Advancing Thyroid Pathologies Detection with Recurrent Neural Networks and Micro-FTIR Hyperspectral Imaging

Matheus de Freitas Oliveira Baffa
Dept. of Computing and Mathematics
University of São Paulo (USP)
Ribeirao Preto - SP, Brazil
mbaffa@usp.br

Luciano Bachmann
Dept. of Physics
University of São Paulo (USP)
Ribeirao Preto - SP, Brazil
l.b@usp.br

Denise Maria Zezell
Nuclear and Energy Research Institute
São Paulo - SP, Brazil
zezell@usp.br

Thiago Martini Pereira
Dept. of Science and Technology
Federal University of São Paulo (UNIFESP)
São José dos Campos - SP, Brazil
t.pereira@unifesp.br

Thomas Martin Deserno
Peter L. Reichertz
Institute for Medical Informatics
Technische Universität Braunschweig
Braunschweig, Germany
thomas.deserno@plri.de

Joaquim Cezar Felipe
Dept. of Computing and Mathematics
University of São Paulo (USP)
Ribeirao Preto - SP, Brazil
jfelipe@ffclrp.usp.br

2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS) | 979-8-3503-1224-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/CBMS58004.2023.00288

***Abstract*—Thyroid disorders are a complex group of diseases that require an accurate diagnosis for effective treatment. Fine-needle aspiration biopsies can assist in detecting many thyroid diseases. These materials can be analyzed visually using traditional computer vision methods, despite the limitations of complex samples. To address this problem, we propose a novel approach that uses hyperspectral imaging (HSI) to analyze thyroid biological samples. HSI measures the absorbance of infrared light by biological samples using a micro Fourier transform infrared spectroscopy (micro-FTIR) and converts this data into hyperspectral images. In this study, we used HSI to train and validate a recurrent neural network to classify thyroid samples as healthy, cancerous, or goiter. Our experiments, based on the k-fold cross-validation, achieved an overall accuracy of 96.88%, a sensitivity of 96.87%, and a specificity of 98.45%. These results demonstrate the potential of hyperspectral imaging as a tool to assist pathologists in the diagnosis of thyroid disease.**

***Index Terms*—classification, computational pathology, deep learning, hyperspectral imaging, thyroid cancer**

I. Introduction

The thyroid is a gland located in the anterior neck and it is responsible for producing hormones such as T3, T4, and calcitonin. It contributes to the body's metabolism, growth, and concentration of electrolytes in the serum [1]. Uncontrollable growth of cells able to avoid cellular apoptosis and invade distant organs characterize Thyroid cancer [2].

In 2023, the American Cancer Society estimates 43,720 new cases of thyroid cancer, of which 12,540 are in men and 31,180 are in women. For the same period, 2,120 deaths are expected, which represents an increase of 0.6% from the previous year [3].

This disease usually appears as a lump in the neck, sometimes growing quickly, along with pain in the front of the neck, hoarseness, swelling, and difficulties in swallowing and breathing [4]. After the first symptoms, the patient may be guided to undergo image tests to find suspicious areas, but the actual diagnosis is only made using biopsy, of which a small portion of the suspicious area is removed using fine-needle aspiration and studied in the laboratory [5]. Inflammatory and autoimmune diseases, such as Goiter and Thyroiditis also affect the thyroid and may be histologically detected [6].

Traditional computer vision methods use digital images acquired from sample slides stained with hematoxylin and eosin (H&E) [7], [8]. These methods represent the image's content in a format of feature vector and use machine learning to find patterns and build a classification model. These methods tend to perform well on small samples but they have difficulties when the samples are larger and contain different pathologies.

