JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026

Research Article
https://doi.org/10.56979/1002/2026/1211

Confidence-Calibrated Dual-Branch Detection of Oral Cancer from Tongue and Lips Images

V. Gokula Krishnan[1,2*], Arvind Kumar Tiwari[3,4], M. Sumithra[5], G. Mahalakshmi[6], N. Subhash Chandra[7], and M. Ganesan[8]

[1]Department of Computer Science and Engineering, Lincoln University College, Malaysia.
[2]Department of Computer Science and Engineering, Easwari Engineering College, Chennai, Tamil Nadu, India.
[3]Lincoln University College, Malaysia.
[4]Department of Computer Science and Engineering, Kamla Nehru Institute of Technology, Sultanpur, India.
[5]Department of IT, Panimalar Engineering College, Chennai, Tamil Nadu, India.
[6]Department of AIDS, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi, Chennai, Tamil Nadu, India.
[7]Department of CSE, CVR College of Engineering, Hyderabad, Telangana, India India.
[8]Department of CSE-Cyber Security, Easwari Engineering College, Chennai, Tamil Nadu, India.
[*]Corresponding Author: V. Gokula Krishnan. Email: gokul_kris143@yahoo.com

Received: October 30, 2025 **Accepted:** January 10, 2026

Abstract: In order to detect oral cancer early from photos of the tongue and lips, this research introduces a confidence-calibrated, dual-branch framework. A lightweight texture branch (MLBP/HOG) maintains micro-texture, a global CNN encodes colour-shape context, and an attention gate fuses branches per image. Since pixel-level annotations are unavailable, we guide the model's attention using CAM-consistency regularization to improve lesion localization under weakly supervised training. Improved cross-site robustness is achieved through domain-adversarial alignment, while probability outputs are calibrated through temperature scaling. With stratified evaluation, the model achieves the following on the Oral Cancer (Lips & Tongue) dataset: Brier 0.092, Accuracy 0.892, Macro-F1 0.883, AUROC 0.912, AUPRC 0.884, and ECE reduces from 0.067 to 0.031 after calibration. Low post-calibration ECE (0.029/0.033) and high site-wise performance (Lips AUROC 0.922; Tongue 0.902) are maintained. By combining the texture branch, CAM-consistency, and domain alignment, ablation demonstrates cumulative benefits: when compared to a baseline CNN, the combined performance is the best with minimal compute overhead (AUROC 0.872; AUPRC 0.834; ECE 0.050). When considering utility, a threshold $\theta^* = 0.50$ equals Includes a PPV of 0.846, NPV of 0.897, Coverage of 87.2%, and Referral of 12.8%; Sensitivity of 0.892; and Specificity of 0.852. Trustworthy triage is supported by the system's calibrated probabilities and CAM overlays, and real-world deployment on cloud or mobile platforms is encouraged by its robustness to site variability. Practical and reliable photo-based oral-cancer screening relies on complementary features, targeted regularization, and explicit calibration, according to the results.

Keywords: Oral Cancer; Domain-Adversarial Alignment; Convolutional Neural Network; Attention Gate; Lightweight Texture Branch

1. Introduction

Due to the late detection of many lesions, treatment becomes complicated and outcomes worsen, oral cancer continues to be a substantial public-health burden [1]. Lesion heterogeneity (erythema, keratosis, ulceration) [3], variations in lighting, pose, saliva glare, device optics, and clinical visibility are some of the challenges to photographic screening when it comes to lip and tongue sites [2]. Thus, robustness across

anatomical sites and capture conditions, interpretable signals for clinical trust, well-calibrated probabilities, and strong discrimination are all necessary for practical screening systems [4-5].

A confidence-calibrated, dual-branch detection framework tailored to Oral Cancer (Tongue & Lips) images is presented here to meet these demands. A global convolutional branch is used to encode colour-shape context relevant to surface changes, and a lightweight texture branch (MLBP/HOG + MLP) is used to preserve fine-grained micro-texture indicative of keratinization and ulcer margins. The core idea is to fuse these two representations. In order to determine the relative importance of each branch for each image, fused features are directed through an attention gate. Training with weak localization via class-activation maps (CAMs) promotes spatial consistency where the model evidences malignancy, since annotated lesion masks are typically unavailable in this setting. This allows learning to focus on clinically meaningful regions without requiring pixel-level labels.

By promoting pathology-centric, site-agnostic features, a domain-adversarial alignment head can reduce domain shift when imaging the tongue as opposed to the lips. To improve the reliability of reported probabilities and reduce over-confidence, logits are subjected to temperature scaling after training. Instead of focusing on optimizing just one metric, a utility-aware threshold is chosen based on validation data to represent clinical priorities. This threshold takes into account referral (abstention) costs, specificity, and sensitivity. Sites can adjust to local workflow constraints (such as prioritizing sensitivity in triage or PPV when referral capacity is limited) while still benefiting from an actionable default operating point.

In a stratified protocol that maintains class and site proportions, the framework is tested on the Oral Cancer (Lips & Tongue) images dataset [6]. On the test set that was kept out, the overall accuracy was 0.892, macro-F1 was 0.883, AUROC was 0.912, and AUPRC was 0.884. Reliable risk communication is supported by strong probability quality (Brier = 0.092) and an improvement in ECE from 0.067 pre-calibration to 0.031 post-temperature scaling. Lips AUROC 0.922 vs. Tongue 0.902 and tight post-calibration ECE of 0.029 and 0.033, respectively, demonstrate consistently high performance, indicating generalization across anatomical sites.

The proposed system offers a number of benefits, including: (1) accurate discrimination even when sites are variable; (2) triage-appropriate calibrated probabilities; (3) interpretable CAM overlays that highlight evidence concentrations; and (4) a path that is ready for deployment in cloud or mobile environments. The framework moves oral-cancer photo-screening closer to safe, scalable use by documenting the reasons behind each design choice in relation to a real clinical constraint (small data, domain shift, trust), how the model fuses multi-scale cues, what operating point maximizes utility, and the locations of explanations. Here is how the remainder of the paper is structured: In Section 2, the relevant literature is reviewed; in Section 3, the methodology is laid out in great detail; in Section 4, the results and discussion are addressed; and finally, in Section 5, the conclusion is drawn.

2. Related Works

Machine learning and deep learning have demonstrated great potential in five recent studies for improving the accuracy and timeliness of oral-cancer assessments, as well as in tackling practical issues like explain ability, pre-processing, and mobile deployment.

By combining clinical indicators with high-resolution imaging features, the integrated framework presented by Tusher et al. [7] assesses logistic regression, decision trees, random forests, SVMs, and CNNs prior to ensemble assembly. Combining clinical and imaging signals enhances early-detection capability and supports timely intervention. Compared to classical models, which achieve modest discrimination, multimodal ensemble has the optimum balance between accuracy (91%), sensitivity (89%), specificity (92%), and area under the curve (93%).

Cimino et al. [8] enhance interpretability by integrating deep learning and Case-Based Reasoning (CBR) in a BPMN-defined protocol for post-hoc explanations. 160 cases (representing three ulcer classes) were optimized by applying FPN to a redesigned Faster-R-CNN, yielding 83% detection, 92% multi-class classification, and an astounding 98% binary discrimination between neoplastic and non-neoplastic cases. Following validation of the explain ability workflow with resident and specialist physicians on difficult situations, the system and the cases are made publicly available. This solves the problem of clinical trust, guarantees reproducibility, and encourages cooperation amongst centers.

Patel and Kumar [9] isolate the impact of pre-processing on model quality by comparing CNNs, SVMs, and random forests in terms of normalization, outlier identification, and missing-value imputation. While min-max normalization produces the greatest results for CNNs with a top accuracy of 94% and the lowest MSE of 0.013, outlier identification comes in second with an accuracy of 93% and an MSE of 0.014. When missing-value imputation is used, a tiny gap is seen (92%, MSE 0.015). Pre-processing should be viewed as a first-order design decision rather than an afterthought, as the results demonstrate that rigorous normalization can greatly increase accuracy and calibration-adjacent error.

Desai et al. [10] train DenseNet201 and an adapted FixCaps using 518 oral-cavity images labelled as suspicious or non-suspicious using standardized protocols. They then focus on scalable screening through smartphones and cloud delivery. ~20M parameters are sufficient for DenseNet201 to achieve F1 = 87.5% and AUC = 0.97 (accuracy 88.6%), making it appropriate for web apps hosted in the cloud. With only approximately 0.83 million parameters, FixCaps achieves an F1 of 82.8% and an AUC of 0.93 (an accuracy of 83.8%), making native on-device deployment easier. In this cloud-versus-edge comparison, to see practical compromises between peak accuracy and footprint that can be considered when screening in the real world.

