Journal of Dental Sciences 20 (2025) 54–60

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

Deep learning system for the differential diagnosis of oral mucosal lesions through clinical photographic imaging

An-Yu Su [a], Ming-Long Wu [b,c], Yu-Hsueh Wu [a,d,e*]

[a] School of Dentistry, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[b] Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan
[c] Institute of Medical Informatics, National Cheng Kung University, Tainan, Taiwan
[d] Institute of Oral Medicine, School of Dentistry, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[e] Department of Stomatology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Received 7 October 2024; Final revision received 15 October 2024
Available online 28 October 2024

KEYWORDS
Artificial intelligence;
Computer-assisted diagnosis;
Deep learning;
Oral diagnosis;
Oral mucosa

Abstract *Background/purpose:* Oral mucosal lesions are associated with a variety of pathological conditions. Most deep-learning-based convolutional neural network (CNN) systems for computer-aided diagnosis of oral lesions have typically concentrated on determining limited aspects of differential diagnosis. This study aimed to develop a CNN-based diagnostic model capable of classifying clinical photographs of oral ulcerative and associated lesions into five different diagnoses, thereby assisting clinicians in making accurate differential diagnoses.
Materials and methods: A set of clinical images were selected, including 506 images of five different diagnoses. The images were pre-processed and randomly divided into two sets for training and testing the CNN model. The model architecture was composed of convolutional layers, batch normalization layers, max pooling layers, the dropout layer and fully-connected layers. Evaluation metrics included weighted-precision, weighted-recall, weighted-F1 score, average specificity, Cohen's Kappa coefficient, normalized confusion matrix and AUC.
Results: The overall performance for the image classification showed a weighted-precision of 88.8%, a weighted-recall of 88.2%, a weighted-F1 score of 0.878, an average pecificity of 97.0%, a Kappa coefficient of 0.851, and an average AUC of 0.985.
Conclusion: The model achieved a decent classification performance (overall AUC = 0.985), showing the capacity to discern between benign and malignant potential lesions, and laid the foundation of a novel tool that can help clinical differential diagnosis of oral mucosal

* Corresponding author. Institute of Oral Medicine, School of Dentistry, College of Medicine, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan.
E-mail address: z10908025@ncku.edu.tw (Y.-H. Wu).

https://doi.org/10.1016/j.jds.2024.10.019

lesions. The main challenges were the small and imbalanced dataset. Enlarging the minority classes, incorporating more oral mucosal lesion diagnoses, employing transfer learning and cross-validation might be included in future works to optimize the image classification model.

Introduction

Oral mucosal lesions (OMLs) are common oral pathology encountered in daily clinical practice, ranging from infectious diseases, immunological disorders, to premalignancy. They could manifest as ulcers, which are truly epithelial and connective tissue defects, or as mixed red-and-white lesions that mimic erosions or ulcers.[1–3] Oral mucosal lesions, including oral ulcers, may be associated with distinct etiological factors so it is important for clinicians to differentiate one disease from another.[3,4] Aphthous ulcer (AU) is the most common oral ulcerative lesions and presents as one or several painful punched-out sores on oral mucous membrane. Most aphthae are self-limiting or can be relieved by eliminating the predisposing factors.[5] On the other hand, oral lichen planus (OLP), as an autoimmune disease, may present as erosions and ulcers with intermixed white striae and patches.[6] Oral precancerous lesions (OPLs), such as speckled leukoplakia, is a mixed red-and-white lesion with a high risk of malignant transformation.[7–9] Hence, making a correct diagnosis is essential for guiding proper treatment.

Recently, increasing studies demonstrate that deep-learning-based convolutional neural networks (CNNs), particularly image recognition, have found extensive applications in computer-aided diagnosis (CAD) for various medical purposes, including skin disease recognition, breast cancer diagnosis, caries detection, tooth segmentation and pulp exposure prediction. Furthermore, the performance in disease detection and diagnosis indicates promising results.[10–17] Diagnosis of oral mucosal lesions greatly relies on inspecting clinical presentations, including the color, size, and distribution of the lesions. Thus, the diagnosis process might be challenging and largely depend on the clinician's experience at present. There have been several studies working on CNN systems in computer-aided oral lesion diagnosis; however, most of them focused on differentiating limited aspects of disease entity.[18] On the grounds that artificial intelligence (AI) has gained increasing progress in the field of medicine, employing the deep learning algorithm in the automate oral-mucosal-lesion-image-diagnosis might be an efficient and novel way.[19–21]

The aim of this study was to develop a CNN-based diagnosis model capable of classifying clinical photographs of oral ulcerative lesions or some associated oral lesions into five distinct diagnoses; further, this CNN-based diagnosis model was expected to be applied in clinical use in the future.

