Original Article

Deep Learning and Transfer Learning for Automated Image-Based Classification of Oral Epithelial Dysplasia: A Comparative Study of Convolutional Neural Network Architectures

M.Kuppusamy[1], Manish Gupta[2], Rajendra Singh[2]

[1]Ph.D Scholar, Department of Anatomy, IMCH&RC, Indore, Malwanchal University, Indore, Madhya Pradesh, India. [2]Assistant Professor, Department of Anatomy, IMCH&RC, Indore, Malwanchal University, Indore, Madhya Pradesh, India

Abstract

Background: Oral epithelial dysplasia (OED) is a spectrum of any architectural and cytological abnormality with an increased potential for malignant transformation. While the histopathological assessment continues to be a key element in the diagnosis and grading process, there is a risk of some subjectivity due to the complexity of morphological features and interobserver variability. With the rise of the digital pathology, incorporation of deep-learning algorithms for automated analysis of histopathological images has become possible. Convolutional neural networks (CNNs) are able to learn hierarchical morphological representations directly from image data in contrast to conventional machine-learning approaches that rely on manually selected morphometric variables. **Material and Methods:** A total of 81 histopathologically evaluated specimens were included, comprising 27 NBM, 27 LR, and 27 HR cases. Representative histopathological images were digitally acquired under standardized microscopic conditions and pre-processed for computational analysis. A 12-layer CNN was developed for binary classification, while a separate 13-layer CNN was developed for multiclass classification of NBM, LR, and HR. Transfer-learning experiments were subsequently performed using pretrained ResNet-50, VGG16, VGG19, MobileNet-V2, and Inception-V3 architectures. Model performance was assessed using accuracy, precision, recall, and F1-score. **Results:** The proposed multiclass CNN achieved an overall classification accuracy of 89.74%. The HR group demonstrated the strongest class-specific performance, with a precision of 0.92, recall of 0.94, and F1-score of 0.93. The LR group demonstrated comparatively lower performance, with precision, recall, and F1-score of 0.76, 0.71, and 0.73, respectively. Among the evaluated transfer-learning architectures, ResNet-50 demonstrated the highest multiclass accuracy (91.86%), followed by VGG19 (88.53%), VGG16 (86.92%), Inception-V3 (84.75%), and MobileNet-V2 (83.68%). **Conclusion:** The results showed that the deep-learning image analysis algorithm could be used to automatically classify OED and was feasible. The transfer learning, especially ResNet-50, showed good classification results in the training set. While deep-learning methods could be used as computational aids for objective evaluation of dysplastic lesions, further larger multicentric data sets, case-level validation, independent external testing and prospective clinical evaluation are needed to implement in clinical practice.

Keywords: Deep Learning, convolutional neural network, transfer learning, ResNet-50, oral epithelial dysplasia, oral leukoplakia, digital pathology, AI, histopathology, oral potentially malignant disorders.

Received: 25 July 2026 **Revised: 10 August 2026** **Accepted: 23 August 2026** **Published: 29 August 2026**

Introduction

Oral epithelial dysplasia is a spectrum of architectural and cytological changes that represent disturbances of epithelial maturation, differentiation and organization. Histopathological examination is still the primary approach to diagnosis and grading of dysplastic lesions, and will continue to be a key part of the risk assessment of OPDs. However, interobserver variation can affect grading due to the need to identify and weigh a number of morphological features when conducting a microscopic assessment.[1-3]

With the growing popularity of digital pathology, computational assessment of histopathological specimens is becoming arising an opportunity. Typical machine-learning methods mostly rely on handcrafted features such as nuclear size, cell size, texture features, and shape features. Such methods can yield quantitative data but only if the features chosen by the investigator are of satisfactory quality and extent.[4,5]

In this alternative computational paradigm, deep learning, the relevant image representations are obtained directly from image data. Within the domain of deep-learning architectures, convolutional networks are gaining even more significance for image analysis, as it is increasingly possible to learn increasingly complex spatial representations with subsequent convolutional layers. In histopathological use cases, CNNs can be able to learn features that are not well captured by a small

Address for correspondence: Dr. Rajendra Singh, Assistant Professor, Department of Anatomy, IMCH&RC, Indore, Malwanchal University, Indore, Madhya Pradesh, India.
E-mail: rajendrasingh211@gmail.com

