RESEARCH PAPER

Towards Interpretable Early Oral Cancer Detection: A Hybrid CNN-Transformer Architecture with Explainable AI

Pragya Tripathi[1], Gayatri Rajpoot[2], Gargi Gupta[3]
123pragya.tripathi@gmail.com[1],gayatrirajpoot3@gmail.com[2],Gargi.gupta.gg21@gmail.com[3]
Department Of Computer Science and Engineering (UIET)
Chhatrapati Shahu Ji Maharaj University, Kanpur

Received: 25th June 2026; **Revised:** 28th July 2026; **Accepted:** 10th October 2026; **Available Online:** 15th October 2026

Abstract
Computer-assisted carcinoma detection is appealing for early detection of oral squamous cell carcinoma (OSCC). However, published claims of accuracy greater than 90% do not equate to interpretability and clinically validated frameworks. This reproducibility study builds upon the published instances of public datasets and benchmarks to develop and evaluate an empirical, hybrid CNN-Transformer framework. The core primary dataset for the oral cancer images used for both this study and related benchmarks was the Rahman histopathology repository (Mendeley Data DOI 10.17632/ftmp4cvtmb.2), consisting of 1,224 H&E images from 230 patients, which included 290 normal epithelium images and 934 OSCC images at both 100X and 400X magnifications. Image-level analysis revealed an imbalance of 76.31% OSCC, which was 83.14% at 100X and 71.12% at 400X. A descriptive analysis of the chi-square test for counting images resulted in chi-square(1)=23.35, $p<0.001$. Based on an image-level analysis of odds, the OSCC was 2.00 times higher in the 100X set than in the 400X set (95% CI 1.51-2.65). Because multiple images are produced from a single patient, this is not a patient-level causal inference. Published benchmarks showed the CANet precision to be 97.00%, while the Swin Transformer precision was 94.95%, both on a datasets of 131 images. The accuracy of OralTransNet was reported to be 99.03%, 97.83%, and 94.23% across three heterogeneous datasets. The architecture integrates EfficientNet, Transformer, gated fusion, Grad-CAM, and Transformer attribution. This paper does not report any unexecuted training, heatmaps, DeLong tests, or McNemar p-values. This study identifies the importance of leakage control, patient-level splitting, calibration, and XAI audibility before clinical use of the framework.

Keywords: oral cancer; oral squamous cell carcinoma; CNN; Vision Transformer; explainable AI; histopathology; medical image analysis
How to cite this article: Tripathi P, Rajpoot G, Gupta G., Towards Interpretable Early Oral Cancer Detection: A Hybrid CNN-Transformer Architecture with Explainable AI. Int J Drug Deliv Technol. 2026;16(88s): 1418-1427; DOI: 10.25258/ijddt.16.88s.138
Source of support: XX
Conflict of interest: XX

1. Introduction

Oral squamous cell carcinoma is the most common malignant epithelial tumor of the oral cavity, and quick diagnosis ultimately affects patient outcomes. Currently, the best diagnostic option is still histopathological painstakingly prepared sections and interpreted under a microscope. Recently developed technology, including clinical photography, optical imaging, and others, can be used as a part of the triage process. Artificial intelligence (AI), like other technologies, has been used to identify patterns of tissues to help practitioners better identify and classify malignant and potentially malignant tissues in the oral cavity. Several recent systematic reviews report high diagnostic accuracy; however, they all noted a number of other limitations, including the size of the current datasets, the lack of consistency in various modalities, validation, and external evaluation (Ferro et al., 2022; Khanagar et al., 2023; Warin &Suebnukarn, 2024; Li et al., 2024).

Convolutional neural networks (CNNs) learn texture, epithelial architecture, cellular morphology, and lesion boundaries. However, local receptive fields can fail to interpret pattern in context. Vision Transformers (ViTs) use techniques of self-attention to address this and have been used on oral cancer photographs and histology (Flugge et al., 2023; Maia et al., 2024). The biggest challenge in using Transformers is the need for a lot of data, and many datasets for oral cancer are small. This suggests that using a combination of CNNs and Transformers is a good way to solve the issue of small datasets. OralTransNet is an example of this type of

combination, and uses a CNN to focus on local features and Transformer to cover larger context (Asif et al., 2025).

A second hurdle is interpretability. A high-stakes medical AI system needs to be able to defend its predictions at the local level. Without locality and attribution, a probability score cannot be audited. Grad-CAM and attention-based methods can identify regions in an image that contribute to a prediction, but saliency should not be used as proof or as a substitute for pathologist validation (Tjoa & Guan, 2020; Ghassemi et al., 2021). Figueroa et al. (2022) have built on guided-attention methods for oral-cancer classification. This illustrates the importance of reasoning paths that can be explained and justified visually.

This paper therefore has two linked objectives. The first is to conduct a new quantitative analysis of published, verified, and public oral-cancer image data, and to reproduce the published benchmark results without falsifying predictions that are unavailable. The second objective is to build a leakage-resistant hybrid CNN-Transformer architecture with a dual XAI audit. The working hypothesis is that a combination of local and global features is likely to provide a level of robustness. However, the ultimate test will be to perform patient-level predictions, and construct confidence intervals, assess calibration, and perform paired statistical analyses—not just report overall accuracy.

