



In Vivo Multimodal Optical Imaging: Improved Detection of Oral Dysplasia in Low-Risk Oral Mucosal Lesions

Eric C. Yang^{1,2}, Richard A. Schwarz¹, Alexander K. Lang³, Nancy Bass³, Hawraa Badaoui⁴, Imran S. Vohra¹, Katelin D. Cherry¹, Michelle D. Williams⁵, Ann M. Gillenwater⁴, Nadarajah Vigneswaran³, and Rebecca R. Richards-Kortum¹

¹Department of Bioengineering, Rice University, Houston, Texas

²Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas

³The University of Texas Health Science Center at Houston School of Dentistry, Houston, Texas

⁴Department of Head and Neck Surgery, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

⁵Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Abstract

Early detection of oral cancer and oral premalignant lesions (OPLs) containing dysplasia could improve oral cancer outcomes. However, general dental practitioners have difficulty distinguishing dysplastic OPLs from confounder oral mucosal lesions in low-risk populations. We evaluated the ability of two optical imaging technologies, autofluorescence imaging (AFI) and high-resolution microendoscopy (HRME), to diagnose moderate dysplasia or worse (ModDys+) in 56 oral mucosal lesions in a low-risk patient population, using histopathology as the gold standard, and in 46 clinically normal sites. AFI correctly diagnosed 91% of ModDys+ lesions, 89% of clinically normal sites, and 33% of benign lesions. Benign lesions with severe inflammation were less likely to be correctly diagnosed by AFI (13%) than those without (42%). Multimodal imaging (AFI+HRME) had higher accuracy than either modality alone; 91% of ModDys+ lesions, 93% of clinically normal sites, and 64% of benign lesions were correctly diagnosed. Photos of the 56 lesions were evaluated by 28 dentists of varied training levels, including 26 dental residents. We compared the area under the receiver operator curve (AUC) of clinical impression alone to clinical impression plus AFI and clinical impression plus multimodal imaging using *k*-Nearest Neighbors models. The mean AUC of the dental residents was 0.71 (range: 0.45-0.86). The addition of AFI alone to clinical impression slightly lowered the mean AUC (0.68; range: 0.40-0.82), whereas the addition of multimodal imaging to clinical impression increased the mean AUC (0.79; range:

Corresponding author: Rebecca R. Richards-Kortum, PhD, Rice University, Department of Bioengineering MS142, 6100 Main Street, Houston, Texas 77005; Phone: 713-348-3823; Fax: (713) 348-5877; rkortum@rice.edu.

Potential Conflicts of Interest:

R. Richards-Kortum, A. M. Gillenwater, and R. A. Schwarz are recipients of licensing fees for intellectual property licensed from the University of Texas at Austin by Remicalm LLC. All other authors disclosed no potential conflicts of interest relevant to this publication.

0.61-0.90). Based on these findings, multimodal imaging could improve the evaluation of oral mucosal lesions in community dental settings.

Introduction

Oral cancer is a significant contributor to the global cancer burden. Worldwide, there are over 300,000 new cases of oral cancer each year, resulting in 145,000 deaths [1]. Rates of oral cancer are particularly high in South Asia due to betel quid chewing [2]. Despite advances in management [3], rates of survival have not improved significantly, with late stage diagnosis being a major reason. In the United States, 65% of diagnoses occur after regional metastasis; these patients have a 50% five-year survival rate. In contrast, the five-year survival rate is 83% if oral cancer is diagnosed when localized [4]. Globally, survival rates for oral cancer are even lower than in the US [5]; late diagnosis is even more common in low- and middle-income countries.

Most oral cancers are preceded by oral premalignant lesions (OPLs), a group of oral mucosal lesions including leukoplakia, erythroplakia, and submucous fibrosis [6,7]. OPLs containing dysplasia have a particularly high risk of malignant progression. Dysplasia ranges in grade from mild to moderate to severe, with worse grades having higher risk [8]. Early detection of oral cancers and dysplastic OPLs could dramatically improve outcomes.

Over 60% of adults visit a dentist each year [9], making routine dental care a promising setting in which to improve early detection. General dental practitioners (GDPs) face two major barriers to improved early detection. First, the majority of oral mucosal lesions identified during routine dental visits are not OPLs and instead are benign confounders – innocuous inflammatory or reactive oral mucosal lesions [10–16]. Most GDPs lack the skills to distinguish OPLs from these benign confounders. In one study, 32 Spanish GDPs shown 50 photos of oral mucosal lesions and their clinical history discriminated oral cancer and OPLs from benign confounders with a 57.8% sensitivity and 53% specificity [17]. Second, even in the hands of expert providers, conventional oral exam (COE) cannot accurately distinguish dysplastic OPLs from non-dysplastic OPLs. A recent meta-analysis concluded that COE had a 93% sensitivity but only a 31% specificity to distinguish dysplasia and cancer from benign lesions [18]. Therefore, biopsies are required for definitive diagnosis, but they are highly invasive, resource intensive, do not provide immediate results, and often result in underdiagnosis due to sampling bias. Additionally, many GDPs are reluctant to perform biopsies because of unfamiliarity with biopsy technique and lack of faith in choosing representative biopsy sites [19–21]. Diagnostic adjuncts capable of distinguishing OPLs from benign confounders, distinguishing dysplastic OPLs from non-dysplastic OPLs, and guiding biopsies to the most abnormal part of a lesion could help GDPs better identify patients with high-risk OPLs for diagnosis and referral.

Current guidelines do not recommend commercially available point of care adjuncts such as toluidine blue, acetowhitening, and autofluorescence imaging (AFI) [22,23], in part due to low accuracy in distinguishing OPLs from benign confounders. For example, AFI consists of the noninvasive, macroscopic imaging of native tissue fluorescence under blue excitation light. AFI has a high sensitivity for the detection of dysplasia and cancer, which are

associated with loss of fluorescence. As dysplasia develops, increased epithelial metabolic activity, thickness, and scattering combined with stromal microvascularization and collagen crosslink degradation result in decreased blue-green fluorescence and occasional increased red fluorescence associated with protoporphyrins [24,25]. Our group has found that dysplasia and cancer are therefore associated with an elevated red to green fluorescence ratio relative to the same ratio in normal oral mucosa (called the normalized RG ratio) [26]. The major limitation of AFI is that benign lesions frequently demonstrate a loss of fluorescence similar to dysplasia and cancer, leading to false positives. Stromal inflammation, which is common in benign confounders, may be responsible [27]. In one study of 126 patients with suspicious lesions, AFI had a sensitivity of 84.1% for oral dysplasia but a specificity of only 15.3% [28]. This limitation is critical in community dental clinics, where benign confounders are more prevalent.

Our group has developed a device called the high-resolution microendoscope (HRME) capable of identifying dysplasia and cancer in the oral mucosa, cervix, and esophagus [29–34]. The HRME is a flexible fiber optic fluorescence microscope that can image epithelial nuclei using the topical contrast agent proflavine [35]. Nuclear features of dysplasia and cancer such as increased nuclear to cytoplasm (NC) ratio and nuclear crowding are easy to assess in HRME images. The HRME images only the surface epithelium, resulting in improved specificity even in the presence of stromal inflammation [31,34]. However, the HRME has a <1 mm field of view, which is too small to assess an entire lesion.

We hypothesize that combining AFI with HRME (“multimodal imaging”) could allow clinicians to exploit the advantages of each modality. With its large field of view, AFI could identify suspicious regions with high sensitivity followed by HRME imaging within those regions to improve specificity. Previous studies showed that the combination of normalized RG ratio and nuclear features from AFI and HRME images, respectively, accurately distinguished benign oral mucosa from moderate to severe dysplasia and cancer in patients receiving surgery for oral lesions, most of which contained cancer or high-grade dysplasia [31,34].

The goal of this study was to explore the use of AFI and HRME in evaluating more common, low-risk oral mucosal lesions typically encountered in a general dental practice. First, we assessed the diagnostic performance of AFI and HRME individually and combined to detect moderate dysplasia to cancer in oral mucosal lesions, using histopathology as the gold standard, and in clinically normal mucosa. Then, dentists of varied training levels evaluated photos of these same lesions; we compared the area under the receiver operator curve (AUC) of their clinical impression alone to that of clinical impression plus AFI and to clinical impression plus both imaging modalities.

