Original Article

Automatic tissue segmentation of hyperspectral images in liver and head neck surgeries using machine learning

Fernando Cervantes-Sanchez[1], Marianne Maktabi[2], Hannes Köhler[2], Robert Sucher[3], Nada Rayes[3], Juan Gabriel Avina-Cervantes[4], Ivan Cruz-Aceves[5], Claire Chalopin[2]

[1]Department of Computer Science, Centro de Investigación en Matemáticas, A.C., Guanajuato 36023, Mexico.
[2]Innovation Center Computer Assisted Surgery, University of Leipzig, Leipzig 04103, Germany.
[3]Department of Visceral, Transplantation, Vascular and Thoracic Surgery, University Hospital of Leipzig, Leipzig 04103, Germany.
[4]Engineering Division, Universidad de Guanajuato Campus Irapuato-Salamanca, Guanajuato 36885, Mexico.
[5]Department of Computer Science, CONACYT-Centro de Investigación en Matemáticas, A.C., Guanajuato 36023, Mexico.

Correspondence to: Dr. Fernando Cervantes-Sanchez, Department of Computer Science, Centro de Investigación en Matemáticas, A.C., Jalisco S/N, Col. Valenciana, Guanajuato 36023, Mexico. E-mail: fernando.cervantes@cimat.mx

Abstract

Aim: Proper identification in real time of different types of tissues during intraoperative procedures represents a vital and challenging task. This paper addresses tissue segmentation in two different medical applications using hyperspectral imaging (HSI) and machine learning in two main steps.

Methods: The first step consists of data preprocessing designed to overcome the most common problems linked with HSI, involving inter- and intra-patient variability of the tissue spectra and the high dimensionality of the spectra. The preprocessing step involves outlier removal, spectral smoothing, normalization, and dimensionality reduction using principal component analysis applied in the spectral domain of HSI data. In the spatial domain, multiple levels of analysis are performed using Gaussian filters. The second step consists of tissue segmentation using an optimized machine learning model. The most suitable model was selected under statistical comparison of seven machine learning models involving three different levels of spatial analysis.

Results: According to the experimental results, the U-Net achieves the highest precision (0.908) for detection of liver, bile duct, artery, and portal vein tissues in a set of 18 HSI data, while the logistic regression with the elasticnet

regularization combined with multiscale spatial analysis obtains the highest F1-score (0.673) and segmentation accuracy (0.803) for thyroid and parathyroid glands segmentation in a set of 21 HSI data.

Conclusion: In addition to the computational experiments, combining machine learning with HSI represents a promising approach to perform image-guided surgery.

Keywords: Hyperspectral imaging, tissue segmentation, machine learning, deep learning

INTRODUCTION

Identification of distinct types of tissues is a challenging task carried out visually by surgeons during intraoperative procedures. In practice, differentiation of tissues depends on the anatomical knowledge and experience of the surgeon. Consequently, a computerized system for real-time segmentation of tissues combined with a non-invasive imaging method could be of high interest to support the surgeon during image-guided surgery.

Hyperspectral imaging (HSI) is a technique for electromagnetic spectrum sensing that has been used in multiple medical applications in the last decades. The most common applications include image-guided surgery[1], monitoring of physiologic tissue parameters[2,3], and disease diagnosis[4]. Moreover, HSI provides a high-dimensional representation of the tissue spectra along with their corresponding spatial information, which makes it suitable for tissue localization in intraoperative applications. Many studies focus on the intraoperative detection of tumors, such as colorectal, brain, liver, or head and neck tumors, mainly based on *ex vivo* tissue specimens[5-10], while few studies address the problem of characterization and enhancement of healthy tissues, structures, and organs[6,11-14]. The purpose of such works is the intraoperative discrimination of thin health risk structures that are difficult to detect with the eye or under white light imaging, from surrounding tissue such as muscle, fat, and other organs. Nouri *et al*.[6] developed a method to visually enhance the contrast of the ureter and facial nerve by combining relevant spectral channels. Cooney *et al*.[12] analyzed the spectral data of the liver, bile duct, and gall bladder in *ex vivo* human and *in vivo* animal tissues with the goal to differentiate them later. They showed that hemoglobin contributes mainly to the spectral signature of the liver, while bile pigment biliverdin and structural protein collagen are components of the absorbance spectra of the bile duct and gall bladder.

However, problems linked with HSI represent a challenge for automatic tissue localization. The most common problems involve high intra- and inter-patient variance of the tissue spectra, spectral anomalies caused by external lighting conditions, blood presence, and influence of drugs administrated to patients on the tissues' optical properties. In the literature, object localization in HSI data has been addressed using machine learning models, which are convenient to work with when dealing with high-dimensional data[15]. Again, most of the applications concern the automatic detection of tumors[10]. Two examples are given concerning thyroid and liver tumors, which represent the organs of interest of this paper. Halicek *et al*.[9] evaluated an inception-v4 convolutional neuronal network (CNN) to classify thyroid healthy tissue and tumors based on *ex vivo* tissue. Segmentation performed on HSI data and RGB images synthesized from the HSI data were compared. A more accurate pixelwise classification was obtained with the HSI data with an area under the curve of 0.90. Recently, Zhang *et al*.[8] evaluated a saliency-weighted method for the selection of the most relevant spectral channels, followed by a multitask U-Net framework to classify liver tumors. Overall sensitivity and specificity values of 94.48% and 87.22% were obtained on 36 *ex vivo* tumors from 19 patients. A limited number of studies evaluated such models for the automatic discrimination of healthy tissue and anatomical structures based on HSI. Schols *et al*.[11] used a support vector machine (SVM) approach to classify *in vivo* nerves and adipose tissue based on 36 gradient and amplitude difference

features computed on the spectra. The leave-one-out cross-validation (LOOCV) performed on the HSI data of 18 patients acquired during thyroid and parathyroid surgery provided a mean accuracy of 95%. A train-test evaluation performed on the HSI data acquired during five carpal tunnel release procedures provided a mean accuracy of 100%. The same working group later evaluated the same classification approach for the discrimination of *in vivo* parathyroid, thyroid, and adipose tissue[14]. They obtained accuracy values of 97% and 82% considering the classification of parathyroid *vs*. adipose tissue and parathyroid *vs*. thyroid, respectively. Maktabi *et al*.[13] evaluated different standard machine learning methods for the automatic segmentation of in vivo thyroid and parathyroid glands and recurrent laryngeal nerve. The best results were obtained using SVM with a radial basis function kernel and LOOCV, where a mean accuracy of 68% was obtained.

In this paper, machine learning methods are combined with HSI to address the tissue segmentation in two medical applications: liver and thyroid image-guided surgery. The goal is the automatic discrimination using HSI of the bile duct from the gallbladder and liver and the parathyroid gland from the thyroid gland, respectively. The bile duct and parathyroid gland are small and thin healthy structures that are hard to identify visually during surgery. Previous studies showed that their spectra are different from surrounding tissue[12]. Seven machine learning models including a CNN were evaluated using a LOOCV and four evaluation efficiency metrics. To our best knowledge, the application for the automatic identification of the bile duct is not reported in the literature yet. Moreover, significant improvements were achieved in the method in comparison to the work of Maktabi *et al*.[13] from the same working group. The data pre-processing includes two additional steps: the removal of spectra outliers and the detection of image background. A principal component analysis (PCA) is performed to reduce the spatial dimension. Finally, spatial information is introduced along with the spectral data to improve tissue segmentation performance.

