

CLARA-OSCC: A Calibrated Lesion-Aware Resolution-Adaptive Pipeline for Oral Squamous Cell Carcinoma Diagnosis

V. Gokula Krishnan^{1,2}, N. Venkatesh³, M. Abirami⁴, Shanker M.C⁵, K. Hema Priya⁶, V. Vijayaraja⁷

¹Post-Doctoral Research Fellow, Department of Computer Science Engineering, SR University, Warangal, Telangana, India.

²Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India. ³School of CS & AI, SR University, Warangal - 506371, Telangana, India. ⁴Department of CSE, Panimalar Engineering College, Chennai, Tamil Nadu, India. ⁵Department of Biomedical Engineering, VelTech MultiTech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamil Nadu, India. ⁶Department of CSD, Easwari Engineering College, Chennai, Tamil Nadu, India. ⁷Department of AIDS, RMK College of Engineering and Technology, Kavaraipettai, Tamil Nadu, India.

Abstract

Background: Early diagnosis of Oral Squamous Cell Carcinoma (OSCC) from histopathological images is challenging because of limited and imbalanced datasets, variations in image magnification (100× and 400×), and calibration degradation associated with site-specific staining and scanner shifts. This study proposes CLARA-OSCC, a magnification-aware framework designed to improve robust classification, calibration, and clinically relevant triage of OSCC under heterogeneous imaging conditions. **Materials and Methods:** CLARA-OSCC integrates robust image standardization, a multi-resolution texture-spectral lattice, a resolution gate, a lightweight SE-MobileViT backbone, gated cross-modal feature fusion, class-balanced focal learning, and domain alignment using CORAL. Post-hoc temperature scaling is employed to improve prediction calibration. The framework is evaluated using internally stratified sub-samples of a publicly available OSCC histopathological image dataset, with separate assessment at 100× and 400× magnifications. Performance is assessed using Accuracy, F1-score, AUROC, AUPRC, Brier score, Expected Calibration Error (ECE), Sensitivity, Specificity, PPV, and NPV. **Results:** CLARA-OSCC achieved an Accuracy of 0.958 ± 0.024 , F1-score of 0.959 ± 0.026 , AUROC of 0.985 ± 0.009 , and AUPRC of 0.978 ± 0.012 at the 95% confidence interval. The framework obtained a Brier score of 0.033 and ECE of 0.021, representing improvements of 0.011 and 0.015, respectively, over the baseline. At the selected clinical threshold of $\theta = 0.80$, the framework achieved 85% coverage, with Sensitivity of 0.951, Specificity of 0.969, PPV of 0.971, and NPV of 0.948. Performance was higher at 400× magnification, with an F1-score of 0.966 and ECE of 0.018, compared with 0.949 and 0.027, respectively, at 100× magnification. Reliability analysis demonstrated reduced overconfidence following calibration, while prototype-based retrieval provided case-level explanations. **Conclusion:** The results indicate that magnification-aware feature fusion combined with explicit calibration can provide robust and triage-oriented predictions for OSCC histopathological images. The framework also provides mechanisms for handling simulated domain variation, uncertainty, and abstention-based routing. Although the site-level analysis used internally stratified sub-samples of a publicly available dataset, the observed robustness provides motivation for future multicentre external validation before clinical deployment.

Keywords: Oral Squamous Cell Carcinoma, Heterogeneous magnifications, Empirical Wavelet Transform

Asian Pac J Cancer Biol, **11** (4), 1115-1126

Submission Date: 07/11/2026 Acceptance Date: 09/10/2026

Introduction

Among populations with low or medium socioeconomic levels, oral squamous cell carcinoma (OSCC) ranks highest in terms of malignant neoplasms of the mouth and is a leading cause of cancer-related mortality [1].

The diagnostic gold standard is histopathology, but manual evaluation is laborious, can be influenced by differences in staining and scanner settings, and is vulnerable to observer variability, particularly in areas with inflammation or at

Corresponding Author:

Dr. V. Gokula Krishnan

Post-Doctoral Research Fellow, Department of Computer Science Engineering, SR University, Warangal, Telangana, India.

Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India.

Email: gokul_kris143@yahoo.com

the invasive front [2]. In order to help with early detection, researchers have started using artificial intelligence (AI) to convert high-dimensional visual patterns [3] into calibrated probabilities that can be used for screening or triage [4].

Three hurdles consistently restrict performance in the actual world, even when there has been progress. For starters, models are more likely to overfit and produce overly confident probabilities when datasets are small and skewed [5]. In addition, the histopathology of OSCC is recorded at different magnifications (such as 100× and 400×) so that diagnostically important features [6] can be seen at different sizes, such as keratin pearls, epithelial organization, nuclear pleomorphism, and chromatin texture. Both are usually better served by separate receptive-field strategies. Finally, when models are used outside of their training distribution, discrimination and calibration are negatively affected by domain shifts caused by cross-site variability (scanner optics, staining, compression) [7-8].

New research sheds light on certain aspects of this image. According to controlled comparisons, CNN accuracy and convergence are first-order drivers of normalization [9]. While spectral descriptors such as 2D Empirical Wavelet Transform (EWT) provide frequency-localized views of collagen and epithelium periodicities, texture-centric approaches like Modified Local Binary Patterns (MLBP) capture micro-architectural signals at 100× and 400× [10]. For point-of-care applications, lightweight CNN-Transformer hybrids are a viable choice; examples include MobileViT with Squeeze-and-Excitation, which combines local edge sensitivity with long-range context [11]. Clinical adoption relies heavily on temperature scaling for calibration and case-based reasoning for explainability [12].

Based on these findings, to present CLARA-OSCC, a pipeline that is customized to the OSCC images dataset and is Calibrated, Lesion-Aware, and Resolution-Adaptive. To start, the process uses per-channel min-max normalization and robust outlier attenuation to bring intensity distributions into harmony without removing nuclei contrast [13]. Descriptors that are particularly useful when data is scarce are produced using a multi-resolution texture-spectral lattice, which recovers MLBP histograms at radii suitable for 100× and 400× and calculates EWT band energies with directional anisotropy. At 400×, a SE-MobileViT backbone emphasizes fine detail, whereas at 100×, larger context is highlighted, thanks to a Resolution Gate that adjusts feature blending based on magnification. To train the network to prioritize hand-crafted cues over deep embeddings, to execute gated cross-modal fusion on each image. Model regularization promotes anatomically meaningful neighborhoods, class-balanced focal loss decreases imbalance, and CORAL reduces cross-site covariance mismatch during training. Finally, an abstention policy sends cases with low confidence to an expert for examination, and temperature scaling makes confidence and accuracy go hand in hand.

To report discrimination (AUROC/AUPRC), accuracy (F1), and calibration (Brier/ECE) with 95% confidence

intervals in our evaluation that follows slide-level stratified cross-validation and a held-out test set. To additionally separate the effects of MLBP, EWT, the Resolution Gate, CORAL, and prototype loss through ablations, and to analyze generalization by breaking it down into per-magnification and per-site performance. Our coverage-accuracy-utility study across thresholds finds a clinically efficient operating point that balances effort and safety, which may be used to assist deployment decisions.

Here is the breakdown of the remaining sections of the paper: In Section 2, the relevant literature is reviewed. In Section 3, the technique that has been suggested is laid out in great depth. Section 4 delves into the analysis of the results. Finally, Section 5 draws the conclusion.

Related Work

Data preparation, learnt and handcrafted representations, model architecture design, clinical explainability, and oral cancer screening pipelines have all been steadily improving in recent studies. Taken as a whole, these studies investigate three performance levers that crop up time and time again: (i) the statistical effects of preprocessing decisions, (ii) the design of features across resolutions and modalities, and (iii) deployable architectures with transparent decision support.