Materials and Methods

Human Subjects

The study was performed at the UTHealth School of Dentistry in Houston, Texas (UTSD-Houston) in accordance with recognized ethical guidelines. Protocols were approved by the Institutional Review Boards at UTSD and Rice University. Patients 18 years or older with at

least one oral mucosal lesion were recruited. Written informed consent was obtained from all subjects.

Instrumentation

The AFI and HRME devices have been previously described [31,34]. The AFI device images a 4.5 cm diameter field of view (FOV) with a 100 μm lateral spatial resolution and collects images in two modes: reflectance imaging under white-light illumination, and autofluorescence imaging under 405 nm excitation light. The HRME is a compact epifluorescence fiber optic microscope with a 720 μm diameter field of view and 4.4 μm lateral spatial resolution. Imaging occurs with 455 nm excitation light following topical application of 0.01% w/v proflavine, a fluorescent contrast agent that stains cell nuclei. The HRME collects data in 3-second videos at 10 frames per second. The Both devices were built using off-the-shelf components with a combined cost of ~\$5000 and interface with a consumer grade laptop.

Clinical Data Collection

Clinical Evaluation—An oral pathologist with expertise in oral mucosal diseases (‘expert clinician’, NV) inspected the oral mucosa of subjects and identified one or more clinically abnormal areas defined as oral mucosal lesions. In some subjects, the expert clinician also identified areas of clinically normal mucosa, adjacent or distant to the lesion(s). The expert clinician then provided a clinical impression at each lesion and area of clinically normal mucosa using the following scale: 1) normal oral mucosa, 2) oral mucosal lesion, not suspicious for dysplasia or cancer, 3) oral mucosal lesion, suspicious for dysplasia or cancer, and 4) cancer.

Imaging Procedure—The AFI device acquired a series of images at each lesion and area of clinically normal mucosa (henceforth known as “sites”) using both white-light reflectance and autofluorescence modes. Room lights were turned off to minimize the effects of ambient light. Then, proflavine was applied to each site with a cotton-tipped swab. The clinician placed the HRME probe directly on each site and acquired a series of videos. Because the probe has a small field of view, only a portion of the site was imaged. The HRME is insensitive to ambient light, so imaging was performed with room lights on.

Histopathology—Biopsies were performed according to standard of care, independent of imaging. Tissue specimens were evaluated histopathologically using standard UTSD-Houston procedures and later reviewed by an expert oral pathologist (NV). Each biopsy was categorized as benign (no dysplasia or mild dysplasia) or ModDys+ (moderate dysplasia, severe dysplasia, or oral squamous cell carcinoma). Mild dysplasia was considered benign due to its low risk for malignant progression and the difficulty of distinguishing mild dysplasia from benign reactive epithelial atypia in this patient population. The presence of stromal inflammation was also recorded and classified as severe or not severe. No clinically normal sites were biopsied.

Image Analysis

Analyzed Sites—Lesions were analyzed if and only if an incisional or excisional biopsy of the lesion was obtained on the same day as imaging. All imaged clinically normal sites were analyzed.

Widefield Autofluorescence—A single autofluorescence image for each site was selected by a reviewer blinded to histopathology results based on lack of motion artifact, lack of saturated pixels, good focus, and visible oral mucosa surrounding the site. Then, the normalized RG ratio of the site was calculated as previously described [34].

Briefly, the normalized RG ratio was defined as the average ratio of the red intensity to the green intensity at each pixel in the site, divided by the same quantity in a region of normal mucosa. The normal region was defined as the 65×65 square of mucosa with the lowest RG ratio and was identified using an automated algorithm as follows. First, the mucosa in the image was outlined to exclude teeth and dental instruments. Then, the RG ratio at each pixel within the mucosa was calculated and smoothed with a 65×65 mean filter. The pixel with the lowest value after smoothing was set as the center of the normal region. All calculations were performed using an automated MATLAB (The Mathworks, Natick, Massachusetts) script.

One lesion consisted of two discrete regions of differing autofluorescence. A histogram-based method was used to segment the lesion into two regions, and the region with the higher RG ratio was selected to represent the site. This choice mimics histopathology, in which the area within a biopsy with the worst histopathologic findings is most representative.

HRME—At least one high quality image was selected from the HRME videos corresponding to each site by a reviewer blinded to histopathology results. In cases where HRME images were obtained from different locations within a site, multiple high quality HRME images were identified. The quality of an image was determined subjectively by a combination of factors including 1) good focus, 2) lack of motion artifact, and 3) visible nuclei in >50% of the FOV. Sites without an image of sufficient quality were excluded from analysis. The nuclei in each selected image were segmented to calculate the nuclear-to-cytoplasm ratio (NC ratio) using a custom MATLAB script. To mimic histopathology, sites with more than one selected image were represented by the image with the highest NC ratio.

Evaluation of Diagnostic Accuracy—To assess the diagnostic performance of AFI and HRME, sites with images exceeding a threshold normalized RG ratio or NC ratio, respectively, were classified as ModDys+. The thresholds were set to correspond to 91% sensitivity, and specificity was evaluated at those thresholds. Similarly, the performance of AFI plus HRME was assessed using a linear discriminant based on the normalized RG ratio and NC ratio corresponding to a 91% sensitivity. The specificity was evaluated for this classifier. Positive predictive value (PPV) and negative predictive value (NPV) of AFI, HRME, and AFI plus HRME were also calculated for lesions. The Wilson score interval was used to calculate 95% confidence intervals (CIs), and McNemar's test with Yates continuity correction was used to compare specificities. Histopathology was the gold

standard for lesions, and clinically normal sites were assumed to be benign. The sensitivity and specificity of lesions were used to calculate positive and negative likelihood ratios (LR +, LR-), with 95% CIs calculated as described by Simel et al [36]. Finally, two-tailed chi-square tests were used to assess whether benign lesions with severe stromal inflammation were more likely to be false positives by AFI, HRME, and AFI plus HRME than those without severe stromal inflammation.

Dentist Clinical Impression Survey

A survey of 26 dental residents of varied specialties at UTSD-Houston ('non-experts') and two faculty-level oral pathology and oral medicine specialists ('other experts') was conducted to obtain additional clinical impression data for each analyzed lesion. Participants were shown the reflectance white-light images of all analyzed lesions along with the patients' age and sex. They then rated each site with the same clinical impression scale used by the 'expert clinician'. AUCs were calculated using the 1-4 scale to assess each of the 28 surveyed dentist's ability to distinguish ModDys+ from benign lesions. Sensitivity, specificity, PPV, NPV, LR+, and LR- were also calculated, considering clinical impressions ≥ 3 as ModDys+. The same metrics were calculated for the expert clinician, for a total of 29 dentists.

Clinical Impression Combined with Imaging

To assess whether the addition of imaging information could improve the 29 dentists' diagnostic ability, two diagnostic algorithms were developed using the k -Nearest Neighbors (k -NN) model. k -NN models set the probability that a lesion is ModDys+ equal to the percentage of the k lesions with the most similar clinical impression and imaging results that are ModDys+. The first algorithm combined clinical impression with the normalized RG ratio from AFI, and the second algorithm combined clinical impression with the normalized RG ratio from AFI and the NC ratio from HRME. The AUC of each dentist after incorporating imaging information was determined with the probabilities calculated by the k -NN algorithms. Two-tailed paired t-tests were used to determine if the changes in the mean AUC of the non-experts due to imaging information were statistically significant. They were not performed for the expert clinician or additional experts due to low sample size.

To avoid overfitting, AUCs were calculated by averaging 50 runs of 10-fold stratified cross-validation (Figure 1). The full dataset consisted of 29 clinical impressions paired with imaging feature(s) and a pathologic diagnosis for each analyzed lesion. For each fold, the training set (Figure 1, white squares) consisted of all 29 clinical impressions paired with imaging feature(s) and pathologic diagnosis for nine-tenths of the lesions. The corresponding validation set (Figure 1, black squares) contained the same data for the remaining one-tenth of the lesions. Each training set was used to train a k -NN algorithm, which predicted the probability that the lesions in the corresponding validation set were ModDys+ for each of the 29 dentists. These probabilities were used to calculate each dentist's AUC with imaging information incorporated, which were averaged across the 10 validation sets. The final AUC of each dentist was the average of all 50 cross-validation runs. Calculations were performed with a custom MATLAB script, and the k -NN algorithms

were trained with MATLAB's *fitcknn* function using the Euclidian distance metric and $k = 350$.