This paper is organized as follows. The medical HSI database is introduced in the Methods section, along with a description of the preprocessing steps. The Results section presents the statistical results obtained from comparing seven machine learning models combined with three levels of spatial analysis for tissue segmentation in HSI data. Finally, concluding remarks from the computational experiments carried out in the present work are provided in the Discussion section.

METHODS

Patients database
The HSI database used in the present work is formed by 18 HSI data of seven liver surgery patients and 21 HSI data of seven different thyroid surgery patients. The current database was acquired using the TIVITA Tissue hyperspectral imaging system (Diaspective Vision GmbH, Am Salzhaff, Germany)[3]. This system generates HSI data with a spatial resolution of 640 × 480 pixels and a spectral resolution of 5 nm in the electromagnetic wavelength range from 500 to 1000 nm (100 channels). The regions containing tissues of interest have been annotated by a surgeon in both datasets to form the ground truth. The tissues labeled in the datasets involve liver, bile duct, artery, and portal vein for liver surgery and thyroid, parathyroid, and muscle for thyroid surgery. It is important to note out that the ground truth is sparsely labeled because only some regions have been annotated by the surgeon in each image. The main reason is that full labeling of the images is time consuming and exhaustive labor. Figure 1 illustrates two annotated hyperspectral images from each application addressed in this work.

HSI data preprocessing
The proposed preprocessing pipeline used in HSI data prior to the tissue segmentation step is illustrated in Figure 2. The pipeline consists of five main operations involving outlier removal, background identification, spectral smoothing, data normalization, dimensional reduction, and spatial smoothing. Each operation of the pipeline is described in detail below.
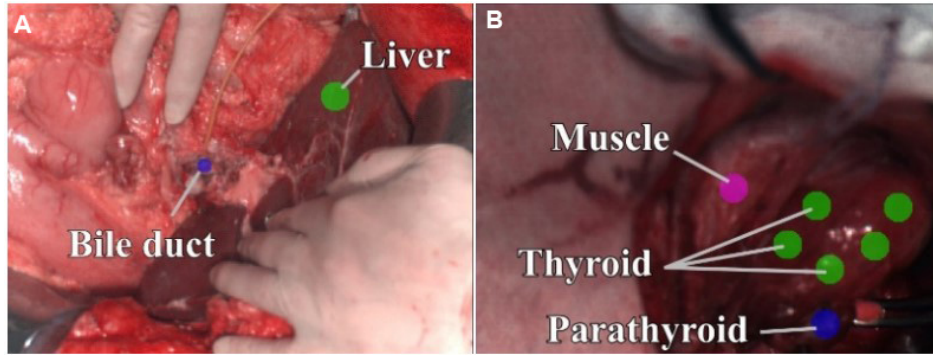
Figure 1. Color representation of hyperspectral images from (A) one liver surgery patient and (B) one thyroid surgery patient, with their respective annotated tissues of interest.
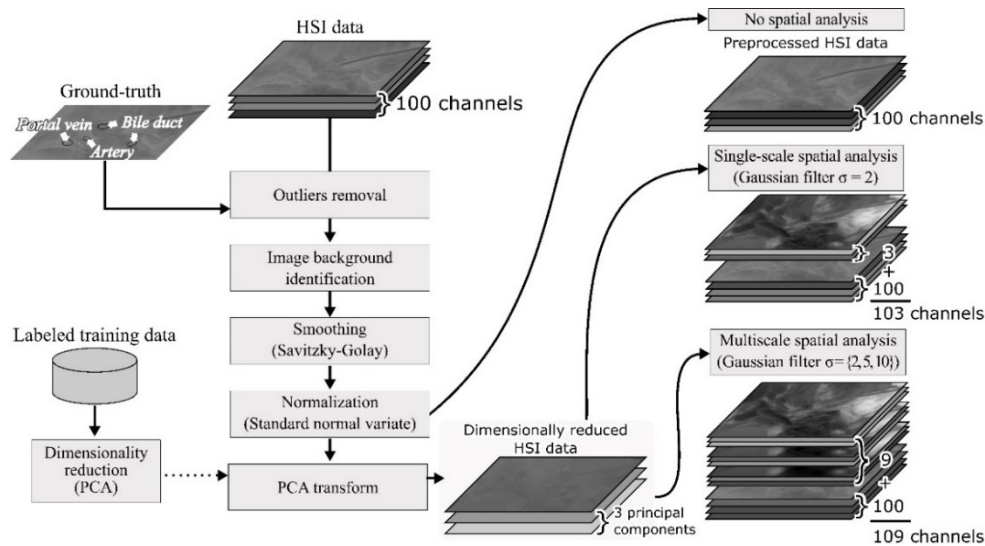

Figure 2. Proposed HSI data preprocessing pipeline.

Outlier removal

A frequent problem present in HSI data is that the spectra pattern of the different tissues can be affected by multiple factors during the acquisition process. Spectral abnormalities can be related to different lighting conditions, and tissues can be covered by blood. Besides, ground truth data can present unintended mislabeled regions that introduce variability to the spectra patterns of the corresponding tissues. In general, those reflectance spectra patterns can be considered as outliers. In the present method, a statistical analysis is carried out to determine reflectance spectra boundaries for the identification of outliers.

The lower boundary $L_k(w)$ and upper boundary $U_k(w)$ corresponding to tissues of class k are defined per wavelength unit w as follows:

$$L_k(w) = max\left(0,\ Q_1\big(R_k(w)\big) - 1.5\ IQR\big(R_k(w)\big)\right) \tag{1}$$

$$U_k(w) =\ Q_3\big(R_k(w)\big) + 1.5\ IQR\big(R_k(w)\big), \tag{2}$$

where Q_1 and Q_3 are the first and third quartiles and *IQR* the interquartile range of the reflectance data R_k for the wavelength unit w.

Background identification
Identification of image background can reduce the number of irrelevant regions falsely detected by a segmentation method as tissues of interest. This is achieved by defining a single set of boundaries that works as threshold for the reflectance spectra of relevant features. The threshold is computed for each wavelength unit w using the maximum of the upper boundaries $U_k(w)$ as $U'(w)$ and the minimum of the lower boundaries $L_k(w)$ as $L'(w)$ of the k classes as follows:

$$U'(w) = max\,(U_k(w)/\forall\,k)$$

(3)

and

$$L'(w) = min\,(L_k(w)/\forall\,k),$$

(4)

A pixel is identified by the mask M_1 as image background if it presents a reflectance pattern outside the $U'(w)$ and $L'(w)$ boundaries for any wavelength unit as follows:

$$M_1(x,y) = \begin{cases} 1 \;\; if\; L'\,(w) \le R(x,y,w) \le U'\,(w), \forall\, w \in [500,1000] \\ 0 \qquad\qquad\qquad\qquad otherwise \end{cases},$$

(5)

where $R(x, y, w)$ is the reflectance corresponding to the wavelength w at the position (x, y).