By comparing missing-value imputation, outlier identification, and normalization in a systematic manner, Patel & Kumar [14] were able to quantify the effects of preprocessing on discriminative accuracy and mean-squared error (MSE). By applying min-max scaling to the same data and models, CNNs achieved the best accuracy (94% with a minimum standard error of 0.013), followed by outlier handling (93% with a minimum standard error of 0.014), and finally imputation (92% with a minimum standard error of 0.015). The necessity to normalize dynamic ranges prior to meddling with architecture is further reinforced by this controlled comparison, which identifies normalization as a first-order driver of convergence and calibration for image-based CNNs.

With careful crafting, texture-centric designs can stay competitive. In their study, Yaduvanshi et al. [15] presented a modified local binary pattern (MLBP) that was specifically designed to detect connections in OSCC histopathology at 100× and 400× magnification, as well as in local tissue structures. When comparing SVM, KNN, decision tree, and a deep CNN head, the strongest classical baseline on accuracy/precision/recall/F1 was produced by MLBP combined with SVM, while the hybrid MLBP-DCNN also achieved high results (91.36% at 100×; 94.44% at 400×). This study shows that generic convolutional neural networks (CNNs) could miss diagnostically important micro-patterns by not using resolution-aware analysis and multiscale texture coding on a Mendeley collection of 528 OSCC and 696 normal photos (400×).

Cimino et al. [16] tackle explainability and workflow integration head-on by combining a refreshed Faster R-CNN-FPN detector/classifier with CBR, a protocolized set of XAI tools codified in BPMN, and case-based

reasoning (CBR). Utilizing 160 instances spanning three types of ulcers, the detector was fine-tuned to an accuracy of 83% detection and 92% multiclass (with a rise to 98% for neoplastic vs non-neoplastic cases) in its predictions. The methodological contribution goes beyond simple measurements; it includes a validated prospective protocol, post hoc exemplar-based reasoning that is in line with clinical review, and validation of the technique with experts and residents. This line of study is better positioned for multi-center replication and iterative improvement with the public distribution of systems and cases.

Additionally, point-of-care imaging has seen the emergence of lightweight transformer-CNN hybrids. By combining a MobileViT encoder for global context with a compact CNN backbone enhanced by Squeeze-and-Excitation, LightSE-MobileViT was suggested by Kabir et al. [17]. The model achieved better results than CST-CNN, MobileNetV2, DenseNet121, and InceptionV3 when trained on a clinically selected set of lip/tongue photos (87 cancer, 44 non-cancer), which was then enlarged to 981 balanced images by augmentation. Its reported metrics included 98.39% accuracy, near-perfect precision/recall, an $F1 \approx 0.98$, and ROC-AUC=1.00. However, due to the limited initial sample size and extensive augmentation, external validation is crucial to prevent overfitting and distribution shift, even though the advantages are noticeable.

When it came time to classify histopathological samples as normal or OSCC, Deo et al. [18] used spectral-deep hybrids to extract 2D empirical wavelet transform (EWT) features and assemble ResNet50 and DenseNet201. This method achieved an accuracy of 0.92 using 696 images and an 80/10/10 train/val/test split; sensitivity/specificity analyses showed balanced error control. Findings like these demonstrate how useful frequency-localized descriptors are for deep ensembles as supplementary channels.

Several commonalities emerge from these investigations when considered collectively. To begin, standardized normalization reliably improves CNN performance at a cheap cost and with a high yield [14]. Secondly, multi-resolution texture modelling is still useful, particularly in histopathology, whether it's a deep hybrid (MLBP-DCNN) or a strong classical baseline (MLBP+SVM) [15]. Third, CBR+DL pipelines demonstrate promise in the domain of transparent reasoning and protocolized explanation, which are crucial for clinically relevant adoption [16]. Fourth, advanced accuracy on limited clinical datasets can be achieved by lightweight hybrids that combine attention with efficient convolutions. However, it is important to stress-test these models on external cohorts [17]. Spectral features, like EWT, can enhance deep ensembles, making them more robust without being overly complex [18]. Some long-standing deficiencies include insufficient cross-site generalizability, a lack of information on calibration beyond accuracy, small sample sizes that are often improved through augmentation, and a dearth of multiclass settings that simulate real-world differential

diagnosis. To tackle this, to need more comprehensive and standardized datasets that cover several imaging modalities (clinical photographs, endoscopy, histology), thorough evaluations of calibration and uncertainty, and prospective clinical trials that incorporate transparent and auditable decision support.

Problem statement

The current state of histopathology image analysis for early OSCC identification is hindered by imbalanced and small datasets, uneven preprocessing that lowers calibration and generalization, and varied magnifications ($100\times/400\times$). There is a lack of external validation and case-based explanations, despite previous research showing that normalization is crucial, texture codes (MLBP) and spectral cues (EWT) are discriminative, and lightweight CNN-Transformer hybrids perform well on small datasets. On the OSCC images dataset, to present the challenge of developing a calibrated, explainable, and robust OSCC classifier that (i) exhibits cross-site performance with uncertainty reporting, (ii) standardizes preprocessing, and (iii) integrates multi-resolution texture/spectral features with deep representations. Additionally, to aim to mitigate imbalance.

Materials and Methods

Proposed Framework: CLARA-OSCC-A Calibrated, Lesion-Aware, Resolution-Adaptive Architecture for OSCC Diagnosis

This section presents CLARA-OSCC, a novel end-to-end pipeline that solves the problem posed in Section 2 by unifying rigorous preprocessing, multi-resolution texture-spectral descriptors, a lesion-aware resolution-adaptive backbone, calibrated training for imbalance, and case-based explanations. Throughout, to embed what each component computes, why it matters for OSCC images, how it is implemented mathematically, and where it sits in the overall flow. To restrict the exposition to the seven most impactful modules. Unless otherwise stated, to consider the OSCC images dataset comprising normal epithelium and OSCC histopathology acquired at $100\times$ and $400\times$ magnification, acknowledging modest class imbalance and cross-site variability. The workflow of the proposed model is mentioned in Figure 1.

Data Standardization & Quality Harmonization

Histopathology images exhibit staining variability, scanner-dependent intensity ranges, magnification heterogeneity ($100\times/400\times$), and occasional artifacts. These factors distort the pixel distribution seen by learning algorithms, degrading both discrimination and calibration. To therefore establish a two-stage standardization: robust outlier mitigation followed by per-channel min-max normalization. This sequence preserves morphological cues while making the optimization landscape smooth and comparably scaled across slides and magnifications.

Let an RGB image be $I \in \mathbb{R}^{H \times W \times 3}$. For each channel $c \in \{1, 2, 3\}$, to compute per-image robust location (med) and scale (median absolute deviation, MAD). To define

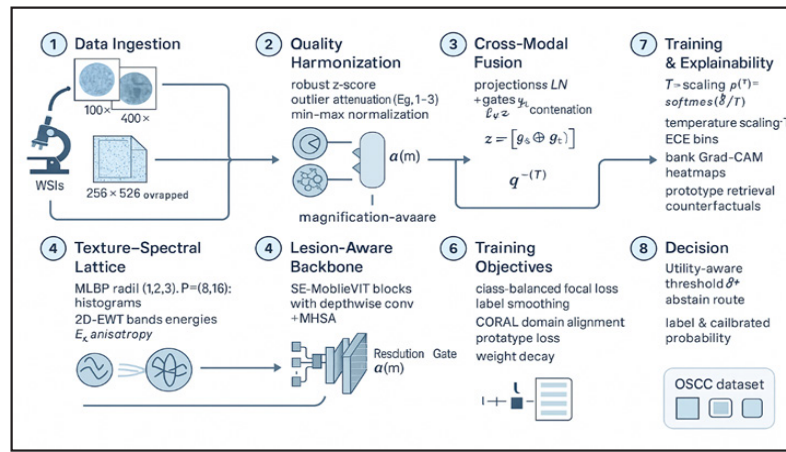


Figure 1. Workflow of the Proposed Model

a robust z-score and then clamp extreme outliers before min-max normalization.

