



Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning

MARTIN HALICEK,^{1,2}  JAMES D. DORMER,¹ JAMES V. LITTLE,³
AMY Y. CHEN,⁴ AND BAOWEI FEI^{1,5,*} 

¹The University of Texas at Dallas, Department of Bioengineering, Richardson, TX 75080, USA

²Emory University and Georgia Institute of Technology, Department of Biomedical Engineering, Atlanta, GA 30332, USA

³Emory University School of Medicine, Department of Pathology and Laboratory Medicine, Atlanta, GA 30322, USA

⁴Emory University School of Medicine, Department of Otolaryngology, Atlanta, GA 30322, USA

⁵The University of Texas Southwestern Medical Center, Department of Radiology, Dallas, TX 75080, USA

*bfei@utdallas.edu

Abstract: The performance of hyperspectral imaging (HSI) for tumor detection is investigated in *ex-vivo* specimens from the thyroid ($N = 200$) and salivary glands ($N = 16$) from 82 patients. Tissues were imaged with HSI in broadband reflectance and autofluorescence modes. For comparison, the tissues were imaged with two fluorescent dyes. Additionally, HSI was used to synthesize three-band RGB multiplex images to represent the human-eye response and Gaussian RGBs, which are referred to as HSI-synthesized RGB images. Using histological ground truths, deep learning algorithms were developed for tumor detection. For the classification of thyroid tumors, HSI-synthesized RGB images achieved the best performance with an AUC score of 0.90. In salivary glands, HSI had the best performance with 0.92 AUC score. This study demonstrates that HSI could aid surgeons and pathologists in detecting tumors of the thyroid and salivary glands.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Thyroid cancer incidence has significantly increased worldwide from 1970 to 2012, despite the fact that mortality from thyroid cancer has decreased [1]. Surgery is the standard treatment for thyroid cancers, and the 5-year survival rate for localized or regional thyroid cancers (excluding anaplastic variant) is above 90% [2]. The most common malignant tumor of the thyroid is papillary thyroid carcinoma (PTC), comprising 70% of thyroid cancers, and there are several variants of PTC, including conventional, follicular, tall-cell, and oncocytic [3]. The initial diagnosis of thyroid tumors is with fine-needle aspiration (FNA) biopsy and histological evaluation of the specimen [3]. Follicular tumors are another cytological type of thyroid neoplasms, which include follicular adenoma, a benign tumor, and follicular thyroid carcinoma (FTC), the malignant form. The requisite diagnostic criterion for follicular carcinoma versus adenoma is definitive invasiveness; no cytological features can provide the diagnosis of FTC, so FNA is therefore useless in making the distinction [4]. Medullary thyroid carcinoma (MTC) is a rare form of thyroid cancer, comprising only 4% of thyroid cancers, that occurs sporadically in most cases, but can be associated with a familial germline mutation [5].

During thyroid tumor resections, intraoperative frozen section (FS) analysis and pathologist consultation can be useful for determining extent of the disease and, according to recent American guidelines, may occasionally confirm malignancy and escalate treatment from partial to total thyroidectomy [6]. For example, in thyroid tumors, 15-30% of preoperative FNA biopsies may

be indeterminate [7]. It remains controversial in thyroid tumor surgeries whether the practice of intraoperative FS can provide relevant diagnostic information, as it can be prone to misdiagnosis [7]. In 4% cases with benign intraoperative FS reports, clinically significant malignancy was found, compared to 6.8% in cases where no FS was performed [8]. This translates to a sensitivity of 22% for identifying malignancy in patients with benign FNA [8]. In the literature, it is suggested that the practice of intraoperative FS may lead to over or under treatment of thyroid tumors [7].

Salivary tumors involve the salivary glands, which are a system of exocrine glands in the mouth that produce saliva to initiate digestion. The major salivary glands are the parotid, the submandibular, and the sublingual salivary glands [9]. The classification of benign and malignant salivary tumors is complex, with over 20 distinct entities according to the most recent standard proposed by the World Health Organization [9–11]. Overall, more than 80% of primary tumors of the salivary glands arise in the parotid gland, which is the largest salivary gland [9,11]. Pleomorphic adenoma is the most common benign tumor of the salivary glands (60%) and typically occurs in the parotid glands [11]. Mucoepidermoid carcinoma is the most common malignant neoplasm of the parotid gland [9]. Adenoid cystic carcinoma is a malignant tumor that can occur with equal likelihood in the submandibular and parotid glands [9]. Polymorphous low-grade adenocarcinoma (PLGA) is a rare malignant tumor, commonly found in minor salivary glands of the hard or soft palate [12]. In surgical resection of salivary tumors, the sensitivity of intraoperative FS for detecting malignant parotid gland tumors with benign FNA was only 33%, suggesting difficulty in diagnosing low grade tumors [13]. Moreover, FS for salivary tumors carries the risk of tumor seeding and may not provide definitive diagnosis [14]. Nonetheless, the combination of preoperative FNA and intraoperative FS leads to high diagnostic accuracy overall for salivary tumors [15].

With the goal of image-guided surgery, hyperspectral imaging (HSI) is an emerging technology in biomedicine that has been used for cancer detection studies both *ex-vivo* and *in-vivo* [16–18]. HSI has been explored for brain cancer detection *in-vivo* [19,20]. Additionally, HSI has been proposed for laparoscopic cancer detection in colorectal surgeries with demonstrated potential [21,22]. The ability of HSI to identify ideal transection margins for colorectal tissues has been demonstrated after devascularization by blocking vascular anastomoses [23]. Our group reported HSI studies of head and neck cancer using *ex-vivo* human surgical specimens [24–28].

In order to leave the parathyroid glands intact during surgery, Barberio et al. demonstrated that HSI may be beneficial in detecting parathyroid glands from thyroid tissue during thyroidectomy [29]. For surgeries of salivary tumors, one challenge is leaving the facial nerve intact, which runs through the parotid gland and can cause facial paresis if injured. Wisotzky et al. showed that HSI can identify the facial nerve in the parotid gland [30]. The submandibular and sublingual salivary glands are surrounded by an anatomical variety of normal tissues in the oral cavity. Previous work from our group has demonstrated that HSI can distinguish amongst normal tissues in the oral cavity, such as stratified squamous epithelium, normal salivary gland, and skeletal muscle [27].

In this large study of 82 patients, we perform tumor detection in 200 thyroid tissue specimens from 76 patients in inter-patient testing experiments, and salivary gland tumor detection was investigated using 16 salivary gland tissue specimens from 6 patients. This is the most comprehensive study to date of tumor detection in thyroid and salivary glands to thoroughly assess the feasibility of label-free, non-contact, and non-ionizing HSI-based imaging modalities for computer aided tumor detection. The outcomes of this work will help guide future HSI and autofluorescence studies and determine the specific benefits that HSI may offer for tumor detection in thyroid and salivary gland tissues.

2. Methods

In this study, *ex-vivo* tissue specimens from the thyroid and salivary glands were imaged with optical imaging modalities; histological sections were prepared from the specimens for ground truths; patients were categorized and used to train, validate, and test deep learning algorithms; and performance was calculated to compare the methods.

2.1. Ex-vivo surgical specimen dataset

For this study, 216 surgical specimens were acquired from 82 patients undergoing routine resection of thyroid tumors or salivary gland tumors at the Emory University Hospital Midtown, who were recruited by giving written, informed consent to an institutional research coordinator. Table 1 shows the categorization of patients and tissue specimens. All patient data were de-identified by the research coordinator. The Institutional Review Board (IRB) of Emory University approved all research protocols and imaging methods. Three types of fresh, *ex-vivo* surgical specimens were obtained from the surgical pathology department during clinical service. We aimed to acquire a sample of normal tissue (N), tissue from the primary tumor (T), and a specimen of the tumor-involved margin that contains both tumor and normal tissue (TN), all of which were confirmed by histopathological analysis. The size of the tissue specimens was approximately 10×6×2 mm on average. Additionally, the final clinical pathology report was made available after de-identification.

Table 1. Number of patients and tissue specimens for this study. Patients are divided by cohort and sub-group. The tissue specimens are categorized into tumor (T), tumor-normal margin (TN), and normal (N).