In addition, by analyzing the reflectance spectra of the HSI data of both applications, three simple rules were defined to identify specific objects in the scene. It was found that relevant tissues in both medical applications show reflectance spectra lower than 0.7 between 520 and 570 nm. Similarly, pixels in dark regions and shadows commonly present reflectance spectra less than 0.1 in the entire acquisition range (500-1000 nm). Moreover, metal instruments commonly present almost constant reflectance spectra in the ranges [520, 570] and [650, 710] nm in both medical applications. In this work, metallic objects are identified if their mean reflectance ratio between range [520, 570] and range [650, 710] nm is greater than 0.9. The three rules defined in this work can be expressed in mask M_2 as follows:

$$M_2(x,y) =$$

(6)

$$\begin{cases} 0 \;\; if\; \dfrac{\bar{R}_{[520,570]}(x,y)}{\bar{R}_{[650,710]}(x,y)} > 0.9 \;or\; \bar{R}_{[520,570]}(x,y) > 0.7 \;or\; \bar{R}_{[500,1000]}(x,y) < 0.1 \\ 1 \qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases},$$

where $\bar{R}_{[520,570]}(x, y)$, $\bar{R}_{[650,710]}(x, y)$, and $\bar{R}_{[500,1000]}(x, y)$ are the mean reflectance values in the ranges [520, 570], [650, 710], and [500, 1000] nm at the position (x, y), respectively. Finally, the image foreground is defined as the union of both masks $M_1(x,y)$ and $M_2(x, y)$ at each position (x, y).

Spectral smoothing and normalization
The Savitzky-Golay smoothing operator[16] is applied to reduce the noise introduced by distinct factors (e.g., noise of HSI camera system) during the acquisition of reflectance spectra. In addition, the standard normal variate transform is used to homogenize the data collected from hyperspectral images acquired under distinct conditions or from different patients. These two operations are applied pixelwise, allowing each hyperspectral image to be preprocessed independently.

Dimensionality reduction and spatial analysis
PCA is used to reduce the HSI dimensionality. This allows working with models of lower complexity for classification tasks[1,6,17]. In this work, PCA is used to extract three principal components from the HSI data, which on average explain more than 95% of the data variance. Subsequently, spatial filtering is applied using a Gaussian filter over each of the resulting three principal components of the HSI data. Additionally, multiscale spatial analysis can be performed using multiple values of the Gaussian filter parameter σ at one time. This allows the detection of anatomical structures of different sizes.
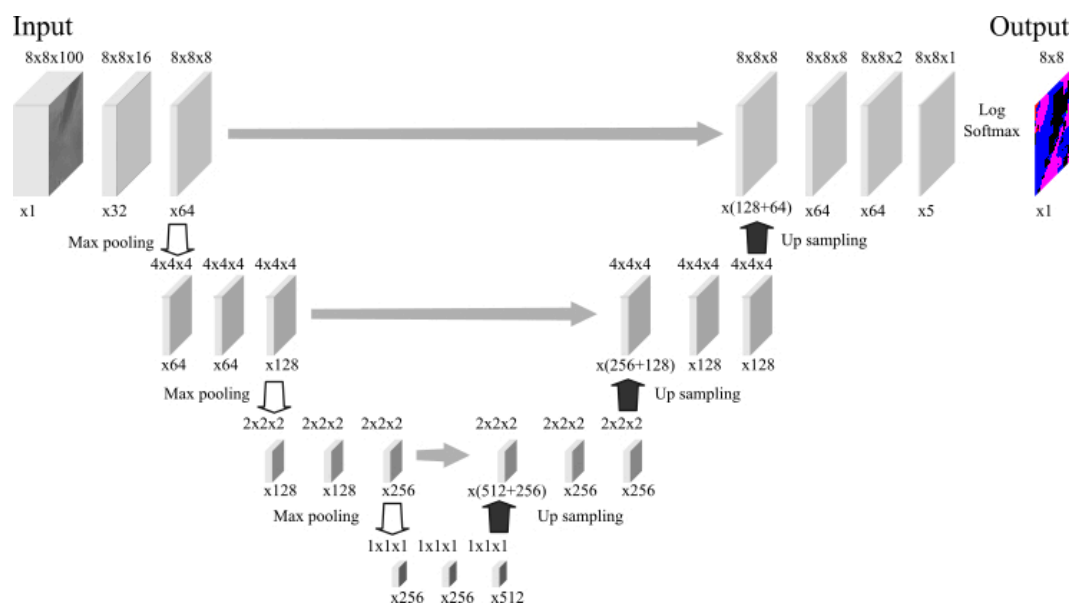
Figure 3. U-Net architecture for tissues segmentation in HSI data.

Tissue segmentation using machine learning

Machine learning models

The seven machine learning models considered in the present work involve logistic regression (LR) with L1, L2, and *elasticnet* regularization techniques, SVM with linear and radial basis function (RBF) kernels, a multilayer perceptron (MLP), and the convolutional neural network U-Net[18]. With the exception of the U-Net, each model is combined with three levels of spatial analysis: no spatial analysis, single-scale analysis ($\sigma = 2$), and multiscale analysis ($\sigma = \{2, 5, 10\}$). The U-Net is excluded from the spatial analysis schemes because it performs an analysis in the spatial domain of the HSI data through its convolution operations. The spatial analysis constructs new features that are concatenated to the original 100 channels corresponding to the spectra of the HSI data for each pixel. The single-scale spatial analysis provides three new features (103 channels), whereas the multiscale analysis produces nine new features (109 channels). The three levels of spatial analysis applied to HSI data are illustrated on the right of Figure 2. The tissues segmentation task is performed pixel-wise, using the complete set of features constructed with each spatial analysis to classify any of the tissue classes of interest.

The logistic regression models and the support vector machine with linear kernel are trained using stochastic gradient descent. The SVM with the RBF kernel is trained using a quadratic programming algorithm. The architecture of MLP consists of 32 neurons in the first hidden layer and 16 neurons in the second hidden layer, which were selected in preliminary experiments. The hyperbolic tangent is used as the activation function in each layer and the *L-BFGS* algorithm is used to fit the artificial neural network parameters.

