieved an AUC of 0·983 (95% CI 0·973−0·991), sensitivity of 94·9% (0·915−0·978), and specificity of 88·7% (0·845−0·926) on the internal validation dataset (*n* = 401), and an AUC of 0·935 (0·910−0·957),

---

* Corresponding author at: Department of Oral and Maxillofacial Surgery, School and Hospital of Stomatology, Wuhan University, Wuhan, Hubei 430079, China.
** Corresponding author at: School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei 430078, China.
E-mail addresses: wanlin@cug.edu.cn (L. Wan), xiongxuepeng@whu.edu.cn (X. Xiong).
[1] These authors contributed equally.

sensitivity of 89·6% (0·847−0·942) and specificity of 80·6% (0·757−0·853) on the external validation dataset (*n* = 402). For a secondary analysis on the internal validation dataset, the algorithm presented an AUC of 0·995 (0·988−0·999), sensitivity of 97·4% (0·932−1·000) and specificity of 93·5% (0·882−0·979) in detecting early-stage OCSCC. On the clinical validation dataset (*n* = 666), our algorithm achieved comparable performance to that of the average oral cancer expert in terms of accuracy (92·3% [0·902−0·943] *vs* 92·4% [0·912−0·936]), sensitivity (91·0% [0·879−0·941] *vs* 91·7% [0·898−0·934]), and specificity (93·5% [0·909−0·960] *vs* 93·1% [0·914−0·948]). The algorithm also achieved significantly better performance than that of the average medical student (accuracy of 87·0% [0·855−0·885], sensitivity of 83·1% [0·807−0·854], and specificity of 90·7% [0·889−0·924]) and the average non-medical student (accuracy of 77·2% [0·757−0·787], sensitivity of 76·6% [0·743−0·788], and specificity of 77·9% [0·759−0·797]).

Interpretation: Automated detection of OCSCC by deep-learning-powered algorithm is a rapid, non-invasive, low-cost, and convenient method, which yielded comparable performance to that of human specialists and has the potential to be used as a clinical tool for fast screening, earlier detection, and therapeutic efficacy assessment of the cancer.

1. Introduction

Oral cancer is one of the common malignancies worldwide. There were an estimated 354,864 new cases and 177,384 deaths occurring in 2018, which represented 2% cancer cases and 1.9% cancer related deaths respectively [1]. Of all oral cavity cancer cases, approximately 90% are squamous cell carcinoma (SCC) [2]. Despite various emerging treatment modalities adopted over the past decades, the overall mortality of oral cavity squamous cell carcinoma (OCSCC) has not decreased significantly since the 1980s due to the relatively limited effort towards screening and early detection, which accounts for the stubborn rate of diagnosis with advanced-stage diseases [3].

The early detection of OCSCC is essential. As the estimated 5-year survival rate for OCSCC demonstrates a distinct decrease from 84% if detected in its early stages (stages I and II) to about 39% if detected in its advanced stages (stages III and IV) [3] And even worse, patients with advanced-stage diseases have to undergo poorer postoperative quality of life as a result of the suffering and costly process of multi-modal therapy including surgery, adjuvant radiation therapy with or without chemotherapy [4].

Unlike other internal organs, oral cavity allows for easy visualization without the need of special instruments. In clinical practice, specialists tend to make suspected diagnoses of oral cancer during visual inspection according to their own experience and knowledge on visual appearances of cancerous lesions [5,6]. Generally, OCSCC lesions often appear first as white or red patches, or mixed white-red patches, the mucosal surface usually exhibits an increasingly irregular, granular, and ulcerated appearance (see appendix p2 for details) [7,8]. Nevertheless, such visual patterns are easily mistaken for signs of ulceration or other oral mucous membrane diseases by non-specialist medical practitioners [8]. For a long time, there is no well-established vision-based method for oral cancer detection. The diagnosis of OCSCC has to rely on invasive oral biopsy which is not only time-consuming, but also not guaranteed in primary care or community settings, especially in developing countries [9,10]. Thus, quite often OCSCC patients cannot receive timely diagnosis and referrals [11,12].

There is growing evidence that deep learning techniques have matched or even outperformed human experts in identifying subtle visual patterns from photographic images, [13] including classifying skin lesions, [14] detecting diabetic retinopathy, [15] and identifying facial phenotypes of genetic disorders [16]. These impressive results inspire us to believe that deep learning also might have a potential to capture fine-grained features of oral cancer lesions, which is beneficial to the early detection of OCSCC.