Group	Patients	Number of Tissue Specimens		
		T	TN	N
Thyroid Tumor Cohort				
Papillary Thyroid Carcinoma (PTC)	54	41	38	59
Medullary Thyroid Carcinoma (MTC) & Insular Carcinoma	6	4	6	8
Follicular Adenoma & Carcinoma	13	14	10	12
Poorly Differentiated Carcinoma	3	4	2	2
Salivary Tumor Cohort				
Parotid Gland Tumor	3	2	3	3
Other Salivary Gland Tumor	3	4	2	2
Total	82	69	61	86

The tissue samples collected for this study were categorized by an experienced pathologist into six groups according to tumor subtype, divided into two broad cohorts: thyroid tumors and salivary gland tumors. The thyroid tumor cohort was comprised of 200 tissue specimens from 76 patients. The malignant tumors included in this cohort were PTC ($N=54$), MTC ($N=5$), insular carcinoma ($N=1$), follicular carcinoma ($N=8$), and poorly differentiated thyroid carcinoma ($N=3$). The benign tumors of the thyroid were follicular adenoma ($N=5$). The only thyroid cohort tissues excluded from this study were six patients with benign thyroid hyperplasia/goiter.

The cohort of salivary gland tumors was comprised of 16 tissue specimens from 6 patients. Two patients had benign pleomorphic adenoma ($N=2$) of the parotid gland. Four patients had malignant tumors of the salivary glands: mucoepidermoid carcinoma ($N=1$), salivary duct carcinoma of the parotid gland ($N=1$), PLGA of the hard palate ($N=1$), and adenoid cystic carcinoma ($N=1$). The patient demographics and relevant cancer properties are shown in Table 2.

Table 2. Patient demographics and cancer properties for the patients recruited for this study. Intraoperative pathologist consultations (IPCs) using intraoperative frozen section (FS) analysis are also reported.

Property	Number	Percentage
Demographics		
Average Age (years)	49.7	-
Male	28	34%
Female	54	66%
Tobacco History	26	32%
Ethnicity		
White	47	57%
Black	19	24%
Asian	6	7%
Other / Unknown	10	12%
Primary Location / Tumor Type		
<i>Thyroid Gland</i>		
Papillary Carcinoma	54	66%
Follicular Carcinoma	8	10%
Follicular Adenoma	5	6%
Medullary Carcinoma	5	6%
Poorly Differentiated Carcinoma	3	4%
Insular Carcinoma	1	1%
<i>Salivary Gland</i>		
Pleomorphic Adenoma	2	2%
Adenoid Cystic Carcinoma	1	1%
Mucoepidermoid Carcinoma	1	1%
Adenocarcinoma (PLGA)	1	1%
Salivary Duct Carcinoma	1	1%
Cancer Stage		
pT1	19	23%
pT2	19	23%
pT3	26	32%
pT4	9	11%
Avg. T Size (cm)	3	-
N+ Status	33	40%
IPC (Averages)		
Surgeries with IPC	19	23%
IPC/Surgery	1.6	-
Tissues/Surgery	1.1	-
Time/IPC (min)	19.1	-
Time/Tissue (min)	17.3	-

2.2. Optical imaging modalities

To assess the ability of HSI for tumor detection, several other optical imaging modalities were acquired for comparison, including both label-free and fluorescent dye-based methods. It was

hypothesized that HSI would outperform fluorescence methods due to lack of sufficient target specificity in 2-NBDG and proflavin. In the following sections, the image acquisition systems are described for hyperspectral reflectance imaging, HSI-synthesized RGB multiplex imaging, autofluorescence imaging, and two fluorescent dye-based imaging techniques: 2-NBDG and proflavin.

2.2.1. Hyperspectral imaging

A CRi Maestro HS system (Perkin Elmer Inc., Waltham, Massachusetts) was used to acquire HSI of the *ex-vivo* specimens. The HS system performs spectral scanning from 450 to 900 nm using a Xenon light source and liquid crystal tunable filter (LCTF) with 5 nm spectral resolution [24,31]. The image size of the HSI was 1040×1392×91 pixels (height×width×spectral bands), and the corresponding specimen-level spatial resolution was 25 μm per pixel. Acquisition time for an HSI was approximately one minute.

The raw HS data (I_{raw}) were normalized band-by-band (λ) by subtracting the inherent dark current of the sensor and dividing by a white reference disk for all pixels (x,y), according to the following equation.

$$I_{norm}(x, y, \lambda) = \frac{I_{raw}(x, y, \lambda) - I_{dark\ current}(x, y, \lambda)}{I_{white\ ref}(x, y, \lambda) - I_{dark\ current}(x, y, \lambda)}$$

The average spectral signatures after white-dark calibration are shown for all groups included in this paper by cohort in Fig. 1.

2.2.2. HSI-Synthesized RGB images

A multiplex image is a synthetic composite image generated from a hyperspectral image. For this work, several three-band (RGB) multiplex images were synthesized from the normalized reflectance HSI hypercubes. The first synthetic RGB was generated from the HSI by applying a Gaussian kernel in each color region, which is referred to as HSI-synthesized Gaussian RGB composite. The second RGB image was constructed from human color perception curves originally proposed by Judd et al. 1951 [32] and expanded by Vos 1978 [33]. For some tissues, a standard RGB image was also captured for comparison by an RGB camera. Figure 2 shows a representative surgical tissue specimen of thyroid cancer from an RGB-captured camera, HSI-synthesized Gaussian RGB multiplex, and HSI-synthesized RGB with human eye perception.

In this paper, we use HSI to simulate three-band images using RGB multiplex imaging. However, the sensors of RGB cameras typically employ the Bayer filter [34, 35] for adapting RGB color spectrums similar to human-eye. While the spectrum of blue and green are typically consistent, different RGB camera sensor types have differing sensitivity to the red channel components between 400 and 500 nm [34, 35]. Therefore, for a subset of thyroid tumor specimens, the component of red channel response between 400 and 500 nm was manipulated to simulate if this would have an effect on performance from the HSI-synthesized human eye RGB multiplex images.

2.2.3. Autofluorescence imaging

Autofluorescence imaging uses the emission from intrinsic fluorophores in biological tissue that are stimulated to fluoresce by external excitation light at specific wavelengths. The autofluorescence images were captured using a 455 nm excitation source and a 490 nm long-pass filter using the CRi Maestro imaging system. The long-pass filter removes any of the external light from the source that would be reflected into the image and allows capturing images of emission-only photons, according to Stokes' theorem. The autofluorescence images were acquired from 500 to 720 nm in 10 nm increments to produce a hypercube with 23 spectral bands and final size of 1040×1392×23 pixels.

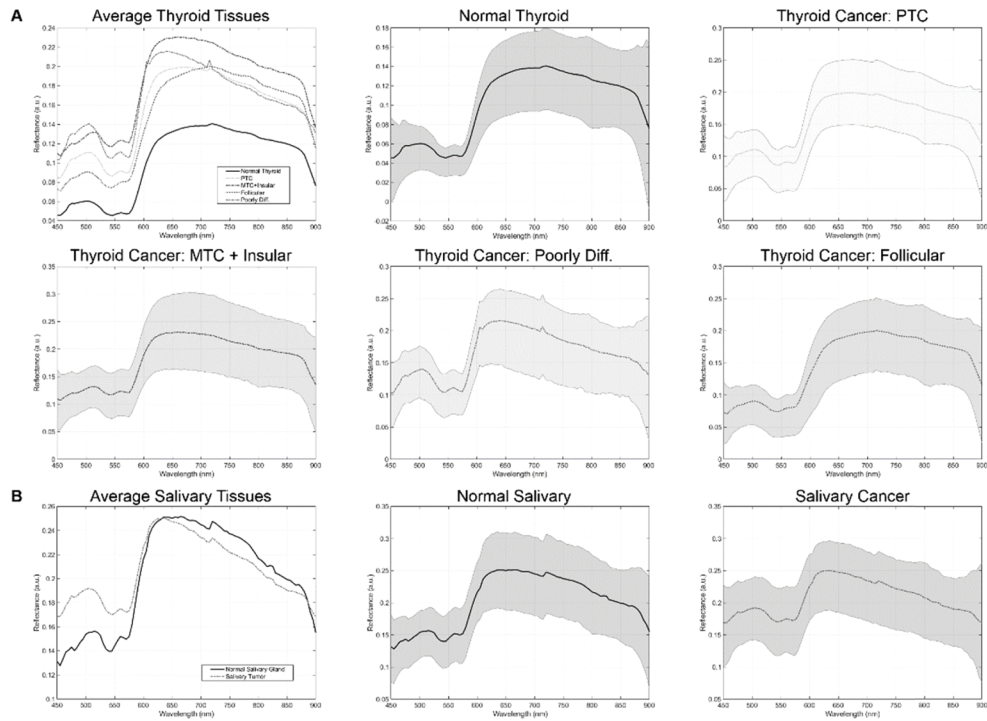


Fig. 1. Average hyperspectral signatures for the tissues in the thyroid cohort (a) and the salivary cohort (b). Subplots show the spectral signature with standard deviation for each tissue sub type in both cohorts.