Materials and methods

Data collection

Photographic images in this study were collected from the patients that were diagnosed in the Department of Stomatology of National Cheng Kung University Hospital, from January 2003 to May 2023. Additional images from classical articles and textbooks were also included to enhance the reliability.[22–29] The selection criteria included the following five diagnoses: AU (including Beçhet's disease), OLP, oral candidiasis (OC), OPL, and oral submucous fibrosis (OSF). In addition to those diseases present with apparent oral ulcers or lesions mimicking oral ulcers, OSF was also included since it was proved to have malignant potential, just as OPL was. Subsequently, an image set representing the aforementioned diagnoses from different anatomical sites in oral cavity was acquired. These images were reviewed by a well-trained researcher to ensure the correct diagnosis, then cropped to the boundaries of the oral cavity. The image dataset was sorted into different groups by respective diagnosis. Finally, the five groups, a total of 506 images, comprised 91 images of AU (including Beçhet's disease), 114 images of OC, 119 images of OLP, 104 images of OPL and 78 images of OSF, were selected and presented in Table 1. This study was reviewed and approved by the Institutional Review Board of National Cheng Kung University Hospital (A-ER-112-147).

Image pre-processing

To standardize every image, the technique of data transformation was utilized to transform the data into a suitable

Table 1 Disease classes and the number of images included within each class.

Classes	Number of images
AU[a]	91
OC	114
OLP	119
OPL	104
OSF	78
Total	506

Abbreviations: AU, aphthous ulcer; OC, oral candidiasis; OLP, oral lichen planus; OPL, oral precancerous lesion; OSF, oral submucous fibrosis.
[a] Including Beçhet's disease.

format so that it could be more accessible and informative for further analysis or modeling.[20,30] Other than that, data augmentation was employed due to the imbalanced quantity of different groups of data. The purpose of augmentation is to expand the size of a dataset by creating modified versions of images in the dataset, such as flipping the image and contrast adjustment. This helps prevent overfitting and can lead to improved model generalization, especially when the original dataset is relatively small.[30,31] In this study, horizontal flipping and image rotation (within $\pm 10^{\circ}$) were adopted. After image transformation and augmentation, the dataset of 544 images were randomly divided by the algorithm into training set and test set in a ratio of 7:3.[32] Thus, 384 and 160 of images were included in the training set and test set, respectively. The training set was the major data for training the CNN model, while the model's performance was evaluated by the test set.

CNN model building

The architecture of the model consisted of several steps, including four convolution layers with a kernel size of 5×5, four batch normalization layers after each convolution layer, two max-pooling layers with a kernel size of 2×2, the dropout layers, and three fully connected layers. The four convolution layers had 12, 12, 24 and 24 filters, respectively. Dropout is a regularization approach that prevents the CNN model encountering overfitting by randomly dropping units during the training.[33] The input size of the images was $128 \times 128 \times 3$ pixels. The workflow of the CNN model was illustrated in Fig. 1.[34]

Evaluating metrics

To evaluate the performance of classification tasks of CNN model, accuracy, weighted-precision, weighted-recall, weighted-F1 score, average specificity, Cohen's Kappa coefficient, normalized confusion matrix, receiver operating characteristics curve (ROC) and area under ROC (AUC) with 95 % confidential intervals were utilized. Since "weighted-" suggests that the scores are calculated depending on the number of samples available in that class, these metrics are suitable for evaluating a multiclass image classification model which has a class-imbalanced dataset. Cohen's Kappa coefficient is a score that expresses the level of agreement between two annotators on a classification problem. The mathematical expressions show as follows:

1. Weighted-Precision (Positive Predictive Value) $= \sum_{i=1}^{C} w_i * Precision_i$
   Where when given C classes,
   (1) $Precision_i$ is the precision of class i, while the precision being the accuracy of positive predications: $\frac{True\ Positives\ (TP)_i}{True\ Positives\ (TP)_i + False\ Positives\ (FP)_i}$
   (2) w_i is the *weight* for class i, typically the proportion of true instances of class i in the dataset.
2. Weighted-Recall (True Positive Rate) $= \sum_{i=1}^{C} w_i * Recall_i$.
   Where when given C classes,
   (1) $Recall_i$ is the recall of class i, while the recall being the fraction of correctly identified positive predictions: $\frac{TP_i}{TP_i + False\ Negatives\ (FN)_i}$
   (2) w_i is the weight for class i, typically the proportion of true instances of class i in the dataset.
3. Weighted-F1 score $= \sum_{i=1}^{C} w_i * F1_i$
   Where when given C classes.
   (1) $F1_i$ is the F1 score for class i, while F1 score being the harmonic mean of precision and recall
   (2) $F1_i = 2 * \frac{Precision_i * Recall_i}{Precision_i + Recall_i}$
4. Average Specificity (True Negative Rate) $= \frac{1}{C} \sum_{i}^{C} Specificity_i$
   Where when given C classes,
   (1) $Specificity_i$ is the specificity for class i, while the specificity being $\frac{True\ Negatives\ (TN)_i}{True\ Negatives\ (TN)_i + False\ Positives\ (FP)_i}$
5. Cohen's Kappa coefficient $= \kappa = \frac{p_o - p_e}{1 - p_e}$.
   (1) p_o is the observed agreement and p_e is the expected agreement when both annotators assign labels randomly.
   (2) $p_e = \sum_{i=1}^{C} p_{true,i} * p_{pred,i}$
      i. $p_{true,i}$ is the proportion of true instances in class i.
      ii. $p_{pred,i}$ is the proportion of predicted instances in class i.

A confusion matrix can help summarize the performance of a classification model by comparing predicted and actual values, and thus evaluate whether the model is biased toward the majority class or consistently misclassifying certain classes. For multi-class classification, the ROC curve is normally plotted for each class against the others (One vs Rest, OvR).

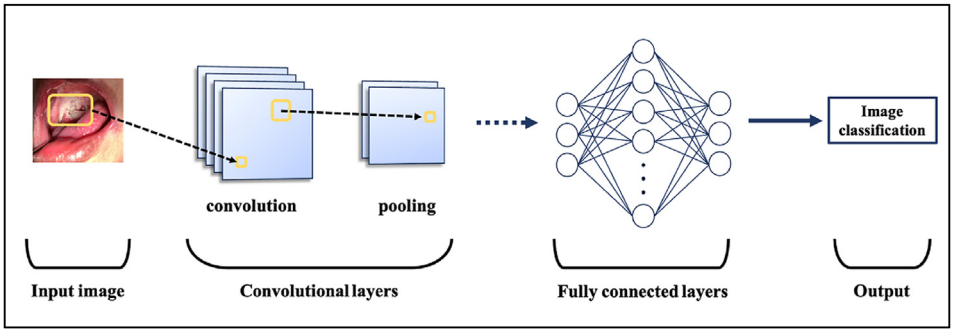

Figure 1 The workflow of the convolutional neural network (CNN) model. The model consisted of 4 convolution layers, each with a batch normalization layer following, 2 max-pooling layers, the dropout layers, and 3 fully connected layers.

Results

Class-based performance in the test set

The precision, recall, and F1-score for each disease class in the test set were shown in Table 2. Among all classes, the OLP group had the highest precision (97.0 %), followed by the AU group (96.0 %), the OSF group (90.0 %), the OPL group (84.4 %), and the OC group with the lowest (78.6 %). In terms of recall, the AU group ranked the highest (100 %), followed by the OC group (97.1 %), the OLP group (94.1 %), the OPL group (84.4 %), and the OSF group with the lowest (64.3 %). For the F1-score, which is the harmonic mean of precision and recall, the AU group led with 0.980, followed by the OLP group (0.955), the oral candidiasis group (0.868), the OPL group (0.844), and the OSF group (0.750).