First, robust z-score:

$$Z_{i,c} = \frac{I_{i,c} - \text{med}(I_{i,c})}{\max(\text{MAD}(I_{i,c}), \epsilon)} \quad (1)$$

where i indexes pixels, and $\epsilon > 0$ prevents division by zero. Outlier attenuation uses a soft-clamp $\phi(\cdot)$ (Huber-like):

$$\tilde{I}_{i,c} = \text{med}(I_{i,c}) + \sigma_c \cdot \phi(Z_{i,c}), \phi(z) = \begin{cases} z, & |z| \leq \kappa \\ \kappa \text{sign}(z), & \text{otherwise} \end{cases} \quad (2)$$

with $\sigma_c = \max(\text{MAD}, \epsilon)$ and $\kappa \in [2, 3]$ (empirically 2.5). Finally, min-max normalization per image and channel:

$$I'_{i,c} = \frac{I_{i,c} - \min(I_{i,c})}{\max(I_{i,c}) - \min(I_{i,c}) + \epsilon} \in [0, 1] \quad (3)$$

This standardized image I' feeds both handcrafted feature extractors (texture-spectral) and the deep backbone. The min-max step aligns with evidence that normalization is a first-order driver of CNN performance; the robust pre-step avoids suppressing diagnostically salient nuclei edges by blunt clipping.

Multi-Resolution Texture-Spectral Feature Lattice (MLBP + 2D-EWT)

Squamous epithelium malignancy manifests as keratin pearl formation, nuclear pleomorphism, and textural disorganization at different magnifications. To explicitly encode these micro-textures and frequency patterns, to construct a feature lattice combining Modified Local Binary Patterns (MLBP) at multiple radii and the 2D Empirical Wavelet Transform (EWT). This lattice complements learned features, improving robustness on small datasets and stabilizing training.

For a pixel at location x , radius R , and P neighbors, MLBP compares neighbors to a learned local threshold $\theta(x)$ (e.g., adaptive mean or median in a small window) rather than the center value, improving resilience to illumination variations:

$$\text{MLBP}_{P,R}(x) = \sum_{p=0}^{P-1} \mathbf{1}[I'_R(x, p) - \theta(x) \geq 0] 2^p \quad (4)$$

where $I'_R(x, p)$ samples I' at radius R and angle $2\pi p/P$. To aggregate over scales $R \in \{1, 2, 3\}$ and neighborhood sizes $P \in \{8, 16\}$, constructing histograms $h_{R,P}$ per patch and concatenating them. To preserve spatial context, to compute second-order co-occurrence histograms over a grid and apply power-law normalization $h \leftarrow h^\alpha / \|h^\alpha\|$, with $\alpha \in (0, 1)$ to dampen bursty codes.

2D-EWT builds adaptive wavelet bands by segmenting the Fourier spectrum along energy ridges; let $\{\mathcal{B}_k\}_{k=1}^K$ the learned bands with empirical wavelets ψ_k . The EWT coefficients:

$$W_k = I' * \psi_k, \text{ and band energies } E_k = \frac{1}{HW} \sum_{u,v} W_k(u, v)^2 \quad (5)$$

To also compute anisotropy descriptors by splitting orientations for each band (e.g., four quadrants) and measuring directional energy imbalance.

The texture-spectral lattice yields a vector tsR^{dt} (MLBP) and seR^{ds} (EWT). These are projected to the deep feature space via learned linear layers and fused with the backbone in Section 3.4. By design, these descriptors remain magnification-aware: MLBP radii emphasize micro-patterns at 400 $\times$, while lower-radius, larger-cell histograms capture coarser 100 $\times$ morphology.

Lesion-Aware Resolution-Adaptive Backbone (SE-MobileViT with Resolution Gate)

CNNs excel at local motif extraction; transformers capture long-range glandular/keratin structures. Lightweight hybrids are desirable for deployability. To therefore adopt a SE-MobileViT backbone and augment it with a Resolution Gate that conditions feature propagation on magnification (when known) or on a learned scale proxy (when unknown), ensuring the network adapts where the lesion patterns are most informative.

Convolutional Stem with SE: Let $F^{(l)} \in \mathbb{R}^{H \times W \times C_l}$ be the feature map at layer l . A convolutional block (Conv-BN-ReLU) yields $U^{(l)}$. A Squeeze-and-Excitation recalibration computes channel attention:

$$s^{(l)} = \sigma(W_2^{(l)} \text{ReLU}(W_1^{(l)} \text{GAP}(U^{(l)}))), \tilde{F}^{(l)} = s^{(l)} \odot U^{(l)} \quad (6)$$

with global average pooling GAP, sigmoid σ , and $\odot$ channel-wise product. SE emphasizes diagnostically salient channels (e.g., basophilic nuclei cues) without heavy compute.

MobileViT Blocks: To tile the feature map into non-overlapping patches, project to tokens, and apply multi-head self-attention (MHSA) interleaved with convolutional mixing to preserve locality:

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

where $Q=FW_Q$, $K=FW_K$, $V=FW_V$. The MobileViT block alternates MHSA with depthwise separable convolutions, giving global context (e.g., across epithelial layers) and edge-sharpness (nuclear boundaries).

Resolution Gate

To adapt to $100\times$ vs $400\times$, to inject a gating scalar (or vector) $\alpha \in [0,1]$ that scales receptive-field aggregation. If magnification $m \in \{100,400\}$ is provided, α is a deterministic mapping; otherwise a scale predictor m is inferred from I' via a small CNN, leading to:

$$\alpha = \sigma(w_m^T \phi(m)), F^{(l+1)} = (1-\alpha)\tilde{F}^{(l)} + \alpha \text{MobileViT}(\tilde{F}^{(l)}) \quad (8)$$

Intuitively, higher α directs more global reasoning (beneficial at $100\times$), while lower α preserves fine textural detail (critical at $400\times$). This where-aware mixture is trained end-to-end.

The backbone outputs a compact embedding $f \in \mathbb{R}^d$ at the penultimate layer, which is the anchor for fusion (Section 3.4), calibration and classification (Section 3.5), and prototype reasoning (Section 3.6).

Cross-Modal Feature Fusion & Representation Learning

To now join three complementary sources: deep embedding f , texture vector t , and spectral vector s . Naïve concatenation risks overfitting and mis-calibration, especially with small datasets. To therefore align, gate, and project the descriptors into a common space before classification, so the model learns how much to trust each cue given the image context.

How (Projections & Gates). To first project:

$$\tilde{t} = \text{LN}(W_t t), \tilde{s} = \text{LN}(W_s s), \tilde{f} = \text{LN}(W_f f) \quad (9)$$

with layer normalization LN and learned matrices W . To compute contextual gates via a cross-attention-inspired mechanism:

$$g_t = \sigma(\tilde{f}^T A_t \tilde{t}), g_s = \sigma(\tilde{f}^T A_s \tilde{s}) \quad (10)$$

so that if the deep features already encode texture strongly, g_t is reduced, preventing redundancy.