With the assumption that deep neural networks could identified specific visual patterns of oral cancer like human experts, we developed a deep learning algorithm using photographic images for fully automated OCSCC detection. We evaluated the algorithm

Research in context

Evidence before this study

We searched PubMed on Jan 4, 2020, for articles that described the application of deep learning algorithm to detect oral cancer from images, using the search terms "deep learning" OR "convolutional neural network" AND "oral cavity squamous cell carcinoma" OR "oral cancer" AND "images", with no language or date restrictions. We found that previous researches were mainly limited to highly standardized images, such as multidimensional hyperspectral images, laser endomicroscopy images, computed tomography images, positron emission tomography images, histological images, and Raman spectra images. There were only two reports of artificial intelligence-enabled oral lesion classification using photographic images, which were published on Oct 10 and Dec 5, 2018, respectively. However, both of them suffered from extreme scarcity of data (<300 images in total) and depended heavily on specialized instruments that generated autofluorescence and white light images. In summary, we identified no research to allow direct comparison with our algorithm.

Added value of this study

To our knowledge, this is the first study to develop a deep learning algorithm for detection of OCSCC from photographic images. The high performance of the algorithm was validated in various scenarios, including detecting early oral cavity cancer lesions (diameters less than two centimetres). We compared the performance of the algorithm with that of oral cancer specialists on a clinical validation dataset and found its competence is comparable to or even beyond that of the oral cancer specialists. The deep learning algorithm was trained and tested with ordinary photographic images (for example, smartphone images) alone and did not require any other highly standardized images via a specialized instrument or invasive biopsy. Specifically, we developed a smartphone app on the basis of our algorithm to provide real-time detection of oral cancer. Our approach outputted reasonable scores for one OCSCC lesion during different cycles of chemotherapy, which exhibited a steady decline in parallel with the chemotherapy shrunk the lesion.

Implications of all the available evidence

Our study reveals that OCSCC lesions carry discriminative visual appearances, which can be identified by deep learning algorithm. The ability of detecting OCSCC in a point-of-care, low-cost, non-invasive, widely available, effective manner has significant clinical implications for OCSCC detecting.

performance on the internal and external validation datasets, and compared the model to the average performance of seven oral cancer specialists on a clinical validation dataset.

2. Methods

2.1. Datasets

We retrospectively collected 44,409 clinical oral photographs from 11 hospitals in China between April 12, 2006, and Nov 25, 2019. We included all biopsy-proven OCSCC photographs and normal controls by performing image quality control to remove intraoperative, postoperative and blurry photographs, and photographs of same lesion from approximate angles. We randomly selected 5775 photographs (development dataset) to develop the algorithm and used the remaining 401 photographs (internal validation dataset) for validation. We also included all photographs of early-stage OCSCC (lesion's diameter less than two centimetres) in the internal validation dataset to evaluate the algorithm performance in the early detection of OCSCC [17]. The corresponding pathological reports were used as the gold standard to develop and validate the deep learning algorithm.

We also curated an external validation dataset comprising 420 clinical photographs from six representative journals in the field of dentistry, and oral maxillofacial surgery (listed in the appendix, p 17−28), which were published between Jan, 2000 and Aug, 2019. We removed black-and-white, intraoperative, and dyed lesions photographs.

We acquired a clinical validation dataset from the outpatient departments of Hospital of Stomatology, Wuhan University between Nov 4, 2010, and Oct 8, 2019. This dataset contained 1941 photographs of OCSCC, other diseases or disorders of oral mucosa (see appendix, p 8−9 for details), and normal oral mucosa. We included all biopsy-proven photographs and normal controls by performing similar image quality control as mentioned above. All photographs involved in this study were stored in a jpg format. We classified photographs of normal mucosa as negative controls.

This study is reported according to STROBE guideline recommendations and approved by the Institutional Review Board (IRB) of the Ethics Committee of Hospital of Stomatology, Wuhan University (IRB No. 2019-B21). Informed consent from all participants was exempted by the IRB because of the retrospective nature of this study.

2.2. Algorithms development process

We developed an automated deep learning algorithm using cascaded convolutional neural networks to detect OCSCC from photographic images. A detection network firstly took an oral photograph as input and generated one bounding box that located the suspected lesion. The lesion area was cropped as a candidate patch according to the detection results returned by the first step. The candidate patch was then fed to a classification network which produced a list of two confidence scores in range of 0−1 for classification of patients with OCSCC and controls. The backbone networks of detection and classification were initialised with a pre-trained model that had been trained with tens of millions of images in the ImageNet dataset and further finetuned on the development dataset [18]. More details were described in the appendix (pp 3−7).

We also augmented our data to generate more training samples through image pre-processing like scaling, rotation, horizontal flipping and adjustment of the saturation and exposure (detailed in the appendix, p 6). Data augmentation was not executed on datasets used for validation. We developed the deep learning algorithm based on transfer learning (detailed in the appendix, p 5), which benefits to shorten the network training time and alleviate overfitting.

