PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Thyroid carcinoma detection on whole histologic slides using hyperspectral imaging and deep learning

Tran, Minh Ha, Ma, Ling, Litter, James, Chen, Amy, Fei, Baowei

Minh Ha Tran, Ling Ma, James V. Litter, Amy Y. Chen, Baowei Fei, "Thyroid carcinoma detection on whole histologic slides using hyperspectral imaging and deep learning," Proc. SPIE 12039, Medical Imaging 2022: Digital and Computational Pathology, 120390H (4 April 2022); doi: 10.1117/12.2612963

SPIE.

Event: SPIE Medical Imaging, 2022, San Diego, California, United States

Thyroid Carcinoma Detection on Whole Histologic Slides Using Hyperspectral Imaging and Deep Learning

Minh Ha Tran[a,b], Ling Ma[a,b], James V. Litter[c], Amy Y. Chen[d], and Baowei Fei[a,b,e]*

[a] Univ. of Texas at Dallas, Dept. of Bioengineering, Richardson, TX
[b] Univ. of Texas at Dallas, Center for Imaging and Surgical Innovation, Richardson, TX
[c] Emory Univ. School of Medicine, Dept. of Pathology and Laboratory Medicine, Atlanta, GA
[d] Emory Univ. School of Medicine, Dept. of Otolaryngology, Atlanta, GA
[e] Univ. of Texas Southwestern Medical Center, Department of Radiology, Dallas, TX

*Corresponding author: bfei@utdallas.edu, Website: https://fei-lab.org

ABSTRACT

Hyperspectral imaging (HSI), a non-invasive imaging modality, has been successfully used in many different biological and medical applications. One such application is in the field of oncology, where hyperspectral imaging is being used on histologic samples. This study compares the performances of different image classifiers using different imaging modalities as training data. From a database of 33 fixed tissues from head and neck patients with follicular thyroid carcinoma, we produced three different datasets: an RGB image dataset that was acquired from a whole slide image scanner, a hyperspectral (HS) dataset that was acquired with a compact hyperspectral camera, and an HS-synthesized RGB image dataset. Three separate deep learning classifiers were trained using the three datasets. We show that the deep learning classifier trained on HSI data has an area under the receiver operator characteristic curve (AUC-ROC) of 0.966, higher than that of the classifiers trained on RGB and HSI-synthesized RGB data. This study demonstrates that hyperspectral images improve the performance of cancer classification on whole histologic slides. Hyperspectral imaging and deep learning provide an automatic tool for thyroid cancer detection on whole histologic slides.

Keywords: Hyperspectral imaging, whole-slide histologic imaging, thyroid cancer, deep learning, digital pathology

INTRODUCTION

Head and neck cancer (HNC) refers to a broad class of cancer types that occur within the oropharynx, the tongue, the nasal cavity, the paranasal, the thyroid, and the larynx. It is the seventh most common type of cancer worldwide with 890,000 new cases reported annually [1]. HNC cases are on the rise globally, due to the increase in alcohol use, smoking, and HPV-associated oropharyngeal cancer [1, 2]. The majority of HNC are head and neck squamous cell carcinoma (HNSCC) [3]. Surgery remains the primary curative method of squamous cell HNC, often in combination with radiotherapy and chemotherapy [1, 2, 4]. During surgery, the surgeon determines the tumor margin and tried to remove the entire tumor. A tumor margin of as close as 1 mm has been shown to significantly increase the remission rate [5]. A more acceptable surgical margin ranges from 5 mm to 20 mm [6]. On the other hand, as much non-tumor tissue should be preserved as possible, so that normal functions are minimally disrupted. Intraoperative pathology consultants (IPC) are known to produce inconsistent results when it comes to the tumor margin [5, 6]. Errors come from sampling errors and inconsistent definitions of "safe" surgical margins [5, 7-9]. Estimation of true negative margins varies but ranges from 50 to 80% [8, 10]. Furthermore, IPC can be a time consuming procedure, taking 25-45 minutes per sample [6]. Computer-aided diagnosis (CAD) may help reduce the time during IPC. In a CAD-based workflow, the whole slide image (WSI) is processed, and a cancer tumor margin is produced.

CAD algorithms often use machine learning algorithms such as support vector machines (SVM) [11-14] and convolutional neural networks (CNNs). SVM requires hand-crafted features such as color and textures for quality classification. If an image has too many correlated features, dimensionality reduction methods such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE) can be applied to produce more valuable features

Medical Imaging 2022: Digital and Computational Pathology, edited by John E. Tomaszewski, Aaron D. Ward, Proc. of SPIE Vol. 12039, 120390H · © 2022 SPIE
1605-7422 · doi: 10.1117/12.2612963

from existing ones [11, 12]. On the other hand, CNNs can work with highly non-linear inputs such as images. As such, CNN-based CAD do not rely as heavily on pre-processing as SVM-based CAD workflows. The popularity of CNN-based CAD also comes from pretrained networks, which are networks that have weights already set from training on millions of images prior. Many pretrained networks architectures such as Inception, LeNet, ResNet, and VGG are publicly available [15, 16]. For HNC, Halicek et al. [16] trained a CNN-based classifier using microscopic H&E-stained images as inputs. The classifier achieves a patch-based area under the receiving operating characteristic curve (AUC-ROC) of 0.916 for all SCC patients. Gupta et al. [17] trained a CNN-based classifier on microscopic H&E-stained slides to grade severity of dysplasia in the oral squamous region. Their classifier achieved 89.3% grading accuracy. Chu et al. [18] reported a review of various machine learning and deep learning-based classifiers used in oral SCC classification and detection.

