

DEEP LEARNING-BASED CLASSIFICATION OF ORAL LESIONS USING CONVOLUTIONAL NEURAL NETWORKS: A COMPARATIVE STUDY OF CNN, RESNET50, AND CONVNEXT-SMALL ARCHITECTURES

Mustafa Kareem Alioui^{1*}, Mohammed Kudhur Abduljalel², Mohammed Salah Ibrahim³

¹Oral and Maxillofacial Surgery Department, College of Dentistry, University of Anbar, Ramadi 31001, Anbar, Iraq. Email: mustafakareemalioui@gmail.com, Phone:07713087390, ORCID ID: <https://orcid.org/0000-0002-1589-4785>

²Oral and Maxillofacial Surgery Department, College of Dentistry, University of Anbar, Ramadi 31001, Anbar, Iraq.

³Department of Artificial Intelligence, College of Computer Science and Information Technology, University of Anbar, Ramadi 31001, Anbar, Iraq

ABSTRACT

Background: oral mucosal lesions include various types of mucosal abnormalities, such as oral cancer (OCA), oral potentially malignant disorders (OPMDs) and benign lesions. Proper classification at early stage is essential for precise clinical management; however, conventional diagnosis depends on subjective visual examination.

Objectives: To investigate and compare three deep learning models-a customized convolutional neural network (BasicCNN), ResNet50, and ConvNeXt-Small-for automated four-category classification of oral lesions from clinical images and to determine the impact of dataset augmentation, regularization, and statistical analysis on model performance.

Methods: A dataset contains 1,304 clinically validated oral images curated from Piyaathne et al.(2024) along with the addition of 117 external OCA images from parola et al. (2024) to address the significant OCA underrepresentation (4.8%). One co-author collected 20 image as external validation dataset and performed an excisional biopsy for five cases. Three models were trained: BasicCNN (from scratch), ResNet50 (pretrained on ImageNet), and ConvNeXt-Small (pretrained, AdamW, CosineAnnealingLR, dropout, label smoothing $\epsilon=0.1$, and 5-view test-time augmentation). Performance was assessed using macro F1, Cohen kappa, sensitivity, and specificity per class.

Results: ConvNeXt-Small achieved the highest accuracy of 71.7% (95% CI: 65.9–77.0%), macro F1 of 0.719, and OCA F1 of 0.91 outperforming ResNet50 (69.8%, F1=0.697) and BasicCNN (53.2%, F1=0.538). OCA achieved AUC=0.950 (macro-AUC=0.881), with a sensitivity of 91.4% and a specificity of 96.9%. Pretrained models were statistically significantly better than BasicCNN (McNemar $p < 0.001$); the ConvNeXt vs. ResNet50 difference was not significant ($p = 0.590$). Dataset enrichment was essential: OCA F1 improved from 0.00 to 0.63 following enrichment.

Conclusions: The ConvNeXt-Small architecture with progressive regularization and test-time augmentation outperformed other models in four-class oral lesion classification. Transfer learning played a vital role in boosting performance. The use of enriched data is crucial in order to detect OCA. These results can prove the possible application of deep learning in medical diagnostics.

KEYWORDS: ConvNeXt, convolutional neural networks; deep learning; transfer learning; oral cancer; oral lesion classification; test-time augmentation.

INTRODUCTION

Globally, oral cancer accounts for approximately 2% of all cancer cases and is associated with high mortality, with a 5-year survival rate of less than 60% [1]. Oral squamous cell carcinoma (OSCC) is the most prevalent malignancy of the oral cavity, representing the majority of oral cancer cases worldwide [2,3]. According to GLOBOCAN 2020, oral cancer ranked as the 16th most common cancer globally, with 377,713 new cases and 177,757 deaths [4]. Approximately two-thirds of newly diagnosed cases in 2020 occurred in Asian countries [5]. The worldwide prevalence of oral potentially malignant disorders (OPMDs) is estimated at 4.67% (95% CI: 3.62–5.84), with a higher prevalence in Asia (5.44%) [6].

In Iraq, 722 OSCC cases were reported between 2014 and 2018, with an estimated incidence rate of 0.4 per 100,000 population, predominantly affecting males and those over 40 years of age [7]. A crude mortality rate of 0.20 per 100,000 population was recorded in 2023 with 321 cases of oral cancer (0.73% of all malignancies) and 88 deaths [8]. The majority of patients are diagnosed at advanced stages, and survival rates are still between 50% and 55% over the past decades despite these epidemiological loads [9].

The gold standard is histopathological diagnosis based on incisional or excisional biopsy, but clinical examination is still the primary evaluation method (10). However, there are a number of limitations to this approach: while biopsy complications, including bleeding and infection, have been documented, the procedure is time-consuming, invasive, and prone to inter-observer variability [11–14].

Additionally, clinical judgment depends mainly on inspection and palpation, which are subjective and operator-dependent, decreasing diagnostic accuracy, particularly for early-stage OPMDs [10, 11, 15]. These restrictions in environments of limited resources highlight the demand for more consistent and objective techniques to enhance clinical decision-making [16, 17].

