CONVOLUTIONAL NEURAL NETWORK-BASED BINARY CLASSIFICATION FOR ORAL EPITHELIAL DYSPLASIA

Muhammad Talha Hassan[1], Majid Jehangir[2], Fatma E.A. Hassanein[3], Lubbabah Ibrahim[4], Madieha Ambreen[5], Muhammad Zohaib Hassan Shah[*6]

[1]*Department of Oral and Maxillofacial Pathology, Superior University, Lahore, Pakistan.*
[2]*Department of Oral and Maxillofacial Pathology, Superior University, Lahore, Pakistan.*
[3]*Faculty of Dentistry, King Salman International University, El-tor, Egypt.*
[4]*Rashid Latif Medical College, Lahore, Pakistan*
[*5]*Department of Biology, Minhaj University, Lahore, Pakistan.*
[*6]*International Center for Primate Brain Research, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China*

[*6]zohaib.hassan@icpbr.ac.cn

Corresponding Author: *
Muhammad Zohaib Hassan Shah
DOI: https://doi.org/10.5281/zenodo.18035748

Received	Accepted	Published
19 October 2025	30 November 2025	19 December 2025

ABSTRACT

Objective: *This study aims to advance digital pathology in oral diagnostics by developing an artificial intelligence (AI) tool capable of detecting oral epithelial dysplasia (OED) and identifying its key histopathological features.*
Methods: *We employed a retrospective analytical design using archived hematoxylin and eosin–stained histopathological slides from previously diagnosed cases obtained from a public hospital. Slides were digitized and used to train and evaluate a convolutional neural network (CNN) for binary classification (OED vs. non-OED). Model performance was assessed using accuracy, sensitivity, specificity, precision, F1-score, and Cohen's kappa (κ) to measure agreement with expert histopathological diagnoses.*
Results: *CNN achieved an overall diagnostic accuracy of 88.1%, with a sensitivity of 87.5% and specificity of 88.7%. Cohen's κ of 0.76 indicated substantial inter-rater agreement between the AI predictions and the reference diagnoses. We have used CNN-based model, which demonstrates the strong potential for accurate, automated detection of OED in histopathological images. By enhancing diagnostic consistency and mitigating interobserver variability, this AI-driven approach may contribute to earlier detection, improved risk stratification, and better clinical outcomes in oral pathology.*
Keywords: *Artificial intelligence; Oral epithelial dysplasia; Convolutional neural network; Digital pathology; Detection and diagnosis*

INTRODUCTION

Artificial intelligence (AI) has revolutionized digital pathology by enabling automated, precise, and reproducible tissue assessment (Shephard et al., 2024). In the context of oral epithelial dysplasia (OED), a key precancerous lesion within the spectrum of oral potentially malignant disorders (OPMDs). This technological advancement holds promise. OED, characterized by cytologic and architectural abnormalities in stratified squamous epithelium, represents an early step in the multistep progression toward oral squamous cell carcinoma (OSCC) (Warin et al., 2024).

Histopathological evaluation remains the gold standard for diagnosing and grading OED, yet it is inherently subjective. Conventional grading systems, despite recent adoption of binary classifications to improve consistency (Xu et al., 2025), still suffer from significant inter- and intra-observer variability, limiting their predictive accuracy for malignant transformation (Peng et al., 2024).

Beyond traditional diagnostic criteria, emerging evidence suggests that quantitative morphometric features—such as cellular density, nuclear size and shape, nuclear-to-cytoplasmic ratio, and epithelial thickness may offer valuable prognostic insights. However, these parameters are rarely standardized or systematically quantified in routine practice due to the challenges of manual visual assessment under light microscopy (Rapado-González et al., 2024). Consequently, histopathological grading alone may be insufficient for guiding optimal clinical management.

Recent machine learning-based studies have demonstrated that methylation signatures can discriminate against oral cancer cases with high accuracy, offering a minimally invasive pathway for early detection (Sahoo et al., 2024). Parallel to molecular biomarker development, rapid progress in AI has transformed image-based diagnostic research. Deep learning models have shown considerable ability to extract complex texture and morphological features from clinical images, enhancing the precision of early cancer detection (Parola et al. 2024). A recent comprehensive review has further highlighted the expanding role of AI-enabled photographic analysis for screening and early diagnosis of oral squamous cell carcinoma (Bowyer, 2000). The integration of AI into dysplasia prediction has advanced substantially with the development of deep learning systems capable of analyzing routine oral photographs. Artificial intelligence model review images that outperforming most human clinicians reviewing the same images (Sahoo et al., 2024).