Results

Analyzed Sites

Seventy-two lesions in 69 patients were imaged with the AFI and HRME and received a clinically indicated incisional or excisional biopsy on the same day. Of these, one lesion was excluded from further analysis due to a histopathologic diagnosis of koilocytic dysplasia, and another was excluded due to a diagnosis of lymphoma. Koilocytic dysplasia in the oral epithelium is thought to be associated with HPV, but its potential for malignant transformation is unknown [37]. While lymphoma is malignant, the purpose of the study was to diagnose oral squamous cell carcinoma and its precursors. Fourteen lesions were excluded due to lack of a sufficient quality HRME image. Forty-nine clinically normal sites in 48 patients were imaged with the AFI and HRME; of these, three were excluded due to lack of a sufficient quality HRME image.

In sum, data from 102 sites in 68 patients were analyzed, including 56 lesions in 54 patients and 46 clinically normal sites in 45 patients (Table 1). Twenty percent of the lesions were ModDys+ (11/56), and 80% of the lesions were benign (45/56). The anatomic sites of the 56 oral lesions were: buccal (20), dorsal tongue (3), gingiva (9), palate (5), ventrolateral tongue (19). The anatomic locations of the 46 clinically normal sites were: buccal (18), dorsal tongue (3), gingiva (5), palate (3), ventrolateral tongue (17).

Representative Sites

Figure 2 shows the images obtained from four representative sites. The first column depicts a clinically normal site (Figure 2A, white arrow) on the right buccal mucosa posterior to a histopathologically diagnosed benign fibroma. No loss of fluorescence is visually apparent in the corresponding autofluorescence image (Figure 2B) when comparing the site (red rectangle) and the automatically selected normal region (green square) based on the manually outlined mucosa (green polygon). The normalized RG ratio was 1.24. The nuclei in the HRME image (Figure 2C) are small, circular, and evenly spaced, consistent with benign epithelium. The NC ratio was 0.04.

The second column shows a lesion (Figure 2D) on the dorsal tongue with an expert clinician clinical impression of “oral mucosal lesion, not suspicious for dysplasia or cancer”. The lesion does not demonstrate loss of fluorescence (Figure 2E), and the normalized RG ratio was 1.01. The nuclei in the HRME image (Figure 2F) are small and evenly spaced, with an NC ratio of 0.14. The site was diagnosed histopathologically as a benign fibroma negative for severe inflammation, consistent with the expert clinician's clinical impression, AFI, and HRME.

The third column shows a lesion (Figure 2G) on the right buccal mucosa with an expert clinician clinical impression of “oral mucosal lesion, suspicious for dysplasia or cancer” and visible loss of fluorescence on the AFI image (Figure 2H). The normalized RG ratio was 2.04. However, the nuclei on the HRME image (Figure 2I) are small and evenly spaced with

an NC ratio of 0.17, consistent with benign mucosa. The histopathologic diagnosis was lichenoid mucositis, a benign lesion. Therefore, this site was a false positive by the expert clinician's clinical impression and AFI but was correctly diagnosed by HRME. The lesion was positive for severe stromal inflammation, consistent with the hypothesis that inflammation is associated with loss of autofluorescence leading to false positive results.

The fourth column shows a lesion (Figure 2J) on the right gingiva with an expert clinician clinical impression of cancer. The site is located in the dark region of the autofluorescence image (Figure 2K), indicating loss of fluorescence compared to the brighter normal region. The normalized RG ratio was 2.41. Unlike the other three sites, the nuclei in the HRME image (Figure 2L) are large, eccentric, and irregularly spaced with an NC ratio of 0.18, consistent with ModDys+. The site was histopathologically diagnosed as moderately differentiated squamous cell carcinoma, consistent with expert clinician clinical impression, AFI, and HRME. The site was positive for severe inflammation.

Diagnostic Performance of Imaging

To assess performance more systematically, the NC ratio vs. normalized RG ratio of each site was plotted (Figure 3A-B). A normalized RG ratio threshold of 1.83 (Figure 3A, vertical dotted line) correctly diagnosed 91% (10/11; 95% CI: 62%-98%) of ModDys+ lesions and 89% (41/46; 95% CI: 77%-95%) of clinically normal sites, but only 33% (15/45; 95% CI: 21%-48%) of benign lesions. The PPV and NPV for lesions were 25% (10/40; 95% CI: 14%-40%) and 94% (15/16; 95% CI: 72%-99%), respectively. The LR+ and LR- were 1.36 (95% CI: 1.03-1.80) and 0.27 (95% CI: 0.04-0.51), respectively. An NC ratio threshold of 0.18 (Figure 3A, horizontal dotted line) correctly diagnosed 91% (10/11; 95% CI: 62%-98%) of ModDys+ lesions, 85% (39/46; 95% CI: 72%-92%) of clinically normal sites, and 49% (22/45; 95% CI: 35%-63%) of benign lesions. The PPV and NPV for lesions were 30% (10/33; 95% CI: 17%-47%) and 96% (22/23; 95% CI: 79%-99%), respectively. The LR+ and LR- were 1.78 (95% CI: 1.26-2.50) and 0.19 (95% CI: 0.03-0.31), respectively. The multimodal imaging linear threshold (Figure 3A, diagonal solid line) incorporating both AFI and HRME correctly diagnosed 91% (10/11; 95% CI: 62%-98%) of ModDys+ lesions, 93% (43/46; 95% CI: 83%-98%) of clinically normal sites, and 64% (29/45; 95% CI: 50%-77%) of benign lesions. The PPV and NPV for lesions were 38% (10/26; 95% CI: 22%-57%) and 97% (29/30; 95% CI: 83%-99%), respectively. The LR+ and LR- were 2.56 (95% CI: 1.65-3.95) and 0.14 (95% CI: 0.02-0.21), respectively. The difference in percentage of correctly classified benign lesions between AFI only and AFI plus HRME was statistically significant ($p = 0.001$); the other comparisons did not reach statistical significance. These results are summarized in Figures 3C and 3D.

To assess the association between a false positive diagnosis and inflammation, a scatterplot of the NC ratio vs. the normalized RG ratio for the benign lesions stratified by inflammation status is shown in Figure 3B. The three mild dysplasia sites were excluded because cellular atypia could explain a suspicious imaging result. Nearly all of the benign lesions (39/42) contained stromal inflammation, so only severe stromal inflammation (16/42) was considered positive. The normalized RG ratio threshold of 1.83 (Figure 3B, vertical dashed line) correctly classified a lower percentage of inflammation-positive benign lesions (13%,

2/16) than inflammation-negative benign lesions (42%, 11/26) ($p = 0.042$). The NC ratio threshold of 0.1801 (Figure 3B, horizontal dashed line) correctly classified 44% (7/16) of inflammation-positive benign lesions and 54% (14/26) of inflammation-negative benign lesions ($p = 0.525$). The multimodal imaging linear threshold (Figure 3B, diagonal solid line) correctly classified 50% (8/16) of inflammation-positive benign lesions and 73% (19/26) of inflammation-negative benign lesions ($p = 0.130$).

Clinical Impression

As shown in Figure 4, of the three groups of dentists providing clinical impressions of the 56 analyzed lesions, the ‘expert clinician’ had the highest performance to diagnose ModDys+ (AUC = 0.88), followed by the two ‘other experts’ (mean AUC 0.77, range: 0.76–0.78), and the 26 ‘non-experts’ (mean AUC 0.71, range: 0.45–0.86). If clinical impressions ≥ 3 were considered ModDys+, the ‘expert clinician’ correctly diagnosed 100% of ModDys+ lesions and 73% of benign lesions, corresponding to a PPV of 48%, NPV of 100%, LR+ of 3.70 (95% CI: 2.29 to 5.99) and LR- of 0. The ‘other experts’ correctly diagnosed 77% (range: 73%–82%) of ModDys+ lesions and 77% (range: 73%–81%) of benign lesions, corresponding to a PPV of 46% (range: 43%–50%), NPV of 93% (range: 92%–94%), LR+ of 3.35 (95% CI: 1.79 to 6.25) and LR- of 0.30 (95% CI: 0.10 to 0.42). Finally, the ‘non-experts’ correctly diagnosed 79% (range: 30%–100%) of ModDys+ lesions and 65% (range: 29%–89%) of benign lesions, corresponding to a PPV of 35% (range: 15%–62%), NPV of 87% (range: 79%–100%), LR+ of 2.26 (95% CI: 1.37–3.37) and LR- of 0.32 (95% CI: 0.10–0.48). These results are summarized in Figure 5A–B.