The U-Net architecture is adapted to work with the high dimensionality of the HSI data. Figure 3 presents the proposed configuration of the U-Net, where the main difference with the original architecture is that two-dimensional convolutions are replaced with three-dimensional ones. Additionally, the U-Net architecture is trained using a patch-based scheme, where the tissues segmentation task is performed using an overlapping sliding window of 8 × 8 pixels, with a stride of 2 pixels. This scheme is used because the number of HSI data available is insufficient to perform the proper training of a deep learning model over complete images. Once trained, the U-Net can be used to segment full images using the convolutional

Table 1. Algorithms and machine learning models used to perform HSI data preprocessing and tissues segmentation

Algorithm/model	Primary use	Description
Outlier removal	HSI preprocessing	Identifies and removes spectra abnormalities in HSI data to avoid their use to train the machine learning models
Background identification	HSI preprocessing	Discriminates background from foreground in HSI data. This allows the reduction of false positives during the tissue segmentation step
Spectral smoothing and normalization	HSI preprocessing	Reduces noise introduced to HSI during data acquisition and simplifies the combination of HSI data obtained from different patients
Dimensional reduction and spatial analysis	HSI preprocessing	Aggregates information, in a low-dimensional space, from the neighborhood of each HSI data pixel. Because this operation introduces spatial context into the feature space, the pixel-wise tissue classification performance of machine learning models can be improved
Linear regression (LR)	Pixel-wise tissue segmentation	Combines the spectra features with a linear transform, followed by a sigmoid function to compute the probability of each pixel belonging to a certain class of tissue. A sparse selection of the spectra features is obtained using the L1 norm regularization, while low sensitivity to changes is achieved with the L2 norm. Moreover, a combination of both behaviors can be obtained by the elasticnet regularization
Support Vector Machine (SVM)	Pixel-wise tissue segmentation	Defines a hyperplane using a set of support vectors, selected from the training data that separates the tissue classes in the features space. A kernel function can be used to map the spectra features into a space where classes could be linearly separable
Multilayer Perceptron (MLP)	Pixel-wise tissue segmentation	Maps the spectra features of each pixel into a space where classes could be linearly separable. This mapping is performed using two layers of connected computing units (neurons). Neurons apply a linear transform to their respective input. A squashing function (e.g., a sigmoid or hyperbolic tangent function) is applied to the neuron responses in order to provide a nonlinear mapping of the feature space
U-Net	Semantic tissue segmentation	Performs the semantic segmentation using a process involving encoding and decoding of the HSI data. The encoding step obtains a high-dimensional, low-resolution representation of the hyperspectral image using a sequence of convolution and downsampling operations. The decoding step upsamples the encoded representation to obtain an image of the same spatial resolution as the original HSI data. A bridge operation transfers information obtained through the encoding step to the decoding step. This allows keeping important features obtained at different resolutions to be used during the decoding step and final segmentation

kernels learned with the patch-based approach. Moreover, the weighted cross-entropy loss function is used for the error assessment of the segmentation result in each patch.

The weight assigned to each coordinate allows considering only labeled regions inside the current patch[19]. For unlabeled pixels, the cross-entropy weight $wce_{ij} = 0$, otherwise it is defined as follows:

$$wce_{ij} = L_x/(C_x C_{i,j}), \quad (7)$$

where L_x is the number of labeled pixels in the patch x, C_x is the number of different classes present in the patch, and $C_{i,j}$ is the number of pixels labeled as the pixel at coordinate (i,j) in the current patch. The contribution of each parameter to the loss error is computed using a back-propagation algorithm. In this work, the *ADAM* method is used to fit the convolutional neural network parameters.

Table 1 presents a summary of the algorithms and machine learning models used in this work. Because the algorithms and models are used to perform different tasks, a brief description and primary use of each of them is provided in this table.

Evaluation metrics
To evaluate the performance of the segmentation models, a leave-one-out cross-validation scheme was adopted in the experiment. For each fold, the training dataset was formed with the HSI data of six patients and validated with the annotated information of the remaining patient. The leave-one-out scheme is illustrated in Figure 4.

The efficiency of the different machine learning models to carry out tissue segmentation in HSI data was evaluated in terms of the following four classification metrics:

$$Accuracy = (TP + TN)/(TP + FP + TN + FN), \quad (8)$$

$$Recall = TP/(TP + FN), \quad (9)$$
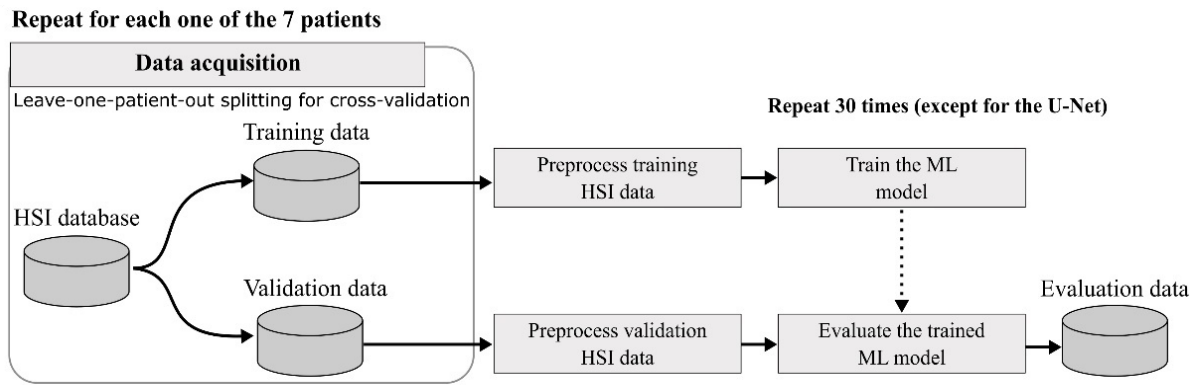
$$Precision = TP/(TP + FP), \quad (10)$$

Figure 4. Schematic representation of the experimental setup.

and

$$F_1 = 2\ TP/(2\ TP + FP + FN), \qquad\qquad (11)$$

where *TP*, *TN*, *FP*, and *FN* are the total of pixels evaluated as true positives, true negatives, false positives, and false negatives according to the annotated regions in the ground truth of the corresponding HSI data, respectively. The metrics were computed for each HSI datum in the validation dataset independently. The overall performance achieved by each segmentation model is defined as the average value obtained for each of the four metrics in the validation dataset. Because the data are sparsely annotated, only the labeled regions in each hyperspectral image can be used to evaluate the segmentation performance. The segmentation of tissues outside the annotated regions can only be assessed qualitatively due to the lack of their respective labels.

The use of stochastic-based training methods introduces randomness to the performance obtained by a segmentation model. To overcome the introduced randomness, a statistical analysis was performed by repeating the training stage for each model 30 times, except for U-Net, for which the training execution time made its repetition prohibitive. In addition, the mean efficiency for the identification of specific tissues was evaluated using the average value obtained for each metric in the complete validation dataset.

RESULTS

This section presents the most relevant experimental results obtained from the statistical comparison of the machine learning models for tissue segmentation in the two medical applications. The segmentation methods were implemented using *Python* programming language and the *Scikit-learn* package[20], on an *Intel Xeon E5*, with 64 GB of RAM and a 2.20 GHz processor. The U-Net was implemented using the *PyTorch* library[21] on a *NVIDIA Tesla K40c* with 12 GB of V-RAM Graphics Processing Unit (GPU).

Tissue segmentation results

The possible outliers were removed from the present dataset by using the feature identification method described in the methods section. The proportion of identified outliers in the annotated HSI data corresponds to 0.58% in liver surgery and 1.47% in thyroid surgery. Figure 5 illustrates the distribution of the reflectance spectra of the predominant classes of tissues, along with the set of feature boundaries *U'(w)* and *L'(w)* computed for each medical application.

