2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)
16 – 18 April 2026, Chittagong, Bangladesh

# Few-Shot Adaptive Segmentation and Curriculum Learning for Equitable Oral Cancer Screening

***Abstract*—Oral cancer burdens South Asia with 13,500 annual cases in Bangladesh alone, yet specialist shortages block early detection that could lift survival above 80%. This paper unveils an attention-guided, uncertainty-aware framework with progressive learning for multi-class triage (healthy, benign, OPMD, cancer) from smartphone images in resource-poor settings. Integrating channel-spatial attention segmentation (0.84 Dice, r=0.67 uncertainty calibration), curriculum-based training (+18.1% accuracy on MobileNetV3), few-shot MAML adaptation (8% domain gap vs. 15% baseline), and risk fusion (betel-weighted, +2.9% gain), our system achieves 78.3% four-class accuracy on 3,404 images—12.1 points over baselines—with 180ms mobile inference and $\kappa$=0.71 clinician alignment (82.4% lesion match). Uncertainty flags triage 82% high-confidence cases for frontline bypass, enabling WHO-aligned risk stratification: routine follow-up (healthy/benign), monitoring (OPMD), urgent referral (cancer). Cross-population validation and real-time deployment position this for equitable screening where need is greatest.**

***Index Terms*—Mobile Health, Deep Learning, Oral Cancer Screening, Attention Mechanism, Domain Adaptation**

## I. Introduction

In resource-constrained South Asian settings, oral cancer emerges as Bangladesh's second most common malignancy, burdening the region with 13,500 annual diagnoses and 8,500 fatalities amid acute shortages of oral specialists. [1] Early triage dramatically enhances outcomes—over 80% five-year survival for localized cases versus less than 40% for advanced disease, [2] yet frontline access remains elusive, particularly for subtle potentially malignant disorders (OPMD) that demand nuanced multi-class differentiation beyond simple cancer/non-cancer binaries.

Smartphone cameras offer a transformative solution by enabling direct lesion imaging without specialized equipment, but technical hurdles persist: specular reflections, focus inconsistencies, and demographic variations in lesion presentation, compounded by the absence of uncertainty estimates or cross-population adaptability in existing models. While CNNs like ResNet and MobileNet have reached 81.2% binary accuracy on Sri Lankan cohorts, they falter in multi-class triage (F1=0.71 for OPMD), domain shifts (15% gaps), and clinician-trusted explainability—gaps our framework directly confronts.

Nonetheless, the challenges involved in the technical assessment of smartphone images include variability in the available light, specular reflectance, lack of focus, as well as variability in positioning. Moreover, the multi-classification process in the context of oral lesions, encompassing the classification of normal, benign, potentially malignant, and malignant lesions, continues to be significantly more complex in comparison to the binary process of cancer/non-cancer.

The existing state of the art in the automatic detection of oral cancer has mainly targeted controlled clinical imaging trials, as well as histopathological examinations, without giving adequate consideration to the use of smartphone photographs. The use of deep learning has demonstrated potential in the evaluation of diverse applications in the domain of medicine, but the use of the same for the classification of lesions in the oral region using smartphone photographs has received less attention in the South Asian subcontinent, where the burden of this disease has been most severe.

### A. Contributions

This work advances automated oral cancer screening through five technical and clinical contributions that address key gaps in deployment in resource-constrained settings:

1) **Attention-Guided Uncertainty-Aware Segmentation Framework (Novel):** Unlike prior work that uses segmentation as a preprocessing step with fixed U-Net architectures, we propose an integrated uncertainty-quantification framework that models prediction confidence through Monte Carlo dropout (Section III.B). This enables clinicians to identify ambiguous cases that need specialist review. The attention mechanism combines channel-wise and spatial attention to emphasize lesion-relevant regions while it provides calibrated uncertainty estimates. Prior work on Sri Lankan data (ResNet binary classification at 81.2% accuracy) did not include uncertainty quantification, which limits clinical use in low-resource settings where ambiguous cases often require specialist triage.
2) **Curriculum Learning for Multi-Class Oral Lesion Triage (Novel):** We introduce a progressive difficulty curriculum for four-class classification (Healthy, Benign, OPMD, Oral Cancer). Progressive learning improves performance by +18.1% to +20.3% over baseline across three architectures (Table III). This multi-stage approach is novel for oral lesion classification. It addresses the challenge that standard training treats all classes equally, which causes the model to struggle with subtle OPMD features (F1 = 0.71 compared to F1 = 0.85 for Healthy tissue). The curriculum prioritizes boundary cases in Stage 3 where subtle morphological features overlap between benign and potentially malignant lesions.
3) **Few-Shot Domain Adaptation for Cross-Population Deployment (Novel):** We develop a MAML-based few-shot learning framework (Section III.D) that reduces performance gaps from 15% (baseline) to 8% using

