

FULL ARTICLE

Detection and delineation of squamous neoplasia with hyperspectral imaging in a mouse model of tongue carcinogenesis

Guolan Lu¹ | Dongsheng Wang² | Xulei Qin³ | Susan Muller⁴ | Xu Wang² | Amy Y. Chen⁴ |
Zhuo Georgia Chen² | Baowei Fei^{1,3,5,6*}

¹The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia

²Department of Hematology and Medical Oncology, Emory University, Atlanta, Georgia

³Department of Radiology and Imaging Sciences, Emory University, Atlanta, Georgia

⁴Department of Otolaryngology, Emory University School of Medicine, Atlanta, Georgia

⁵Department of Mathematics & Computer Science, Emory University, Atlanta, Georgia

⁶Winship Cancer Institute of Emory University, Atlanta, Georgia

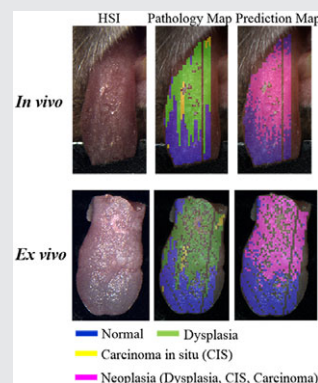
*Correspondence

Baowei Fei, Department of Radiology and Imaging Science, Emory University School of Medicine, Atlanta, Georgia.
Email: baowei.fei@emory.edu

Funding information

National Institutes of Health, Grant/Award numbers: CA176684, CA156775, CA204254; Developmental Funds from the Winship Cancer Institute of Emory University, Grant/Award number: P30CA138292

Hyperspectral imaging (HSI) holds the potential for the noninvasive detection of cancers. Oral cancers are often diagnosed at a late stage when treatment is less effective and the mortality and morbidity rates are high. Early detection of oral cancer is, therefore, crucial in order to improve the clinical outcomes. To investigate the potential of HSI as a noninvasive diagnostic tool, an animal study was designed to acquire hyperspectral images of *in vivo* and *ex vivo* mouse tongues from a chemically induced tongue carcinogenesis model. A variety of machine-learning algorithms, including discriminant analysis, ensemble learning, and support vector machines, were evaluated for tongue neoplasia detection using HSI and were validated by the reconstructed pathological gold-standard maps. The diagnostic performance of HSI, autofluorescence imaging, and fluorescence imaging were compared in this study. Color-coded prediction maps were generated to display the predicted location and distribution of premalignant and malignant lesions. This study suggests that hyperspectral imaging combined with machine-learning techniques can provide a noninvasive tool for the quantitative detection and delineation of squamous neoplasia.



KEYWORDS

4NQO-induced tongue carcinogenesis model, ensemble LDA, hyperspectral imaging, linear discriminant analysis, linear support vector machine, random forest, squamous neoplasia, supervised learning

1 | INTRODUCTION

Oral cancers are among the most common cancers globally, with an estimated 300 400 new cases and 145 400 deaths in 2012, and thus posing a significant health problem [1]. Although the oral cavity is easily accessible, lesions in the oral cavity often go unnoticed. In the United States, approximately 65% of patients diagnosed with oral cavity and pharynx cancer have regional spread or distant metastases

[2]. If these cancers are diagnosed at a localized stage, most of the patients can be cured with a 5-year survival rate of 83% [2]. Late-stage cancer identification can lead to poor survival with speech, swallowing and cosmetic problems. Moreover, patients who survive the initial occurrence of oral cancer have an increased risk of developing a second primary tumor [3]. Therefore, early detection of premalignant and malignant lesions remains the most promising approach for improving the clinical outcomes [4].

The conventional screening method for oral cancer begins with visual inspection and palpation of the tissue surface using incandescent light. Next, highly suspicious tissue regions are biopsied for pathological examination to make a definitive diagnosis [5]. However, visual examination is highly dependent on a clinician's training and experience. It is often a difficult task, even for an experienced physician, to identify the most malignant tissue region for biopsy due to the heterogeneity of the lesions. In addition, biopsy is invasive, costly and label-intensive. Pathology diagnosis is inconsistent due to intra- and inter-observer variations [6, 7]. Noninvasive alternatives are, therefore, desirable in order to improve the early diagnosis of oral cancer, and thus decreasing the morbidity and mortality of the disease.

Most of the lesions in the oral cavity are squamous-cell carcinoma, which originates from the oral mucosal lining of the oral cavity [8]. The easy access of the oral lesions makes it possible to detect superficial tumors with optical imaging which offers a variety of noninvasive tools utilizing intrinsic and extrinsic tissue contrast for the early detection of oral cancer [5]. The intrinsic approaches involve the development of optical devices to specifically probe the alterations of tissue absorption and scattering properties by reflectance imaging and to reveal the levels of endogenous chromophores, such as reduced nicotinamide adenine dinucleotide (NADH) and flavin adenine dinucleotide (FAD), and with autofluorescence accompanying malignant progression. The extrinsic methods rely on the exogenous molecular imaging agents which are topically applied on the tissue surface to enhance the detection of fluorescence changes. One study reported the use of a point spectroscopy device combining intrinsic fluorescence, diffuse reflectance and scattering signals to collect signals from multiple sites within the oral cavity. A 96% sensitivity and 96% specificity for the distinction of cancerous/dysplastic from normal tissue, and 65% sensitivity and 90% specificity for the discrimination of dysplastic from cancerous tissue were reported [9]. Although the results were promising, only a very limited number of measurements were made from selected sites of the oral cavity. In addition, it would be difficult to screen the entire oral cavity with small optic fibers, and this limits the clinical use of spectroscopy for oral cancer screening. In another study, multispectral imaging with reflectance wavelengths corresponding to hemoglobin spectral features was shown to maximize microvasculature visualization and the contrast of oral tissue [10]. However, only 2 wavelengths at 530 and 600 nm were explored, and without quantitative evaluations of their diagnostic performance.

Two, fluorescence vital-dyes have been previously tested for the detection of oral neoplasia [11]. The first dye is a fluorescent deoxyglucose molecule called 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose (2-NBDG, Cayman Chemical, Ann Arbor, Michigan), which can be used to assess the metabolic activity of cells. The

second fluorescence dye, proflavine, can nonspecifically stain cellular structures and enable observation of the nuclear morphology. The standard deviation of the proflavine fluorescence intensity and the mean fluorescence intensity of 2-NBDG have been reported to discriminate non-neoplastic regions of interest from neoplasia, including moderate dysplasia, severe dysplasia and carcinoma, with a sensitivity and specificity of 91% [11]. However, this method was shown to be only 42% sensitive when used to classify mild dysplasia.

Hyperspectral imaging (HSI) is an emerging optical technique that integrates wide-field imaging with spectroscopy to simultaneously acquire both spectral and spatial information, and this makes it possible to interrogate large tissue surfaces in a noncontact and noninvasive way. A wavelength-scanning hyperspectral imaging system generates a 3-dimensional (3-D) dataset called hypercube by spectrally splitting the light reflected from a sample surface with a dispersive device and then collecting the light using a 2-dimensional (2-D) detector array [12]. The alterations of the reflectance spectrum are associated with the tissue structural and biochemical changes. Each pixel in a hypercube has a spectral fingerprint that is associated with the structural and biochemical properties of tissue. To fully exploit the diagnostic potential of hyperspectral imaging, machine-learning techniques offer quantitative tools to mine the rich, spectral-spatial information in a hypercube. In our previous studies, we have demonstrated the feasibility of using HSI for the noninvasive detection of prostate cancer [13] and head and neck cancers (HNC) [14–18] in subcutaneous, xenograft tumor models.