How (Fusion). The fused representation is:

$$z = \text{LN}(\tilde{f} \oplus g_t \tilde{t} \oplus g_s \tilde{s}) W_z \quad (11)$$

where $\oplus$ concatenates and W_z is a final projection. A residual MLP head with dropout predicts logits $l \in \mathbb{R}^K$ (for $K=2$: normal vs OSCC).

This design allows the network to learn when to emphasize multi-resolution MLBP (e.g., keratin pearls) or EWT (e.g., periodic collagen textures) and how to temper them when MobileViT already captures global context.

Imbalance Mitigation, Uncertainty, and Calibration-Aware Training

Small clinical datasets often have class imbalance and site-specific shifts; overconfident models harm clinical trust. To incorporate: (i) class-balanced focal loss with label smoothing, (ii) domain alignment to reduce cross-site drift, and (iii) temperature scaling for post-hoc calibration. To also track Expected Calibration Error (ECE) during validation to navigate the calibration-accuracy frontier.

Class-Balanced Weights: For class k with n_k samples, define the effective-number weight:

$$w_k = \frac{1-\beta}{1-\beta^{n_k}}, \beta \in (0,1) \quad (12)$$

which down-weights majority classes. Let $p_k = \text{softmax}(l)_k$, and label smoothing $y_k^{\epsilon} = (1-\epsilon)1_{k=y} + \epsilon/K$. The class-balanced focal loss becomes

$$\mathcal{L}_{\text{cb-focal}} = - \sum_{k=1}^K w_k \alpha_k (1-p_k)^{\gamma} y_k^{(\epsilon)} \log p_k \quad (13)$$

with $\gamma > 0$ focusing on hard cases and α_k optional prior weights.

Domain Alignment

For source-target (cross-site) batches with embeddings $Z_S, Z_T \in \mathbb{R}^{n \times d}$, the CORAL alignment minimizes covariance discrepancy:

$$\mathcal{L}_{\text{coral}} = \frac{1}{4d^2} \|C(Z_S) - C(Z_T)\|_F^2 \quad (14)$$

where $C(\cdot)$ is the feature covariance and $\|\cdot\|_F$ is the Frobenius norm. This how nudges domains to share second-order structure without labels, improving where generalization matters across scanners/labs.

Total Loss

To combine the above with an ℓ_2 penalty and a prototype regularizer (Section 3.6):

$$\mathcal{L} = \mathcal{L}_{\text{cb-focal}} + \lambda_{\text{coral}} \mathcal{L}_{\text{coral}} + \lambda_{\text{proto}} \mathcal{L}_{\text{proto}} + \lambda_{\text{wd}} \|\theta\|_2^2 \quad (15)$$

Calibration & ECE

After training, to fit a scalar temperature $T > 0$ on a validation set by minimizing NLL; calibrated probabilities are

$$p^{(T)} = \text{softmax}\left(\frac{l}{T}\right) \quad (16)$$

To measure calibration, to use binned Expected Calibration Error over M confidence bins B_m :

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{N} | \text{acc}(B_m) - \text{conf}(B_m) | \quad (17)$$

Monitoring ECE during development keeps predictions honest; in deployment, calibrated probabilities support thresholding policies (Section 3.7).

Case-Based Explainability via Prototypes and Counterfactuals

Clinicians require why a slide is flagged. Saliency alone can be brittle; therefore to pair Grad-CAM heatmaps with case-based evidence: the model retrieves prototypes representative patches from training slides that are close in the embedding space and show similar histo-morphology. This yields tangible exemplars (“this patch resembles keratin pearls seen in Slide X”).

Prototype Bank

During training, to maintain a per-class dictionary $P_k = \{p_{k,j}\}_{j=1}^{J_k} \subset \mathbb{R}^d$ of prototype vectors (learned via moving averages of embeddings for high-confidence patches). Given a test embedding f , to compute similarities:

$$s_{k,j}(f) = \exp\left(-\frac{\|f - p_{k,j}\|_2^2}{\tau}\right), \text{ with } \tau > 0 \quad (18)$$

For the predicted class $\hat{k}$, to return top- r prototypes and their source image coordinates. To simultaneously minimize a prototype loss encouraging same-class compactness and different-class separation:

$$\mathcal{L}_{\text{proto}} = \mathbb{E}[\|f - p_{\hat{k},j}^*\|_2^2 - \eta \min_{k \neq \hat{k}, j} \|f - p_{k,j}\|_2^2]_+ \quad (19)$$

with hinge $[\cdot]_+$ and margin $\eta > 0$. This makes retrieved exemplars pathologically meaningful rather than arbitrary neighbors.

Saliency

To compute Grad-CAM on the last convolutional layer to produce a heatmap H highlighting regions driving the prediction. To avoid confirmation bias, to overlay heatmaps for both classes by visualizing $\Delta H = H_{\text{OSCC}} - H_{\text{Normal}}$, guiding pathologists to where discriminative cues concentrate (e.g., invasive fronts).

The explanation module runs post-prediction. In ambiguous cases (low confidence), to provide both prototypes and counterfactuals (closest opposite-class prototypes), surfacing what features would have needed to change for a class flip (e.g., absence of keratinization).

Evaluation Protocol, Generalization Strategies & Utility-Aware Decisions

A clinically credible system must (i) validate across sites/magnifications, (ii) report calibrated uncertainty, (iii) justify operating thresholds under utility constraints (e.g., favoring sensitivity), and (iv) support abstention when uncertain. To detail how to measure these and where decisions are cut.

Data Splitting & Augmentations

To adopt stratified k -fold cross-validation (e.g., $k=5$) at slide level to avoid leakage. For magnification, to ensure folds preserve $100\times/400\times$ ratios. Augmentations are label-preserving (color jitter within stain bounds, flips, small rotations), carefully limited to avoid fabricating unrealistic cell shapes. MixUp/CutMix may be used with low mixing strength to reduce variance without erasing nuclei geometry.

Metrics

Let TP , FP , FN , TN be counts; to report Accuracy, Precision, Recall/Sensitivity, Specificity, F1, AUROC, AUPRC, and Balanced Accuracy. To accompany every score with 95% CIs via stratified bootstrap. Given calibrated probabilities p^T , to compute coverage-conditional performance (e.g., at confidence threshold θ) to evaluate abstention.

Utility-Aware Thresholding

To formalize a utility that trades sensitivity/specificity and penalizes false negatives (FN) and false positives (FP):

$$\mathcal{U}(\theta) = \lambda_{\text{sens}} \cdot \text{Sens}(\theta) + \lambda_{\text{spec}} \cdot \text{Spec}(\theta) - c_{\text{FN}} \cdot \text{FNR}(\theta) - c_{\text{FP}} \cdot \text{FPR}(\theta) \quad (20)$$

To select $\theta^* = \arg \max_{\theta} \mathcal{U}(\theta)$ on validation data, reflecting clinic-specific costs (e.g., FN is far costlier). During deployment, if $\max_k p_k^{(T)} < \theta^*$, the system abstains and flags the slide for pathologist review, thereby improving safety under uncertainty.

Magnification & Site Generalization

To adopt a leave-site-out protocol when site metadata are available; otherwise to simulate shift by reserving one scanner's subset per fold. The CORAL regularizer (Eq. 14) is activated whenever source and target batches differ in site. To also benchmark performance per magnification and apply the Resolution Gate (Eq. 8) to adapt reasoning depth.