To address these limitations, we designed a two-part study: first, to develop a digital quantitative framework for evaluating key morphological features in OED; and second, to correlate these features with established dysplasia grades and normal epithelium to identify objective biomarkers of progression risk. While convolutional neural networks (CNNs) (Camalan et al., 2021) have demonstrated success in identifying subtle architectural patterns missed by human observers—and models like DeepLabv3 (Yurtkulu et al. 2019) and U-Net (Araujo et al., 2023) have shown efficacy in complex histological analysis (Abdul et al., 2024), their computational demands often hinder deployment in resource-limited clinical settings.

In contrast, our study leverages MobileNetV2 (Sandler et al., 2018), a lightweight, computationally efficient CNN architecture. Its optimized feature extraction pipeline and reduced parameter count make it uniquely suited for real-world implementation, particularly in academic and clinical environments where hardware constraints exist. We hypothesize that MobileNetV2 (Sandler et al., 2018) can reliably quantify morphological features associated with dysplastic severity, thereby enhancing diagnostic objectivity and supporting more accurate risk stratification in OED.

Materials and Methodology

This section outlines the materials and methodology employed for the binary classification of oral epithelial lesions. The deep learning model, namely, MobileNetV2 (Sandler et al., 2018), was evaluated using a curated histopathological dataset comprising 101 whole-slide images of hematoxylin- and eosin-stained tissue sections, categorized into two diagnostic groups: normal oral epithelium (n = 48) and oral epithelial dysplasia (n = 53). More details can be seen in the following subsections.

2.1. Data Collection

This analytical observational study was conducted at the Department of Histopathology, FMH College of Medicine and Dentistry, Lahore, Pakistan. Ethical approval was obtained prior to commencement of data collection.

Data were retrospectively collected from archived histopathological specimens spanning a five-year period (2020–2025). All materials — including hematoxylin- and eosin-stained glass slides and corresponding formalin-fixed paraffin-embedded tissue blocks — were retrieved from the department's archives, which adhere to international standards for biospecimen storage and preservation.

Each case included in the study had been initially diagnosed by a board-certified pathologist during routine clinical practice. For quality assurance and diagnostic validation, all selected slides underwent independent re-evaluation by a second experienced pathologist prior to inclusion. Discrepancies were resolved through consensus review.

A total of 101 tissue sections — comprising both normal oral epithelium and dysplastic lesions - were ultimately included in the analysis. Representative regions of interest (ROIs) were identified based on established histopathological criteria: in dysplastic epithelium, the basal and parabasal layers of the lower third of the epithelium were prioritized due to their consistent demonstration of diagnostic morphological alterations (Ghezloo, 2024). This targeted approach ensured focused evaluation of key diagnostic features while optimizing image acquisition efficiency and minimizing computational load without compromising accuracy.

Images were captured using a high-resolution Leica digital microscope camera equipped with 20× and 40× objectives. Only areas most representative of the lesion's histopathological characteristics were selected for imaging.

The cohort was divided into two groups:

Dysplastic Group (n=53): Specimens exhibiting characteristic cytological and architectural abnormalities of OED, including nuclear pleomorphism, hyperchromatism, increased nuclear-to-cytoplasmic ratio, prominent nucleoli, and frequent mitotic figures — some atypical in morphology (Hadilou et al., 2025).

Normal Group (n=48): Specimens displaying orderly epithelial maturation, well-preserved basal cell polarity, uniform nuclear size and chromatin pattern, absence of suprabasal mitoses, and normal stratification without evidence of abnormal keratinization or cellular disarray.

2.2 MobileNetV2-based deep learning architecture

MobileNetV2 (Sandler et al., 2018) is a highly efficient convolutional neural network architecture designed for mobile and edge devices, leveraging depth wise separable convolutions and linear bottlenecks to significantly reduce computational cost while preserving representational power (Sandler et al., 2018). Its core building block is the inverted residual structure, which first expands the channel dimension using a 1×1 convolution, applies lightweight depth wise convolution in the expanded space, and then projects back to a lower-dimensional space, retaining rich feature information with minimal parameters. Originally pre-trained on ImageNet for large-scale image classification, MobileNetV2 (Sandler et al., 2018) offers robust, transferable feature extractors ideal for medical imaging tasks with limited data. A typical pipeline data analysis using MobileNetV2 is shown in Fig. 1.