Different from the traditional digital image acquisition process, hyperspectral imaging (HSI) uses a special camera called a spectrometer to generate images, characterized by their large number of channels [9]. Using a light source, the spectrometer measures how the biological material interacts with the light at different frequencies, including those not visible to humans. Some of the most common measurements are reflectance, transmittance, emission, and absorbance. The absorbance spectra can provide information about the molecular and histochemical composition of the tissue, which can be used for diagnostic purposes. In medical applications, HSI has been successfully applied to detect breast cancer [10], ovarian cancer [11], and thyroid cancer [12].

Over the literature, other researchers approached thyroid tissue characterization using hyperspectral imaging. In 2017, Halicek et al. [12] developed a convolutional neural network (CNN) to differentiate head and neck cancer using hyperspectral imaging. Their method consists in obtaining HSI from

2372-9198/23/$31.00 ©2023 IEEE
DOI 10.1109/CBMS58004.2023.00116

tissues, such as healthy thyroid and squamous-cell carcinoma using reflectance spectroscopy over a bandwidth from 450 to 900cm^{-1}. Overall their method achieved an accuracy of 80%, a sensitivity of 81%, and a specificity of 78% in classifying the tissue between healthy and cancer. For thyroid-only experiments, his method based on CNNs achieved an overall accuracy of 90%, a sensitivity of 83%, and a specificity of 93%.

Later on, Lu et al. [13] developed a work for head and neck cancer detection on histological samples. The authors acquired HSI from 36 patients using a reflectance spectrometer over a bandwidth of 450 to 900 cm^{-1}, with a 5cm^{-1} interval. Several conventional machine learning methods were benchmarked to check the best classification method for this data. The ensemble linear discriminant analysis (LDA) achieved the highest results for thyroid tissue, with an accuracy of 91%, a sensitivity of 91%, and a specificity of 93% in classifying them between cancer and healthy tissue.

In 2020, Maktabi et al. [14] developed an automatic classification method to detect goiter and healthy tissue. Their method is based on reflectance spectroscopy which was acquired in the spectral range of 500 to 1000cm^{-1}. Using a support vector machine, they obtained a patient-based accuracy of 68%.

Edwards et al. [15] evaluated HSI to predict the aggressiveness of papillary thyroid carcinoma using multiparametric radiomics. They extracted 120 features to represent the HSI content in a form of a feature vector. This feature vector is used to train and benchmark several conventional machine learning methods, of which the linear support vector machine (SVM) and the quadratic discriminant analysis (QDA) achieved the highest results. Both classifiers achieved an accuracy of 83% and an area under the curve (AUC) of 85%.

Recently, Tran et al. [16] developed a deep learning-based method to detect thyroid carcinoma on whole histological slides. They acquired 33 HSI from 23 patients using reflectance spectroscopy over the wavelength of 467 to 721cm^{-1}. The classification was performed using a VGG-19 network and achieved an overall accuracy of 93.80%, an F1-Score of 86.72%, and an AUC of 96.60%.

Therefore, this paper addresses thyroid tissue classification using HSI acquired with absorbance spectroscopy in the infrared spectrum (micro-FTIR Spectroscopy). The proposed method uses a deep learning-based algorithm, called recurrent neural network (RNN), to detect patterns over the infrared spectrum and develop a classification model for further applications.

A. Our Contributions

The majority of the articles explore thyroid tissue classification using reflectance spectroscopy methods over a visible spectrum bandwidth. Reflectance spectroscopy measures how much light is being reflected by the material when exposed to light. Also, most papers approach a binary classification problem, between cancer and healthy tissue, or goiter and healthy tissue. Although they reach good results in their proposal, it is common to observe slides with multiple diagnoses, mostly containing all three types above mentioned. There is still a gap in this research field, which includes (i) investigating the performance of HSI acquired using other optical methods, such as absorbance and transmission; (ii) studying other spectral ranges, such as infrared from 700 to 1800cm^{-1}; and (iii) exploring deep learning methods, such as fully-connected neural networks and recurrent neural networks to detect patterns over the spectral information. Therefore, the main contributions of this paper are stated below:

- HSI acquisition process using absorbance spectroscopy over the infrared spectrum (micro-FTIR), from 750 to 1800cm^{-1}.
- Multi-classification model for cancer, healthy, and goiter tissue.
- Development of a custom recurrent neural network (RNN) architecture to detect patterns over the spectrum.