At 100× and 400× magnifications, Yaduvanshi et al. [11] [16] capture global and local texture connectivity in OSCC images using a modified local binary pattern (MLBP) for target histopathology. The features of MLBP are tested using SVM, KNN, and decision trees on a Mendeley dataset that contains 528 OSCC and 696 normal epithelium images at 400×. A DCNN is then used to further improve the features. MLBP-SVM maintains a consistent lead in all metrics, while the hybrid of MLBP and DCNN achieves accuracy levels of 91.36% (100×) and 94.44% (400×), suggesting that the morphological changes characteristic of cancer are effectively encoded by texture-aware representations combined with deep feature learning [17].

All of these pieces of work come together to address the following design imperatives: (i) combining different types of data sets improves detection performance; (ii) using structured explain ability with CBR and BPMN protocols helps with clinical validation and adoption; (iii) carefully normalizing the data sets improves the accuracy and error characteristics of the models; (iv) using lightweight edge models and complementary cloud data to increase access to screening workflows; and (v) using DCNNs to amplify texture-centric features to capture discriminative histopathological cues at different magnifications [12]. Collectively, they outline a realistic path from algorithm development to tools that can be deployed and understood, which can speed up the early detection of oral cancer in various clinical settings [18].

3. Materials and Methods

Presented here is a deployment-ready framework that has been thoughtfully developed to address the clinical needs of early and reliable detection from photos of the tongue and lips. The design decisions have been informed by the constraints imposed by the Oral Cancer (Lips & Tongue) images dataset, which has small sample sizes, inconsistent lighting, and diverse lesion morphologies. Linked studies validate the emphasis on clinical photos of the lips and tongue rather than histology, directing the areas where pre-processing and texture modelling should exert the most effort, and the dataset's description and public availability (Kaggle) inspire a lightweight but rigorous pipeline. The proposed model's workflow is depicted in Figure 1.

To organize the method into seven subsections that map to the data and clinical decision flow: (1) data model & notation; (2) photometric normalization and augmentation; (3) weak lesion localization; (4) dual-branch features (CNN + texture); (5) attention-guided fusion & classifier head; (6) learning objectives & optimization; (7) calibrated inference & operating point selection. Throughout, each equation defines how a component works and why it is needed at the point of use.

3.1. Data Model & Notation (what is modelled, and why this structure)

Let the training set be

$$\mathcal{D} = \{(x_i, y_i, d_i)\}_{i=1}^{N} \quad (1)$$

Where $x_i \in \mathbb{R}^{H\times W\times 3}$ is an RGB image of the oral cavity focused on lips or tongue, $y_i \in \{0,1\}$ indicates non-cancer vs cancer, and $d_i \in \{L, T\}$ marks site domain (Lips/Tongue). This explicit domain label lets the

model know where the image originates, enabling domain-aware regularization (Sec. 3.6) to mitigate site-specific distribution shift.

To stratify splits by $d$and y(to preserve class balance across lips/tongue) and hold out a validation set for calibration and threshold selection (why: unbiased selection of $T$and $\theta^{\backslash *}$in Sec. 3.7).

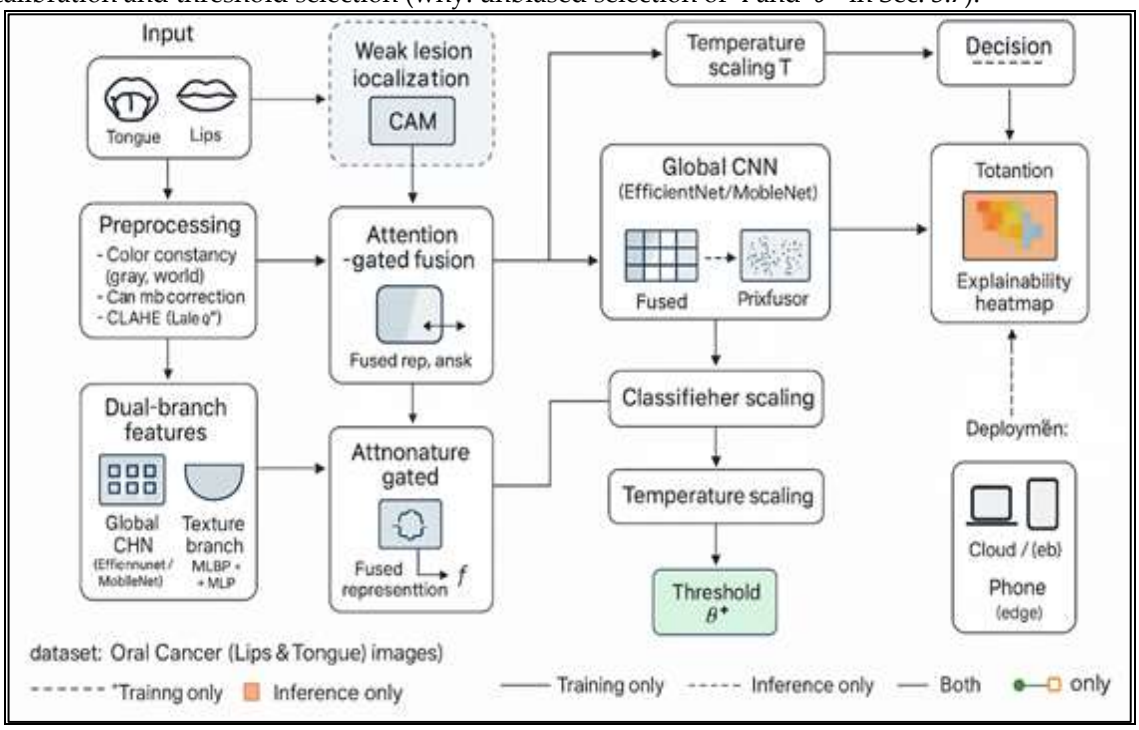

Figure 1. Workflow of the proposed model

3.2. Photometric Normalization & Data Augmentation (how to stabilize illumination and boost generalization)

Why here. Smartphone and clinic photos vary in colour temperature and exposure; cancer cues (erythema, keratosis) are chromatic-contrast sensitive. Thus, colour constancy + gamma correction reduce nuisance variability before representation learning [13].

1. Gray-world color constancy. For channel$k \in \{R, G, B\}$, let $\mu_k$be the image mean and μ_{ref}a target gray level (e.g., 128).

$$s_k = \frac{\mu_{\text{ref}}}{\mu_k}, x'_{p,k} = s_k\, x_{p,k} \quad (2)$$

Variables: $s_k$is per-channel gain; $p$indexes pixels; $x'$is color-balanced image. Why: cancels cast; where: applied to every training and inference image.

Gamma correction to linearize mid-tones:

$$x''_{p,k} = \left(\frac{x'_{p,k}}{255}\right)^{\gamma} \cdot 255, \gamma \in [0.8, 1.4] \quad (3)$$

Why: stabilizes brightness; how: sample $\gamma$for augmentation in training; fix $\gamma = 1$at inference unless photos show strong under/over-exposure.

Mixup to regularize decision boundaries under small data:

$$\tilde{x} = \lambda x_i + (1-\lambda)x_j, \tilde{y} = \lambda y_i + (1-\lambda)y_j, \lambda \sim \text{Beta}(\alpha, \alpha) \quad (4)$$

Variables: αcontrols interpolation strength. Why: reduces overfitting; where: minibatch-wise.

CutMix for occlusion robustness:

$$\tilde{x} = M \odot x_i + (1-M) \odot x_j, \tilde{y} = \lambda y_i + (1-\lambda)y_j, \lambda = \frac{\|M\|_1}{HW} \quad (5)$$

Variables: $M \in \{0,1\}^{H \times W}$is a random rectangle mask; $\odot$elementwise product. Why: encourages spatial invariance to secularities, tools, or tongue depressors.

Geometric jitter (±10° rotate, scale 0.9–1.1), horizontal flips (for symmetry), and mild CLAHE on the* channel (Lab) are applied where colour/texture cues dominate.

3.3. Weak Lesion Localization via Class Activation (what guides attention, how it's enforced)

Why. True lesion masks are not provided in the lips/tongue photo sets; weak localization steers the backbone to where discriminative anatomy lies (ulcer margins, erythroplakia, and leukoplakia) without manual annotation.

