Computer Methods and Programs in Biomedicine 247 (2024) 108100

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Deep neural networks can differentiate thyroid pathologies on infrared hyperspectral images

Matheus de Freitas Oliveira Baffa [a,*], Denise Maria Zezell [b], Luciano Bachmann [c], Thiago Martini Pereira [d], Thomas Martin Deserno [e], Joaquim Cezar Felipe [a]

[a] Department of Computing and Mathematics, University of São Paulo, Ribeirão Preto, SP, Brazil
[b] Nuclear and Energy Research Institute, São Paulo, SP, Brazil
[c] Department of Physics, University of São Paulo, Ribeirão Preto, SP, Brazil
[d] Department of Science and Technology, Federal University of São Paulo, São José dos Campos, SP, Brazil
[e] Peter L. Reichertz Institute for Medical Informatics, Technische Universität Braunschweig, Braunschweig, Germany

ARTICLE INFO

Keywords:
Computer-aided diagnosis
Deep learning
Hyperspectral imaging
Computational pathology
Thyroid cancer

ABSTRACT

Background and objective: The thyroid is a gland responsible for producing important body hormones. Several pathologies can affect this gland, such as thyroiditis, hypothyroidism, and thyroid cancer. The visual histological analysis of thyroid specimens is a valuable process that enables pathologists to detect diseases with high efficiency, providing the patient with a better prognosis. Existing computer vision systems developed to aid in the analysis of histological samples have limitations in distinguishing pathologies with similar characteristics or samples containing multiple diseases. To overcome this challenge, hyperspectral images are being studied to represent biological samples based on their molecular interaction with light.
Methods: In this study, we address the acquisition of infrared absorbance spectra from each voxel of histological specimens. This data is then used for the development of a multiclass fully-connected neural network model that discriminates spectral patterns, enabling the classification of voxels as healthy, cancerous, or goiter.
Results: Through experiments using the k-fold cross-validation protocol, we obtained an average accuracy of 93.66 %, a sensitivity of 93.47 %, and a specificity of 96.93 %. Our results demonstrate the feasibility of using infrared hyperspectral imaging to characterize healthy tissue and thyroid pathologies using absorbance measurements. The proposed deep learning model has the potential to improve diagnostic efficiency and enhance patient outcomes.

1. Introduction

The thyroid is an endocrine gland situated in the lower front region of the neck, near the trachea. This gland is important for the production of hormones T3 and T4, which play significant roles in various body regulatory processes, including heart rate, energy level, and metabolism. Pathologies of the thyroid encompass a wide variety of disorders, all of which can adversely impact these hormonal productions. These disorders can range from benign conditions, such as thyroiditis and goiter, to more severe ones, including hyperthyroidism, hypothyroidism, and even thyroid cancer [1].

Goiter refers to the enlargement of the thyroid gland, a condition indicating that the gland's volume exceeds normal size. It can manifest in different forms, such as diffuse, nodular, or multinodular. Typically, the thyroid gland grows anteriorly in the neck due to minimal constraints from the anterior cervical muscles, subcutaneous tissue, or skin [2].

On the other hand, thyroid cancer is a condition characterized by the growth of tumor cells within the thyroid gland tissue [3]. Typical symptoms of thyroid cancer include the presence of a neck lump (or nodule), difficulty in breathing or swallowing, pain, and voice hoarseness. Projections indicate that by 2023, an estimated 1958,310 new cases of thyroid cancer are expected, and approximately 609,820 patients are projected to die from the disease [4].

Diagnosis of these diseases relies on screening techniques such as ultrasound, CT scans, and MRI, complemented by a biopsy of the thyroid gland. The fine-needle aspiration biopsy technique, in particular, extracts multiple tissue samples from the thyroid, aiding in the

* Corresponding author.
E-mail address: mbaffa@usp.br (M.F.O. Baffa).

https://doi.org/10.1016/j.cmpb.2024.108100
Received 7 December 2023; Received in revised form 12 February 2024; Accepted 23 February 2024
Available online 24 February 2024

determination of cancer type and stage. Furthermore, it facilitates treatment follow-up, confirms the complete removal of cancer, and provides an evaluation of the patient's overall health [5].

Conventional computer vision techniques have been employed to aid in diagnosing diseases from biopsy slides [6,7]. These techniques analyze digital images of slides stained with hematoxylin and eosin (H&E). While these methods have achieved satisfactory results, their accuracy tends to decrease when analyzing samples from tumor margins, where a variety of pathological states can be observed on a single slide.

To address this challenge, hyperspectral imaging (HSI) has emerged as a promising imaging modality to represent biological specimens [32]. It has been effectively applied to various tissue types, including breast [8,9], oral cavity [10,11], and gastrointestinal cancer [12,13]. HSI, resulting from vibrational spectroscopy processes, may be created using several measurements like absorbance, reflectance, and transmission. With a spectrometer, we modulate the frequency of the electromagnetic wave to acquire a sequence of measurements (spectrum) [14].

Biological tissues, with their distinct histochemical compositions, leave unique spectral signatures that can be captured and interpreted. In our initial observations, as illustrated in Fig. 1, the infrared light absorption patterns of various thyroid tissues revealed significant nuances. While the cancerous tissue demonstrated characteristic intensity contrasts in its absorption spectrum (Fig. 1(a)), the spectra of goiter and healthy tissues shared notable similarities (Fig. 1(b) and 1 (c)). However, when averaged across all three tissue types, certain regions of the spectrum emerged as more distinctive markers for each tissue type (Fig. 1(d)). These intricate patterns, although subtle and challenging for the human eye to discern, suggest the potential for computational techniques, like machine learning, to detect and interpret them.

Various HSI acquisition methods for thyroid tissue classification have been explored in the literature. In 2017, Lu et al. [15] utilized ensemble linear discriminant analysis (LDA) in a classification model to distinguish thyroid cancer tissues from healthy ones using reflectance HSI, achieving an overall accuracy of 92 %. Later in 2019, Taylor et al. [16] and O'Dea et al. [17] explored the efficiency of HSI derived from Raman spectroscopy. Both studies proposed binary classifications to detect cancerous tissue, employing agglomerative hierarchical clustering (AHC) and principal component analysis (PCA), respectively. Taylor et al. reported an overall accuracy of 89.8 %, while O'Dea et al. achieved an accuracy of 88 %.