In recent years, artificial intelligence (AI) and deep learning approaches have attracted considerable attention in medical and dental research [18]. Several medical imaging techniques, including fundus, endoscopic, histopathological, CT, and MRI, have successfully employed deep learning [19]. AI-based systems for OPMD and oral cancer screening utilizing deep learning image identification have been studied in the field of oral health [20]. Conventional machine learning techniques depend on manual feature extraction, restricting their capability to capture complex visual patterns [21]. Convolutional neural networks (CNNs) address this by automatically extracting hierarchical image features [22]. For lesion localization in medical imaging, deep learning-based detection frameworks such as YOLO and Faster R-CNN have also been used [23, 24].

Deep learning for automated oral lesion detection has been reported by several studies. Fu et al. utilized 5,775 images and achieved an external validation AUC of 0.935 [25]. Warin et al. applied CNN to 700 images and reported a sensitivity of 82.14% [26]. Using Vision Transformers for binary OSCC classification, Flügge et al. achieved accuracy of 98.6% and AUC of 99% [27]. While using a CNN model, Lee et al. worked on 1,810 tongue images and reached an accuracy of 94.44% [28]. More recent multi-class work includes Tanriver et al. (F1=0.85, n=684) [29]; Talwar et al. (F1=0.84, n=2,178 smartphone images) [30]; Kouketsu et al. (OSCC sensitivity 93.9%, n=1,043) [31]; and Hsu et al. (F1=0.688, n=6,903, YOLOv7) [32].

Despite these advances, research on simultaneous four-class oral lesion classification—encompassing healthy mucosa, benign lesions, OPMDs, and OCA—with rigorous statistical comparison of multiple architectures remains limited. The present study addresses this gap by training and comparing three deep learning architectures for four-class oral lesion classification, with systematic evaluation of dataset enrichment, progressive regularization, and pairwise statistical significance testing of model differences.

METHODS

The flow diagram of the proposed system is shown in Fig. 1. The process starts with sources of datasets, then merges and enhances the main dataset for OCA enrichments. After that, the dataset is split into training and testing to be prepared for the next step. The next step is model training, and then finally model evaluation. The whole process is explained in the next sections.

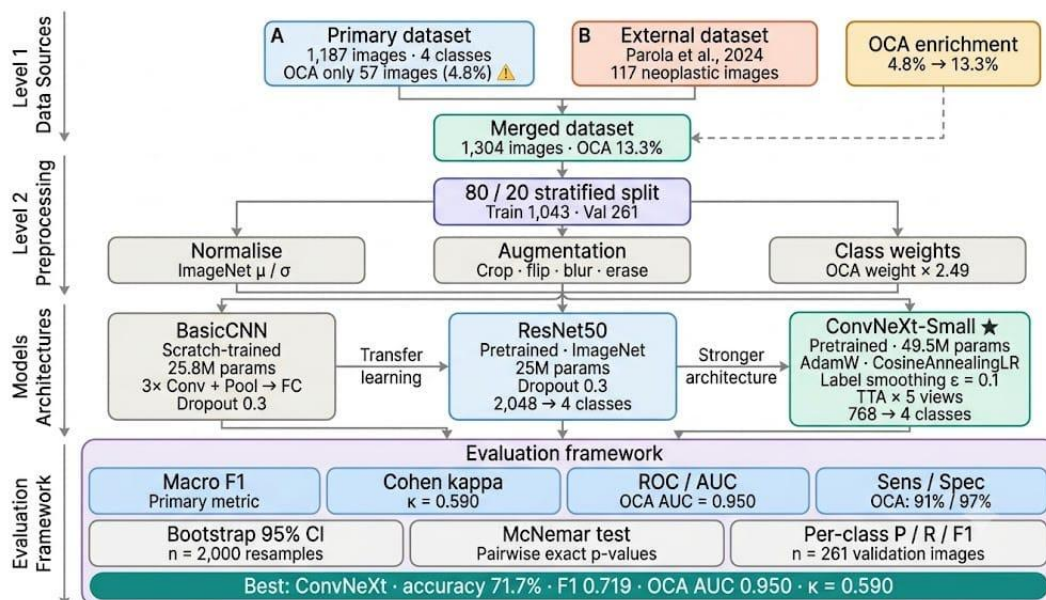


Fig. 1: The flow diagram of the proposed system

Study Design and Dataset

This retrospective diagnostic modeling study utilized clinical oral lesion images publicly available through Zenodo (DOI: 10.5281/zenodo.10664056) and described by Piyaathne et al. [33]. The dataset contains 3,000 clinical photographs acquired using mobile phones from 714 patients at the Teaching Hospital of Peradeniya, Sri Lanka (2022–2023), categorized into four classes: Benign, Healthy, OCA, and OPMD, by two dentists supervised by oral medicine specialists. Histopathological confirmation was available for 100% of OCA cases, 56.4% of OPMD cases, and 15.4% of benign lesions. After file-level validation excluding 15 corrupt files, the usable primary dataset comprised 1,187 images.

External Data Enrichment for OCA Class

The OCA class accounted for only 57 images (4.8%), yielding 11 OCA validation samples under an 80/20 stratified split. Preliminary experiments confirmed this was clinically insufficient: a Basic CNN trained on the original dataset achieved OCA F1=0.00 regardless of augmentation. To address this, supplementary images were obtained from an external oral lesion dataset developed by Parola et al. at the University of Pisa [34, 35] accessed under a research-use agreement. From 125 extracted neoplastic images, 117 usable files were retained after validation. The final merged dataset comprised 1,304 validated images as shown in table 2, with OCA representation increasing from 57 image (4.8%) to 174 image 13.3% following external enrichment. Independent external validation was performed on 20 histopathologically confirmed clinical photographs collected by the authors, with five cases undergoing excisional biopsy performed by one of the co-authors.