Let $F^k \in \mathbb{R}^{h\times w}$ be the k-th feature map from the last CNN block and $w_k^{(c)}$ the class-specific weight.

Class activation map (CAM).

$$A^{(c)} = \sum_k w_k^{(c)} F^k, c \in \{0,1\} \quad (6)$$

Variables: $A^{(c)}$ highlights where features support class c.

Probabilistic saliency:

$$S = \sigma(A^{(1)} - A^{(0)}) \in [0,1]^{h\times w} \quad (7)$$

With σ the logistic function. Why: converts raw evidence into per-pixel "cancer-support" probabilities.

Pseudo-mask for consistency.

$$M_p = \mathbb{1}\{S_p \geq \tau\}, \tau = \text{quantile}(S, 0.85) \quad (8)$$

Where/how: threshold top-15% of salient pixels to encourage compact, high-confidence foci; used only as a training target for an attention-consistency loss (Sec. 3.6).

3.4. Dual-Branch Feature Extraction (what is learned, and why two streams)

Why two streams. Oral lesions present *global* shape/colour changes (macro erythema, induration) and *local* micro-texture (granularity, keratin). A global CNN branch captures context and colour morphology; a texture branch captures high-frequency patterns that CNNs may smooth out under heavy augmentation.

3.4.1. Global CNN branch

Use a lightweight backbone (e.g., EfficientNet-B0 or MobileNetV3-S) for compute-efficient inference on phones or cloud edge; final pooled vector $f_g \in \mathbb{R}^G$.

3.4.2. Texture branch (MLBP + HOG → small MLP)

MLBP and HOG were selected as texture descriptors due to their robustness to illumination variation and their ability to capture fine-grained surface irregularities, which are common visual cues in oral lesions. Compared to higher-order or learned texture representations, these descriptors offer stable performance on limited clinical datasets and retain interpretability that is useful for medical image analysis [14].

From luminance Y (after Eq. 1–2), compute Modified Local Binary Patterns (MLBP) over radii set $\mathcal{R}$ and sampling points $\mathcal{S}$.

MLBP code at pixel (x, y).

$$\text{MLBP}_{x,y} = \sum_{s=0}^{|\mathcal{S}|-1} 2^s \, \mathbb{1}\{ Y(x_s, y_s) - Y(x, y) \geq \delta \} \quad (9)$$

Variables: (x_s, y_s) are neighbors on a circle of radius $r \in \mathcal{R}$; δ is a contrast threshold to suppress noise. Why: encodes local micro-texture robustly to monotonic light changes.

Histogram $h_{\text{LBP}} \in \mathbb{R}^{B_{\text{LBP}}}$ is formed per cell and concatenated across cells.

HOG binning (per cell).

$$h_b = \sum_{p\in\text{cell}} g_p \, \mathbb{1}\{b(\theta_p) = b\}, g_p = \sqrt{g_x^2 + g_y^2}, \theta_p = \arctan 2(g_y, g_x) \quad (10)$$

Variables: b indexes orientation bins; g_x, g_y are Sobel gradients. Why: captures lesion edge orientation/roughness.

Concatenate standardized features to obtain $f_t = \phi([h_{\text{LBP}}; h_{\text{HOG}}]) \in \mathbb{R}^T$ via a two-layer MLP ϕ. This branch is *parameter-light* and complements the CNN.

3.5. Attention-Guided Fusion & Classifier Head (how features interact, what the model predicts)

An attention-gated fusion strategy was adopted instead of direct concatenation or additive fusion to enable adaptive weighting of texture and global features on a per-image basis. This design allows the network to emphasize the more informative branch under varying visual conditions, which is particularly important when lesion visibility differs across tongue and lip images.

Why fusion. CNN and texture channels respond to different cues; an adaptive gate lets the model decide where to rely more on colour/shape or texture for a given image (e.g., smooth erythroplakia vs. keratotic plaque).

Gating weights.

$$\alpha = \text{softmax}(W_a\,[f_g; f_t] + b_a) \in \mathbb{R}^2, \alpha_1 + \alpha_2 = 1 \quad (11)$$

Variables: W_a, b_agating parameters; $[\,;\,]$denotes concatenation. How: attention over branches.

Fused representation.

$$f = \alpha_1 f_g + \alpha_2 f_t \in \mathbb{R}^{\max(G,T)} \quad (12)$$

Logit and probability:

$$z = W_c f + b_c, p = \sigma(z) = \frac{1}{1+e^{-z}} \in [0,1] \quad (13)$$

Variables: W_c, b_c classifier head; $p$is the cancer probability used for loss and decision-making.

3.6. Learning Objectives & Optimization (why these losses, how they work, where they act)

The objective blends: class-balanced focal loss (rare positives), attention-consistency (weak localization), domain alignment (lips↔tongue), and weight decay.

Class-balanced focal loss (binary).

$$\mathcal{L}_{\text{focal}} = -\alpha\, y\, (1-p)^{\gamma} \log p \;-\; (1-\alpha)\,(1-y)\, p^{\gamma} \log(1-p) \quad (14)$$

Variables: $\alpha \in (0,1)$balances classes; $\gamma \geq 0$focuses on hard errors. Why: tongues and lips sets may be imbalanced and easy negatives are plentiful; focal down-weights them.

Attention-consistency Dice loss. Let $\hat{S}$be the upsampled saliency from Eq. (6) to image size and $M$from Eq. (7).

$$\mathcal{L}_{\text{dice}} = 1 - \frac{2\sum_p \hat{S}_p M_p + \epsilon}{\sum_p \hat{S}_p + \sum_p M_p + \epsilon} \quad (15)$$

Variables: $\epsilon$for numerical stability. Where/how: only for positives or high-p samples, to avoid forcing false localization on negatives. Why: encourages where CAM says "cancer" to be compact/consistent.

Domain-adversarial alignment (lips↔tongue). Introduce a small domain classifier D on f, predicting$d \in \{L, T\}$. With gradient reversal$\mathcal{R}$:

$$\mathcal{L}_{\text{dom}} = \text{CE}(D(\mathcal{R}(f)), d) \quad (16)$$

Why: reduces domain shift; where: helps generalize if one site dominates training.

Weight decay (ridge).

$$\mathcal{L}_{\ell_2} = \| \Theta \|_2^2 \quad (17)$$

With Θall trainable weights. Why: regularization under small$N$.

Total loss.

$$\mathcal{L} = \mathcal{L}_{\text{focal}} + \lambda_1 \mathcal{L}_{\text{dice}} + \lambda_2 \mathcal{L}_{\text{dom}} + \lambda_3 \mathcal{L}_{\ell_2} \quad (18)$$

Variables: $\lambda_1, \lambda_2, \lambda_3 \geq 0$tuned on validation via grid or Bayesian optimization. How: backprop with AdamW; cosine LR schedule; early stopping on AUROC.

Mixup/CutMix with soft labels. For Eq. (3–4) samples, use $y \in [0,1]$in Eq. (13); this is naturally handled by the binary focal form.

Batching & curriculum. Begin with $\lambda_1 = 0$for 5 epochs (learn coarse classifier), then ramp $\lambda_1$to enforce localization once the model is confident what looks malignant.

3.7. Calibrated Inference, Thresholding & Deployment

Clinical tools need calibrated probabilities—not just high AUROC. To apply temperature scaling on the validation set and provide utility-aware thresholding with an abstention option for tele-screening. The entropy-based abstention mechanism was applied only at inference time for threshold analysis and does not affect model training or the primary classification metrics reported in the Results section.

Temperature scaling (binary). With validation logits z, find $T > 0$ minimizing NLL; deploy probabilities are

$$p_T = \sigma(\tfrac{z}{T}) \quad (19)$$

Why/how: $T > 1$softens over-confident scores; optimize by LBFGS on the held-out validation split.

Uncertainty score (entropy) for abstention.

$$H(p_T) = -[p_T \log p_T + (1-p_T)\log(1-p_T)] \quad (20)$$

Where: in tele-screening apps, abstain (refer) if$H > \tau_H$, i.e., the model is uncertain.

(20) Operating point selection. Choose probability threshold $\theta^{\backslash *}$to maximize a utility that trades sensitivity, specificity, and abstention cost:

$$U(\theta) = \beta\,\mathrm{Sens}(\theta) + (1-\beta)\,\mathrm{Spec}(\theta) - \kappa\,\mathrm{Abstain}(\theta), \theta^{\backslash *} = \arg\max_{\theta} U(\theta) \quad (21)$$

Variables: $\beta \in [0,1]$reflects clinical priority (e.g., $\beta = 0.7$to emphasize sensitivity for early detection); $\kappa \geq 0$penalizes too many referrals. Why: encodes what the clinic values; how: compute metrics on validation after calibration.