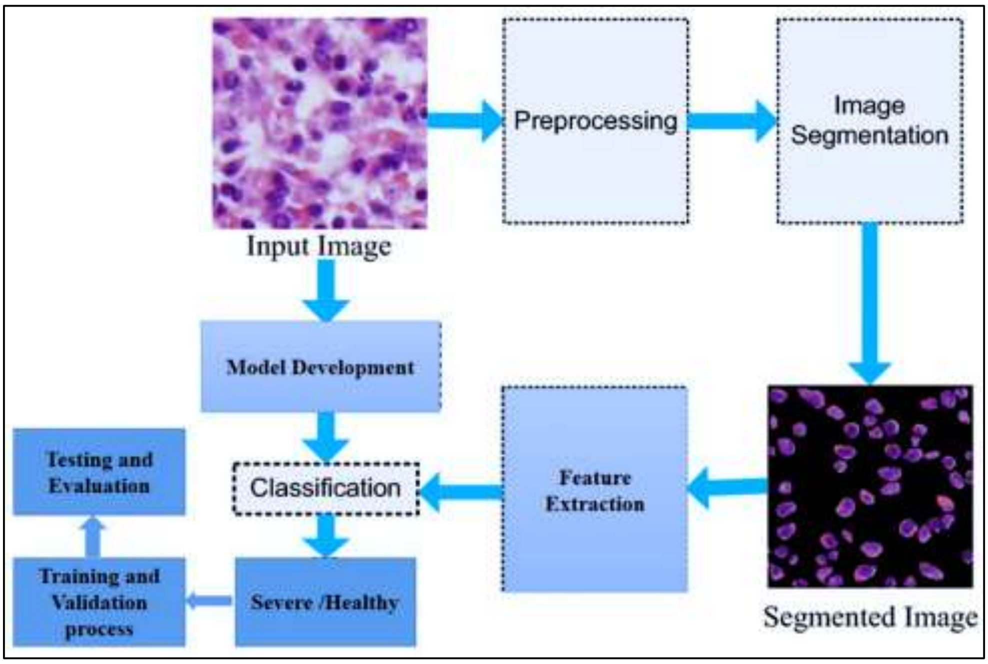


Figure 1: Data analysis flowchart based on Mobile NetV2 architecture.

To adapt MobileNetV2 (Sandler et al., 2018) for binary classification in this study, the original 1,000-class output layer was removed and replaced with a custom head consisting of a global average pooling layer followed by a single dense neuron with a sigmoid activation function. Figure 1 shows the typical architecture of MobileNetV2 (Sandler et al., 2018) adopted in this study.

This outputs a probability score between 0 and 1, corresponding to the likelihood of a specimen belonging to one of two diagnostic categories (e.g., dysplastic vs. normal oral epithelium). By preserving the pre-trained convolutional backbone and only fine-tuning the final layers, the model achieves high diagnostic accuracy with reduced memory footprint and faster inference—making it particularly well-suited for deployment in resource-constrained digital pathology workflows. The computational analysis of the data is done with the help of QuPath (Bankhead, 2017).

2.3 Statistical Analysis

The diagnostic performance of the proposed model was rigorously evaluated using established classification metrics: accuracy, sensitivity (recall), specificity, precision (positive predictive value), negative predictive value, and F1-score (Alotaibi et al., 2024). Each metric was computed on the held-out test set and reported with its corresponding 95% confidence interval (CI), estimated using the Wilson score method to account for binomial uncertainty in proportion-based measures (Bowyer, 2000). Statistical significance was defined as $p < 0.05$ for all inferential analyses.

All histopathological images were pre-processed prior to model input: they were resized to 224 × 224 pixels and normalized using ImageNet-derived mean and standard deviation values to align with the input requirements of the pre-trained MobileNetV2 (Sandler et al., 2018) backbone. To improve generalization and mitigate overfitting—particularly given the limited sample size—on-the-fly data augmentation was applied during training, including random horizontal and vertical flipping, rotation (±15°), and brightness adjustments (±20%).