In 2020, Halicek et al. [18] developed a method to detect tumors on thyroid and salivary samples. They processed 216 surgical specimens from 82 patients, capturing HSI data and measuring reflectance over the visible electromagnetic spectrum, ranging from 450 to 900 cm^{-1}. Their method based on convolutional neural network (CNN) reached an accuracy of 78 $\pm$ 2 % for all thyroid tumor classifications.

Edwards et al. [19] examined the use of HSI and multiparametric radiomics to predict the aggressiveness of papillary thyroid carcinoma. They extracted 120 features to form a feature vector, which was used to train various conventional machine learning models. Among these, the linear support vector machine (SVM) and quadratic discriminant analysis (QDA) demonstrated the best performance, each achieving an accuracy of 83 % and an area under the ROC curve (AUC) of 85 %.

Recently, Tran et al. [20] proposed a video transformer network for detecting thyroid cancer in hyperspectral histological images. After collecting HSI from 49 whole slides of thyroid cancer and employing five data augmentation techniques to enrich the dataset, their method achieved an accuracy of 89.64 % and an F1 score of 88.1 % on the test database.

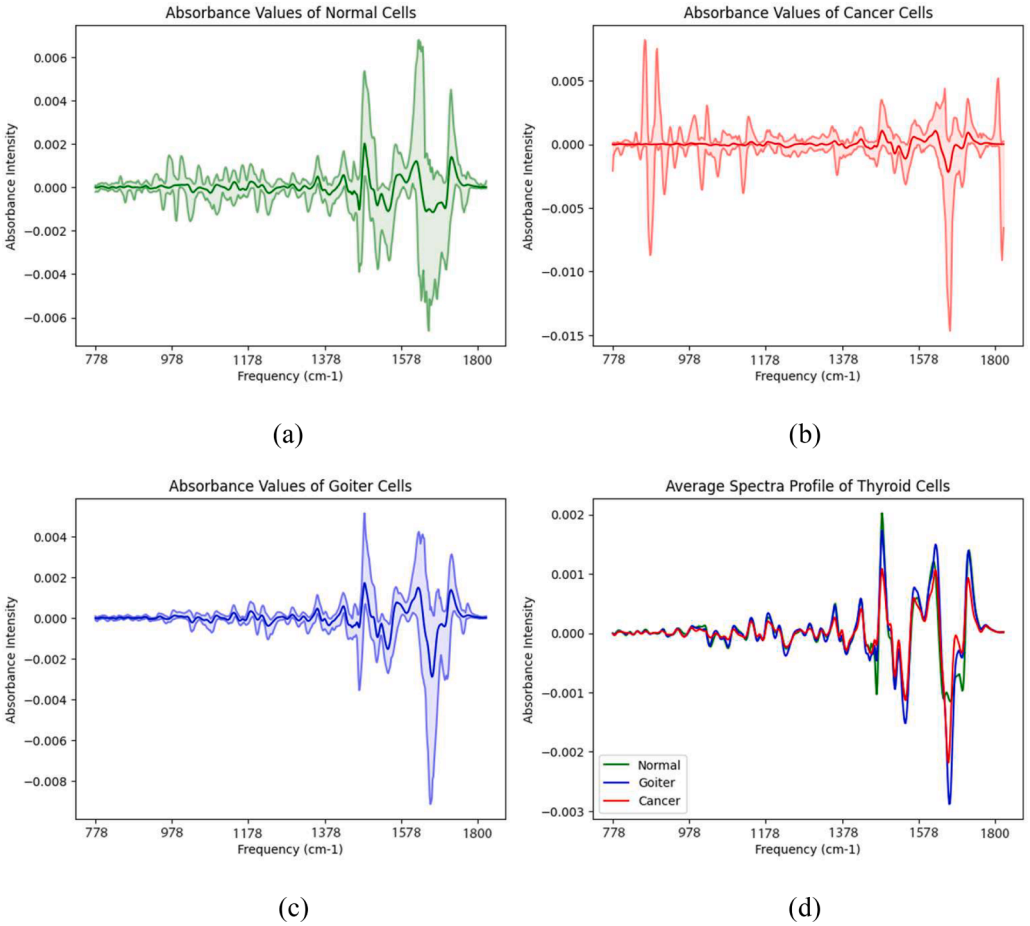


Fig. 1. Minimum, maximum, and average spectrum for each tissue type.

In contrast to the existing literature (Table 1), this study expands the tissue analysis to include both cancerous and goitrous tissues, in light of the high prevalence of goiter. This approach allows for a more comprehensive understanding of thyroid pathologies and their varied presentations. Our methodology comprises a comprehensive pipeline, including data acquisition, preprocessing, and a custom voxel-based classification model developed using a fully-connected neural network (FCNN). Furthermore, we compare our proposed method against traditional machine learning algorithms through experimentation to discern the most effective AI model for this task.

While current research focuses primarily on the binary classification of thyroid tissue using reflectance or Raman spectroscopy within the visible spectrum range, histological slides often contain multiple diagnoses, including inflammatory diseases, that may not be adequately addressed by a binary model applied to an entire sample. Furthermore, while convolutional neural networks (CNNs) and traditional machine learning algorithms are commonly employed to detect patterns and develop classification models for HSI samples, a variety of other deep learning approaches can provide additional opportunities for data modeling and pattern recognition, thereby enhancing data representation and comprehension. As such, there exist several research approaches that could be explored to bridge these gaps, including: (i) investigating the utility of alternative optical measurements such as absorbance and transmission; (ii) exploring other spectral ranges, including the infrared range with wavelengths from 700 to 1800 cm^{-1}, and far-infrared frequencies exceeding 1800 cm^{-1}; (iii) assessing the effectiveness of other deep learning architectures, such as FCNN, recurrent neural networks (RNN), and 1-dimensional convolutional neural networks (1D-CNN) in identifying patterns within spectral data; and (iv) evaluating various data modeling options, such as voxel-based and volume-based classifications.