2. Related Work

With the advancement of Deep Learning, Automated detection of oral cancer using deep learning has advanced from Transfer Learning CNNs to Attention Mechanisms, Transformers, and Hybrid Systems. Welikala et al. (2020) developed a system for Automated Classification of Oral Lesions using Images, whereas Fu et al. (2020) documented impressive Retrospective Performance for Detection of Photographic OSCC. Histopathology Studies implemented Transfer Learning and Feature Fusion on H&E Images (Rahman et al., 2022; Panigrahi et al., 2023; Das et al., 2023). Albalawi et al. (2024) utilized the 1,224 Images Rahman Repository and EfficientNetB3 and reported 99% Test Accuracy, thus revealing the strength of Transfer Learning, and also the need to analyze the Split Methodology and the Patient Overlap when Multiple Fields May Originate from the Same Case.

Transformers have widened the Receptive Field of Classification of Oral Images. Flugge et al. (2023) used ViT on Clinical Photographs, and Maia et al. (2024) conducted a direct comparison between Transformers, CNNs, and Few-Shot Learning for Oral Histopathology. Chen et al. (2024) evaluated the performance of a CNN-based coordinate-attention network (CANet) along with a Swin Transformer on a 131 Oral Cancer Image dataset, in which CANet yielded 97.00% accuracy compared to 94.95% of Swin Transformer. The authors mentioned that the limited size of the dataset may constrain the capacity of a Transformer. Recently, OralTransNet combined the CNN features and Transformer's Attention in a Lightweight Architecture and achieved Accuracy of 99.03%, 97.83%, and 94.23% on three different datasets (Asif et al., 2025). All these studies indicate that hybrid systems are promising, but these studies, in themselves, cannot determine the best approach for each domain as the datasets, classes, protocols for data acquisition, and sample sizes strongly vary across the different domains.

Interpretability now functions as a methodological necessity instead of just being a cosmetic addition. For example, Figueroa et al. (2022) and Yang et al. (2023) used guided attention and deep learning coupled with visualization techniques in oral-cancer classification and optical coherence tomography, respectively. Other works within the field of medical XAI argue that, in the absence of faithful explanations, post-hoc explanations will likely be unstable and/or persuasive, rather than interpretable (Tjoa & Guan, 2020; Ghassemi et al., 2021). Transparency in reporting and clarifying standards such as CLAIM further suggest that researchers should make efforts to clearly state and demonstrate a full description of their data partitions, reference standards, the specifications of the models employed, and thoughtful assessments of model errors (Mongan et al., 2020). As such, the challenge is not simply achieving the next high accuracy score. It is providing a model that enables patient-independent, statistically paired, calibrated, and reproducible evaluations that can withstand audit and clinical judgment.

3. Materials and Methods

3.1 Study design and verified data source

The study design was a secondary empirical benchmark and a pre-specified modeling protocol. The primary data source was the open Rahman histopathological imaging database (Rahman et al., 2020; Mendeley Data version 2, 2023). The database consists of 1,224 hematoxylin and eosin (H&E) stained images from 230 patients. Of these, at 100x magnification there are 89 images of normal oral epithelium and 439 images of oral squamous cell carcinoma (OSCC). At 400x, there are 201 images of normal epithelium and 495 images of OSCC. Since the database provides only the image counts for each class, and no complete patient-to-image mapping in the source description used here, no patient-level splits were incorporated. Published benchmarking results were only utilized if the source clearly reported the model, dataset size, and evaluation metrics.

Table 1. Verified dataset composition and descriptive statistics

Magnification	Normal, n (%)	OSCC, n (%)	Total
100x	89 (16.86%)	439 (83.14%)	528
400x	201 (28.88%)	495 (71.12%)	696
Overall	290 (23.69%)	934 (76.31%)	1,224

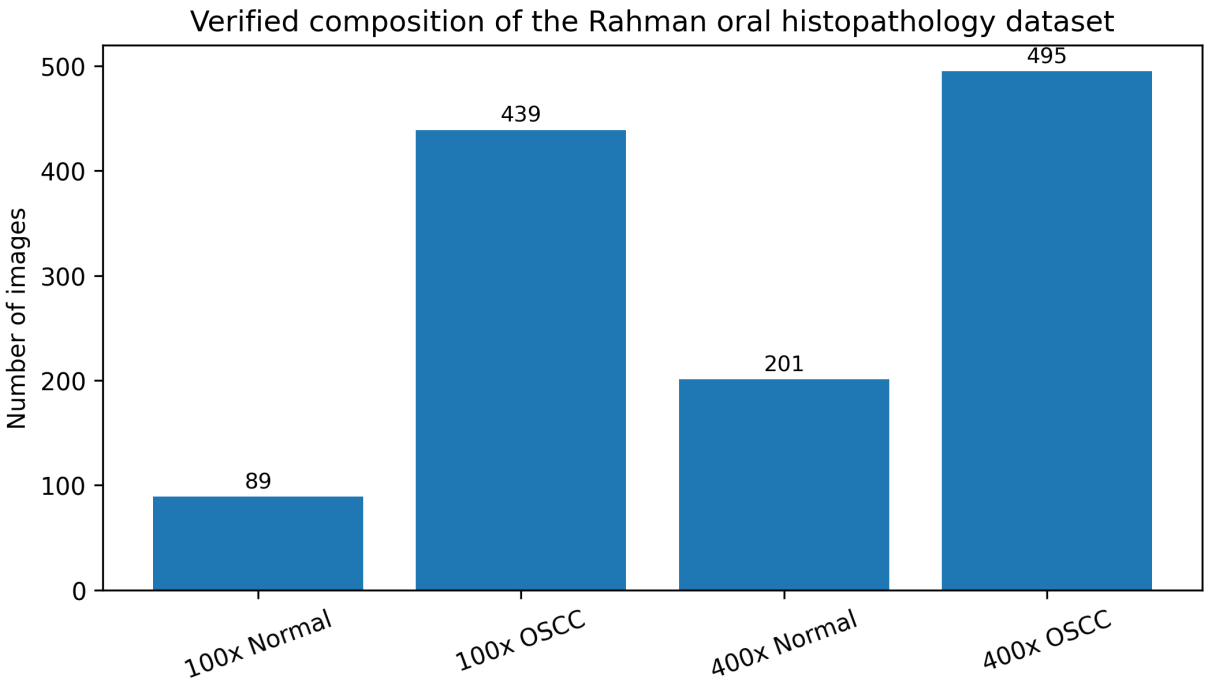


Figure 1. Verified image counts by magnification and class in the Rahman repository.