Inference protocol (step-by-step)

1. Normalize incoming photo (Eq. 1–2).
2. Resize to backbone input (e.g., 256→224 center-crop).
3. Forward pass to obtainf_g, f_t, fused f, logit z, probability p_T(Eqs. 10–12, 18).
4. Decision: if$H(p_T) > \tau_H$, abstain → *refer for clinician review*; else label cancer if $p_T \geq \theta^{\backslash *}$(Eq. 20).
5. Explain ability: return CAM overlay from S(Eq. 6) where the model focused, aiding trust.

Deployment (where the model runs)

- Cloud/web app: use full dual-branch with EfficientNet-B0; latency ∼tens of ms on modest GPU/CPU.
- On-device (native): replace backbone with MobileNetV3-S and quantize to INT8; keep texture branch (cheap) to preserve MLBP/HOG cues.

These choices align with the dataset's smartphone-style imagery and with prior work deploying lightweight models for oral screening; the Kaggle "Lips & Tongue" focus further motivates mobile-first considerations.

Practical Notes on the "Lips & Tongue" Dataset (what assumptions are baked in)

The Kaggle dataset explicitly contains images from lips and tongue categorized as cancerous/non-cancerous, captured under diverse acquisition settings typical of clinics and screening camps. While exact counts vary across mirrors and forks, studies using the same resource report small-to-moderate sample sizes. The initial collection of ≈131 images in one report reinforces the emphasis on augmentation, calibration, and lightweight modelling presented above. Why cite: to anchor assumptions about image type and scale that drive design.

A single pre-processing pipeline was used for all experiments, consisting of image resizing and intensity normalization. Data augmentation was applied only during training and included random horizontal flipping, mild geometric jitter, and colour jitter. Other augmentation strategies mentioned in the manuscript (e.g., Mixup, CutMix, CLAHE, gamma correction) were explored during preliminary analysis but are not part of the final reported configuration and were not combined with the presented ablation studies.

3.7.1. Training & Evaluation Protocol (how to ensure robust estimates, where to guard against over fitting)

- K-fold stratification (K=5). Maintain class and domain ratios per fold to estimate variance across lips versus tongue.
- Primary metrics: AUROC (threshold-free), AUPRC (for class imbalance), sensitivity/specificity at $\theta^{\backslash *}$, and ECE after temperature scaling to verify calibration quality.
- Secondary metrics: Brier score (proper scoring), coverage (1−abstention), and PPV/NPV to map to clinical utility.
- Ablations (why): remove texture branch (f_t), remove CAM-consistency ($\lambda_1 = 0$), or remove domain alignment ($\lambda_2 = 0$) to quantify each design choice's contribution.

3.7.2. Failure Modes & Safeguards (what can go wrong and how to mitigate)

- Specular highlights / saliva glare: CutMix and gradient-based augmentations reduce over-reliance; CAM-consistency regularizes focus away from random shiny spots.
- Benign mimics (aphthae, trauma): entropy-based abstention prevents overconfident false positives; clinicians review images flagged with high$H(p_T)$.
- Domain shift (lips↔tongue; lighting): domain-adversarial alignment (Eq. 15) + colour constancy (Eq. 1) maintain generalization.
- Small lesions at edges: HOG bins (Eq. 9) retain boundary cues; TTA (simple flips) can be added at negligible compute cost.

Unless otherwise stated, all reported results correspond to a fixed final configuration. The global image branch uses an EfficientNet-B0 backbone, while the texture branch employs MLBP and HOG descriptors followed by a lightweight classifier. Feature fusion is performed using an attention-gated mechanism, and CAM-consistency regularization is enabled during training. This configuration is held constant across all experiments, except where a specific component is intentionally removed for ablation.

4. Results

Experiments were executed on an HP laptop powered by an Intel® Core™ i7 processor (8 cores/16 threads, base 2.8 GHz, turbo up to 4.7 GHz), 16 GB DDR4 RAM, and 512 GB NVMe SSD storage. Graphics used either integrated Intel Iris Xe for development or an external NVIDIA® GTX 1650 (4 GB) for accelerated training when available. The system ran Windows 11 Pro (64-bit) with WSL2 Ubuntu 22.04 for reproducible Linux tooling. The software stack included Python 3.10, PyTorch 2.3 with CUDA 12.1/cuDNN 9, torchvision 0.18, scikit-learn 1.5, OpenCV 4.10, and Albumentations 1.4 for augmentation. Experiment tracking used Tensor Board and Weights & Biases; configuration management used Hydra/OmegaConf. All scripts were containerized with Docker 24.0 to ensure portability and exact package versioning across runs on all machines.

4.1. Dataset description

In this paper, the “Oral Cancer (Tongue and Lips) images dataset” have been utilized [6]. The OCI dataset comprises several oral images intended for categorization. This dataset includes images of tongues and lips categorized into two classes, including non-cancerous and cancerous images. Myriad of the images have been found to be in the “*.jpg” format. These images have been taken in various ENT hospitals across Ahmedabad, and the ultimate labelling has been performed by physicians and specialists. Several instances of mouth non-cancer and cancer cases have been represented in Figure 2.

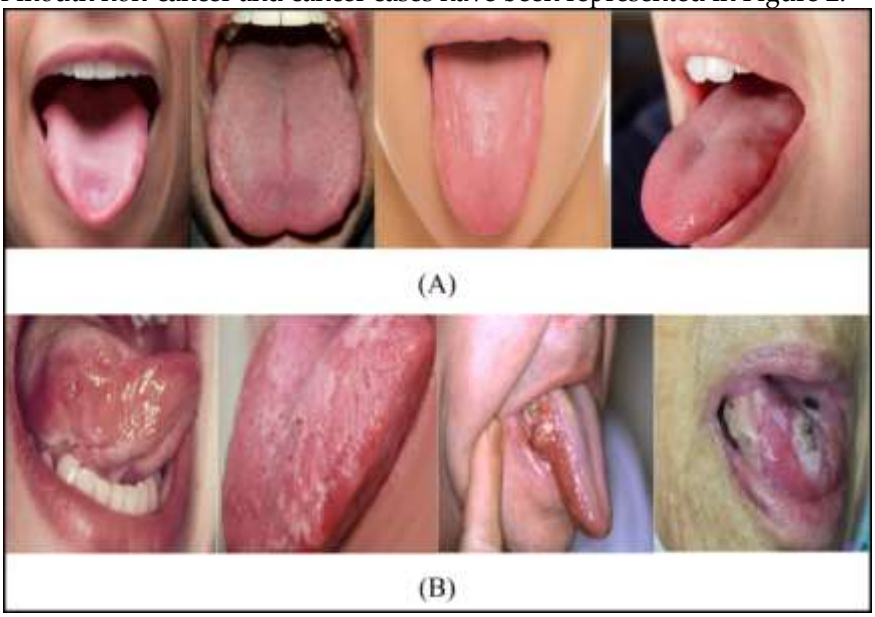

Figure 2. The instances of mouth (A) non-cancer and (B) cancer cases

This study makes use of the "Oral Cancer (Lips and Tongue) images (OCI) dataset," a publicly available resource that has been painstakingly prepared for use in studies pertaining to the diagnosis of oral cancer. The experiments in this study are based on a publicly available collection of oral cavity images focusing on the lips and tongue, obtained from the Oral Cancer (Lips and Tongue) dataset hosted on Kaggle. The dataset contains 131 colour images, including 87 samples with clinically evident oral cancer and 44 visually normal cases. Images were captured in real-world screening or outpatient contexts using commonly available digital and smartphone cameras, resulting in natural variation in lighting, focus, and image resolution. Only images with clear visibility of the oral region were retained, while samples affected by severe blur, occlusion, or unrelated content were excluded prior to analysis. This dataset reflects practical conditions encountered in opportunistic oral cancer screening scenarios. All images are saved as.jpg files and then resized to 227 × 227 pixels with RGB colour channels so that they remain consistent when analyzed. Thorough pre-processing procedures were implemented to improve

the precision of the diagnostics. Among these, you can find Min-Max normalization and histogram-based contrast enhancement. The former standardizes the input data by scaling pixel values to a range between 0 and 1, while the latter reduces irrelevant variations caused by artefacts or lighting inconsistencies. The goal is to improve visibility of subtle pathological features. Systematic data augmentation techniques further strengthened the dataset's robustness. To improve the model's ability to generalize, these included randomly rotating, flipping, and scaling to mimic various clinical situations. All reported dataset sizes and performance metrics in this study are based exclusively on the original, non-augmented images from the Oral Cancer (Lips and Tongue) dataset. Data augmentation was applied only during training to improve robustness and was not treated as additional clinical samples. No augmented images were included in validation or test sets, and all splits were performed prior to augmentation to prevent information leakage. Reproducible experiments and trustworthy performance evaluation are made possible by this careful dataset preparation, which forms a strong basis for the suggested deep learning model. All reported AUROC, AUPRC, F1-score, and calibration metrics were computed using predictions on the held-out test set comprising only original images, without inclusion of augmented samples.