Data Preprocessing and Augmentation

All images were resized to 224×224 pixels in RGB format. For pretrained models, ImageNet normalization was applied (mean= [0.485, 0.456, 0.406], std= [0.229, 0.224, 0.225]). The dataset was partitioned 80/20 using stratified random sampling (random_state=42), yielding training (n=1,043) and validation (n=261) sets. BasicCNN and ResNet50 used basic augmentation (horizontal flip, ±10° rotation, colour jitter ±0.15). ConvNeXt-Small used a stronger pipeline: RandomResizedCrop (scale= [0.7,1.0]), vertical flip (p=0.2), rotation ±15°, colour jitter ±0.30, GaussianBlur (k=3), and RandomErasing (p=0.3). Validation images received deterministic preprocessing only.

Model Architectures

Three architectures were implemented and evaluated:

- (1)BasicCNN: a custom three-block convolutional network trained from scratch (32→ 64 →128 channels, Conv2D + ReLU + MaxPool2D per block), followed by Linear (100,352→256) + ReLU + Dropout (0.3) + Linear (256→4), ~25.8M parameters.
- (2)ResNet50: ImageNet-pretrained, with the original head replaced by Dropout (0.3) + Linear (2,048→4), fully fine-tuned [36].
- (3)ConvNeXt-Small: ImageNet-pretrained, ~49.5M parameters, with Dropout (0.3) + Linear (768→4) head [37]. Observed overfitting (>400× train/val loss divergence) motivated four regularisation improvements: AdamW with weight decay 5×10^{-3} ; learning rate 5×10^{-5} with CosineAnnealingLR; label smoothing ($\epsilon=0.1$); and early stopping (patience=7 on macro F1). At inference, five-view TTA was applied (original crop, horizontal flip, vertical flip, larger-scale crop, and corner crop), with softmax probabilities averaged before argmax prediction. Detailed hyperparameters are in Table 1.

Table 1. Training hyperparameters per model architecture.

Parameter	BasicCNN	ResNet50	ConvNext-small
Input size	224×224	224×w224	224×224
Learning rate	1×10^{-4}	1×10^{-4}	5×10^{-5}
Optimizer	Adam	Adam	AdamW
Weight decay	–	–	5×10^{-3}
Batch size	16	8	8
LR Scheduler	–	–	CosineAnnealingLR
Max epochs	50	50	30
Dropout	0.3 (FC layer)	0.3(head)	0.3 (head)
Test-Time augmentation	No	No	Yes (×5 views)
Early stopping	No	No	Yes (patience=7)
Loss function	Weighted CE	Weighted CE	Weighted CE + label smoothing ($\epsilon=0.1$)

Statistical Analysis

Performance was assessed using macro F1 (primary metric; equal weight to all classes), per-class precision/recall/F1, overall accuracy, and Cohen kappa (κ , Landis and Koch criteria). Sensitivity and specificity were computed per class in a one-vs-rest manner. ROC curves and AUC were computed for the ConvNeXt-Small final model using softmax probability outputs. Bootstrap 95% CIs were computed using the non-parametric percentile method (n=2,000 resamples, seed=42). The McNemar paired test was applied to all enriched-dataset model pairs (R3–R7) using actual per-sample prediction vectors, with an exact binomial test when discordant pairs ≤ 25 . All analyses used Python (scikit-learn, statsmodels). Significance: $\alpha=0.05$.

RESULTS

The final merged dataset comprised 1,304 validated images (Table 2). OCA representation increased from 57 (4.8%) to 174 images (13.3%) following external enrichment.

Table 2. Dataset class distribution and train/validation partition after OCA enrichment.

Category	Source	Total	Train (80%)	Val (20%)
Healthy	Primary dataset	459	367	92
OPMDs	Primary dataset	333	267	66
Benign	Primary dataset	338	270	68
OCA	Primary+External (parola et al.)		139	35
Total	–	1304	1043	261

BasicCNN trained from scratch on the enriched dataset achieved an accuracy of 53.2% (95% CI: 47.1–59.4%), a macro of F1 0.538, and $\kappa=0.362$ (fair agreement). OCA detection was achieved (F1=0.63), a critical improvement over the original dataset where OCA F1=0.00 (only 11 validation samples). ResNet50 achieved substantially higher performance: accuracy 69.8% (95% CI: 64.0–75.1%), macro F1 0.697, $\kappa=0.579$ (Moderate agreement), and OCA sensitivity 88.6%. ConvNeXt-Small with advanced regularization and TTA achieved the highest performance: accuracy 71.7% (95% CI: 65.9–77.0%), macro F1 0.719, and OCA F1 = 0.91. Detailed per-class metrics are in Table 3.

Training curves for the final ConvNeXt model are presented in Fig. 2, showing controlled train/validation loss convergence with label smoothing preventing loss collapse.

Confusion matrices for all three models are shown in Fig. 3.

Table 3. Per-class classification metrics for all three models on the validation set (n=261). Values for ConvNeXt-Small (v3+TTA) are derived from actual Kaggle prediction outputs (Figure 2). Bootstrap 95% CI computed with n=2,000 resamples. κ interpreted per Landis and Koch criteria.