Therefore, the main contributions of this paper are:

- Deployment of an HSI acquisition process using absorbance spectroscopy over the infrared spectrum (micro-FTIR), ranging from 750 to 1800 cm^{-1}.
- Development of a multi-classification model capable of distinguishing between cancerous, healthy, and goiter tissue.
- Inclusion of a binary evaluation for tissue classification in addition to multi-classification.
- Evaluation and implementation of a voxel-based HIS classification method.
- Design of a custom fully-connected neural network to detect patterns across the spectrum.
- Benchmarking of traditional machine learning methods against the proposed FCNN.

2. Background

Spectroscopy is a research area that investigates the interaction of light with the atoms and molecules within a specific material, resulting in emission, absorption, reflection, or transmission. Emission spectroscopy characterizes the quantity of electromagnetic radiation emitted by a substance. Absorption, reflection, and transmission spectroscopy, on the other hand, reveal how a material interacts with light, providing crucial information about its composition. At specific frequencies, these interactions can be utilized to identify a material based on its histochemical reaction [14].

Micro-Fourier Transform Infrared (Micro-FTIR) absorbance spectroscopy is a specific vibrational spectroscopic technique that utilizes infrared radiation, which is absorbed by the tissue's molecules. This technique generates unique spectral patterns to characterize each tissue structure. These patterns are specific to the structure and composition of the sample, thereby facilitating the precise differentiation and identification of various substances [21]. In the context of our study, Micro-FTIR absorbance spectroscopy enabled the differentiation between healthy and pathological thyroid tissues based on their distinct absorbance spectra. The key advantage of this method lies in its high sensitivity and specificity, which offer detailed insights into the molecular structures of biological tissues without requiring any intrusive procedures or staining.

Researchers study and measure the interaction of light with materials using specialized equipment known as a spectrometer. This device generally includes a light source along with a prism or a similar component that separates this light into various frequencies. The spectrometer then directs these frequencies onto the material under examination. Following this, a sensor records the intensity of light that the material absorbs, reflects, transmits, or emits, providing valuable insights into its properties.

A typical digital image consists of three-color channels — green, red, and blue (RGB) — each corresponding to the intensity of wavelengths within a range from approximately 400 to 700 cm^{-1}. In contrast, a hyperspectral image (Fig. 2) contains a substantially larger number of

Table 1
Literature overview on hyperspectral imaging for thyroid tissue classification.

Article	Optical Method	Spectral Band	Classification Method
Lu et al. [15]	Reflectance	450 to 900 cm^{-1}	LDA
Taylor et al. [16]	Raman	400 to 3000 cm^{-1}	AHC
O'Dea et al. [17]	Raman	400–1800 cm^{-1}	PCA
Halicek et al. [18]	Reflectance	450 to 900 cm^{-1}	CNN
Edwards et al. [19]	Not specified	450 to 900 cm^{-1}	SVM and QDA
Tran et al. [20]	Transmittance	470 to 720 cm^{-1}	Transformers
Our work	**Absorbance**	**750 to 1800 cm^{-1}**	**FCNN**

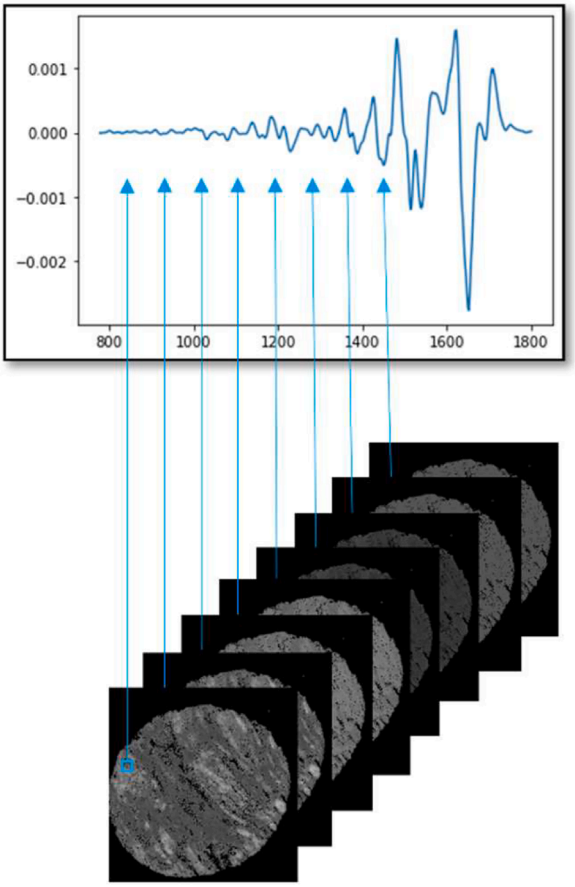


Fig. 2. Absorbance spectrum of a single HSI voxel.

channels, such as 400 to 500 channels, approximately one channel for each minor wavelength or interval, usually around 300 to 400 intensities along the wavelength. This structure allows each voxel in the hyperspectral image to store a spectrum of intensities corresponding to a portion of the examined material, hence providing a much richer representation of its properties [22].

After the acquisition of the spectral data using one of the above-mentioned spectroscopy methods, such as Micro-FTIR, we can generate a hyperspectral image. This image is a three-dimensional data structure where each voxel encapsulates the spectrum for a specific point in the specimen, capturing not only spatial information but spectral details about the composition and structure of the tissue [14,23].

3. Methodology

Our method comprises several steps for hyperspectral image analysis. First, we preprocess the hyperspectral image by eliminating irrelevant data such as colloids through a high pass filter applied in the frequency domain (Fig. 3– Step 1). We then enlarge the cell area using a morphological operation (Fig. 3- Step 2). After that, we create a binary mask (Fig. 3- Step 3) to select valid voxels over the HSI and generate a voxel dataset. Later, this dataset is used to train a classification method (Fig. 3- Steps 4 to 6) that detects patterns across the voxels. To ensure real-world applicability, we also developed a data augmentation technique using Gaussian noise to effectively increase the database's variability. In the following sections, we provide an in-depth discussion of the implementation and significance of these steps.