2.3. Human readers versus the algorithm

We compared the performance of the algorithm with that of 21 human readers on the clinical validation dataset. Readers employed in our study were divided into three panels according to their professional backgrounds and clinical experiences. The specialist panel consisted of seven oral cancer specialists from five hospitals. The medical student panel contained seven postgraduates who major in oral and maxillofacial surgery; and the non-medical student panel recruited seven non-medical undergraduates (readers' detailed information listed in the appendix, p 12). None of these readers participated in the clinical care or assessment of the enrolled patients, nor did they have access to their medical records.

For the purposes of the present study, we classified the photographs of OCSCC, non-OCSCC malignancies and oral epithelial dysplasia in the clinical validation dataset as oral cancer lesions; others (benign lesions and normal oral mucosa) as negative control. We chose this classification because both precancerous and cancerous oral lesions should be detected without delay in clinical practice.

Each reader was tested independently on the clinical validation dataset. We asked them to read each photograph in the dataset and record their judgements on the answer sheet. The photograph presentation order was randomized, and the answer sheet used in the test was shown in the appendix (pp 13). The performance of readers was assessed by comparing their predictions with corresponding pathological reports. We aggregated the final results and calculated the overall accuracy, sensitivity, and specificity of each panel.

2.4. Statistical analysis

We used the receiver operating characteristic (ROC) curve to evaluate the performance of the deep learning algorithm in discriminating OCSCC lesions from controls. The ROC curve was plotted by calculating the true positive rate (sensitivity) and the false positive rate (1-specificity) with different predicted probability thresholds, and we calculated AUC values [19]. We calculated 95% bootstrap CIs for accuracy, sensitivity, and specificity using 10,000 replicates [20]. Sensitivity was calculated as the fraction of photographs of oral cancer patients which were correctly classified, and specificity was calculated as the fraction of photographs of non-cancer individuals which were correctly classified. We used the average accuracy, sensitivity, and specificity of human readers when comparing with that of the model. We also employed statistical t-distributed Stochastic Neighbour Embedding (t-SNE) to demonstrate the effectiveness of our deep neural networks on differentiating OCSCC from non-OCSCC oral diseases [21]. All statistical analyses were done using scipy (version 0.22.1) and scikit-learn (version 1.4.1) python packages.

2.5. Role of the funding source

There was no funding source for this study. XPX and LW had full access to all the data and had final responsibility for the decision to submit for publication.

3. Results

In the development and internal validation datasets, 28,064 photographs (non-OCSCC diseases [n = 19,271] and non-biopsy-proven [n = 8793]) were excluded. Of these, another 9989 photographs were removed after image quality control, including intraoperative and postoperative photographs (n = 8551), photographs of same lesion from approximate angles (n = 1167), and blurry photographs (n = 271). We selected 402 of 420 photographs in the external validation dataset after removing 18 photographs, including the black and white (n = 12), intraoperative (n = 4), and dyed lesions (n = 2) photographs. In the clinical validation dataset, 233 photographs were
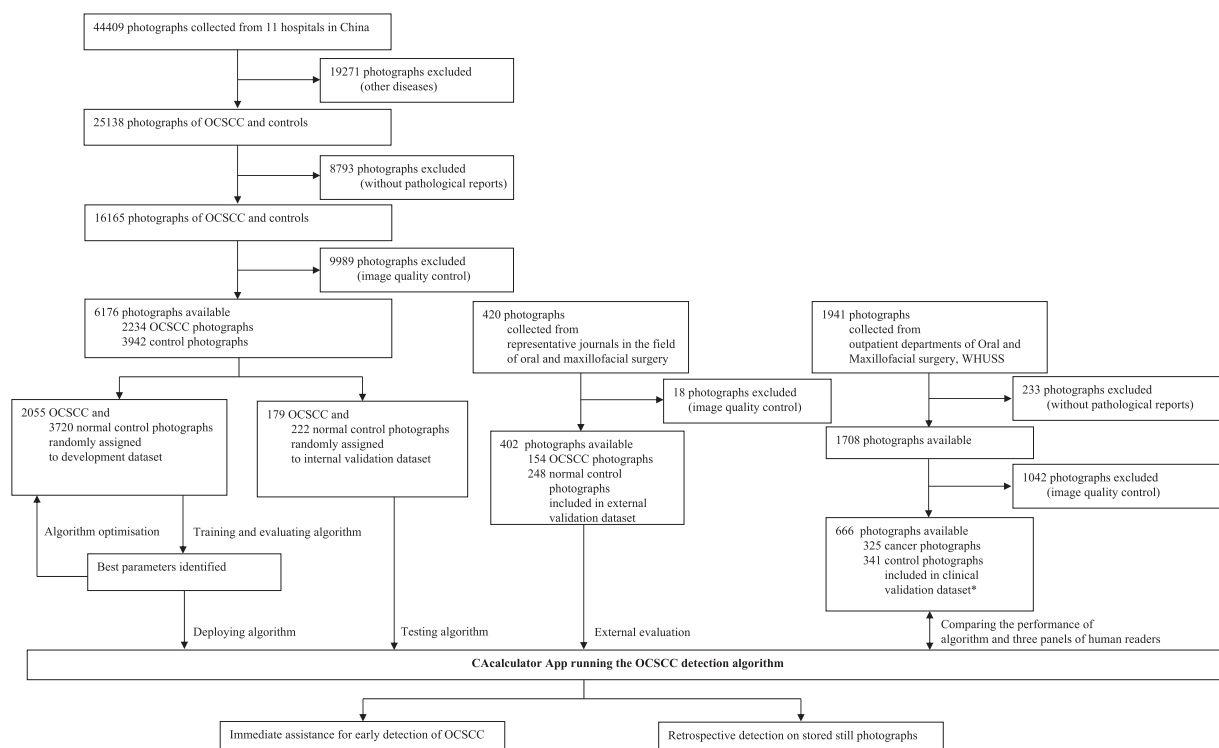
Fig. 1. Workflow diagram for the development and evaluation of the OCSCC detection algorithm
　*Cancer photographs were images of OCSCC, other malignancies, and epithelial dysplasia while control photographs were images of benign lesions and normal oral mucosa for the clinical validation dataset. OCSCC=oral cavity squamous cell carcinoma. WHUSS=School and Hospital of Stomatology, Wuhan University.