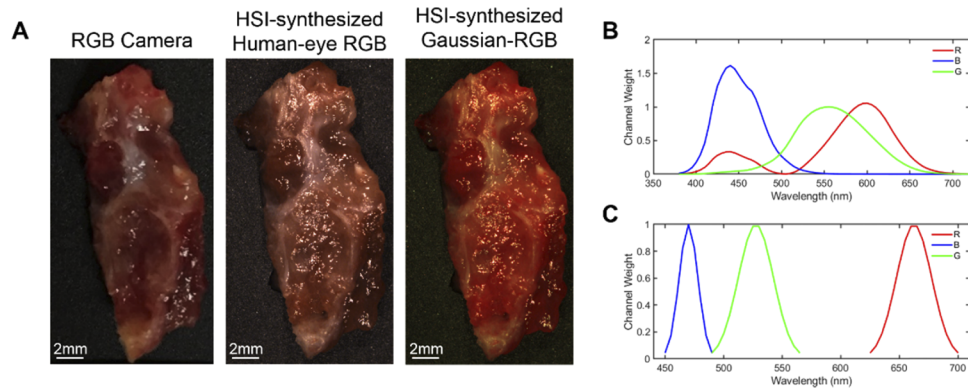


Fig. 2. A representative specimen of thyroid cancer. (a) Left to right: RGB image from standard RGB camera; HSI-synthesized RGB human-eye multiplex image made from reflectance HSI using Vos et al. 1978 method; HSI-synthesized Gaussian RGB multiplex image made from reflectance HSI. (b) Spectral signatures of human-eye color perception of red (R), green (G), and blue (B) colors proposed by Vos et al. 1978. (c) Gaussian kernels used for generating Gaussian RGB multiplex images.

2.2.4. 2-NBDG imaging

A fluorescently tagged glucose molecule, 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose (2-NBDG), is a dye that targets cancer regions by producing a stronger signal measured

from regions with higher metabolic glucose uptake. After the hyperspectral and autofluorescence imaging methods described above, the tissues were incubated for 20 minutes in a 160 μM 2-NBDG solution (Cayman Chemical, Ann Arbor, MI, USA) at 37 degrees Celsius, quickly rinsed in 1 \times phosphate buffered solution (PBS) to remove excess dye, and fluorescence imaging was performed using the CRi Maestro. The images were acquired with the same excitation light source at 455 nm and a long-pass filter at 490 nm from 500 to 720 nm in 10 nm increments, producing a hypercube that has 23 spectral bands.

2.2.5. Proflavin imaging

The second dye used for fluorescence imaging was proflavin dye, which is unaffected by previous 2-NBDG dye because it has a significantly stronger optical signal comparatively. Proflavin fluorescent dye binds to DNA and thus allows visualization of nuclear morphology, which can improve the ability of machine learning based cancer detection methods [36]. Keratin is also a target of proflavin dye, but this should not affect the glandular tissues involved in this study [36,37]. For proflavin imaging, the tissue samples were incubated for 2 minutes in a 0.01% proflavin solution (Sigma Aldrich, St. Louis, MO, USA) at room temperature, and the tissues were rinsed in PBS before imaging with the CRi Maestro. The images were acquired with an excitation light source at 455 nm and a long-pass filter at 490 nm from 500 to 720 nm in 10 nm increments, producing a hypercube that contains 23 spectral bands.

2.3. Histological ground truth

The ground truths for the optical imaging modalities were achieved using digitized histology imaging. After acquiring all images, the tissue specimens were inked at the top, bottom, left and right edges, and back surface of the tissue to identify tissue orientation in histological sections. Tissues were then fixed in formalin, paraffin embedded, and sectioned with a microtome, and 5 μm slices were made from the surface that was optically imaged. The first high quality slice was kept to serve as the histology ground truth, processed with hematoxylin and eosin staining, and digitized using whole-slide scanning at 40 \times objective [38]. A board-certified pathologist with expertise in head and neck pathology annotated the tumor and normal areas on the digital histology images.

A binary mask was made of the contoured digital histology images, which served as the ground truth for the optical imaging modalities. Due to tissue deformations during histological processing and slide preparation, the histology ground truth masks needed to be registered to the gross-level optical images. The digital histology slide was registered in a semi-automated method according to a previously established pipeline of affine, land-mark, and deformable registration to the gross-level HSI [39,40]. The transformation was applied to the binary histology mask, and thus a ground truth mask was generated matching the gross-level optical images of the tissue specimens.

2.4. Experimental design

The deep learning experimental designs involved training, validation, and testing. However, for the two cohorts, thyroid tumors ($N=76$ patients) and salivary tumors ($N=6$ patients), there were different designs of the data partitioning, which was required because of the significant differences in sample size. Therefore, the thyroid cohort is used to produce fully-independent inter-patient results. However, the salivary cohort is used to produce intra-patient training and testing, as described below in detail.

2.4.1. Thyroid tumors

Tumor detection of the thyroid gland was performed in fully-independent patients, divided across 5 folds. Each fold served as the fully-independent testing group, while training and validation

was performed on the patients in the remaining 4 folds, as depicted in Fig. 3. This design was selected to allow test-level performance metrics for all 76 thyroid patients.

Pt. 01 – 15	Fold 1	Training & Validation	...	Testing
Pt. 16 – 30	Fold 2			Training & Validation
Pt. 31 – 45	Fold 3			
Pt. 46 – 60	Fold 4			
Pt. 61 – 76	Fold 5	Testing		

Fig. 3. Schematic depicting the experimental design of fully-independent training, validation, and testing paradigms for the 76 patient thyroid tumor cohort.

2.4.2. Salivary gland tumors

Tumor detection in the limited sample size of salivary gland tumors was performed using intra-patient experiments. Training and validation was performed on the patients' primary tumor (T) and normal (N) tissues, and testing was performed on the tumor-normal (TN) margin tissue. Figure 4 shows a schematic diagram of the training and testing paradigm. The salivary gland cohort was separated into tumors of the parotid gland ($N = 3$ patients) and other salivary glands ($N = 3$), as shown in Table 1.

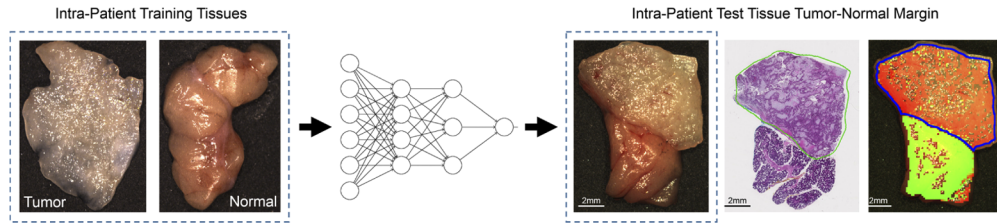


Fig. 4. Flow diagram of intra-patient experiments of the salivary gland, with representative tumor of the parotid gland. Intra-patient T and N tissues were used for MLP (multilayer perceptron) training, and TN tissue specimens were used for testing. The histological ground truth is shown with tumor contour in green. The predicted tumor heat-map overlay onto the RGB image is shown with tumor predictions (red) and normal predictions (green). Areas of specular glare in the heat-map are not classified, and the ground-truth tumor contour is in blue.

2.5. Convolutional neural network

For thyroid tumor detection using 200 tissue specimens from 76 patients, a convolutional neural network (CNN) was developed for the effective classification of thyroid tissue into tumor and normal using a patch-based approach. The inception-v4 CNN architecture [41] was selected because it is one of the top performing CNNs on standard tasks like Image-Net, yet has a manageable number of hyperparameters. HSI data has several unique challenges due to data size. Therefore, the CNN required modification for HS data pre-processed into image-patches of size $25 \times 25 \times C$ pixels, where C represents the number of spectral bands. The first convolutional layers were modified for the smaller patch-size necessitated by HS data, and the operating resolution in the modular inception blocks was reduced to allow more efficient training and classification using the CNN. Additionally, squeeze-and-excitation modules were added to increase the performance of the CNN [42]. The implemented CNN architecture schematic is detailed in Fig. 5.