3.1. Hyperspectral imaging acquisition

In this study, we obtained 60 micro-array thyroid samples from different patients, each representing healthy, goiter, and cancerous tissues (20 samples each). The goiter tissue group contained samples of both hyperplastic and colloid goiter, while the cancer tissue group comprised samples of papillary and follicular carcinomas. These histological specimens were acquired from a specialized company, US Biomax Inc., located in Rockville, USA. The samples were pre-labeled by the company's pathologists and were later confirmed and reviewed by two highly experienced pathologists specialized in thyroid tissue, one from the Department of Pathology at the Ribeirão Preto School of Medicine of the University of São Paulo, and the other from the Pathology Service of the Clinical Hospital of Ribeirão Preto.

We utilized *ex vivo* samples, which were collected in two adjacent slices. One slice from each sample was prepared specifically for spectral analysis, while the other was processed for standard histological examination, including hematoxylin and eosin staining. This dual-preparation approach ensured that the hyperspectral imaging analysis was conducted on samples in a state that closely simulates what is typically encountered in a clinical or surgical setting. Furthermore, it enabled us to effectively correlate the hyperspectral data with traditional histological findings, thereby enhancing the robustness and clinical relevance of our analysis.

Our study aims to differentiate histological tissues based on their absorbance intensity within the infrared spectrum. To perform these measurements, we used a Perkin Elmer Spotlight 400 micro-FTIR spectrometer (PerkinElmer, Waltham, Massachusetts, U.S.A.) equipped with a liquid nitrogen-cooled mercury−cadmium−telluride (MCT) detector, connected to a microscope, to capture the absorbance intensity of thyroid tissue within an infrared spectrum range of 778 to 1800 cm^{-1}. Utilizing the spectral mapping technique, we collected measurements at two frequency intervals, resulting in 512 intensity readings per voxel. The system provides a spectral resolution of 4 cm^{-1} and a spatial resolution of 6.25 μm. To ensure the accuracy of our measurements, we performed the HSI acquisition under controlled dry-air conditions, keeping relative humidity below 5 %. The Spectrum Image software (PerkinElmer) automatically subtracted a background spectrum, which was obtained from the CaF2 window through 240 accumulations, from every image collected.

3.2. Preprocessing and data augumentation

In the initial preprocessing stage, we removed noise such as humidity, cosmic rays, and background from the raw signal using a two-step process. First, we applied the second derivative to the signal, which helps reveal important features that might otherwise be hidden. After that, we applied the Savitzky-Golay filter to smooth the data using polynomial fitting techniques [24]. We chose this filter specifically for its effectiveness in smoothing the spectral signal, with the main goal of eliminating noise and attenuating narrow water bands within the 1350 to 1800 cm^{-1} range.

After preprocessing, we excluded data that was not pertinent to our study, specifically colloids. Colloids, found in thyroid follicles, store hormones but do not contribute to our research objective of distinguishing between cancerous, goiter, and healthy tissue. Therefore, we implemented a segmentation process (Fig. 4) to isolate the regions of interest, which are cellular-based structures. This process involved creating a binary mask, using a high-pass filter in the frequency domain to emphasize the cellular borders (Fig. 4(c)), and expanding the detected region using opening morphological operation (Fig. 4(d)).

To enhance the robustness, we employed Standard Scalar normalization [25]. This technique adjusts the values within the dataset to have a mean of zero and a standard deviation of one, thus standardizing the range and distribution of the data. By doing so, we mitigated the potential impact of outlier values and scale disparities among frequencies, thereby ensuring a more balanced and representative analysis.

Also, to reduce the dimensionality of our problem and to avoid potential negative influences from irrelevant frequencies, we utilized the SelectKBest algorithm. This feature selection method computes the chi-square statistic between each frequency and the target class, allowing us to concentrate on the 100 most impactful frequencies — as determined by their chi-square scores — out of the total 512 measured.

To create a robust classification model that avoids overfitting and that is applicable to real-world cases, it is essential to establish a database encompassing a broad range of scenarios. Data augmentation enhances data variability, aiding the development of a more generalized classification model. In this study, we adopted a data augmentation approach that involved the addition of noise based on the Gaussian distribution. This approach simulates natural variations in data by generating noise following the Gaussian, or normal, distribution and adding it to the original data. The mean and standard deviation of the distribution were calculated from the original data, ensuring the noise was relevant and representative. This method helps increase the

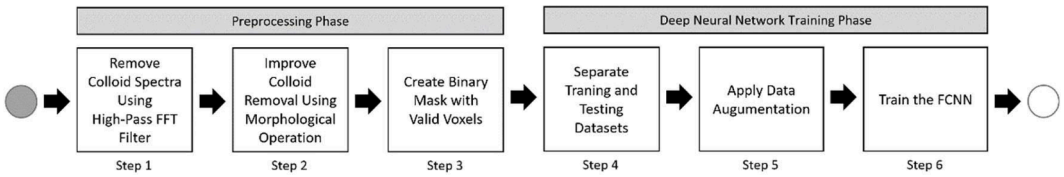
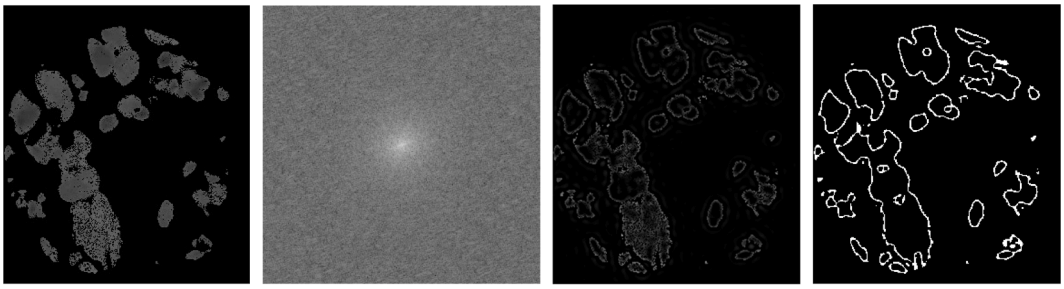


Fig. 3. Overview of the proposed method.