excluded for unavailable pathological reports. We performed similar quality control by removing intraoperative and postoperative photographs (n = 529), photographs of same lesion from approximate angles (n = 456), and blurry photographs (n = 57). Fig. 1 summarises the workflow diagram for the development and evaluation of the deep learning algorithm.

Baseline characteristics for the development and three validation datasets are summarised in Table 1. In the development dataset, 2055 photographs of OCSCC lesions were included while 3720

normal oral mucosa photographs were used as negative controls. The internal validation dataset contained 179 photographs of OCSCC lesions and 222 normal controls. The clinical validation dataset included 274 photographs of OCSCC lesions, 77 photographs of non-OCSCC oral diseases, and 315 photographs of normal oral mucosa. The external validation dataset consisted of 154 photographs of OCSCC lesions and 248 normal controls. Statistics for the sites of occurrence of OCSCC lesions were conducted according to the International Classification of Diseases 11th Revision (ICD-11) [22].

Table 1
Baseline characteristics.

	Development dataset	Internal validation dataset	Clinical validation dataset	External validation dataset	p value
Number of photographs	5775	401	666	402	..
Stage					0.033
Number of T1 OCSCC patients	459	101	51	..	..
Number of T2 OCSCC patients	471	27	54	..	..
Number of T3 OCSCC patients	110	7	18	..	..
Number of T4 OCSCC patients	82	1	8	..	..
Number of photographs for which age was unknown	3735	224	316	402	..
Mean age, years (range)	55 (19−88)	58 (26−89)	55 (21−83)	..	<0.0001
Lesion location					0.005
Squamous cell carcinoma of lip	99 (2%)	8 (2%)	6 (1%)	22 (6%)	..
Squamous cell carcinoma of tongue	901 (16%)	83 (20%)	120 (18%)	37 (9%)	..
Squamous cell carcinoma of gum	272 (5%)	21 (5%)	43 (7%)	34 (8%)	..
Squamous cell carcinoma of floor of mouth	202 (3%)	16 (4%)	9 (1%)	12 (3%)	..
Squamous cell carcinoma of palate	112 (2%)	10 (3%)	12 (2%)	10 (2%)	..
Squamous cell carcinoma of pharynx	40 (1%)	5 (1%)	18 (3%)	1 (1%)	..
Squamous cell carcinoma of other or unspecified parts of mouth	429 (7%)	36 (9%)	66 (10%)	38 (9%)	..
Non-OCSCC oral mucosal diseases*	0	0	77 (11%)	0	..
Normal oral mucosa	3720 (64%)	222 (56%)	315 (47%)	248 (62%)	..