DOI:
10.21276/amit.2026.v13.i2.992

How to cite this article: Kuppusamy M, Gupta M, Singh R. Deep Learning and Transfer Learning for Automated Image-Based Classification of Oral Epithelial Dysplasia: A Comparative Study of Convolutional Neural Network Architectures. Acta Med Int. 2026;13(2):1974-1980.

number of manually designed variables, such as nuclear morphology, cellular organization, epithelial architecture, tissue texture, and spatial relationships.[4,6,7]

A major limitation of deep-learning applications in medical imaging is the requirement for sufficiently large and well-annotated datasets. Histopathological datasets are frequently limited by the availability of carefully classified specimens. Transfer learning provides a potential solution by allowing pretrained neural networks to be adapted to specialized medical-image classification tasks. Architectures such as ResNet, VGG, MobileNet, and Inception have consequently been investigated for image-based classification across a range of medical and histopathological applications.[7-11]

The present study was undertaken to evaluate deep-learning-based classification of NBM, LR, and HR and to compare a proposed CNN architecture with established transfer-learning models. The study specifically aimed to determine whether direct image-based learning could provide effective discrimination among the three histopathological categories.

Materials and Methods

The present retrospective observational study was conducted in the Department of Oral and Maxillofacial Pathology at Index Medical College, Hospital & Research Centre, Indore, in collaboration with computational imaging resources

The study comprised 81 histopathologically evaluated specimens, equally distributed among three diagnostic categories: normal buccal mucosa (NBM; n = 27), low-risk epithelial dysplasia (LR; n = 27), and high-risk epithelial dysplasia (HR; n = 27). Diagnostic categorization was performed according to the histopathological classification adopted in the parent thesis. Representative histopathological fields from each specimen were digitally captured under standardized microscopic conditions, with consistent image-acquisition parameters maintained to minimize variations related to image quality and imaging conditions.

The acquired digital images were subsequently prepared for deep-learning analysis. As specified in the study methodology, the original image dimensions of 6000 × 4000 × 3 pixels were reduced to 600 × 400 × 3 pixels to facilitate computational processing. Each image was assigned to its corresponding diagnostic category (NBM, LR, or HR) and subsequently partitioned into training, validation, and testing subsets for model development and performance assessment.

Convolutional Neural Network Development: A 12-layer two-dimensional convolutional neural network (CNN) was developed for binary classification of NBM and HR dysplasia. The architecture incorporated convolutional layers for hierarchical feature extraction, max-pooling operations for dimensionality reduction, batch normalization to improve training stability, dropout for regularization, and fully connected layers for classification.

For multiclass classification, a separate 13-layer CNN architecture was developed to discriminate among NBM, LR, and HR. The multiclass architecture incorporated an additional convolutional layer and employed a softmax output layer to generate class probabilities for the three diagnostic categories. This architecture enabled direct image-based learning of morphological representations relevant to the classification of the three histopathological groups.

Transfer-Learning-Based Classification: To further evaluate the applicability of established deep-learning architectures, a transfer-learning framework was implemented using pretrained convolutional neural networks. The evaluated architectures included ResNet-50, VGG16, VGG19, MobileNet-V2, and Inception-V3. The pretrained networks were adapted to the present histopathological classification task and fine-tuned for multiclass discrimination of NBM, LR, and HR images. This comparative approach enabled assessment of the relative classification performance of different CNN architectures within the same study framework.

Model Performance Evaluation: The performance of the proposed CNN and transfer-learning models was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Overall accuracy was used to quantify the proportion of correctly classified images across the diagnostic categories, whereas precision, recall, and F1-score were used to assess class-specific discriminatory performance. These measures enabled comparative evaluation of the proposed CNN and pretrained transfer-learning architectures for automated image-based classification of NBM, LR, and HR.