II. Methodology and Development

Our method involves the classification of each HSI voxel and the generation of a classification map that identifies all the present classes in a given sample. It was implemented in several steps (Figure 1). First, we acquired micro-FTIR HSI data and processed it to remove unwanted information (Fig. 1, Steps 1-4). Then, we created a descriptive database by collecting the spectra from the region of interest and separated it into a training and testing dataset (Fig. 1, Step 5). After that, we improved the data variability using data augmentation (Fig. 1, Step 6). Finally, we trained a recurrent neural network to detect patterns in the spectra and classify the samples. To explain these steps in detail, this section is divided into three parts: (i) data acquisition, (ii) preprocessing and data augmentation, and (iii) classification methodology. These parts provide a comprehensive description of each step of our method, along with the explanation behind the choices we made.

A. Data Acquisition

We obtained thyroid biopsies from a specialized company (US Biomax Inc., Rockville, USA) that acquires histological material and provides histochemical analysis services. We acquired sample slides with healthy, cancer, and goiter thyroid tissue. The biopsies are pre-labeled by two Biomax pathologists and confirmed by two pathologists from the Department of Pathology, School of Medicine of Ribeirão Preto and the Pathology Service from the Clinical Hospital, Ribeirão Preto. In total, we acquired 60 samples from 60 different patients, 20 of each tissue type.

The HSI was acquired using a micro-FTIR spectrometer (Perkin Elmer Spotlight 400) connected to a microscope. This system measures the intensity of absorbance over the infrared spectrum (E.g., Figure 2) ranging from 778cm^{-1} to 1800cm^{-1}, using the spectral mapping technique. The measurements were taken every two frequencies, resulting in 512 intensity readings for each pixel. The system provides 4cm^{-1} spectral resolution and 6.25μm spatial resolution. The

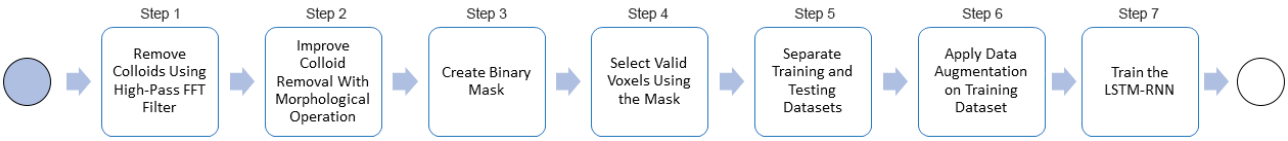

Fig. 1. Overview of the proposed method.

HSI was acquired in a controlled environment with dry air, and the relative humidity was kept below 5%.

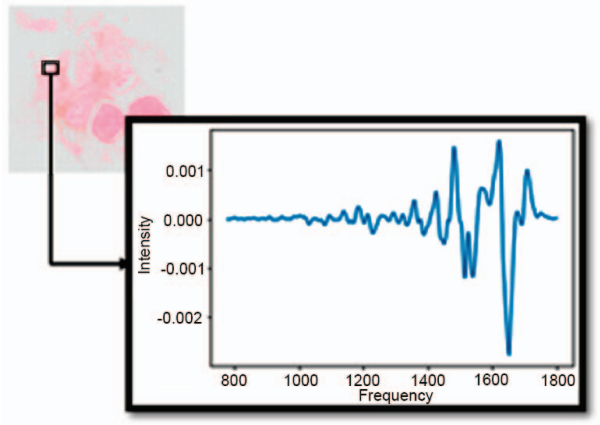

Fig. 2. Absorbance intensity over the infrared spectrum collected using micro-FTIR spectroscopy.