OCSCC = oral cavity squamous cell carcinoma.
Data are n (%), unless otherwise stated.
　* Non-OCSCC oral mucosal diseases included non-OCSCC malignancies, epithelial dysplasia and benign lesions that were detailed in the appendix.
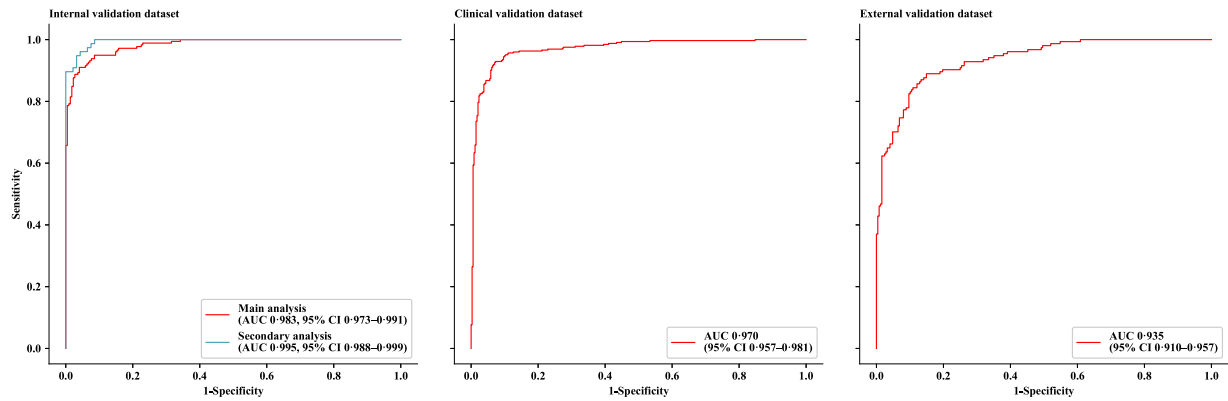
Fig. 2. ROC curves for the deep learning algorithm on three validation datasets
In the main analysis, all photographs in the internal validation dataset were used. In the secondary analysis, only photographs of early-stage oral cavity squamous cell carcinoma (lesion's diameter less than two centimetres) and random selected negative controls in the internal validation dataset were used. ROC=receiver operating characteristic. AUC=area under the curve.

Table 2
Algorithm performance.

	AUC	Sensitivity	Specificity	Accuracy
Internal validation dataset (n = 401)	0·983 (0·973−0·991)	94·9% (91·5−97·8)	88·7% (84·5−92·6)	91·5% (88·8−94·3)
Secondary analysis* (n = 170)	0·995 (0·988−0·999)	97·4% (93·2−100·0)	93·5% (88·2−97·9)	95·3% (91·8−98·2)
External validation dataset (n = 402)	0·935 (0·910−0·957)	89·6% (84·7−94·2)	80·6% (75·7−85·3)	84·1% (80·3−87·6)
Clinical validation dataset (n = 666)	0·970 (0·957−0·981)	91·0% (87·9−94·1)	93·5% (90·9−96·0)	92·3% (90·2−94·3)

Data in parentheses are 95% CIs.
  * In the secondary analysis, only photographs of early-stage oral cavity squamous cell carcinoma (lesion's diameter less than two centimetres) and random selected negative controls in the internal validation dataset were used.
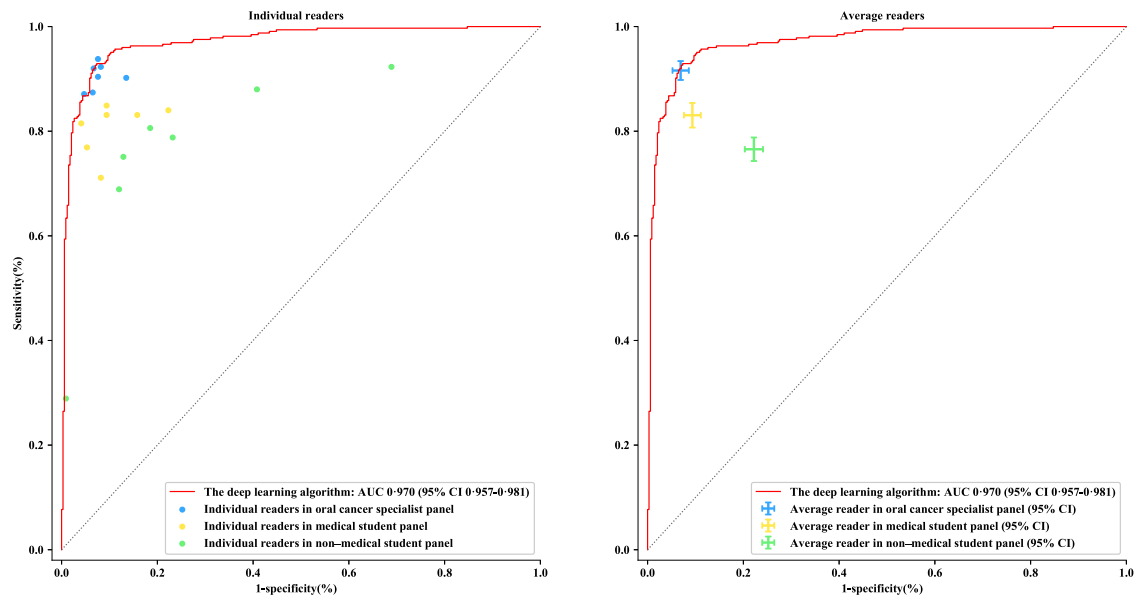
Fig. 3. Comparisons between the deep learning algorithm and three panels of human readers
The dots in the left subgraph indicate the performance of each individual. The crosses in the right subgraph demonstrate the average performance and corresponding error bar of each panel. OCSCC=oral cavity squamous cell carcinoma. AUC=area under the curve.