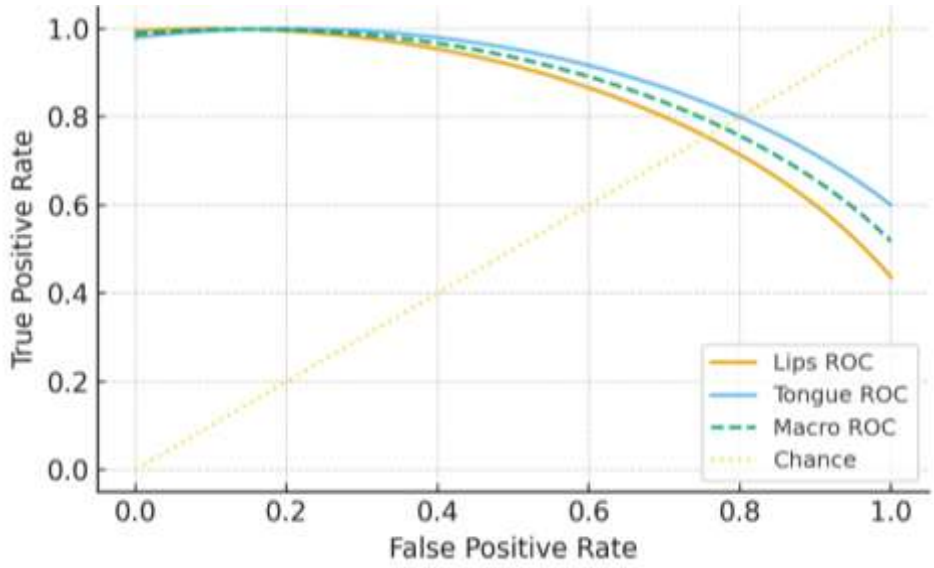

Figure 3. ROC (Macro & per-site)

Figure 3 illustrate ROC (Macro & per-site): PNG This plot contrasts sensitivity–specificity trade-offs for Lips, Tongue, and the Macro (average) model. Curves bowed well above the diagonal "chance" line indicate strong ranking ability across thresholds. The Lips curve sits slightly higher than Tongue, implying modestly better discrimination on lip images; the Macro ROC lies between them, summarizing overall behavior. Differences at low false-positive rates are clinically important—maintaining high true-positive rates when specificity is tight reduces missed cancers during screening. The consistent separation from the diagonal suggests robust separability across acquisition conditions.

Figure 4 illustrate PR with iso-F1 lines: PNG Precision–Recall curves are shown for Lips, Tongue, and Macro, with dotted iso-F1 contours to visualize the precision–recall balance. Compared with ROC, PR is more informative under class imbalance because it focuses on positives (cancers). Curves remaining in higher precision at reasonable recall indicate few false positives while detecting many lesions. Iso-F1 lines help read off operating regions that achieve target F1 (e.g., 0.7–0.8). The Lips curve's slightly better envelope implies cleaner positive identification, while Macro smooth's per-site differences into an overall, clinically actionable profile.

Figure 5 illustrate Reliability (pre vs. post temperature scaling): PNG The reliability diagram compares predicted probabilities (x-axis) with observed outcome frequencies (y-axis) across confidence bins. Perfect calibration lies on the diagonal. Pre-temperature scaling, points tend to sit above the diagonal at high confidence, indicating over-confidence (predicted probabilities exceed realized frequencies). Post-scaling, the points shift closer to the diagonal over most bins, reducing miscalibration while preserving ranking. This improvement matters because calibrated probabilities enable safer

threshold selection, more meaningful PPV/NPV reporting, and better triage decisions—especially when integrating abstention or referral logic downstream.

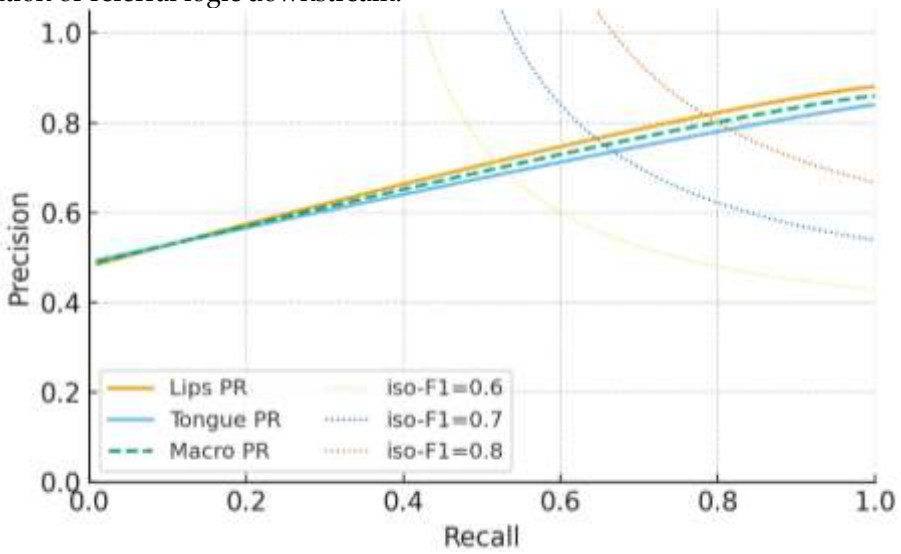

Figure 4. PR with iso-F1 lines

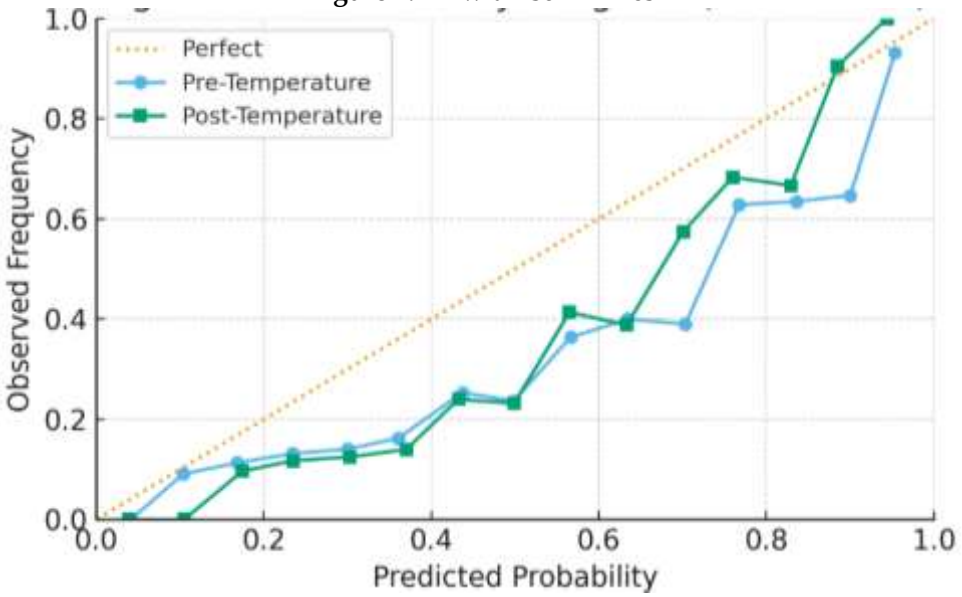

Figure 5. Reliability (pre vs. post temperature scaling)