Image patches ($25 \times 25 \times C$) were generated using a sliding window approach with a stride of 13 pixels, and the data pre-processing of HSI was performed in MATLAB (MathWorks, Inc., Natick,

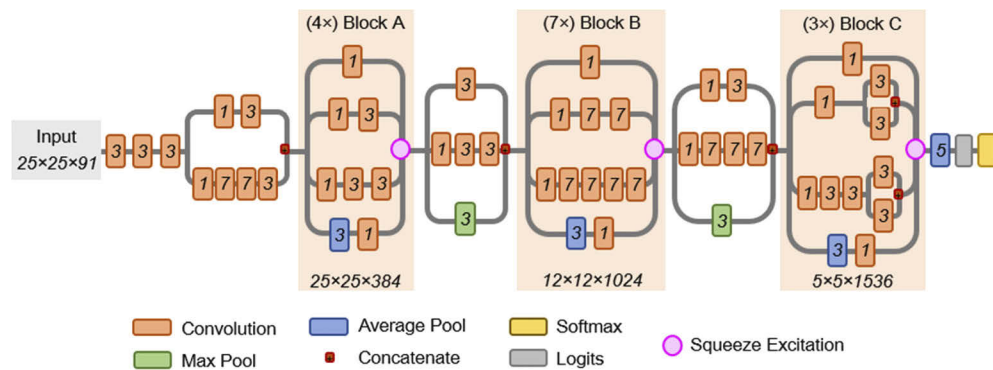


Fig. 5. Schematic diagram of the modified inception v4 CNN architecture. The CNN was customized to operate on the $25 \times 25 \times 91$ patch-size selected. The receptive field size and number of convolutional filters is shown at bottom of each inception block. The convolutional kernel size used for convolutions is shown in italics inside each convolution box. Squeeze-and-excitation modules were added to the CNN to increase performance.

MA). All deep learning programming was done in the TensorFlow python software package [43] on an Ubuntu machine and accelerated with CUDA execution on Titan-XP NVIDIA GPUs (Nvidia Corp., Santa Clara, CA). The CNN loss function was cross-entropy, the optimizer utilized was Adadelta with an initial learning rate of 1.0, and validation performance was calculated every 2 epochs of training data. Training each CNN model was performed for 14 epochs of $8 \times$ augmented (reflections and rotations) training data, which took about 23 hours to train. Deployment of a fully-trained CNN model on a single GPU to classify a new thyroid tissue specimen, which consisted of hundreds of patches, was 20 ± 8 seconds (avg. \pm st. dev.) for all imaging modalities. The heatmaps were produced by averaging the results of overlapping pixel regions in image-patches, since the 25×25 patches were produced with a stride of 13 pixels, which was used to produce a smoother and less coarse final result.

2.6. Multilayer perceptron

For salivary gland tumor detection of 16 tissue specimens from 6 patients, a simplified artificial neural network, called a multilayer perceptron (MLP), was used for intra-patient detection with spectral information only. The MLP consisted of a 91 unit spectral vector input, a single hidden layer with 128 neurons, and an output layer of 2 nodes (normal or tumor). This simplified MLP was applied only to the salivary gland tumor cohort and selected to limit overfitting in this small dataset. The salivary gland cohort was separated into parotid gland tumors ($N = 3$ patients) and other salivary gland tumors ($N = 3$ patients).

The spectral signatures of tissue were extracted by local averaging of 5×5 pixel blocks to reduce noise. Image pre-processing was used to remove the specular glare pixels from both training and testing. For each group, the normal (N) and tumor-only (T) specimens' spectra were used for training (85%) and a subset for validation (15%), and the tumor-normal (TN) margin tissue spectra was used for testing. For the parotid group and other salivary gland group separately, all patients' training samples were combined into one training group (6 tissues), and the three TN test specimens were classified. Training was performed on the order of a few minutes, and testing was produced in about one second.

2.7. Performance evaluation

The principal evaluation metric used for this study was area under the curve (AUC) of the receiver operator curve (ROC). The AUC score was selected because it is robust to class imbalances

within tissues and provides an estimate of performance at all possible thresholds of separating the normal and tumor classes. Additionally, the accuracy, sensitivity, and specificity were calculated and reported using the tumor probability threshold from the validation data. All results were calculated on a tissue specimen level and averaged. Additionally, for the final testing results of both cohorts, the 10 pixels at the edge of tissue, corresponding to 0.25 mm, were excluded from performance calculations. The imaging protocol for ex-vivo tissue specimens was performed using a flat imaging surface, so the tissue edges created unnatural curvature where the tissue was too thin to provide an adequate imaging signal. Statistical significance was calculated for the test results using Student's *t*-test and a 0.05 *p*-value threshold.

3. Results

3.1. Thyroid tumors

Tumor detection for the thyroid cohort with all cancer types combined ($N = 76$) demonstrated that the HSI-synthesized RGB multiplex images generated from the HSI were the best performing results in terms of average AUC scores with 0.89 and 0.90 for HSI-synthesized Gaussian RGB multiplex and the HSI-synthesized human-eye RGB multiplex, respectively. HSI performed with an average AUC score of 0.86. Full results of thyroid tumor detection by AUC score, accuracy, sensitivity, and specificity for all imaging methods are shown in Table 3 and separated by cancer type.

The average and median AUC scores are presented in Fig. 6(A) and (B) across all thyroid tumor types with statistical significance. Combining all thyroid tumors, both implementations of three-band, HSI-synthesized RGB multiplex imaging (average AUC score of 0.89 for Gaussian-RGB multiplex and 0.90 for human-eye RGB multiplex) outperformed autofluorescence (0.85 AUC score), 2-NBDG (AUC score of 0.86), and proflavin (0.83 AUC score) to a degree of statistical significance (all $p < 0.05$). Additionally, HSI-synthesized human-eye RGB multiplexing also significantly outperformed HSI ($p < 0.05$). For the PTC sub-group ($N = 54$), both HSI-synthesized RGB multiplex images statistically outperformed autofluorescence, 2-NBDG, and proflavin (all $p < 0.05$). For the MTC group ($N = 6$), both HSI-synthesized RGB multiplex images significantly outperformed 2-NBDG and proflavin (all $p < 0.05$). For the follicular tumor group, autofluorescence outperforms the other methods in AUC score, but the difference is not significant ($p > 0.05$). Lastly, poorly differentiated thyroid carcinomas were classified with the highest AUC score from HSI-synthesized human-eye RGB multiplex imaging, but not significantly ($p > 0.05$).

The different imaging modalities and respective probability heat-maps for tumor detection are shown in Fig. 7 for all groups of thyroid tumors. As can be seen, HSI shows the most consistent heat-maps around regions of specular glare, compared to the HSI-synthesized RGB multiplex methods. As shown in Fig. 6(B), the median AUC scores show that HSI (0.95) and the two HSI-synthesized RGB multiplex methods (0.95 and 0.96) have approximately equivalent median performance for combined thyroid tumors.

The median AUC scores are substantially greater than the averages, which indicates that the distribution of performance results tends to be more accurate than the average reflects. Histogram analysis of percent difference of the HSI and HSI-synthesized RGB multiplex imaging methods shows that HSI-synthesized RGB outperforms HSI in a relatively small number of tissue specimens, which causes the average AUC scores for HSI-synthesized RGB multiplexing to be greater than HSI. Figure 8(A) and (C) show the histograms of percent difference in tissues from HSI to both Gaussian-RGB multiplex and the human-eye RGB multiplex, respectively. Additionally, Fig. 8(B) and (D) show the tissue specimens with the largest differences in AUC scores compared to HSI, where HSI-synthesized RGB multiplexing still works quite well.

The three-band HSI-synthesized RGB multiplex images from HSI are meant to represent RGB imaging. However, these multiplex images are still constructed from HSI data. Standard RGB camera sensors have different responses to the component of the red channel between 400 and

Table 3. Performance results of the optical imaging modalities for the thyroid tumor cohort (average \pm SEM). The best performing modality for each groups' evaluation metrics is bolded.