Figures 6 and 7 show the results of the proposed background identification method applied on liver and thyroid HSI data, respectively. It can be seen that in both applications most of the instruments, gauze, and the surgeon hands are correctly identified as image background.
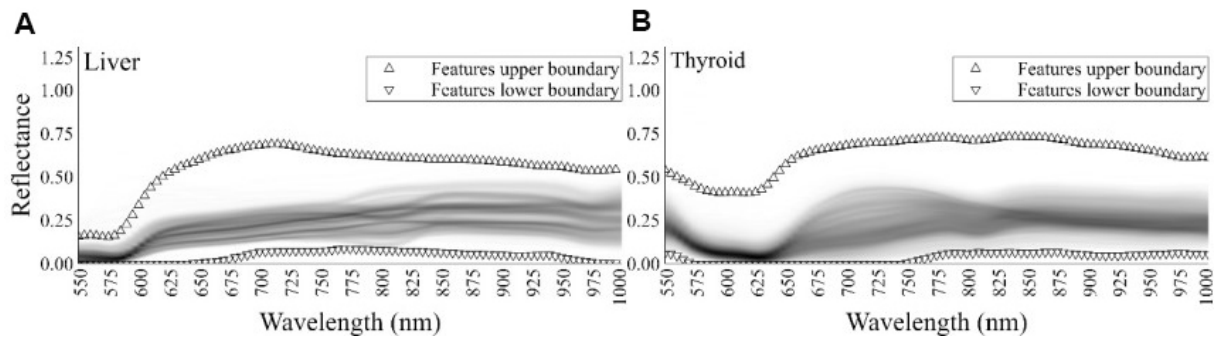
Figure 5. Distribution of the reflectance spectra of (A) liver tissue and (B) thyroid tissue in HSI data.
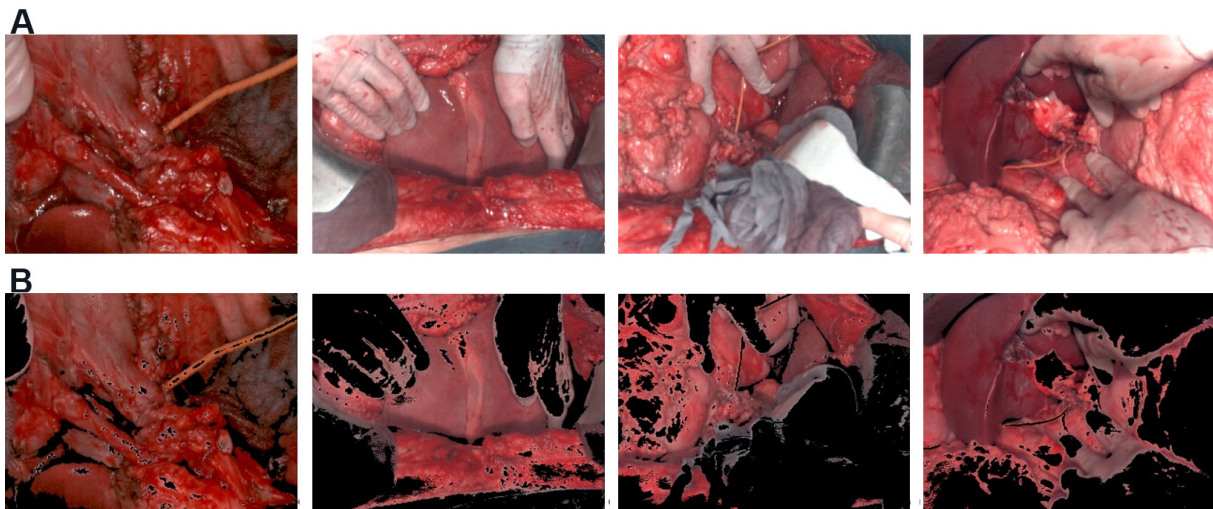

Figure 6. (A) Color representation of four HSI data from liver surgery and (B) their respective foreground identification obtained with the method proposed in this work.

Segmentation of tissues in liver surgery HSI data

The seven machine learning methods combined with three levels of spatial analysis were applied for tissue segmentation in HSI data of liver surgery. The median segmentation efficiency achieved by each model is presented in Table 2. According to the experimental results, the U-Net achieves the highest segmentation efficiency. Table 3 presents the average segmentation efficiency obtained using the U-Net model. The deep learning model can identify liver, bile duct, portal vein, and artery tissues with high accuracy (≥ 0.841). The second best performance was obtained with the SVM model with RBF kernel and combined with the multiscale spatial analysis. SVM is known to provide good results for the segmentation of HSI data[22]. This result demonstrates that the multiscale analysis can classify anatomical structures of different sizes in the data and can accurately extract thin and elongated structures such as the bile duct and blood vessels.

In Figure 8, the segmentation results of the U-Net with HSI data of three patients during liver surgery are illustrated. By visual inspection of Figure 8B, it can be noticed that U-Net provides correct identification of liver, bile duct, portal vein, and artery tissues.

Segmentation of tissues in thyroid surgery HSI data

For tissue identification in HSI data of thyroid surgery, Table 4 presents the median segmentation performance obtained by the seven machine learning models when combined with three levels of spatial analysis. In this application, the logistic regression with the elasticnet regularization (LR elasticnet), combined with multiscale spatial analysis, provides the highest segmentation efficiency.

Table 2. Median segmentation performance of the seven machine learning models combined with three levels of spatial analysis applied to HSI data of liver surgery

Model	Spatial filtering	F1-score	Accuracy	Recall	Precision
U-Net	-	0.815	0.836	0.81	0.908
SVM RBF	σ = {2, 5, 10}	0.763	0.788	0.756	0.917
SVM RBF	-	0.745	0.779	0.734	0.927
SVM RBF	σ = 2	0.656	0.718	0.65	0.886
MLP	-	0.655	0.698	0.652	0.875
MLP	σ = {2, 5, 10}	0.618	0.631	0.597	0.899
LR L2	-	0.617	0.66	0.608	0.857
LR L1	-	0.617	0.663	0.613	0.866
LR elasticnet	-	0.615	0.656	0.606	0.86
SVM	σ = 2	0.589	0.653	0.627	0.816
SVM	σ = {2, 5, 10}	0.58	0.64	0.615	0.833
LR L1	σ = 2	0.57	0.627	0.586	0.877
LR L1	σ = {2, 5, 10}	0.56	0.617	0.571	0.836
SVM	-	0.555	0.632	0.601	0.829
MLP	σ = 2	0.554	0.623	0.537	0.847
LR elasticnet	σ = {2, 5, 10}	0.542	0.606	0.556	0.848
LR L2	σ = {2, 5, 10}	0.542	0.605	0.555	0.845
LR L2	σ = 2	0.537	0.605	0.558	0.866
LR elasticnet	σ = 2	0.537	0.605	0.558	0.865

Table 3. Mean efficiency of the U-Net for identification of tissues in HSI data of liver surgery