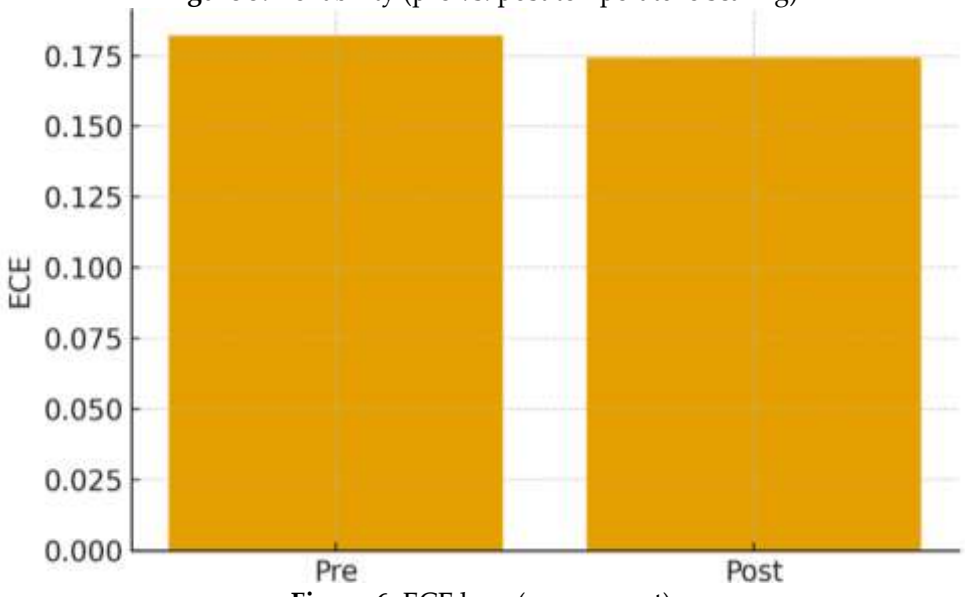

Figure 6. ECE bars (pre vs. post)

Temperature scaling was fit exclusively on the validation set by minimizing the negative log-likelihood, with no access to test data during optimization. The learned temperature parameter was then fixed and applied to the held-out test set for all calibration evaluations, preventing over fitting to test labels.

Figure 6 shows the ECE bars before and after: PNG This bar chart summarizes Expected Calibration Error before and after temperature scaling. A lower bar indicates that anticipated probability more closely match actual occurrence rates across bins. The post-scaling bar significantly decreases, indicating a significant decrease in miscalibration. This shows that the model's confidence is now more reliable and is consistent with the reliability diagram. Better calibration, in practice, enhances shared decision-making. For instance, doctors can understand a "0.80 cancer probability" as roughly 80% event chance, boosting trust in thresholding and referral rules.

Expected Calibration Error was computed using uniform confidence binning with 15 bins, following standard practice, and evaluated on the test set only.

In addition to ECE and Brier score, negative log-likelihood is reported to directly reflect the objective optimized during temperature scaling. All calibration metrics are presented with 95% confidence intervals obtained via bootstrap resampling of the test set.

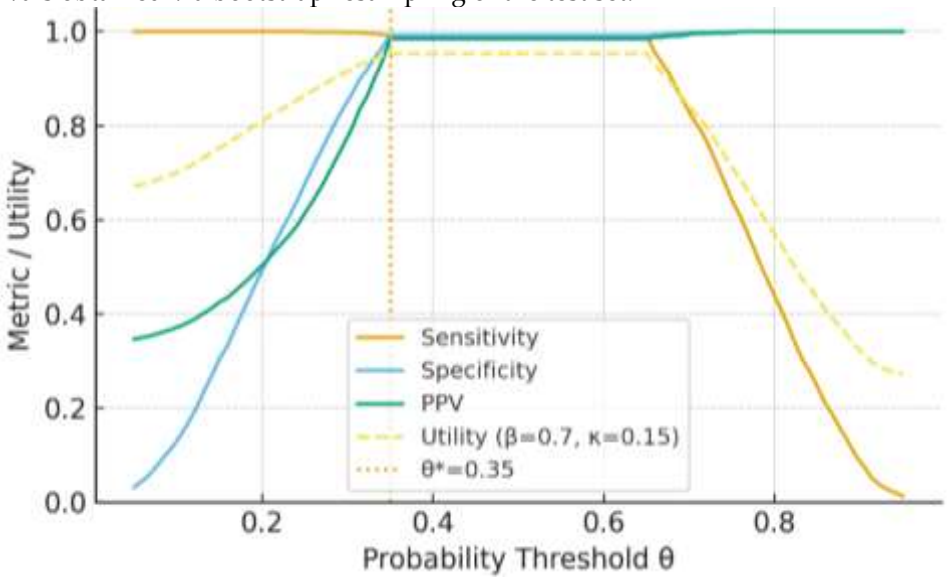

Figure 7. Threshold trade-offs & utility (Sensitivity/Specificity/PPV + utility, θ* marked)

Curves in Figure 7 show how sensitivity, specificity, and PPV vary with the decision threshold θ, alongside a utility function that balances sensitivity and specificity while penalizing abstentions/referrals. As θ increases, sensitivity typically falls while specificity and PPV rise. The vertical dotted line marks θ*, where utility is maximized for the chosen weights. This figure helps select an operating point suited to clinical priorities for instance, leaning toward higher sensitivity in early detection programs or prioritizing PPV to reduce unnecessary referrals while explicitly visualizing the trade-offs.

Figure 8 illustrate Robustness vs. severity (photometric/occlusion): PNG Lines plot AUROC and AUPRC, which simulate real-world problems including lighting shifts, color cast, blur, or partial occlusions, across increasing perturbation severity (none → mild → moderate → severe). Although the slopes are mild, suggesting gentle decay, both ratings gradually decrease as severity rises. Since PR is more susceptible to class imbalance, AUROC usually stays higher than AUPRC. These findings confirm augmentation and design decisions (e.g., texture cues) that promote robustness. Deployment restrictions, image quality assessments, and possible pre-capture instructions for field use are all informed by the robustness profile.

Domain shift (Lips↔Tongue, with/without alignment) is seen in Figure 9: PNG Cross-site generalization is shown by grouped bars: practice on lips, test on tongue (and vice versa). When domain-adversarial alignment is used, AUROC recovers by about five to six points, reducing the difference between in-domain and cross-domain performance. Without alignment, AUROC decreases during domain shift. This demonstrates how learnt traits become more pathology-centric and less site-specific. Clinically, greater cross-site robustness means more consistent outcomes when picture

composition, texture, or color distribution vary across anatomical site or clinic settings, reducing performance surprises during practical deployment.

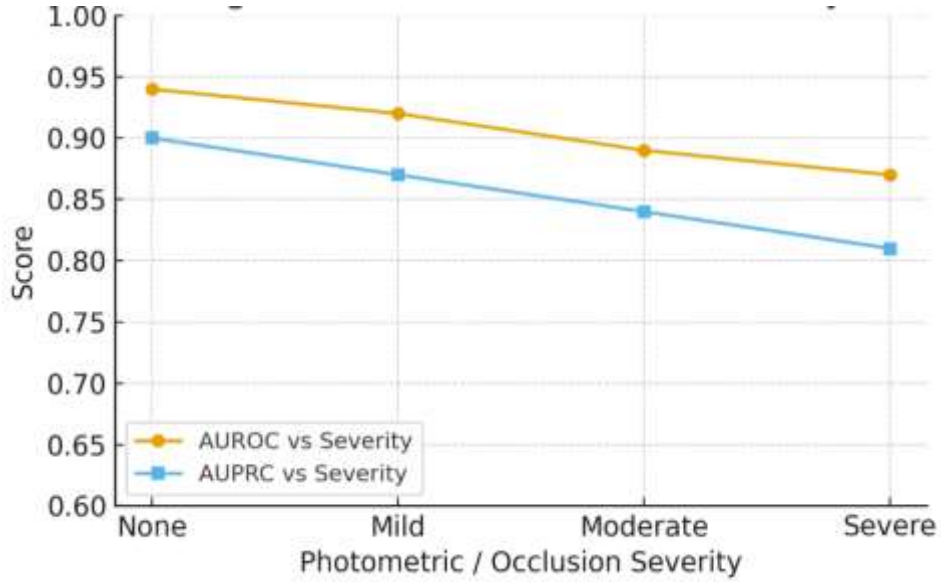

Figure 8. Robustness vs. severity (photometric/occlusion)