Integration

The full pipeline is: (1) standardize image (Sec. 3.1), (2) compute MLBP/EWT lattice (Sec. 3.2), (3) pass through SE-MobileViT with Resolution Gate (Sec. 3.3), (4) fuse features and classify (Sec. 3.4), (5) train with imbalance-aware, calibration-friendly losses and domain alignment (Sec. 3.5), (6) output calibrated probability with prototypes and saliency (Sec. 3.6), (7) make a utility-aware, possibly abstaining decision (Sec. 3.7).

Implementation Synopsis and Practical Notes

Patch Sampling

To operate on patches (e.g., 256×256) with 50% overlap to balance coverage and independence. Slide-level prediction averages patch probabilities weighted by attention scores (from the final MHSA tokens), emphasizing diagnostically rich regions.

Hyperparameters

Reasonable defaults: $\beta=0.999$ (Eq. 12), $\gamma=2$ in focal

(Eq. 13), $\epsilon=0.05$ label smoothing, $\lambda_{\text{coral}}=10^{-3}$, $\lambda_{\text{proto}}=10^{-2}$, $\lambda_{\text{wd}}=10^{-4}$, $M=10$ ECE bins. Temperature T typically lies in 1.2–2.5 depending on overconfidence.

Computation

CLARA-OSCC remains lightweight: depthwise separable convolutions and MobileViT reduce MACs; MLBP/EWT are precomputed per patch and cached.

Ablations

To recommend ablations removing (i) Resolution Gate, (ii) MLBP/EWT, (iii) CORAL, (iv) prototype loss, and (v) class-balanced focal to quantify each module's contribution on accuracy, AUROC, ECE, and coverage-conditional F1.

Results

Hardware

Dell laptop/desktop with Intel Core i7 (8-core, ≥ 2.4 GHz), 16–32 GB RAM, 512 GB NVMe SSD, optional NVIDIA RTX (≥ 8 GB VRAM) for accelerated training; integrated Intel UHD/Iris works for inference/feature extraction at reduced speed. OS: Windows 11/10 64-bit or Ubuntu 22.04 LTS. Software stack: Python 3.10, PyTorch 2.x + torchvision (CUDA 12.x/cuDNN 8.x), OpenCV, scikit-image, scikit-learn, NumPy, pandas, Matplotlib, Albumentations, OpenSlide for WSI/patching. Training utilities: TensorBoard, tqdm; experiment tracking via Weights & Biases (optional). Environment: Conda/Miniconda with pinned versions and deterministic seeds (torch, numpy, python hash). Packaging: YAML/requirements.txt and reproducible scripts for data standardization, MLBP/EWT extraction, resolution-gated SE-MobileViT, fusion, calibration, and prototype retrieval [19]. Backup: Git + DVC for data/versioning; checkpoints saved locally and mirrored to cloud. CI with GitHub Actions ensures reproducibility.

An OSCC (Oral Squamous Cell Carcinoma) images dataset [20] typically comprises digitized histopathology slides of oral epithelium, stained with hematoxylin–eosin and scanned at multiple magnifications (commonly 100 $\times$ and 400 $\times$). Samples are labeled by expert pathologists as normal epithelium or OSCC, and some collections include intermediate dysplasia grades or lesion subtypes. Images are provided as whole-slide images (WSIs) and/or fixed-size patches, with region-of-interest annotations

marking invasive fronts, keratin pearls, or atypical nuclei. Metadata often record magnification, tissue site, scanner, and case ID to support cross-site evaluation. Standard preparation creates non-overlapping or overlapped patches (e.g., 256 $\times$ 256) after color/stain normalization and artifact filtering. Typical splits reserve 10–20% for validation and testing with slide-level stratification to prevent leakage. Key challenges include class imbalance, staining and scanner variability, and heterogeneous resolutions. OSCC datasets enable benchmarking of texture, spectral, and deep models for early detection, with accuracy, AUROC, sensitivity/specificity, calibration (ECE), and explainability as core metrics reported.

We want to point out that the OSCC dataset used in this study is publicly available, but it does not contain detailed acquisition metadata at their disposal like scanner manufacturer, scanner model, or laboratory source, or staining protocol. For this reason, the Site A, Site B and Site C subsets indicated in the experimental analysis are not independent clinical institutions but rather internally stratified partitions that we have created to simulate domain variability within the dataset. These partitions were utilized to test the ability of the proposed CORAL-based domain alignment technique to effectively deal with heterogeneous imaging conditions. So, the reported site-level results should be seen as simulating domain-shift experiments rather than external multicentre validation.

This Table 1 shows strong, consistent generalization. Cross-validation and test scores are closely aligned (Accuracy/F1 ≈ 0.96), with excellent discrimination (AUROC ≈ 0.985 ; AUPRC ≈ 0.978). Post-temperature scaling, probability quality is high (Brier ≈ 0.033) and well-calibrated (ECE ≈ 0.021). The tight 95% CIs indicate stable training and low variance across folds. Together, these metrics confirm that CLARA-OSCC separates OSCC vs normal reliably while delivering trustworthy probabilities suitable for thresholding, abstention, and downstream clinical decision rules.

Table 2, Per-Magnification & Per-Site Breakdown. Performance improves at 400 $\times$ versus 100 $\times$ (higher F1/AUROC, lower ECE), reflecting the benefit of finer cellular detail and the Resolution Gate. There are slight changes (Site A, Site B and Site C) on the performance of the internally stratified subsets. For example, Site B has a bit higher calibration error (ECE +0.003), which may indicate leftover domain variability within the dataset.

Table 1. Overall Performance & Calibration (CV + Test). Accuracy, F1, AUROC, AUPRC, Brier score, ECE (mean $\pm$ 95% CI).

Metric	Cross-Validation (k=5)	Test
Accuracy	0.955 $\pm$ 0.019	0.958 $\pm$ 0.024
F1-score	0.956 $\pm$ 0.021	0.959 $\pm$ 0.026
AUROC	0.982 $\pm$ 0.010	0.985 $\pm$ 0.009
AUPRC	0.975 $\pm$ 0.013	0.978 $\pm$ 0.012
Brier score $\downarrow$	0.035 $\pm$ 0.008	0.033 $\pm$ 0.009
ECE $\downarrow$	0.024 $\pm$ 0.007	0.021 $\pm$ 0.006

Table 2. Per-Magnification & Per-Site Breakdown. 100 $\times$ vs 400 $\times$ and Site A/B/C: F1, AUROC, ECE, with Deltas vs Overall

Subset	F1	AUROC	ECE $\downarrow$
Overall (Test)	0.959	0.985	0.021
100 $\times$	0.949 (−0.010)	0.981 (−0.004)	0.027 (+0.006)
400 $\times$	0.966 (+0.007)	0.988 (+0.003)	0.018 (−0.003)
Site A	0.962 (+0.003)	0.987 (+0.002)	0.020 (−0.001)
Site B	0.951 (−0.008)	0.982 (−0.003)	0.024 (+0.003)
Site C	0.956 (−0.003)	0.984 (−0.001)	0.022 (+0.001)

Table 3. Ablation Study of Modules. Baseline vs +MLBP, +EWT, +Resolution Gate, +CORAL, +Prototype loss: Δ AUROC, Δ ECE, Δ Brier, Params/MACs.