The standard inputs for microscopic images have been hematoxylin and eosin (H&E) stained slides, captured by a regular RGB camera [15, 16, 19, 20]. Halicek et al. [15] showed that using autofluorescence images produce better classifier compared to using images of proflavine dye and using RGB images of tissues. Fakurnejad et al. [21] used a fluorescing anti-epidermal growth factor receptor contrast agent to identify high-grade dysplasia. Hyperspectral imaging (HSI) is a non-invasive, non-dye method that captures the reflectance spectra of the tissue in a very small increment (*e.g.*, <5 nm). For microscopic images, HSI inputs showed improvements in identification of cancer over regular RGB images. Ortega et al. [22] compared the performance of HSI and RGB inputs over microscopic images of breast cancer. They found an AUC-ROC of 0.90 for HSI inputs and suggested that uses of HSI inputs outperform those that use RGB inputs. More recently, in the HNC domain, Ma et al. [23] compared the performances of two classifiers: one uses HSI of H&E-stained microscopic images as input, the other one uses synthesized RGB images as input. They found that HSI inputs trained a better classifier. Still, there is a lack of concrete literature that compares the performance between HSI, RGB, and synthesized RGB as inputs across a wide variety of microscopic imaging classification tasks.

Our paper seeks to innovate in both image acquisition and classification aspects. For image acquisitions, we propose a new method of whole slide microscopic image acquisition for hyperspectral imaging. Ortega et al. [24] produced a small hyperspectral image database of microscopic brain tissue. Liao et al. [25] captured tetra-pixel whole slide microscopic images in hyperspectral domain using slit arrays. In classification, we propose a tumor margin determination algorithm that use the entire microscopic whole slide image in the HSI domain as input. Ma et al. [6] proposed a tumor margin assessment algorithm using hyperspectral images as input. Ma et al. [26] also proposed a patch-based HNSCC detection algorithm using hyperspectral images as input. To the best of our knowledge, this study is the first work to investigate HNC on the whole slide level using hyperspectral images as inputs.

METHODS

1.1 Histologic Slides from Thyroid Cancer Patients

The histologic slides and their acquisitions were described in our previous studies [6, 23, 27]. Briefly, 33 head and neck histologic slides of the thyroid region taken from 23 cancer patients were used. All histologic slides from our collection are of follicular type; this is in contrast with other types of HNC which are majority squamous cells carcinomas [28]. The histologic slides were fixed and underwent hematoxylin and eosin (H&E) staining. Each histologic slides were imaged in two different methods, from which three datasets were created: the hyperspectral image dataset, the conventional RGB image dataset, and the HSI-synthesized RGB image dataset. From each dataset, we train a different classifier. The following sections detail the creation of each dataset.

1.2 Conventional RGB Image Dataset

All slides were digitized at 40× magnification using a whole slide scanner that produces RGB digital slides with spatial resolution of 0.23 μm [29]. From these whole slide images, we produced non-overlapping patches with the size of 400×400×3, which we then downsampled to 200×200×3. The cancerous regions in the digitized images were manually annotated by a board-certified pathologist. This annotation serves as our only ground truth for both RGB image dataset and HSI image dataset. We trained and fine-tuned the "RGB classifier" on this dataset.

1.3 Hyperspectral Image Dataset

We used a custom HSI system to acquire whole slide hyperspectral images of the tissue slides. The system includes an optical microscope, a compact hyperspectral snapshot camera, and a high-precision motorized stage as reported in [6]. For

whole slide images, we developed a program in LabView that synchronizes stage movement with camera acquisition. The program divides the whole slide into rectangular grid, and the camera goes through each segment of the grid to capture the image. Afterward, a phase correlation algorithm was used to account for any overlap between each image segment. All images were acquired at an objective of 20×. The image's raster size was 2000 pixel × 2000 pixel × 84 bands, acquired within the wavelength range from 467 nm to 721 nm. The raw hyperspectral data (I_{raw}) were normalized at each wavelength by using the inherent dark current (I_{dark}) image and the white reference image (I_{white}) with the following equation [30]:

$$I_{norm}(\lambda, s) = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}}$$

From the whole slide image, we produce non-overlapping image patches of size 400×400×84, which we then down-sample to 200×200×84 to train our neural network. Because stitching all hyperspectral images together would produce a WSI with very large raster size (100,000 × 100,000 × 84), we produced patches directly from each individual image. Figure 1 shows the complete acquisition and patch productions workflow. We trained and finetuned the "HSI classifier" on this dataset. Because we used two different systems to produce the RGB dataset and the hyperspectral image dataset, the patches we produced in each dataset are not the same. The number of patches used in the train/validation/test split for RGB dataset is 15,287/7145/6098, whereas for the HSI dataset they are 15,168/7339/6301. We trained and fine-tuned the "HSI classifier" on this dataset.