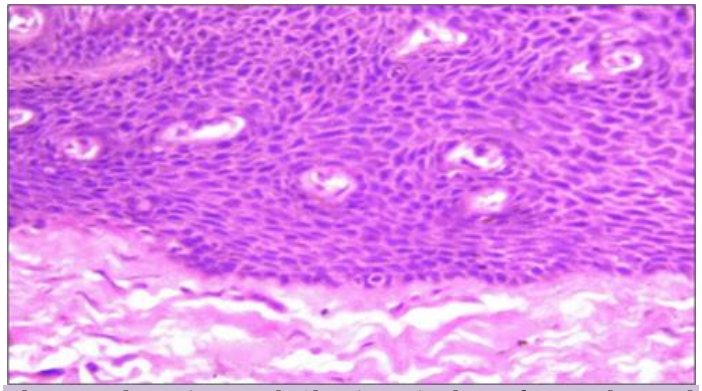

Photo 1: Photomicrograph Showing Histology of Normal Buccal Mucosa Under High Magnification (H and E Stain; 40x Magnification)

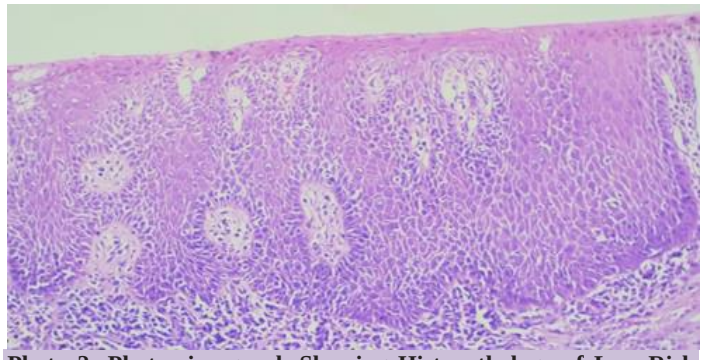

Photo 2: Photomicrograph Showing Histopathology of Low-Risk Epithelial Dysplasia Under High Magnification (H and E Stain; 40x Magnification)

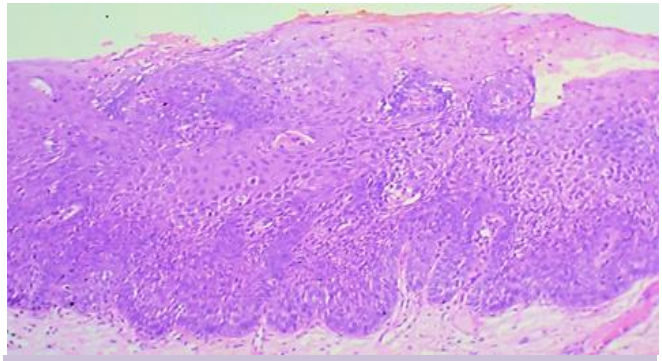

Photo 3: Photomicrograph Showing Histopathology of High-Risk Epithelial Dysplasia Under High Magnification (H and E Stain; 40x Magnification).

RESULTS

A total of 81 histopathologically confirmed specimens were included in the present study for computational assessment and classification of oral epithelial dysplasia. The specimens were equally distributed across three diagnostic categories, comprising 27 (33.3%) cases of normal buccal mucosa (NBM), 27 (33.3%) cases of low-risk epithelial dysplasia (LR), and 27 (33.3%) cases of high-risk epithelial dysplasia (HR). The balanced distribution of specimens across the diagnostic groups provided comparable representation of each category and facilitated standardized comparative evaluation of the computational models. The distribution of the study population is presented in [Table 1 and Figure 1]

Table 1: Distribution of the study population

Diagnostic group	Number of cases	Percentage (%)
Normal buccal mucosa (NBM)	27	33.3
Low-risk epithelial dysplasia (LR)	27	33.3
High-risk epithelial dysplasia (HR)	27	33.3
Total	81	100.0

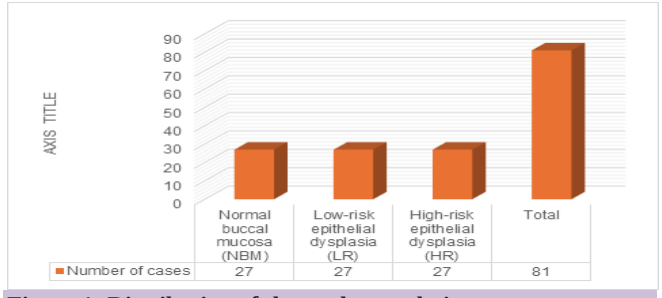

Figure 1: Distribution of the study population

Performance of the Proposed CNN: The proposed multiclass CNN demonstrated an overall classification accuracy of 89.74% for discrimination among NBM, LR, and HR.