The MobileNetV2 (Sandler et al., 2018) architecture was selected as the base model due to its proven efficacy in histopathological image analysis, favourable balance between accuracy and computational efficiency, and suitability for deployent in resource-constrained clinical settings (Mori et al., 2022). Its lightweight design, built on inverted residual blocks with depth wise separable convolutions, enables rapid inference without compromising diagnostic fidelity. MobileNetV2 (Sandler et al., 2018) architecture, specifically designed for resource-constrained environments like mobile devices, has emerged as a particularly relevant tool for this application. Its lightweight nature and efficient computational performance, achieved using depth wise separable convolutions, make it ideally suited for deployment in Health applications in primary care settings, potentially enabling non-specialist healthcare workers to accurately screen and classify OED lesions from simple clinical photographs (Ghezloo, 2024).

a) *b)*

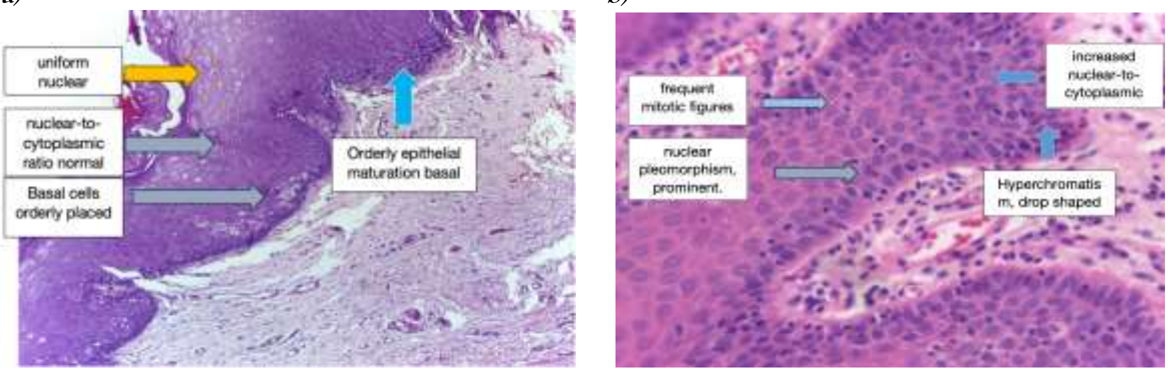


Figure 2: a) Normal oral mucosa. b) Severe epithelial dysplasia.

Results

Histopathological evaluation confirmed the presence of a full spectrum of epithelial changes across the study samples. Normal oral mucosa displayed regular epithelial architecture with uniform basal cell morphology and appropriate maturation toward the surface layers as shown in Fig. 2a. In contrast, severe epithelial dysplasia was characterized by pronounced cytological abnormalities, including nuclear pleomorphism, hyperchromasia, and increased mitotic activity, accompanied by architectural disturbances such as loss of polarity and disordered epithelial stratification is shown in Fig. 2b. These findings were used as the diagnostic gold standard against which clinical and analytical assessments were compared.

Table 1: Confusion-matrix-based quantitative assessment of model performance.

Class	Sensitivity	Accuracy	Precision	F1 score
Severe Dysplasia	0.875	0.881	0.875	0.875
Normal	0.887	0.881	0.887	0.887
Average	0.881	0.881	0.881	0.881

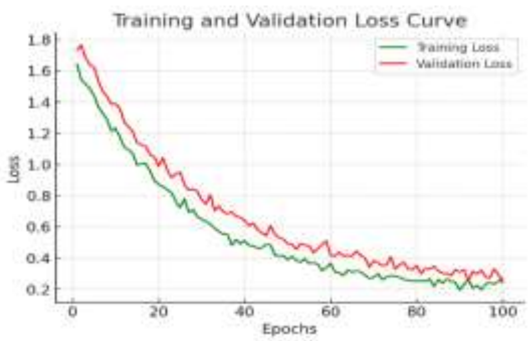
The proposed MobileNetV2-based (Sandler et al., 2018) pipeline demonstrated high diagnostic performance in distinguishing dysplastic from normal oral epithelium on a test set of 101 histopathological images as shown in Table 1. Based on confusion matrix analysis, the model achieved a sensitivity of 87.5%, correctly identifying 42 out of 48 dysplastic cases, while six were misclassified as normal. Conversely, 47 of the 53 normal epithelial samples were accurately classified, yielding a specificity of 88.7%. The overall accuracy was 88.1%, reflecting balanced performance across both diagnostic categories. Precision and F1-score were each 87.5% for the dysplastic class and 88.7% for the normal class, indicating consistent and reliable classification with minimal false-positive and false-negative predictions. Notably, Cohen's kappa (Bowyer, 2000), coefficient of 0.76 ($p < 0.001$) revealed substantial agreement between the model's predictions and expert histopathological diagnoses, a level of concordance consistent with prior studies evaluating deep learning systems in oral lesion assessment. These results, summarized in Table 2, support the potential of this lightweight deep learning framework as a robust and accurate assistive tool for the detection of oral epithelial dysplasia in digital pathology settings.