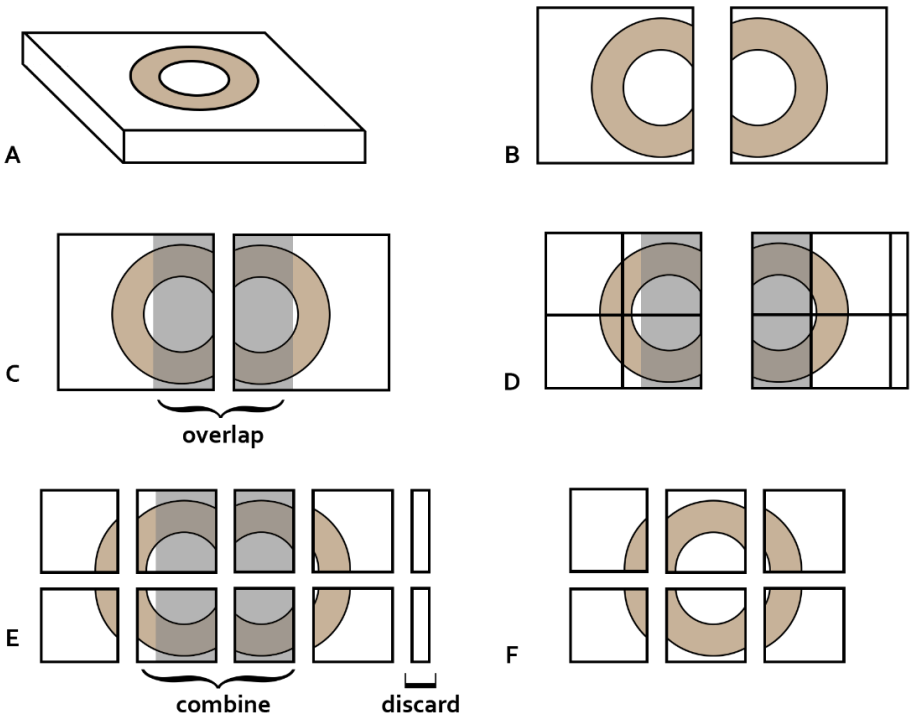

Figure 1. The workflow for our whole slide patch-based acquisition algorithm. With this method, patches for the entire slide can be produced without directly producing the whole slide image. (A) The sample in acquisition. (B) The individual images captured by the camera. (C) Overlapping regions are determined using phase correlation algorithm. (D) and (E) patches coordinates were created using overlapping regions determined in the previous part. (F) overlapping patches are combined by averaging the values of the pixels.

1.4 HSI-synthesized-RGB Image Dataset

Using techniques previously used in [23], we synthesized RGB images from hyperspectral images. The algorithm uses the spectral response of human perception to colors to generate pseudo-RGB images. We trained and finetuned the "HSI-

synthesized-RGB classifier" on this dataset. Because this dataset is synthesized directly from the HSI dataset, there is a one-to-one match between the original HSI and synthesized-RGB datasets.

1.5 Convolutional Neural Network

We trained three convolutional neural networks based on the VGG-19 architecture [31]. VGG-19 consists of two sections: a convolutional network (16 layers) and a feedforward network (3 layers). We set the initial weights using a pretrained VGG-19 network that has been trained on >1 million RGB images. All networks were modified to output only one binary class. To fit the network to our specific application, we applied several modifications. Spatially, the original pretrained VGG-19 network accepts 224×224×3 images as inputs. A workaround for this is an adaptive pooling layer prior to the flatten layer, so that the input into the feedforward dense network always stay the same. VGG-19 network originally has 1000 class outputs; we changed this to a single binary output range from 0 to 1. Because the conventional RGB image dataset and the HSI-synthesized-RGB image dataset already have input dimension of 200 pixels×200 pixels×3 channels, no other modification was needed to train the "RGB classifier" and the "HSI-synthesized-RGB classifier." For the "HSI classifier," we proposed a modified architecture that preserves the advantages of pretrained networks and accepts 200 pixels×200 pixels×84 channels images as inputs. The 84-channel image is divided into 27 sequential groups of 200 ×200 ×3-channel images. Each group corresponds to a wavelength section. In the forwarding step, each 3-channel image will be fed sequentially into a VGG-19 network, producing 27 sequential values. We add a perceptron layer, which is a single feedforward layer that combines 27 inputs and produce a final binary output from 0 to 1. In the backward pass, the gradient is cumulative across all groups to move the network parameter. Figure 2 shows the modified architecture. Table 1 shows the output size of each layer/block. To regularize the network, dropout value of 0.5 was employed in the feedforward network.

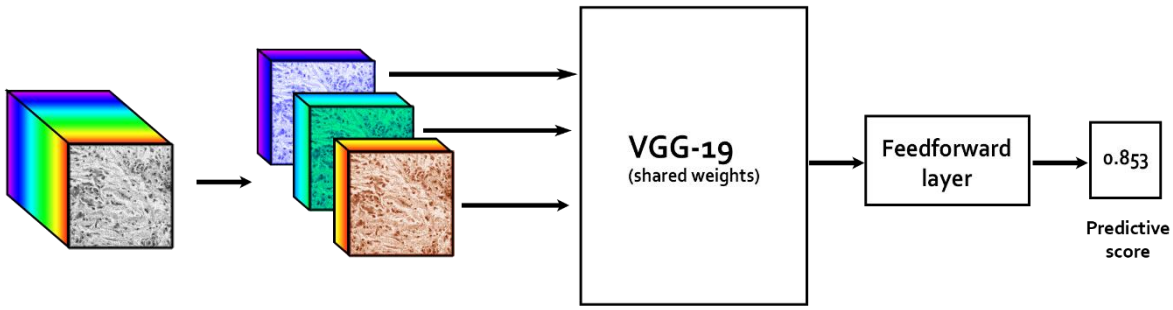

Figure 2. Network architecture for the HSI classifier network. The HSI is sectioned into 27 3-channels sub-images. Each sub-image is fed into VGG-19 individually. The final feedforward layer combines all outputs to provide a cancer predictive probability ranging from 0 to 1.