To further evaluate the diagnostic potential of HSI, we designed a longitudinal study for both intrinsic extrinsic HSI of a 4-nitroquinoline-1-oxide (4NQO)-induced, oral carcinogenesis model. 4NQO, a synthetic water-soluble carcinogen, has been widely used in murine models to investigate all stages of oral carcinogenesis [19]. One advantage of this model is that 4NQO-induced lesions exhibit similar histological and molecular changes as in human oral carcinogenesis [19]. This animal model mimics human, oral neoplastic transformation with reproducible isolation of all stages, including dysplasia, carcinoma in situ (CIS), and squamous-cell carcinoma (SCC), and therefore providing an excellent opportunity to investigate the application of HSI for the noninvasive detection of squamous neoplasia.

The goal of our study is to develop supervised learning methods for the distinction of neoplasia (dysplasia, CIS and SCC) and non-neoplastic tongue tissue with hyperspectral imaging and to validate the diagnostic performance of HSI using the histopathology gold standard. Since HSI can provide images relying on intrinsic tissue contrast alone (reflectance, autofluorescence) or using exogenous contrast agents mapping the expression of biomarkers (fluorescence), we compared the diagnostic performance of label-free HSI with

autofluorescence and vital-dye fluorescence imaging of 2-NBDG and proflavine for the detection and delineation of squamous neoplasia.

2 | MATERIALS AND METHODS

2.1 | Instrumentation

We used a hyperspectral imaging system called CRI Maestro (PerkinElmer Inc., Waltham, Massachusetts). This instrument is comprised of a flexible fiber-optical light system with a Xenon light source, a solid-state liquid crystal tunable filter (LCTF) as the wavelength dispersion device, a spectrally optimized lens, and a 12-bit charge-coupled device (CCD) as the area detector [20, 21]. This system can be used to acquire both reflectance images and fluorescence.

2.2 | Mouse tongue carcinogenesis model

Thirty, female CBA/J mice were purchased from Jackson Laboratory (Bar Harbor, Maine) and were then separated into an experimental group ($N = 24$) and a control group ($N = 6$). Mice in the experimental group were treated with drinking water mixed with 4NQO powder (Sigma Aldrich, Saint Louis, Missouri) for 16 consecutive weeks (concentration: 100 $\mu\text{g/mL}$) in order to induce tongue carcinogenesis. Mice in the control group received normal drinking water without 4NQO. The mice in each group were monitored weekly for body weight and water consumption. The experiment was terminated at 24 weeks. All of the animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee (IACUC) of Emory University.

2.3 | Data acquisition

During the experiment, we randomly selected several mice from both the experimental group and the control group at weeks 8, 12, 20 and 24 for hyperspectral imaging using the following procedures. First, we acquired white and dark reference hypercubes before imaging the animals [16, 22–24]. Second, we anesthetized the selected mouse with ketamine after the reference hypercubes were acquired. To image the tongue in vivo, we placed the mouse in a supine position, gently pulled out the mouse tongue and taped it on a sterilized imaging stage. Next, we acquired reflectance hyperspectral images of the in vivo tongue from 450 to 950 with 5 nm intervals and then acquired autofluorescence images of the tongue with 455 nm excitation and a 490-nm long-pass emission filter. After in vivo imaging was complete, we euthanized the mouse by cervical dislocation and procured the tongue specimen for a series of ex vivo imaging, including HSI, autofluorescence imaging and fluorescence imaging of 2-NBDG (Cayman Chemical) and proflavine

(Sigma Aldrich). Both 2-NBDG and proflavine were topically applied on the resected tongues using the procedures described in Ref. [11]. Briefly, tongue specimens were first incubated in a 160 μM solution of 2-NBDG in $1\times$ PBS for 20 minutes at 37°C , and were then washed once with PBS. Fluorescence images of 2-NBDG were obtained using the blue excitation and 490 long-pass emission. Next, tongues were incubated in a 0.01% w/v solution of proflavine in $1\times$ PBS for 2 minutes at room temperature, and were then washed once with PBS. Fluorescence images of proflavine were obtained using blue excitation and 490 nm emission. The fluorescence signal from proflavine staining is much brighter than that of 2-NBDG, and thus allowing for imaging proflavine-stained tissue after 2-NBDG staining.

2.4 | Histology correlation

Immediately after ex vivo imaging of the dissected tongue, the ventral surface of the tongue was inked to allow the identification and correct orientation of the tongue during subsequent processing. The inked tongue specimens were placed in 10% buffered formalin overnight for fixation. Each fixed tongue was further processed through dehydration, clearing and wax infiltration. Next, the specimen was embedded in a cassette to form a paraffin block which was then clamped into a microtome for tissue sectioning. Tongue tissues procured at weeks 8 and 12 were sectioned sagittally into a series of 5- μm slices. Specimens procured at weeks 20 and 24 were first bisected longitudinally along the midline groove and then embedded in paraffin blocks which were sagittally sectioned into a series of 5- μm slices. The interval between the 2 tissue sections was 200 μm for week 12, and 100 μm for week 20 and 24. Immediately after sectioning, tissue slices were laid onto each glass for histological staining with hematoxylin and eosin (H&E), and the H&E stained slides were digitized for pathology diagnosis.

2.4.1 | Pathology diagnosis

An clinically experienced pathologist (SM) reviewed the H&E slides and graded each tongue slice according to the most severe pathology it contained. The dorsal surface of each tongue slice was further segmented into regions of normal (including healthy and hyperplastic tissue), dysplasia, CIS and SCC, as demonstrated in Figure 1. The grading and diagnosis of tongue cancer is based on the presence of architectural and cytological changes of the epithelial layer based on microscopic examination of H&E stained sections [25].

2.4.2 | Generation of pathology gold standard maps

It was a very challenging task to map the pathology diagnosis onto the dorsal surface of the tongue as the gold standard in order to validate the cancer prediction results as hyperspectral images and fluorescence images were the projections of the dorsal tongue surface, while the histology slides

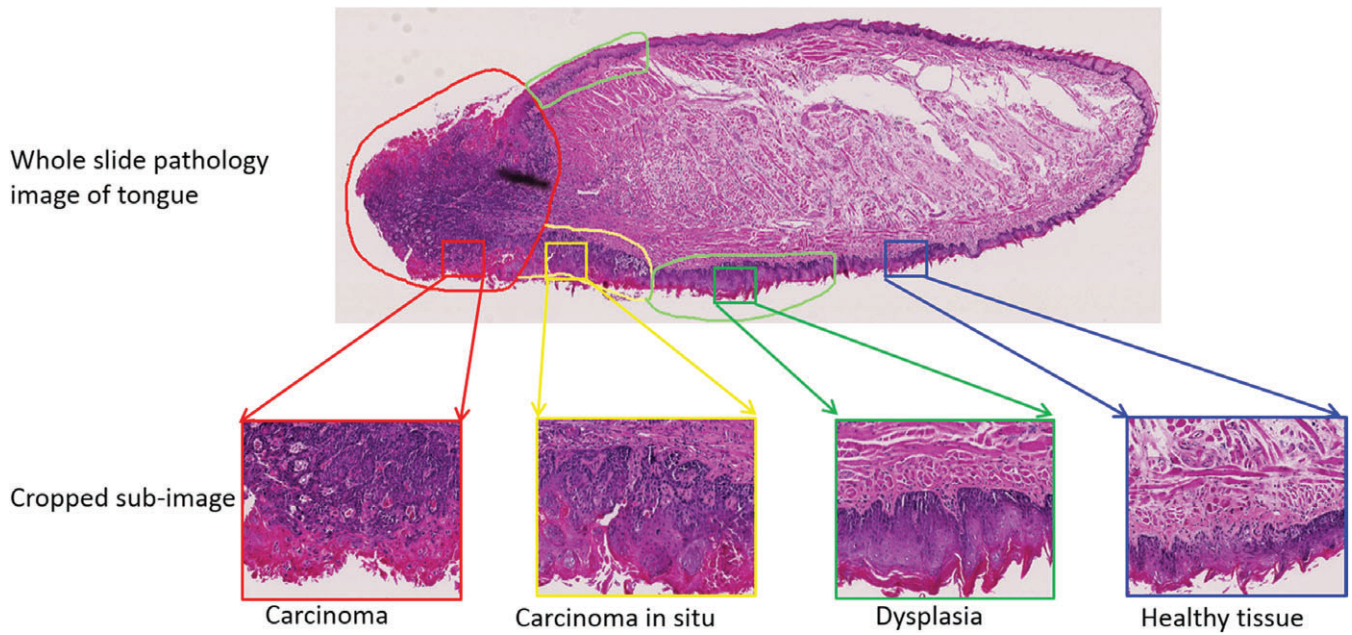


FIGURE 1 An example of pathology grading and annotation in a whole slide image of tongue specimen