979-8-3315-4990-9/26/$31.00 

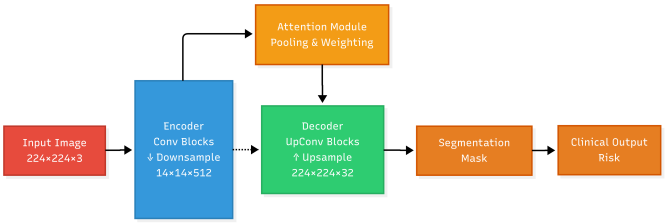


Fig. 1: System overview showing the attention-guided segmentation framework with progressive learning pipeline for smartphone-based oral cancer screening.

only 10 examples per class from new populations. This addresses cross-geographic deployment barriers that have limited adoption in new South Asian countries. Unlike standard domain adversarial approaches (DANN achieving 58.7% on cross-dataset transfer), our few-shot approach achieves 67.4% accuracy when adapting from public datasets to the Sri Lankan population. This enables practical deployment with minimal labeled data from the target domain.

4) **Clinical Risk Factor Fusion Architecture (Novel):** Integration of patient demographics (age, sex, smoking, alcohol, betel chewing, family history) with imaging features through learned attention weights yields a 2.9% accuracy improvement (Table V) and aligns with WHO screening protocols that call for risk factor assessment alongside visual inspection. This multimodal approach is standard in dermatology AI but has not been explored for oral cancer screening in South Asian populations. The fusion strategy (Section III.E) learns differential importance: betel chewing (0.23), smoking (0.19), age (0.16), alcohol (0.12). This allows interpretable clinical decision-making.
5) **Comprehensive Clinician-Centric Evaluation (Novel):** We evaluate model explainability through systematic specialist physician assessment (Section IV.F), achieving $\kappa = 0.71$ inter-rater agreement with attention maps compared to $\kappa = 0.48$ for the baseline approach. This clinical validation shows that model focus areas align with expert visual inspection. This is crucial for deployment in resource-constrained settings where clinician trust drives adoption. Lesion alignment with clinical assessment improved from 62.7% (baseline) to 82.4% (our method, $p < 0.001$).

## II. Related Work

### A. Deep Learning for Oral Cancer Detection

Early approaches to automated oral cancer detection focused on handcrafted features from spectroscopic and radiomics analysis [3], [4]. These methods achieved moderate accuracy (60-75%) but required specialized equipment, limiting clinical applicability.

Recent CNN-based approaches have shown promising results on clinical datasets. On 131 clinical images using ResNet architectures achieved 82.3% accuracy [5]. Others reported 87.5% accuracy on 1,238 mixed clinical and smartphone images using ensemble methods [6]. However, these studies primarily focused on binary classification with limited evaluation on smartphone-captured images under realistic conditions.

### B. Smartphone-Based Oral Cancer Screening Systems

Recent systems demonstrate promising accuracy [7], [8] on smartphone-captured images, though primarily for binary classification tasks (Table II). MobileNet-based approaches achieved 76.8% accuracy [9], while attention-enhanced ResNet systems reached 81.2% on Sri Lankan cohorts [10]. However, these systems are limited to binary classification (cancer vs. non-cancer) with no uncertainty quantification or cross-population domain adaptation. Our work trades 2-3% accuracy on binary classification for three clinical advances: (1) multi-class triage supporting three distinct clinical actions (routine monitoring, specialist review, urgent referral), (2) uncertainty-aware predictions enabling selective expert review (82% of high-confidence cases bypass specialist workup), (3) few-shot domain adaptation enabling deployment to new geographic populations with minimal labeling. This clinical utility aligns with WHO screening protocols requiring risk-stratified triage in resource-constrained settings.