The HR group demonstrated the strongest class-specific performance, with a precision of 0.92, a recall of 0.94, and an F1-score of 0.93. In contrast, the LR group demonstrated comparatively lower performance, with a precision of 0.76, a recall of 0.71, and an F1-score of 0.73. These findings indicate that HR lesions were more readily distinguished from the other diagnostic categories than LR lesions. [Table 2 & Figure 2]

Figure 2: Performance Evaluation of the Proposed CNN Model for Multiclass Classification Using Precision, Recall, F1-Score, and Accuracy.

Table 2: Performance Evaluation of the Proposed CNN Model for Multiclass Classification Using Precision, Recall, F1-Score, and Accuracy

Diagnostic group	Precision	Recall	F1-score	Overall Accuracy (%)
NBM	0.82	0.84	0.83	89.74%
LR	0.76	0.71	0.73	
HR	0.92	0.94	0.93	

Performance of Transfer-Learning Models: Among the evaluated transfer-learning architectures, ResNet-50 achieved the highest multiclass classification accuracy of 91.86%. The proposed CNN demonstrated the second-highest accuracy at 89.74%, followed by VGG19 at 88.53% and VGG16 at 86.92%. Inception-V3 and MobileNet-V2 achieved accuracies of 84.75% and 83.68%, respectively. [Table 3]

Table 3: Comparative classification accuracy of deep-learning architectures

Model	Classification accuracy (%)
ResNet-50	91.86
Proposed CNN	89.74
VGG19	88.53

VGG16	86.92
Inception-V3	84.75
MobileNet-V2	83.68

Discussion

It shows that deep-learning image analysis is able to classify oral epithelial dysplasia effectively and is feasible in the current study. The proposed multiclass CNN achieved an accuracy of 89.74% while the transfer-learning experiments showed that ResNet-50 achieved the highest overall classification accuracy of 91.86%. Available findings show that the histopathological images possess enough visual information that computational model can confidently classify the images to NBM, LR or HR within the study data set.

The finding is especially significant because, unlike traditional feature-based machine-learning models, deep-learning models fundamentally rely on different types of processes. Unlike manually designed measurements, CNNs naturally extract hierarchical representations from image data that may be used simultaneously to analyze nuclear morphology, cellular organization, epithelial architecture, tissue texture and spatial relationships.[4,5]

This proposed multiclass CNN achieved an overall accuracy of 89.74%, which shows that the CNN has a fairly good discriminatory power among the three diagnostic groups. This performance was better than the best conventional feature-based machine-learning model that was evaluated in the related study, Random Forest whose accuracy was 86.73%. The difference implies that direct image-based learning is capable of some visual information that is not fully covered by the selected manually extracted morphometric variables.

CNN-based analysis may offer the benefit of representing images at multiple levels, which may be learned. Early layers in the convolutional network can detect rather simple visual patterns, and deeper layers can combine the patterns into more complex structures. Such representations might include the architectural organisation at the tissue level, cellular arrangement, epithelial organisation, and nuclear characteristics in histopathological pictures.

But this increased accuracy with CNN does not imply that deep learning is always better than traditional machine learning. Several factors, such as dataset characteristics, image preprocessing, feature representation, class distribution, training strategy and validation methodology, influence the performance of the models. The comparison of direct comparison is thus meaningful only if the models are examined under the same experimental conditions.

The HR category achieved the best classification results with precision equal to 0.92, recall equal to 0.94 and F1 score equal to 0.93. LR exhibited, on the other hand, relatively low values of 0.76, 0.71 and 0.73.

The increased recognition of HR lesions may be related to the increased severity of architectural and cytological abnormalities. In advanced dysplasia, features like nuclear enlargement, pleomorphism, irregularity, abnormal maturation of epithelial cells and an increased N:C ratio can create more distinct visual patterns.[2,12]

The significant difference in performance between LR and the other two is especially relevant. Low-risk dysplasia is an intermediate stage of dysplasia morphologically, and the microscopic features of low risk dysplasia can have overlapping features with normal epithelium and with higher dysplasia. Thus, the separation between LR and nearby categories may not be as clear cut.

It shows the importance of not only overall accuracy but also class specific metrics. Overall accuracy may be high, but discriminatory ability may be lower for an intermediate diagnostic category for a model. Future studies should thus report class-specific precision, recall, F1-score, sensitivity and specificity where applicable and confusion matrix.

The transfer-learning approach which has achieved the best classification accuracy was ResNet-50 with 91.86%. It outperforms the proposed CNN and all the other architectures tested for transfer learning.