were the cross-sections of the tongue with both the dorsal and ventral surfaces. To reconstruct the pathology map, we tracked each step of the histological processing so as to ensure correct tongue orientation during embedding, numbering of each tissue slice, and the specific layout of tissue slices on each glass slide during sectioning. As shown in Figure 2, we sectioned each tongue specimen into sagittal slices of 5 μm with 100 μm or 200 μm intervals between sections. Since each H&E slice corresponded to one straight line along the midline of the dorsal tongue surface, we measured the width of each pathological region along the dorsal surface and then mapped these pathology readings back proportionally onto the corresponding regions of the tongue images. The last image in Figure 2 shows an example of the 2D, color-coded pathology map overlaid with an *ex vivo* tongue image. Some regions were labeled as black lines on the pathology map and thus indicating the missing gold standard.

2.5 | Classification of hyperspectral image

2.5.1 | Pre-processing of hyperspectral image

The pre-processing of hyperspectral images involves 2 steps. First, we converted the raw data into a percent reflectance value in a pixel-wise manner, as described in [16, 22–24]. The purpose of this step was to eliminate the illumination nonuniformity and the influence of the dark current. Next we identified and removed glare pixels from the normalized hyperspectral images. Glare regions were formed due to specular reflection from the moist tongue surfaces which did not contain useful diagnostic information. We observed that as glare pixels were characterized with very bright reflectance intensity, they fell into the long tail region in the intensity histogram of the sum image for all the spectral

bands. To identify a threshold to detect these pixels, we developed an adaptive thresholding method which fitted the histogram of the sum image of all the spectral bands with a log-logistic distribution and then used the method described in [24] to compute the intensity threshold for glare pixel detection and removal.

2.5.2 | Feature extraction

Reflectance hyperspectral imaging.

For reflectance hyperspectral images, we evaluated and compared the pixel-wise and block-based feature extraction (FE) methods: (1) the pixel-wise spectral method extracted the normalized reflectance intensity of each pixel on the gold standard lines of the pathology map, and each pixel was labeled with a pathology type, and (2) A block-based spectral method first gridded each tongue hypercube ($M \times N \times K$; M , image height; N , image width; $K = 101$, number of wavelengths) into many blocks with a size $m \times n \times K$ ($m = 5$, $n = 5$, $K = 101$) and then computed the mean spectra of all the pixels from each block as the spectral feature. The physical size of each pixel is 26 μm . Each block is labeled with the most frequent pathology type in this block.

Autofluorescence imaging

For autofluorescence imaging, we extracted the average fluorescence intensity from 500 to 720 nm with 5 nm increments within each block for classification. We also extracted the red-to-green (red: 650 nm, green: 510 nm) fluorescence for comparison, which was reported to be effective in distinguishing neoplastic and normal areas of the oral cavity [26].

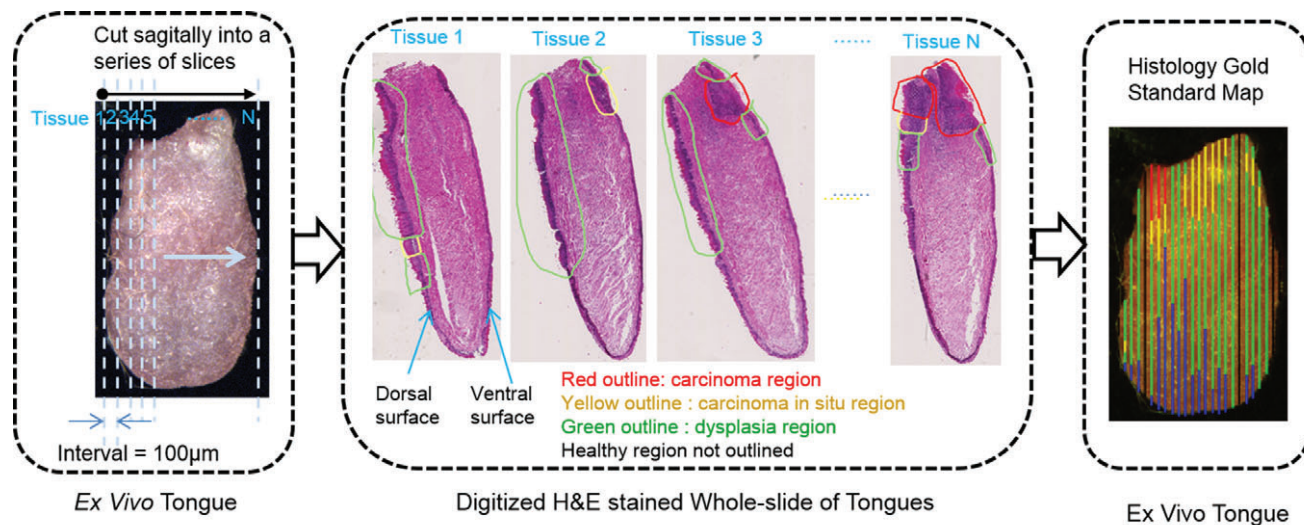


FIGURE 2 Schematic diagram of the pathological gold-standard map generation

Vital-dye fluorescence imaging

For 2-NBDG and proflavine imaging, we extracted the average fluorescence intensity of 2-NBDG or proflavine from 500 to 720 nm with a 5-nm increment within each block for classification. Moreover, the mean fluorescence intensity at the maximum emission band (540 nm) of 2-NBDG and the standard deviation of fluorescence intensity at the peak emission (515 nm) of proflavine were extracted as features for comparison. These 2 features were shown to discriminate between neoplastic and non-neoplastic regions of interest with 91% sensitivity and specificity [11].

2.5.3 | Supervised classification

Supervised classification can be used to ascertain the underlying patterns of training data and build predictive models to differentiate between neoplastic and non-neoplastic tissue from hyperspectral images of mouse tongues. The neoplastic class included lesions histopathologically diagnosed as dysplasia, CIS and SCC. To build a robust prediction model with low bias and low variances, we evaluated 7 classifiers, that is, linear and quadratic discriminant analysis classifiers (LDA and QDA), ensemble learning methods including ensemble LDA, random forests, RUSBoost and support vector machines (SVMs) with linear and radial basis kernel (RBF). These classifiers were implemented in Matlab R2015b. Their characteristics were briefly summarized as follows:

Discriminant analysis (LDA and QDA)

Discriminant analysis intends to find the projections of the original high dimensional space to a lower dimensional space in order to maximize the class separation [27]. Both LDA and QDA classification models assume that the data has a multivariate normal distribution [27, 28]. The difference between them is that the LDA model assumes the same covariance matrix for each class with varying means,

while QDA assumes an individual mean and covariance matrix for each class [28, 29]. LDA is advantageous in several aspects. First, it does not require tuning of free parameters. Secondly, it has been shown to perform well in classification tasks even when the assumption of the common covariance matrix among groups and normality is violated [30]. Finally, the decision boundary obtained by LDA is equivalent to the binary SVM on the set of support vectors [31].

Ensemble LDA

Ensemble learning methods build a high-quality ensemble predictor by combining results from many weak learners, such as decision trees and discriminant learners. Ensemble LDA improves the accuracy of the LDA with random subspace ensembles [32] which involve training multiple LDA learners with randomly selected subsets of features without replacing and classifying the testing data by taking the average of the scores predicted by the weak learners. Two parameters in this model need to be tuned for optimal classification performance, i.e. the number of feature dimensions to sample in each learner and the number of learners in the ensemble.