The normalized confusion matrix to evaluate the performance of the CNN model in the test set was illustrated in Fig. 2, indicating that the CNN model reached a normalized true positive rate of 1.00 for the AU group, 0.97 for the OC group, 0.94 for the OLP group, 0.84 for the OPL group and 0.64 for the OSF group. The OSF group was the most commonly misclassified group (recall = 0.64), often being labeled as OC or OPL by the algorithm. The OPL group was prone to misclassification as OC and OSF, while OLP tended to be misclassified as OC and OPL.

Class-based ROC curves and AUC values

The ROC curves for each class with the relevant AUC values were shown in Fig. 3. The respective AUC for each class were: 1.00 for the AU group, 0.99 for the OC group, 0.99 for the OLP group, 0.98 for the OPL group, and 0.98 for the OSF group.

Overall performance for the oral-mucosal-lesion-image classification model

The overall performance of the CNN model achieved a weighted-precision of 88.8 %, a weighted-recall of 88.2 %, a weighted-F1 score of 0.878, an average specificity of 97.0 %, a Kappa coefficient of 0.851, and an average AUC of 0.985.

Table 2 Precision, recall, and F1-score of each disease class in the test set.

	Precision (%)	Recall (%)	F1-score
AU[a]	96.0	100.0	0.980
OC	78.6	97.1	0.868
OLP	97.0	94.1	0.955
OPL	84.4	84.4	0.844
OSF	90.0	64.3	0.750

Abbreviations: AU, aphthous ulcer; OC, oral candidiasis; OLP, oral lichen planus; OPL, oral precancerous lesion; OSF, oral submucous fibrosis.

[a] Including Beçhet's disease.

Discussion

Diagnosis of oral mucosal lesions might be challenging and largely depend on the clinician's experience in current. Progressively, deep learning system has gained its popularity in aiding diagnosis of oral mucosal lesions. Zhou et al. reported a CNN model using pretrained algorithm for classification and detection of recurrent aphthous ulcerative lesions based on non-invasive oral images.[32] The presented deep learning CNN image classification model aimed to help clinical work using five different diagnoses of oral mucosal lesion photographs. Accurate diagnosis via CNN model not only can guide the proper treatment but also greatly palliate medical burden. In this study, the model achieved a final AUC of 0.985, which could be regarded as an excellent differential classification performance.

The model performed outstanding on AU (including Beçhet's disease) since it achieved high precision (96.0 %) and recall (100 %), suggesting the capacity of correctly identifying the images of this class and rarely misclassified other classes as AU. The recall was high for oral candidiasis (97.1 %), which meant the model identified most instances of this class. However, the precision was relatively low (78.6 %), indicating that the model easily misclassified other classes as oral candidiasis. The model performed well on OLP due to the strong precision (97.0 %) and recall (94.1 %). The model's performance on OPL was balanced, with a fairly high precision (84.4 %) and recall (84.4 %). For OSF, there were a high precision (90.0 %) but a low recall (64.3 %), suggesting that the model was good at predicting OSF when the images actually were. On the other hand, the model had the tendency of missing actual OSF instances.

According to the confusion matrix, OLP, OPL, and OSF were most likely to be misclassified as oral candidiasis, which explains the low precision of oral candidiasis. OSF was often mislabeled as oral candidiasis, OPL, and OLP, contributing to its low recall. These were probably due to the small dataset of OSF and the similar features between oral candidiasis and other classes, which affected recall and precision. To improve the results, increasing the OSF dataset and enhancing feature extraction from images might be necessary. All in all, as the respective ROC curve and AUC for each class indicated, the model still seemed to have a decent overall performance and certainly had the capacity to discern between benign and malignant-potential lesions (Fig. 3). It was notable that although OPL and OSF had identical AUC values (0.98), they varied significantly in recall, with OPL at 84.4 % and OSF lower at 64.3 %. A high AUC but low recall is not an uncommon phenomenon, especially in imbalanced datasets, and it arises from the different approaches in which AUC and recall evaluate the model's performance. The high AUC score for OSF indicated that the model was good at distinguishing between the class and other classes across different thresholds. However, the lower recall value suggested that at specific threshold, the model missed more true positive instances for OSF, leading to a higher false-negative rate.