Table 1. CNN architecture for the HSI classifier network.

Layer/Block	Output size (B×H×W×C)
Input	200×200×84
Divide	27×200×200×3
Conv2D+ReLU	27×200×200×64
Conv2D+ReLU	27×200×200×64
MaxPool	27×100×100×64
Conv2D+ReLU	27×100×100×128

Conv2D+ReLU	27×100×100×128
MaxPool	27×50×50×128
Conv2D+ReLU	27×50×50×256
Conv2D+ReLU	27×50×50×256
MaxPool	27×25×25×256
Conv2D+ReLU	27×25×25×512
Conv2D+ReLU	27×25×25×512
Conv2D+ReLU	27×25×25×512
Conv2D+ReLU	27×25×25×512
MaxPool	27×12×12×512
Conv2D+ReLU	27×12×12×512
Conv2D+ReLU	27×12×12×512
Conv2D+ReLU	27×12×12×512
Conv2D+ReLU	27×12×12×512
AdaptivePool	27×7×7×512
Flatten	27×25,088
Dense	27×4096
Dense	27×1
Dense (“Sigmoid”)	1

1.6 Metrics and Network Training

The network was implemented in PyTorch and trained on Titan XP NVIDIA 16GB GPU. The optimization algorithm is mini-batch gradient descent. Each training dataset consists of the same 23 WSI, each of which contain entirely tumor or normal tissue. Each validation dataset consists of the same 5 WSI, each of which contain a tumor margin. Each test dataset consists of the same 5 WSI, each of which contain a tumor margin. Early stopping was used, which means training will stop after the validation metrics do not improve after a certain number of epochs.

The metric we use is the F-1 score, which is the average of the precision and recall. Precision and recall are calculated from the true positive (TP), true negative (TN), and false positive (FP) rates. Precision is the ratio of true positive values over all positive predicted values, while recall (sensitivity) is the ratio of true positive values over all actual positive values [32].

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1\ score = \frac{Precision + Recall}{2}$$

We fine-tuned each network by varying the learning rate and the batch size, keeping all other parameters constant. The network with the best F-1 score on each validation dataset will be selected. We found the optimal batch size and learning rate for training both conventional RGB images and HSI-synthesized RGB images to be 64 and 5×10^{-3}. The optimal batch size and learning rate for training HSI images is 4 and 1×10^{-4}.

RESULT

1.7 Training and Validation Result

We evaluate the validation accuracy by setting the classification probability cutoff for tumor at 50 percent. All classifiers achieved good training fit with an F-1 score of > 0.99 and a classification accuracy of > 99%. Table 2 shows the validation F-1 score and the AUC-ROC for the best classifier within each image set. Even though the selection of best performing classifier is based on F-1 score, we observed that F-1 score and classification accuracy during training are highly correlated. If we count all images in the validation dataset, all three classifiers have similar F-1 scores, and the HSI classifier had a slightly lower AUC-ROC. However, there is a large disparity in classification quality from patient to patient within the validation dataset. Most notably, in Patient V4, all classifiers underperform compared to that in other patients. If we remove Patient V4, the F-1 score in conventional RGB, HSI-synthesized RGB, and HSI validation dataset will become 0.8589, 0.8644, and 0.8677, respectively. Visual inspection of Patient V4 WSI shows that the ground truth includes a large region of thyroid follicle, which all classifiers predict as normal.

Table 2. Performance of each classifier on the validation dataset. The weighted average takes the number of patches of each patient into consideration.

Patient	Conventional RGB		HSI-synthesized RGB		HSI	
	F-1	AUC-ROC	F-1	AUC-ROC	F-1	AUC-ROC
V1	0.9102	0.9844	0.8458	0.9695	0.9116	0.9798
V2	0.8391	0.9587	0.8343	0.9592	0.8817	0.9758
V3	0.8122	0.9187	0.8464	0.9338	0.8152	0.9109
V4	0.6689	0.8945	0.7646	0.9335	0.6813	0.8332
V5	0.9154	0.9573	0.9088	0.9371	0.8931	0.9374
Weighted Average	0.8427	0.9472	0.8472	0.9497	0.8460	0.9380
Unweighted Average	0.8314	0.9435	0.8412	0.9471	0.8382	0.9292

