

Deep Learning Analysis of in Vivo Hyperspectral Images for Automated Intraoperative Nerves Detection

Manuel Barberio (✉ manuel.barberio@ircad.fr)

IRCAD

Toby Collins

IRCAD

Valentin Bencteux

IHU-Strasbourg

Richard Nkusi

IRCAD Africa

Eric Felli

IHU-Strasbourg

Massimo Viola

Ospedale Card. G. Panico, Tricase (LE)

Jacques Marescaux

IRCAD

Alexandre Hostettler

IRCAD

Michele Diana

IRCAD

Research Article

Keywords: Vivo Hyperspectral, Intraoperative Nerves, Convolutional Neural Network

Posted Date: April 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-393233/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Nerves are difficult to recognize during surgery and inadvertent injuries may occur, bringing catastrophic consequences for the patient. Hyperspectral imaging (HSI) is a non-invasive technique combining photography with spectroscopy, allowing biological tissue property quantification. We show for the first time that HSI combined with deep learning allows nerves and other tissue types to be automatically recognized in-vivo at the pixel level. An animal model is used comprising eight anesthetized pigs with a neck midline incision, exposing several structures (nerve, artery, vein, muscle, fat, skin). State-of-the-art machine learning models have been trained to recognize these tissue types in HSI data. The best model is a Convolutional Neural Network (CNN), achieving an overall average sensitivity of 0.91 and specificity of 0.99, validated with leave-one-patient-out cross-validation. For the nerve, the CNN achieves an average sensitivity of 0.76 and specificity of 1.0. In conclusion, HSI combined with a CNN model is suitable for in vivo nerve recognition.

Introduction

Minimally invasive surgery offers obvious advantages to the patients, including faster postoperative recovery and shorter hospital stay. To improve the surgical performances, intraoperative imaging modalities might enhance the surgical vision and assist the surgeon in the tasks of recognizing important anatomical structures that need to be safeguarded or selectively removed.

Specifically, nerves are fundamental structures that need to be preserved and an inadvertent nerve injury might result in an important function loss with a consistent negative impact on the patient's quality of life. Iatrogenic nerve injury might occur during minimally invasive digestive surgical procedures. Despite the technical improvements in high-definition laparoscopic systems, nerves remain difficult to visualize. In particular, after laparoscopic colorectal resections the incidence of fecal incontinence resulting from damage to the pelvic autonomic nerves remains high with a rate of 40-60% [1].

Precise nerve identification is a challenging task in both open and minimally invasive surgery. For example, the identification of the recurrent laryngeal nerve is a key step during open thyroidectomies. Although permanent vocal cord palsy caused by a laryngeal recurrent nerve injury is rare, ranging from 0.5 to 1.5%, it still remains a considerable burden, considering that thyroidectomy is one of the most frequently performed surgical procedures worldwide [2].

The difficulty of identifying nerves intraoperatively is principally given by their small size and tubular shape. Furthermore, in their original anatomical positions, nerves are often embedded into a sheet of connective and fatty tissue, together with blood vessels: the neurovascular bundle.

An intraoperative guidance tool able to precisely and objectively discriminate and identify the nerves would provide an important assistance to the surgeon to prevent iatrogenic injuries during open and laparoscopic operations. In this view, hyperspectral imaging (HSI) represents a promising technology.

HSI combines a photo camera and a spectrometer, providing both spatial and spectral information of the analyzed structures. The tissue/light interaction (reflectance, absorption, scattering) generates specific spectral signatures, theoretically enabling for an optical biopsy and consequently discrimination of structures with different biochemical composition. However, HSI produces a massive amount of data and in order to extract and classify the discriminative information, machine learning (ML) techniques are required. Likewise, HSI could be easily integrated into current surgical optical navigation tools, such as laparoscopes[3], endoscopes[4] or even robotic systems, hence it has captured the attention of the surgical community as a potential innovative guidance tool. Its usefulness in the biomedical field has been already extensively proven[5]. HSI technology has been previously applied in digestive surgery to quantify intestinal perfusion before anastomosis during several procedures[6-9], as well as in case of mesenteric ischemia[10, 11], or to quantify liver perfusion[12]. A number of previous works focused successfully on the ability of HSI to discriminate between normal and tumoral tissue, in particular prostate cancer[13], colorectal cancer[14, 15], gastric cancer[16, 17], glioblastoma[18] and head and neck cancers[19-22]. In the oncological field, advances in hyperspectral signature classification have been remarkable and lead to the successful use of sophisticated deep learning algorithms[16-19, 21, 22]. However, these previous works have been conducted *ex vivo*: either using surgical specimens or histopathological slides. Furthermore, only binary tissue recognition has been carried out: cancer versus the cumulative non-cancer class. Several groups have attempted to employ HSI as surgical navigation tool for discriminating intraoperatively key anatomical structures, such as bile duct[23], esophagus or tracheal tissue[24], parathyroid gland[25], nerve[26] and ureter[27]. However, those previous tissue recognition works were conducted using either simple feature discrimination algorithms[23-25] or band selection methods[26, 27].