ResNet-50's performance might be connected with its residual learning structure. Residual connections are used to optimize deeper networks by allowing information to flow more efficiently through multiple layers of the network. An architectural feature that can facilitate learning in hierarchy-based visual representations in complex image analysis tasks.

Within histopathological image classification, the information that can be obtained can span several spatial scales from nuclear morphology to the structure of the entire epithelium. It can thus be argued that deep residual networks' power to learn hierarchical representations might play a role in their success in this context.

However, the findings regarding the superiority of ResNet-50 need to be considered in the context of the specific dataset and experimental setting of this study. This performance will only be confirmed if external datasets can be found that show it is reproducible in other patient populations, laboratories, staining protocols, and imaging systems.

VGG19 achieved an accuracy of 88.53%, while VGG16 achieved 86.92%. Both architectures showed fairly good performance, but still lagged behind ResNet-50.

VGG family is relatively simple sequential conv-net, shown usefulness in image classification tasks. This excellent result in the current study indicates the potential of relatively simple deep CNN models for learning diagnostically relevant representations from histopathology images.

That VGG19 performed better than VGG16 might be due to increased depth and representational power of VGG19. This difference should not be extrapolated without independent evidence, however, as it can be greatly affected by the properties of both the training set and optimization method.

Inception-V3 had an accuracy of 84.75% which is still lower than ResNet-50, VGG19, and VGG16, but higher than MobileNet-V2. The multiscale feature extraction used in inception architectures is conceptually similar to the way one interprets histopathology, in which the information one needs for diagnosis can be found at various spatial scales.

The better performance of Inception-V3 in the current set of

data could be due to the architectural differences, learning dynamics, or representation of features. Importantly, the result does not suggest that Inception-V3 is by itself inferior in terms of histopathological analysis, but it was not as successful as the other models analyzed in this study in these circumstances.

The worst performance was the accuracy of MobileNet-V2 which was 83.68%. However, due to its less high classification accuracy, MobileNet-V2 has a comparatively light architecture and is designed to be fast in computational.

Thus, whether the model is selected for future clinical use may not solely rely on classification accuracy. Other factors to consider are the amount of computational necessary, the time necessary to make inferences, memory usage, deployment environment, interpretability, and robustness. A slightly less accurate model may have practical benefits if that deployment is to be on limited-resource hardware.

The results are generally comparable to previous studies which have shown the potential for deep-learning methods to perform oral histopathological image analysis. Nguyen et al. tested Inception-V3 for grading the degree of epithelial dysplasia and found moderate to substantial agreement with the human observers.[7] Similarly, the application of CNN-based methods for classification of oral histopathological images has proven to be beneficial in other investigations.[4,6]

In the present study, these observations are extended by making them comparable with multiple well-known architectures that have been previously used for transfer learning on the same experiment. This type of comparison can be beneficial since the neural network structure can vary in its sensitivity to the number of data points, resolution, preprocessing techniques, and morphologic complexity.

The results should be considered as proof of computational feasibility and not as firm proof of superiority in clinical results. Reported accuracy values across studies are of questionable value when they are compared directly due to the variation in data and in the methodology used to validate the data in different studies.

One of the major contributions of the current work is the comparison between the direct image based deep learning and traditional feature based machine learning approaches. The feature-based study showed that the conventional machine learning classifiers Random Forest gave the best accuracy of 86.73%. The proposed CNN attained 89.74% accuracy, and ResNet-50 attained 91.86% accuracy.

The slight improvement seen with the deep-learning methods indicates that direct image analysis may be able to extract more information from the image morphology than the morphometrics selected. While conventional morphometry is highly sensitive to what features are used for analysis, CNNs have the potential for discovering new features within an image that are not pre-determined.

This benefit needs to be weighed against the increased complexity and interpretability limitations of deep-learning models, though. While feature-based models allow for direct analysis of their classification variables, CNNs have internally acquired representations which are hard to interpret without specific clarifying techniques.

Importance of Explainability:

Interpretability represents an important consideration in the clinical translation of deep-learning systems. A highly accurate model that cannot provide meaningful information regarding the basis of its prediction may be difficult to integrate into routine pathology workflows.