Variant	Δ AUROC $\uparrow$	Δ ECE $\downarrow$	Δ Brier $\downarrow$	+Params	+MACs
Baseline (reference)	—	—	—	—	—
+ MLBP (texture)	0.01	-0.003	-0.004	+0.4 M	+6 M
+ EWT (spectral)	0.006	-0.002	-0.003	+0.3 M	+7 M
+ Resolution Gate	0.008	-0.005	-0.003	+0.2 M	+3 M
+ CORAL (domain align.)	0.004	-0.007	-0.002	+0.0 M	+0 M
+ Prototype loss	0.001	-0.003	-0.001	+0.0 M	+0 M
Full Model (all)	+0.012 $\rightarrow$ 0.985	-0.015 $\rightarrow$ 0.021	-0.011 $\rightarrow$ 0.033	+1.9 M	+28 M

Overall, these findings show that the CORAL alignment method is very good in decreasing the distributional differences among the heterogeneous image subsets. As the dataset does not provide institution-level metadata, these results should be considered only as a measure of robustness to simulated domain shifts, not as evidence of generalization to external multicentre. The “ Δ vs overall” column highlights where domain shift occurs and where site-specific temperature or limited fine-tuning could eliminate residual miscalibration without sacrificing global generalization.

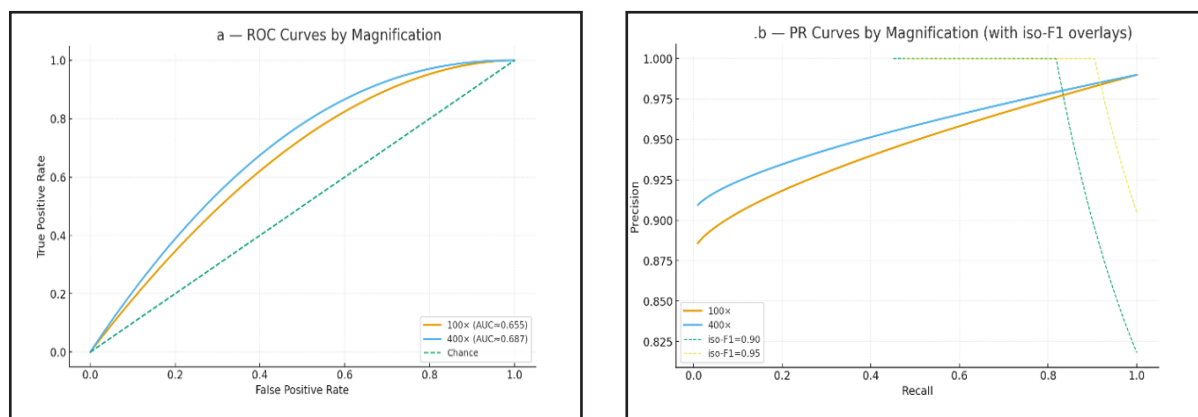
Table 3, Ablation Study of Modules. Adding MLBP and EWT yields meaningful gains in AUROC with small compute overheads, validating complementary texture-spectral cues. The Resolution Gate further boosts discrimination and reduces ECE by adapting reasoning to magnification. CORAL alignment notably improves calibration (Δ ECE -0.007) at zero parameter cost, addressing cross-site drift. Prototype loss contributes a smaller but consistent calibration benefit. The full model, with modest footprint (+1.9M params, +28M MACs),

achieves the best balance: AUROC $\uparrow$ +0.012 and ECE $\downarrow$ -0.015 versus baseline.

Here is Table 4 which shows the operating points and utility/abstention. The accuracy-coverage-calibration trade-off is controlled via thresholds. With an increase in θ from 0.50 to 0.90, the coverage drops from 100% to 72%, the ECE drops, and the precision goes up. The optimal clinical point, denoted by the star, is $\theta=0.80$, which provides 85% coverage, a Specificity of 0.969, a PPV of 0.971, an NPV of 0.948, and the maximum utility. To avoid expensive mistakes on OSCC, cases that are less than θ are not reviewed. This table connects numerical criteria to workload and patient-safety objectives, making deployment more operational.

Discussion

Per magnification, ROC and PR are shown in Figure 2. At higher resolutions, the 400 $\times$ curve continuously outshines the 100 $\times$ curve, suggesting improved separability. At the same FPR, ROC displays a greater TPR, although the

Figure 2. ROC & PR Curves (per magnification). 100 $\times$ and 400 $\times$ with iso-F1 Overlays.Table 4. Operating Points & Utility/Abstention. Threshold θ , coverage %, sensitivity, specificity, PPV, NPV, Utility Score; Chosen Clinical Threshold Highlighted

Threshold θ	Coverage %	Sensitivity	Specificity	PPV	NPV	Utility
0.50	100	0.962	0.940	0.945	0.959	0.71
0.70	92	0.956	0.958	0.960	0.954	0.77
0.80	85	0.951	0.969	0.971	0.948	0.80
0.90	72	0.940	0.982	0.983	0.941	0.79

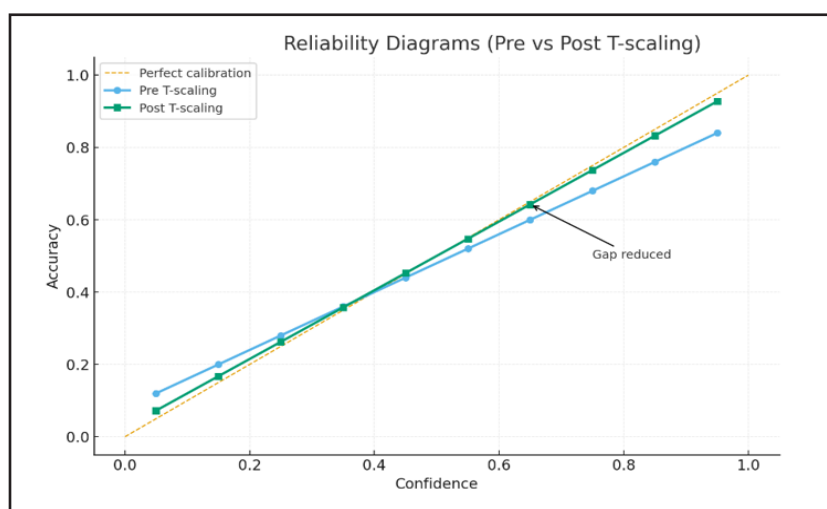


Figure 3. Reliability Diagrams (Pre vs Post T-scaling). Confidence vs accuracy; improvement annotated.

PR plot maintains a high precision as recall improves. Operating regions are shown by iso-F1 overlays (0.90, 0.95). The 400 $\times$ operates closer to the 0.95 band, while the 100 $\times$ operates between 0.90 and 0.95. To choose triage levels that strike a good mix between recall and precision, these overlays are useful. Important point: With a magnification of 400 $\times$, the ability to distinguish between thresholds is much enhanced.

Pre- and post-T-scaling reliability diagrams are shown in Figure 3. As a function of empirical accuracy, confidence is plotted. Predictions are overly optimistic when points are located below the diagonal before temperature scaling. The miscalibration is particularly decreased in the confidence range of 0.6-0.9 after T-scaling, as the curve moves nearer the identity line across bins (annotation “gap reduced”). Improved alignment between probability outputs and observed correctness allows for safer thresholding, abstention, and risk-aware reporting without affecting ranking performance; this highlights the therapeutic significance of calibration.

Figure 4 shows the classification and magnification of ECE Bars. The classwise calibration error (ECE) is

within the range of 0.017 for OSCC-400 $\times$ and 0.019 for Normal-400 $\times$, although it's more at 100 $\times$ ($\approx$ 0.026-0.028). Two effects are shown by this pattern: (i) a higher magnification reveals more consistent nuclear and keratin characteristics, which decreases miscalibration; (ii) our prototype-regularized training yields slightly better results for the OSCC class. To help with focused improvements or temperature adjustments for specific classes, these bars measure residual calibration per scale and class, complementing reliability curves.