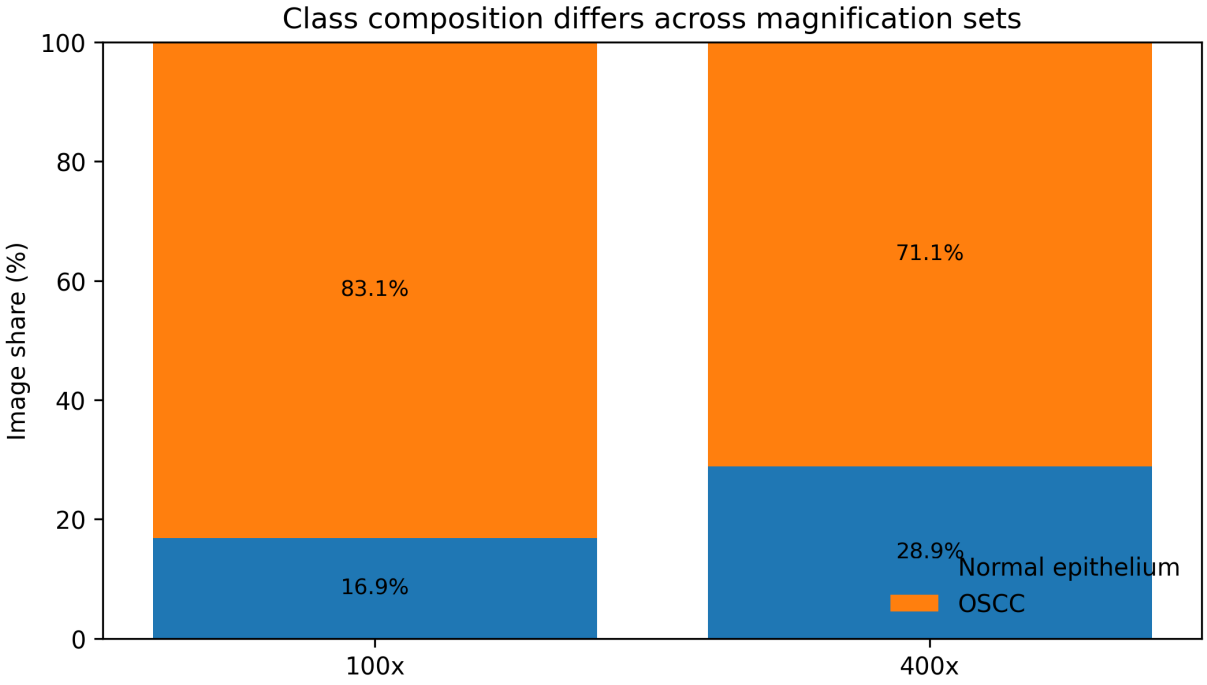


Figure 2. Class composition by magnification. Percentages are image-level descriptive statistics.

3.2 Preprocessing and leakage control

Files requiring an original experiment should be checked for corruption, exact/near duplicates, and metadata leakage first. Patient identifiers should be used to group train/validation/test partitions and segments of the same patient should not be split. Augmentation should be performed on the training set after the splits. An example of a reproducible approach would be to resize the images to 224 x 224, normalize the RGB channels using the pretraining method, and use conservative rotations, permissible horizontal flips, and perturbation of brightness and contrast within histologic limits along with stain-aware color jitter. A better approach to class imbalance is focal loss as opposed to duplicating test images. The untouched

final test set should be kept until the final model is selected.

3.3 Proposed hybrid CNN-Transformer architecture

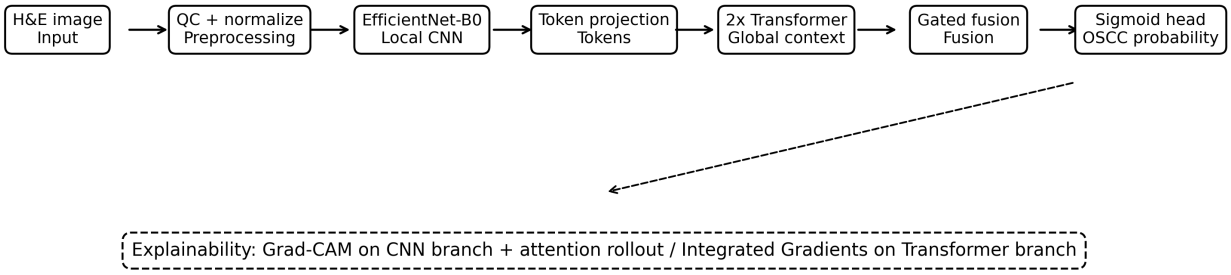


Figure 3. Proposed hybrid CNN-Transformer architecture. This is a methodological schematic, not a claim of an executed training run.