Table 2: Assessment of agreement between the CNN model and histopathological findings.

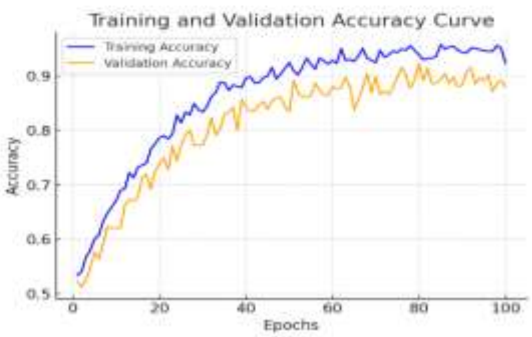
Quantity	Formula	Value
Po (Observed agreement)	(TP+TN)/N	0.881
Pe (Expected by chance)	$(48\times48+53\times53)/101^2$	0.495
K	$(P_o - P_e)/(1 - P_e)$	0.762

The findings in Table 2 indicate that deep learning–assisted histopathological evaluation can reliably detect OED with a high level of diagnostic accuracy. Cohen's value of 0.76 observed in this study reflects a substantial degree of agreement between the AI-generated classifications and expert histopathological assessments. This result is consistent with earlier validation studies in which CNN models achieved comparable kappa scores when applied to the grading of oral mucosal lesions, reinforcing the reliability of AI-based diagnostic tools in this domain.

a)



b)



c)

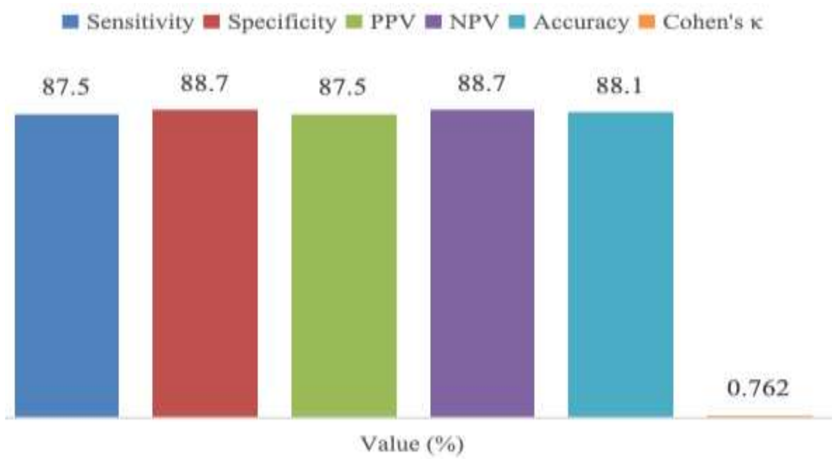


Figure 3: a) Loss curves for training and validation data. b) Accuracy curves for training and validation data. c) Performance metrics of the method.

In Fig. 3a and Fig. 3b, the training and validation loss and accuracy curves of the MobileNetV2 model, plotted across 100 epochs, demonstrated a steady improvement in performance. Both curves began to plateau after approximately 30 epochs, suggesting that the model had reached optimal learning with minimal evidence of overfitting. To further assess its generalization capability, the trained model was evaluated on an independent test set consisting of 15 normal and 15 dysplastic images, where it maintained consistent diagnostic performance. Fig. 3c shows the corresponding quality measures such as sensitivity, specificity, negative predictive value (NPV), negative predictive value (PPV) accuracy and Cohen's κ. All the values are consistently higher, which further signifies the applicability of method.