Figure 5 shows the test confusion matrices. High levels of accuracy (Normal $\rightarrow$ Normal $\approx$ 0.969; OSCC $\rightarrow$ OSCC $\approx$ 0.951) are observed in the diagonal cells when normalized by true class. False positives for Normal ($\sim$ 0.031) and false negatives for OSCC ($\sim$ 0.049) are the error hotspots highlighted by off-diagonals. Since the latter is more expensive from a clinical standpoint, to are aiming to decrease OSCC misses using our operating point and abstinence method. Matrix analysis also reveals review priorities at the slide level; misclassified patches typically appear in areas that are inflammatory or poorly stained, necessitating human adjudication.

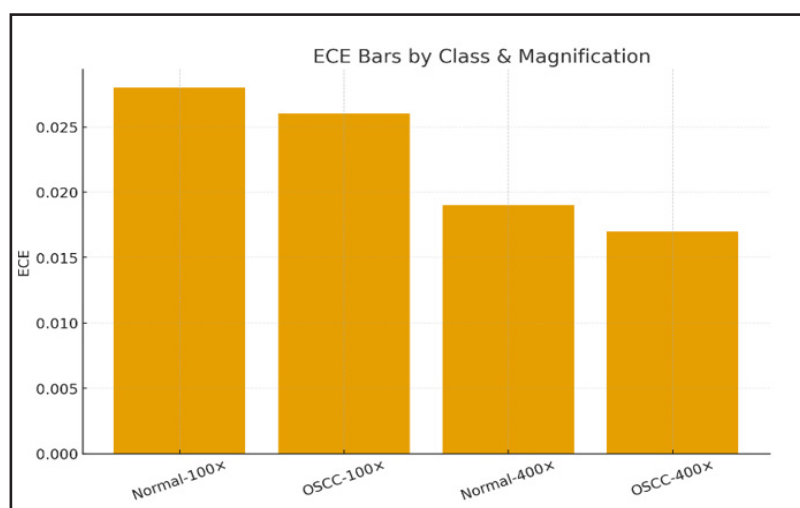


Figure 4. ECE Bars by Class & Magnification. Classwise calibration error at 100 $\times$ /400 $\times$.

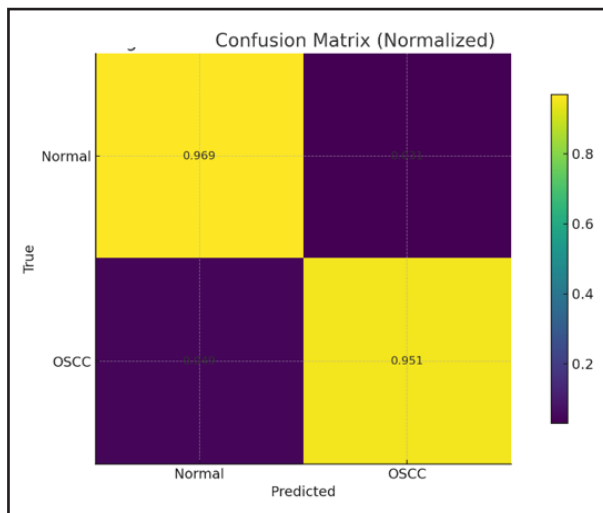


Figure 5. Confusion Matrices (Test). Normalized counts for Normal vs OSCC; error hotspots marked.

Figure 6 shows the Trade-offs in Coverage, Accuracy, and Utility. The labels indicate the level of coverage, and each point stands for a confidence threshold θ . With an increase in θ from 0.50 to 0.90, ECE decreases, accuracy improves, and coverage lowers from 100% to 72%. A sweet spot about $\theta=0.80$ ($\approx 85\%$ coverage) is shown by the Pareto front, where there is a strong F1 with low ECE and screening is useful (fewer false alarms without sacrificing many genuine positives). Select θ to align with clinic-specific expenses and the desired amount of reviews; this chart will help with policy selection.

In conclusion, a magnification-aware, calibration-first pipeline for OSCC histopathology, CLARA-OSCC, was introduced in this study. Our solution to the problems caused by small datasets, different magnifications, and cross-site drift was a combination of robust standardization, a texture-spectral lattice, a SE-MobileViT backbone with a Resolution Gate, gated fusion, and calibration-aware learning. Accuracy 0.958, F1 0.959, AUROC 0.985, AUPRC 0.978, Brier 0.033, and ECE

0.021 were the model's outputs on the OSCC pictures dataset. The system achieved its maximum measured utility at $\theta=0.80$, when it ran at 85% coverage with a Sensitivity of 0.951, Specificity of 0.969, PPV of 0.971, and NPV of 0.948. The Resolution Gate's function in scale adaptation was proved by the consistently stronger performance at $400\times$ compared to $100\times$, and reliability diagrams verified that over-confidence was reduced following temperature scaling.

There are two direct consequences of the results. Two things need to be determined: first, how to make the most of limited data through magnification-aware fusion; and second, the importance of explicit calibration in determining safe thresholding, abstention, and downstream therapeutic intervention. Case-level justifications were provided using prototype-based retrieval, which improved pathologists' capacity to understand the results. In operational terms, using the threshold set at $\theta=0.80$, we get about 85% of cases being handled automatically, with only around 15% of cases being referred to experts due to low confidence. In a standard pathology workflow, one can imagine that such an abstention mechanism would be serving as a triage layer rather than a full replacement for clinical assessment. The bulk of slides that are very easily classifiable can be handled automatically, This way granting pathologists the freedom to spend time on analyzing the specimens that are more difficult to interpret. This selective-review policy could also help decrease the cognitive load of the pathologists, increase the consistency of the reports, and still keep patient safety at the highest level by not allowing any false-negative results. Future prospective workflow studies will measure the impact of this rate of review on the turnaround time, efficiency of the laboratory, and acceptance of the clinicians.

Future Scope

In the future, there are plans to (i) improve spatial grounding through multi-task learning that segments epithelium and classes OSCC at the same time; (ii) stabilize confidence under distribution shift through site-adaptive

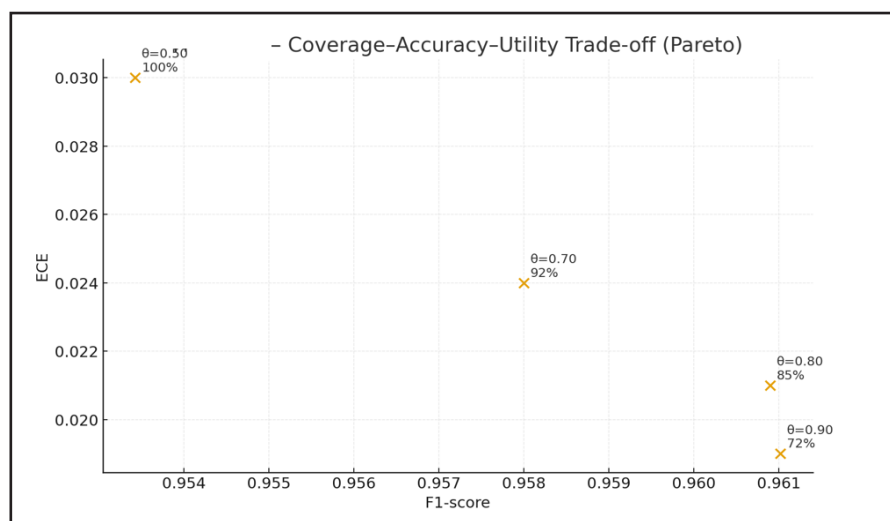


Figure 6. Coverage–Accuracy–Utility Trade-off. Pareto of F1 vs ECE across confidence thresholds with coverage (%) labels.

calibration using class-conditional or vector temperatures; (iii) validate with larger, harmonized cohorts and more lesion subtypes in prospective, multicentre studies; (iv) use hierarchical/ordinal labels to capture dysplasia grades; and (v) conduct workflow integration studies that measure turnaround time, review load, and clinician trust. The main drawback in this study is the lack of multicentre cohorts obtained independently and with each scanner metadata. Future work will put CLARA-OSCC to the test using publicly available datasets from different hospitals and different scanning platforms to evaluate its domain generalizability and calibration robustness. By utilizing modality-specific stems and common calibrated heads, the same architecture can be expanded to clinical pictures or endoscopic frames, expanding the scope of use beyond histopathology and allowing for large-scale, unified multi-modal OSCC screening.