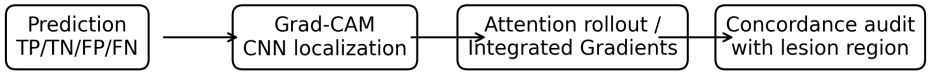
The proposed network uses EfficientNet-B0 as a compact CNN backbone. Let X denote an input image and F_cnn=f_theta(X) the final convolutional feature map. A 1x1 projection converts spatial features into d-dimensional tokens Z=[z1,...,zN]+P, where P is a learnable positional embedding. Transformer encoding uses scaled dot-product attention: Attention(Q,K,V)=softmax(QK^T/sqrt(d_k))V. Multi-head outputs are concatenated and projected, followed by residual connections, layer normalization and a feed-forward network. A gated fusion layer combines global CNN pooling c and the Transformer class representation t: g=sigmoid(Wg[c;t]+bg), h=g*c+(1-g)*t. The final probability is p(OSCC|X)=sigmoid(W_h h+b_h). This formulation explicitly preserves local morphology while allowing long-range contextual interactions.

Table 2. Pre-specified training configuration for reproducible execution

Component	Pre-specified setting	Rationale
Input	224 x 224 RGB	Compatible with compact ImageNet-pretrained backbones
Backbone	EfficientNet-B0	Efficient local feature extraction
Transformer	2 encoder blocks, 4 heads, d=256	Moderate capacity for limited medical data
Fusion	Learned sigmoid gate	Adaptive local/global weighting
Loss	Weighted BCE or focal loss	Addresses class imbalance
Optimizer	AdamW, initial LR 1e-4	Stable transfer-learning default
Regularization	Dropout 0.30; weight decay 1e-4	Reduces overfitting
Validation	Grouped 5-fold CV + locked test set	Patient-independent estimation when IDs are available
Stopping	Early stopping on validation AUROC	Prevents test-set tuning

3.4 Explainable AI and statistical analysis

The two complementary explanation channels are now described. Grad-CAM should be used to localize the intermediate filter bank responding to the morphological level of the final CNN convolutional block. Either Transformer attention rollout or Integrated Gradients should provide token-level/global explanations. These should be reviewed separately for true positives, true negatives, false positives, and false negatives. If the expert lesion masks are available, localization can be measured using Dice coefficient, intersection-over-union, or pointing-game accuracy; otherwise qualitative concordance must be labeled explicitly. Saliency maps should not be used for diagnosis.



Pre-specified interpretability audit: report agreement, disagreement, and failure modes; do not equate saliency with causal explanation.

Figure 4. Pre-specified XAI audit workflow; no synthetic patient heatmap is presented as an experimental result.

Statistics were calculated from verified image counts. Wilson 95% confidence intervals were used to estimate the normal image proportion. An image-level odds ratio (OR) was calculated and a 2x2 chi-square test was performed to examine the image-class composition across the magnification sets. The test is descriptive, and consequently likely in violation of the assumption of independence. In order to make inferences at the patient level, images would have to be linked to cases. When the models described above are carried out, primary diagnostic endpoints should focus on AUROC, sensitivity, specificity, positive predictive value, negative predictive value, F1-score, Matthews correlation coefficient, and balanced accuracy with bootstrap 95% confidence intervals. DeLong testing should be used to evaluate AUROC, and McNemar testing should be used to evaluate paired errors, both on the same held-out patients. Calibration should be evaluated using the Brier score and calibration plots with expected calibration error.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (TN + FP)$$

$$\text{Precision} = TP / (TP + FP)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

4. Results

4.1 Dataset composition and statistical analysis

The verified repository contains 1,224 images, 934 (76.31%) of which are OSCC and 290 (23.69%) are normal epithelium. With regard to normal images, the Wilson 95% CI is 21.40%-26.15%. For the 100x set, normal images represent 16.86% (95% CI: 13.90%-20.29%) and, for the 400x set, they represent 28.88% (95% CI: 25.64%-32.35%). Therefore, class imbalance is not the same at all levels of magnification. For image-level analysis at 2x2 in the 400x set, the result of chi square statistics was 23.35, p=1.35x10^-6. If we express these results in terms of odds, then for the 400x set, the ratio of OSCC to normal epithelium is 1/4 that of the 100x set (OR=2.00, 95% CI: 1.51-2.65). These results provide observable evidence of class imbalances at different levels of magnification from the verified repository, but cannot be interpreted to provide evidence of patient-level biological imbalance of classes due to the presence of nesting of images within patients.

4.2 Published same)-dataset CNN/Transformer benchmark

Chen et al (2024) evaluated four previously published models on a 131 image oral cancer dataset which consisted of 87 cancer images and 44 non-cancer images. On this dataset, CANet achieved the best average accuracy of 97.00%, sensitivity of 97.82%, specificity of 97.82%, and an F1-score of 96.61%. Swin Transformer achieved 94.95%, 95.37%, 95.52%, and 94.66%, respectively. Therefore, CANet outperformed Swin Transformer by 2.05 percentage points in accuracy and by 1.95 percentage points in F1-score. This comparison is useful, as the models were tested in the same study. However, the very small source dataset constrains the precision and generalization of the results, and since no new paired predictions are available, McNemar and DeLong statistics cannot be calculated.