As in the primary analysis on all photographs in the internal validation dataset, the deep learning algorithm achieved an AUC of 0·983 (95% CI, 0·973−0·991), accuracy of 91·5% (88·8−94·3), sensitivity of 94·9% (91·5−97·8), and specificity of 88·7% (84·5−92·6) in detecting OCSCC lesions (Fig. 2 and Table 2). A secondary analysis was performed on all photographs of early-staged OCSCC lesions (n = 77) and randomly selected normal controls (n = 93) in the same dataset, which achieved an AUC of 0·995 (0·988−0·999) with accuracy of 95·3% (91·8−98·2), sensitivity of 97·4% (93·2−100·0), and specificity of 93·5% (88·2−97·9). Similarly, the model also achieved promising performance on the external validation dataset with an AUC of 0·935 (95% CI, 0·910−0·957), accuracy of 84·1% (80·3−87·6), sensitivity of 89·6% (84·5−94·1), and specificity of 80·6% (75·5−85·4).

The test results for the algorithm and three panels of human readers on the clinical validation dataset are shown in Fig. 3. The algorithm achieved an AUC of 0·970 (95% CI, 0·957−0·981) with accuracy of 92·3% (90·2−94·3), sensitivity of 91·0% (87·9−94·1), and specificity of 93·5% (90·9−96·0) in detecting oral cancer. Among the human readers, the accuracy of specialist panel was slightly higher than that of the algorithm at 92·4% (95% CI, 91·2−93·6) whereas 87·0%

(85·5−88·5) and 77·2% (75·7−78·7) for the medical and non-medical student panel, respectively. The sensitivity and specificity varied greatly among three panels: the model achieved comparable results to the specialist panel (sensitivity of 91·7% [95% CI, 89·8−93·4], and specificity of 93·1% [91·4−94·8]) and demonstrated significantly higher results than the medical student panel (sensitivity of 83·1% [80·7−85·4], and specificity of 90·7% [88·9−92·4]) and non-medical student panel (sensitivity of 76·6% [74·3−78·8], and specificity of 77·9% [75·9−79·7]). Results of each human reader were listed in the appendix (pp 14).

4. Discussion

In this study, we developed a deep learning algorithm that performs well (AUC 0·980 and 0·935 for the clinical and publication validation datasets, respectively) in automated detection of OCSCC from oral photographs, which is comparable favourably with the performance of seven oral cancer specialists. Our finding that deep learning can capture fine-grained visual patterns of OCSCC in cluttered oral image background with a speed and reliability matching or even beyond the capabilities of human experts validates what is, to our knowledge, the first fully automated, photographic-image-based approach for oral cancer lesion precise localization and recognition. This is of particular importance since such a non-invasive, rapid, and easy-to-use tool has significant clinical implications for early diagnosis or screening for suspected patients in countries that are lack of medical expertise.

Delays in referrals to cancer specialists is a primary cause of late presentation of quite a proportion of OCSCC patients (approximately 60%−65% at advanced stages), leading to a worse prognosis of this cancer [23,24]. Early detection of OCSCC is challenging because of poor public awareness and knowledge about oral cancer, particularly its clinical presentation. Furthermore, it is really hard for patients and even non-specialist healthcare professionals to perceive subtle visual signs of OCSCC from the variability in the appearance of oral mucosa lesions [11,12]. For example, appearances of tumours could be erythematous and ulcerative lesions since the onset, typically producing no prominent signs and discomfort until they progress. Trained health workers-based screening program conducted in previous studies did reduce oral cancer mortality but was costly, time-consuming, labor-intensive, and inefficient due to a large fraction of inexperienced non-medical undergraduate employees in the task [25−27]. In comparison, our work is novel in that no specific training or expert experience is required, the artificial intelligence (AI)-powered algorithm enables OCSCC lesions to be discriminated easily and automatically, just from one ordinary smartphone photo containing the suspicious region, achieving performance on par with human experts and far outperforming medical/non-medical school students (Fig. 3). Furthermore, we also built a mobile app on the basis of our OCSCC-recognition algorithm (see appendix pp 9−11), which might provide effective, easy, and low-cost medical assessments for more individuals in need than is possible with existing healthcare systems.