		AUC	Accuracy	Sensitivity	Specificity
HSI	All Thyroid Tumors ($N = 82$)	0.86 \pm 0.02	78 \pm 2%	80 \pm 3%	74 \pm 3%
	PTC ($N = 54$)	0.86 \pm 0.02	76 \pm 2%	79 \pm 3%	71 \pm 4%
	MTC & Insular Ca. ($N = 6$)	0.81 \pm 0.09	85 \pm 5%	85 \pm 5%	72 \pm 11%
	Follicular Ad. & Ca. ($N = 13$)	0.90 \pm 0.04	80 \pm 4%	80 \pm 7%	82 \pm 5%
	Poorly Diff. Ca. ($N = 3$)	0.90 \pm 0.08	81 \pm 15%	73 \pm 23%	96 \pm 4%
Autofluorescence	All Thyroid Tumors ($N = 82$)	0.85 \pm 0.02	76 \pm 2%	83 \pm 2%	68 \pm 3%
	PTC ($N = 54$)	0.81 \pm 0.03	72 \pm 2%	79 \pm 3%	62 \pm 4%
	MTC & Insular Ca. ($N = 6$)	0.86 \pm 0.06	80 \pm 5%	83 \pm 7%	78 \pm 6%
	Follicular Ad. & Ca. ($N = 13$)	0.95 \pm 0.02	87 \pm 3%	93 \pm 3%	81 \pm 5%
	Poorly Diff. Ca. ($N = 3$)	0.98 \pm 0.01	95 \pm 1%	95 \pm 3%	93 \pm 1%
2-NBDG Dye	All Thyroid Tumors ($N = 82$)	0.86 \pm 0.02	78 \pm 2%	75 \pm 3%	80 \pm 3%
	PTC ($N = 54$)	0.84 \pm 0.02	76 \pm 2%	72 \pm 3%	79 \pm 3%
	MTC & Insular Ca. ($N = 6$)	0.84 \pm 0.05	80 \pm 5%	74 \pm 8%	72 \pm 10%
	Follicular Ad. & Ca. ($N = 13$)	0.92 \pm 0.03	84 \pm 4%	82 \pm 6%	89 \pm 4%
	Poorly Diff. Ca. ($N = 3$)	0.91 \pm 0.08	79 \pm 11%	81 \pm 15%	89 \pm 11%
Proflavin Dye	All Thyroid Tumors ($N = 82$)	0.83 \pm 0.02	75 \pm 2%	68 \pm 3%	80 \pm 3%
	PTC ($N = 54$)	0.82 \pm 0.02	73 \pm 2%	66 \pm 4%	77 \pm 4%
	MTC & Insular Ca. ($N = 6$)	0.70 \pm 0.08	71 \pm 4%	60 \pm 9%	71 \pm 8%
	Follicular Ad. & Ca. ($N = 13$)	0.91 \pm 0.03	81 \pm 4%	73 \pm 7%	93 \pm 2%
	Poorly Diff. Ca. ($N = 3$)	0.98 \pm 0.01	95 \pm 2%	91 \pm 5%	97 \pm 2%
HSI-synthesized Gaussian-RGB	All Thyroid Tumors ($N = 82$)	0.89 \pm 0.02	79 \pm 2%	77 \pm 2%	82 \pm 3%
	PTC ($N = 54$)	0.87 \pm 0.02	77 \pm 2%	76 \pm 3%	79 \pm 4%
	MTC & Insular Ca. ($N = 6$)	0.95 \pm 0.03	88 \pm 4%	91 \pm 4%	82 \pm 8%
	Follicular Ad. & Ca. ($N = 13$)	0.90 \pm 0.02	77 \pm 3%	67 \pm 6%	91 \pm 2%
	Poorly Diff. Ca. ($N = 3$)	0.98 \pm 0.01	94 \pm 3%	92 \pm 4%	95 \pm 4%
HSI-synthesized Human-Eye RGB	All Thyroid Tumors ($N = 82$)	0.90 \pm 0.02	79 \pm 2%	80 \pm 2%	79 \pm 3%
	PTC ($N = 54$)	0.88 \pm 0.02	78 \pm 2%	80 \pm 3%	76 \pm 4%
	MTC & Insular Ca. ($N = 6$)	0.96 \pm 0.02	88 \pm 3%	93 \pm 3%	85 \pm 6%
	Follicular Ad. & Ca. ($N = 13$)	0.92 \pm 0.02	76 \pm 3%	68 \pm 7%	86 \pm 4%
	Poorly Diff. Ca. ($N = 3$)	0.99 \pm 0.01	91 \pm 3%	92 \pm 3%	93 \pm 5%

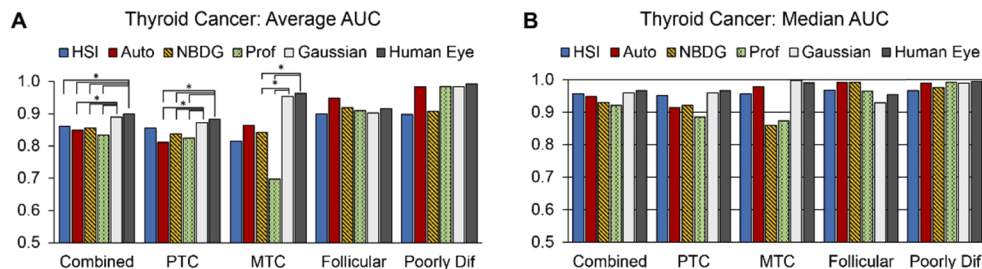


Fig. 6. Average and median AUC scores from thyroid tumor detection. (a) average AUC scores for thyroid tumor detection across all tissue specimens grouped by tumor subtype; statistical significance (*, $p < 0.05$) is shown above. (b) median AUC scores of tumor subtype detection.

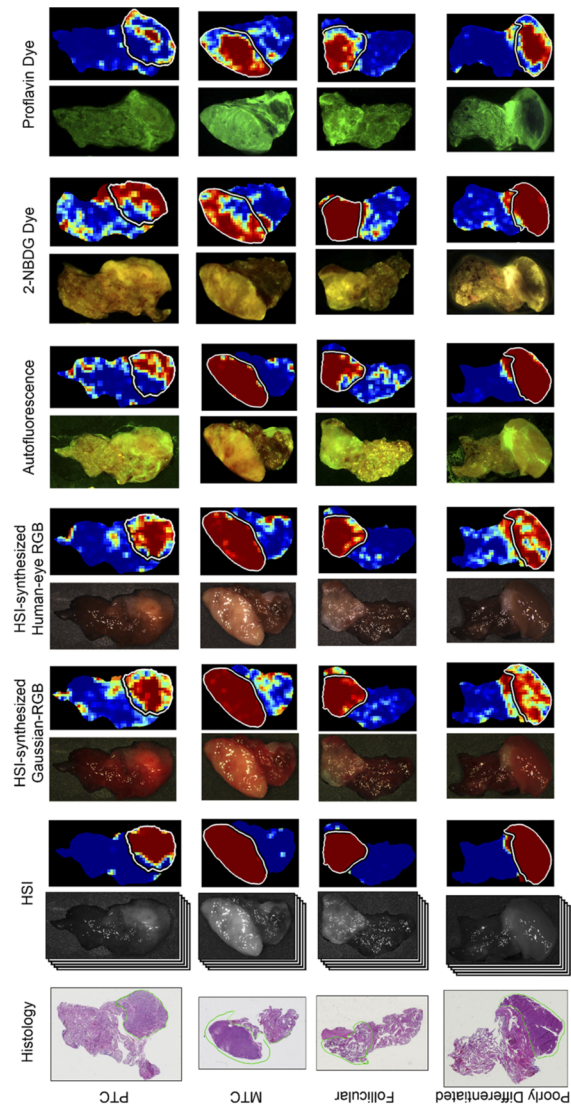


Fig. 7. Representative tissue images and corresponding classification heat-maps from all modalities from patients with thyroid carcinoma. Columns from left to right: histology, HSI with heat-map, HSI-synthesized Gaussian-RGB multiplex with heat-map, HSI-synthesized human-eye RGB multiplex with heat-map, autofluorescence with heat-map, 2-NBDG dye image with heat-map, Proflavin dye image with heat-map. Rows from top to bottom: papillary thyroid carcinoma (PTC), medullary thyroid carcinoma (MTC), follicular thyroid carcinoma, and poorly differentiated thyroid carcinoma. The contours in white (in heat-maps) and green (on histology) outline the cancerous regions. Predicted tumor heat-maps range from dark blue (predicted normal) to dark red (predicted cancer).