Model	Class	Precision	Recall	F1-score	n	Accuracy (95% CI)
BasicCNN	Benign	0.42	0.40	0.41	68	0.532
	OCA	0.63	0.63	0.63	35	Macro F1: 0.538
	Healthy	0.64	0.53	0.58	92	[0.471–0.594]
	OPMDs	0.48	0.62	0.54	66	$\kappa = 0.362$ (Fair)
ResNet50	Benign	0.62	0.65	0.63	68	0.698
	OCA	0.89	0.69	0.77	35	Macro F1: 0.697
	Healthy	0.74	0.82	0.77	92	[0.640–0.751]

	OPMDs	0.64	0.59	0.61	66	$\kappa = 0.579$ (Moderate)
ConvNeXt-small	Benign	0.61	0.60	0.60	68	0.717
	OCA	0.91	0.91	0.91	35	Macro F1: 0.719
(V3+TTA)	Healthy	0.72	0.67	0.70	92	[0.659–0.770]
	OPMDs	0.70	0.71	0.71	66	$\kappa = 0.590$
						Moderate–Substantia

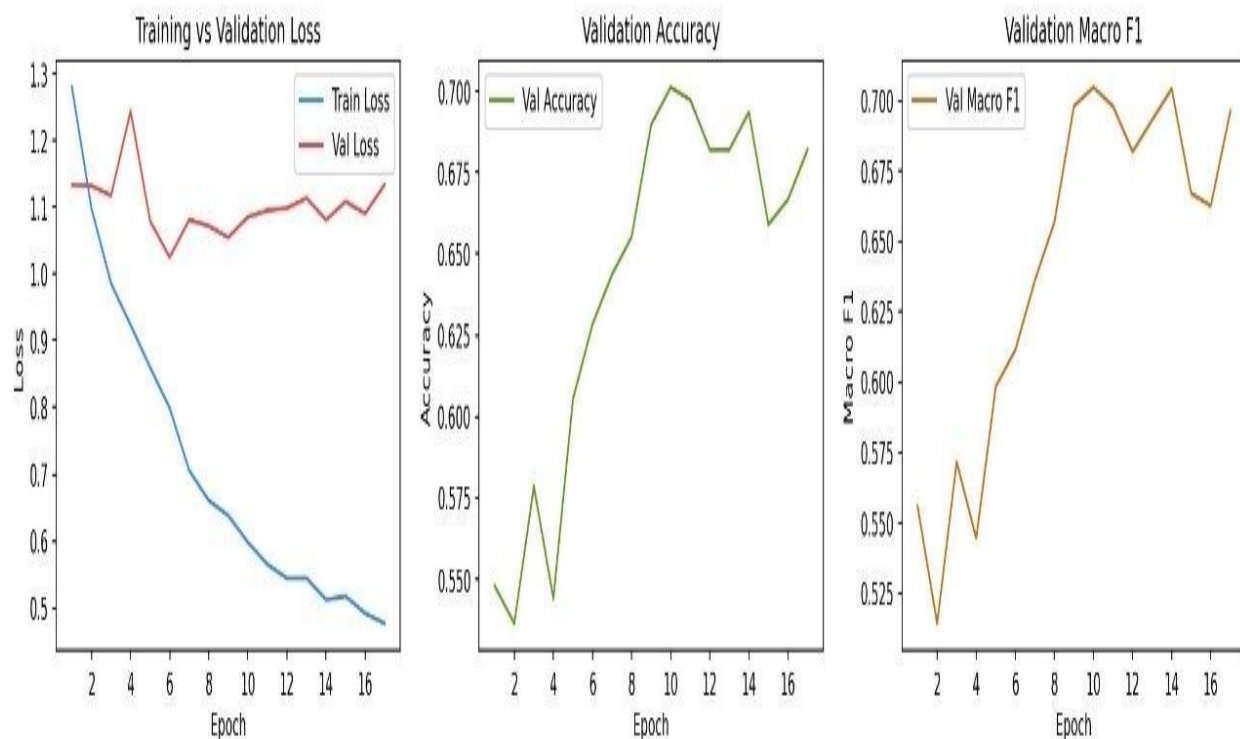


Fig. 2: Training and validation loss (left) and validation accuracy (centre) and macro F1 (right) curves for the ConvNeXt-Small v3 + TTA model over 17 epochs. Label smoothing ($\epsilon=0.1$) prevents training loss from collapsing to near zero, maintaining a controlled train/validation gap throughout training. Early stopping was triggered at epoch 17.