Apart from the photographic variability problem, identifying oral cancer from ordinary photographic images is a far trickier task than classifying skin-lesion diseases, [15] because OCSCC lesions are often hidden or masked in complex background by overlapping teeth, buccal mucosa, tongue, palate, and lip. Here we present a two-step deep learning algorithm to detect OCSCC in a 'coarse-to-fine' way. The detection network Single Shot MultiBox Detector (SSD) firstly spotted highly suspicious areas with OCSCC visual patterns in given photographs by filtering out unrelated contents [28]. DenseNet121 then classified those targeted regions into OCSCC or not [29]. Our deep neural networks achieve 92·3% (95% CI 0·902−0·943) overall accuracy whereas seven tested experts attain 92·4% (0·912−0·936) average accuracy. In addition, it also shows better generalization performance for varied datasets contained photographs taken by different cameras

(see appendix, p 2). These results imply that application-oriented deep neural network architecture is more effective to improve overall performance than simply stacking more layers into one network.

Our deep learning algorithm also generalizes well for early cancer lesions. Recognizing early-stage oral cavity cancer lesions, which are smaller than two centimetres and carried few visual features, [17] can be very difficult, but is effective to improve the curative effect, as the World Health Organization (WHO) stated. We found our deep neural networks to be helpful in identifying these very small OCSCC lesions in high-risk individuals, achieving a promising result (AUC 0·995) during the secondary analysis on internal validation dataset.

Another noteworthy finding is that our approach might potentially be used as a quantitative tool in aiding assessment of efficacy of therapeutic regimens. The deep learning algorithm outputs a score in terms of visual features extracted from lesion photos, which might be considered as a measure of the severity of cancer. It could be possibly helpful for assisting human specialists in rating the curative effects of non-surgical treatment modalities. For instance, the downward trend among outputted scores corresponded to a triple of oral photos taken for an OCSCC patient who received two cycles of docetaxel/cisplatin/5-fluorouracil (TPF) induction chemotherapy at three time points: before the chemotherapy, after the first chemotherapy cycle, and after the second chemotherapy cycle, which shows that the treatment was effective (see appendix pp 10−11). We believe that such a finding should be a fundamental basis of further clinical research and practice.

Despite recent advances in applying deep learning techniques to medical-imaging interpretation tasks, large datasets are remained as one prerequisite for achieving the performance of human-based diagnosis [14,15,16,30]. Unlike computed tomography (CT) and magnetic resonance imaging (MRI) image and electrocardiogram, taking oral photographs is not mandatory before treatment. Hence it is extraordinarily difficult to collect large amounts of photographs. The development dataset (5775 images) used in our study is not enough to train a robust deep learning model from scratch. Therefore, we adopt transfer learning by finetuning a pre-trained model trained on large-scale image datasets. Furthermore, data augmentation techniques are utilized to increase the size of our training set. Additional tricks, such as multi-task learning, hard example mining, etc., are also employed to improve our model (see appendix pp 6−7).

A limitation of our study is that the algorithm cannot make definite predictions for other oral diseases, mainly because the photographs used to train the deep neural networks may not fully represent the diversity and heterogeneity of oral disease lesions. Despite this, the internal features learned by the neural networks show our algorithm has the promising potential not only to distinguish OCSCC from non-OCSCC oral diseases, but also to differentiate between non-OCSCC oral disease and normal oral mucosal (Appendix Fig. 1.6). On the resulting plots of t-SNE representations of these three lesion classes, [22] each point represents one oral photo projected from the 1024-dimensional output of the last hidden layer of our neural network into two dimensions. We see the points of the same lesion class are aggregated into one cluster with the same color while OCSCC, non-OCSCC oral diseases and normal oral mucosal are well separated into three clusters. But still, the proposed algorithm fails in distinguishing several visually confusing cases such as epulis (see appendix pp 15 for more cases of where the algorithm and human experts failed to assess the malignancy correctly). Much larger diverse training dataset might be one possible solution to tackle this issue and will be tested in next clinical trials.

In conclusion, we report that deep learning methods may offer opportunities for automatically identifying OCSCC patients with the performance matching or even beyond that of skilled human experts. The developed algorithm with good generalization capability could be used as a handy, non-invasive, and cost-effectiveness tool for non-specialist people to detect OCSCC lesions as soon as possible, thereby enabling early treatment.

