Journal of Dental Sciences 18 (2023) 322–329

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

Oral squamous cell carcinoma diagnosis in digitized histological images using convolutional neural network

Kaori Oya [a†], Kazuma Kokomoto [b†], Kazunori Nozaki [b], Satoru Toyosawa [a,c*]

[a] *Division of Clinical Laboratory, Osaka University Dental Hospital, 1-8 Yamadaoka, Suita, Osaka, Japan*
[b] *Division for Medical Informatics, Osaka University Dental Hospital, 1-8 Yamadaoka, Suita, Osaka, Japan*
[c] *Department of Oral Pathology, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita, Osaka, Japan*

Received 1 August 2022; Final revision received 19 August 2022; accepted 19 August 2022
Available online 8 September 2022

KEYWORDS
Computational neural networks;
Artificial intelligence;
Digital image processing;
Oral squamous cell carcinoma

Abstract *Background/purpose:* Diagnostic methods of oral squamous cell carcinoma (SCC) using artificial intelligence (AI) and digital-histopathologic images have been developed. However, previous AI training methods have focused on the cellular atypia given by the training of high-magnification images, and little attention has been paid to structural atypia provided by low-power wide fields. Since oral SCC has histopathologic types with bland cytology, both cellular atypia and structural atypia must be considered as histopathologic features. This study aimed to investigate AI ability to judge oral SCC in a novel training method considering cellular and structural atypia and their suitability.

Materials and methods: We examined digitized histological whole-slide images from 90 randomly selected patients with tongue SCC who attended a dental hospital. Image patches of 1000 × 1000 pixels were cut from whole-slide images at 0.3125-, 1.25-, 5-, and 20-fold magnification, and 90,059 image patches were used for training and evaluation. These image patches were resized into 224 × 224, 384 × 384, 512 × 512, and 768 × 768 pixels, and the differences in input size were analyzed. EfficientNet B0 was utilized as the convolutional neural network model. Gradient-weighted class activation mapping (Grad-CAM) was used to elucidate its validity.

Results: The proposed method achieved a peak accuracy of 99.65% with an input size of 512 × 512 pixels. Grad-CAM suggested that AI focused on both cellular and structural atypia of SCC, and tended to focus on the region surrounding the basal layer.

* Corresponding author. Department of Oral Pathology, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: toyosawa@dent.osaka-u.ac.jp (S. Toyosawa).
† These authors contributed equally to the work.

https://doi.org/10.1016/j.jds.2022.08.017

Conclusion: Training AI regarding both cellular and structural atypia using various magnification images simultaneously may be suitable for the diagnosis of oral SCC.

Introduction

The incidence of oral cancers continues to increase.[1] In 2018, 246,000 cases of lip and oral carcinoma and 177,000 related deaths were reported worldwide.[2] Oral squamous cell carcinoma (SCC) is the most common type of oral cancer, accounting for more than 90% of oral cancers[3] that occur on the oral mucosal surface. Despite the development of new treatment modalities for cancer, the prognosis remains poor, with a low 5-year survival rate.[4–6] Early diagnosis of cancerous lesions can prevent progression through effective early treatment. However, small cancerous lesions often clinically resemble other noncancerous lesions[7] such as stomatitis. Underestimation by the clinician or patient himself/herself leads to delayed diagnosis and disease advancement. With respect to pathological diagnosis, there is also the potential for underdiagnosis that leads to disease progression. The oral mucosa easily displays significant cellular atypia (e.g., nuclear swelling) from inflammatory stimulation; meanwhile, well-differentiated oral SCC sometimes does not show remarkable cellular atypia.[8] In addition, histopathological diagnosis has the possibility of being influenced by subjective view.[9]

In this context, artificial intelligence (AI) is expected to assist in achieving a more reliable, objective, and accurate diagnosis that leads to better clinical outcomes.[9] Some studies have already used convolutional neural networks (CNNs), a major approach for AI-learning, for analyzing digital-histopathologic images of oral SCC. However, although these CNNs showed a high classification accuracy of more than 90%,[10–12] the training data set used in these previous studies had high magnification but a narrow range for histological images,[11,12] that is, their AI studied cellular atypia only. In general, pathologists diagnose oral SCC with cellular and structural atypia in both low- and high-power fields for the reasons described above. Structural atypia, which is observable in low-power fields, is important for diagnosis. Small-field or cell-by-cell observation has limitations in diagnosing oral SCC; therefore, a novel approach for training AI in structural atypia for carcinoma and cellular atypia is needed.