The aim of this study is to apply HSI in combination with several advanced machine-learning algorithms, in particular convolutional neural networks (CNNs) and Support Vector Machines (SVMs) to differentiate the nerve from a number of surrounding classes of tissues in an *in vivo* porcine model. These classes are artery, fat, muscle, skin and vein. We also include metallic objects (instruments) as an additional class to recognize by the ML models.

Results

2.1 Model Configurations

We compare performance of different models to measure performance across two axes. The first axis is model type (SVM with radial basis kernel versus CNN), and the second is with and without Standard Normal Variate (SNV) normalization. This produces 4 combinations, named as follows:

1. **CNN**: CNN trained without SNV normalization
2. **CNN+SNV**: CNN trained with SNV normalization
3. **SVM**: SVM trained without SNV normalization
4. **SVM+SNV**: SVM trained with SNV normalization

These were selected to measure prediction performance across the two model classes (SVM and CNN), and to determine the impact of SNV normalization on each model. SNV normalization is commonly used to reduce variation in imaging conditions e.g. surface orientation with respect to the camera, yet is also represents a loss of hyperspectral information. It is not clear whether normalization methods such as SNV would benefit or harm in vivo tissue recognition from HSI data using machine learning models. We evaluate performance with confusion matrices, Receiver Operator Curves (ROCs) and three performance statistics: sensitivity, specificity and Sørensen–Dice coefficient score (DCS). DCS is commonly used to evaluate classification accuracy of machine learning models at the pixel level in the medical imaging community and it is equivalent to the F1 score and represents the harmonic mean of a classifier's positive predictive value and its sensitivity.

2.2 Confusion Matrices

The confusion matrices give a compact representation of multi-class classification performance in **Figure 1**. One confusion matrix is given per model, and an item (i,j) in each matrix gives the proportion of pixels of class i that are predicted to be of class j . The confusion matrices are normalized such that each row adds to 1. Consequently, the diagonal entries give the true positive rate (TPR) of each class, equivalently its sensitivity. The off-diagonal entries (i,j) with $j>i$ give the false positive rate (FPR, Type 1 error) for class i with respect to class j . The off-diagonal entries (i,j) with $i>j$ give the false negative rate (FNR, Type 2 error) for class i with respect to class j . A confusion matrix of a perfectly performing model has zeros everywhere except ones along the diagonal.

The CNN model achieves a sensitivity of 89.4% and above for all tissue classes except for nerve, which has a sensitivity of 76.3%. Skin tissue has the lowest class confusion among all tissue classes with a sensitivity of 99.7%. Considering the CNN versus the SVM model, the CNN achieves a much higher sensitivity for fat (93.6% versus 78.4%), nerve (76.3% versus 56.4%) and vein (96.1% versus 80.9%) tissues. The remaining classes (artery, skin, muscle and metal) have similar sensitivities to within 3% for either model. These results indicate that SNV normalization has in general *harmed* the performance of the CNN. Instead of improving performance thanks to a reduction of intra-class variability (for which SNV normalization is designed), it has had harmed performance of the CNN because it filters out useful information that can be exploited by the CNN via deep learning. One can see that this lost information is useful to the CNN to obtain good classification performance especially for nerve and vein tissues. Consequently, the CNN model has learned useful spatio-spectral features via deep learning that are able to handle inter-class variability without SNV normalization. In contrast, the benefit of SNV normalization on the SVM model is mixed. SNV normalization improves the sensitivity of the vein class with the SVM model (80.9% versus 87.8%) but for the other classes the sensitivity is similar and within 2% for either model. Therefore, the SVM, which does not learn deep features, can sometimes achieve better results by reducing intra-class variation with SNV normalization, but there is not a systematic improvement for all tissue classes.

2.3 Performance Visualizations

The outputs of the best model (CNN) are visualized in **Figure 2**. This figure is arranged into rows and columns as follows. Each row corresponds to a different subject and there are four images per row. From left-to-right these are as follows: (a) the RGB simulated from the HSI image, (b) the corresponding ground-truth tissue annotations provided by the surgeon (overlaid on greyscale images for better visibility), (c) the predicted tissue annotations computed by the model, and (d) an error map visualization. The error map shows where predictions from the model deviated with the ground-truth annotations. Pixels in the error map that are black indicate perfect agreement between the predicted and ground-truth tissue classes. Colored pixels in the error map indicate a prediction error, and a color is provided of the class that was incorrectly predicted. A perfectly performing model has a completely black error map. When a mistake is made by the model at a spatial location, we visualize the incorrectly predicted class.