Declaration of Competing Interest

Funding

Contributions

XPX and LW designed and co-supervised the study. ZHL and QYJ implemented the deep learning algorithm and developed the CACalculator app. XPX, HJW, CYH, YS, HKW, YFW, XQW, MZ, YMH, JJ, BL, JBX, SLC, JJZ, QYF, and KXL contributed to data collection. QYF and JHB collated data and checked data sources. QYF, HXZ, JYL, and HuiLiu coordinated the human-AI competition experiment and reader recruitment. XPX, LW, ZHL, QYF, and YHSC drafted the manuscript. QYF, YHSC, and HanLiu did the statistical analysis, data interpretation, and constructed all tables and figures. YFZ and TS reviewed and revised the draft. All authors did several rounds of amendments and approved the final manuscript.

Data sharing

All photographic images used in the study are available from the corresponding authors on reasonable request.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.eclinm.2020.100558.

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.

[2] Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. Mayo Clin Proc 2008;83:489–501.

[3] Howlader N, Noone AM, Krapcho M, Miller D, Bishop K, Kosary CL, et al. SEER cancer statistics review, 1975-2014. Bethesda, MD Natl Cancer Inst 2017 2018.

[4] Hammerlid E, Bjordal K, Ahlner-Elmqvist M, Boysen M, Evensen JF, Biörklund A, et al. A prospective study of quality of life in head and neck cancer patients. Part I: at diagnosis. Laryngoscope 2001;111:669–80.

[5] der Waal I, de Bree R, Brakenhoff R, Coebegh JW. Early diagnosis in primary oral cancer: is it possible? Med Oral Patol Oral Cir Bucal 2011;16:e300–5.

[6] Kundel HL. History of research in medical image perception. J Am Coll Radiol 2006;3:402–8.

[7] Chi AC, Day TA, Neville BW. Oral cavity and oropharyngeal squamous cell carcinoma—An update. CA Cancer J Clin 2015;65:401–21.

[8] Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. Oral Oncol 2010;46:414–7.

[9] Moy E, Garcia MC, Bastian B, Rossen LM, Ingram DD, Faul M, et al. Leading causes of death in nonmetropolitan and metropolitan areas - United States, 1999-2014. MMWR Surveill Summ 2017;66:1–8.

[10] Pagedar NA, Kahl AR, Tasche KK, Seaman AT, Christensen AJ, Howren MB, et al. Incidence trends for upper aerodigestive tract cancers in rural United States counties. Head Neck 2019;41:2619–24.

[11] Gigliotti J, Madathil S, Makhoul N. Delays in oral cavity cancer. Int J Oral Maxillofac Surg 2019;48:1131–7.

[12] Liao DZ, Schlecht NF, Rosenblatt G, Kinkhabwala CM, Leonard JA, Ference RS, et al. Association of delayed time to treatment initiation with overall survival and recurrence among patients with head and neck squamous cell carcinoma in an underserved urban population. JAMA Otolaryngol - Head Neck Surg 2019;145:1001–9.

[13] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[14] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–8.

[15] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA J Am Med Assoc 2016;316:2402–10.

[16] Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 2019;25:60–4.

[17] Woolgar JA. Histopathological prognosticators in oral and oropharyngeal squamous cell carcinoma. Oral Oncol 2006;42:229–39.

[18] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115:211–52.

[19] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

[20] Efron B. Better bootstrap confidence intervals. J Am Stat Assoc 1987;82:171–85.

[21] Van Der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–625.

[22] International Classification of Diseases 11th Revision. May 25, 2019. https://icd.who.int/en/(accessed Jan 4, 2020)

[23] Scott SE, Grunfeld EA, Main J, McGurk M. Patient delay in oral cancer: a qualitative study of patients' experiences. Psycho-Oncol J Psychol Soc Behav Dimens Cancer 2006;15:474–85.

[24] De Vicente JC, Recio OR, Pendás SL, López-Arranz JS. Oral squamous cell carcinoma of the mandibular region: a survival study. Head Neck J Sci Spec Head Neck 2001;23:536–43.

[25] Sankaranarayanan R, Ramadas K, Thomas G, Muwonge R, Thara S, Mathew B, et al. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. Lancet 2005;365:1927–33.

[26] 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.

[27] Mathew B, Sankaranarayanan R, Sunilkumar KB, Kuruvila B, Pisani P, Krishnan Nair M. Reproducibility and validity of oral visual inspection by trained health workers in the detection of oral precancer and cancer. Br J Cancer 1997;76:390–4.

[28] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: single shot multibox detector. In: Proceeding of the european conference on computer vision; 2016 published online September 17. doi: 10.1007/978-3-319-46448-0_2.

[29] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017 published online June 22. doi: 10.1109/CVPR.2017.243.

[30] Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet 2018;392:2388–96.