B. Preprocessing and Data Augumentation

The raw HSI was preprocessed to reduce noise and the influence of humidity. To do so, we calculated the second derivative of the signal and applied the Savitzky-Golay filter, with a window with size 11 and a second-order polynomial. The Savitzky-Golay filter is a digital filter that smooths data points in a spectrum using a polynomial fitting approach [17]. We opted to utilize this filter based on its effective spectral smoothing and its function here is to remove noise and thin bands of water, among the region of 1350 and 1800cm^{-1}.

We remove unwanted thyroid structures, such as colloids, that may not contain cancer, healthy, or goiter signal. To do so, we observed that the cellular information is visible as high frequencies on grayscale images. Therefore, to remove such structures and to keep only the region of interest (ROI), we developed a binary mask (Figure 3) using a high-pass filter based on a fast Fourier transform (FFT) algorithm along with the opening morphological operation to remove noisy pixels and enlarge the cellular thickness in the mask.

In order to improve generalization in the network and consequently the effectiveness of applying the classification model to real-world problems, we used a data augmentation technique to increase the data variability. The technique relies on adding Gaussian noise to the data used in the training dataset, doubling its size, and increasing the robustness of the method. Adding noise to the input of a neural network can

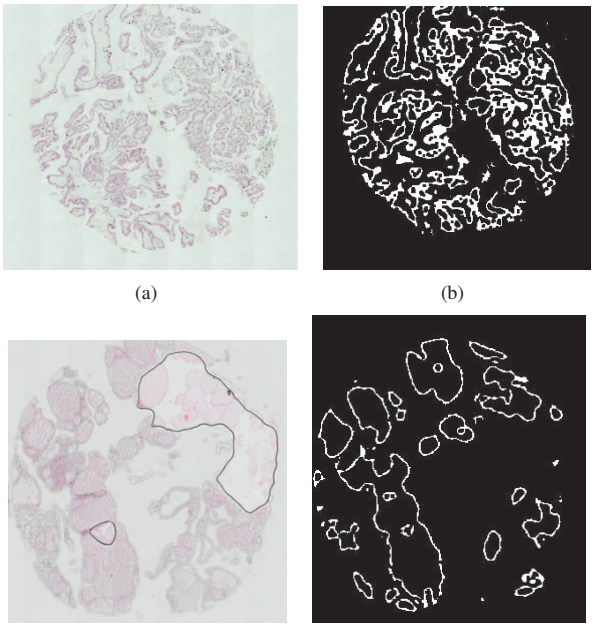

Fig. 3. H&E samples and their respective HSI binary masks.

be considered a type of data augmentation [18]. The Gaussian noise was implemented using the NumPy framework and may be calculated using the equation

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (1)$$

where σ is the standard deviation, μ is the mean, and the σ^2 is the variance.

We developed a voxel dataset containing 104,107 spectra from a total of 60 hyperspectral images, not including the augmented data. It is important to notice that the data augmentation is only applied to the training dataset and the separation between training and test subsets considers different patients among them. Therefore, no patient is repeated in between the subsets. Besides, the number of entries in each class during the training process is balanced based on the number of voxels in the minority class.

C. Classification Methodology

Due to the nature of the problem, in which a variable (intensity) changes over the frequencies, the recurrent neural network (RNN) was the deep learning architecture that bests

suits the proposal. RNN is a special type of neural network that can take advantage of this type of data [19]. Therefore, this neural network is able to detect the pattern of changes in absorption intensity along the frequencies of the spectrum.

There is a common problem with RNN called vanishing gradient. This problem occurs during the detection of patterns using a very large sequence. While the algorithm detects patterns throughout the sequence, the intensity values seen at the beginning of the training can be forgotten, reducing the classification model's effectiveness. Long short-term memory (LSTM) is a specific type of recurrent layer that includes a memory module, capable of reducing the vanishing gradient problem.