Explainable artificial intelligence techniques such as Grad-CAM, saliency mapping, and attention visualization may assist in identifying histopathological regions that contribute to model predictions.[13-15] Such techniques could potentially determine whether the model is focusing on diagnostically meaningful regions, such as dysplastic epithelial areas, nuclear abnormalities, or architectural changes.

Integration of explainability methods would therefore represent an important next step in developing clinically meaningful AI systems for OED. Rather than simply reporting a predicted diagnostic category, future systems could potentially provide both the classification and a visual representation of the image regions contributing to that classification.

Dataset Size and Validation Considerations

The study included 81 histopathologically evaluated specimens, which represent a relatively limited dataset for training and evaluating deep-learning models. Although digital image analysis can generate many image observations from individual specimens, images originating from the same patient or specimen may not be statistically independent.

This issue is particularly important in medical image analysis because random image-level partitioning can potentially result in images from the same case being represented in both training and testing datasets. Such an arrangement may lead to overly optimistic estimates of model performance if the model learns case-specific visual characteristics.

Future investigations should therefore emphasize case-level data partitioning, ensuring that images from an individual specimen are restricted to a single data partition. Independent external datasets should subsequently be used to assess generalizability.

Technical Reproducibility

Deep-learning performance may be affected by image acquisition and preprocessing. Variations in staining intensity, tissue preparation, section thickness, illumination, image resolution, camera characteristics, and preprocessing procedures can alter the visual characteristics of histopathological images.

Consequently, standardized image-acquisition protocols and preprocessing pipelines are important for reproducibility. Models developed using images from a single institution may not necessarily perform similarly when applied to images generated using different scanners, microscopes, staining protocols, or laboratory procedures.

The development of multicentric datasets with standardized quality-control procedures would therefore represent an important step toward robust clinical translation.

Clinical Implications: The present findings indicate that deep-learning models may have potential as adjunctive computational tools for oral epithelial dysplasia classification. Automated image-based classification could potentially assist pathologists by providing an independent computational assessment of

histopathological images.

However, the current evidence does not support replacing expert histopathological assessment with automated classification. Histopathological diagnosis requires integration of morphological findings with clinical information, specimen quality, anatomical context, and diagnostic expertise. Deep-learning systems should therefore initially be considered decision-support tools rather than autonomous diagnostic systems.

The relatively strong performance observed for HR dysplasia is particularly encouraging for risk-oriented classification. Nevertheless, the lower performance for LR highlights the continuing difficulty of recognizing intermediate morphological phenotypes and emphasizes the need for more extensive training datasets.

Limitations: The principal limitation of the present investigation is the relatively small number of histopathologically evaluated cases. Although the dataset included three balanced diagnostic groups, 81 specimens remain limited for robust development and validation of deep-learning models.

A second limitation relates to potential dependence among multiple images derived from the same specimen. Future studies should ensure patient- or specimen-level separation between training, validation, and testing datasets.

Third, the study evaluated a limited number of CNN and transfer-learning architectures. Other contemporary architectures and hybrid approaches may provide different performance characteristics.

Fourth, the present analysis focused primarily on classification performance. Additional evaluation of computational efficiency, calibration, robustness, explainability, and external generalizability would provide a more comprehensive assessment of clinical utility.

Finally, the reported results were obtained from the study dataset and should not be interpreted as evidence of performance in an independent clinical population.

Future Directions: Future research should focus on developing larger, multicentric and prospectively curated histopathological datasets representing diverse patient populations and imaging conditions. Case-level data partitioning and independent external validation should be incorporated into the study design.

Further studies should also investigate explainable deep-learning approaches using Grad-CAM, saliency maps, attention mechanisms, and related methods to determine whether model predictions are based on diagnostically meaningful histopathological features.[13,14,15]

Integration of deep-learning predictions with quantitative morphometry may represent another promising direction. A hybrid framework combining interpretable morphometric measurements with learned image representations could potentially provide both strong classification performance and greater clinical interpretability.

Ultimately, prospective clinical studies will be required to determine whether deep-learning-assisted classification improves diagnostic reproducibility, risk stratification, and clinical decision-making in patients with oral potentially malignant disorders.

Conclusion

The present study demonstrates the feasibility of deep-learning-based image classification of oral epithelial dysplasia using both a proposed CNN and established transfer-learning architectures. The proposed multiclass CNN achieved an overall accuracy of 89.74%, while ResNet-50 demonstrated the highest classification accuracy of 91.86% among the evaluated models.