Thus, this study aimed to develop AI that can recognize both cellular and structural atypia to diagnose the entire SCC phenotype, including well-differentiated SCC with bland cytology. Towards this goal, broad views (large image patches) of various magnifications were used for training datasets that may resemble the pathologist's approach to diagnosis.

Materials and methods

Histological image processing

The histological dataset comprised primary tongue SCC samples with hematoxylin-eosin (HE) staining. The samples were acquired from biopsy or operative treatment of 90 patients randomly chosen from patients diagnosed with SCC during 2018–2020 at a university dental hospital. NanoZoomer (Hamamatsu Photonics K.K., Shizuoka, Japan) was used for the digitization of whole-slide images, and images were taken at 20-fold magnification (objective lens). The original images were too large for our computer; therefore, they were divided into patches of 1000 × 1000 pixels at 0.3125-,1.25-,5-, and 20-fold magnification. Image patches were categorized into "cancer, non-cancer, suspicious," and "do not know." Images with cancer, even if only minute, were also categorized as "cancer," whereas image patches with no cancer, such as muscle, connective tissue, minor salivary gland, and granulation tissue, were categorized as "non-cancer" (Fig. 1).

Images that could not be evaluated as "cancer" or "non-cancer" despite having indications of cancerous lesions were categorized as "suspicious." Images impossible to evaluate because of issues such as defocus were categorized as "do not know." For example, ulcer-like lesions that could not be diagnosed at low magnification were thus categorized as "do not know." These image classifications were performed by three board-certified pathologists. Finally, 90,059 image patches comprising 8332 cancer and 81,727 non-cancer sets were prepared as datasets. Other categories were not used for training.

Generally, when the dataset is imbalanced, oversampling or undersampling are commonly used for balancing data distributions. However, generating minority data or excluding majority data often leads to the deletion of important samples or inclusion of meaningless new objects.[13,14] In addition, given the nature of cancer, it seems unnatural to align the numbers of cancer and non-cancer. Therefore, we did not perform oversampling or undersampling, and trained our model in a straightforward manner.

Ethics approval statement

This study was approved by the appropriate ethical review board. Patients were informed of the opportunity to opt out on the website of our institution. Informed consent was not obtained.

Classification model training

EfficientNet is a CNN-based, high-performance image classification model that achieves state-of-the-art images on the ImageNet dataset[15] with better accuracy and efficiency than previous famous convolutional networks such as ResNet,[16] DenseNet,[17] Inception,[18] and Xception.[19] EfficientNet scales up from base B0 to B7 step by step, and the image recognition accuracy also improves as the base