The LSTM-RNN architecture developed for this work consists of two LSTM layers with 256 and 128 units, respectively. Each LSTM layer is followed by a dropout layer that randomly turns off 20% of the neurons to reduce the risk of overfitting and improve the model's generalization. The input for the neural network is the entire spectrum, which includes 512 intensity values for each observed frequency. After the LSTM layers, we used a dense layer with 512 neurons, and the ReLU activation function is used to aid in pattern recognition. A dropout layer of 20% follows the dense layer. Finally, the output layer contains three neurons representing cancer, healthy, and goiter classes, using the softmax activation function.

We used the categorical cross-entropy function along with the Adam optimization algorithm to calculate the loss between the predicted and expected values during training. The model was trained for 200 epochs, which were found to be sufficient for convergence to the best result. To fit into the available two GPUs' memory, we optimized the batch size to 1000. The architecture was designed empirically with the goal of maximizing the efficiency of the classification model.

III. Experiments and Results

To implement the method we used Python programming language, in version 3.9, along with Tensorflow 2.10 [20] and Scikit-Learn 1.2 [21]. The development and experiments run into a server with Linux Ubuntu 22.04, two Intel Xeon Silver processors, 192 GB of RAM DDR4, and two NVIDIA RTX A4000 graphics card.

The experiments were conducted in an inter-patient k-fold cross-validation protocol. The whole spectrum database was divided into ten parts ($K = 10$) of which, each experiment used nine parts for training ($K-1$) and one part for testing. As the experiment is inter-patient, there were exclusive patients in between the subsets, therefore no patients were in training and testing subsets at the same time. Additionally, to validate externally, one patient of each class was carried from both test and train separation to visually validate the model.

To evaluate the effectiveness, we measured the accuracy, sensitivity, specificity, precision, and AUC for each iteration (Table 1). Overall, the majority of the folds achieved accuracies above 95%, with the highest accuracy reaching 99.42% in folds 3 and 99.34% in fold 4, and the lowest at 89.01% in fold 8. The average accuracy across all folds was 96.88%, indicating the network's strong ability to extract relevant information and build an efficient classification model.

TABLE I
LSTM-RNN Performance Over the K-Fold Cross-Validation.

Fold	Accuracy	Precision	Sensitivity	Specificity	AUC
0	94.54%	94.61%	94.54%	97.30%	97.94%
1	99.73%	99.77%	99.73%	99.89%	99.95%
2	98.51%	98.52%	98.50%	99.26%	99.47%
3	99.42%	99.42%	99.42%	99.71%	99.91%
4	99.34%	99.34%	99.34%	99.67%	99.89%
5	98.99%	98.99%	98.98%	99.50%	99.73%
6	95.20%	95.32%	95.17%	97.66%	98.30%
7	96.86%	96.86%	96.86%	98.43%	99.07%
8	89.01%	89.04%	88.98%	94.52%	95.57%
9	97.21%	97.22%	97.21%	98.61%	98.94%
Mean	**96.88%**	**96.90%**	**96.87%**	**98.45%**	**98.87%**
Std. Dev.	±3.13	±3.11	±3.13	±1.55	±1.28

To assess the generalization ability of the developed RNN-based classification model on samples not seen in the train-test separation, we performed a validation experiment on three independent samples, one from each class. These samples were obtained from patients that were not included in the training or test subsets. Figure 4 displays the H&E images of the samples and their corresponding classification maps, where green, blue, and red pixels represent healthy, goiter, and cancer spectra, respectively, detected by the classification model. The results show that the classification model achieved an average accuracy of 96%, indicating its high performance in identifying thyroid tissue classes. This validation experiment demonstrates the potential of the RNN-based model for diagnosing thyroid diseases in a real-world application.

A. Discussion

Compared to other deep learning approaches, the RNN architecture is particularly well-suited to the natural shape of the data. Its ability to take advantage of the structure of the data allows it to capture patterns in the way that absorption intensity changes across the spectrum. This is the main reason why we chose to use RNN in our approach to solving the problem.