Table 3. Published benchmark metrics reported by Chen et al. (2024)

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)
IDL-OSCDC	92.86	90.00	95.00	91.81
Swin Transformer	94.95	95.37	95.52	94.66
OIDCNN-OPMDD	96.30	97.37	97.37	95.71
CANet	97.00	97.82	97.82	96.61

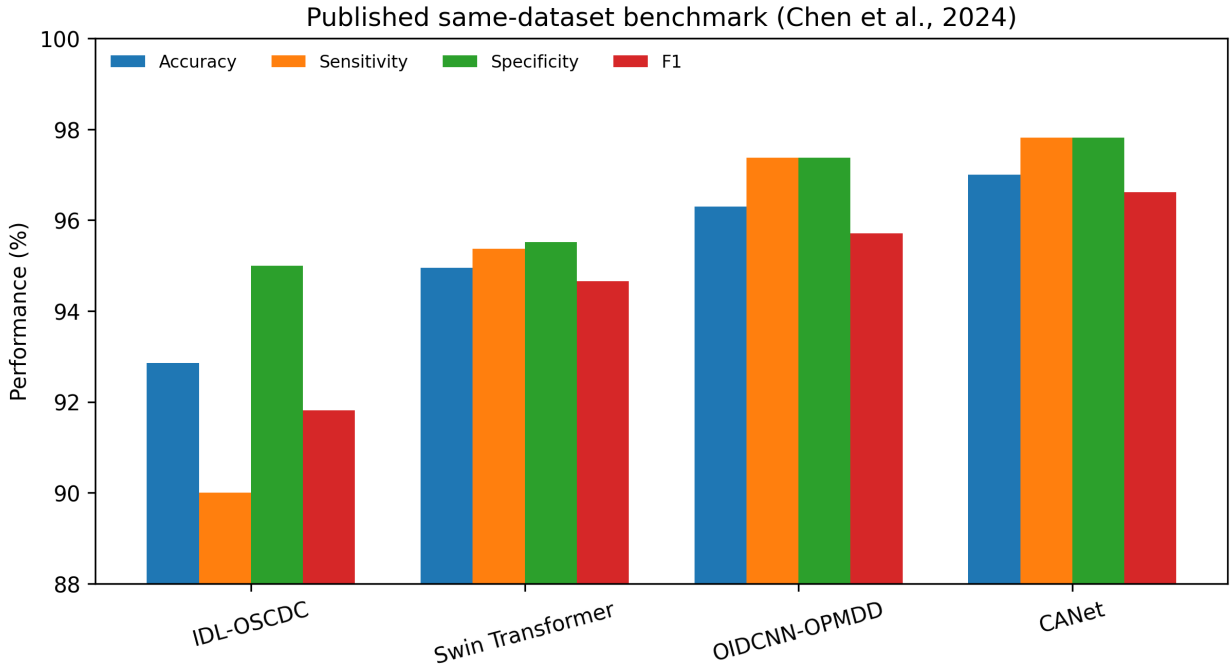


Figure 5. Same-study performance comparison reproduced from numerical results reported by Chen et al. (2024).

4.3 Published hybrid benchmark and secondary descriptive analysis

OralTransNet (Asif et al., 2025) is a recent addition to the external benchmarks for the hybrid design principle. The authors report 99.03% accuracy on a 5,143-image, seven-class mouth-and-oral disease dataset, 97.83% accuracy on a 241-image binary oral cancer dataset, and 94.23% accuracy on a 5,192-image binary histopathological oral cancer dataset. The mean accuracy across these three datasets was 97.03% (sample SD 2.50 percentage points; range 94.23%-99.03%). The accuracy on the histopathology dataset was 3.60 percentage points lower than on the binary oral cancer dataset (241 images), demonstrating that, among other things, (and perhaps most importantly) the complexity of the dataset and the modality can strongly influence performance. These values should not be combined as if they came from a single randomized comparison.

Table 4. Published OralTransNet dataset-level accuracy (Asif et al., 2025)

Dataset	Images	Classes	Published accuracy (%)
Mouth and oral diseases	5,143	7	99.03
Oral cancer	241	2	97.83
Histopathological oral cancer	5,192	2	94.23

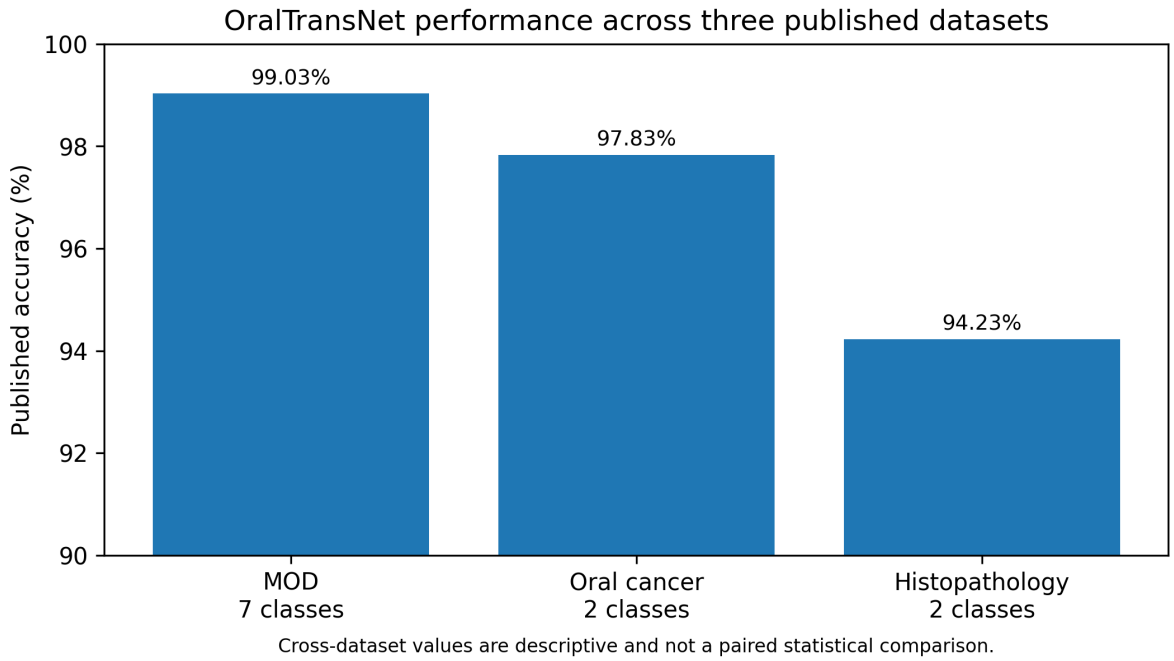


Figure 6. Published OralTransNet accuracies. Values are cross-dataset descriptive benchmarks, not paired tests.