RUS boost

RUS Boost is an ensemble learning method that is especially effective at classifying an imbalanced dataset [33]. Random Under Sampling (RUS) is a commonly used data sampling method that randomly removes examples from the majority class in order to adjust the class distribution of the training data set. AdaBoost is a popular boosting technique that has been shown to improve the classification performance of weak learners. RUSBoost combines RUS with AdaBoost techniques in order to improve the classification performance for a skewed dataset. This method first generates a balanced training dataset for each weak learner in the

ensemble, by taking N , the number of instances in the minority class as the basic unit for sampling and selecting a subset of the majority classes with N instances, and then following the boosting procedure to iteratively create an ensemble of weak learners and make predictions according to the weighted vote of individual weak learners. Here, a decision tree is chosen as the weak learner. We control the depth of a decision tree by adjusting the number of observations per leaf node and the maximal number of branch node splits per tree. We also select the optimal number of tree learners in the ensemble by cross validation.

Random forest

A random forest is a classification algorithm that uses an ensemble of decision trees [34]. Random forest combines bagging and random feature selection to yield an ensemble that can achieve both low bias and low variance. Each tree is grown on a training set randomly drawn with replacement from the original dataset. In each bootstrap training set, approximately one-third of the instances are left out for out-of-bag estimates. At each node of the tree, a random subset of features was selected from the input features to split on. The final classification decision is made by computing the mean class probabilities from all the trees. To generate the forest trees, 2 parameters need to be optimized, including the number of trees to be produced and the number of features to be selected. Random forest classifier has been widely used in the remote sensing field due to its classification accuracy [35]. Our previous work also demonstrated the successful use of random forests for tongue cancer detection [36].

SVM

SVM is a binary classifier that intends to find the optimal separating hyperplane which maximizes the distance between the closest training sample and the hyperplane. The simplest form of SVM utilizes a linear decision function to separate linearly separable data. Nonlinear SVM generalizes the linear SVM by nonlinearly mapping the input data through kernel transformation into a higher dimensional feature space where they are linearly separable. SVM has been widely used in the remote sensing field for hyperspectral image classification due to its high classification accuracy, very good generalization capability, and effectiveness within the presence of heterogeneous classes even with few training samples available [37]. In this study, we chose the LIBSVM software package [38] for both linear and kernel SVMs.

Performance evaluation

We assessed the performance of the classifiers with a variety of metrics, including receiver operating characteristic (ROC) curves, the areas under the ROC curve (AUC), accuracy, sensitivity and specificity. The ROC curve offers a

graphic interpretation of the trade-off between sensitivity and specificity for a range of possible cut-off points. AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. At the optimal operating point of an ROC curve, we can obtain the accuracy, sensitivity and specificity from the confusion matrix as defined in Refs. [15, 16].

3 | RESULTS

3.1 | Mouse tongue carcinogenesis model

Over the course of the neoplastic progression, a number of mice from 4NQO-treated group and the control group were selected for hyperspectral imaging and were euthanized after imaging at various periods of observation. Mice in the control group were healthy throughout the experiment, while mice exposed to 4NQO with observations up to 24 weeks developed premalignant and malignant tongue lesions, as shown in Table 1. At 8 weeks after the 4NQO treatment, 6 of the 6 mice developed epithelial dysplasia of the tongue. CIS was induced in one of 6 mice at 12 weeks, in 2 of 5 mice at 20 weeks, and in one of 3 mice at 24 weeks. SCC was detected in 3 of 5 at 20 weeks and in 2 of 3 mice at 24 weeks. Four of 24 mice from the 4NQO-treated group died during the experiment and were not included in Table 1. 4NQO treatment for up to 16 weeks induced heterogeneous lesions in the mouse tongue as well as the esophagus.

3.2 | Characteristics of hypercube

Each hypercube consists of a series of 2-D, grayscale spectral images at wavelengths ranging from 450 to 950 nm. Figure 3A shows some representative spectral bands from a reflectance hypercube of an ex vivo tongue. It can be very difficult to visually identify lesions on tongue surfaces from individual grayscale images. Different spectral bands obtained optical information at varying tissue depths. The image intensities varied across the wavelength region, and thus reflecting the variations of tissue absorption and scattering properties at different wavelengths. Figure 3B shows a representative average spectral curve of neoplastic and non-neoplastic tissue. The dominant chromophores in the VIS region are hemoglobin which has been used to

TABLE 1 Histological diagnosis of all the mice in the experimental group

Week	Number of mice			Total
	Dysplasia	CIS	SCC	
8	6			6
12	5	1		6
20		2	3	5
24		1	2	3

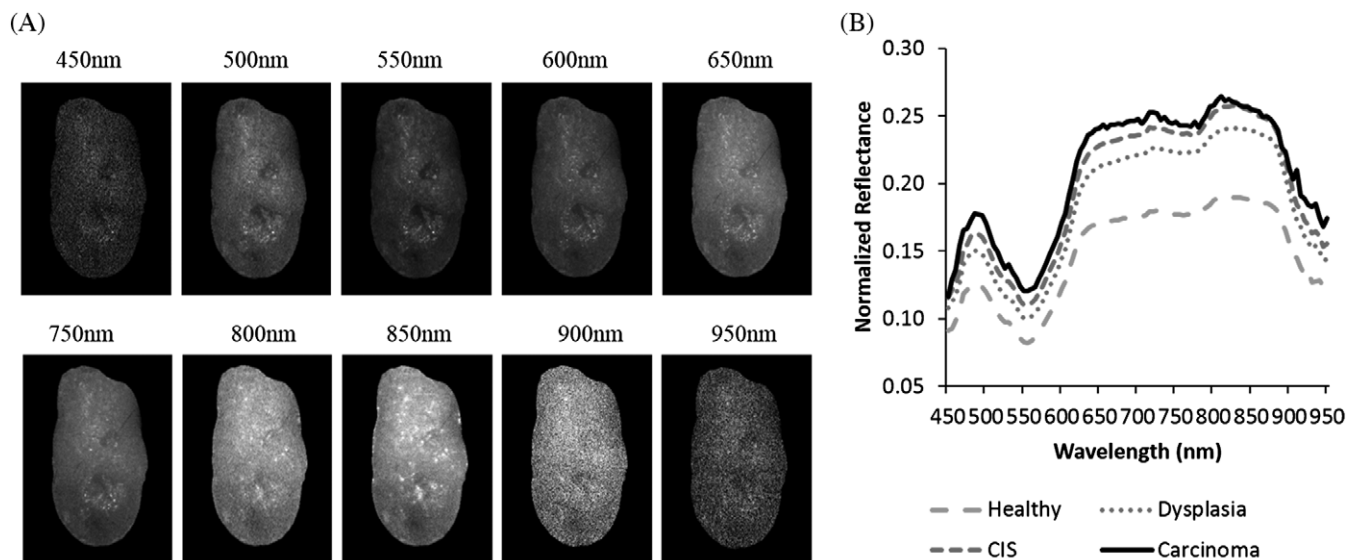


FIGURE 3 Characteristics of the reflectance hypercube for an ex vivo tongue. (A) Representative spectral bands at multiple wavelengths. (B) Average spectra of healthy, dysplasia, CIS, and carcinoma tissue

visualize increased vasculature in the oral cavity during malignant progression [39, 40]. It seems that the most prominent difference between different pathologies are in the reflectance intensity. However, there are many intra- and inter-mouse variances in the spectral curves and these factors further complicate the detection of neoplasia.