1.8 Testing Result

Conventional RGB classifier achieves an AUC-ROC of 0.9333 (Table 3). HSI-synthesized RGB classifier achieves slightly better AUC-ROC at 0.9471. HSI classifier achieves the highest AUC-ROC at 0.9660. Figure 3 shows the ROC curve of each classifier. If we set the tumor cutoff at 50%, HSI classifier will achieve the highest overall F-1 score (0.8672) compared to those of HSI-synthesized RGB (0.8427) and conventional RGB classifier (0.8310). In terms of F-1 score, all classifiers performed the worst on Patient T5. Visual inspection of T5 WSI (Figure 4) shows a negative region of thyroid gland that is predicted by all classifiers as tumor with high probability. Another notable WSI is T4, which achieve high F-1 score in HSI-synthesized RGB and HSI classifier but low F-1 score in conventional RGB. Visual inspection of T4 (Figure 5) WSI shows that the HSI classifier produce margins that stay closest to the ground truth, whereas conventional RGB classifier under predicts the cancer margin. This speaks to the robustness of the HSI classifier in consistently making predictions that are close to the ground truth compared to conventional RGB image classifier. Also note that the thyroid cancer RGB dataset has been used prior in training an Inception-V4 network and achieved a test AUC-ROC of 0.954 and an F-1 score of 0.894 [29] where a much larger dataset of 153 WSI were used in the training.

1.9 Visualizing HSI Network

Due to the architecture of the HSI network, all weights in the convolutional and the feedforward network are shared. The final layer is a feedforward layer that calculates the final prediction score based on the output from each wavelength range. By extracting the weights of this final layer, we can visualize how much importance the network assigns to each wavelength in making the final prediction. Figure 6 shows that the network gives the most weights of prediction to the

wavelength range 500-650 nm and the least toward the wavelength range 650-720nm. Figure 6 also shows representative images at different wavelengths. At higher wavelengths, the image is noisier and potentially is less valuable for the image classifier.

Table 3. Performance of each classifier on the test dataset. The weighted average takes the number of patches of each patient into consideration.

Patient	Conventional RGB		HSI-synthesized RGB		HSI	
	F-1	AUC-ROC	F-1	AUC-ROC	F-1	AUC-ROC
T1	0.8461	0.9648	0.8207	0.9623	0.8222	0.9605
T2	0.9569	0.9923	0.9428	0.9907	0.9277	0.9869
T3	0.9444	0.9831	0.9444	0.9886	0.9559	0.9912
T4	0.7079	0.8860	0.8273	0.9334	0.8923	0.9755
T5	0.6978	0.9391	0.6392	0.9130	0.6791	0.9418
Weighted Average	0.8310	0.9333	0.8427	0.9471	0.8672	0.9660
Unweighted Average	0.8307	0.9498	0.8362	0.9559	0.8574	0.9703

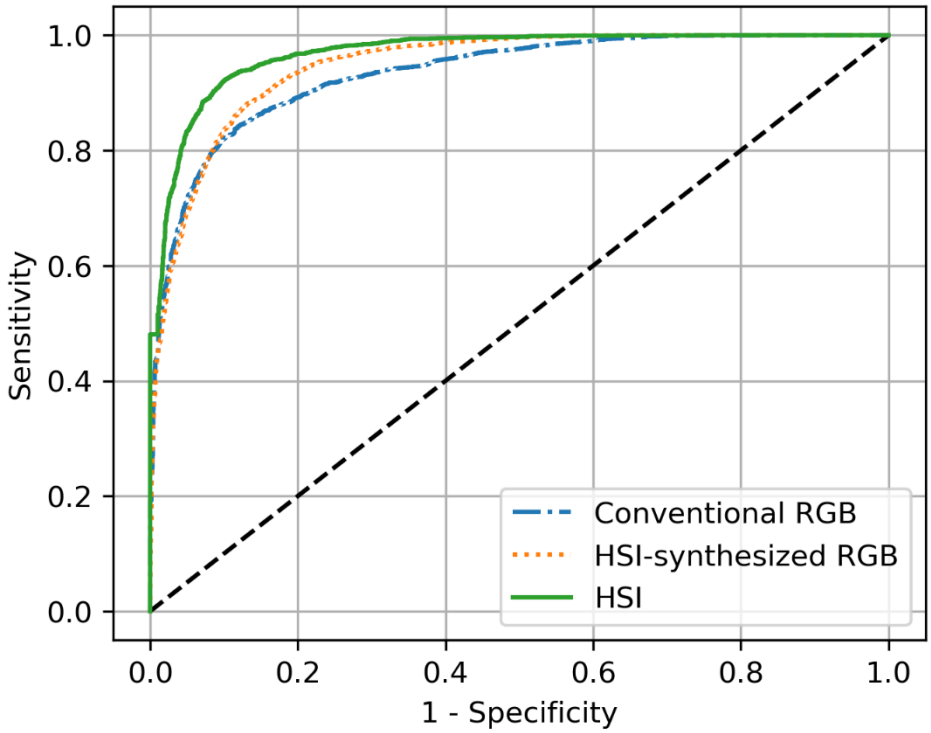

Figure 3. Receiving operating characteristic curves for three classifiers on the test dataset.