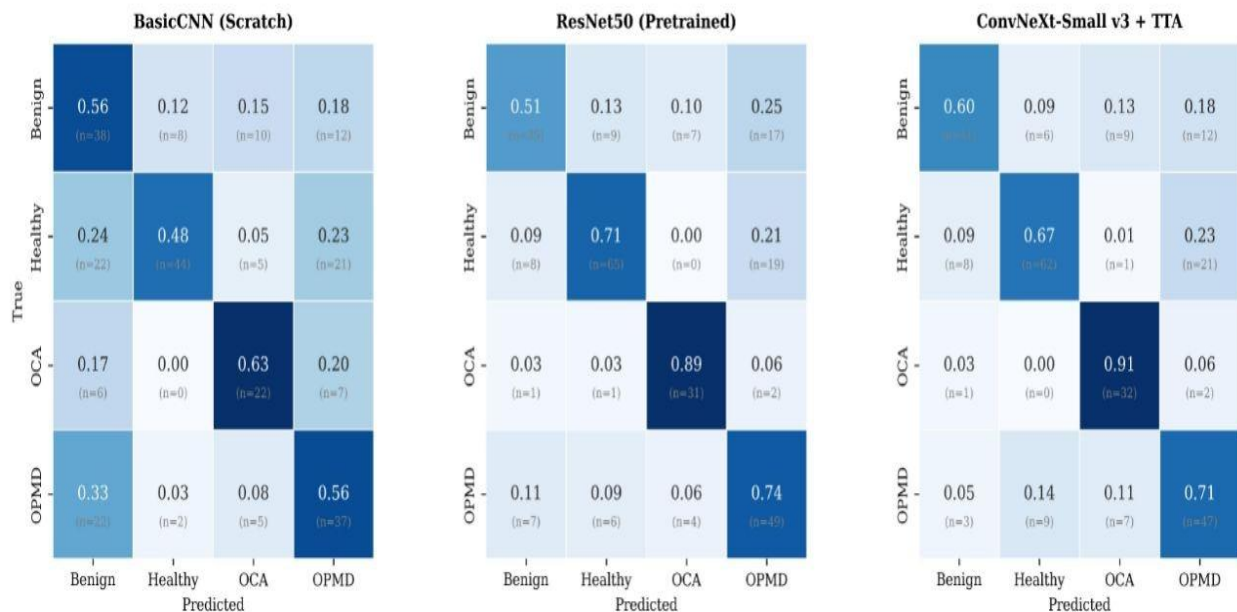


Figure 3. Normalized confusion matrices (row-normalized recall) for all three models on the validation set (n=261). Values represent the proportion of each true class predicted as each category; raw counts in parentheses. OCA recall improves from 0.63 (BasicCNN) to 0.89 (ResNet50) to 0.91 (ConvNeXt-Small v3+TTA). The Benign–OPMD boundary is the most consistently confusing pair across all three models.

For the ConvNeXt-Small final model, one-vs-rest sensitivity and specificity were Benign (0.60, 0.86), Healthy (0.67, 0.90), OCA (0.914, 0.969), and OPMD (0.71, 0.89). ROC/AUC analysis (Figure 4) yielded: Benign=0.825, Healthy=0.908, OCA=0.950, OPMD=0.831, macro average=0.881.

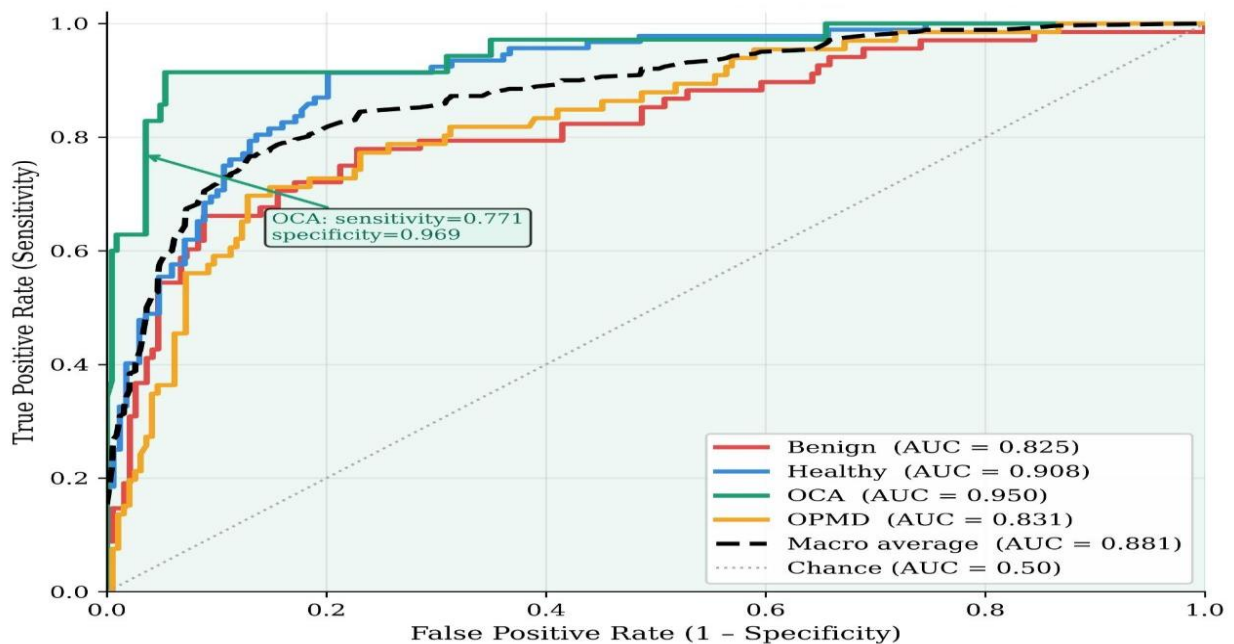


Figure 4. ROC curves for the ConvNeXt-Small v3+TTA final model using a one-vs-rest strategy (each class vs all others), computed from softmax probability outputs. OCA achieves the highest AUC (0.950), indicating excellent discrimination of malignant from non-malignant cases across all decision thresholds.