4.4 Integrity of inferential claims

Table 5. What can and cannot be claimed from the available evidence

Analysis	Status in this paper	Interpretation
Dataset class/magnification statistics	Computed from verified repository counts	Real secondary data analysis; image-level clustering caveat applies
Published model metrics	Reanalysed from cited papers	Valid as reported benchmarks; not new training results
DeLong AUROC comparison	Not computed	Requires per-image probabilities on the same held-out cases
McNemar error comparison	Not computed	Requires paired binary predictions on identical cases
Calibration / Brier score	Not computed	Requires model probabilities
Grad-CAM / attention heatmaps	Protocol specified only	No synthetic heatmap is presented as patient evidence

5. Discussion

Two practical findings come from the empirical study of this paper. The first is that the verified Rahman dataset is imbalanced and class composition is affected by magnification. When this imbalance is present and no control is implemented, cues in the dataset might be learned as class shortcuts. Thus, on a biological level, an image-level OR of 2.00 is less meaningful. Instead, the number is more of a warning concerning the design of the experiment because stratification must be done at the level of diagnosis and magnification, and grouping must be done at the patient level, as opposed to random splitting of images. This concern mirrors the current advice regarding the reporting of medical-AI, and is also concerned with the issue of the provenance of the data and independence from one another of the various components of a system (Mongan et al., 2020).

The second empirical finding is that the currently available evidence does not support the dominance of any architecture. Chen et al. (2024) conclude that a coordinate-attention CNN outperformed the Swin Transformer on a small clinical dataset. Flugge et al. (2023), Maia et al. (2024), and Asif et al. (2025) conclude that components of Transformers can outperform when the dataset and architecture are appropriate. Based on the available literature, the best interpretation is that complementary architectures can be used in combination, as CNNs provide locality and induce bias, while Transformers provide global context.

Interpretability must be treated with the same rigor as classification. A Grad-CAM overlay may look compelling, but it may be unstable, diffuse, or sensitive to non-lesional artifacts. Brain attention weights do not guarantee explanations, either. A dual-view audit can be useful as

long as it asks a narrow question: do independent attribution schemes focus on plausible lesion regions, and do they behave differently in false-negative/false-positive cases? Visual plausibility can be converted into localization validity with the help of expert review and, if possible, spatial annotations (Figueroa et al., 2022; Ghassemi et al., 2021).

In the clinical setting, sensitivity deserves the most attention, since a false-negative screen can delay necessary referral, and specificity controls the cost of unnecessary biopsies and consultations. However, a high level of discrimination is insufficient. A model with high AUROC can be poorly calibrated, and its performance can deteriorate because of domain shift from different stains, scanners, or cameras. For this reason, several factors should be addressed before deployment, including prospective multicenter validation, uncertainty estimation, calibration, subgroup analysis and evaluation of the operational impact of the system. The system should be clearly presented as an assistive screening or decision-support system, since definitive OSCC diagnosis remains the responsibility of the specialist and a histopathological examination.

6. Limitations

This study prefers verifiability over simulated completeness. The raw Rahman image archive was not trained in the current execution environment, and the available repository description did not provide a case-level mapping sufficient to construct a verified patient-disjoint split. Therefore, no original confusion matrix, AUROC, calibration curve, Grad-CAM heatmap, DeLong test, McNemar test, or ablation result is claimed. Published benchmarks differ in modality and dataset composition, and should not be meta-analytically pooled without raw predictions or a formal review protocol. The image-level chi-square analysis also treats images as observations despite within-patient clustering, so it is descriptive rather than patient-level inferential evidence.

7. Conclusion

A clinical-grade oral cancer detection AI system requires more than high accuracies. Independent testing sets is a necessity. Along with this, pairwise statistical analysis, calibration, interpretable failure analysis, and prevention against data leakage are necessary. The Rahman dataset contains 1,224 images with predominantly OSCC. Further analysis of the dataset suggests a class imbalance developing with the enlargement of the dataset. This suggests the laptop is susceptible to learning shortcuts. Published literature shows the CANet and Swin Transformer models achieving better results on a clinical dataset using the OralTransNet model. The OralTransNet model, when combined with the CANet model, will be helpful for achieving the objectives of this architecture. The next step in the development is an external patient dataset used to lock the testing dataset, along with a confidence interval based on the DeLong and McNemar tests. In the absence of such a study, the system should be positioned as a research framework for decision support, not as a substitute for a clinical biopsy or specialist examination.

Future Directions

Considering the architecture, the first priority is to obtain and use multicenter patient-disjoint cohorts and include external stains and scanner domains. This should be followed by experiments to remove different components of the architecture and assess the impact of uncertainty and calibration. Also, pathologist-defined regions should be used to assess the model's salience. As a longer-term goal, consider self-supervised learning and combine histopathology images with clinical images to enhance the system. A departure from the research framework should only occur if the system has been proven to be safe for deployment on mobile devices.