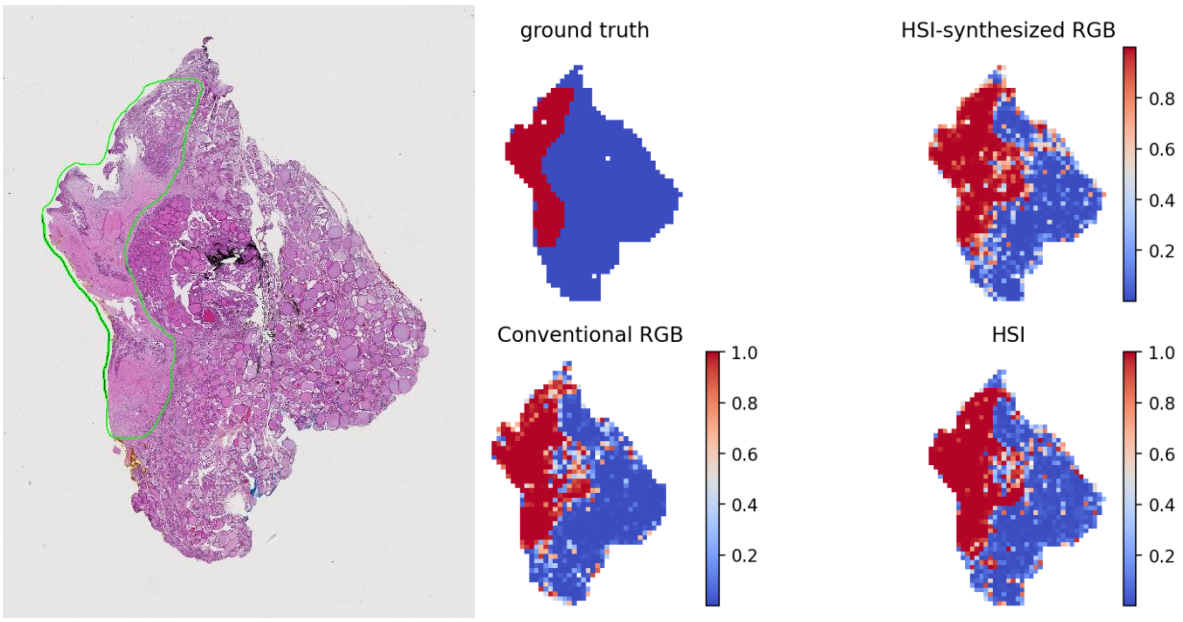

Figure 4. Left. WSI of patient T5 using whole slide RGB scanner. Tumor margin is outlined by a pathologist in green. Right. Cancer probability predictions made on patient T5 on a scale of 0 (normal, blue) to 1 (tumor, red).

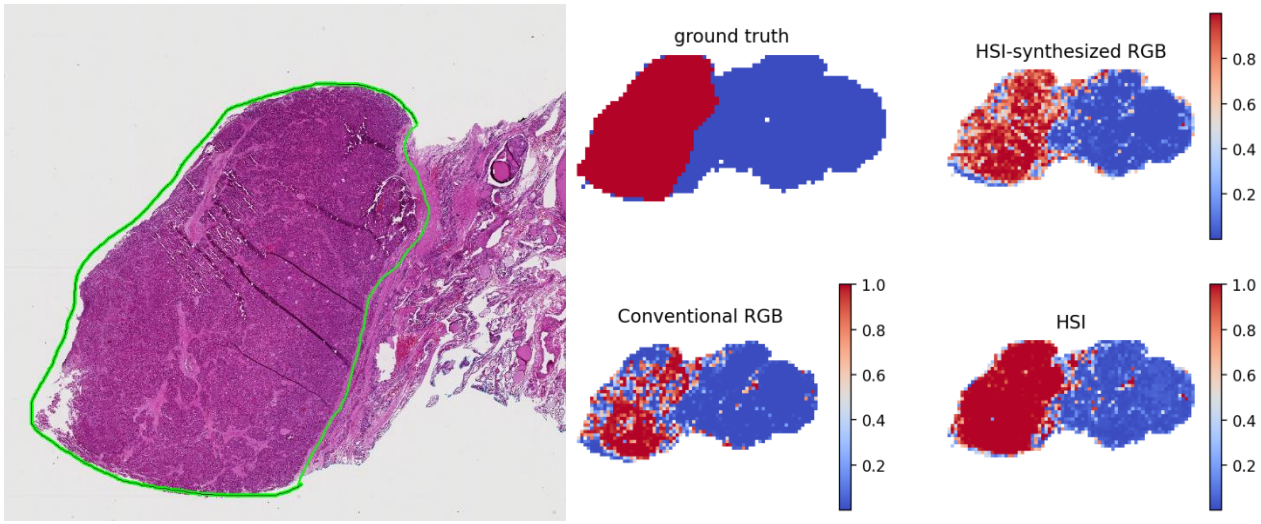

Figure 5. Left. WSI of patient T4 using whole slide RGB scanner. Tumor margin is outlined by a pathologist in green. Right. Cancer probability predictions made on patient T4 on a scale of 0 (normal, blue) to 1 (tumor, red).