Tissue	F1-score	Accuracy	Recall	Precision
Liver	0.841	0.841	0.844	0.981
Bile duct	0.787	0.901	0.849	0.78
Portal vein	0.702	0.855	0.803	0.774
Artery	0.567	0.854	0.510	0.848
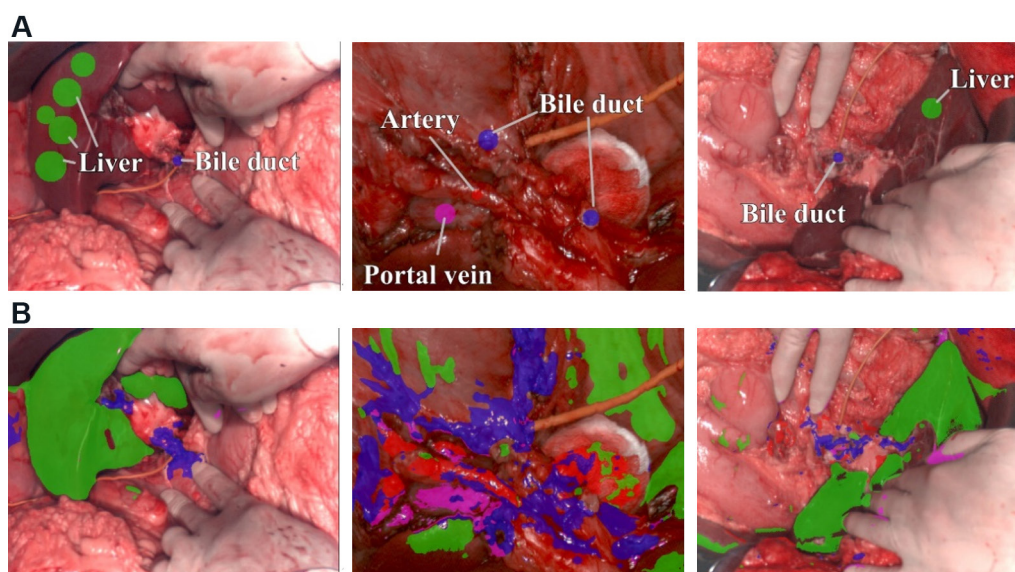
Figure 8. (A) Color representation of the HSI data of three patients during liver surgery, with the annotated tissues of interest; and (B) the resulting tissue segmentation provided by U-Net (liver in green, bile duct in blue, portal vein in pink, and artery in red).

In Table 5, the discrimination performance of this model is presented. The experimental results show that thyroid and parathyroid glands and muscle tissues can be detected in HSI data with high accuracy (≥ 0.696). The lower precision on the detection of parathyroid glands is related to a relatively higher number

Table 4. Median segmentation performance of the seven machine learning models combined with three levels of spatial analysis applied to HSI data of thyroid surgery

Model	Spatial filtering	F1-score	Accuracy	Recall	Precision
LR elasticnet	σ = {2, 5, 10}	0.673	0.803	0.675	0.825
U-Net	-	0.668	0.811	0.674	0.845
LR L2	σ = {2, 5, 10}	0.664	0.796	0.666	0.822
SVM	σ = {2, 5, 10}	0.664	0.791	0.675	0.811
LR L1	σ = {2, 5, 10}	0.661	0.799	0.665	0.819
MLP	σ = 2	0.639	0.755	0.641	0.798
SVM	σ = 2	0.62	0.766	0.637	0.811
LR L2	σ = 2	0.618	0.763	0.629	0.827
LR elasticnet	σ = 2	0.618	0.763	0.63	0.825
SVM RBF	σ = {2, 5, 10}	0.598	0.751	0.611	0.815
LR elasticnet	-	0.598	0.755	0.616	0.804
LR L2	-	0.597	0.756	0.615	0.802
SVM	-	0.597	0.759	0.615	0.797
MLP	σ = {2, 5, 10}	0.595	0.755	0.603	0.775
MLP	-	0.583	0.731	0.586	0.759
SVM RBF	-	0.57	0.734	0.596	0.763
LR L1	σ = 2	0.57	0.731	0.592	0.809
SVM RBF	σ = 2	0.566	0.729	0.589	0.754
LR L1	-	0.558	0.723	0.584	0.798

Table 5. Mean efficiency of the logistic regression with the elasticnet regularization for identification of tissues in HSI data of thyroid surgery

Tissue	F1-score	Accuracy	Recall	Precision
Thyroid	0.663	0.888	0.657	0.75
Parathyroid	0.476	0.696	0.591	0.499
Muscle	0.524	0.873	0.53	0.725

of false positives than true positives in the segmentation results. U-Net provided slightly lower median values of the F1-score and recall than those of the LR model with elasticnet but larger median values of accuracy and precision. The performance of U-Net and LR elasticnet are similar. Moreover, the second best performances were obtained with different models but using the multiscale analysis. This shows again the relevance of the multiscale analysis.

The tissue segmentation results obtained by the LR elasticnet with multiscale spatial analysis are presented in Figure 9 for the HSI data of three patients during thyroid surgery. Visual inspection of the results shows that this model provides correct identification of thyroid, parathyroid, and muscle tissues. Misclassification of image background regions as thyroid and parathyroid glands can be attributed to the similarity of their spectra after the standard normal variate normalization.

Table 6 presents a statistical analysis of the execution time taken by the segmentation models considered in this work to segment a single hyperspectral image. In this table, the different models appear only once because their execution time is similar when applied in both applications. The only exceptions are the models based on SVM and the SVM with RBF kernel, the execution times of which varied according to the number of support vectors required during the segmentation process. The multiscale analysis increases slightly the running time, but it remains lower than 0.5 s. The SVM model with RBF kernel is the exception and requires longer computing time, between 4 and 14 min. The running time of U-Net is nearly 5 s, which is acceptable for its application during surgical interventions.

U-Net, which performs tissue segmentation in 4.954 (± 0.030) s when executed using a GPU, and the LR elasticnet with multiscale spatial analysis, which provides tissue segmentation in 0.393 (± 0.035) s, presented the highest tissue segmentation efficiency for HSI data for liver and thyroid surgery, respectively.
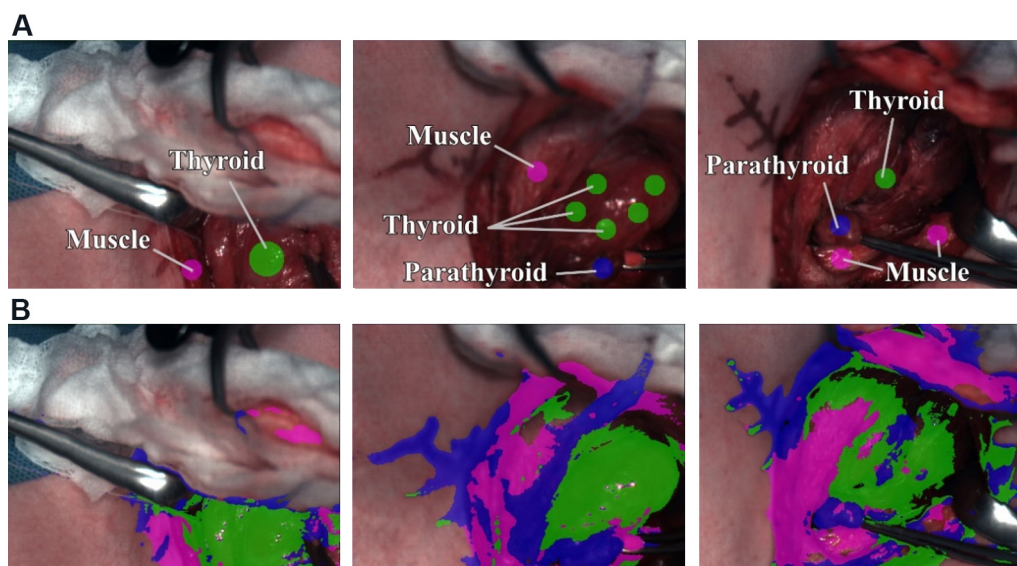
Figure 9. (A) Color representation of the HSI data of three patients during thyroid surgery, with the annotated tissues of interest; and (B) the resulting tissue segmentation provided by the logistic regression with the elasticnet regularization combined with multiscale spatial analysis (thyroid in green, parathyroid in blue, and muscle in pink).