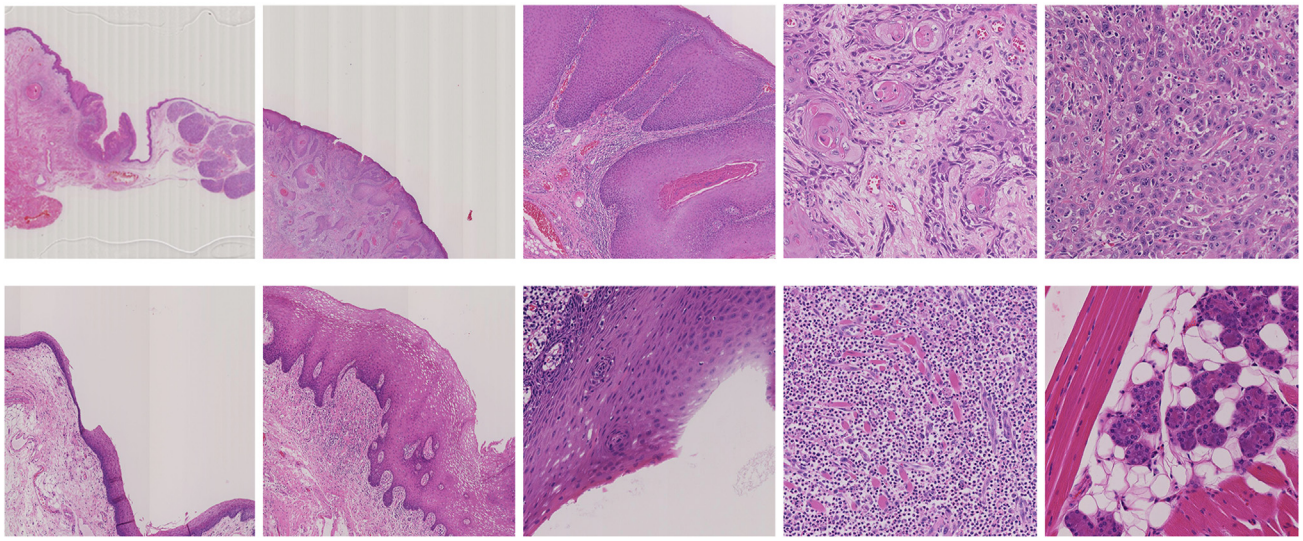

Figure 1 Histological image samples are used for training and evaluation in which the resolution is reduced to image sizes of 224 × 224, 384 × 384, 512 × 512, and 768 × 768 pixels after tagging.
Top: examples of cancer obtained at low to high magnifications (magnifications: ×0.3125, ×1.25, ×5, ×20, and ×20 from left to right). Bottom: image patches of non-cancer, including stratified squamous epithelium, skeletal muscle, inflammatory tissue, and salivary glands (magnifications: ×5, ×5, ×20, ×20, and ×20 from left to right).

increases. However, because the parameters and computational complexity increase exponentially as EfficientNet scales up, we utilized the lightest EfficientNet B0 as our classification model, and its parameters were initialized by a pretrained imagenet weight. Given that imagenet weight was not pretrained on histopathological images, it was considered undesirable to replace the fully connected head layers or freeze some layers. Therefore, we performed fine-tuning on all layers of EfficientNet B0. Using a larger input size has been reported to effectively increase image recognition performance.[20] EfficientNet B0 was trained with input sizes of 224 × 224, 384 × 384, 512 × 512, and 768 × 768 pixels, and the results were compared. Insufficient training data can lead to overfitting,[21] and our study had limited data. Thus, data augmentation, which included random choices of bellows (horizontal flip, vertical flip, hue, saturation, contrast, brightness, cropping, rotation, zoom, and shift), was utilized.

Evaluation

A 5-fold stratified cross-validation (5CV) (Fig. 2) was performed so that each fold could have the same proportion of cancer and non-cancer images. Training datasets were split into 5 folds, and EfficientNet B0 was trained with 4 folds of data. Subsequently, the remaining fold was evaluated with a trained model. We iterated this training five times, and our training metrics used out-of-fold predictions. AI diagnosis was visualized using gradient-weighted class activation mapping (Grad-CAM).[22] Two pathologists with 6 and 30 years of experience evaluated the results of Grad-CAM to determine the efficiency of the AI for recognizing both cellular and structural atypia in histological images.

Results

Accuracy, precision, and recall improved with increased image size to 512 × 512 pixels. The image patches of tongue normal tissue and SCC were classified with a peak accuracy of 99.65% (Fig. 3A). The precision and recall in the cancer group were 97.83% and 98.36%, respectively (data were rounded off to the second decimal place) (Fig. 3B).

Grad-CAM revealed that AI focused on reasonable features of cancer for judgment as follows. In high-magnification images of cancer, AI focused on cancer pearl and ectopic keratinization which are the most common features of well-differentiated SCC (Fig. 4A and B). Invasive tumor nests and atypical cells were reasonably focused on in cases of moderately (Fig. 4C and D) to poorly differentiated SCC (Fig. 4E and F). In the low-power field of ≤5-fold magnification, AI focused on anomalous structures, such as irregular overgrowth of epithelial components, abnormal distribution of stroma (Fig. 4G and H), and ectopic keratinization (Fig. 4I and J).

Furthermore, we found a tendency for AI to focus on the region surrounding the basal layer. In high-magnification images of cancer, especially when the cancer lesion showed sheet-like proliferation and there was no notable keratinized nest, the outer edges of the cancerous parenchyma were the focus (Fig. 5A and B). In the non-cancer images, AI focused on the boundary area between the epithelium and juxta-epithelial stroma (Fig. 5C and D) at every magnification. If the picture contained all layers of the stratified squamous epithelium at high magnification, the focus was dispersed throughout all layers of the epithelium (data not shown).

Grad-CAM also revealed that the focus points for evaluation were sharpened with increasing image size. As shown in Fig. 5E–H, AI recognized cancerous or non-