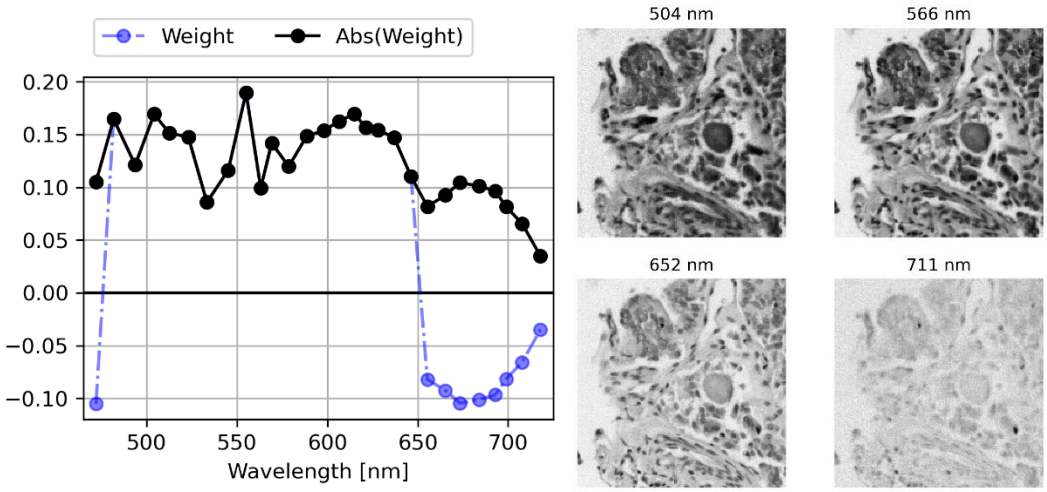

Figure 6. Left. Weight assigned by the HSI classifier to each wavelength. Right. Sample images from different wavelengths. At higher wavelengths, the image tends to get noisier.

CONCLUSIONS AND DISCUSSIONS

We demonstrate the acquisition of a hyperspectral histology whole slide image database (33 whole slide images). We also demonstrate the use of hyperspectral imaging to identify cancer margin on whole slide histology images. By comparing the performance of the HSI network with an equivalent network trained on RGB images, we demonstrate the improvements of hyperspectral imaging data over other alternatives. The study is necessary to validate the role of hyperspectral imaging in improving classification results.

One limitation of the study is the difference in the number of patches in the training, validation, and test dataset between the RGB and the HSI image dataset. This difference makes it hard to use statistical methods such as McNemar's test to directly compare the classifiers. However, we believe that the differences in the number of patches do not contribute much to the differences in network performance. If we look at the test performance in patient T4 and T5, we see that these patients contribute the most differences in test performance. We believe that even if we match the patches between HSI and RGB dataset, the classification margin would stay the same, and so would the classification result.

Another topic worthy of discussion is the difference between validation and test performance. In validation data, all three classifiers achieve a similar F-1 score, with the HSI classifier having the lowest AUC-ROC. Whereas in the test data, HSI classifier achieve both the highest F-1 score and AUC-ROC. A potential explanation for this is because of the variations between patients. However, if we look at the patient-by-patient performances in both the test and validation data, we see that the HSI classifier performs the best in situations where all three classifiers perform badly. This shows HSI classifier performs more consistently than RGB image classifier. We believe that the improvement in consistency comes to the unique design of the network, which utilizes an ensemble learning layer to combine outputs from all wavelengths. Further work needs to be done in studying and improving the performance of hyperspectral image classifiers.

DISCLOSURES

This work has not been submitted for presentation or publication elsewhere. The authors have no potential conflicts of interest to disclose. Informed consent was obtained from all patients in accordance with Emory Institutional Review Board policies under the Head and Neck Satellite Tissue Bank (HNSB, IRB00003208) protocol.

ACKNOWLEDGEMENTS

This research was supported in part by the U.S. National Institutes of Health (NIH) grants (R01CA156775, R01CA204254, R01HL140325, and R21CA231911), by the Cancer Prevention and Research Institute of Texas (CPRIT) grant RP190588, and by the Eugene McDermott Graduate Fellowship 202009 at the University of Texas at Dallas.

REFERENCES

[1] L. Q. Chow, "Head and neck cancer," *New England Journal of Medicine,* vol. 382, no. 1, pp. 60-72, 2020.

[2] S. Marur and A. A. Forastiere, "Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment," in *Mayo Clinic Proceedings*, 2016, vol. 91, no. 3: Elsevier, pp. 386-396.

[3] A. Argiris, M. V. Karamouzis, D. Raben, and R. L. Ferris, "Head and neck cancer," *The Lancet,* vol. 371, no. 9625, pp. 1695-1709, 2008.

[4] R. Sullivan *et al.*, "Global cancer surgery: delivering safe, affordable, and timely cancer surgery," *The lancet oncology,* vol. 16, no. 11, pp. 1193-1224, 2015.

[5] K. K. Tasche, M. R. Buchakjian, N. A. Pagedar, and S. M. Sperry, "Definition of "close margin" in oral cancer surgery and association of margin distance with local recurrence rate," *JAMA Otolaryngology–Head & Neck Surgery,* vol. 143, no. 12, pp. 1166-1172, 2017.

[6] L. Ma *et al.*, "Pixel-level tumor margin assessment of surgical specimen in hyperspectral imaging and deep learning classification," in *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, 2021, vol. 11598: International Society for Optics and Photonics, p. 1159811.

[7] K. T. Robbins *et al.*, "Surgical margins in head and neck cancer: Intra-and postoperative considerations," *Auris Nasus Larynx,* vol. 46, no. 1, pp. 10-17, 2019.