(a) (b) (c) (d)

Fig. 4. Process of acquiring the region of interest (ROI) throughout FFT High-Pass Filter and enhancing using opening morphological operation. In (a) the original HIS image in the frequency 1604; (b) is its respective FFT Image; in (c) is the preliminary ROI detection after the application of the high-pass filter; and in (d) is the final ROI after morphology.

robustness of our model by effectively expanding the variability within the dataset, thereby allowing our model to learn and generalize better from more diverse data. The Gaussian noise is calculated according to the following equation:

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

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

The addition of inputs with noise, a form of data augmentation, contributes to the creation of a more generalized classification model [26,27]. Data augmentation is applied only to the training set, with the test set composed of voxels from patients not previously seen during training. This approach ensures no overlap of patients between the training and testing subsets.

After preprocessing the dataset, we created a voxel dataset containing all valid voxels from the HSI image. Excluding augmented data, background, and colloids, our dataset contains a total of 104,107 voxels from 60 hyperspectral images. The dimensions of the hyperspectral images typically vary around 300 × 300 pixels in width and height.

3.3. Deep neural network architecture

Our spectral dataset is formatted as an array, with each voxel as a separate sample. To effectively detect patterns and develop a classification model, we utilize a FCNN, also known as a dense neural network. FCNN is very effective for tasks involving array-formatted data, as its architecture allows each neuron in one layer to be connected to all neurons in the subsequent layer. This characteristic enables a comprehensive exploration of interconnections and potential patterns within the data, making FCNN a fitting choice for our classification task.

The FCNN architecture developed in our study consists of eight layers, of which one is the input layer, six are hidden layers, and one is the output layer. The input layer has 100 neurons, corresponding to the 100 frequencies identified earlier by the SelectKBest algorithm. Each of the six hidden layers features 86 neurons, all employing the rectified linear unit (ReLU) activation function. To prevent overfitting and to improve model generalization, a dropout layer is added after each dense layer, set to randomly deactivate 40 % of the neurons. Finally, the output layer comprises three neurons, each representing one of the study's classes: cancerous, goiter, and healthy tissues. This layer employs the softmax activation function to provide a probability distribution over these three classes.

In addition to the aforementioned hyperparameters, we employ the RMSProp optimization algorithm and used the categorical cross-entropy function as the loss function to compute the difference between actual and predicted values during the training process. We ran the model for 2000 epochs which were sufficient for the model to converge to optimal accuracy, and the batch size of 15,000 was chosen to maximize data utilization within the GPU's VRAM capacity.

In order to apply the FCNN for binary evaluation, we reduced the number of neurons in the last layer to two, corresponding to the two classes being evaluated. The softmax activation function, previously used for the multiclass evaluation, remained the same. Furthermore, we adjusted the loss function to binary cross-entropy.

3.4. Benchmark methodology

To compare the proposed deep learning model and to identify the classification method best suited to this problem, we propose a benchmark with traditional machine learning algorithms. In total, eleven algorithms were tested: (i) quadratic discriminant analysis (QDA), (ii) naive Bayes, (iii) decision tree, (iv) AdaBoost classifier, (v) k-nearest neighbors (KNN), (vi) random forest, (vii) extreme gradient boosting (XGBoost), (viii) support vector machine with a linear-based function (linear SVM), (ix) SVM with a polynomial-based function (polynomial SVM), (x) SVM with a radial-based function (radial SVM), and (xi) multilayer perceptron (MLP).

All the data processing steps used for deep learning were also applied in this benchmark, including best frequency selection and standard scalar normalization. The algorithms were implemented using the default parameters in the Scikit-Learn library, except for the KNN algorithm where the value of K was set to 76.

4. Experimental design

The development of the proposed method and the experiments were conducted on a Linux Ubuntu 22.04 server, with two Intel Xeon Silver processors, 192 GB of RAM DDR4, and two NVIDIA RTX A4000 video cards. We implemented the method using Python 3.9 programming language with the Tensorflow 2.10 [28] and Scikit-Learn 1.2 [29] libraries.

Thyroid HSI were evaluated under a three-experiment scheme: multiclass (healthy, goiter, and cancer) evaluation with deep learning, binary (all versus all in pairs) evaluations with deep learning, and multiclass evaluation with traditional machine learning methods. In order to prove the initial hypothesis of the work, the objective of these experiments is to evaluate (i) the performance of the FCNN in detecting patterns in a multiclass approach; (ii) the performance of FCNN in detecting patterns in a binary approach; (iii) compare the complexity and differences between binary and multiclass approaches; (iv) evaluate the performance of deep learning against traditional machine learning techniques.

The experiments we conducted followed the K-fold cross validation experiment protocol in an inter-patient manner. In this protocol, the entire voxel dataset is separated into K parts, with K equal to 10, of which at each interaction, $K - 1$ parts are used for training and one is

used for testing. As this model follows the inter-patient format, the patients contained in the test are different from the patients contained in the training base, with no repetition or data leakage between the bases.

The evaluation of the methods consists of calculating five efficiency metrics: accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic curve (AUC). These metrics measure the overall effectiveness of the model in addition to in-class detection efficiency and enable comparison in the format of a benchmark.

5. Results

The quantitative evaluation of the multiclass FCNN achieves an average accuracy of 93.66 % and an average AUC of 98.62 %, demonstrating the method's overall efficacy (Table 2). Furthermore, our deep learning-based method achieved a sensitivity of 93.47 %, indicating a robust ability to detect the positive classes. Among the experiments, the highest success rates were achieved in fold 4 (97.28 %) and fold 3 (96.63 %). The lowest result, found in fold 0, achieved an accuracy of 89.82 %, which is still considerably higher than 33.33 %, the expected accuracy of a random guess in a three-class problem. This illustrates the model's capacity to distinguish between the three classes.