The probable overfitting of a model, which is a common challenge in CNN model training, usually indicates that the

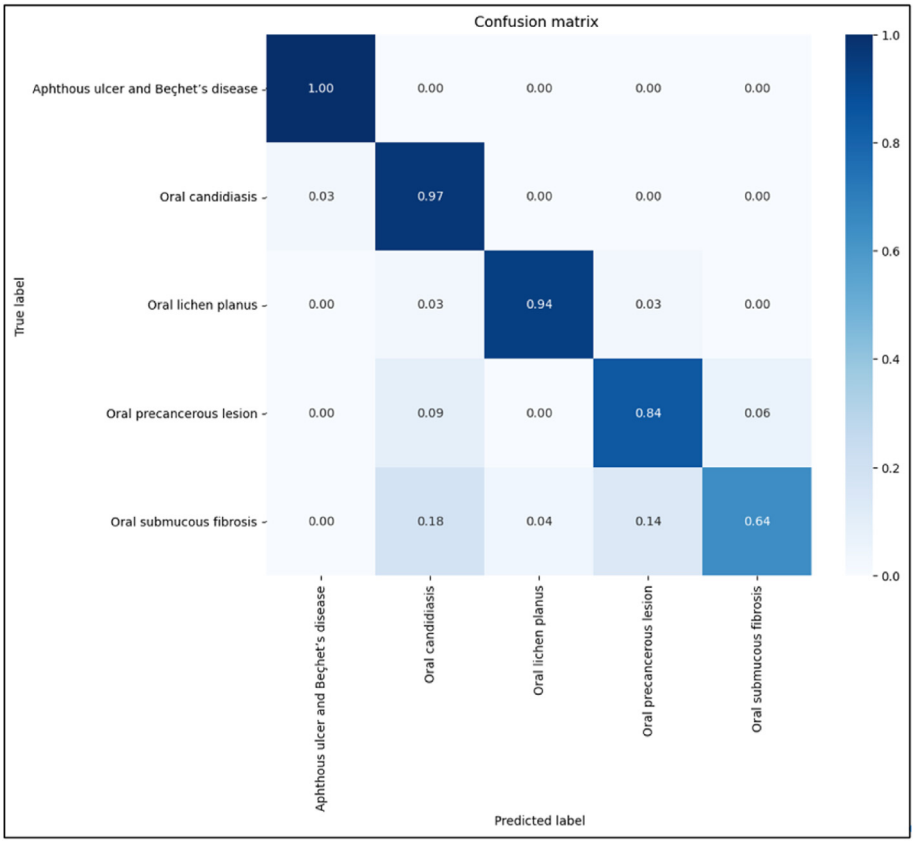

Figure 2 The normalized confusion matrix of the test set.