For all subjects we clearly see excellent recognition performance for skin, subcutaneous fat, carotid artery, jugular vein and metal classes, with very low errors. For muscle and vagal nerve classes, we observe strong performance for subjects 1, 2, 4, 5, 7 and 8 with few errors (indicated by mostly black pixels in the error visualization maps). There are two main sources of error: in subject 3 nerve has been mistaken for muscle, and subject 6 muscle has been mistaken for nerve and artery.

2.4 Further Quantitative Analysis and Statistics

Model performance is further analyzed with Receiver Operator Characteristic (ROC) curves and other performance metrics: Dice Similarity Coefficient (DSC), sensitivity and specificity. A ROC curve measures the performance of a classifier in terms of false positive and true positive rates. These are generated for each model and for each tissue class as follows. For each class, the prediction scores for all pixels in the test images are computed, then the false positive and true positive rate is computed using a varying detection threshold. For a lower threshold, a higher false positive rate is obtained, but at the cost of a lower true positive rate. The ROC curve is generated by sweeping the detection thresholds from low (true positive rate of zero) to high (true positive rate of one). ROC curves are plotted for the four models in **Figure 3**. High recognition performance is indicated by a high Area Under Curve (ROC-AUC), ROC-AUC summarizes the balance between false positives and true positives, and it is also insensitive to high class imbalance, which is the case with this dataset where some classes e.g. skin are significantly more represented than others e.g. nerve. ROC-AUC is also equivalent to the probability that a randomly selected sample will have a higher classification score for its true class compared to another classes. For the CNN model, all ROC-AUCs are above 0.99, which is considered outstanding performance[28]. For the other models, ROC-AUCs are similar but the ROC-AUC for the nerve class with SVM and SVM+SNV models are consistently lower compared to CNN and CNN+SVM.

DCS (also called the Dice coefficient or F1 score), sensitivity and specificity performance of the four models are plotted in **Figure 4**. These are commonly used performance metrics used in the medical image processing literature. DCS gives the harmonic mean of a classifier's precision and sensitivity[29]. Solid bars represent the average performance metric for each model and for each class, computed by averaging the performance metric across each of the 8 images. Error bars represent one Standard Error.

Statistically significant differences were computed with a two-tailed paired t-test ($\alpha=0.05$) similarly to previous works in HSI classification analysis[30]. Significance is indicated by the bracketed bars in Figure 4 with one star for $p<0.05$ and two stars for $p<0.01$. Regarding DSC, the CNN model has generally a better average DSC compared to CNN+SNV, but the difference is not statistically significant. In contrast, the CNN model obtains significantly better DSC performance compared to SNV for the nerve ($p=0.010$) and vein ($p=0.044$) classes. No other models (CNN+SNV, SVM, SVM+SNV) obtain significantly better DSC compared to the CNN model. Concerning sensitivity, there are no significant differences between the CNN and CNN+SNV models. In contrast, the CNN model has significantly better sensitivity compared the CNN and CNN+SNV models for the nerve class ($p=0.0038$ and $p=0.0011$) respectively. Concerning specificity, there is no significant difference between the CNN and CNN+SNV models. The CNN model has significantly better specificity compared to the SVM model for skin tissue ($p=0.021$) and better specificity compared to the SVM+SNV model for muscle tissue ($p=0.036$). DSC, sensitivity and specificity were averaged across all tissue classes, shown in the right-most bars in Figure 4. The CNN model obtains the highest average scores for all three metrics.

Discussion

In the current work, despite the relatively small sample size (8 pigs), we were able to recognize 6 different tissue classes with a high degree of accuracy using HSI in combination with advanced machine learning algorithms in an *in vivo* porcine model. CNN proved to be the best performing model. There are three key advantages of CNNs that make them very popular for image classification: 1) they automatically learn a hierarchical representation of the data tailored to the specific classification task at hand. Unlike classical methods such as SVMs or decision trees, the CNN learns the best features from the raw image data. This eliminates tedious feature engineering tasks. 2) CNNs are general purpose and highly flexible, and model complexity can be modified with different CNN designs. 3) CNNs can be trained efficiently using highly parallelized implementations such as Pytorch[31]. The main disadvantage of CNNs is that they typically have many parameters to learn (several millions is not uncommon), and this can make them susceptible to poor performance if training data is limited. Various techniques exist to overcome this, including using smaller models, simulating additional training data (data augmentation), training by randomly deactivating neurons (drop-out), or by reusing parameters from a CNN trained on a related task with much more data (transfer learning). The CNN model used in this study is relatively small with 32,628 parameters and we were able to train well-performing models that in general outperformed the SVM models. Consequently, it is possible to solve multi-class tissue classification with HSI using a relatively small CNN, and therefore to benefit from automatic discovery of relevant spatio-spectral features within the HSI data using deep learning.