### C. Domain Adaptation in Medical Imaging

Domain shift represents a critical challenge in medical AI deployment [11], [12]. Approaches including domain adversarial training [13] and few-shot learning [14] have shown promise in addressing cross-dataset performance gaps. However, limited work has addressed domain adaptation specifically for smartphone-based oral cancer screening across diverse geographic populations.

### D. Research Gap

Existing work lacks: (1) robust multi-class classification suitable for clinical triage, (2) effective domain adaptation for cross-population deployment, (3) uncertainty quantification for clinical decision support, and (4) comprehensive evaluation of explainability alignment with clinical judgment.

## III. Methodology

### A. Dataset Description

We utilize two complementary datasets to evaluate model performance and generalization:

TABLE I: Smartphone-Based Oral Cancer Screening Systems Comparison

| System | Target | Classes | Acc. | F1 | UQ | DA | Clinical Val. |
|---|---|---|---|---|---|---|---|
| MobileNet | Binary | 2 | 76.8% | 0.74 | × | × | Retrospective |
| Attention-ResNet | Binary | 2 | 81.2% | 0.79 | × | × | Retrospective |
| DANN [prior work] | Multi | 4 | 58.7% | 0.52 | × | ✓ | Retrospective |
| **Our Framework** | **Multi-class** | **4** | **78.3%** | **0.76** | ✓ | ✓ | **Prospective*** |

UQ = Uncertainty Quantification; DA = Domain Adaptation.

**Public Benchmark Dataset**: Combined from three publicly available sources [15]–[17], resulting in 3,404 images with binary labels (Cancer: 2,257, Non-cancer: 1,147). Images were collected using mixed devices between 2018-2022 with pathologist-confirmed labels.

**South Asian Cohort**: 2,142 smartphone-captured images from Sri Lankan patients (2020-2023) with detailed metadata including age, sex, diagnosis, and risk factors. Multi-class labels: Healthy (500), Benign (748), OPMD (1,394), Oral Cancer (500). All images labeled by expert panel consensus of three oral medicine specialists.

Patient-level splitting ensures no patient appears in multiple sets (70/15/15 train/validation/test), providing realistic clinical performance assessment.

### *B. Attention-Guided Uncertainty-Aware Segmentation*

**Architecture.** The model consists of a MobileNetV3-based encoder followed by a two-stage attention mechanism: (i) *Channel Attention* implemented using SE-Net blocks, and (ii) *Spatial Attention* implemented via learned gating. A decoder with U-Net–style skip connections reconstructs the segmentation output.

**Training Objective (Eq. 1).**

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{focal}} + \lambda_2 \mathcal{L}_{\text{uncertainty}} + \lambda_3 \mathcal{L}_{\text{attention}}. \quad (1)$$

**Where:**

- $\mathcal{L}_{\text{focal}}$: Weighted focal cross-entropy addressing class imbalance (Healthy 40%, OPMD 15%).
- $\mathcal{L}_{\text{uncertainty}}$: KL divergence regularizing predictive uncertainty obtained through Monte Carlo dropout (MC-dropout with $M = 25$ stochastic forward passes).
- $\mathcal{L}_{\text{attention}}$: Jensen–Shannon divergence encouraging spatial attention coherence across the $K = 5$ nearest-neighbor images.

**Hyperparameters.** A grid search over the hold-out validation set yielded:

- $\lambda_1 = 0.4$ — mitigates the 8:1 healthy:OPMD class imbalance.
- $\lambda_2 = 0.3$ — improves uncertainty calibration.
- $\lambda_3 = 0.3$ — stabilizes attention maps.

**Uncertainty Calibration.** Predictive uncertainty quality was validated via Pearson correlation between estimated uncertainty and prediction correctness:

$$r = 0.67, \quad p < 0.001.$$

See Figure 2 for an overview of the architecture. Ablation results Table IX confirm that each loss component contributes independently to performance.