This average accuracy rate can also be observed in the classification of external validation samples. Fig. 5 contains three distinct tissue samples: malignant tumor (Fig. 5(a) – (c)), healthy (Fig. 5(d) – (f)) and goiter (Fig. 5(g) – (i)). Each sample has its respective voxel classification map generated by our model, providing a visual interpretation of the classification process. These three samples were left out of the training and testing process in order to visually validate the classification generated by the model. From the classification maps, we observe that 93 % of the voxels are correctly classified.

Upon analyzing the misclassifications, two distinct cases emerged: the model occasionally misidentified cancerous voxels as goiter, and it also incorrectly classified some normal voxels as goiter. This pattern can also be observed from the binary analysis of the classifier (Table 3). The differentiation between healthy and cancerous tissue reached an average accuracy of 98.73 % while between cancer and goiter, and healthy and goiter reached an overall accuracy of approximately 94 % each. We can infer from this observation that the goiter spectrum tends to approximate both healthy and cancerous spectra, which makes the classification process harder.

Regarding the benchmark with traditional methods, the MLP, XGBoost, and Radial SVM achieved performance above 90 %, showing a good ability to detect patterns in the absorbance intensity of the infrared light (Table 4). Although they achieve lower performance compared to deep learning, these methods stand out as possible methods to integrate an ensemble learning system.

6. Discussion

Our analysis of hyperspectral images presents a unique approach to the diagnosis of thyroid pathologies on histological specimens. While most existing literature is focused on binary classification using reflectance spectroscopy, we have introduced a novel method using a multiclass approach in the context of hyperspectral imaging. This methodology has the potential to be extended to other subtypes of thyroid pathologies, such as inflammatory diseases, thereby enhancing its potential to make a significant impact in the field of medical diagnostics.

Furthermore, our method brings another novelty, a voxel-based analysis in hyperspectral imaging. This approach allows the detection of multiple or mixed pathologies within a single sample. By breaking down the hyperspectral image into individual voxels and classifying each one independently, we can pinpoint distinct areas of the tissue sample, effectively overcoming the challenge of similar-looking pathologies and improving the overall diagnostic accuracy.

Table 2
Quantitative evaluation of the multiclass FCNN.

Fold	ACC	PRE	SEN	SPE	AUC
0	89.82%	90.17%	89.68%	95.11%	97.38%
1	98.40%	98.47%	98.32%	99.24%	99.80%
2	94.21%	94.35%	94.04%	97.18%	99.37%
3	96.63%	96.70%	96.57%	98.35%	99.71%
4	97.28%	97.42%	97.06%	98.71%	99.72%
5	91.50%	91.62%	91.41%	95.82%	97.41%
6	91.33%	91.47%	91.13%	95.75%	97.32%
7	92.25%	92.27%	92.17%	96.14%	97.88%
8	90.21%	90.69%	89.63%	95.40%	98.47%
9	95.03%	95.25%	94.77%	97.64%	99.15%
Mean	**93.66%**	**93.84%**	**93.47%**	**96.93%**	**98.62%**
Std. Dev	**±2.92**	**±2.84**	**±2.97**	**±1.41**	**±0.99**

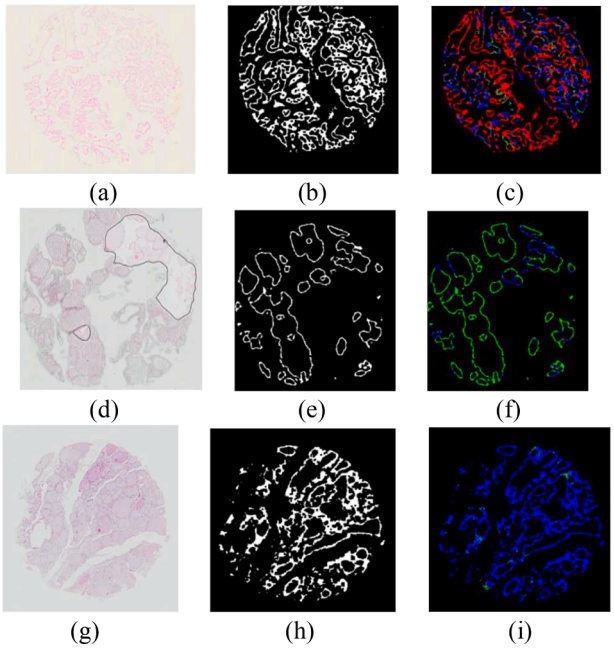


Fig. 5. Classification Map of External Validation Samples. Each pixel color denotes a diagnosis, with blue for goiter-classified voxels, green for healthy-classified voxels, and red for cancer-classified voxels.

Table 3
Average results for the binary FCNN.

Classification Type	ACC	PRE	SEN	SPE	AUC
Cancer & Healthy	98.73%	98.73%	98.73%	98.73%	99.53%
Cancer & Goiter	94.95%	94.95%	94.95%	94.95%	97.02%
Healthy & Goiter	94.43%	94.43%	94.42%	94.43%	97.43%

Table 4
Benchmark with traditional machine learning algorithms.

Classifier	ACC	PRE	SEN	SPE	AUC
QDA	49.41%	77.92%	49.01%	49.01%	74.83%
Naive Bayes	74.12%	74.20%	75.13%	75.13%	89.10%
Decision Tree	79.67%	78.12%	79.59%	79.59%	84.84%
Adaboost	81.70%	81.54%	83.00%	83.00%	91.98%
KNN	84.60%	83.94%	85.23%	85.23%	95.48%
Random Forest	87.31%	86.64%	87.87%	87.87%	96.95%
SVM Linear	88.31%	89.89%	89.97%	89.97%	95.50%
SVM Polynomial	88.34%	88.43%	89.91%	89.91%	95.10%
MLP	90.37%	91.35%	92.15%	92.15%	99.24%
XGBoost	90.64%	90.47%	91.28%	91.28%	98.61%
SVM Radial	92.59%	92.78%	93.62%	93.62%	98.19%

Data augmentation played an important role in enhancing the performance of our model. With our dataset being relatively small, there was a risk of overfitting, which could have led to a less generalized model. However, by employing data augmentation techniques along with a dropout layer, we increased the data variability of our training data by around 3–4 %, strengthening the robustness of our model.