The AUC of each of the 29 dentists after incorporating imaging information was estimated using cross-validation of k -NN models (Figure 1). The addition of the normalized RG ratio to clinical impression lowered the AUC for all three groups; the mean AUCs were 0.84, 0.74 (range: 0.70–0.77), and 0.68 (range: 0.40–0.82; $p = 0.018$) for the ‘expert clinician’, ‘other experts’, and ‘non-experts’, respectively. In contrast, the combination of clinical impression, the normalized RG ratio from AFI, and the NC ratio from HRME increased the mean AUC of all three groups. The AUC of the ‘expert clinician’ increased from 0.88 to 0.90, the mean AUC of the two ‘other experts’ increased from 0.77 to 0.86 (range: 0.84–0.87), and the mean AUC of the twenty-six non-experts increased from 0.71 to 0.79 (range: 0.61–0.90; $p < 10^{-6}$).

Discussion

In this study, we explored the diagnostic value of AFI and HRME to classify dysplastic or cancerous and benign oral mucosal lesions, including many non-OPL confounder lesions, as well as clinically normal mucosa in a low-risk population typical of community dental clinics. Then, we assessed the potential benefit of imaging to the clinical impression of dentists of varied levels with training.

AFI accurately diagnosed ModDys+ lesions (91% correct) and clinically normal sites (89% correct), but was ineffective for benign lesions (33% correct). Benign lesions with severe inflammation were significantly more likely to be false positives. These results suggest that AFI can visualize oral mucosal lesions but cannot distinguish ModDys+ lesions from benign lesions in the community setting. Accordingly, the VELscope (LED Dental, Atlanta,

Georgia), an AFI device, is FDA approved to “enhance the identification and visualization of oral mucosal abnormalities that may not be apparent or visible to the naked eye” but not to predict pathology [38].

The HRME was more accurate than AFI, correctly diagnosing 91% of ModDys+ lesions, 85% of clinically normal sites, and 49% of benign lesions, although the differences were not statistically significant. It is unclear why the accuracy was lower for benign lesions than for clinically normal sites. Unlike AFI, severe stromal inflammation did not significantly affect the false positive rate of benign lesions. The combination of HRME and AFI (“multimodal imaging”) improved the diagnostic ability of either modality alone, correctly diagnosing 91% of ModDys+ lesions, 93% of clinically normal sites, and 64% of benign lesions. The increase in performance for benign lesions was statistically significant compared to AFI alone.

We also surveyed dentists of varied training levels to assess their ability to distinguish ModDys+ lesions from benign lesions. The oral pathologist treating the patient had the highest performance (AUC = 0.88), followed by the two specialists (AUC = 0.77) and the 26 dental residents (AUC = 0.71) viewing white-light images of the lesions. In a general practice setting, oral lesions are ideally evaluated with high NPV, but not at the expense of an excessively low PPV. At this operating point, the dental residents had worse NPV (97% vs. 87%), PPV (38% vs. 35%), LR+ (2.26 vs. 2.56), and LR- (0.32 vs. 0.14) than AFI and HRME combined, although these values were within the 95% confidence intervals for multimodal imaging. The large ranges in AUC (0.45-0.86), sensitivity (30%-100%), specificity (29%-89%), PPV (15%-62%), and NPV (79%-100%) of the dental residents, a population with training similar to GDPs, highlights the potential benefit of automated algorithms to reduce variation.

Finally, we assessed the impact of adding imaging information to clinical impression. The addition of AFI to clinical impression slightly decreased the AUC of the expert clinician, other specialists, and dental residents. This decrease was statistically significant for the dental residents, indicating that AFI does not provide diagnostic information helpful to GDPs for this population. On the other hand, the addition of HRME and AFI to clinical impression improved the AUC of all three groups, with the increase for dental residents being statistically significant. These results indicate that HRME provides information unavailable through visual assessment.

Previous studies of these imaging modalities in high-risk surgical patients at MD Anderson Cancer Center found that AFI provided significant diagnostic value in that setting [31,34]. The patients in this study were from a community dental clinic at UTSD-Houston and reflect the low-risk patients presenting to community dentists, including a diverse array of non-OPL oral mucosal lesions. The difference in AFI’s performance between the two settings underscores the importance of studying diagnostic adjuncts in a variety of populations.

Overall, our results point to the potential value of AFI and HRME to GDPs. Only 11 of the 45 biopsied lesions were ModDys+; the other 34 biopsies were theoretically unnecessary. AFI and HRME could help reduce these unnecessary biopsies, and identify ModDys+ in

lesions that would not have been biopsied. In the US, a typical biopsy costs several hundred dollars. With a one-time instrumentation cost of a few thousand dollars, cost savings could be achieved quickly by reducing the number of unnecessary biopsies. Improved early diagnosis could also lead to significant cost savings. Other potential benefits include biopsy site guidance and improved surveillance. In practice, GDPs can interpret HRME images subjectively and/or objectively with automated algorithms. Two studies have found that non-pathologists can accurately distinguish HRME images of cancerous tissue from benign tissue with high inter-rater agreement and little training [39,40]. Although low-quality images and dysplasia images were excluded from both studies, they suggest that GDPs could learn the fundamentals of subjective HRME image interpretation. The performance metrics in this study were based on objective algorithms, establishing their feasibility in this setting.

A weakness of this study is that only 11 ModDys+ lesions were imaged, a challenge in low-prevalence populations. A larger study could validate these results and provide data to optimize the image analysis algorithms without overfitting. The HRME algorithm could be improved with better methods to identify and exclude debris or keratin from segmentation, utilization of additional features, and more complex classification models. A second weakness is that surveyed dentists were unable to palpate the lesions and did not have a clinical history, so the results may differ from their true clinical performance.

Since this study was conducted, we have made improvements to the HRME. In this study, HRME data were collected as videos, and selecting image frames required significant manual labor. A newly developed second generation HRME features a foot pedal that can pause and unpause the image feed [30]. The clinician can pause at a quality image and save it for real-time analysis, eliminating the need for manual selection. Alternatively, we have developed an algorithm that automates frame selection [41]. It was also difficult to visualize nuclei in lesions with a superficial keratin layer. We are testing if a mechanical tool similar to a brush biopsy could remove surface keratin to allow for successful imaging.

In the future, AFI and HRME could be integrated such that AFI first identifies high-risk regions within a large lesion followed by HRME imaging at those regions. This procedure would rapidly assess an entire field of mucosa and help clinicians select a biopsy site. With these advances, multimodal imaging has the potential to improve the evaluation of oral lesions in the community and help address the global oral cancer burden.

Acknowledgments

This work was supported by National Institutes of Health grants R01 CA103830 (to R. Richards-Kortum); R01 CA185207 (to R. Richards-Kortum); R01 DE024392 (to N. Vigneswaran); F30 CA213922 (to E. Yang) and by the Cancer Prevention and Research Institute of Texas (CPRIT) grant RP100932 (to R. Richards-Kortum). We would like to thank Sharon Mondrik, M.K. Quinn, and Travis King, previously of Rice University, for their contributions to data collection, and Mark E. Wong, DDS, of UTSD-Houston, for his assistance with the dental resident survey. Additionally, we would like to thank Melody T. Tan of Rice University, Jessica Rodriguez, PA, and Justin Jacob of MD Anderson Cancer Center, for their insight and feedback.