Due to the LSTM-RNN complexity and the number of frequencies used, the training time was not short, around 2 hours and 30 minutes for each fold on the such hardware configuration. This would be a disadvantage of such a robust method. On the other hand, with the capacity of representing the data as its original shape, with all the frequencies, we were able to create a classification method that generalized well. Additionally, the data augmentation method played an important role in improving performance, as we observed an increase of around 7% in accuracy and 9% in sensitivity when we added Gaussian Noise to increase the number of training samples.

IV. Conclusion

In this paper, we have demonstrated the potential of hyperspectral imaging for the accurate and efficient diagnosis of thyroid diseases. Considering the limitations of traditional

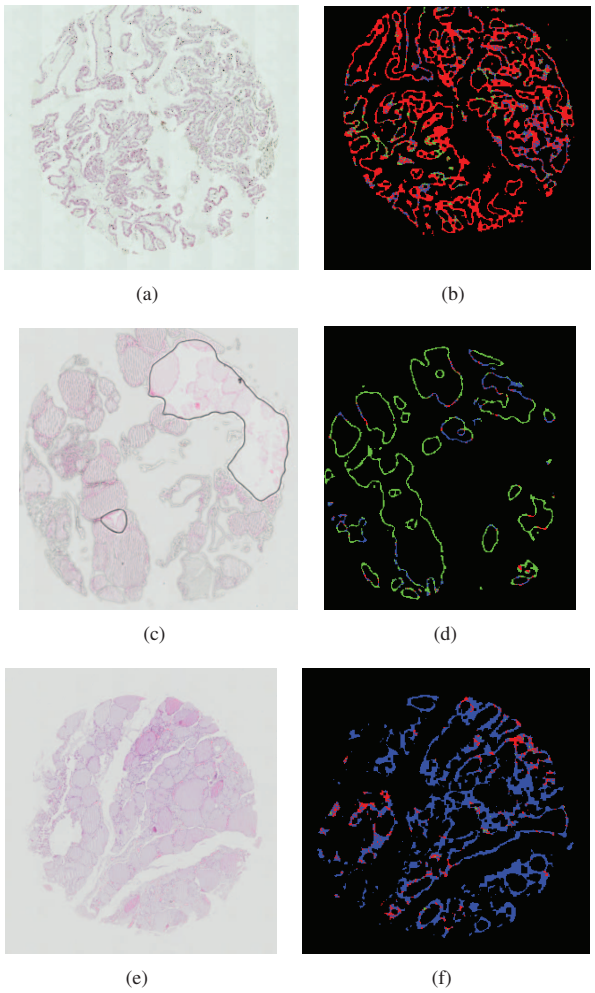

Fig. 4. Classification maps of three tissue samples from different patients: (a) cancerous tissue, (c) healthy tissue, (e) goiter tissue, and (b), (d), and (f) their corresponding classification map. Green pixels indicate healthy tissue, blue pixels indicate goiter, and red pixels indicate cancer.

computer vision methods, such as samples with multiple diagnostics and diseases with similar aspects, hyperspectral images are capable to represent these tissues using their histochemical composition by measuring how they interact with light. The absorbance spectroscopy over the infrared spectrum showed as an effective way to differentiate the samples between healthy, cancerous, and goiter contributing to improving the accuracy and speed of thyroid disease diagnosis, and treatment. In future work, we intend to incorporate the proposed AI model into a diagnostic tool that will aid pathologists in the detection and classification of thyroid diseases.

Acknowledgment

This work is supported in part by The São Paulo Research Foundation (FAPESP) under grant 2021/00633-0. M.F.O.B. is supported in part by Coordination for the Improvement of Higher Education Personnel (CAPES) under grant 88887.498626/2020-00 and 88887.695355/2022-00.