During our experiments, we methodically evaluated the influence of varying frequency counts on our model's performance. It was observed that selecting fewer than 100 frequencies led to a deterioration in the model's diagnostic accuracy while increasing this number beyond 100 did not yield any notable improvements. Consequently, we identified the optimal count to be 100 frequencies, which exhibited the highest scoring rate based on their chi-square output. These frequencies, ranging from approximately 810 to 1728 cm^{-1}, serve as valuable markers for disease detection in thyroid tissues. This wide span comprehensively covers various biochemical components that contribute to the overall histological composition of the thyroid tissue. The literature indicates that these frequencies correspond to specific vibrational modes of molecules prevalent in pathological conditions, such as carbohydrates around 1020 to 1155 cm^{-1}; lipids from 1220 to 1470 cm^{-1}; proteins on 1204, 1240, 1280 and 1400 cm^{-1}; deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), from 1666 to 1717 cm^{-1} [30,31].

The limited size of our dataset impacted the complexity of the machine learning model we could train for detecting patterns in the HSI spectrum. Acquiring and labeling hyperspectral images is a complex task that demands the time and expertise of both physicians and pathologists. The addition of augmented data did improve the model's generalizability, suggesting that with an expanded dataset our proposed method could be scaled up to process larger data collections, potentially achieving even better results.

A limitation of our work, which also opens up opportunities for further studies, lies in goiter detection. We noticed that goiter is typically harder to differentiate from the other two classes as their binary assessment reaches 4 % less in the evaluated metrics. This finding paves the way for further research into improved methods of representing goiter data, such as combining additional optical methods, expanding the range of frequencies measured, or even evaluating the performance of other deep learning models such as recurrent neural networks and 1D convolutional neural networks, besides the ensemble of them.

Besides our approach for thyroid HSI classification, it may achieve potential clinical applications beyond thyroid pathology. Micro-FTIR spectroscopy with a multiclass FCNN could be adaptable to diagnose a range of other diseases, including breast, oral cavity, and lung cancer. The voxel-based analysis enables precise detection of multiple pathologies within a single sample, which could be very practical in clinical diagnostics. This approach could enhance disease detection accuracy, potentially leading to improved patient outcomes.

The integration of real-time HSI with advanced AI analysis presents a tool for surgeons, especially during intricate procedures. This combination has the unique capability to assist in the precise identification of various tissue types, offering a level of detail that exceeds what standard imaging techniques can provide. One advantage of HSI is its efficacy in analyzing biopsies that are inadequately prepared for traditional histopathological examination. Different from traditional computer vision methods, which depend on well-stained and properly processed samples, HSI leverages the distinct spectral signatures inherent in tissues. This allows for a comprehensive characterization of tissue types, even from samples that are poorly stained or not optimally prepared.

Moreover, our paper tackles the thyroid tissue problem using a multiclass approach, which, despite increasing complexity, has yielded very promising results. This strategy also opens avenues for exploring a broader range of thyroid tissue types in future research. For instance, subsequent studies could focus on more detailed classifications, such as determining whether the tissue is a goiter and further categorizing it into specific types, like hyperplastic or colloid. Similarly, there is potential to specify the type of carcinoma, whether it be papillary or follicular.

The voxel-based analysis of micro-FTIR spectroscopy spectra, utilizing an FCNN, marks a significant advancement in the classification of thyroid tissues, providing a sophisticated means of detecting patterns indicative of cancer, goiter, or healthy states. The principal advantage of our method lies in its capacity to analyze hyperspectral images at the voxel level, facilitating the detection of multiple or mixed pathologies within a single specimen—an accomplishment that exceeds the capabilities of traditional binary classification methods. Despite these promising advantages, a primary limitation stems from the method's complexity and the requirement for specialized instrumentation in the clinical setting, specifically a spectrometer. Additionally, the differentiation of goiter from other thyroid pathologies poses a challenge, as reflected by a modest decrease in performance metrics.

For future work, we plan to evaluate the performance of an ensemble method based on the findings from our benchmark and to integrate the proposed method into a practical tool designed to assist pathologists in their daily diagnostic routine. Additionally, we aim to explore the potential of integrating additional patient metadata along with pathological images to further enhance the model's predictive capabilities.

7. Conclusion

Through hyperspectral imaging using micro-FTIR, our study confirmed the ability to distinguish between thyroid pathologies in histological samples. Our deep learning model adeptly identified healthy, cancerous, and goiter-affected tissues, presenting a groundbreaking tool for pathologists. Despite dataset size constraints, our results evidence the diagnostic potential and precision this approach offers.

CRediT authorship contribution statement

Matheus de Freitas Oliveira Baffa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Denise Maria Zezell:** Resources, Methodology, Data curation, Conceptualization. **Luciano Bachmann:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Thiago Martini Pereira:** Writing – review & editing, Validation, Methodology, Investigation, Funding acquisition, Conceptualization. **Thomas Martin Deserno:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Joaquim Cezar Felipe:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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] Johns Hopkins Medicine, Thyroid Disorders, 2023 [Online]. Available: https://www.hopkinsmedicine.org/health/conditionsand-diseases/disorders-of-the-thyroid.

[2] A.S. Can, A. Rehman, Goiter, 2023 [online]. Available at: https://www.ncbi.nlm.nih.gov/books/NBK562161/.

[3] Thyroid Cancer Treatment, National Cancer Institute, 2023 [Online]. Available: https://www.cancer.gov/types/thyroid/patient/thyroidtreatment-pdq.