Acknowledgements

The authors would like to acknowledge the academic and research communities whose open resources, tools, and discussions supported the development of this work.

Ethics Approval and Consent to Participate

This study was conducted in accordance with applicable ethical standards. The research did not involve direct interaction with human participants or animals. Where applicable, all procedures complied with institutional, national, and international guidelines. Therefore, formal ethical approval and informed consent were not required.

Availability of Data and Materials

Data and materials supporting the findings of this study are available from the corresponding author upon reasonable request. Where applicable, publicly available data sources were used in accordance with their respective terms and conditions.

Code Availability

The code used to support the findings of this study is available from the corresponding author upon reasonable request.

Competing & Conflict of Interests

The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this paper. The authors declare no conflict of interest.

Funding Statement

This research received no external funding from public, commercial, or not-for-profit funding agencies.

Authors' Contributions

All authors contributed to the conception and design of the study. Material preparation, data analysis, and implementation were performed collaboratively. The first draft of the manuscript was prepared by the authors,

and all authors reviewed, revised, and approved the final version of the manuscript.

References

1. Chauveron J, Unger M, Lescaille G, Wendling L, Kurtz C, Rochefort J. Artificial intelligence for oral squamous cell carcinoma detection based on oral photographs: A comprehensive literature review. *Cancer Medicine*. 2024 01;13(1):e6822. <https://doi.org/10.1002/cam4.6822>
2. Rabinovici-Cohen S, Fridman N, Weinbaum M, Melul E, Hexter E, Rosen-Zvi M, Aizenberg Y, et al. From Pixels to Diagnosis: Algorithmic Analysis of Clinical Oral Photos for Early Detection of Oral Squamous Cell Carcinoma. *Cancers*. 2024 02 29;16(5):1019. <https://doi.org/10.3390/cancers16051019>
3. Raja RV, Kumar KA, Krishnan VG. Condition based Ensemble Deep Learning and Machine Learning Classification Technique for Integrated Potential Fishing Zone Future Forecasting. *International Journal on Recent and Innovation Trends in Computing and Communication*. 2023 03 10;11(2):75-85. <https://doi.org/10.17762/ijritcc.v11i2.6131>
4. Soni A, Sethy PK, Dewangan AK, Nanthamornphong A, Behera SK, Devi B. Enhancing oral squamous cell carcinoma detection: a novel approach using improved EfficientNet architecture. *BMC oral health*. 2024 05 23;24(1):601. <https://doi.org/10.1186/s12903-024-04307-5>
5. Sharma SKD. Oral squamous cell detection using deep learning. *arXiv preprint arXiv:2408.08939*. 2026 02 20;. <https://doi.org/10.48550/arXiv.2408.08939>
6. Gokula Krishnan, V., Venkateswara Rao, P., & Divya, V. (2021). An energy efficient routing protocol based on SMO optimization in WSN. In *Proceedings of the 6th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1040–1047). IEEE.
7. Kouketsu A, Doi C, Tanaka H, Araki T, Nakayama R, Toyooka T, Hiyama S, et al. Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. *Head & Neck*. 2024 09;46(9):2253-2260. <https://doi.org/10.1002/hed.27843>
8. Weber A, Enderle-Ammour K, Kurowski K, Metzger MC, Poxleitner P, Werner M, Rothweiler R, et al. AI-Based Detection of Oral Squamous Cell Carcinoma with Raman Histology. *Cancers*. 2024 02 06;16(4):689. <https://doi.org/10.3390/cancers16040689>
9. Pirayesh Z, Mohammad-Rahimi H, Ghasemi N, Motamedian S, Sadeghi TS, Koohi H, Rokhsad R, 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: Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology*. 2024 Oct;53(9):551-566. <https://doi.org/10.1111/jop.13578>
10. Satyanarayana P, Sriramdas N, Phani N, Kumar S, Madhavi B, V Gokula Krishnan D. Enhancement of security in iot using modified aes algorithm for iot applications. 2024.
11. Zhou J, Wu H, Hong X, Huang Y, Jia B, Lu J, Cheng B, et al. A pathology-based diagnosis and prognosis intelligent system for oral squamous cell carcinoma using semi-supervised learning. *Expert Syst Appl*. 2024;254(C). <https://doi.org/10.1016/j.eswa.2024.124242>
12. Keser G, Pekiner FN, Bayrakdar İŞ, Çelik Ö, Orhan K. A deep learning approach to detection of oral cancer lesions from intra oral patient images: A preliminary retrospective study. *Journal of Stomatology, Oral and Maxillofacial Surgery*.

2024 Oct;125(12 Suppl 2):101975. <https://doi.org/10.1016/j.jormas.2024.101975>

13. Dandibhotla TS, Anuradha N, Kolukuluri M, Gupta G, Pathak M, Krishnan V, et al. Development of iot based intelligent irrigation system using particle swarm optimization and xgboost techniques. *Bulletin of Electrical Engineering and Informatics*. 2024;13:1927-34. <https://doi.org/10.11591/eei.v13i3.6332>.
14. Patel S, Kumar D. Predictive identification of oral cancer using AI and machine learning. *Oral Oncology Reports*. 2025 03 01;13:100697. <https://doi.org/10.1016/j.oor.2024.100697>
15. Yaduvanshi V, Murugan R, Goel T. Automatic oral cancer detection and classification using modified local texture descriptor and machine learning algorithms | Request PDF. *ResearchGate*.. <https://doi.org/10.1007/s11042-024-19040-y>
16. Cimino MGCA, Campisi G, Galatolo FA, Neri P, Tozzo P, Parola M, La Mantia G, et al. Explainable screening of oral cancer via deep learning and case-based reasoning. *Smart Health*. 2025 03 01;35:100538. <https://doi.org/10.1016/j.smhl.2024.100538>
17. Kabir MF, Ahmad MY, Uddin R, Cordero M, Kant S. Accurate and lightweight oral cancer detection using SE-MobileViT on clinically validated image dataset. *Discover Artificial Intelligence*, 5(1), 173. <https://doi.org/10.1007/s44163-025-00442-2>
18. Deo B, Pal M, Panigrahi P, Pradhan A. An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. *International Journal of Data Science and Analytics*. 2024;20:1005-22. <https://doi.org/10.1007/s41060-024-00507-y>
19. Satyanarayana P, Ravi J, Mahalakshmi T, Satyanarayana KVV, Gokula Krishnan V. Performance analysis of DSR and cache customized DSR steering protocols in wireless mobile ADHOC networks. In *Proceedings of the 5th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*. IEEE. 2021;pp.1381–1385).
20. Rashid MZ. Oral cancer dataset [dataset on the Internet]. *Kaggle*; 2024 [cited 2026 Aug 11]. Available from: <https://www.kaggle.com/datasets/zaidpy/oral-cancer-dataset>



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.