3.3 | Predictive analysis

3.3.1 | Dataset

Imaging data of 10 mice fulfilling the following criteria were selected for quantitative analysis: (1) the mouse has detailed gold standard of the epithelium for the dorsal tongue outlined by the pathologist, (2) the mouse has all 6 types of images available, including in vivo reflectance hyperspectral images, autofluorescence images as well as ex vivo reflectance hyperspectral images, autofluorescence, 2-NBDG and proflavine fluorescence images, (3) satisfactory resection without severe damage to the tongue regions, and (4) good imaging quality without significant motion artifacts. Images acquired at 8 weeks were excluded from analysis due to severe motion artifacts and tissue folds in pathology slices. All of the selected mice developed heterogeneous lesions on different sites of their tongues following exposure to 4NQO for up to 16 weeks.

3.3.2 | Comparison of feature extraction methods

To begin with, we compared the predictive performance of pixel-wise feature extraction and block-based feature extraction with all of the spectral bands of ex vivo tongues. As shown in Figure 4, the block-based method was found to be more accurate than the pixel-wise method in terms of all the performance metrics and classifiers ($P < .05$). In addition, as the block size was chosen as 5×5 , the number of samples for training and testing using the block-based method

was approximately 4 times less than the number of samples using the pixel-based method and thus significantly reducing the computation time. Therefore, the results of block-based method are reported in the following sections.

3.3.3 | Optimal classifier selection

For cancer detection in ex vivo tongues, we selected a classifier with minimal generalization errors by comparing the predictive performance of 7, different classifiers. Since nested cross validation (CV) has been proven to give an almost unbiased estimate of the true error on independent (unseen) test dataset, we conducted nested CV with k -fold internal CV on the training dataset in order to tune model parameters and we used leave-one-out external CV to estimate the generalization error [41–43]. For each fold of the external CV, we built a training model on the data from 9 out of 10 mice and tested the model on the remaining mouse data. On each of the training datasets, we performed an additional k -fold CV to search for the optimal model parameters. We tuned the optimal parameters of the individual classifiers as follows:

1. *Ensemble LDA*: We determined 2 parameters in a sequential search on the training dataset. First, the optimal number of features to sample in each weak learner was searched with 5-fold CV in the range of 1:10:101 (the spectral dimension of each hypercube is 101). Next, the optimal number of weak learners used in the ensemble was searched from 1 to 500 with 5-fold CV;
2. *RUSBoost*: We first searched for the minimum number of observations per leaf node $N_{\text{leaf}} \in 10^{[0.50, 0.67, 0.83, 1.00, 1.17, 1.33, 1.50, 1.67, 1.83, 2.00]}$ with 5-fold CV. Next we did a grid search to identify the maximum number of branch node splits per tree ($2^{\lceil \log_2(N_{\text{tr}} - 1) \rceil}$, N_{tr} is the number of training samples) and

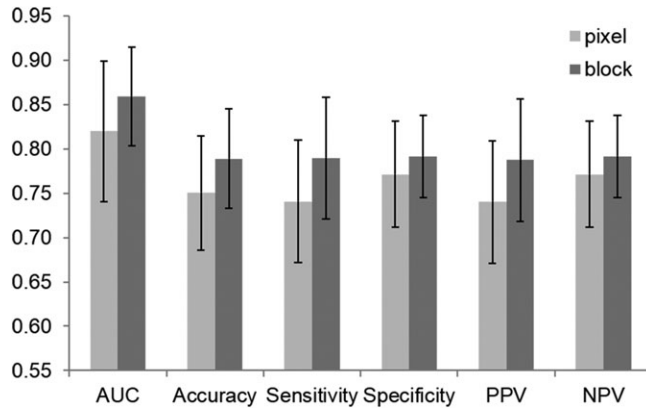


FIGURE 4 classification performance of pixel-wise and block-wise feature extraction methods

the optimal number of tree learners (1:1200) in the ensemble by 5-fold CV;

3. *Random forest*: We tuned 2 parameters to grow the optimal classification trees. The optimal number of decision trees were selected from 1 to 2000 that produced the lowest out-of-bag error. The number of features to be selected was fixed to be the square root of the total number of input features [35];
4. *Linear SVMs*: We conducted a 3-fold CV to search for the optimal cost parameter $C \in 2^{[-10, -9, \dots, 0, \dots, 9, 10]}$;
5. *SVMs with Gaussian RBF kernel*: We performed a grid search with 3-fold CV to determine the optimal combination of the cost parameter C and kernel parameter γ over the range of $C \in 2^{[-5, -4, \dots, 0, \dots, 4, 5]}$ and $\gamma \in 2^{[-5, -4, \dots, 0, \dots, 4, 5]}$.

As shown in Table 2, among all the classifiers, linear SVM, ensemble LDA and LDA performed the best with an AUC value of 0.86, while SVM with RBF kernel only yielded an AUC value of 0.80. Ensemble LDA slightly enhanced the performance of LDA, which was better than QDA and random forest. RUSBoost had the highest sensitivity, although the variance was very high. In the remaining results, we only reported classification with linear SVM.

TABLE 2 Predictive performance of different classification models for ex vivo tongue cancer detection

Classifier	AUC	Accuracy	Sensitivity	Specificity
Linear SVM	0.86 ± 0.06	$79\% \pm 6\%$	$79\% \pm 7\%$	$79\% \pm 5\%$
Ensemble LDA	0.86 ± 0.06	$79\% \pm 6\%$	$78\% \pm 7\%$	$79\% \pm 5\%$
LDA	0.86 ± 0.06	$78\% \pm 6\%$	$78\% \pm 8\%$	$80\% \pm 4\%$
Random forest	0.84 ± 0.08	$77\% \pm 7\%$	$77\% \pm 7\%$	$77\% \pm 8\%$
QDA	0.82 ± 0.05	$76\% \pm 5\%$	$76\% \pm 6\%$	$75\% \pm 4\%$
RBF SVM	0.80 ± 0.07	$75\% \pm 7\%$	$77\% \pm 9\%$	$75\% \pm 8\%$
RUSBoost	0.79 ± 0.12	$71\% \pm 15\%$	$80\% \pm 6\%$	$69\% \pm 20\%$

3.4 | Diagnostic performance of reflectance, autofluorescence and fluorescence imaging

In this study, we compared the diagnostic performance of reflectance HSI and autofluorescence imaging of both in vivo and ex vivo tongues, as well as 2-NBDG fluorescence and proflavine fluorescence imaging of the ex vivo tongues. As shown in Table 3, among all the imaging methods, 2-NBDG fluorescence imaging yielded the best performance, with an average AUC, sensitivity and specificity of 0.91, 85% and 84%, respectively, for ex vivo tongue neoplasia detection. Label-free HSI without any contrast agent achieved an average AUC, sensitivity and specificity of 0.86, 79% and 79%. Although the AUC values of autofluorescence and proflavine fluorescence imaging were slightly better than that of the reflectance method, the differences were not statistically significant ($P < .05$). As for neoplasia detection for in vivo tongues, label-free HSI based on reflectance produced similar performance compared with that of autofluorescence imaging. Features extracted from one or 2 spectral band, including the red-to-green ratio of autofluorescence intensity, the mean intensity of 2-NBDG fluorescence at 540 nm, and the standard deviation of proflavine fluorescence at 515 nm, did not provide better prediction performance than multiband autofluorescence or fluorescence features.