[4] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, Cancer statistics, 2023, CA Cancer J. Clin. 73 (1) (2023) 17–48.

[5] Thyroid Cancer Treatment (PDQ) – Health Professional Version, National Cancer Institute, 2023 [Online] Available at: https://www.cancer.gov/types/thyroid/hp/thyroid-treatment-pdq.

[6] Q. Guan, Y. Wang, B. Ping, D. Li, J. Du, Y. Qin, H. Lu, X. Wan, J. Xiang, Deep convolutional neural network vgg-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study, J. Cancer 10 (20) (2019) 4876.

[7] P. Chen, X. Shi, Y. Liang, Y. Li, L. Yang, P.D. Gader, Interactive thyroid whole slide image diagnostic system using deep representation, Comput. Methods Programs Biomed. 195 (2020) 105630.

[8] S. Parawira, Classification of Hyperspectral Breast Images For Cancer Detection, 2009 [Online]. Available at: https://cs229.stanford.edu/proj2009/Parawira.pdf.

[9] S. Ortega, M. Halicek, H. Fabelo, R. Guerra, C. Lopez, M. Lejeune, F. Godtliebsen, G.M. Callico, B. Fei, Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images, in: Medical Imaging 2020: Digital Pathology 11320, SPIE, 2020, pp. 206–214.

[10] J. Jayanthi, G. Nisha, S. Manju, E. Philip, P. Jeemon, K. Baiju, V. Beena, N. Subhash, Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive screening technique for early detection of malignant changes in the oral cavity, BMJ Open. 1 (1) (2011) e000071.

[11] D.G. Thiem, P. Romer, M. Gielisch, B. Al-Nawas, M. Schluter, B. Plaß, P. W. Kammerer, Hyperspectral imaging and artificial intelligence to detect oral malignancy–part 1-automated tissue classification of oral muscle, fat and mucosa using a light-weight 6-layer deep neural network, Head. Face Med. 17 (1) (2021) 1–9.

[12] H. Akbari, K. Uto, Y. Kosugi, K. Kojima, N. Tanaka, Cancer detection using infrared hyperspectral imaging, Cancer Sci. 102 (4) (2011) 852–857.

[13] F.B. Muniz, M.d.F.O. Baffa, S.B. Garcia, L. Bachmann, J.C. Felipe, Histopathological diagnosis of colon cancer using micro-FTIR hyperspectral imaging and deep learning, Comput. Methods Programs Biomed. 231 (2023) 107388.

[14] J.M. Hollas, Modern spectroscopy. Chichester, 4, John Wiley & Sons, England, 2004.

[15] G. Lu, J.V. Little, X. Wang, H. Zhang, M.R. Patel, C.C. Griffith, M.W. El-Deiry, A. Y. Chen, B. Fei, Detection of head and neck cancer in surgical specimens using quantitative hyperspectral imaging hyperspectral imaging for head and neck cancer detection, Clin. Cancer Res. 23 (18) (2017) 5426–5436.

[16] J.N. Taylor, K. Mochizuki, K. Hashimoto, Y. Kumamoto, Y. Harada, K. Fujita, T. Komatsuzaki, High-resolution raman microscopic detection of follicular thyroid cancer cells with unsupervised machine learning, J. Phys. Chem. B 123 (20) (2019) 4358–4372.

[17] D. O'Dea, et al., Raman spectroscopy for the preoperative diagnosis of thyroid cancer and its subtypes: an in vitro proof-of-concept study, Cytopathology 30 (1) (2019) 51–60.

[18] M. Halicek, J.D. Dormer, J.V. Little, A.Y. Chen, B. Fei, Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning, Biomed. Opt. Express. 11 (3) (2020) 1383–1400.

[19] E. KaToria, 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 11597, SPIE, 2021, pp. 517–524.

[20] M.H. Tran, O. Gomez, B. Fei, A video transformer network for thyroid cancer detection on hyperspectral histologic images, in: Medical Imaging 2023: Digital and Computational Pathology 12471, SPIE, 2023, pp. 32–41.

[21] C.C.S. Bandeira, K.C.R. Madureira, M.B. Rossi, J.F. Gallo, A.P.M.A. da Silva, V. L. Torres, V.A. de Lima, N.K. Junior, J.D. Almeida, R.M. Zerbinati, et al., Micro-fourier-transform infrared reflectance spectroscopy as tool for probing igg glycosylation in covid-19 patients, Sci. Rep. 12 (1) (2022) 4269.

[22] G. Lu, B. Fei, Medical hyperspectral imaging: a review, J. Biomed. Opt. 19 (1) (2014) 010901. –010901.

[23] L. He, J. Li, C. Liu, S. Li, Recent advances on spectral–spatial hyperspectral image classification: an overview and new guidelines, IEEE Trans. Geosci. Remote Sens. 56 (3) (2017) 1579–1597.

[24] R.W. Schafer, What is a savitzky-golay filter? IEEe Signal. Process. Mag. 28 (4) (2011) 111–117.

[25] T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2, Springer, 2009.

[26] C.M. Bishop, Training with noise is equivalent to tikhonov regularization, Neural Comput. 7 (1) (1995) 108–116.

[27] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT press, 2016.

[28] M. Abadi, et al., TensorFlow: Large-Scale Machine Learning On Heterogeneous Systems, 2015 [Online]. Available: https://www.tensorflow.org/.

[29] 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.

[30] Z. Movasaghi, S. Rehman, D.I. ur Rehman, Fourier transform infrared (FTIR) spectroscopy of biological tissues, Appl. Spectrosc. Rev. 43 (2) (2008) 134–179.

[31] T.M. Pereira, D.M. Zezell, B. Bird, M. Miljkovic, M. Diem, The characterization of normal thyroid tissue by micro-FTIR spectroscopy, Analyst 138 (23) (2013) 7094–7100.

[32] Q. Li, X. He, Y. Wang, H. Liu, D. Xu, F. Guo, Review of spectral imaging technology in biomedical engineering: achievements and challenges, J. Biomed. Opt. 18 (10) (2013) 100901. -100901.