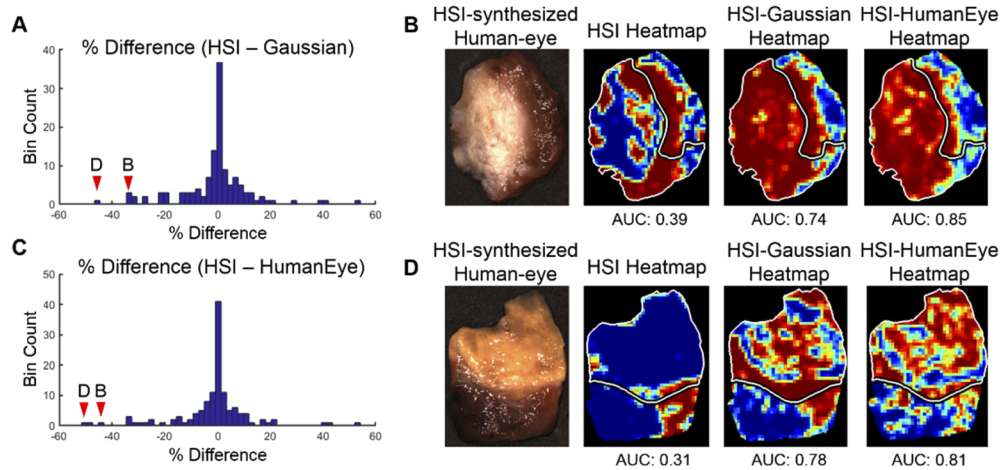


Fig. 8. Differences in AUC score performance comparing HSI against HSI-synthesized Gaussian-RGB multiplex and HSI-synthesized human-eye RGB multiplexing. (a) Histogram of percent difference in AUC scores of tissue specimens between HSI and HSI-synthesized Gaussian-RGB multiplex imaging. The arrows show the bins that contain the patient specimens shown in the right panels, which are the two worst performing tissues. (b) RGB image of tissue specimen with large difference in AUC score performance between heat-maps produced from HSI, HSI-synthesized Gaussian multiplex, and HSI-synthesized human-eye multiplex image. (c) Histogram of percent difference in AUC scores of tissue specimens between HSI and HSI-synthesized human-eye RGB multiplex imaging. The arrows show the bins that contain the patient specimens shown in the right panels, which are the two worst performing tissues. (d) RGB image of the tissue specimen with the largest difference in AUC score performance between heat-maps produced from HSI, HSI-synthesized Gaussian multiplex, and HSI-synthesized human-eye multiplex image. The tumor margin is delineated in white.

500 nm. Therefore, for fold 1 of the HSI-synthesized human-eye RGB multiplex thyroid tumor detection, the red channel component from 400 to 500 nm was multiplied by half and by zero, and two more CNNs were trained. The results are plotted in Fig. 9. The original human-eye multiplex result for fold 1 using the original 400-500 nm red component was an AUC score of 0.90 for thyroid tumor detection. Completely eliminating the 400-500 nm red component (by multiplying by zero) in human-eye multiplex still resulted in an equivalent AUC score of 0.90 for thyroid tumor detection ($p > 0.05$). Lastly, an equivalent AUC score of 0.89 for thyroid tumor detection was obtained when the 400-500 nm red component was set to half the original value for human eye multiplexing ($p > 0.05$).

To further investigate how CNN methods can utilize HSI and the relevant wavelengths for correctly predicting normal and thyroid tissues, we incorporated the gradient class-activated maps (grad-CAM) algorithm [44]. Briefly, the method is used for tracing the most relevant gradients from the input data to the class of interest, either normal or tumor, which is used to infer spectral feature saliency. The mean spectral signatures and class-activated gradients were averaged for 89 tissues that were correctly classified with high AUC scores and separated into normal thyroid and tumor (Fig. 10). As can be observed in Fig. 10(a), the salient spectral features for correctly classifying normal thyroid tissues were from 570-700 nm. In Fig. 10(b), the most salient spectral features for correctly classifying thyroid tumors were also in the range of 550-700 nm, with additional bands near 500 and 750 contributing some lesser importance for classification.

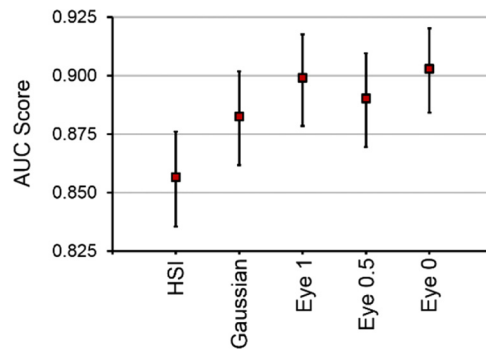


Fig. 9. AUC score results from one fold of the testing data comparing different methods using HSI. The HSI-synthesized RGB multiplex images represent RGB imaging with different parameters. From left to right in the plot: original HSI method, 3-band Gaussian-RGB from HSI, original HSI-synthesized human-eye RGB from HSI, human-eye RGB from HSI synthesized with half of the 400-500 nm red component, and last the human-eye RGB from HSI synthesized with none of the 400-500 nm red component. Values shown are average AUC score from all tissues in one fold of testing data with 95% confidence interval error bars.

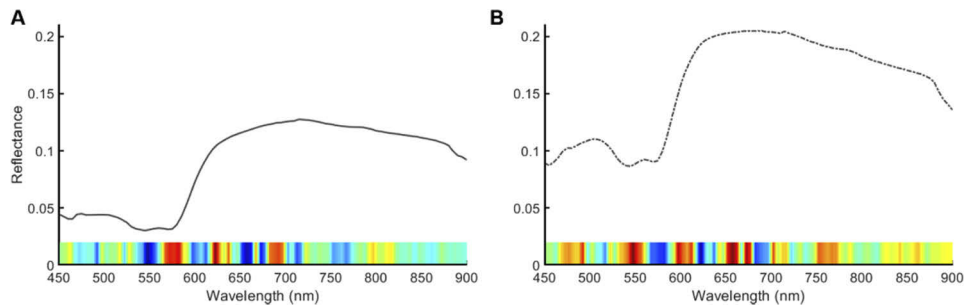


Fig. 10. Mean spectral signatures of correctly-classified normal thyroid tissues (a) and thyroid tumors (b). The saliency of spectral features is identified below each plot using the grad-CAM technique. Red hues represent the most important features for correctly predicting each tissue class, and blue color hues represent less important wavelengths for correctly predicting each class.

3.2. Salivary gland tumors

The intra-patient tumor detection results of the salivary gland tumors cohort are separated into parotid gland tumors and other salivary gland tumors. For parotid gland tumors, HSI was the best performing imaging modality with an AUC score of 0.92, accuracy of 88%, sensitivity of 90%, and specificity of 79% (all differences were not significant, $p > 0.05$). For tumors of other salivary glands, autofluorescence was the best performing imaging modality with an AUC score of 0.80, accuracy of 84%, sensitivity of 77%, and specificity of 85% (all differences were not significant, $p > 0.05$). The full results are shown in Table 4.

Table 4. Performance results from all optical imaging modalities for the salivary tumor cohort, separated by salivary gland group (average \pm SEM). The best performing modality for each groups' evaluation metrics is bolded.