References

Albalawi, E., Thakur, A., Ramakrishna, M. T., Bhatia Khan, S., SankaraNarayanan, S., Almarri, B., & Hadi, T. H. (2024). Oral squamous cell carcinoma detection using EfficientNet on histopathological images. Frontiers in Medicine, 10, 1349336. https://doi.org/10.3389/fmed.2023.1349336

Alanazi, A. A., Khayyat, M. M., Khayyat, M. M., Elamin Elnaim, B. M., & Abdel-Khalek, S. (2022). Intelligent deep learning enabled oral squamous cell carcinoma detection and classification using biomedical images. Computational Intelligence and Neuroscience, 2022, 7643967. https://doi.org/10.1155/2022/7643967

Ananthakrishnan, B., Shaik, A., Kumar, S., Narendran, S. O., Mattu, K., & Kavitha, M. S. (2023). Automated detection and classification of oral squamous cell carcinoma using deep neural networks. Diagnostics, 13(5), 918. https://doi.org/10.3390/diagnostics13050918

Asif, S., Wang, V. Y., & Xu, D. (2025). OralTransNet: A novel hybrid model integrating transformer attention and CNN features for accurate diagnosis of mouth and oral diseases. Engineering Applications of Artificial Intelligence, 159, 111609. https://doi.org/10.1016/j.engappai.2025.111609

Ahuja R, Agarwal BB, Singh K. Leveraging Transfer Learning for the Early Diagnosis of Autism in Infants via Facial Feature Classification. Int J Drug Deliv Technol. 2026;16(39s): 787-802. DOI: 10.25258/ijddt.16.39s.98

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. Information, 11(2), 125. https://doi.org/10.3390/info11020125

Chang, X., Yu, M., Liu, R., Jing, R., Ding, J., Xia, J., et al. (2023). Deep learning methods for oral cancer detection using Raman spectroscopy. Vibrational Spectroscopy, 126, 103522. https://doi.org/10.1016/j.vibspec.2023.103522

Chen, R., Wang, Q., & Huang, X. (2024). Intelligent deep learning supports biomedical image detection and classification of oral cancer. Technology and Health Care, 32(1_suppl). https://doi.org/10.3233/THC-248041

Das, M., Dash, R., & Mishra, S. K. (2023). Automatic detection of oral squamous cell carcinoma from histopathological images of oral mucosa using deep convolutional neural network. International Journal of Environmental Research and Public Health, 20(3), 2131. https://doi.org/10.3390/ijerph20032131

Deif, M. A., Attar, H., Amer, A., Elhaty, I. A., Khosravi, M. R., & Solyman, A. A. A. (2022). Diagnosis of oral squamous cell carcinoma using deep neural networks and binary particle swarm optimization on histopathological images: An AIoMT approach. Computational Intelligence and Neuroscience, 2022, 6364102. https://doi.org/10.1155/2022/6364102

Elmakaty, I., et al. (2022). Accuracy of artificial intelligence-assisted detection of oral squamous cell carcinoma: A systematic review and meta-analysis. Critical Reviews in Oncology/Hematology, 178, 103777. https://doi.org/10.1016/j.critrevonc.2022.103777

Fatapour, Y., Abiri, A., Kuan, E. C., & Brody, J. P. (2023). Development of a machine learning model to predict recurrence of oral tongue squamous cell carcinoma. Cancers, 15(10), 2769. https://doi.org/10.3390/cancers15102769

Ferro, A., Kotecha, S., & Fan, K. (2022). Machine learning in point-of-care automated classification of oral potentially malignant and malignant disorders: A systematic review and meta-analysis. Scientific Reports, 12, 13797. https://doi.org/10.1038/s41598-022-17489-1

Figueroa, K. C., Song, B., Sunny, S., et al. (2022). Interpretable deep learning approach for oral cancer classification using guided attention inference network. Journal of Biomedical Optics, 27(1), 015001. https://doi.org/10.1117/1.JBO.27.1.015001

Flugge, T., Gaudin, R., Sabatakakis, A., Troltzsch, D., Heiland, M., van Nistelrooij, N., et al. (2023). Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. Scientific Reports, 13, 2296. https://doi.org/10.1038/s41598-023-29204-9

Fu, Q., Chen, Y., Li, Z., Jing, Q., Hu, C., Liu, H., et al. (2020). A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. EClinicalMedicine, 27, 100558. https://doi.org/10.1016/j.eclinm.2020.100558

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745-e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Han, K., Wang, Y., Chen, H., et al. (2022). A survey on vision transformer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1), 87-110. https://doi.org/10.1109/TPAMI.2022.3152247

Haq, U., Ahmed, M., Assam, M., Ghadi, Y. Y., & Algarni, A. (2023). Unveiling the future of oral squamous cell carcinoma diagnosis: An innovative hybrid AI approach for accurate histopathological image analysis. IEEE Access, 11, 118281-118290. https://doi.org/10.1109/ACCESS.2023.3326152

Hunter, B., Hindocha, S., & Lee, R. W. (2022). The role of artificial intelligence in early cancer diagnosis. Cancers, 14(6), 1524. https://doi.org/10.3390/cancers14061524

Khanagar, S. B., et al. (2023). Application and performance of artificial intelligence (AI) in oral cancer diagnosis and prediction using histopathological images: A systematic review. Biomedicines, 11(6), 1612. https://doi.org/10.3390/biomedicines11061612