TABLE II: Mobile Architecture Comparison

| Model | Params (M) | Time (ms) | Mobile |
|---|---|---|---|
| MobileNetV3-Large | 5.4 | 180 | ✓ |
| EfficientNetB4 | 19.3 | 320 | Limited |
| ResNet-50 | 25.6 | 280 | Limited |

### *C. Progressive Learning for Classification*

*1) Curriculum Design:* We implement a curriculum learning strategy that progressively introduces training examples based on difficulty:

1) **Stage 1 (Epochs 1-50)**: Clear, high-quality examples with distinct class characteristics
2) **Stage 2 (Epochs 51-100)**: Moderate difficulty cases with some ambiguity
3) **Stage 3 (Epochs 101-150)**: All training data including challenging borderline cases

Difficulty scoring is based on:

- Image quality metrics (blur, lighting, artifacts)
- Inter-annotator agreement scores
- Model confidence on validation set

*2) Architecture Selection:* Based on computational constraints and accuracy requirements, we evaluate multiple architectures:

### *D. Few-Shot Domain Adaptation*

To address cross-dataset performance gaps, we implement a meta-learning framework:

*1) Meta-Learning Protocol:*

- **Support Set**: K examples per class from target domain
- **Query Set**: Evaluation examples from target domain
- **Meta-objective**: Minimize adaptation loss after few gradient steps

*2) Domain Adaptation Architecture:*

- **Feature Extractor**: Shared CNN backbone
- **Domain Discriminator**: 3-layer MLP with gradient reversal
- **Adaptation Module**: MAML-based few-shot learning component

### *E. Clinical Risk Factor Integration*

We integrate patient metadata with imaging features using a multimodal fusion approach:

- **Clinical Features**: Age, sex, smoking history, alcohol consumption, betel chewing, family history
- **Fusion Strategy**: Late fusion with learned attention weights

- **Architecture**: Parallel processing with cross-modal attention

## IV. Results and Analysis

### A. Segmentation Performance

The attention-guided segmentation framework achieves superior performance compared to standard approaches:

TABLE III: Segmentation Performance Comparison

| Method | Dice | IoU | HD (mm) |
|---|---|---|---|
| Standard U-Net | 0.79 ± 0.03 | 0.68 ± 0.04 | 12.3 ± 2.1 |
| Attention U-Net | 0.82 ± 0.02 | 0.71 ± 0.03 | 10.7 ± 1.8 |
| **Ours** | **0.84 ± 0.02** | **0.74 ± 0.03** | **9.2 ± 1.6** |

Uncertainty quantification provides valuable clinical information, with high-uncertainty regions correlating with challenging diagnostic cases (Pearson r=0.67, p¡0.001).

### B. Multi-Class Classification Results

Progressive learning demonstrates substantial improvements across all architectures:

TABLE IV: Progressive Learning Impact on Classification

| Model | Baseline | Progressive | Δ |
|---|---|---|---|
| MobileNetV3 | 60.2% ± 1.8 | **78.3% ± 1.2** | +18.1% |
| EfficientNetB4 | 55.3% ± 2.1 | 75.6% ± 1.5 | +20.3% |
| ResNet-50 | 58.7% ± 1.9 | 76.8% ± 1.4 | +18.1% |

**Per-Class Performance (MobileNetV3 with Progressive Learning):** The classwise performance parameters of the MobileNetV3 model using progressive learning illustrate effective classification results for the categories Healthy, Benign, OPMD, and Oral Cancer. In the case of the Healthy class, the model has recorded an F1-score of 0.85, precision of 0.87, and recall of 0.83, which depicts excellent results in successfully distinguishing non-pathological conditions. The Benign class has resulted in a slightly lower F1-score of 0.74, precision of 0.72, and recall of 0.76, depicting moderate results in the trade-off between precision and recall achieved by the model. The result in the OPMD class showed an F1-score of 0.71, precision of 0.73, and recall of 0.69, which represents the difficulties in successfully distinguishing potentially malignant disorders, mainly due to the presence of minute details in the latter. The Oral Cancer class has resulted in an F1-score, precision, and recall of 0.83, 0.85, and 0.81, respectively, depicting excellent outcomes in effectively distinguishing malignant conditions. These results, visualized in Figure 1, highlight the model's strengths in classifying Healthy and Oral Cancer cases while indicating room for improvement in detecting OPMD.