[8] M. Amit *et al.*, "Improving the rate of negative margins after surgery for oral cavity squamous cell carcinoma: a prospective randomized controlled study," *Head & neck,* vol. 38, no. S1, pp. E1803-E1809, 2016.

[9] M. W. Kubik *et al.*, "Intraoperative margin assessment in head and neck cancer: a case of misuse and abuse?," *Head and neck pathology,* vol. 14, no. 2, pp. 291-302, 2020.

[10] F. A. Kerker *et al.*, "Anatomical locations in the oral cavity where surgical resections of oral squamous cell carcinomas are associated with a close or positive margin—a retrospective study," *Clinical oral investigations,* vol. 22, no. 4, pp. 1625-1630, 2018.

[11] W. Deng *et al.*, "Head and neck cancer tumor segmentation using support vector machine in dynamic contrast-enhanced MRI," *Contrast media & molecular imaging,* vol. 2017, 2017.

[12] H. Chung, G. Lu, Z. Tian, D. Wang, Z. G. Chen, and B. Fei, "Superpixel-based spectral classification for the detection of head and neck cancer with hyperspectral imaging," in *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2016, vol. 9788: International Society for Optics and Photonics, p. 978813.

[13] M. Mete, X. Xu, C.-Y. Fan, and G. Shafirstein, "Head and neck cancer detection in histopathological slides," in *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)*, 2006: IEEE, pp. 223-230.

[14] G. Lu *et al.*, "Detection of head and neck cancer in surgical specimens using quantitative hyperspectral imaging," *Clinical Cancer Research,* vol. 23, no. 18, pp. 5426-5436, 2017.

[15] M. Halicek *et al.*, "Hyperspectral Imaging of Head and Neck Squamous Cell Carcinoma for Cancer Margin Detection in Surgical Specimens from 102 Patients Using Deep Learning," *Cancers (Basel),* vol. 11, no. 9, Sep 14 2019, doi: 10.3390/cancers11091367.

[16] M. Halicek *et al.*, "Head and Neck Cancer Detection in Digitized Whole-Slide Histology Using Convolutional Neural Networks," *Sci Rep,* vol. 9, no. 1, p. 14043, Oct 1 2019, doi: 10.1038/s41598-019-50313-x.

[17] R. K. Gupta, M. Kaur, and J. Manhas, "Tissue level based deep learning framework for early detection of dysplasia in oral squamous epithelium," *Journal of Multimedia Information System,* vol. 6, no. 2, pp. 81-86, 2019.

[18] C. S. Chu, N. P. Lee, J. W. Ho, S.-W. Choi, and P. J. Thomson, "Deep Learning for Clinical Image Analyses in Oral Squamous Cell Carcinoma: A Review," *JAMA Otolaryngology–Head & Neck Surgery,* vol. 147, no. 10, pp. 893-900, 2021.

[19] J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, "Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018: IEEE, pp. 770-773.

[20] M. Shaban *et al.*, "A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma," *Scientific reports,* vol. 9, no. 1, pp. 1-13, 2019.

[21] S. Fakurnejad *et al.*, "Fluorescence molecular imaging for identification of high-grade dysplasia in patients with head and neck cancer," *Oral oncology,* vol. 97, pp. 50-55, 2019.

[22] S. Ortega *et al.*, "Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images," *Proc SPIE Int Soc Opt Eng,* vol. 11320, Feb 2020, doi: 10.1117/12.2548609.

[23] L. Ma *et al.*, "Hyperspectral microscopic imaging for the detection of head and neck squamous cell carcinoma on histologic slides," in *Medical Imaging 2021: Digital Pathology*, 2021, vol. 11603: International Society for Optics and Photonics, p. 116030P.
[24] S. Ortega, H. Fabelo, R. Camacho, M. De la Luz Plaza, G. M. Callicó, and R. Sarmiento, "Detecting brain tumor in pathological slides using hyperspectral imaging," *Biomedical optics express*, vol. 9, no. 2, pp. 818-831, 2018.
[25] J. Liao *et al.*, "Terapixel hyperspectral whole-slide imaging via slit-array detection and projection," *Journal of biomedical optics*, vol. 23, no. 6, p. 066503, 2018.
[26] L. Ma, M. Halicek, X. Zhou, J. Dormer, and B. Fei, "Hyperspectral microscopic imaging for automatic detection of head and neck squamous cell carcinoma using histologic image and machine learning," in *Medical Imaging 2020: Digital Pathology*, 2020, vol. 11320: International Society for Optics and Photonics, p. 113200W.
[27] M. Halicek *et al.*, "Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning," *Cancers*, vol. 11, no. 9, p. 1367, 2019.
[28] J. P. Shah, "Thyroid carcinoma: epidemiology, histology, and diagnosis," *Clinical advances in hematology & oncology: H&O*, vol. 13, no. 4 Suppl 4, p. 3, 2015.
[29] M. Halicek *et al.*, "Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks," *Scientific reports*, vol. 9, no. 1, pp. 1-11, 2019.
[30] L. G and F. B, "Medical hyperspectral imaging: a review," *J Biomed Opt*, vol. 19, no. 1, p. 010901, 2014.
[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[32] L. Ma *et al.*, "Hyperspectral microscopic imaging for head and neck squamous cell carcinoma detection in histologic images," presented at the Medical Imaging 2021: Digital Pathology, 2021.