Discussions

Artificial intelligence is rapidly transforming digital pathology by enabling objective, reproducible, and scalable analysis of histological tissue. Among AI approaches, CNNs have consistently surpassed conventional machine learning and handcrafted feature-based methods in the detection and grading of oral potentially malignant disorders, including OED (Parola et al., 2024). The present study contributes to this growing body of evidence by demonstrating that a lightweight, transfer learning–based CNN—specifically, a modified MobileNetV2 (Sandler et al., 2018) architecture—can achieve high diagnostic accuracy in distinguishing dysplastic from normal oral epithelium, even with a limited, single-institution dataset.

MobileNetV2 (Sandler et al., 2018) was selected for its computational efficiency and proven efficacy in resource-constrained medical imaging tasks (Hadilou et al., 2025). In our implementation, the model achieved a sensitivity of 87.5% and specificity of 88.7%, with an overall accuracy of 88.1% and a balanced F1-score across both classes. These metrics are comparable to those reported in prior digital pathology studies (Rapado-González et al., 2024) and fall within the performance range typically observed among experienced pathologists interpreting digitized H&E-stained slides. Critically, the high sensitivity reduces the risk of false negatives—a vital consideration in early dysplasia detection, where missed diagnoses can delay intervention and increase malignant transformation risk. The corresponding specificity minimizes unnecessary referrals due to false positives, enhancing clinical utility.

The substantial inter-rater agreement between the model and expert pathologists—quantified by a Cohen's κ of 0.76—further validates the model's diagnostic reliability. This level of concordance mirrors the inter-observer variability commonly seen in histopathological practice, suggesting that the AI system functions as a consistent and trustworthy diagnostic adjunct. Our findings align with recent studies by (Aguirre-Urizar JM, 2024, Parola, 2024, Li X, 2024), who demonstrated that CNNs can reliably grade OED and even predict malignant transformation potential with accuracy rivalling human experts.

A key methodological strength of this work lies in its rigorous data curation strategy. Rather than relying on randomly sampled image patches—a common limitation in prior studies—we leveraged expert-guided ROI selection. All biopsy specimens were processed into standard H&E-stained slides, independently reviewed by a qualified pathologist, and annotated to highlight diagnostically relevant epithelial zones (particularly the basal and parabasal layers, where dysplastic changes are most pronounced). This pathologically informed approach ensures that the model learns from histologically meaningful features, closely mirroring real-world diagnostic workflows. Our strategy builds upon concepts like visual attention used by Ghezloo et al. 2024, but grounds ROI definition in direct microscopic evaluation rather than algorithmic inference alone, thereby improving the biological and clinical fidelity of model training.

Nevertheless, this study has limitations. Its retrospective, single-centre design and modest sample size (n = 101) may constrain generalizability. Although the dataset was carefully curated, larger, multi-institutional cohorts—including diverse ethnicities, staining protocols, and scanner types—are needed to validate robustness across varied clinical settings. Additionally, the current binary classification framework does not differentiate between grades of dysplasia (mild, moderate, severe), which is clinically relevant for risk stratification. Future work should expand the model into a multiclass system capable of fine-grained OED grading, incorporate explainability modules for clinician trust, and explore prospective integration into diagnostic pipelines to assess real-time impact on workflow efficiency and patient outcomes.

In conclusion, we have used the MobileNetV2 (Sandler et al., 2018) based model which demonstrates that a lightweight, expert-informed deep learning system can achieve human-level performance in the binary classification of oral epithelial dysplasia. By combining computational efficiency with histopathological rigor and tools such as QuPath (Bankhead, 2017), this approach holds promise as a scalable tool for early detection, particularly in settings with limited access to specialized pathologists.

Conclusion

This study demonstrates that a lightweight, deep learning–based model leveraging MobileNetV2 can achieve high diagnostic accuracy in distinguishing oral epithelial dysplasia from normal epithelium. The model exhibited substantial agreement with expert histopathological assessment (Cohen's κ = 0.76), underscoring its reliability as a diagnostic aid. By delivering consistent, objective, and efficient analysis of H&E-stained slides, this approach holds significant promise for improving early detection of oral potentially malignant disorders, reducing interobserver variability, and ultimately supporting better clinical decision-making and patient outcomes in oral pathology.

Acknowledgment

The authors acknowledge the Department of Histopathology, FMH College of Medicine and Dentistry, Lahore, Pakistan, for academic and technical support. No external funding was received for this study.
The authors used ChatGPT-4 (OpenAI) to assist with drafting the initial manuscript. All content was rigorously reviewed, verified, and approved by the authors, who assume full responsibility for the work.