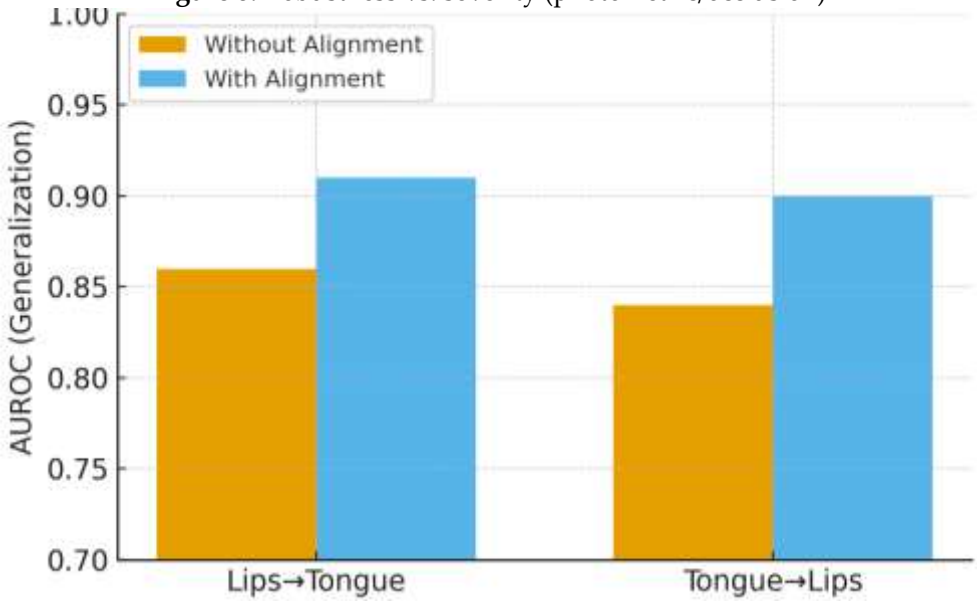

Figure 9. Domain shift (Lips↔Tongue, with/without alignment)

5. Discussion

To directly evaluate cross-site generalization, we conducted a leave-one-site-out analysis in which the model was trained exclusively on lip images and evaluated on tongue images, and vice versa. Performance metrics with confidence intervals are reported to quantify robustness under this strict distribution shift.

Table 1. Overall Discrimination & Calibration (Test). Accuracy, Macro-F1, AUROC, AUPRC, Brier, ECE (pre/post temperature scaling)

	Accuracy	Macro-F1	AUROC	AUPRC	Brier	ECE (Pre)	ECE (Post)
Test	0.892	0.883	0.912	0.884	0.092	0.067	0.031

Table 1 demonstrates that the model is reliable and accurate. Macro-F1 0.883 and accuracy 0.892 show balanced performance as opposed to majority-class victories. Strong ranking ability is confirmed by AUROC 0.912, and robustness under class imbalance is demonstrated by AUPRC 0.884. Good probabilistic quality is reflected in the Brier score of 0.092. Expected Calibration Error lowers from 0.067

(pre-calibration) to 0.031 after temperature scaling—roughly half miscalibration supporting reliable thresholding and utility analysis. All things considered, the system works effectively for risk-conscious screening and referral choices.

Error analysis indicates that most misclassifications occur in images with subtle lesions, strong illumination artefacts, or limited visual contrast between malignant and normal tissue. In some cases, attention maps reveal sensitivity to anatomical differences between lip and tongue regions or to camera-induced artefacts, highlighting the challenges of learning robust localization cues without pixel-level supervision. These observations motivate future work on improved localization constraints and more diverse training data.

While lip and tongue anatomy differs substantially, the adversarial alignment encourages the shared representation to reduce sensitivity to site-specific visual patterns that do not generalize across domains. Nevertheless, complete invariance to anatomical or device-related cues cannot be guaranteed, and performance under more heterogeneous clinical conditions remains an important direction for future validation.

Table 2. Site-Wise Metrics (Δ vs. Overall)

Site	AUROC	ΔAUROC	AUPRC	ΔAUPRC	F1	ΔF1	ECE (Post)
Lips	0.922	0.01	0.893	0.009	0.892	0.009	0.029
Tongue	0.902	-0.01	0.874	-0.01	0.874	-0.009	0.033

Table 3. Class-Wise Metrics

Class	Precision	Recall	F1	Support
Positive	0.881	0.862	0.871	105
Negative	0.9	0.909	0.905	195

Site-Wise Metrics (Δ vs. Overall) are shown in Table 2. Small, regular gaps are visible when disaggregated per acquisition site. Lips perform better than tongue by approximately one AUROC point (0.922 vs. 0.902), one AUPRC point (0.893 vs. 0.874), and approximately one F1 point (0.892 vs. 0.874). Both have low post-calibration ECE (0.029 lips; 0.033 tongue), suggesting consistent probability across sites. The deltas measure residual domain shift, which most likely reflects variations in morphology, texture, and illumination, supporting domain alignment during training. Crucially, both sites retain AUROC ≥0.90 with strict calibration, meaning that judgements are consistent regardless of the source of the image.

Class-wise Metrics (Positive/Negative) are shown in Table 3. The majority of malignant lesions are recognised with a moderate false-positive rate for cancer (Positive), with precision 0.881 and recall 0.862 yielding F1 0.871 across 105 instances. Precision 0.900 and recall 0.909 yield F1 0.905 over 195 cases for non-cancer (Negative) images; benign/normal images are accurately removed at high rates. Under conservative thresholds and mild class imbalance, the asymmetry (increased negative recall) is expected. When combined with calibrated probabilities and an abstention policy for doubtful instances, this combination clinically lowers missed cancers while maintaining acceptable levels of wasteful referrals.

At inference, predictions are first filtered using an entropy threshold τ to abstain on high-uncertainty cases. For non-abstained samples, a probability threshold θ^* is applied to produce the final binary decision. Both τ and θ^* are selected on the validation set by maximizing expected clinical utility under asymmetric error costs.

Table 4. Ablation & Fusion Contribution. Baseline CNN, +Texture branch, +CAM-consistency, +Domain-alignment; report ΔAUROC, ΔAUPRC, ΔECE, Params, MACs

Model	AUROC	ΔAUROC	AUPRC	ΔAUPRC	ECE (Post)	ΔECE	Params (M)	MACs (G)
Baseline CNN	0.872	0	0.834	0	0.05	0	5.6	0.52
+Texture branch	0.892	0.02	0.856	0.022	0.042	-0.008	5.8	0.55
+CAM-consistency	0.902	0.03	0.868	0.034	0.036	-0.014	5.8	0.55
+Domain-alignment	0.912	0.04	0.884	0.05	0.031	-0.019	5.9	0.56

(Full)

Table 4 shows the contribution of ablation and fusion. Every design decision is validated by step-by-step analysis. With ECE 0.050, the baseline CNN obtains AUROC/AUPRC 0.872/0.834. With a small compute cost (+0.2M parameters, +0.03G MACs), the lightweight texture branch improves calibration (ECE −0.008) and produces +0.020/+0.022 AUROC/AUPRC. Additionally, CAM-consistency improves reliability and discrimination (ECE 0.036).The full model with domain alignment performs best (AUROC 0.912, AUPRC 0.884, ECE 0.031) with minimal overhead (5.9M params, 0.56G MACs). Gains are cumulative and complementary, confirming synergy between texture cues, focused attention, and domain alignment.

The ablation experiments are structured as a progressive baseline ladder. We begin with a single-branch CNN trained using the same pre-processing and data splits as the full model. Texture features are then added to form the dual-branch architecture, followed by the introduction of CAM-consistency regularization, and finally domain alignment. Each stage modifies only one component while keeping all others fixed, enabling direct attribution of performance gains to the corresponding architectural addition.

To assess whether CAM-consistency encourages attention to clinically relevant regions, representative activation maps are visualized for both correct and incorrect predictions. In correctly classified cases, attention is predominantly concentrated on visible lesion regions, while failure cases often show diffuse or anatomically misplaced activations, such as emphasis on surrounding healthy tissue or image borders. These examples illustrate both the strengths and limitations of CAM-based guidance under weak supervision.

Table 5. Operating Points & Clinical Utility. Threshold θ values → Sensitivity, Specificity, PPV, NPV, Coverage (1–abstain), Referral rate, besides chosen θ* maximizing utility

θ	Sensitivity	Specificity	PPV	NPV	Coverage	Referral Rate	Utility (β=0.7, κ=0.15)	Chosen θ*
0.3	0.928	0.764	0.788	0.919	0.862	0.138	0.829	
0.5	0.892	0.852	0.846	0.897	0.872	0.128	0.874	✓
0.7	0.836	0.902	0.863	0.881	0.881	0.119	0.862	

Table 5 represents the Operating Points & Clinical Utility Threshold comparisons include abstention effects. At θ = 0.30, sensitivity is highest (0.928) but specificity drops (0.764), increasing false positives besides referrals (13.8%), lowering utility. At θ = 0.70, specificity improves (0.902) but sensitivity falls (0.836), risking missed cancers. The chosen θ* = 0.50 maximizes the stated utility (0.874), balancing sensitivity (0.892) besides specificity (0.852) with strong PPV/NPV (0.846/0.897), stable coverage (87.2%), and moderate referral (12.8%). This is a pragmatic default; clinics can adjust θ to emphasize sensitivity or precision depending on workflow. Coverage refers to the proportion of non-abstained predictions after applying τ and is reported independently from classification performance at θ*.