Group	Imaging	AUC	Accuracy	Sensitivity	Specificity
Parotid ($N = 3$)	HSI	0.92 ± 0.04	$88 \pm 4\%$	$90 \pm 4\%$	$79 \pm 7\%$
	Autofluorescence	0.60 ± 0.30	$87 \pm 2\%$	$99 \pm 1\%$	$26 \pm 26\%$
	2-NBDG Dye	0.77 ± 0.12	$87 \pm 2\%$	$97 \pm 3\%$	$31 \pm 31\%$
	Proflavin Dye	0.64 ± 0.32	$89 \pm 2\%$	$96 \pm 2\%$	$54 \pm 27\%$
	Gaussian-RGB	0.66 ± 0.15	$81 \pm 5\%$	$97 \pm 2\%$	$38 \pm 24\%$
	Human-Eye	0.72 ± 0.02	$78 \pm 15\%$	$95 \pm 5\%$	$35 \pm 20\%$
Other Salivary ($N = 3$)	HSI	0.59 ± 0.14	$76 \pm 12\%$	$55 \pm 29\%$	$85 \pm 15\%$
	Autofluorescence	0.80 ± 0.14	$84 \pm 10\%$	$77 \pm 13\%$	$85 \pm 6\%$
	2-NBDG Dye	0.52 ± 0.08	$76 \pm 11\%$	$49 \pm 28\%$	$84 \pm 16\%$
	Proflavin Dye	0.62 ± 0.10	$77 \pm 11\%$	$60 \pm 30\%$	$78 \pm 22\%$
	Gaussian-RGB	0.45 ± 0.03	$70 \pm 10\%$	$62 \pm 31\%$	$57 \pm 43\%$
	Human-Eye	0.78 ± 0.15	$84 \pm 9\%$	$76 \pm 12\%$	$87 \pm 6\%$

4. Discussion

The results of this extensive study suggest that label-free HS-based imaging and autofluorescence does indeed outperform the two fluorescent dye-based imaging methods tested here for thyroid tumor detection, but not to a significant degree. Interestingly, we discovered that HSI-synthesized RGB multiplex imaging significantly outperforms all imaging methods tested for thyroid tumor detection, including HSI and autofluorescence ($p < 0.05$). For salivary tumor detection, HSI performs best in the parotid gland and autofluorescence in performs best in other salivary glands, but no difference was significant. As can be observed in Fig. 6(B), one main conclusion from this work is that with sufficiently large datasets, many different optical imaging modalities can be used to create deep learning algorithms for tumor detection with median AUC scores of 0.90 and upwards. This phenomenon can be seen specifically for the thyroid tumors combined cohort.

The experiments for the thyroid cohort and salivary gland cohort were processed separately in different ways because of vastly different numbers of tissue samples collected. The thyroid tumor cohort experiment was investigated in fully-independent testing patients because 200 tissue specimens from 76 patients were available. The salivary tumor cohort was comprised of only 16 tissues from 6 patients, so intra-patient experiments were performed using tumor-only and normal specimens for training and testing on tumor-normal margin tissues.

The hypothesis that HSI-based methods would outperform fluorescent dye-based methods was upheld in the thyroid tumor combined category, largely because it was supported in the PTC group ($N = 54$), which comprises 71% of cases. However, it was not supported for MTC, FTC, and poorly differentiated thyroid carcinoma groups (all p values were not significant). Additionally, the thyroid tumor detection results show that HSI-synthesized human-eye RGB multiplex imaging made from HSI statistically outperforms reflectance-based HSI. Despite differences in average AUC scores, the median values are equivalent around 0.95 (see Fig. 6). Exploring this phenomenon further in Fig. 8, it was demonstrated that only a few tissues differ between the HSI and HSI-synthesized RGB multiplex modalities. Moreover, the probability heat-maps from HSI seem to provide more consistent classification around regions of significant specular glare compared to the HSI-synthesized multiplex methods (Fig. 7). These results are consistent with a previous study from our group that was limited to only 11 thyroid patients, in which RGB composite images (AUC score of 0.95) also outperformed HSI (AUC score of 0.92) [24].

The purpose of the three-band HSI-synthesized multiplex images synthesized from HSI was to represent standard RGB imaging from a standard camera. However, these multiplex images are still constructed from HSI data. Additionally, there are differences in the spectral responses to the red channel component from 400-500 nm in standard RGB camera sensors. The impact of this red component value was studied, and no effect was observed in AUC score by altering these values for HSI-synthesized human-eye RGB multiplexing. To provide physical intuition for this conclusion, the grad-CAM method reveals that the most salient spectral features for correctly classifying normal thyroid tissues were from 570-700 nm, well above this range. Future studies are required to investigate if a standard RGB camera would indeed outperform HSI directly. Additionally, future work is needed to capture more thyroid tumor HSI data with higher spatial and spectral resolution HS cameras. It is possible that the spectral resolution of 5 nm in this LCTF spectral-scanning HS system was inadequate for this study.

5. Conclusion

In conclusion, we present an extensive study using 216 tissue samples from 82 patients to evaluate the performance of HSI for tumor detection of the thyroid and salivary glands. For comparison to HSI, the tissues were imaged with label-free autofluorescence and two fluorescent dyes, 2-NBDG and proflavin dye. Additionally, HSI-synthesized three-band multiplex images, representing the human-eye response and Gaussian RGBs, were synthesized from HSI. Several CNNs were developed for tumor detection that perform with median AUC scores of 0.90 and higher for all imaging modalities in combined thyroid tumors. Investigating each group specifically, our results suggest that HSI-synthesized human-eye RGB multiplexing can classify thyroid tumors significantly better than HSI. In salivary glands, label-free HSI and autofluorescence may offer the best performance for tumor detection. This study demonstrates that HSI could aid surgeons and pathologists in tumor detection in glands.

Funding

National Institutes of Health (CA156775, CA176684, CA204254); Cancer Prevention and Research Institute of Texas (RP190588).

Acknowledgements

The authors would like to thank the surgical pathology team at Emory University Hospital Midtown for their help in collecting fresh tissue specimens.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose. Informed consent was obtained from all patients in accordance with Emory IRB policies under the Head and Neck Satellite Tissue Bank protocol.

References

1. C. La Vecchia, M. Malvezzi, C. Bosetti, W. Garavello, P. Bertuccio, F. Levi, and E. Negri, "Thyroid cancer mortality and incidence: a global overview," *Int. J. Cancer* **136**(9), 2187–2195 (2015).
2. "Thyroid Cancer" (American Cancer Society, 2019), retrieved October 7, 2019, 2019, cancer.org/cancer/thyroid-cancer.
3. R. V. Lloyd, D. Buehler, and E. Khanafshar, "Papillary Thyroid Carcinoma Variants," *Head and Neck Pathol* **5**(1), 51–56 (2011).
4. M. Sobrinho-Simoes, C. Eloy, J. Magalhaes, C. Lobo, and T. Amaro, "Follicular thyroid carcinoma," *Mod. Pathol.* **24**(S2), S10–S18 (2011).
5. M. Roy, H. Chen, and R. S. Sippel, "Current Understanding and Management of Medullary Thyroid Cancer," *The Oncologist* **18**(10), 1093–1100 (2013).