McNemar pairwise testing (Table 4) confirmed that both pretrained models were statistically significantly better than enriched BasicCNN (ResNet50 vs. BasicCNN: $n_{01}=70$, $n_{10}=27$, $p<0.001$; ConvNeXt vs. BasicCNN: $n_{01}=74$, $n_{10}=26$, $p<0.001$). The difference between ResNet50 and ConvNeXt-Small was not statistically

significant ($n_{01}=30$, $n_{10}=25$, $p=0.590$), indicating that pretrained transfer learning—rather than specific architecture or regularization—was the dominant driver of improvement.

Table 4. Summary statistical analysis — accuracy, 95% bootstrap CI, Cohen kappa, OCA sensitivity and specificity, ROC/AUC (from Figure 3), and McNemar pairwise test results. AUC values derived from actual softmax probability outputs of the final ConvNeXt-Small checkpoint.

Metric	BasicCNN	ResNet50	ConvNeXt-Small (Final)	P-value
Accuracy	0.532	0.698	0.717	—
95%CI (Accuracy)	[0.471–0.594]	[0.640–0.751]	[0.659–0.770]	—
Macro F1	0.538	0.697	0.719	—
95% CI Sw,(Macro F1)	[0.475–0.597]	[0.639–0.755]	[0.662–0.774]	—
Cohen kappa (k)	0.362 (Fair)	0.579 (Moderate)	0.590 (Moderate)	—
OCA Sensitivity	0.629	0.886	0.914	—
OCA specificity	0.943	0.987	0.969	—
OCA AUC	—	—	0.950	—
Macro AUC	—	—	0.881	—
Vs.BasicCNN (McNemar)	—	$p<0.001^{***}$	$p<0.001^{***}$	—
Vs.ResNet 50 (McNemar)	—	—	$p=0.590$ (n.s.)	—

DISCUSSION

ConvNeXt-Small with progressive regularization and test-time augmentation achieved the highest performance in four-class oral lesion classification (accuracy 71.7%, macro F1 0.719, OCA AUC 0.950). The most important statistical finding, however, is that this advantage over ResNet50 was not significant (McNemar $p = 0.590$), while both pretrained models significantly outperformed the scratch-trained CNN ($p < 0.001$). This confirms that pretrained ImageNet transfer learning is the essential enabling factor for this dataset scale, rather than any architecture-specific feature.

In contrast to the original dataset, where OCA was completely undetectable ($F1=0.00$), the OCA class achieved the highest AUC (0.950) and F1 (0.91) among all classes in the final model.

This shift was caused mostly by dataset enlargement (4.8%→13.3% OCA representation), proving that minority-class representation is the most crucial requirement for clinically safe oral cancer classification. Even the robust architecture would consistently fail to identify malignancy in the absence of data enrichment.

Although statistically non-significant in this situation, the superiority of ConvNeXt-Small over ResNet architectures may be explained by its 7×7 depthwise separable convolutions permitting larger receptive fields, LayerNorm for training stability, and GELU activations [38]. These adjustments improve fine-grained texture discrimination that is required for classifying the oral lesions. However, it should not be exaggerated because the statistical non-significance indicates that these architectural advantages were not reflected by a noticeable accuracy enhancement result on this 261-sample validation set.

In all three models, benign lesions persisted as the most challenging group, highlighting the real clinical difficulty of differentiating benign reactive lesions from early OPMDs. White patches visually identical to leukoplakia may appear in reactive lesions such as frictional keratosis. Previous studies have similarly observed lower performance at the benign-OPMD boundary [31, 32], which suggests it is an intrinsic feature of visual categorization rather than a model limitation.

The OCA specificity of 96.9% is clinically suitable for a monitoring tool, producing false alarms in only 3.1% of non-cancer conditions and thereby reducing excessive referral cases. By having one undetected case in eleven cancer cases, the 91.4% OCA sensitivity remains a barrier for reliable clinical applications. The high OCA AUC (0.950) indicates that threshold scaling may be able to further enhance sensitivity while maintaining satisfactory specificity without the need for retraining.

While comparing previous studies, methodological variations should be considered. Studies demonstrating accuracy >95% [27, 31] usually used binary formulations (OSCC vs. normal), more uniform datasets, or more standardized imaging conditions; therefore, making direct comparison is improper. In this study, the four-class simultaneous classification is inherently more difficult, and the using bootstrap CIs and McNemar testing yields a more cautious and consistent performance estimate than using point-estimate accuracy only.

Study Limitations

First, the generalizability of point estimates is restricted since evaluation relied on a single 261-sample validation split; bootstrap CIs overcome this issue.

Second, there was inadequate histopathological confirmation for OPMD (56.4%) and benign (15.4%) categories, which could lead to label noise. Third, the domain shift was possibly due to cross-dataset domain discrepancies between the Sri Lankan primary images and the Italian external OCA. Fourth, clinical interpretability was restricted as the model has no explainability output (e.g., Grad-CAM). Fifth, an inadequate external validation sample (only 20 cases) for conclusive findings regarding generalizability.

Future research should use multi-institutional datasets, incorporate explainable AI techniques, and prospective validation.