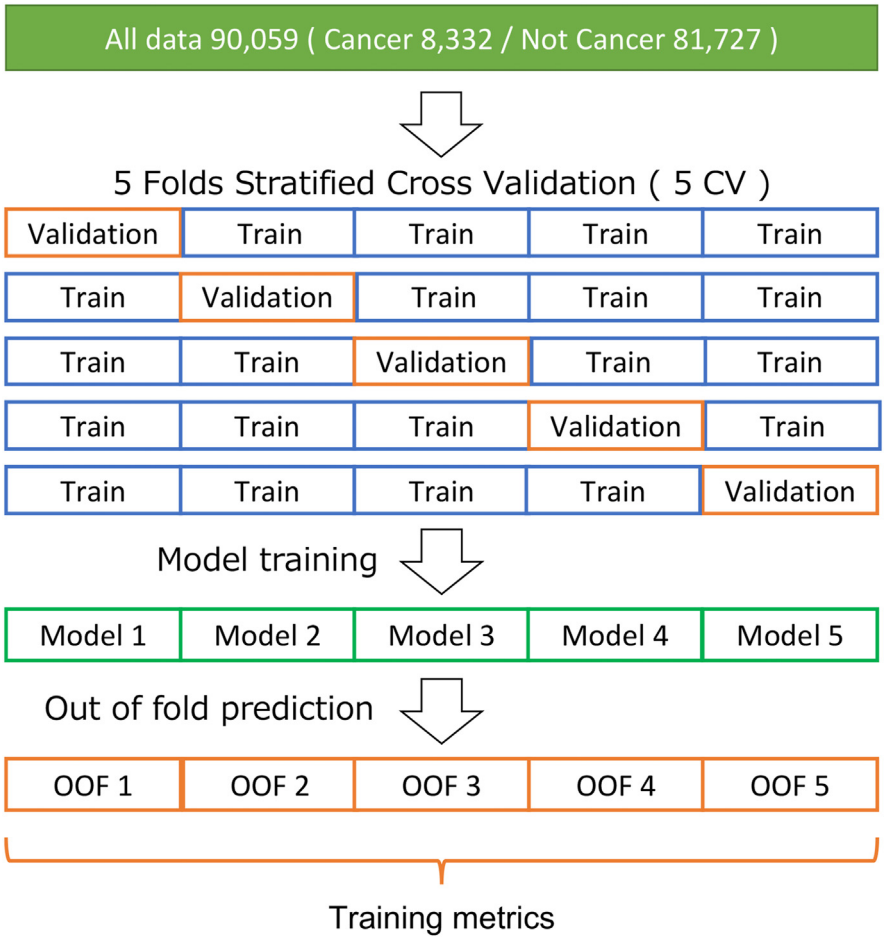

Figure 2 Scheme of classification model training and evaluation.
Training datasets are split into 5 folds, and EfficientNet B0 is trained with 4 folds of data. Subsequently, the remaining fold is evaluated with a trained model. This training is iterated five times, and out-of-fold predictions are used to evaluate our training results.

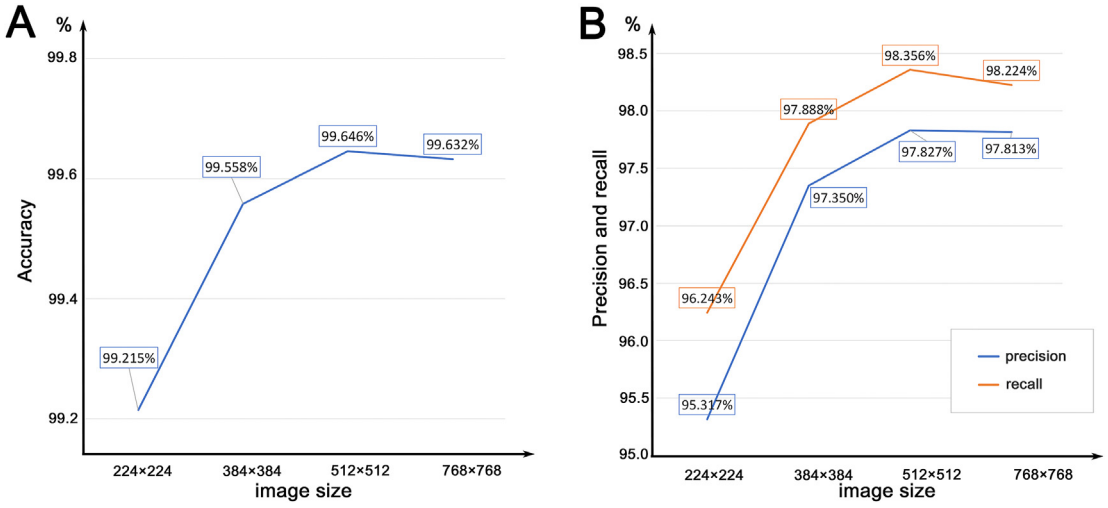

Figure 3 Accuracy, prediction, and recall at each image size.
The accuracy (A), precision, and recall (B) for cancer are improved by using larger (higher resolution) images. However, the performance improvement plateaued at resolutions higher than 512 × 512 pixels.