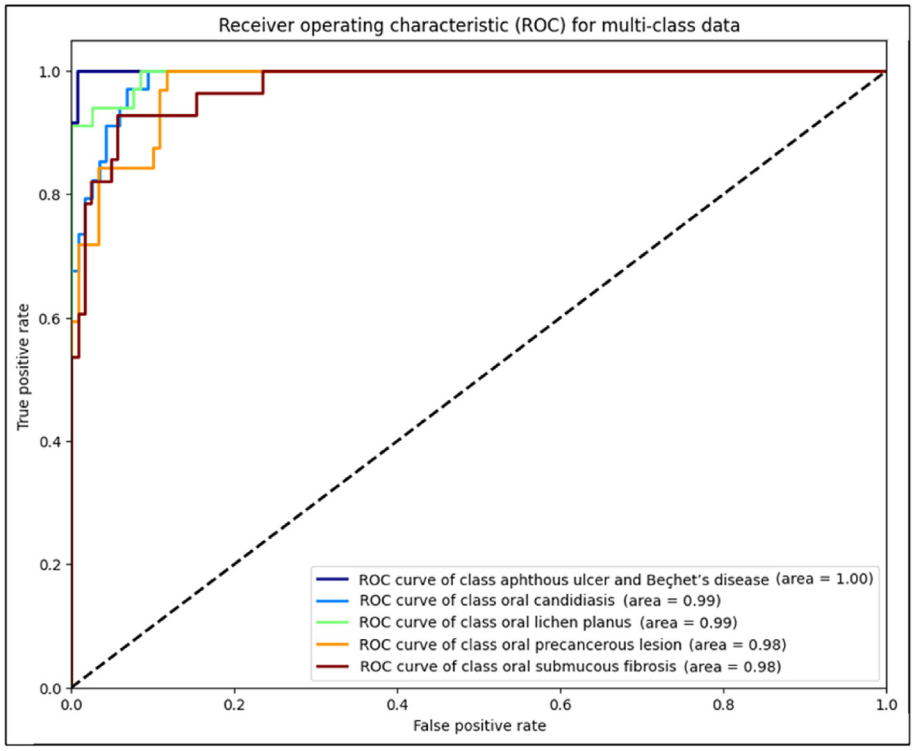

Figure 3 The ROC curve for each class with the relevant AUC value. Through presenting the ROC curve and AUC, the model can be considered capable of discerning between benign and malignant-potential lesions. AUC: area under curve. ROC: receiver operating characteristic.

model is memorizing the training data instead of learning the pattern. This phenomenon is particularly pronounced when the dataset is small and imbalanced. Alternatively, early stopping prevents the model from learning noise in the training data by monitoring the peak of test accuracy and terminating the training session as accuracy starts to decline. Priorly, to handle the problem, the dropout technique and data augmentation had already been adopted. Enlarging the dataset could also be taken into consideration, which was meant for eliminating the imbalanced classes.[35]

Though not presented in this study, there are several ways that can also make improvements on the model's performance. Cross-validation is a data resembling method for assessing the generalization ability of predictive models and preventing overfitting.[36] Stratified K-fold cross-validation, where the dataset is divided into several folds and the class distribution in each fold is identical in the entire dataset, is useful in class-imbalanced classifications. One of its key advantages is to maintain class distribution during the training process, leading to a more robust evaluation of the model's performance.[37] Transfer learning refers to introducing a pre-trained CNN model as the starting point for a model on a new but similar task. This technique is known to enhance model performance and is reported to be used in computer-aided diagnosis.[10,34,38] Therefore, it was worthwhile to enlarge the dataset, and further, to subdivide different clinical subtypes in the training set as well as adopting transfer learning and cross-validation.

The statistical performances of the model reached a final AUC of 0.985 and the model certainly had the capacity to discern between benign and malignant-potential lesions, being suggestive of an excellent role in classification performance. On the whole, the proposed model laid the foundations of a novel tool that might be feasible of helping clinical differential diagnosis of oral mucosal lesions. In terms of enhancing clinical significance, enlarging the minority classes, incorporating more oral mucosal lesion diagnoses, employing transfer learning, and adopting cross-validation might be included in future works.

Declaration of competing interest

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

Acknowledgements

This study was supported by Summer Research Project Grant no. NCKUMCS2023035 from College of Medicine at National Cheng Kung University and National Science and Technology Council, Taiwan (Grant numbers: NSTC 111-2314-B-006 -037 -MY2).