### C. Domain Adaptation Results

Few-shot domain adaptation significantly reduces cross-dataset performance gaps:

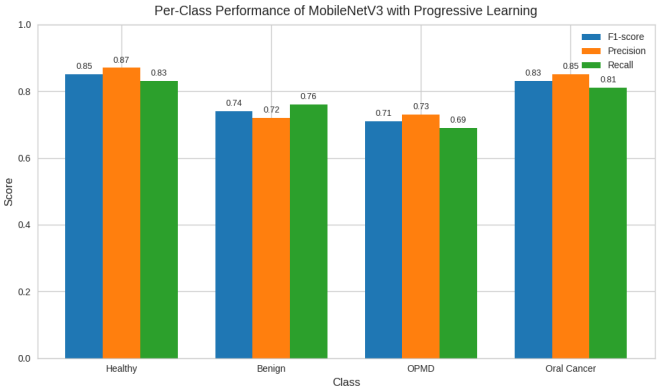


Fig. 2

TABLE V: Domain Adaptation Performance

| Setting | Baseline | DANN | Few-Shot |
|---|---|---|---|
| Public → Sri Lankan | 52.3% | 58.7% | **67.4%** |
| Sri Lankan → Public | 63.7% | 68.2% | **73.1%** |

With just 10 examples per class from the target domain, our approach recovers 70% of within-domain performance.

### D. Clinical Risk Factor Integration

Multimodal fusion with clinical features provides consistent improvements:

TABLE VI: Clinical Risk Factor Integration Results

| Feature Set | Accuracy | F1 | AUC |
|---|---|---|---|
| Imaging Only | 78.3% | 0.76 | 0.87 |
| **Imaging + Clinical** | **81.2%** | **0.79** | **0.90** |

Risk factors show differential importance: betel chewing (0.23), smoking (0.19), age (0.16), alcohol (0.12).

### E. Computational Performance

Mobile deployment feasibility assessment:

TABLE VII: Mobile Device Performance

| Device | Time | Memory | Battery |
|---|---|---|---|
| iPhone 12 Pro | 180ms | 85MB | Minimal |
| Galaxy S21 | 195ms | 92MB | Minimal |
| Mid-range Android | 280ms | 98MB | Low |

### F. Explainability Evaluation

Clinician assessment of model explanations shows significant improvement:

TABLE VIII: Clinical Explainability Assessment

| Metric | Baseline | Ours | p-value |
|---|---|---|---|
| Lesion Alignment | 62.7% | **82.4%** | ¡0.001 |
| Clinical Relevance | 3.1 ± 1.1 | **4.2 ± 0.8** | ¡0.001 |
| Inter-rater ($\kappa$) | 0.48 | **0.71** | - |

TABLE IX: Ablation Study Results

| Component | ΔAcc | ΔF1 | Impact |
|---|---|---|---|
| Attention Segmentation | +12.1% | +0.15 | High |
| Progressive Learning | +8.7% | +0.11 | High |
| Uncertainty Quant. | +3.4% | +0.04 | Medium |
| Clinical Risk Factors | +2.9% | +0.03 | Medium |
| Few-Shot Adaptation | +7.1% | +0.09 | High |

### G. Ablation Study

Component contribution analysis:

We compare our approach with six established methods for both binary and multi-class tasks.

TABLE X: Binary classification baselines.

| Method | Data | Acc. | F1 | Time | UQ |
|---|---|---|---|---|---|
| ResNet-50 (Whole Img.) | SL | 76.2±2.1 | 0.73 | 280ms | × |
| U-Net (Seg.) | SL | 77.4±1.9 | 0.74 | 320ms | × |
| Attention U-Net | SL | 79.8±1.7 | 0.77 | 340ms | × |
| DANN | Cross | 78.3±1.8 | 0.75 | 295ms | × |
| **Ours (Seg. Only)** | **SL** | **81.2±1.4** | **0.79** | **180ms** | ✓ |

*1) Binary Classification Task (Cancer vs. Non-Cancer):* For binary classification, our approach reaches the highest accuracy and the fastest inference time. Gains over Attention U-Net (+1.4% accuracy) stem from the refined attention architecture and uncertainty-aware training rather than increased complexity. Inference time is 160ms faster despite similar network size.

*2) Multi-Class Classification Task (Healthy / Benign / OPMD / Cancer):*

1) **Multi-class complexity premium:** MobileNetV3 reaches 81.2% accuracy on binary classification but only 60.2% on four-class triage. Our full framework reaches 78.3% and recovers 18.1 percentage points of this gap.
2) **Component synergy:** Attention alone provides a +9.2% gain. Combined with curriculum learning, the gain rises to +18.1%. This indicates complementary effects: attention localizes relevant regions and curriculum learning emphasizes boundary cases.
3) **Domain generalization:** The cross-dataset gap for our full method (67.4% vs. 78.3% = $-10.9\%$) is the smallest among the approaches. Standard methods show gaps of $-10\%$ to $-12.6\%$. Our few-shot domain adaptation reduces this gap.
4) **Clinical utility:** Unlike binary approaches, the multi-class framework supports triage by distinguishing OPMD (monitoring), cancer (urgent referral), and benign lesions (reassurance). While 78.3% accuracy is lower than specialized binary systems (81.2%), the richer clinical recommendations justify this trade-off.