References

1. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, DM., Forman, D., Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, Fr Int Agency Res Cancer. 2013. Available from: <http://globocan.iarc.fr/Default.aspx>
2. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*. 2009; 45:309–16. [PubMed: 18804401]
3. Rapidis, AD., Gullane, P., Langdon, JD., Lefebvre, JL., Scully, C., Shah, JP. Major advances in the knowledge and understanding of the epidemiology, aetiopathogenesis, diagnosis, management and prognosis of oral cancer; *Oral Oncol Elsevier*. 2009. p. 299-300. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19411038>
4. Cancer of the Oral Cavity and Pharynx - SEER Stat Fact Sheets. Available from: <http://seer.cancer.gov/statfacts/html/oralcav.html>
5. Sankaranarayanan, R., Black, RJ., Swaminathan, R., Parkin, DM. An overview of cancer survival in developing countries; *IARC Sci Publ*. 1998. p. 135-73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10194635>
6. Warnakulasuriya, S., Johnson, NW., Van Der Waal, I. Nomenclature and classification of potentially malignant disorders of the oral mucosa; *J Oral Pathol Med Blackwell Publishing Ltd*. 2007. p. 575-80. Available from: <http://dx.doi.org/10.1111/j.1600-0714.2007.00582.x>
7. Napier, SS., Speight, PM. *J Oral Pathol Med*. Vol. 37. Blackwell Publishing Ltd; 2008. Natural history of potentially malignant oral lesions and conditions: an overview of the literature; p. 1-10. Available from: <http://dx.doi.org/10.1111/j.1600-0714.2007.00579.x>
8. van der Waal, I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management; *Oral Oncol Elsevier*. 2009. p. 317-23. Available from: <http://dx.doi.org/10.1016/j.oraloncology.2008.05.016>
9. National Center for Health Statistics. Health, United States 2015 With Special Feature on Racial and Ethnic Health Disparities. 2015th. Hyattsville, Maryland: 2015.
10. Saraswathi, TR., Ranganathan, K., Shanmugam, S., Sowmya, R., Narasimhan, PD., Gunaseelan, R. Prevalence of oral lesions in relation to habits: Cross-sectional study in South India; *Indian J Dent Res*. p. 121-5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17176828>
11. SHULMAN JD, BEACH MM, RIVERA-HIDALGO F. The prevalence of oral mucosal lesions in U.S. adults: Data from the Third National Health and Nutrition Examination Survey, 1988–1994. *J Am Dent Assoc*. 2004; 135:1279–86. [PubMed: 15493392]
12. Bouquot, JE. Common oral lesions found during a mass screening examination; *J Am Dent Assoc*. 1986. p. 50-7. Available from: <http://www.sciencedirect.com/science/article/pii/S000281778621017X>
13. Mumcu, G., Cimilli, H., Sur, H., Hayran, O., Atalay, T. *Oral Dis*. Vol. 11. Munksgaard International Publishers; 2005. Prevalence and distribution of oral lesions: a cross-sectional study in Turkey; p. 81-7. [cited 2016 Nov 14] Available from: <http://doi.wiley.com/10.1111/j.1601-0825.2004.01062.x>
14. Mathew, AL., Pai, KM., Sholapurkar, AA., Vengal, M. The prevalence of oral mucosal lesions in patients visiting a dental school in Southern India; *Indian J Dent Res*. p. 99-103. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18445924>
15. Axéll, T. A prevalence study of oral mucosal lesions in an adult Swedish population; *Odontol Revy Suppl*. 1976. p. 1-103. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/186740>
16. Kovac-Kavcic, M., Skaleric, U. *J Oral Pathol Med*. Vol. 29. Munksgaard International Publishers; 2000. The prevalence of oral mucosal lesions in a population in Ljubljana, Slovenia; p. 331-5. Available from: <http://doi.wiley.com/10.1034/j.1600-0714.2000.290707.x>
17. Seoane, J., Warnakulasuriya, S., Varela-Centelles, P., Esparza, G., Dios, P. *Oral Dis*. Vol. 12. Blackwell Publishing Ltd; Oral cancer: experiences and diagnostic abilities elicited by dentists in North-western Spain; p. 2006p. 487-92. Available from: <http://doi.wiley.com/10.1111/j.1601-0825.2005.01225.x>

18. Epstein JB, Güneri P, Boyacioglu H, Abt E. The limitations of the clinical oral examination in detecting dysplastic oral lesions and oral squamous cell carcinoma. *J Am Dent Assoc.* 2012; 143:1332–42. [PubMed: 23204089]
19. Diamanti, N., Duxbury, AJ., Ariyaratnam, S., Macfarlane, TV. *Br Dent J.* Vol. 192. Nature Publishing Group; 2002. Attitudes to biopsy procedures in general dental practice; p. 588-92. Available from: <http://www.nature.com/doi/10.1038/sj.bdj.4801434>
20. Greenwood, LF., Lewis, DW., Burgess, RC. How competent do our graduates feel?; *J Dent Educ American Dental Education Association.* 1998. p. 307-13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9603445>
21. Wan, A., Savage, N. *Aust Dent J.* Vol. 55. Blackwell Publishing Ltd; Biopsy and diagnostic histopathology in dental practice in Brisbane: usage patterns and perceptions of usefulness; p. 2010p. 162-9. Available from: <http://doi.wiley.com/10.1111/j.1834-7819.2010.01210.x>
22. Macey, R., Walsh, T., Brocklehurst, P., Kerr, AR., Liu, JL., Lingen, MW., et al. Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. In: Macey, R., editor. *Cochrane Database Syst Rev.* Chichester, UK: John Wiley & Sons, Ltd; 2015. Available from: <http://doi.wiley.com/10.1002/14651858.CD010276.pub2>
23. Lingen, MW., Abt, E., Agrawal, N., Chaturvedi, AK., Cohen, E., D'Souza, G., et al. *J Am Dent Assoc.* Vol. 148. Elsevier; 2017. Evidence-based clinical practice guideline for the evaluation of potentially malignant disorders in the oral cavity: A report of the American Dental Association; p. 712-727.e10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28958308>
24. Lane, PM., Gilhuly, T., Whitehead, P., Zeng, H., Poh, CF., Ng, S., et al. Simple device for the direct visualization of oral-cavity tissue fluorescence; *J Biomed Opt International Society for Optics and Photonics.* 2006. p. 24006 Available from: <http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.2193157>
25. Inaguma, M., Hashimoto, K. *Cancer.* Vol. 86. John Wiley & Sons, Inc; Porphyrin-like fluorescence in oral cancer; p. 1999p. 2201-11. Available from: <http://doi.wiley.com/10.1002/%28SICI%291097-0142%2819991201%2986%3A11%3C2201%3A%3AAID-CNCR5%3E3.0.CO%3B2-9>
26. Roblyer D, Kurachi C, Stepanek V, Williams MD, El-Naggar AK, Lee JJ, et al. Objective Detection and Delineation of Oral Neoplasia Using Autofluorescence Imaging. *Cancer Prev Res.* 2009; 2
27. Pavlova I, Williams M, El-Naggar A, Richards-Kortum R, Gillenwater A. Understanding the Biological Basis of Autofluorescence Imaging for Oral Cancer Detection: High-Resolution Fluorescence Microscopy in Viable Tissue. *Clin Cancer Res.* 2008; 14 [PubMed: 18413830]
28. Awan KH, Morgan PR, Warnakulasuriya S. Evaluation of an autofluorescence based imaging system (VELscope™) in the detection of oral potentially malignant disorders and benign keratoses. *Oral Oncol.* 2011; 47:274–7. [PubMed: 21396880]
29. Quinn, MK., Bubi, TC., Pierce, MC., Kayembe, MK., Ramogola-Masire, D., Richards-Kortum, R., et al. *PLoS One.* Vol. 7. Public Library of Science; 2012. High-Resolution Microendoscopy for the Detection of Cervical Neoplasia in Low-Resource Settings. Lo AWI, editor; p. e44924 Available from: <http://dx.plos.org/10.1371/journal.pone.0044924>
30. Quang T, Schwarz RA, Dawsey SM, Tan MC, Patel K, Yu X, et al. A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia. *Gastrointest Endosc.* 2016; 84:834–41. [PubMed: 27036635]
31. Pierce, MC., Schwarz, RA., Bhattar, VS., Mondrik, S., Williams, MD., Lee, JJ., et al. Accuracy of In Vivo Multimodal Optical Imaging for Detection of Oral Neoplasia; *Cancer Prev Res.* 2012. p. 801-9. Available from: <http://cancerpreventionresearch.aacrjournals.org/content/5/6/801.abstract>
32. Pierce MC, Guan Y, Quinn MK, Zhang X, Zhang W-H, Qiao Y-L, et al. A Pilot Study of Low-Cost, High-Resolution Microendoscopy as a Tool for Identifying Women with Cervical Precancer. *Cancer Prev Res.* 2012; 5
33. Grant, BD., José, Jü., Scapulatempo-neto, C., Matsushita, GM., Mauad, EC., et al. High-resolution microendoscopy: a point-of-care diagnostic for cervical dysplasia in low-resource settings; *Eur J Cancer Prev European Journal of Cancer Prevention.* 2017. p. 63-70. Available from: <https://insights.ovid.com/pubmed?pmid=26637074>