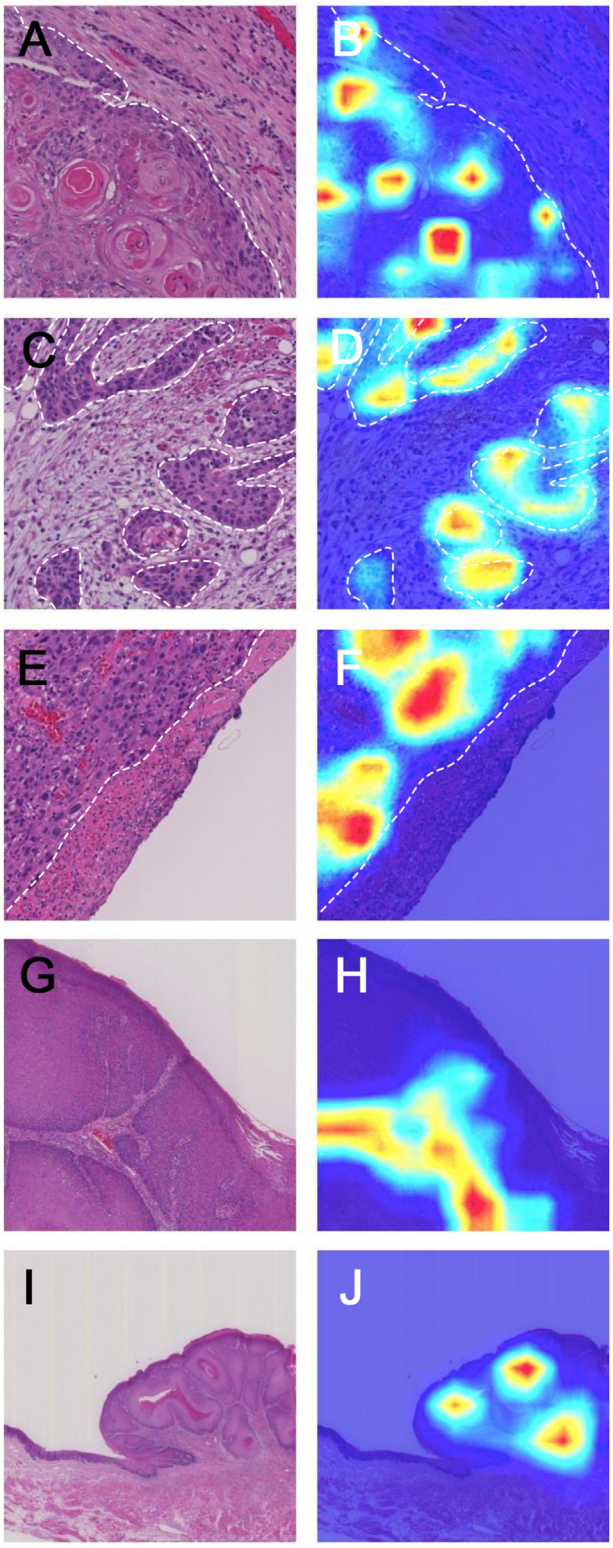

Figure 4 Representative images of cancer (A, C, E, G, I) and corresponding gradient-weighted class activation mapping (B, D, F, H, J).
In the case of well-differentiated squamous cell carcinoma (SCC) (A), cancer pearl and ectopic keratinization receive the most attention (B). Artificial intelligence (AI) focus on individual cancer cells and nests is almost similar between moderately differentiated (C, D) and poorly differentiated SCC (E, F). In low-power fields, abnormal distribution of stroma due to unusual hyperplasia of SCC (G) and ectopic keratinization (I) are focused on as cancer (H, J). Magnifications: ×20 in A—F; ×5 in G, H; and ×1.25 in I, J. Image size: 512 × 512 pixels. The dotted line shows the border between the epithelium and stroma or other components.

cancerous epithelial components, including the wide epithelial or stromal area, even in small-sized images (lower resolution). Further, the sharpness of the focus points was increased in larger images (higher resolution) (Fig. 5A—D). The atypical epithelium formed by the effect of inflammation was not focused on and was correctly classified as non-cancer. Ulcers, granulation tissue, denatured striated muscles with swollen nuclei, macrophages, and foreign body giant cells, often associated with cancer, were almost classified as non-cancer. The thickness of the tissue section and blurring of images did not result in classification errors.