REFERENCES

Shephard AJ, Bashir RMS, Mahmood H, et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. npj Precis Oncol. 2024;1:6.

Rapado-González, Ó., Salta, S., López-López, R., Henrique, R., Suárez-Cunqueiro, M. M., & Jerónimo, C. (2024). DNA methylation markers for oral cancer detection in non-and minimally invasive samples: a systematic review. Clinical Epigenetics, 16(1), 105.

Aguirre-Urizar JM, Lafuente-Ibáñez de Mendoza I, Aguirre-Urizar A, et al. A deep-learning system to predict epithelial dysplasia in oral leukoplakia. J Dent Res. 2024;103(12):1218–1226.

Xu, Z.; Lin, A.; Han, X. Current AI Applications and Challenges in Oral Pathology. Oral 2025, 5, 2.

Sandler, Mark, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

Peng J, Xu Z, Dan H, et al. Oral epithelial dysplasia detection and grading in oral leukoplakia using deep learning. *BMC Oral Health.* 2024;24:434.

Camalan, S., Mahmood, H., Binol, H., Araujo, A. L. D., Santos-Silva, A. R., Vargas, P. A., ... & Gurcan, M. N. (2021). Convolutional neural network-based clinical predictors of oral dysplasia: class activation map analysis of deep learning results. Cancers, 13(6), 1291.

Sahoo, Rakesh Kumar, et al. Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical diagnostic imaging: a systematic review and meta-analysis. Frontiers in Oral Health 5 (2024): 1494867.

Araujo, Anna Luiza Damaceno, et al. The use of deep learning state-of-the-art architectures for oral epithelial dysplasia grading: A comparative appraisal. Journal of Oral Pathology & Medicine 52.10 (2023): 980-987

Warin, Kritsasith, and Siriwan Suebnukarn. Deep learning in oral cancer-a systematic review. BMC Oral Health 24.1 (2024): 212.

Parola, Marco, et al. Towards explainable oral cancer recognition: Screening on imperfect images via Informed Deep Learning and Case-Based Reasoning. Computerized Medical Imaging and Graphics 117 (2024): 102433.

Bankhead, Peter, et al. QuPath: Open-source software for digital pathology image analysis. Scientific reports 7.1 (2017): 1-7.

Mete, Ozgur, and Bruce M. Wenig. Update from the 5th edition of the world health organization classification of head and neck tumors: overview of the 2022 WHO classification of head and neck neuroendocrine neoplasms. Head and neck pathology 16.1 (2022)

Bowyer, Kevin W. Validation of medical image analysis techniques. Handbook of medical imaging 2 (2000): 567-607.

Abdul, Nishath Sayed, et al. Applications of artificial intelligence in the field of oral and maxillofacial pathology: a systematic review and meta-analysis. BMC Oral Health 24.1 (2024): 122.

Alotaibi S, et al. AI in oral medicine: Is the future already here? Br Dent J. 2024;236:745–752.

Li XL, Zhou G. Deep learning in the diagnosis and prognosis of oral potentially malignant disorders. Cancer Screen Prev. 2024;3(4):203–213.

Hadilou, Mahdi, et al. Artificial intelligence based vision transformer application for grading histopathological images of oral epithelial dysplasia: a step towards AI-driven diagnosis. BMC cancer 25.1 (2025): 780.

Pirayesh, Zeynab, et al. Deep Learning-Based Image Classification and Segmentation on Digital Histopathology for Oral Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis. Journal of Oral Pathology & Medicine 53.9 (2024): 551-566.

Ghezloo F, Chang OH, Knezevich SR, Shaw KC, Thigpen KG, Reisch LM, Shapiro LG, Elmore JG. Robust ROI detection in whole slide images guided by pathologists' viewing patterns. *J Imaging Inform Med.* 2024;38(1):439-454.

Mori, Kazuki, et al. Detecting early-stage oral cancer from clinically diagnosed oral potentially malignant disorders by DNA methylation profile." Cancers 14.11 (2022): 2646.

Yurtkulu, Salih Can, Yusuf Hüseyin Şahin, and Gozde Unal. Semantic segmentation with extended DeepLabv3 architecture. 2019 27th signal processing and communications applications conference (SIU). IEEE, 2019.