Furthermore, we examined how many samples of dysplasia, CIS and SCC were correctly classified in the

TABLE 3 Diagnostic performance of hyperspectral reflectance imaging, autofluorescence imaging, 2-NBDG fluorescence imaging and proflavine fluorescence imaging for the distinction of non-neoplastic and neoplastic tissue of ex vivo tongue and the diagnostic performance of hyperspectral reflectance imaging and autofluorescence imaging of in vivo tongue

	Imaging method	AUC	Accuracy	Sensitivity	Specificity
In vivo	HSI	0.84 ± 0.05	$78\% \pm 5\%$	$78\% \pm 5\%$	$78\% \pm 5\%$
	Autofluorescence	0.84 ± 0.06	$78\% \pm 5\%$	$78\% \pm 4\%$	$77\% \pm 9\%$
	Autofluorescence(R/G)	0.43 ± 0.10	$48\% \pm 6\%$	$48\% \pm 6\%$	$47\% \pm 12\%$
Ex vivo	2-NBDG	0.91 ± 0.04	$84\% \pm 3\%$	$85\% \pm 3\%$	$84\% \pm 6\%$
	Proflavine	0.89 ± 0.03	$82\% \pm 3\%$	$83\% \pm 3\%$	$81\% \pm 5\%$
	HSI	0.86 ± 0.06	$79\% \pm 6\%$	$79\% \pm 7\%$	$79\% \pm 5\%$
	Autofluorescence	0.87 ± 0.03	$80\% \pm 4\%$	$80\% \pm 6\%$	$81\% \pm 3\%$
	Autofluorescence(R/G)	0.49 ± 0.19	$50\% \pm 14\%$	$50\% \pm 17\%$	$54\% \pm 12\%$
	2-NBDG (mean)	0.75 ± 0.23	$70\% \pm 17\%$	$70\% \pm 18\%$	$71\% \pm 19\%$
	Proflavine (Std)	0.49 ± 0.15	$54\% \pm 9\%$	$56\% \pm 12\%$	$49\% \pm 20\%$

TABLE 4 Classification accuracy of different imaging methods for the distinction of different pathologies

	Imaging method	Normal	Dysplasia	CIS	SCC
In vivo	HSI	78%	75%	89%	81%
	Autofluorescence	75%	76%	83%	91%
Ex vivo	2-NBDG	84%	84%	91%	100%
	HSI	79%	79%	85%	71%
	Proflavine	81%	82%	87%	98%
	Autofluorescence	81%	78%	88%	83%

neoplastic group, as shown in Table 4. With 2-NBDG fluorescence imaging, 84% of normal tissue, 84% of dysplasia, 91% CIS and 100.0% SCC were correctly classified. Ex vivo, label-free HSI was able to classify 79% of normal tissue, 79% of dysplasia, 85% of CIS and 71% of SCC, while in vivo reflectance HSI was able to classify 78% of normal tissue, 75% of dysplasia, 86% of CIS and 81% of SCC.

Figure 5 shows that the diagnostic performance of HSI varied among different mouse tongues with varying ROC curves. Figure 6 shows a representative example of tongue cancer detection with in vivo and ex vivo imaging. As shown in the first and third column, label-free HSI can accurately detect and delineate the premalignant and malignant lesions with an AUC of 0.88 and 0.92 across the heterogeneous in vivo and ex vivo tongues, respectively. The prediction color map matched well with the gold standard color map. The diagnostic performance of label-free HSI was found to be better than that of autofluorescence imaging and comparable to that of fluorescence imaging of proflavine and 2-NBDG stains in terms of the AUC value for the example tongue.

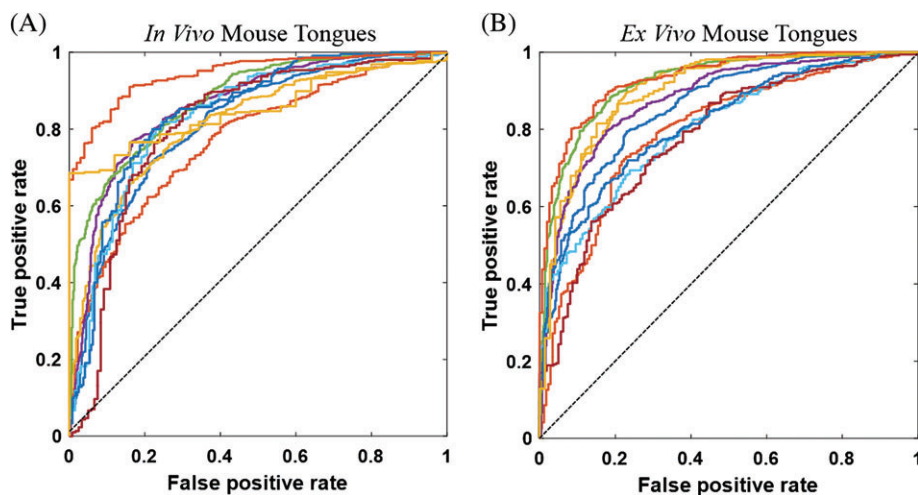
4 | DISCUSSION

Early detection of oral lesions could improve the clinical outcomes because treatment is most effective at an early stage. Furthermore, identification of local recurrences and second primary tumors is important for surveillance of

patients who have survived their initial oral cancer. HSI is a noninvasive optical imaging modality that holds the potential to improve oral cancer diagnosis and reduce cancer-related mortality and morbidity. In this proof-of-principle study, we demonstrated the diagnostic capability of HSI for the detection and delineation of premalignant and malignant lesions in a chemically induced carcinogenesis murine model which was validated by the reconstructed pathological gold-standard maps for hyperspectral, autofluorescence and fluorescence images of ex vivo and in vivo tongues. We found that the prediction results of ex vivo imaging were slightly better than the results of in vivo imaging. This might be due to the fact that it was much more difficult to accurately align the ex vivo histology gold standard with in vivo tongue images due to the stretching and motion of the tongue during in vivo imaging, in conjunction with the tissue deformations during the histological preparation phase.

In terms of feature extraction, we utilized the full reflectance spectrum from 450 to 950 nm in order to capture all of the available spectral-spatial information provided by HSI. We found that spatial averaging of spectral bands over a small block was more accurate than using the spectrum from an individual pixel. This may be attributed to 2 reasons: (1) spatial averaging of the spectra made the spectral features more robust to noises and registration errors between the pathology and hyperspectral images, and (2) spatial averaging reduced the spectral redundancy in neighboring pixels due to the cross-talks inherent in the wide-field imaging approach. Feature selection can be conducted in our future work to rank the diagnostic value of individual wavelengths. In addition, development of spectral-spatial feature extraction may further improve the diagnostic performance of label-free HSI.

We evaluated and compared a variety of machine-learning classifiers in this animal model in order to separate neoplasia from non-neoplastic tongue tissue. Linear SVM performed the best among all of the 7 classifiers included in the study. The classification performance of LDA, ensemble