Table 6. Average execution time per image of the seven models combined with three levels of spatial analysis applied to HSI data

Model	Spatial filtering	Execution time (s)
SVM (Thyroid)	-	0.078 (± 0.017)
LR L1	-	0.083 (± 0.017)
LR elasticnet	-	0.084 (± 0.013)
LR L2	-	0.084 (± 0.015)
SVM (Liver)	-	0.085 (± 0.009)
MLP	-	0.225 (± 0.057)
LR L2	$\sigma = 2$	0.235 (± 0.024)
SVM (Thyroid)	$\sigma = 2$	0.241 (± 0.031)
LR elasticnet	$\sigma = 2$	0.241 (± 0.024)
LR L1	$\sigma = 2$	0.242 (± 0.026)
SVM (Liver)	$\sigma = 2$	0.242 (± 0.020)
MLP	$\sigma = 2$	0.379 (± 0.069)
SVM (Thyroid)	$\sigma = \{2, 5, 10\}$	0.390 (± 0.038)
SVM (Liver)	$\sigma = \{2, 5, 10\}$	0.393 (± 0.035)
LR elasticnet	$\sigma = \{2, 5, 10\}$	0.393 (± 0.035)
LR L2	$\sigma = \{2, 5, 10\}$	0.394 (± 0.034)
LR L1	$\sigma = \{2, 5, 10\}$	0.396 (± 0.035)
MLP	$\sigma = \{2, 5, 10\}$	0.555 (± 0.123)
U-Net	-	4.954 (± 0.030)
SVM RBF (Thyroid)	$\sigma = \{2, 5, 10\}$	240.669 (± 67.069)
SVM RBF (Thyroid)	$\sigma = 2$	294.315 (± 99.520)
SVM RBF (Thyroid)	-	325.903 (± 98.614)
SVM RBF (Liver)	$\sigma = \{2, 5, 10\}$	574.009 (± 89.451)
SVM RBF (Liver)	$\sigma = 2$	645.721 (± 150.838)
SVM RBF (Liver)	-	823.531 (± 123.488)

DISCUSSION

This work presents novel approaches using a combination of machine learning methods and multiscale spatial analysis to perform tissue segmentation on two medical applications. These are the identification of healthy anatomical structures in HSI data acquired intraoperatively during liver and head and neck surgeries. Bile duct and parathyroid glands are thin and small structures that can be hard to be detected

Table 7. Description of the best classification performance results obtained by existing works in the field of HSI in liver and thyroid surgery

Segmentation model	Examined tissue	Spectral range of the HSI system	Patient data	Evaluation method	Classification performance
Inception-v4 CNN (Halicek et al.[9])	*Ex vivo* tumors of the thyroid	450-900 nm	200 tumors of 76 patients	LOOCV	AUC of 0.90
Multitask U-Net framework (Zhang et al.[8])	*Ex vivo* liver tumors	400-1000 nm	36 tumors of 19 patients	LOOCV	Overall sensitivity and specificity of 94.48% and 87.22%
SVM with polynomial kernel (Schols et al.[11])	Intraoperative *in vivo* nerve and adipose tissue	900-1700 nm	18 patients	LOOCV (thyroid and parathyroid surgery) Train-test (carpal tunnel release procedure)	Mean accuracy: 95% (LOOCV) and 100% (train-test)
SVM with polynomial kernel (Schols et al.[14])	Intraoperative *in vivo* parathyroid, thyroid and adipose tissue	350-1830 nm	19 patients	LOOCV	Mean accuracy and sensitivity for the parathyroid-adipose classification of 97% and 100%, and for the parathyroid-thyroid classification of 82% and 86%
SVM with RBF kernel (Maktabi et al.[13])	Intraoperative *in vivo* parathyroid, thyroid and nerve	500-1000 nm	9 patients	LOOCV	Mean patient accuracy including the three tissues: 68% F_1 value per patient for the classification of the parathyroid: 0.17 M F1 M 0.71
U-Net and logistic regression (LR) with multiscale spatial analysis (Our method)	Intraoperative *ex vivo* liver, bile duct and blood vessels Intraoperative *in vivo* parathyroid, thyroid and muscle	500-1000 nm	18 HSI data from 7 patients (liver surgery) and 21 HSI data of 7 patients (liver surgery)	LOOCV	Mean accuracy and F1-scores of 0.901 and 0.787 for the bile duct and of 0.696 and 0.476 for the parathyroid

LOOCV: Leave-one-out cross-validation; RBF: radial basis function; AUC: area under the curve.

visually. Computer-assisted tools can support the surgeon to automatically identify them. Different machine learning methods combined with a multiscale spatial analysis were evaluated. The best performances were obtained with U-Net and the LR model with elasticnet regularization technique. A leave-one-patient-out cross-validation provided median values of the accuracy and F1-score of 0.901 and 0.787 for the segmentation of the bile duct and 0.696 and 0.476 for the segmentation of the parathyroid. For comparison, Table 7 provides the best results of existing studies that performed tissue segmentation based on HSI data in liver and head and neck surgery. The first three studies concern the classification of tumors and nerve, which are not addressed in this paper. Schols et al.[14] obtained robust results for the segmentation of parathyroid, thyroid, and adipose tissue using a SVM model. The HSI system used is a fiber probe equipped with two sensors, allowing the acquisition of wide-band reflectance spectra (350-1830 nm). This explains the discrepancies with our results. Maktabi et al.[13] obtained mean F1-scores and recall computed for each patient of the same dataset of 0.17-0.71 and 0.35-0.93, respectively, for the segmentation of the parathyroid using a SVM model with RBF kernel and a leave-one-patient-out cross-validation. Moreover, the model achieved a mean accuracy of 0.68 including the segmentation of thyroid, parathyroid, and nerve. These results are difficult to compare because, in contrast to Maktabi et al.[13], who presented the results by patient, the performance of each segmentation model was computed overall for all patients in this study. In comparison, we obtained an overall accuracy of 0.803 for the segmentation of thyroid, parathyroid, and muscle. Nevertheless, the multiscale analysis, which was added to the standard SVM, LR, and MLP methods, seems to improve the identification of small structures such as the parathyroid gland. The benefit of this additional analysis is clearer for the segmentation of the bile duct. U-Net, which performs a similar analysis through the convolution operations, provided good results as well. The improvement of the segmentation efficiency obtained by introducing the multiscale spatial analysis could be attributed to the fact that those tissues can be detected based on their apparent size in the scene. Therefore, machine learning models combined with HSI represent a promising method to perform image-guided surgery considering that the most relevant tissues were correctly detected in both medical applications.