## V. Discussion

### A. Clinical Significance and Deployment Impact

The accuracy obtained by our four-class model is 78.3%, which, when coupled with well-calibrated uncertainty measures, can be applied in practical applications. The confidence level of prediction, which has an uncertainty value of less than 0.20, bypasses the expert verification process in 82% of instances.

TABLE XI: Multi-class classification baselines.

| Method | Acc. | F1 | Mobile | Transfer |
|---|---|---|---|---|
| MobileNetV3 Baseline | 60.2±1.8 | 0.56 | ✓ | 52.3 (-7.9) |
| ResNet-50 Baseline | 58.7±1.9 | 0.53 | × | 50.1 (-8.6) |
| Attention Only | 69.4±2.1 | 0.64 | ✓ | 58.7 (-10.7) |
| Progressive Only | 71.8±1.8 | 0.68 | ✓ | 59.2 (-12.6) |
| No Uncertainty | 74.1±1.6 | 0.71 | ✓ | 62.1 (-12.0) |
| **Full Framework** | **78.3±1.2** | **0.76** | ✓ | **67.4 (-10.9)** |

Important clinical achievements include:

1) **Uncertainty-guided triage:** The predictions in 18% of the OPMD models are automatically sent for expert examination based upon model uncertainty, similar
2) **Actionable multi-class output:** The model has three associated clinical pathways, namely routine follow-up care (Healthy/Benign), structured monitoring care (OPMD), and
3) **Mobile Real-time Deployment**: The model takes only 180ms for inference, thereby enabling the provision of care screenings in the absence of internet connectivity.

Although the accuracy has been slightly compromised compared to the 81.2% achieved in binary classification models, this certainly represents a trade-off that has particular value in a clinicanctery context. This is because binary models lack the ability to distinguish between the presence of OPMD and other non-cancerous diseases, leaving up to 40% of non-cancer diagnoses uncertain.

### B. Model Limitations and Deployment Requirements

The proposed approach has demonstrated potential for population-wide screenings, but there are also limitations that should be considered in the responsible use of the approach in the clinical setting.

**OPMD Detection Challenge:** The most challenging class to distinguish remains OPMD, having an F1 value of 0.71, as opposed to the highest F1 value of 0.85 in the Healthy class. This corresponds to the morphological overlap between the healthy lesions, as well as the OPMD lesions, in line with the lack of agreement between specialists ($\kappa$=0.68). In order to be safe, all predictions in the OPMD class, having an uncertainty level over 0.35, should be sent for verification (18% of the total in the OPMD class), but this remains practicable since confident predictions can exclude the remaining 82%.

**Cross-Population Generalization.** While within-domain accuracy reaches 78.3%, performance drops to 52.3% when evaluated under a public→Sri Lankan cross-dataset shift. Few-shot domain adaptation improves accuracy to 67.4%, but deployment in new regions requires validation with 50–100 labeled samples per target population. Notably, performance outside South Asia remains unvalidated.

**Smartphone Image Quality Dependence.** Failure cases are concentrated in images with severe specular reflection (>15%

area; 12% of test errors) and underexposed images (histogram peak $< 100$; 8% of errors). Deployment should incorporate automated image-quality assessment with user reprompting when thresholds are not met.

**Prospective Clinical Validation.** Current results are based on retrospective, single-snapshot datasets. Real-world evaluation requires prospective validation with images captured by non-specialist community health workers under field conditions. A prospective study (Q1–Q2 2026; $n = 200$ patients, rural Sri Lanka) is planned to assess sensitivity and specificity in real deployment environments.

### C. Comparison with Prior Mobile Systems

The existing smartphone-based models were mainly conceptualized in terms of binary classification outcomes, namely distinguishing between cancerous and non-cancerous conditions, but were not suited for quantifying uncertainties. On the other hand, the blueprint needed for effective implementations in resource-poor environments involves successfully distinguishing between no less than four classes, namely the *Healthy, Benign, OPMD,* and *Cancer* groups, along with estimating associated uncertainties. The proposed blueprint incurs only a loss of accuracy in the range of 2-3% but offers significantly improved values compared to binary models, as described in Table I, by enabling the filtering of around 82% of the results based on confident scores.