**FIGURE 5** ROC curve of the diagnostic performance for HSI in both in vivo (A) and ex vivo (B) tongues

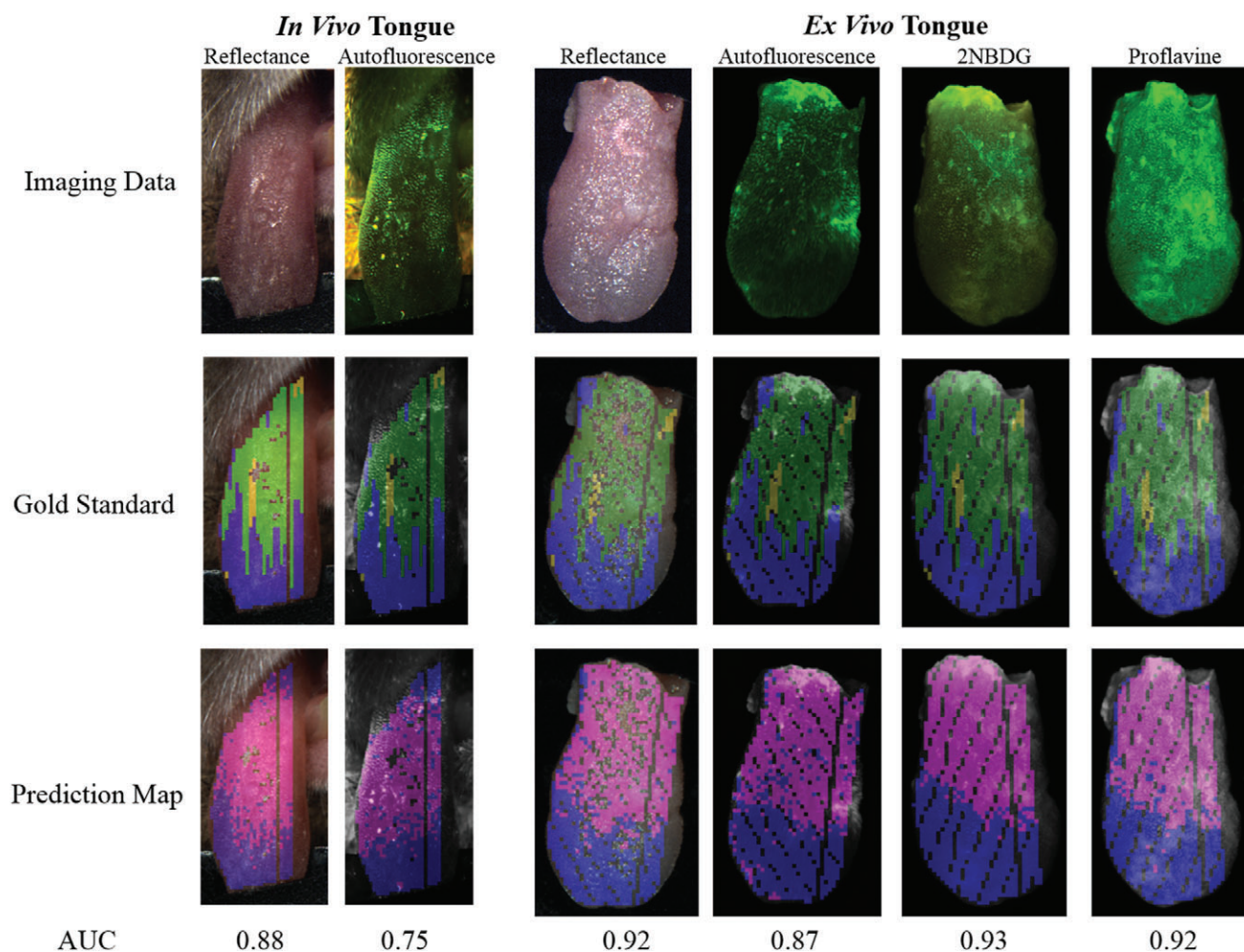


FIGURE 6 Example cancer prediction results for in vivo and ex vivo tongues with different imaging methods. The first rows are the RGB composite images of all imaging data. The second and third rows are gold-standard color maps and predicting color maps corresponding to the block-based feature extraction method, respectively. In the gold-standard color map, blue, green and yellow color indicated normal, dysplasia and CIS. In the prediction color map, magenta and cyan represent neoplastic and non-neoplastic tissue, respectively. The numbers under the prediction map are the AUC values of the corresponding prediction models. In the gold-standard and prediction map, glare pixels excluded from the feature extraction and classification, as well as tissue regions missing the gold standard, were left unlabeled

LDA and random forest were comparable to that of linear SVM. According to [38], linear SVM is a special case of RBF SVM, as linear SVM with parameter C has the same performance as the RBF SVM with some parameter (C, γ). In addition, the choice of linear or kernel SVM depends on the sample size and feature dimension. Linear SVM is faster than RBF SVM as we only need to search for one parameter, that is, the penalty parameter C . Here we showed that linear SVM gave better classification results than RBF SVM for the detection of tongue neoplasia. Our previous work [16, 17] demonstrated that the SVMs with the RBF kernel also produced accurate diagnosis for hyperspectral imaging of xenograft head and neck cancer models in mice. LDA is a fast, efficient, yet accurate classifier that has been widely used in many fields, including the classification of multispectral and hyperspectral images. Our results show that ensemble LDA and LDA produced better classification performance than random forest and QDA. As reported in

Ref. [44], LDA performed better than the K -nearest neighbor (KNN) classifier, decision tree (DT), QDA, ensemble LDA, ensemble KNN, ensemble-DT, etc. in classifying multispectral images of burn wound tissue in a swine model. LDA was also shown to be more accurate than decision trees for classifying multispectral reflectance and autofluorescence images for discriminating cancers and precancers from normal tissue [40].

We grouped dysplasia, CIS and SCC as the neoplastic group and used binary classifiers to distinguish neoplasia and non-neoplastic groups. We showed that dysplastic lesions (including mild dysplasia and moderate dysplasia) were more difficult to differentiate from normal tissue than CIS, most likely due to the histological similarity of normal tissue and dysplasia, as well as the difficulty in making an accurate histological diagnosis of mild dysplasia. Only 2 out of 10 tongues developed small regions of SCC and these regions constituted only a very small part of the

dataset. Therefore, it was more difficult to make a robust and accurate prediction of SCC.

As shown in Figure 6, it is very difficult to directly visualize the abnormal tissue transformations from the RGB images of the tongue. However, label-free HSI was able to detect and delineate lesions on the dorsal surface of the tongue compared to the pathology gold standard, which suggested that wide-field HSI can better capture the neoplastic changes over the heterogeneous tongue surface than point-based spectroscopy methods. It was also found that most of the misclassification errors of HSI occurred in the interface of different pathologies, such as the normal and dysplasia interface and the dysplasia and CIS interface. As seen in the gold-standard map, there were finger-like, protruding dysplasia regions in the interface of dysplasia and normal tongue regions, and where wide-field imaging may not be as sensitive as spectroscopy. In these local regions, spectroscopy-based technology may complement HSI to assist in the identification of the small tumor foci around the tumor-normal interface.

Furthermore, this study compared the diagnostic performance of label-free HSI with autofluorescence and fluorescence of 2 topical dyes in the 4NQO-induced tongue carcinogenesis model. Although autofluorescence and proflavine fluorescence performed slightly better than HSI, the differences were not statistically significant ($P > .05$). In addition, HSI is more favorable for future clinical translation in humans because no contrast agent is needed. Reflectance signal obtained by HSI is a complex interplay of tissue scattering, absorption which is associated with the underlying tissue structure and biochemical properties. Autofluorescence imaging reveals the reduced signal from neoplastic tissue, likely due to decreased collagen crosslinks in the stroma [26]. One FDA-approved, commercial device, called VELscope, detects neoplastic tissue based on their fluorescence visualization loss (FVL) under blue-violet light (400–460 nm). Unfortunately, the sensitivity of the VELscope for detecting malignancy and dysplasia has been reported to be from 30% to 100%, and the specificity ranges from 15.3% to 100% [5]. The mean intensity of 2-NBDG imaging which reflected increased metabolic activity of neoplastic cells, and the standard deviation of proflavine fluorescence intensity which revealed neoplastic transformation by disorganized fluorescence intensity patterns from nuclei accompanying carcinogenesis, were found to be the best discriminative features for neoplasia detection [11]. However, our results suggested that classification with multiband fluorescence intensity was more accurate than the single band in both autofluorescence and fluorescence imaging. One limitation of the topically applied dye on ex vivo tissue lies in the fact that the cut edge of the tissue or any damaged areas on the tissue surface could lead to high fluorescence intensity due to nonspecific uptake, which may confound the detection of neoplastic regions. In the example

case, tissue regions close to the cut edge exhibited very high fluorescence intensity which did not cause confusion because the cut edge region was also neoplastic.

Our ultimate goal is to provide an affordable, noninvasive and accurate tool for oral cancer screening in human patients. There is still a long way to go in order to achieve this goal. Moving forward, we are working on the evaluation of the diagnostic utility HSI on fresh human surgical specimens of tongue SCC and have achieved promising preliminary results. In the future, development of a portable HSI instrument with higher spectral resolution and validation of the HSI device in a large population of human patients could enable clinical translation from bench to bedside and potentially improve oral cancer detection and diagnosis.