34. Quang, T., Tran, EQ., Schwarz, RA., Williams, MD., Vigneswaran, N., Gillenwater, AM., et al. *Cancer Prev Res (Phila)*. Vol. 10. American Association for Cancer Research; 2017. Prospective Evaluation of Multimodal Optical Imaging with Automated Image Analysis to Detect Oral Neoplasia In Vivo; p. 563-70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28765195>
35. Muldoon, TJ., Pierce, MC., Nida, DL., Williams, MD., Gillenwater, A., Richards-Kortum, R. Subcellular-resolution molecular imaging within living tissue by fiber microendoscopy; *Opt Express Optical Society of America*. 2007. p. 16413 Available from: <https://www.osapublishing.org/abstract.cfm?URI=oe-15-25-16413>
36. Simel, DL., Samsa, GP., Matchar, DB. Likelihood ratios with confidence: Sample size estimation for diagnostic test studies; *J Clin Epidemiol Pergamon*. 1991. p. 763-70. Available from: <https://www.sciencedirect.com/science/article/pii/089543569190128V?via%3Dihub>
37. Allen, CM., Fornatora, M., Jones, AC., Kerpel, S., Freedman, P. *Oral Surgery, Oral Med Oral Pathol Oral Radiol Endodontology*. Vol. 82. Mosby; 1996. Human papillomavirus-associated oral epithelial dysplasia (koilocytic dysplasia): An entity of unknown biologic potential; p. 47-56.
38. FDA VELscope. 2007. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf7/K070523.pdf
39. Vila, PM., Park, CW., Pierce, MC., Goldstein, GH., Levy, L., Gurudutt, VV., et al. *Ann Surg Oncol*. Vol. 19. Springer-Verlag; 2012. Discrimination of Benign and Neoplastic Mucosa with a High-Resolution Microendoscope (HRME) in Head and Neck Cancer; p. 3534-9. Available from: <http://www.springerlink.com/index/10.1245/s10434-012-2351-1>
40. Miles BA, Patsias A, Quang T, Polydorides AD, Richards-Kortum R, Sikora AG. Operative margin control with high-resolution optical microendoscopy for head and neck squamous cell carcinoma. *Laryngoscope*. 2015; 125:2308–16. [PubMed: 26059758]
41. Ishijima, A., Schwarz, RA., Shin, D., Mondrik, S., Vigneswaran, N., Gillenwater, AM., et al. *J Biomed Opt*. Vol. 20. International Society for Optics and Photonics; 2015. Automated frame selection process for high-resolution microendoscopy; p. 46014 Available from: <http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JBO.20.4.046014>

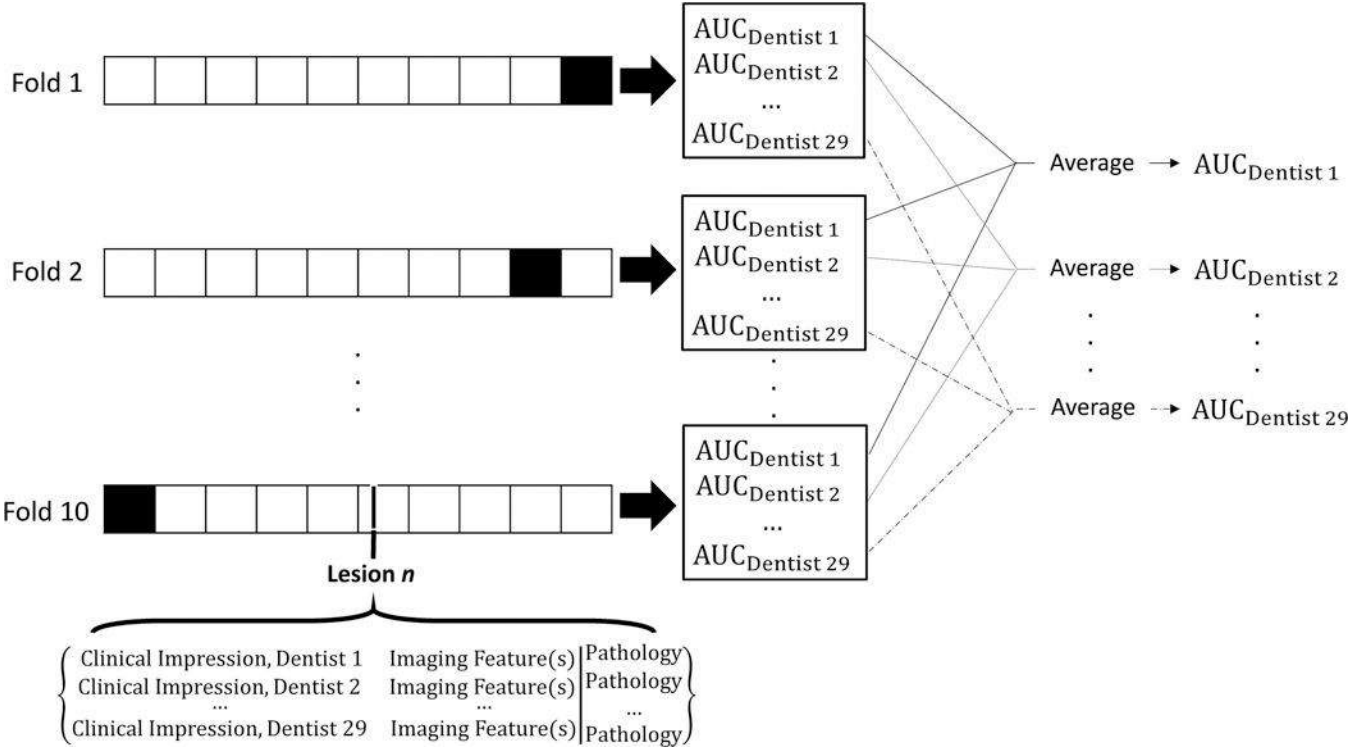


Figure 1. 10-fold cross-validation to estimate the AUC of clinical impression combined with imaging

For each fold, the training set (white squares) contained data from nine-tenths of the lesions, and the validation set (black squares) contained data from one-tenth of the lesions. Data from each lesion consisted of 29 clinical impressions, imaging feature(s), and histopathology. Training sets were used to train k -NN algorithms combining clinical impression with imaging feature(s), which were used to calculate the AUC of each dentist in the validation set. The AUC of each dentist was then averaged across all ten folds. This entire procedure was repeated 50 times, and the AUC of each dentist was averaged across the 50 runs.