References

[1] E. Allen and A. Fingeret, "Anatomy, head and neck, thyroid," in *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK470452/

[2] National Health Service, "Thyroid cancer," 2019. [Online]. Available: https://www.nhs.uk/conditions/thyroid-cancer/

[3] American Cancer Society, "Key statistics for thyroid cancer," 2023. [Online]. Available: https://www.cancer.org/cancer/thyroid-cancer/about/key-statistics.html

[4] American Cancer Society, "Signs and symptoms of thyroid cancer," 2023. [Online]. Available: https://www.cancer.org/cancer/thyroid-cancer/detection-diagnosis-staging/signs-symptoms.html

[5] Instituto Nacional do Câncer, "Thyroid cancer - version for health professionals," 2022. [Online]. Available: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/tireoide

[6] P. Caturegli, H. Kimura, R. Rocchi, and N. R. Rose, "Autoimmune thyroid diseases," *Current opinion in rheumatology*, vol. 19, no. 1, pp. 44–48, 2007.

[7] S. Assaad, D. Dov, R. Davis, S. Kovalsky, W. T. Lee, R. Kahmke, D. Rocke, J. Cohen, R. Henao, L. Carin, and D. E. Range, "Thyroid cytopathology cancer diagnosis from smartphone images using machine learning," *Modern Pathology*, vol. 36, no. 6, p. 100129, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893395223000340

[8] M. de Freitas Oliveira Baffa, L. Bachmann, T. M. Pereira, D. M. Zezell, E. G. Soares, J. D. B. Pádua, and J. C. Felipe, "Histopathological analysis of fine-needle aspiration biopsies of thyroid nodules using explainable convolutional neural networks," *Proceedings of IFMBE International Federation for Medical and Biological Engineering*, 2022.

[9] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, pp. 010 901–010 901, 2014.

[10] I. H. Aboughaleb, M. H. Aref, and Y. H. El-Sharkawy, "Hyperspectral imaging for diagnosis and detection of ex-vivo breast cancer," *Photodiagnosis and Photodynamic Therapy*, vol. 31, p. 101922, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1572100020302763

[11] T. E. Renkoski, K. D. Hatch, and U. Utzinger, "Wide-field spectral imaging of human ovary autofluorescence and oncologic diagnosis via previously collected probe data," *Journal of biomedical optics*, vol. 17, no. 3, pp. 036 003–036 003, 2012.

[12] 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," *Journal of biomedical optics*, vol. 22, no. 6, pp. 060 503–060 503, 2017.

[13] 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 imaginghyperspectral imaging for head and neck cancer detection," *Clinical Cancer Research*, vol. 23, no. 18, pp. 5426–5436, 2017.

[14] M. Maktabi, H. Köhler, M. Ivanova, T. Neumuth, N. Rayes, L. Seidemann, R. Sucher, B. Jansen-Winkeln, I. Gockel, M. Barberio *et al.*, "Classification of hyperspectral endocrine tissue images using support vector machines," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 16, no. 5, pp. 1–10, 2020.

[15] M. Halicek, J. V. Little, A. Y. Chen, B. Fei *et al.*, "Multiparametric radiomics for predicting the aggressiveness of papillary thyroid carcinoma using hyperspectral images," in *Medical Imaging 2021: Computer-Aided Diagnosis*, vol. 11597. SPIE, 2021, pp. 517–524.

[16] M. H. Tran, L. Ma, J. V. Litter, A. Y. Chen, and B. Fei, "Thyroid carcinoma detection on whole histologic slides using hyperspectral imaging and deep learning," in *Medical Imaging 2022: Digital and Computational Pathology*, vol. 12039. SPIE, 2022, pp. 101–111.

[17] R. W. Schafer, "What is a savitzky-golay filter?[lecture notes]," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[19] L. R. Medsker and L. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, pp. 64–67, 2001.

[20] M. Abadi et. al, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org/

[21] L. Buitinck et al., "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.