Li, J., Kot, W. Y., McGrath, C. P., Chan, B. W. A., Ho, J. W. K., & Zheng, L. W. (2024). Diagnostic accuracy of artificial intelligence assisted clinical imaging in the detection of oral potentially malignant disorders and oral cancer: A systematic review and meta-analysis. International Journal of Surgery, 110(8), 5034-5046. https://doi.org/10.1097/JS9.0000000000001469

Liu, Z., Lin, Y., Cao, Y., et al. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012-10022. https://doi.org/10.1109/ICCV48922.2021.00986

Maia, B. M. S., de Assis, M. C. F. R., de Lima, L. M., Rocha, M. B., Calente, H. G., Correa, M. L. A., & Krohling, R. A. (2024). Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. Expert Systems with Applications, 241, 122418. https://doi.org/10.1016/j.eswa.2023.122418

Mongan, J., Moy, L., & Kahn, C. E., Jr. (2020). Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A guide for authors and reviewers. Radiology: Artificial Intelligence, 2(2), e200029. https://doi.org/10.1148/ryai.2020200029

Nagarajan, B., Chakravarthy, S., Venkatesan, V. K., Ramakrishna, M. T., Khan, S. B., Basheer, S., et al.

(2023). A deep learning framework with an intermediate layer using the swarm intelligence optimizer for diagnosing oral squamous cell carcinoma. Diagnostics, 13(22), 3461. https://doi.org/10.3390/diagnostics13223461

Panigrahi, S., Nanda, B. S., Bhuyan, R., Kumar, K., Ghosh, S., & Swarnkar, T. (2023). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. Heliyon, 9(3), e13444. https://doi.org/10.1016/j.heliyon.2023.e13444

Pereira-Prado, V., et al. (2023). Artificial intelligence for image analysis in oral squamous cell carcinoma: A review. Diagnostics, 13(14), 2416. https://doi.org/10.3390/diagnostics13142416

Rahman, A. U., Alqahtani, A., Aldhafferi, N., Nasir, M. U., Khan, M. F., Khan, M. A., et al. (2022). Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. Sensors, 22(10), 3833. https://doi.org/10.3390/s22103833

Rahman, T. Y., Mahanta, L. B., Das, A. K., & Sarma, J. D. (2020). Histopathological imaging database for oral cancer analysis. Data in Brief, 29, 105114. https://doi.org/10.1016/j.dib.2020.105114

Rahman, T. Y., Mahanta, L. B., Das, A. K., & Sarma, J. D. (2020). Study of morphological and textural features for classification of oral squamous cell carcinoma by traditional machine learning techniques. Cancer Reports, 3, e1293. https://doi.org/10.1002/cnr2.1293

Rahman, T. Y., Mahanta, L. B., Das, A. K., & Sarma, J. D. (2023). Histopathological imaging database for Oral Cancer analysis (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/ftmp4cvtmb.2

Sahoo, R. K., et al. (2024). 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, 1494867. https://doi.org/10.3389/froh.2024.1494867

Singh, A., Bhushan, B., (2026) "AI and Machine Learning Enabled Enterprise Cloud Computing for Digital Transformation"" at Dandao Xuebao/Journalof Ballistics,DOI: https://doi.org/10.52783/dxjb.v38.333 , ISSN:

1004-499X. Vol. 38 No. 2, pp169 – 202

Sukegawa, S., Ono, S., Tanaka, F., Inoue, Y., Hara, T., Yoshii, K., et al. (2023). Effectiveness of deep learning classifiers in histopathological diagnosis of oral squamous cell carcinoma by pathologists. Scientific Reports, 13, 11676. https://doi.org/10.1038/s41598-023-38343-y

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2020.3027314

Verma, P., Agarwal, B.B., Gupta N., (2024).Predicting Chronic Kidney Disease and Coronary Heart Disease Risks in NIDDM Patients Using CNNAdam: A Deep Learning Approach. 2024 4th International Conference on Technological

Advancements in Computational Sciences (ICTACS), 994-1000.https://doi.org/10.1109/ICTACS62700.2024.10 8405

44

Verma P, Agarwal BB, Gupta N. Proposed Methods for Diabetes Detection in Medical Decision Making. Int J Drug Deliv Technol. 2026;16(28s):996-1007. DOI: 10.25258/ijddt.16.28s.117

Warin, K., & Suebnukarn, S. (2024). Deep learning in oral cancer - a systematic review. BMC Oral Health, 24, 212. https://doi.org/10.1186/s12903-024-03993-5

Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., et al. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. IEEE Access, 8, 132677-132693. https://doi.org/10.1109/ACCESS.2020.3010180

Wu, Y., Koyuncu, C. F., Toro, P., Corredor, G., Feng, Q., Buzzy, C., et al. (2022). A machine learning model for separating epithelial and stromal regions in oral cavity squamous cell carcinomas using H&E-stained histology images: A multi-center, retrospective study. Oral Oncology, 131, 105942. https://doi.org/10.1016/j.oraloncology.2022.105942

Yang, Z., Pan, H., Shang, J., Zhang, J., & Liang, Y. (2023). Deep-learning-based automated identification and visualization of oral cancer in optical coherence tomography images. Biomedicines, 11(3), 802. https://doi.org/10.3390/biomedicines11030802

Yu, M., Ding, J., Liu, W., Tang, X., Xia, J., Liang, S., et al. (2023). Deep multi-feature fusion residual network for oral squamous cell carcinoma classification and its intelligent system using Raman spectroscopy. Biomedical Signal Processing and Control, 86, 105339. https://doi.org/10.1016/j.bspc.2023.105339