However, there were some false-positive and false-negative classifications. For example, AI missed the cancer component located in the corner, probably because it was too focused on the non-cancer component (Fig. 6A and B). Moreover, minor salivary glands, muscle, or blood vessels located on the outside edge of the image patches were judged as cancer in a few cases (Fig. 6C and D).

Discussion

We investigated the effect of image size (resolution) on classification performance. The results showed that a higher resolution up to a size of 512 × 512 pixels improved the performance but did not increase further at a resolution of 768 × 768 pixels (Fig. 3). This indicates that an image size of 512 × 512 pixels is sufficient to achieve adequate classification performance in EfficientNetB0. In addition, Grad-CAM comparison by image size revealed that limiting the focus area of AI increased its performance. As shown in Fig. 5, focused points were confined to the boundary area between the epithelium and stroma at higher resolutions, whereas focus points were scattered widely at lower resolutions. These results suggest that alteration in the basal layer and juxta-epithelial stroma was recognizable at high resolution and an important factor for evaluation.

The results using images of 512 × 512 pixels were as follows. Trained AI showed high classification ability. There was little response to non-stratified squamous epithelium components (e.g., granulation tissue, denatured muscle, and endothelial cells with swelling nuclei), which are commonly confusing to the uninitiated. Thus, AI must have learned features of SCC. At high magnification, the prepared cancer images included recognizable cancerous cellular atypia, such as nuclear swelling, anisokaryosis, high nuclear-cytoplasmic ratio, hyperchromic nucleus, and large nucleolus. AI focused on such atypical cell with little information on structural atypia (Fig. 4E and F); therefore, AI must have learned individual cellular features of SCC as in previous studies. However, because the images contained

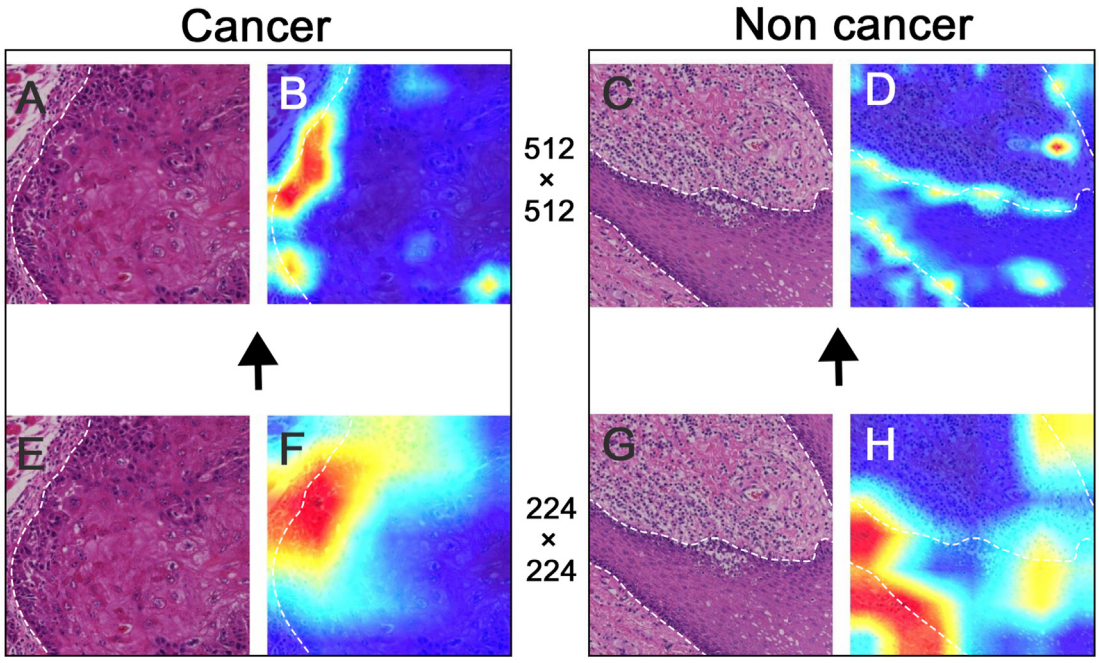

Figure 5 Artificial intelligence (AI) focuses on the outer edges of the cancerous parenchyma (A, B) or the region surrounding the basal layer of normal epithelium (C, D). AI focus is limited as the image size increases. In small-size images (224 × 224 pixels), AI recognizes the epithelial component, including the wide area (E–H). Magnification: ×20 in A–H. Image size: A–D, 512 × 512 pixels; and E–H, 224 × 224 pixels. The dotted line shows the border between the epithelium and stroma.

limited structural atypia, such as ectopic keratinization, AI learned and valued this over cellular atypia in many cases of well-differentiated SCC (Fig. 4A and B). In moderately differentiated SCC, pictures contained cellular and structural atypia; therefore, AI may have focused on both features (Fig. 4C and D).