The stronger classification performance observed for HR dysplasia and the comparatively lower performance for LR dysplasia indicate that the morphological continuum of epithelial dysplasia remains an important challenge for automated classification. Nevertheless, the findings demonstrate that deep-learning models can identify image-level patterns associated with different histopathological categories.

Transfer learning represents a potentially useful strategy for histopathological image analysis when large annotated datasets are difficult to obtain. ResNet-50 demonstrated strong performance within the present study framework.

Deep-learning systems should currently be regarded as adjunctive computational tools rather than replacements for expert histopathological diagnosis. Future research should prioritize larger multicentric datasets, standardized image acquisition, case-level validation, independent external testing, model explainability, robustness assessment, and prospective clinical evaluation before routine diagnostic implementation.

References

1. Speight PM, Abram TJ, Floriano PN, James R, Vick J, Thornhill MH, et al. Inter-observer agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. Oral Surg Oral Med Oral Pathol Oral Radiol. 2015;120(4):474-82.
2. Odell EW, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: recognition, grading and clinical significance. Oral Dis. 2021;27(8):1947-1976.
3. Ng GTE, Phang SC, Yu KS, Tiwari L, Khurram SA, Sloan P, et al. Understanding interobserver variability of pathologists to improve oral epithelial dysplasia grading. Oral Dis. 2025;31(3):838-845.
4. Das M, Dash R, Mishra SK. Automatic detection of oral squamous cell carcinoma from histopathological images using deep convolutional neural networks. Int J Environ Res Public Health. 2023;20(3):2131.
5. Komura D, Ishikawa S. Machine learning approaches for pathologic diagnosis. Virchows Arch. 2018;473(2):131–138.
6. Araújo ALD, da Silva VM, Lopes MA, Vargas PA, Moraes MC, Santos-Silva AR. Deep learning for oral epithelial dysplasia grading. Oral Surg Oral Med Oral Pathol Oral Radiol. 2023;136(1):e83. doi:10.1016/j.oooo.2023.03.308
7. Nguyen PTH, Sakamoto K, Ikeda T. Deep-learning application for identifying histological features of epithelial dysplasia of the tongue. J Oral Maxillofac Surg Med Pathol. 2022;34(4):514–522.
8. Bharti A, Kashyap N, Kamboj M, Joshi K, Hooda S, Mishra D, et al. Deep learning based automated detection and grading of oral epithelial dysplasia: a comparative histopathological image analysis. J Dent. 2026;173:106771. doi:10.1016/j.jdent.2026.106771.
9. Sukegawa S, Ono S, Tanaka F, Inoue Y, Hara T, Yoshii K, et al. Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists. Sci Rep. 2023;13(1):11676. doi:10.1038/s41598-023-38343-y.
10. Pandiar D, Choudhari S, Poothakulath Krishnan R. Application of

InceptionV3, SqueezeNet, and VGG16 convoluted neural networks in the image classification of oral squamous cell carcinoma: a cross-sectional study. Cureus. 2023;15(11):e49108. doi:10.7759/cureus.49108.
11. Dhanya K, Venkata Vara Prasad D, Venkataramana Lokeswari Y. Detection of oral squamous cell carcinoma using pre-trained deep learning models. Exp Oncol. 2024;46(2):119-128. doi:10.15407/exp-oncology.2024.02.119.
12. World Health Organization. WHO Classification of Head and Neck Tumours. 5th ed. Lyon: International Agency for Research on Cancer; 2022.
13. Shephard AJ, Bashir RMS, Mahmood H, Jahanifar M, Minhas F, Raza SEA, et al. A fully automated and explainable algorithm for predicting malignant transformation in oral epithelial dysplasia. NPJ Precis Oncol. 2024; 8:137. doi:10.1038/s41698-024-00624-8.
14. Alajaji SA, Khoury ZH, Jessri M, Sciubba JJ, Sultan AS. An update on the use of artificial intelligence in digital pathology for oral epithelial dysplasia research. Head Neck Pathol. 2024;18(1):38.
15. Ardila CM, Pineda-Vélez E, Díaz-Laclaustra AI, et al. Artificial intelligence-assisted histopathologic diagnosis and grading of oral epithelial dysplasia: a systematic review and functional meta-synthesis. Head Neck Pathol. 2026;20(1):74. doi:10.1007/s12105-026-01937-9