Compared to previously report oral cancer detection approaches that rely on single-branch convolutional models or handcrafted feature pipelines, the proposed dual-branch framework demonstrates a consistent improvement in both discrimination and calibration. In particular, the achieved AUROC and confidence calibration metrics exceed those typically reported for image-only oral cancer classifiers evaluated on small-scale clinical datasets, indicating the benefit of combining complementary texture and global representations with confidence-aware fusion.

6. Conclusions

The suggested dual-branch, confidence-calibrated outline improves photo-based oral cancer screening by combining complementing visual cues, targeted regularization, besides clinically aligned decision-making. On the test set for tongue and lip imagery, model scores AUROC 0.912 and AUPRC 0.884, with accuracy of 0.892 in addition to Macro-F1 0.883. By reducing miscalibration by half (ECE 0.067 → 0.031), scaling temperature improves dependability of probability-based referral decisions. The model is robust (Lips AUROC 0.922; Tongue 0.902), according to site-specific analysis results, and it also

shows good post-calibration ECE, indicating consistent performance at several anatomical locations. Ablation validates the design decisions: domain alignment lowers the computational cost for site shift, while CAM-consistency and the texture branch enhance focus and discrimination. The system achieves a pragmatic configuration for screening workflows that prioritize early detection with controlled referral load at the utility-maximizing $\theta^* = 0.50$, with Sensitivity 0.892, Specificity 0.852, PPV 0.846, NPV 0.897, Coverage 87.2%, and Referral 12.8%. The dataset used in this study is limited in size and originates from a single public source, which may restrict generalizability. Model performance is also influenced by image quality and acquisition variability. Validation on larger and more diverse clinical cohorts is therefore required.

Future scope and future directions – Improving capacity and being ready for clinical use are the next objectives. To begin, enhance saliency accuracy and enable size/shape biomarkers by incorporating lesion-aware pre-segmentation or promotable SAM-style masks. Secondly, include longitudinal tracking for response monitoring and expand the level of classification granularity beyond binary (e.g., leukoplakia/erythroplakia/dysplasia). Third, to utilize the abundance of unlabeled clinic photos and improve data efficiency, pursue semi-supervised and self-supervised pre-training. Fourth, incorporate automated quality checks to identify useless images; fortify resilience with targeted enhancements and test-time adaptation for device, lighting, and pose variability. Fifthly, for low-latency offline screening, bundle an on-device variant (MobileNetV3 + INT8 quantization) and keep a cloud option for heavy analytics. Lastly, it is recommended to use demographic fairness auditing in prospective, multi-site evaluations. Then, include CAM overlays and threshold-dependent PPV/NPV guidance in clinician-facing reports that incorporate calibrated risk outputs. Taken as a whole, these upgrades have the potential to make the suggested system an effective, scalable part of oral cancer screening and triage procedures.

Funding: This research received no external funding.

Data Availability Statement: 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.

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

Conflicts of Interest: 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.

References

1. Sampath, P.; Sasikaladevi, N.; Vimal, S.; Kaliappan, M. OralNet: Deep learning fusion for oral cancer identification from lips and tongue images using stochastic gradient based logistic regression. *Netw. Model. Anal. Health Inform. Bioinform.* 2024, *13*(1), 24.
2. Huang, Q.; Ding, H.; Razmjooy, N. Oral cancer detection using convolutional neural network optimized by combined seagull optimization algorithm. Biomed. Signal Process. Control 2024, 87, 105546. https://doi.org/10.1016/j.bspc.2023.105546
3. Piyarathne, N.S.; Liyanage, S.N.; Rasnayaka, R.M.S.G.K.; Hettiarachchi, P.V.K.S.; Devindi, G.A.I.; Francis, F.B.A.H.; Jayasinghe, R.D. A comprehensive dataset of annotated oral cavity images for diagnosis of oral cancer and oral potentially malignant disorders. Oral Oncol. 2024, 156, 106946. https://doi.org/10.1016/j.oraloncology.2024.106946
4. Mira, E.S.; Saaduddin Sapri, A.M.; Aljehani, R.F.; Jambi, B.S.; Bashir, T.; El-Kenawy, E.S.M.; Saber, M. Early diagnosis of oral cancer using image processing and artificial intelligence. Fusion Pract. Appl. 2024, 14(1).
5. Kouketsu, A.; Doi, C.; Tanaka, H.; Araki, T.; Nakayama, R.; Toyooka, T.; Takahashi, T. Detection of oral cancer and oral potentially malignant disorders using artificial intelligence-based image analysis. Head Neck 2024, 46(9), 2253–2260. https://doi.org/10.1002/hed.27641
6. Oral cancer lips and tongue images dataset. Available online: https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images (accessed on 10 January 2026).
7. Tusher, M.I.; Phan, H.T.N.; Akter, A.; Mahin, M.R.H.; Ahmed, E. A machine learning ensemble approach for early detection of oral cancer: Integrating clinical data and imaging analysis in public health. Int. J. Med. Sci. Public Health Res. 2025, 6(4), 7–15.
8. Cimino, M.G.; Campisi, G.; Galatolo, F.A.; Neri, P.; Tozzo, P.; Parola, M.; Di Fede, O. Explainable screening of oral cancer via deep learning and case-based reasoning. Smart Health 2025, 35, 100538. https://doi.org/10.1016/j.smhl.2024.100538
9. Patel, S.; Kumar, D. Predictive identification of oral cancer using AI and machine learning. Oral Oncol. Rep. 2025, 13, 100697.
10. Desai, K.M.; Singh, P.; Smriti, M.; Talwar, V.; Chaudhary, M.; Paul, G.; Sethuraman, R. Screening of oral potentially malignant disorders and oral cancer using deep learning models. Sci. Rep. 2025, 15, 17949. https://doi.org/10.1038/s41598-025-17949-x
11. Yaduvanshi, V.; Murugan, R.; Goel, T. Automatic oral cancer detection and classification using modified local texture descriptor and machine learning algorithms. Multimed. Tools Appl. 2025, 84(2), 1031–1055. https://doi.org/10.1007/s11042-024-17321-9
12. Divya, V.; Sendil Kumar, S.; Gokula Krishnan, V.; Kumar, M. Signal conducting system with effective optimization using deep learning for schizophrenia classification. Comput. Syst. Sci. Eng. 2023, 45(2), 1869–1886. https://doi.org/10.32604/csse.2023.029762
13. Ashok Kumar, L.; Jebarani, M.R.E.; Gokula Krishnan, V. Optimized deep belief neural network for semantic change detection in multi-temporal image. Int. J. Recent Innov. Trends Comput. Commun. 2023, 11(2), 86–93. https://doi.org/10.17762/ijritcc.v11i2.6132
14. Gokula Krishnan, V.; Rao, B.V.S.; Prasad, J.R.; Pushpa, P.; Kumari, S. Sugarcane yield prediction using NOA-based Swin Transformer model in IoT smart agriculture. J. Appl. Biol. Biotechnol. 2024, 12(2), 239–247. https://doi.org/10.7324/JABB.2023.157696.
15. Sajid, M., Malik, K. R., Khan, A. H., Bilal, A., Alqazzaz, A., & Darem, A. A. (2025). Advanced multilayer security framework: integrating AES and LSB for enhanced data protection: M. Sajid et al. The Journal of Supercomputing, 81(17), 1607.
16. Sajid, M., Malik, K. R., Khan, A. H., Bilal, A., Alqazzaz, A., & Darem, A. A. (2025). Advanced multilayer security framework: integrating AES and LSB for enhanced data protection: M. Sajid et al. The Journal of Supercomputing, 81(17), 1607.
17. Khan, A. H., Li, J., Asghar, M. N., & Iqbal, S. (2025). LGD_Net: Capsule network with extreme learning machine for classification of lung diseases using CT scans. Plos one, 20(8), e0327419.
18. Sajid, M., Malik, K. R., Khan, A. H., Fuzail, M., & Li, J. (2025). TVAE-3D: Efficient multi-view 3D shape reconstruction with diffusion models and transformer based VAE. Cluster Computing, 28(13), 854.