Additionally, we confirmed that AI focused on the outside edges of the cancerous parenchyma, especially when the pictures contained stroma and parenchyma. This showed a sheet-like growth pattern (Fig. 5A and B), which is when the loss of basal cell polarity was emphasized in some cases of well-differentiated SCC. In the non-cancer images, AI also focused on the basal layer, including juxta-epithelial stroma, for judgement. It is also known that an alteration of basement membrane can be observed in cancer,[23] and the stroma affected by cancer becomes fibrotic.[24] These results support the abovementioned idea that alteration in the basal layer and juxta-epithelial stroma was recognizable and was as important as individual epithelial cellular atypia for evaluation. Moreover, when the picture contained all layers of the stratified squamous epithelium at high magnification, the focus was distributed throughout all layers of the epithelium in non-cancer images. This may suggest that differentiation toward the surface was also a focus.

At low-power fields of ≤5-fold magnification, AI also recognized cancer components despite the unclear cellular morphology. Structural atypia, such as irregular overgrowth of epithelial components, ectopic keratinization, and abnormal distribution of stroma due to unusual hyperplasia of the stratified squamous epithelium, was adequately focused on for classifying the lesion image as cancer. These results suggest that AI could distinguish cancer from non-cancer based on structural atypia without detailed information on cellular atypia. Collectively, the results support the notion that AI studied both cellular and structural atypia of SCC using the proposed method.

However, AI did not study images classified as "do not know" or "suspicious," which are images that we could not evaluate using image patches only. For example, some low-power magnification images of ulcers were classified as "do not know." Moreover, even if the image was clearly a part of cancer, it was classified as "suspicious" and was not used for training when the pathologist could not recognize the cellular or structural atypia as carcinoma within the patch. This indicates that AI still has limited applicability at this time. Images classified as "suspicious" provide important data for the diagnosis of carcinoma under AI, even if the pathologist's diagnosis is conflicting. Therefore, we tested the trained AI with the "suspicious" data set. The results varied widely for similar images (Supplemental Fig. 1). The count distribution of cancer probability showed a gradual sigmoid curve (Supplemental Fig. 2), suggesting that the trained AI contradicted the classification task.

As such, we tried to train AI with "suspicious" images as "cancer" to assist in improving its performance. However, the classification accuracy was slightly decreased. Although it had an insignificant effect on recall, precision was reduced in the cancer group. These results suggest that AI had learned the characteristics of cancer similar to pathologists, whereby "suspicious" images remained suspicious, and AI was confused by suspicious data, similar to pathologists. Pathologists do not use such suspicious data for diagnosis. As such, simple AI training that includes

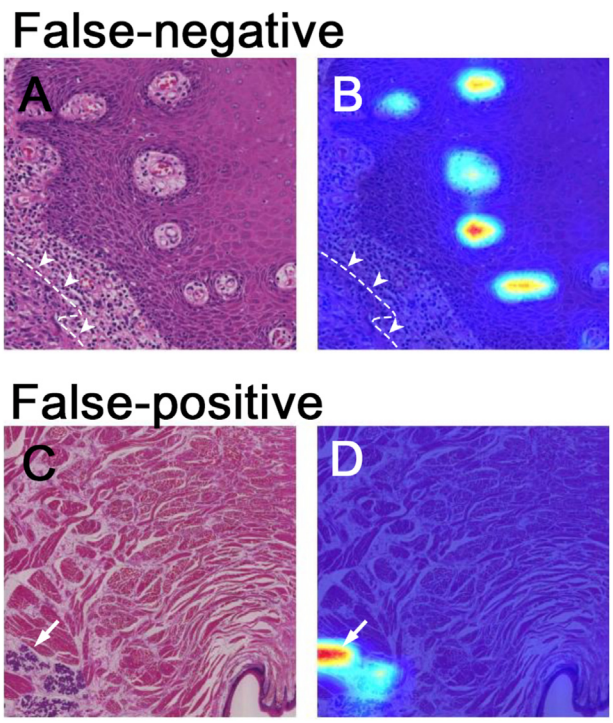

Figure 6 Examples of false-negative (A, B) and false-positive (C, D) findings. Artificial intelligence (AI) misses the cancer component located in the corner (A, B, arrowhead). AI highly focused on normal interstitial tissue (B). Minor salivary glands located on the outside edges of the images (C, arrow) are judged as cancer (D, arrow). Magnifications: ×20 in A, B; and ×1.25 in C, D. Image size: 512 × 512 pixels. The dotted line shows the border between the cancer nest and stroma.