6. B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Y. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, J. A. Sosa, D. L. Steward, R. M. Tuttle, and L. Wartofsky, "2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer," *Thyroid* **26**(1), 1–133 (2016).
7. R. Mallick, T. M. Stevens, T. S. Winokur, A. Asban, T. N. Wang, B. M. Lindeman, J. R. Porterfield, and H. Chen, "Is Frozen-Section Analysis During Thyroid Operation Useful in the Era of Molecular Testing?" *J. Am. Coll. Surgeons* **228**(4), 474–479 (2019).
8. R. W. Berg, T. W. Yen, D. B. Evans, B. Hunt, F. A. Quiroz, S. D. Wilson, and T. S. Wang, "Analysis of an institutional protocol for thyroid lobectomy: Utility of routine intraoperative frozen section and expedited (overnight) pathology," *Surgery* **159**(2), 512–517 (2016).
9. H. Boukheris, R. E. Curtis, C. E. Land, and G. M. Dores, "Incidence of Carcinoma of the Major Salivary Glands According to the WHO Classification, 1992 to 2006: A Population-Based Study in the United States," *Cancer Epidemiol., Biomarkers Prev.* **18**(11), 2899–2906 (2009).
10. V. S. H. To, J. Y. W. Chan, R. K. Y. Tsang, and W. I. Wei, "Review of Salivary Gland Neoplasms," *ISRN Otolaryngology* **2012**, 6 (2012).
11. S. Bobati, B. Patil, and V. Dombale, "Histopathological study of salivary gland tumors," *Int. J. Oral Maxillofac. Pathol.* **21**(1), 46–50 (2017).
12. K. Chatura, "Polymorphous low grade adenocarcinoma," *Int. J. Oral Maxillofac. Pathol.* **19**(1), 77–82 (2015).
13. M. Suzuki, Y. Nakaegawa, T. Kobayashi, T. Kawase, T. Matsuzuka, and S. Murono, "The role of frozen section biopsy for parotid gland tumour with benign fine-needle aspiration cytology," *J. Laryngol. Otol.* **133**(03), 227–229 (2019).
14. D. C. Howlett, E. Skelton, and A. B. Moody, "Establishing an accurate diagnosis of a parotid lump: evaluation of the current biopsy methods – fine needle aspiration cytology, ultrasound-guided core biopsy, and intraoperative frozen section," *Br. J. Oral. Maxillofac. Surg.* **53**(7), 580–583 (2015).
15. S. Grasl, L. Kadletz, S. Janik, A. Riedl, B. Erlacher, M. Formanek, M. C. Grasl, and B. M. Erovic, "Fine-needle aspiration cytology and intraoperative frozen section in parotid gland tumour surgery: A retrospective multicenter analysis of 417 cases," *Clin Otolaryngol* **44**(3), 461–465 (2019).
16. M. Halicek, H. Fabelo, S. Ortega, G. M. Callico, and B. Fei, "In-Vivo and Ex-Vivo Tissue Analysis through Hyperspectral Imaging Techniques: Revealing the Invisible Features of Cancer," *Cancers* **11**(6), 756 (2019).
17. J. Shapey, Y. Xie, E. Nabavi, R. Bradford, S. R. Saeed, S. Ourselin, and T. Vercauteren, "Intraoperative multispectral and hyperspectral label-free imaging: A systematic review of in vivo clinical studies," *J. Biophotonics* **12**(9), e201800455 (2019).
18. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
19. H. Fabelo, M. Halicek, S. Ortega, M. Shahedi, A. Szolna, J. Piñeiro, C. Sosa, A. O'Shanahan, S. Bisshopp, C. Espino, M. Márquez, M. Hernández, D. Carrera, J. Morera, G. Callico, R. Sarmiento, and B. Fei, "Deep Learning-Based Framework for In Vivo Identification of Glioblastoma Tumor using Hyperspectral Images of Human Brain," *Sensors* **19**(4), 920 (2019).
20. H. Fabelo, S. Ortega, A. Szolna, D. Bulters, J. F. Pineiro, S. Kabwama, A. J-O'Shanahan, H. Bulstrode, S. Bisshopp, B. R. Kiran, D. Ravi, R. Lazcano, D. Madronal, C. Sosa, C. Espino, M. Marquez, M. De La Luz Plaza, R. Camacho, D. Carrera, M. Hernandez, G. M. Callico, J. Morera Molina, B. Stanculescu, G.-Z. Yang, R. Salvador Perea, E. Juarez, C. Sanz, and R. Sarmiento, "In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection," *IEEE Access* **7**, 39098–39116 (2019).
21. E. J. M. Baltussen, E. N. D. Kok, S. G. Brouwer de Koning, J. Sanders, A. G. J. Aalbers, N. F. M. Kok, G. L. Beets, C. C. Flohil, S. C. Bruin, K. F. D. Kuhlmann, H. J. C. M. Sterenborg, and T. J. M. Ruers, "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," *J. Biomed. Opt.* **24**(01), 1 (2019).
22. S. J. Leavesley, M. Walters, C. Lopez, T. Baker, P. F. Favreau, T. C. Rich, P. F. Rider, and C. W. Boudreaux, "Hyperspectral imaging fluorescence excitation scanning for colon cancer detection," *J. Biomed. Opt.* **21**(10), 104003 (2016).
23. B. Jansen-Winkel, N. Holfert, H. Kohler, Y. Moulla, J. P. Takoh, S. M. Rabe, M. Mehdorn, M. Barberio, C. Chalopin, T. Neumuth, and I. Gockel, "Determination of the transection margin during colorectal resection with hyperspectral imaging (HSI)," *Int J Colorectal Dis* **34**(4), 731–739 (2019).
24. G. Lu, J. V. Little, X. Wang, H. Zhang, M. R. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging," *Clin. Cancer Res.* **23**(18), 5426–5436 (2017).
25. M. Halicek, J. V. Little, X. Wang, M. Patel, C. C. Griffith, A. Y. Chen, and B. Fei, "Tumor margin classification of head and neck cancer using hyperspectral imaging and convolutional neural networks," *Proc SPIE Int Soc Opt Eng* **10576**, 1–11 (2018).
26. M. Halicek, G. Lu, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *J. Biomed. Opt.* **22**(6), 060503 (2017).
27. M. Halicek, J. V. Little, X. Wang, A. Y. Chen, and B. Fei, "Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks," *J. Biomed. Opt.* **24**(03), 1–9 (2019).

28. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *J. Biomed. Opt.* **22**(08), 1–7 (2017).
29. M. Barberio, M. Maktabi, I. Gockel, N. Rayes, B. Jansen-Winkel, H. Köhler, S. M. Rabe, L. Seidemann, J. P. Takoh, M. Diana, T. Neumuth, and C. Chalopin, "Hyperspectral based discrimination of thyroid and parathyroid during surgery," *Curr. Dir. Biomed. Eng.* **4**(1), 399–402 (2018).
30. E. L. Wisotzky, F. C. Uecker, P. Arens, S. Dommerich, A. Hilsmann, and P. Eisert, "Intraoperative hyperspectral determination of human tissue properties," *J. Biomed. Opt.* **23**(09), 1–8 (2018).
31. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral Imaging of Head and Neck Squamous Cell Carcinoma for Cancer Margin Detection in Surgical Specimens from 102 Patients Using Deep Learning," *Cancers* **11**(9), 1367 (2019).
32. D. B. Judd, "Report of US Secretariat Committee on Colorimetry and Artificial Daylight," Proceedings of the Twelfth Session of the CIE **1**, 11 (1951).
33. J. J. Vos, "Colorimetric and photometric properties of a 2° fundamental observer," *Color Res. Appl.* **3**(3), 125–128 (1978).
34. E. Berra, S. Gibson-Poole, A. MacArthur, R. Gaulton, and A. Hamilton, "Estimation of the spectral sensitivity functions of un-modified and modified commercial off-the-shelf digital cameras to enable their use as a multispectral imaging system for UAVs," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XL-1/W4**, 207–214 (2015).
35. C. Maurer and D. Wueller, *Measuring the spectral response with a set of interference filters*, IS&T/SPIE Electronic Imaging (SPIE, 2009), Vol. 7250.
36. A. Hellebust, K. Rosbach, J. K. Wu, J. Nguyen, A. Gillenwater, N. Vigneswaran, and R. Richards-Kortum, "Vital-dye-enhanced multimodal imaging of neoplastic progression in a mouse model of oral carcinogenesis," *J. Biomed. Opt.* **18**(12), 126017 (2013).
37. N. Thekkekk, T. Muldoon, A. D. Polydorides, D. M. Maru, N. Harpaz, M. T. Harris, W. Hofstetter, S. P. Hiotis, S. A. Kim, A. J. Ky, S. Anandasabapathy, and R. Richards-Kortum, "Vital-dye enhanced fluorescence imaging of GI mucosa: metaplasia, neoplasia, inflammation," *Gastrointestinal Endoscopy* **75**(4), 877–887 (2012).
38. M. Halicek, M. Shahedi, J. V. Little, A. Y. Chen, L. L. Myers, B. D. Sumer, and B. Fei, "Head and Neck Cancer Detection in Digitized Whole-Slide Histology Using Convolutional Neural Networks," *Sci. Rep.* **9**(1), 14043 (2019).
39. M. Halicek, J. V. Little, X. Wang, Z. G. Chen, M. Patel, C. C. Griffith, M. W. El-Deiry, N. F. Saba, A. Y. Chen, and B. Fei, "Deformable registration of histological cancer margins to gross hyperspectral images using demons," *Proc SPIE Int Soc Opt Eng* 10581, 1–8 (2018).
40. M. Halicek, H. Fabelo, S. Ortega, J. V. Little, X. Wang, A. Y. Chen, G. M. Callico, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral imaging for head and neck cancer detection: specular glare and variance of the tumor margin in surgical specimens," *J. Med. Imag.* **6**(03), 1–11 (2019).
41. S. I. Christian Szegedy and Vincent Vanhoucke, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," arXiv: Computational Research Repository **1602.07261** (2016).
42. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," arXiv: Computational Research Repository **1709.01507** (2018).
43. M. Abadi, A. Agarwal, P. Barham, and E. A. Eugene Brevdo, "TensorFlow: Large-scale machine learning on heterogeneous systems," <https://www.tensorflow.org> (2015).
44. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proc IEEE Int Conf Comput Vis*, 618–626 (2017).