CONCLUSION

The most accurate results in four-class oral lesion classification were recorded by ConvNeXt-Small with progressive regularization (AdamW, CosineAnnealingLR, label smoothing $\epsilon=0.1$) and five-view test-time augmentation: accuracy 71.7%, macro F1 0.719, OCA F1 0.91, and OCA AUC 0.950.

Pretrained ImageNet transfer learning was the primary cause for the statistically significant improvement over the scratch-trained CNN ($p < 0.001$).

In order to raise OCA representation from 4.8% to 13.3% and optimize OCA detection from F1=0.00 to 0.91, data enrichment was crucial. Upon achieving the requirement of multi-institutional validation and explainable AI incorporation, these results confirm the potential of pretrained convolutional architectures as adjunctive screening tools for oral lesion triage.

Statements and Declarations

Acknowledgements:

The authors acknowledge Piyarathne et al. for the public availability of dataset (DOI: 10.5281/zenodo.10664056), and Marco Parola and colleagues at the University of Pisa for providing access to the external oral lesion dataset under a research-use agreement.

They also acknowledge Dr.Muhassad al-mudhafar, Dr.Auday Mahmood, Dr.Ahmed jassam, Dr.Loui abdulaziz, Dr.Mustafa abdulhussain, Dr.Hiba , Dr.mariam and Dr.Ali hussain for their help and support throught the study.

FUNDING:

The authors did not receive any external funding for this study.

Competing Interests:

The authors declared no conflict of interest.

Authors contributions:

•Mustafa Kareem alioui (the corresponding author):

Literature research, data acquisition, clinical contribution, manuscript preparation, editing and review,

•**Muhammad khudhur abdaljalel:**

Conceptualization, study design, Definition of intellectual content, manuscript editing and review

•**Muhammad Salah Ibrahim:**

Computational contribution, data analysis, manuscript preparation, editing and review.

Data availability statement:

The primary dataset is publicly available at: <https://zenodo.org/records/10664056>. The supplementary OCA images were obtained from Parola et al. and are accessible at: <https://drive.google.com/drive/folders/1USSphFVqYuA0H9vZbpB-4K3tgIcNyI-d>. The model training code and statistical analysis code are available upon reasonable request from the corresponding author.

Ethical statement:

For the public and supplementary oral cancer datasets, data were obtained and utilized under official research-use agreements from the original authors, and all data were completely anonymized. For the independent external validation dataset, it was collected according to the ethical standards of the institutional research committee.

Ethical approval:

For the primary training dataset, ethical approval was provided by the institutional ethics review committee at the Faculty of Dental Sciences, University of Peradeniya, Sri Lanka (ERC/FDS/2022/03), with written informed consent obtained from all participants. The external supplementary dataset was used under a research-use agreement with the authors (Parola et al.), with terms permitting non-commercial research use with appropriate citation. For the independent external validation dataset comprising 20 author-collected clinical images, ethical approval was granted by the Ethics Committee at the University of Anbar (Ref. No. 135; dated 12/11/2025). The study was conducted in accordance with the principles of the Declaration of Helsinki (2013).

Consent to participate

For the independent external validation dataset, written informed consent to participate was obtained from all participants prior to data collection. For the public datasets, data were completely anonymized, and the requirement for participant consent was covered under the original institutional approvals.

Consent for publication

Since this manuscript includes clinical author-collected images, written informed consent for publication was obtained from all participants prior to data collection. All images were completely anonymized to ensure that no identifiable participant characteristics are visible.

REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–263.
2. Lingen MW, Abt E, Agrawal N, Bhatt AA, Cohen E, Czerninski R, et al. Evidence-based clinical practice guidelines for the evaluation of potentially malignant disorders in the oral cavity. *J Am Dent Assoc*. 2017;148(10):712–727.e10.
3. Glick M, Greenberg MS, Lockhart PB, Challacombe SJ, editors. *Burket oral medicine*. 13th ed. Hoboken (NJ): John Wiley and Sons; 2021.
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
5. Rai P, Ng A, Intekhab I, Sim YF, Lai CWM, Loh JSP. Oral cancer in Asia: A systematic review. *Adv Oral Maxillofac Surg*. 2022; 8:100366.
6. Lonni N, Rosa TS, Pereira CK, Melo G, Warnakulasuriya S, Guerra ENS, et al. Global prevalence of oral potentially malignant disorders: An updated systematic review and meta-analysis. *J Oral Pathol Med*. 2026. doi:10.1111/jop.70146.
7. Alshami ML, Al-Maliky MA, Alsagban AA, Alshaeli AJ. Epidemiology and incidence of oral squamous cell carcinoma in the Iraqi population over 5 years (2014–2018). *Health Sci Rep*. 2023;6(4): e1205.
8. Iraqi Cancer Board. *Cancer registry of Iraq annual report 2023*. Baghdad: Iraqi Ministry of Health; 2023.
9. Ren ZH, Hu CY, HR, Li YJ, Lyu J. Global and regional burdens of oral cancer from 1990 to 2017: Results from the global burden of disease study. *Cancer Commun (Lond)*. 2020;40(2–3):81–92.
10. Warnakulasuriya S, Fennell N, Diz P, Seoane J, Rapidis A. A review of oral cancer and precancer screening programmes in Europe. *J Oral Pathol Med*. 2015;44(8):559–570.
11. Miloro M, Ghali GE, Larsen PE, Waite PD, editors. *Peterson principles of oral and maxillofacial surgery*. 4th ed. Shelton (CT): Elsevier; 2022.