"suspicious" cancer images may have a disadvantage. The proposed method in this study using only cancer and non-cancer images may be suitable for developing AI for the diagnosis of oral SCC. However, in clinical practice, suspicious images should be excluded in AI categorization at the pathologist's discretion. Despite these limitations, our results show that AI can help standardize oral SCC diagnoses. Further research may be needed to construct more reliable and reproducible AI, e.g., training with a larger sample size and validation with independent institutions and pathologists. Moreover, advanced methods using combined data of HE staining and immunostaining should be examined.

In conclusion, AI can be trained to evaluate cellular and structural atypia of oral SCC using our proposed method. Thus, our approach may be suitable for the diagnosis of oral SCC.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgements

We gratefully acknowledge the work of past and present members of the Department of Pathology, Osaka University Graduate School of Medicine, Suita, Osaka, Japan, for generously giving us permission to use NanoZoomer. There is no funder, and our income from pathological diagnosis for another dental hospital was used for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jds.2022.08.017.

References

1. American Cancer Society. *Cancer facts and figures 2021*. Kennesaw, GA: ACS; 2021. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf. [Accessed 3 August 2021].
2. 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 A Cancer J Clin* 2018;68:394–424.
3. Bagan J, Sarrion G, Jimenez Y. Oral cancer: clinical features. *Oral Oncol* 2010;46:414–7.
4. Wang W, Adeoye J, Thomson P, Choi SW. Statistical profiling of oral cancer and the prediction of outcome. *J Oral Pathol Med* 2021;50:39–46.
5. Neville BW, Day TA. Oral cancer and precancerous lesions. *CA A Cancer J Clin* 2002;52:195–215.
6. Iyer SG, Pradhan SA, Pai PS, Patil S. Surgical treatment outcomes of localized squamous carcinoma of buccal mucosa. *Head Neck* 2004;26:897–902.
7. Thompson L, Bishop J. *Head and neck pathology*, 3rd ed. Amsterdam: Elsevier, 2017:220.
8. Gnepp DR, Bishop JA. *Gnepp's diagnostic surgical pathology of the head and neck*, 3rd ed. Amsterdam: Elsevier, 2020:92–111.
9. Mehlum CS, Larsen SR, Kiss K, et al. Laryngeal precursor lesions: interrater and intrarater reliability of histopathological assessment. *Laryngoscope* 2018;128:2375–9.
10. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med* 2020;49:849–56.
11. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Network* 2020;128:47–60.
12. Halicek M, Shahedi M, Little JV, et al. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. *Sci Rep* 2019;9:14043.
13. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Progr Artif Intell* 2016;5:221–32.
14. Sun Y, Wong AKC, Kamel MS. Classification of imbalanced data: a review. *Int J Pattern Recogn Artif Intell* 2009;23:687–719.
15. Tan M, Le Q. *EfficientNet: rethinking model scaling for convolutional neural networks*. Available from: https://arxiv.org/abs/1905.11946. [Accessed 3 August 2021] [Date accessed.
16. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conf CVPR*; 2016:770–8.
17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conf CVPR*; 2017:2261–9.
18. Szegedy C, Vanhoucke V, loffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *2016 IEEE Conf CVPR*; 2016:2818–26.
19. Chollet F. Xception: deep learning with depthwise separable convolutions. In: *2017 IEEE Conf CVPR*; 2017:1800–7.

20. Wei Q, Li X, Wang H, Ding D, Yu W, Chen Y. Laser scar detection in fundus images using convolutional neural networks. In: Jawahar C, Li H, Mori G, Schindler K, eds. *Computer vision — ACCV. ACCV 2018. Lecture notes in computer science*, vol. 11364. Cham: Springer, 2018:2019.
21. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017; 60:84—90.
22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Visual explanations from deep networks via gradient-based localization*. Available from: https://arxiv.org/abs/1610.02391. [Accessed 3 August 2021].
23. Zargaran M, Eshghyar N, Vaziri PB, Mortazavi H. Immunohistochemical evaluation of type IV collagen and laminin-332 γ2 chain expression in well-differentiated oral squamous cell carcinoma and oral verrucous carcinoma: a new recommended cut-off. *J Oral Pathol Med* 2011;40:167—73.
24. Paral KM, Taxy JB, Lingen MW. CD34 and α smooth muscle actin distinguish verrucous hyperplasia from verrucous carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2014;117:477—82.