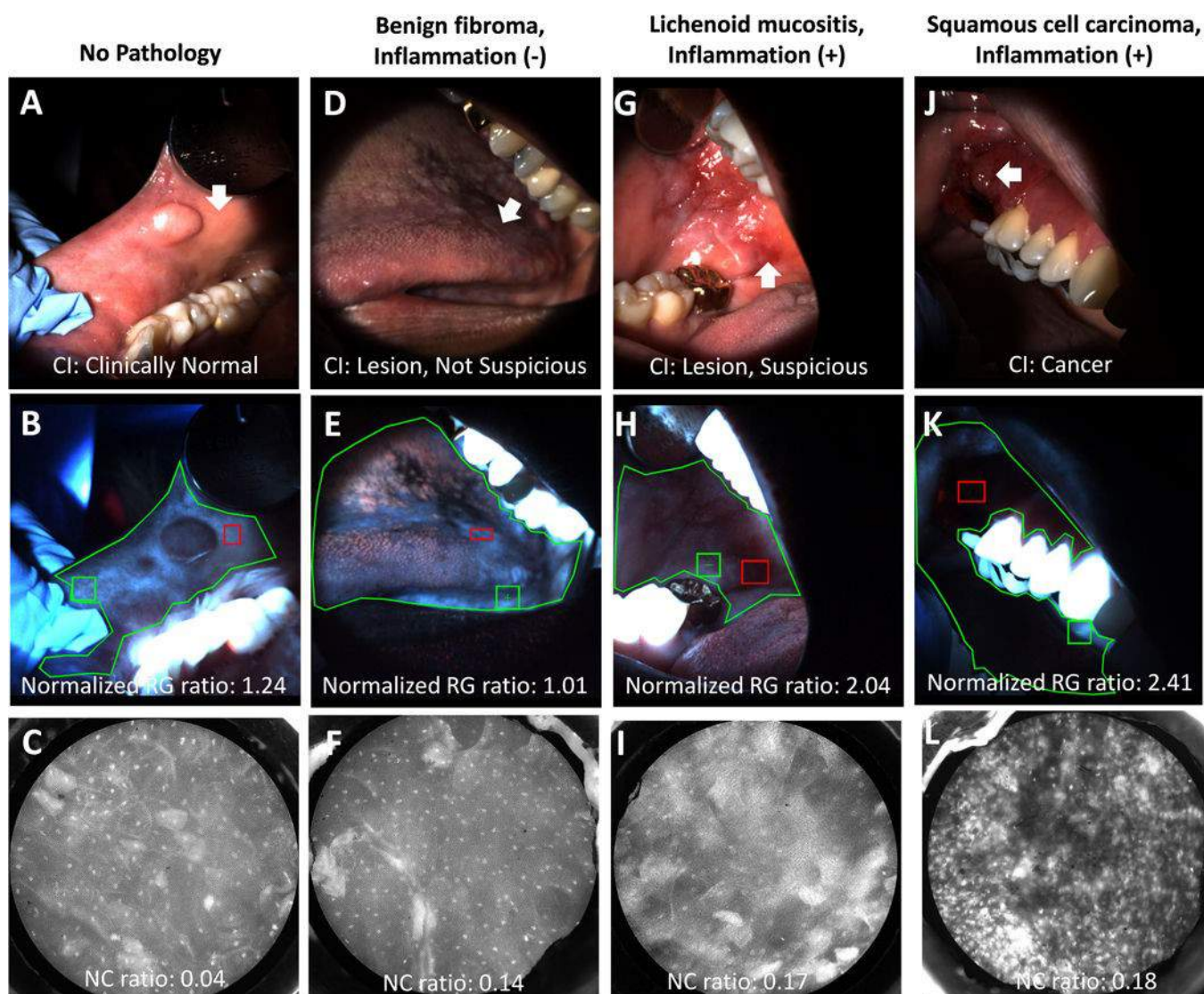


Figure 2. Multimodal images of analyzed sites

Top row: WL images showing sites (white arrows). **Middle row:** AFI images showing sites (red rectangles), normal sites (green squares), and outlined mucosa (green polygons).

Bottom row: HRME images showing nuclei of the superficial epithelium. **(A-C)** Images from a site rated “clinically normal” by the expert clinician that was negative by AFI and HRME. The site was not biopsied. **(D-F)** Images from a site rated “oral mucosal lesion, not suspicious” by the expert clinician that was negative by AFI and HRME. The histopathological diagnosis was a benign fibroma without severe inflammation. **(G-I)** Images from a site rated “oral mucosal lesion, suspicious” by the expert clinician that was positive by AFI but negative by HRME. The histopathological diagnosis was lichenoid mucositis with severe inflammation. **(J-L)** Images from a site rated “cancer” by the expert clinician that was positive by AFI and HRME. The histopathological diagnosis was moderately differentiated squamous cell carcinoma with severe inflammation.

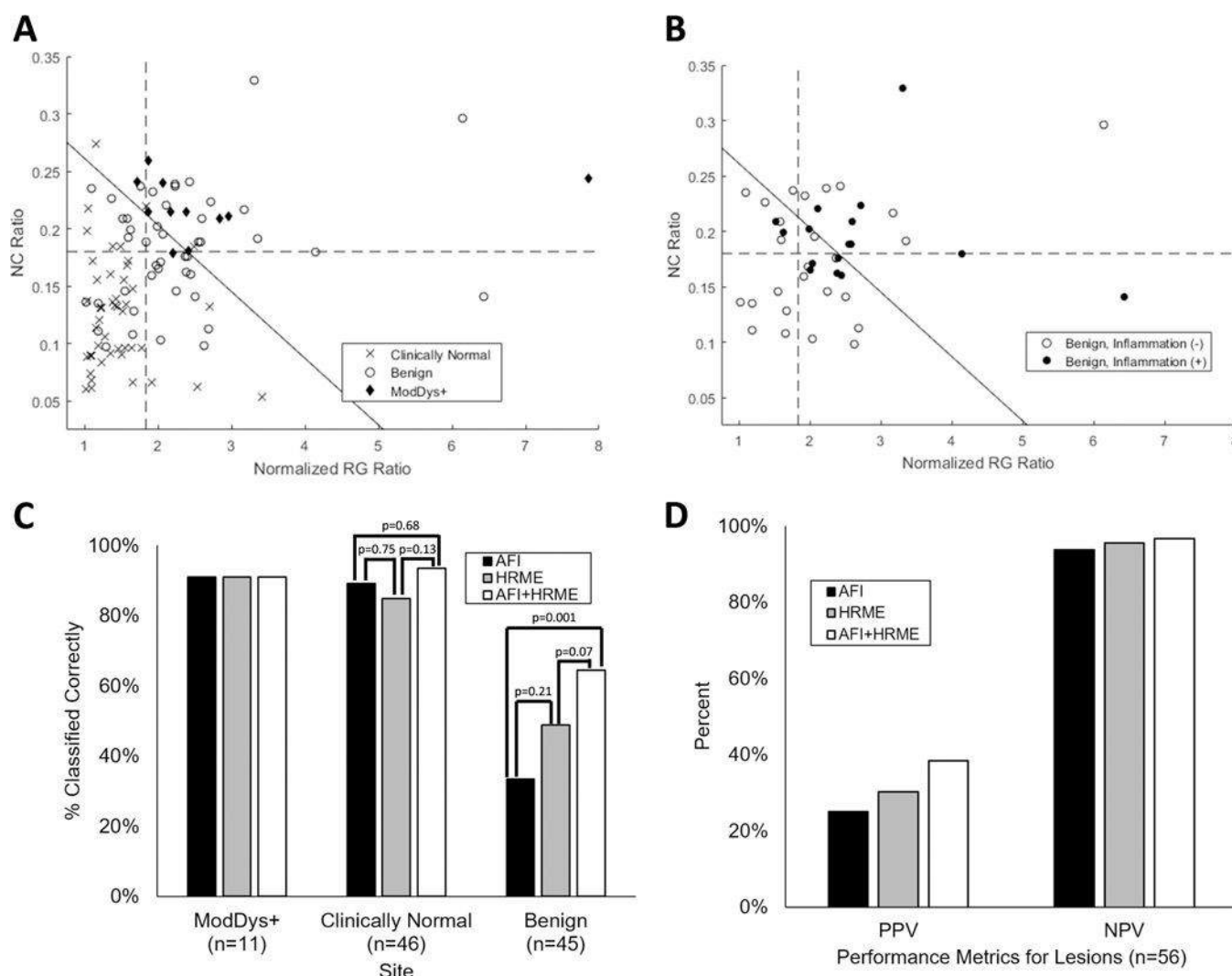


Figure 3. Objective diagnostic performance of multimodal imaging

(A) Scatterplot of NC ratio vs. normalized RG ratio of all analyzed sites stratified by pathology. An NC ratio threshold of 0.18 (horizontal dotted line) and normalized RG ratio threshold of 1.83 (vertical dotted line) correctly classified 91% of ModDys+ lesions. The multimodal imaging classifier combining both imaging methods (diagonal solid line) was also chosen to correctly classify 91% of ModDys+ lesions. (B) Scatterplot of NC ratio vs. normalized RG ratio of benign lesions (excluding mild dysplasia), stratified by inflammation status. Thresholds are identical to (A). (C) Percentage of sites correctly classified by AFI, HRME, and AFI+HRME using the thresholds in (A). (D) PPV and NPV of AFI, HRME, and AFI+HRME using the thresholds in (A) for the 56 lesions.