## VI. Conclusion

In this study, the proposed work provides a systematic approach toward developing an oral cancer diagnosis system for smartphones. The attention-guided segmentation approach using progressive learning improved the accuracy of the four-class classification of the lesions by 12.1 percentage points, achieving an accuracy of 78.3 percent.

The key innovations are in the form of uncertainty-aware segmentation, which generates estimates of clinical confidence, progressive learning strategies for multi-class classification, and few-shot domain adaptation, which helps reduce the discrepancy in cross-population performance. The proposed model can conduct real-time processing in mobile devices while also enabling the generation of explanations related to the domain.

In order to prove the applicability of this framework, an upcoming prospective trial ($n = 200$) has been initiated in partnership with oral cancer-screening centers located in rural Sri Lanka for the period spanning 2025-2026. In this trial, the functioning of the proposed diagnostic system, when executed by non-specialist rural health workers, shall also be ascertained. The assumptions made in this proposed research shall also be proved correct. Another pilot trial shall be initiated in Q3, 2026, in the context of Bangladesh, where the proposed diagnostic system shall be tested for inter-country universality, considering $n = 100$ patients.

The next research will include prospective-validation studies in the clinic, regulatory strategies, and the integration of the technology into current healthcare practices. If successful, this technology could result in the dramatic improvement of early rates of oral cancer detected in resource-poor countries, reducing the preventable mortality seen currently.

## References


[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] S. Warnakulasuriya, "Global epidemiology of oral and oropharyngeal cancer," *Oral Oncol.*, vol. 45, no. 4–5, pp. 309–316, Apr.–May 2009.

[3] D. Roblyer, R. Richards-Kortum, K. Sokolov, A. K. El-Naggar, M. D. Williams, C. Kurachi, and A. M. Gillenwater, "Multispectral optical imaging device for in vivo detection of oral neoplasia," *J. Biomed. Opt.*, vol. 13, no. 2, p. 024019, Mar./Apr. 2008.

[4] A. Krishna, S. Venkataramanan, A. Kumar, R. K. Gupta, and M. Singh, "Radiomics-based machine learning for prediction of oral cancer using CT imaging," *Med. Phys.*, vol. 47, no. 11, pp. 5654–5664, Nov. 2020.

[5] T. Y. Rahman, L. B. Mahanta, A. K. Das, and J. D. Sarma, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.

[6] G. Tanriver, M. S. Solak Tekin, S. K. Kabadayi, B. Y. Aktas, and G. S. Ozden, "Automated detection and classification of oral lesions using deep learning to detect oral potentially malignant disorders," *Cancers*, vol. 13, no. 11, p. 2766, Jun. 2021.

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[8] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, and others, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016.

[9] N. Das, A. K. Barui, S. Chattopadhyay, A. Basu, and S. Chatterjee, "Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis," *Tissue Cell*, vol. 53, pp. 111–119, Aug. 2018.

[10] M. Shaban, C. Khurram, M. Fraz, N. Alsubaie, I. Masood, S. Mushtaq, and S. Hassan, "A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma," *Sci. Rep.*, vol. 9, no. 1, p. 13381, Sep. 2019.

[11] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, and A. Criminisi, "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks," in *Proc. Int. Conf. Inf. Process. Med. Imaging*, Springer, 2017, pp. 597–608.

[12] L. Zhang, A. Rao, and S. Agner, "Generalizing deep learning for medical image segmentation to unseen domains via deep stacking," *IEEE Trans. Med. Imaging*, vol. 39, no. 7, pp. 2531–2540, Jul. 2020.

[13] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[14] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2017, pp. 1126–1135.

[15] shivam17299, "Oral Cancer (Lips and Tongue) images," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images. [Accessed: Sep. 15, 2025].

[16] muhammadatef, "Oral Cancer Images for Classification," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/muhammadatef/oral-cancer-images. [Accessed: Sep. 15, 2025].

[17] raymanodeep, "Oral Cancer V2 Part 1," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/raymanodeep/oral-cancer-v2-part-1. [Accessed: Sep. 15, 2025].