These promising results could still be improved. For some tissues, most of the segmentation models still present low-performance accuracy (e.g., parathyroid in thyroid surgery and arteries in liver surgery HSI data). This could be associated with different factors such as low representation of these tissues in the dataset and high similarity of their reflection spectra with other tissues. The high rate of false positives of the parathyroid tissues can be reduced by using more robust background identification methods. Moreover, because distinct tissues can be indistinguishable after the standard normal variate transform, different alternatives are necessary to perform normalization of HSI data acquired from distinct patients. The limitations of the proposed method are mostly related to extrinsic factors, such as the unintended mislabeling in the ground truth data. The problems related to intrinsic factors, such as the intra- and inter-patient variability of the tissue composition, can be addressed by considering larger datasets of annotated HSI data.

In preliminary experiments, other convolutional neural network architectures were tested to perform tissues segmentation in HSI data. For this approach, the segmentation was performed using a neighborhood of each pixel in the HSI data. However, because of the small size of the database and the high dimensionality of the HSI data, those methods presented lower segmentation performance than the non-deep learning models. Consequently, it is expected that, by using larger datasets of patients for training purposes, the detection efficiency of the machine learning models can be improved.

Conclusions

In this paper, the segmentation of different tissues in HSI data is addressed using machine learning models. Two medical applications are considered: liver and bile duct segmentation in liver surgery and thyroid and parathyroid segmentation in thyroid surgery. In a preprocessing step, spectra pattern outliers are identified and removed from the training data. The high dimensionality of HSI data is reduced using PCA, the presence of noise is reduced with the Savitzky-Golay operator, and the spectra are normalized by the standard normal variate transform. In addition, a statistical analysis of the tissue segmentation efficiency of seven machine learning models combined with three levels of spatial analysis of the HSI data is carried out. According to the computational experiments using 18 HSI data from seven liver surgery patients and 21 HSI data from seven thyroid surgery patients, the U-Net model and the logistic regression with the elasticnet regularization combined with multiscale spatial analysis achieved the highest tissue segmentation performance for each medical application, respectively. Finally, the results obtained in this work suggest that hyperspectral imaging combined with machine learning is suitable for performing tissue segmentation for image guided surgery applications.

DECLARATIONS

Authors' contributions
Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Cervantes-Sanchez F, Maktabi M, Chalopin C
Performed data acquisition, as well as technical, and material support, and data interpretation: Köhler H, Sucher R, Rayes N
Performed administrative support, and made substantial contribution to the conception and design of the study: Cruz-Aceves I, Avina-Cervantes JG

Availability of data and materials
Data is not available due to that the patients of this study did not agree for their data to be shared publicly.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
The procedures involving human participants in this study were performed with the ethical standards of the institutional and/or national research committee, and in accordance with the Declaration of Helsinki. The experimental hyperspectral measurements from patients have obtained the approval of the Ethics Committee of the University Hospital of Leipzig under 026/18-ek. All study participants provided informed consent.

Consent for publication
The individuals participating in this study have expressed their informed consent for the publication of their data in this work.

Copyright
© The Author(s) 2021.

REFERENCES

1. Fabelo H, Halicek M, Ortega S, et al. Deep learning-based framework for *in vivo* identification of glioblastoma tumor using hyperspectral images of human brain. *Sensors (Basel)* 2019;19:920.
2. Holmer A, Tetschke F, Marotz J, et al. Oxygenation and perfusion monitoring with a hyperspectral camera system for chemical based tissue analysis of skin and organs. *Physiol Meas* 2016;37:2064-78.
3. Kulcke A, Holmer A, Wahl P, Siemers F, Wild T, Daeschlein G. A compact hyperspectral camera for measurement of perfusion parameters in medicine. *Biomed Tech (Berl)* 2018;63:519-27.
4. Ortega S, Fabelo H, Iakovidis DK, Koulaouzidis A, Callico GM. Use of hyperspectral/multispectral imaging in gastroenterology. Shedding some-different-light into the dark. *J Clin Med* 2019;8:36.
5. Baltussen EJM, Kok END, Brouwer de Koning SG, et al. Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery. *J Biomed Opt* 2019;24:1-9.
6. Nouri D, Lucas Y, Treuillet S. Hyperspectral interventional imaging for enhanced tissue visualization and discrimination combining band selection methods. *Int J Comput Assist Radiol Surg* 2016;11:2185-97.
7. Shapey J, Xie Y, Nabavi E, et al. Intraoperative multispectral and hyperspectral label-free imaging: a systematic review of in vivo clinical studies. *J Biophotonics* 2019;12:e201800455.
8. Zhang Y, Yu S, Zhu X, et al. Explainable liver tumor delineation in surgical specimens using hyperspectral imaging and deep learning. *Biomed Opt Express* 2021;12:4510.
9. Halicek M, Dormer JD, Little JV, Chen AY, Fei B. Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning. *Biomed Opt Express* 2020;11:1383-400.
10. Halicek M, Fabelo H, Ortega S, Callico GM, Fei B. In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: revealing the invisible features of cancer. *Cancers (Basel)* 2019;11:756.
11. Schols RM, ter Laan M, Stassen LP, et al. Differentiation between nerve and adipose tissue using wide-band (350-1,830nm) in vivo diffuse reflectance spectroscopy. *Lasers Surg Med* 2014;46:538-45.
12. Cooney GS, Barberio M, Diana M, Sucher R, Chalopin C, Köhler H. Comparison of spectral characteristics in human and pig biliary system with hyperspectral imaging (HSI). *Current Directions in Biomedical Engineering* 2020;6:20200012.
13. Maktabi M, Köhler H, Ivanova M, et al. Classification of hyperspectral endocrine tissue images using support vector machines. *Int J Med Robot* 2020;16:1-10.
14. Schols RM, Alic L, Wieringa FP, Bouvy ND, Stassen LP. Towards automated spectroscopic tissue classification in thyroid and parathyroid surgery. *Int J Med Robot* 2017;13:e1748.
15. Signoroni A, Savardi M, Baronio A, Benini S. Deep learning meets hyperspectral image analysis: a multidisciplinary review. *J Imaging* 2019;5:52.
16. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 1964;36:1627-39.
17. Chen Y, Jiang H, Li C, Jia X, Ghamisi P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans Geosci Remote Sensing* 2016;54:6232-51.
18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical image computing and computer-assisted intervention - MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41.
19. Bokhorst JM, Pinckaers H, van Zwam P, Nagtegaal I, van der Laak J, Ciompi F. Learning from sparsely annotated data for semantic segmentation in histopathology images. In: Cardoso MJ, Feragen A, Glocker B, Konukoglu E, Oguz I, Unal G, et al., editors. London, United Kingdom: PMLR; 2019. p. 84-91.

20. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825-30.

21. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8026-37.

22. Ghamisi P, Plaza J, Chen Y, Li J, Plaza AJ. Advanced spectral classifiers for hyperspectral images: a review. *IEEE Geosci Remote Sens Mag* 2017;5:8-32.