12. Essat M, Cooper K, Bessey A, Tonnah N, Sutton A, Wood H, et al. Diagnostic accuracy of conventional oral examination for detecting oral cavity cancer and potentially malignant disorders: A systematic review and meta-analysis. *Head Neck*. 2022;44(4):998–1013.
13. Abbey LM, Kaugars GE, Gunsolley JC, Burns JC, Page DG, Svirsky JA, et al. Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 1995;80(2):188–191.
14. Eversole LR. Dysplasia of the upper aerodigestive tract squamous epithelium. *Head Neck Pathol*. 2009;3(1):63–68.
15. Ilhan B, Lin K, Gunes P, Wilder-Smith P. Improving oral cancer outcomes with imaging and artificial intelligence. *J Dent Res*. 2020;99(3):241–248.
16. Warnakulasuriya S, Kujan O, Aguirre-Urizar JM, Bagan JV, Gonzalez-Moles MA, Kerr AR, et al. Oral potentially malignant disorders: A consensus report from an international seminar convened by the WHO Collaborating Centre for Oral Cancer. *Oral Dis*. 2021;27(8):1862–1880.
17. Coelho KR. Challenges of the oral cancer burden in India. *J Cancer Epidemiol*. 2012; 2012:701932.
18. Maria-Garcia-Pola M, Pons-Fuster E, Suarez-Fernandez C, Lopez-Jornet P, Calvo-Silvosa A, Ayuso-Montero R, et al. Role of artificial intelligence in the early diagnosis of oral cancer: A scoping review. *Cancers (Basel)* 2021;13(18):4600.
19. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med*. 2020;8(11):713.
20. Mirfendereski P, Li GY, Pearson AT, Villa A, Bhatt A, Bhatt DL, et al. Artificial intelligence and the diagnosis of oral cavity cancer and oral potentially malignant disorders from clinical photographs: A narrative review. *Front Oral Health*. 2025; 6:1569567.
21. Pham TD, Teh MT, Chatzopoulou D, Coulthard P, Bell D. Artificial intelligence in head and neck cancer: Innovations, applications, and future directions. *Curr Oncol*. 2024;31(9):5255–5290.
22. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. *Adv Exp Med Biol*. 2020; 1213:3–21.
23. Ding L, Liu G, Zhang X, Du F, Li L, Huang J, et al. A deep learning nomogram kit for predicting metastatic lymph nodes in rectal cancer. *Cancer Med*. 2020;9(23):8809–8820.
24. Thanathornwong B, Suebnukarn S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs using faster regional convolutional neural networks. *Imaging Sci Dent*. 2020;50(2):169–174.
25. Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine*. 2020; 27:100558.
26. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J Oral Pathol Med*. 2021;50(9):911–918.
27. Flugge T, Gaudin R, Sabatakakis A, Naros A, Gutknecht MH, Poxleitner P, et al. Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. *Sci Rep*. 2023; 13:2296.
28. Lee SJ, Oh HJ, Son YD, Kim JH, Kwon IJ, Kim B, et al. Enhancing deep learning classification performance of tongue lesions in imbalanced data. *BMC Oral Health*. 2024; 24:161.
29. Tanriver G, Tekkesin MS, Ergen O. Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders. *Cancers (Basel)* 2021;13(11):2766.
30. Talwar V, Singh P, Mukhia N, Shetty A, Birur P, Desai KM, et al. AI-assisted screening of oral potentially malignant disorders using smartphone-based photographic images. *Cancers (Basel)* 2023;15(16):4120.
31. Kouketsu A, Doi C, Tanaka H, Takahashi A, Michi Y, Tsukinoki K, et al. Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. *Head Neck*. 2024;46(9):2253–2260.
32. Hsu Y, Chou CY, Huang YC, Liu HC, Lee MT, Pan PC, et al. Oral mucosal lesions triage via YOLOv7 models. *J Formos Med Assoc*. 2024. doi: 10.1016/j.jfma.2024.07.010.
33. Piyaathne NS, Liyanage SN, Rasnayaka RMSGK, Hettiarachchi PVKS, Devindi GAI, Francis FBAH, et al. A comprehensive dataset of annotated oral cavity images for diagnosis of oral cancer and oral potentially malignant disorders. *Oral Oncol*. 2024; 156:106946.
34. Parola M, Cimino MGCA, Cantini I, La Mantia G, Campisi G, Di Fede O. Towards explainable oral cancer recognition: Screening on imperfect images via Informed Deep Learning and Case-Based Reasoning. *Comput Med Imaging Graph*. 2024; 117:102433.
35. Parola M, Cimino MGCA, Cantini I, La Mantia G, Campisi G, Di Fede O. Oral cancer recognition on photographic images via deep learning semantic segmentation. In: 2025 IEEE Symposium on Computational Intelligence in Health and Medicine Companion (CIHM Companion); 2025. p. 1–5.
36. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE CVPR*. 2016:770–778.
37. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: *Proceedings of the IEEE/CVF CVPR*. 2022:11976–11986.
38. Song B, Sunny S, Li S, Gurushanth K, Mendonca P, Mukhia N, et al. Bayesian deep learning for reliable oral cancer image classification. *Biomed Opt Express*. 2021;12(10):6422–6430.