References

1. Espinoza I, Rojas R, Aranda W, Gamonal J. Prevalence of oral mucosal lesions in elderly people in santiago, Chile. *J Oral Pathol Med* 2003;32:571—5.
2. Pentenero M, Broccoletti R, Carbone M, Conrotto D, Gandolfo S. The prevalence of oral mucosal lesions in adults from the turin area. *Oral Dis* 2008;14:356—66.
3. Mortazavi H, Safi Y, Baharvand M, Rahmani S. Diagnostic features of common oral ulcerative lesions: an updated decision tree. *Int J Dent* 2016;2016:7278925.
4. Minhas S, Sajjad A, Kashif M, Taj F, Al Waddani H, Khurshid Z. Oral ulcers presentation in systemic diseases: an update. *Open Access Maced J Med Sci* 2019;7:3341—7.
5. Scully C. Aphthous ulceration. *N Engl J Med* 2006;355:165—72.
6. Lavanya N, Jayanthi P, Rao UK, Ranganathan K. Oral lichen planus: an update on pathogenesis and treatment. *J Oral Maxillofac Pathol* 2011;15:127—32.
7. Field EA, Allan RB. Oral ulceration — aetiopathogenesis, clinical diagnosis and management in the gastrointestinal clinic. *Aliment Pharmacol Ther* 2003;18:949—62.
8. Porter SR, Leao JC. Review article: oral ulcers and its relevance to systemic disorders. *Aliment Pharmacol Ther* 2005;21:295—306.
9. Neville BW, Day TA. Oral cancer and precancerous lesions. *CA A Cancer J Clin* 2002;52:195—215.
10. Gao J, Jiang Q, Zhou B, Chen D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: an overview. *Math Biosci Eng* 2019;16:6536—61.
11. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115—8.
12. Cantu AG, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent* 2020;100:103425.
13. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018;77:106—11.
14. Mohammad-Rahimi H, Motamedian SR, Rohban MH, et al. Deep learning for caries detection: a systematic review. *J Dent* 2022;122:104115.
15. Ramezanzade S, Dascalu TL, Ibragimov B, Bakhshandeh A, Bjørndal L. Prediction of pulp exposure before caries excavation using artificial intelligence: deep learning-based image data versus standard dental radiographs. *J Dent* 2023;138:104732.
16. Schwendicke F, Elhennawy K, Paris S, Friebertshäuser P, Krois J. Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study. *J Dent* 2020;92:103260.
17. Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: a systematic review and meta-analysis. *J Dent* 2015;43:924—33.
18. Tiwari A, Gupta N, Singla D, et al. Artificial intelligence's use in the diagnosis of mouth ulcers: a systematic review. *Cureus* 2023;15:e45187.
19. Sharma D, Kudva V, Patil V, Kudva A, Bhat RS. A convolutional neural network based deep learning algorithm for identification of oral precancerous and cancerous lesion and differentiation from normal mucosa: a retrospective study. *Eng Sci* 2022;18:278—87.
20. Welikala RA, Remagnino P, Lim JH, et al. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access* 2020;8:132677—93.
21. Schwendicke F, Golla T, Dreher M, Krois J. Convolutional neural networks for dental image diagnostics: a scoping review. *J Dent* 2019;91:103226.
22. Neville BW, Damm DD, Allen CM, Chi AC. *Oral and maxillofacial pathology-e-book*, 5th ed. St. Louis: Elsevier, 2023:321—95.
23. Regezi JA, Sciubba J, Jordan RCK. *Oral pathology: clinical pathologic correlations*, 7th ed. St. Louis: Elsevier Health Sciences, 2016:38—109.
24. Herzum A, Burlando M, Cozzani E, Parodi A. The 30th birthday of chronic ulcerative stomatitis: a systematic review. *Int J Immunopathol Pharmacol* 2021;35:20587384211052437.

25. Bruch JM, Treister NS. *Clinical oral medicine and pathology*, 2nd ed. Cham: Springer, 2010:61—147.
26. Philipone E, Yoon AJ. *Oral pathology in the pediatric patient: a clinical guide to the diagnosis and treatment of mucosal lesions*. Cham: Springer, 2016:46—76.
27. Scully C, Flint S, Moos K, Bagan J. *Oral and maxillofacial diseases*, 4th ed. London: CRC Press, 2010:100—21.
28. Eversole LR. *Clinical outline of oral pathology: diagnosis and treatment*, 3rd ed. Hamilton: B.C: Decker, 2001:62—78.
29. Cawson RA, Eveson JW. *Oral pathology and diagnosis: color atlas with integrated text*. Philadelphia: Saunders, 1987:1—292.
30. Guo JB, Wang HL, Xue XS, Li MT, Ma ZX. Real-time classification on oral ulcer images with residual network and image enhancement. *IET Image Process* 2022;16:641—6.
31. Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding data augmentation for classification: when to warp?. In: *2016 international conference on digital image computing: techniques and applications (DICTA)*. IEEE: Gold Coast, 2016. Abstract 1609.
32. Zhou MM, Jie WP, Tang F, et al. Deep learning algorithms for classification and detection of recurrent aphthous ulcerations using oral clinical photographic images. *J Dent Sci* 2024;19:254—60.
33. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929—58.
34. Aljuaid H, Alturki N, Alsubaie N, Cavallaro L, Liotta A. Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Comput Methods Progr Biomed* 2022;223:106951.
35. Xue Y. An overview of overfitting and its solutions. *J Phys Conf Ser* 2019;1168:022022.
36. Berrar D. Cross-validation. In: Ranganathan S, Nakai K, Schonbach C, Gribskov M, eds. *Encyclopedia of bioinformatics and computational biology*, 1st ed. Amsterdam: Elsevier, 2019:542—5.
37. Szeghalmy S, Fazekas A. A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sens* 2023;23:2333.
38. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag* 2016;35:1285—98.