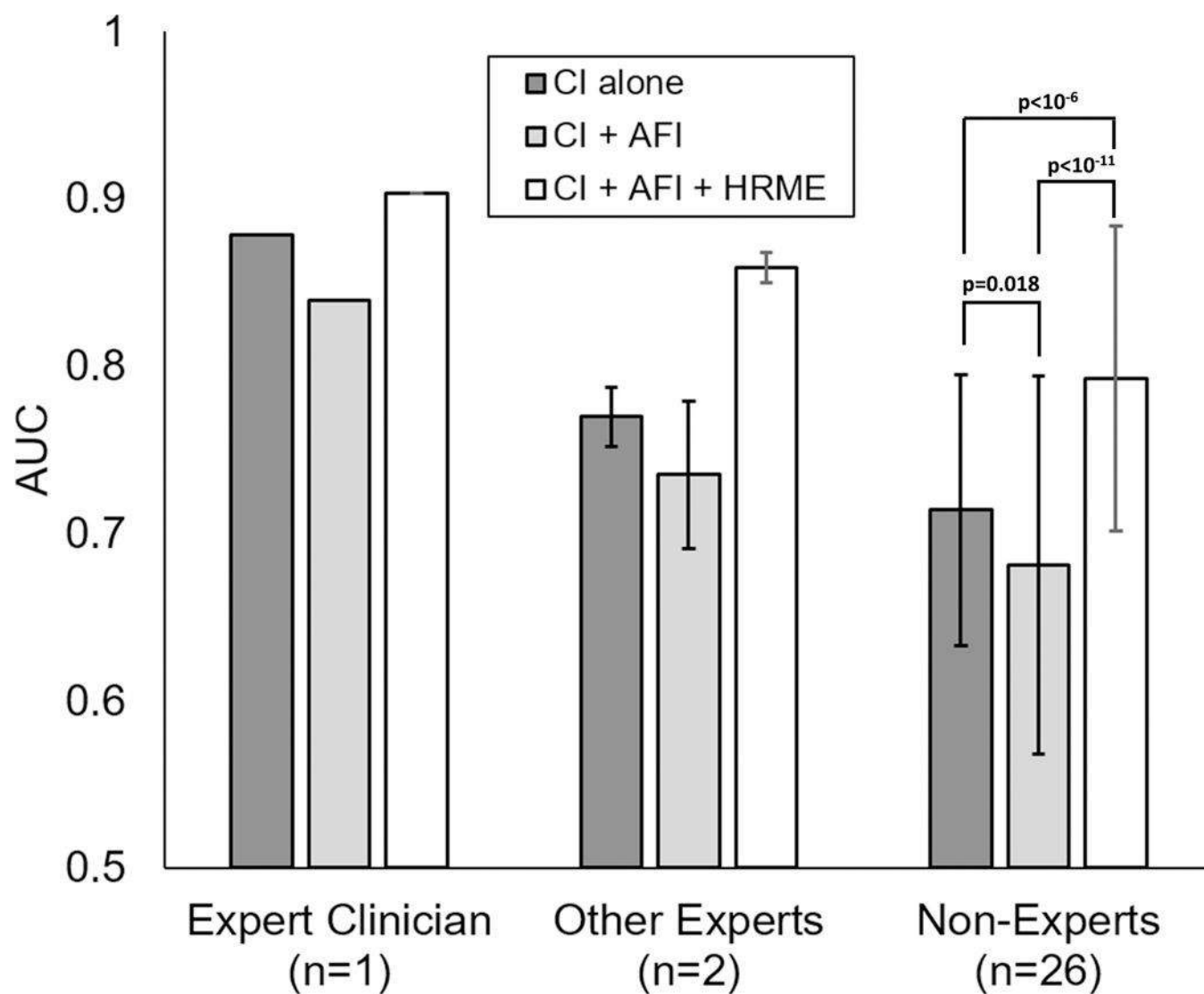


Fig. 4. Mean AUC of dentist clinical impression alone, clinical impression plus AFL, and clinical impression plus both imaging modalities

Error bars represent sample standard deviation. Statistical comparisons were performed with paired two-tailed t-tests.

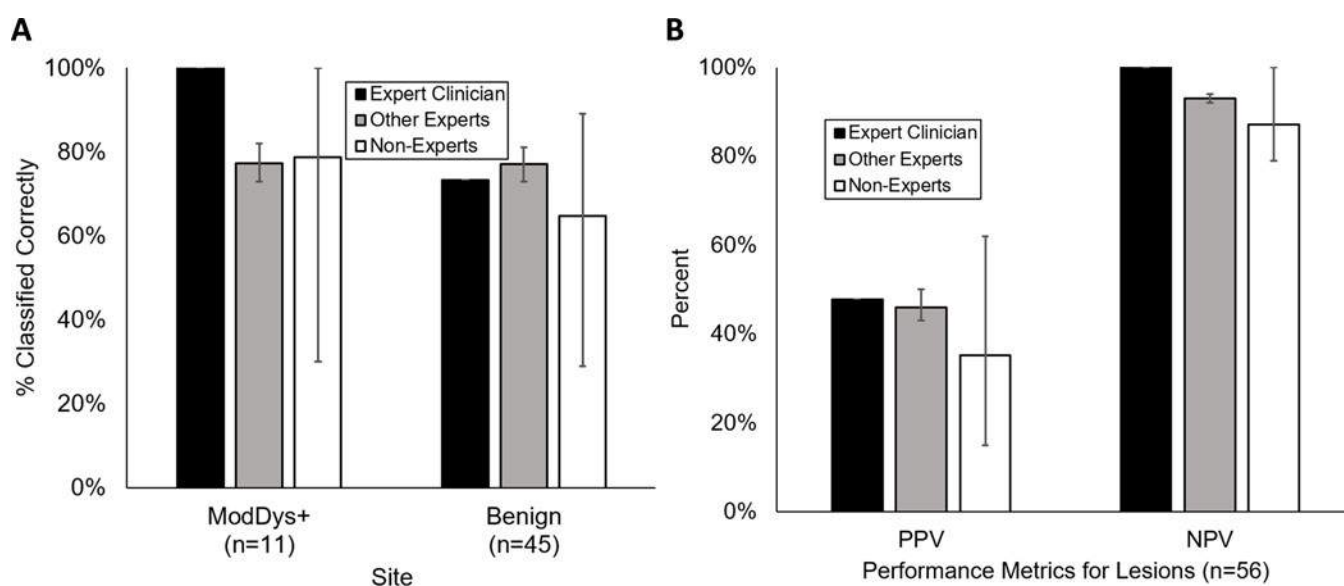


Figure 5. Diagnostic performance of dentist clinical impression

(A) Percentage of sites correctly classified by the expert clinician (n = 1), other experts (n = 2), and non-experts (n = 26). (B) Positive predictive value (PPV) and negative predictive value (NPV) of the expert clinician, other experts, and non-experts for the 56 lesions. Values represent aggregate performance of each group of dentists. Error bars represent range.

Table 1

Summary of Analyzed Sites.

Analyzed Sites	Classification	Pathology
Normal Oral Mucosa, No Biopsy (46)	Clinically Normal (46)	N/A
Biopsied Lesions (56)	Benign (45)	Lichenoid Mucositis (17)
		Fibroma (9)
		Granuloma (2)
		Pemphigus/Pemphigoid (2)
		Hyperkeratosis and/or Hyperplasia (4)
		Lymphoid Epithelial Cyst (1)
		Mucositis / Gingivitis (2)
		Papilloma (1)
		Pseudoepitheliomatous Hyperplasia (1)
		Pyogenic Granuloma (1)
		Submucous Fibrosis (1)
		Ulcer (1)
		Mild Dysplasia (3) ^a
	Moderate Dysplasia+ (11)	Moderate Dysplasia (7)
		Severe Dysplasia (1)
		Oral SCC (3)

^aMild dysplasia was considered benign due to its low risk of malignant progression and the difficulty of distinguishing mild dysplasia from reactive atypia in this population.