5 | CONCLUSION

This proof-of-concept study suggested that hyperspectral imaging is a promising optical modality for quantitative and noninvasive detection and delineation of squamous neoplasia in an animal model. In this study, we designed an animal study to acquire multimodal, hyperspectral images of in vivo and ex vivo mouse tongues from a chemically induced, oral carcinogenesis model. We reconstructed the pathological gold-standard map for the dorsal surface of the tongue in order to validate tumor prediction results. We implemented and validated a variety of machine-learning classifiers, including discriminant analysis, ensemble learning methods, and support vector machines, compared the diagnostic performance in hyperspectral reflectance, autofluorescence and fluorescence images, and generated prediction maps that displayed the location and distribution of neoplasia. In the future, we would like to apply the HSI and machine-learning techniques to the diagnosis of human tongue tumors. Further development of hyperspectral imaging and image quantification methods could provide a noninvasive tool to improve the early detection of oral cancers.

ACKNOWLEDGMENTS

This research was supported in part by NIH grants (CA176684, CA156775, and CA204254). Research reported in this publication was supported in part by Developmental Funds from the Winship Cancer Institute of Emory University under award number P30CA138292. We thank Ms. Jennifer Shelton from the Pathology Core Lab at Winship Cancer Institute of Emory University for her help with histological processing of the tongue specimen.

AUTHOR BIOGRAPHIES

Please see Supporting Information online.

REFERENCES

- [1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, *Int. J. Cancer* **2015**, *136*, E359.
- [2] E. Surveillance, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database in *SEER 18 2005–2011, All Races, Both Sexes by SEER Summary Stage* 2000. , Vol. 2016, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, <http://seer.cancer.gov/statfacts/html/oralcav.html>.
- [3] L. G. T. Morris, A. G. Sikora, S. G. Patel, R. B. Hayes, I. Ganly, *J. Clin. Oncol.* **2011**, *29*, 739.
- [4] D. V. Messadi, *Int. J. Oral Sci.* **2013**, *5*, 59.
- [5] A. Rashid, S. Warnakulasuriya, *J. Oral Pathol. Med.* **2015**, *44*, 307.
- [6] M. W. M. van den Brekel, W. L. Lodder, H. V. Stel, E. Bloemena, C. R. Leemans, I. van der Waal, *Head Neck* **2012**, *34*, 840.
- [7] S. M. Ismail, A. B. Colclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. O'Sullivan, J. M. Summerell, R. G. Newcombe, *Br. Med. J.* **1989**, *298*, 707.
- [8] R. L. Siegel, K. D. Miller, A. Jemal, *CA Cancer J. Clin.* **2016**, *66*, 7.
- [9] M. G. Müller, T. A. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. W. Boone, R. R. Dasari, S. M. Shapshay, M. S. Feld, *Cancer* **2003**, *97*, 1681.
- [10] D. Roblyer, R. Richards-Kortum, K. Sokolov, A. K. El-Naggar, M. D. Williams, C. Kurachi, A. M. Gillenwater, *J. Biomed. Opt.* **2008**, *13*, 024019.
- [11] A. Hellebust, K. Rosbach, J. K. Wu, J. Nguyen, A. Gillenwater, N. Vigneswaran, R. Richards-Kortum, *J. Biomed. Opt.* **2013**, *18*, 126017.
- [12] G. Lu, B. Fei, *J. Biomed. Opt.* **2014**, *19*, 10901.
- [13] H. Akbari, L. V. Halig, D. M. Schuster, A. Osunkoya, V. Master, P. T. Nieh, G. Z. Chen, B. Fei, *J. Biomed. Opt.* **2012**, *17*, 076005.
- [14] R. Pike, S. K. Patton, G. Lu, L. V. Halig, D. Wang, Z. G. Chen, B. Fei, *Proc. SPIE* **2014**, *9034*(9034), 90341w.
- [15] G. Lu, L. Halig, D. Wang, Z. G. Chen, B. Fei, *Proc. SPIE* **2014**, *9034*.
- [16] G. Lu, L. Halig, D. Wang, X. Qin, Z. G. Chen, B. Fei, *J. Biomed. Opt.* **2014**, *19*, 106004.
- [17] R. Pike, G. Lu, D. Wang, Z. G. Chen, B. Fei, *Biomed. Eng. IEEE Trans. PP* **2015**, *1*.
- [18] H. Chung, G. Lu, Z. Tian, D. Wang, Z. G. Chen, B. Fei, *Proc. SPIE* **2016**, *9788*, 978813.
- [19] D. Kanojia, M. M. Vaidya, *Oral Oncol.* **2006**, *42*, 655.
- [20] G. Lu, L. Halig, D. Wang, Z. G. Chen, B. Fei, *Proc. SPIE* **2014**, *90360s*, 9036.
- [21] G. Lu, X. Qin, D. Wang, Z. G. Chen, B. Fei, *Proc. SPIE* **2015**, *94170Q*, 9417.
- [22] D.-W. Sun, *Hyperspectral imaging for food quality analysis and control*, Elsevier, Academic Press Inc., London, UK **2010**.
- [23] B. Park, R. Lu, *Hyperspectral imaging technology in food and agriculture*, Springer, New York **2015**.
- [24] G. Lu, D. Wang, X. Qin, L. Halig, S. Muller, H. Zhang, A. Chen, B. W. Pogue, Z. G. Chen, B. Fei, *J. Biomed. Opt.* **2015**, *20*, 126012.
- [25] P. M. Speight, *Head Neck Pathol.* **2007**, *1*, 61.
- [26] D. Roblyer, C. Kurachi, V. Stepanek, M. D. Williams, A. K. El-Naggar, J. J. Lee, A. M. Gillenwater, R. Richards-Kortum, *Cancer Prev. Res.* **2009**, *2*, 423.
- [27] R. A. Fisher, *Ann. Eugenics* **1936**, *7*, 179.
- [28] Y. Guo, T. Hastie, R. Tibshirani, *Biostatistics (Oxford, England)* **2006**.
- [29] D. Landgrebe, *IEEE Signal Process. Mag.* **2002**, *19*, 17.
- [30] T. Li, S. Zhu, M. Ogihara, *Knowl. Inf. Syst.* **2006**, *10*, 453.
- [31] A. Shashua, *Neural. Process. Lett.* **1999**, *9*, 129.
- [32] H. Tin Kam, *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832.
- [33] C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse, A. Napolitano, *IEEE Trans. Syst. Man Cybern. Syst. Hum.* **2010**, *40*, 185.
- [34] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.
- [35] M. Belgiu, L. Drăguț, *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24.
- [36] G. Lu, X. Qin, D. Wang, MS, ZH, A. Chen, Z. G. Chen, B. Fei, *Proc. SPIE* **2016**, *9788*, 978812.
- [37] F. Melgani, L. Bruzzone, *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778.
- [38] C.-C. Chang, C.-J. Lin, *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1.
- [39] N. Subhash, J. R. Mallia, S. S. Thomas, A. Mathews, P. Sebastian, J. Madhavan, *J. Biomed. Opt.* **2006**, *11*, 014018.
- [40] D. Roblyer, C. Kurachi, V. Stepanek, R. A. Schwarz, M. D. Williams, A. K. El-Naggar, J. J. Lee, A. M. Gillenwater, R. Richards-Kortum, *J. Biomed. Opt.* **2010**, *15*, 066017.
- [41] S. Varma, R. Simon, *BMC Bioinformatics* **2006**, *7*, 91.
- [42] S. Kothari, J. H. Phan, A. N. Young, M. D. Wang, *BMC Med. Imaging* **2013**, *13*, 9.
- [43] S. Kothari, Georgia Institute of Dent. Tech., (2013).
- [44] J. J. Squiers, W. Li, D. R. King, W. Mo, X. Zhang, Y. Lu, E. W. Sellke, W. Fan, J. M. DiMaio, J. E. Thatcher, *Proc. SPIE* **9785**, 9785, 97853L–97853L–97810 **2016**.

How to cite this article: Lu G, Wang D, Qin X, et al. Detection and delineation of squamous neoplasia with hyperspectral imaging in a mouse model of tongue carcinogenesis. *J. Biophotonics*. 2018;11: e201700078. <https://doi.org/10.1002/jbio.201700078>