Hyperspectral imaging is a powerful tool enabling to unravel the invisible in an objective way, relying on curves obtained by the interaction of the emitted light with the biochemical tissue components. Given the contrast-free and non-invasive nature of this versatile technology, it has the potential of becoming a future intraoperative guidance navigation tool. As a result, the rationale behind our experimental study

was to test the reliability of HSI in recognizing multiple tissue classes *in vivo*, in order to develop at a later stage, HSI-based automatic intraoperative navigation system for use in the clinical setting.

More than a decade ago, in a pioneer work, Zuzak et al. successfully differentiated the portal vein from the biliary duct and the hepatic artery, by means of a laparoscopic hyperspectral imager prototype[23]. In their experimental proof of the concept, the spectral curves of the different tissue types were analyzed using the PCA (Principal Component Analysis). However, despite the promising results, the long acquisition time of the HSI laparoscopic prototype and the scarce reproducibility of the PCA-based tissue discrimination, impaired the clinical translation of this technology.

Recently, other authors attempted to discriminate anatomical structures by means of HSI intraoperatively[26, 27]. However, in both works the authors did not rely on the spectral curves for the discriminative analysis, rather attempted to select the best spectral bands in order to obtain an RGB enhanced image, to highlight for the human eye the chosen target structures, such as nerves or ureters. This approach has the potential of allowing for a rapid analysis, once the best band is selected. However, this principle very much resembles the concept of NBI (Narrow Band Image), a band selection modality, largely utilized in diagnostic gastrointestinal endoscopy. In fact, the light emitted in this particular band (415 nm) is able to delineate the surface micro-vessels, which are, due to neoplastic neoangiogenesis, more represented within cancerous lesions, allowing to discriminate normal from cancerous or pre-cancerous mucosa. Despite the remarkable results of both groups, band selection uses only partially the potential of HSI and still requires a substantial interpretation of the operator to discriminate the enhanced structure. The utilization of advanced machine learning algorithms to reliably detect the slight differences within spectral curves of different tissues, reduces enormously the human interpretation bias, heading towards precision surgery.

Enormous advances have been made in this sense, in the field of HSI-based cancer detection. In particular, deep learning has been successfully coupled to HSI to detect with high accuracy gastric cancer from healthy tissue on fresh samples[17] or microscopic slides[16]. Recently, in a remarkable work involving 82 patients, a group could identify salivary gland and thyroid tumors on operative specimens, with impressive precision[22]. Additionally, the authors were able to demonstrate the superiority of HSI in detecting the specific cancer types against fluorescence-dye-based imaging techniques. All these innovative works have the merit to discriminate cancerous lesions with a high degree of accuracy by coupling HSI to deep learning algorithms. However, in contrast to those previous works, in which only cancer has to be recognized over the non-cancer merged classes, in our setup we were able to discriminate precisely among 6 different tissue classes in an *in vivo* large animal model.

Since we had a limited amount of data, we decided to use all data for the training phase and by employing the Leave One Patient Out Cross Validation (LOPOCV) technique we could reliably verify our algorithms' performance. On the other hand, a cross validation in which only a subset of data is used as training and the rest as testing data, as performed in a number of previous works, has the disadvantage of being less reliable of the LOPOCV method.

During our experimental setup the ground truth was provided by annotations made intraoperatively by a fully trained surgeon. As explained in the methods, the annotations were performed in the OR, having the surgical scene still available and magnifying the RGB images associated with each hyperspectral image on the computer's screen. In fact, annotation based on the RGB images alone in a subsequent postprocessing phase, would have certainly led to annotation mistakes. However, only the tissue classes that after accurate dissection could be visually recognized with absolute confidence, and this was always the case of the elements of the neurovascular bundle (carotid artery, jugular vein and vagal nerve) given their defined anatomical position, were annotated. As a consequence, a large part of the image remained unlabeled, since without systematical histopathological mapping (which is hardly practicable in the reality) it is not possible to discriminate visually certain tissue types, such as fat from connective tissue, which might be encountered over large areas.

Despite the limited sample size, we were able to achieve excellent discrimination accuracy. This result supports the fitness of CNN for interpreting objectively HSI datasets.

Previously tissue discrimination was achieved by qualitative analysis of the spectra obtained by means of the probe-based diffuse reflectance spectroscopy (DRS)[32]. As noted in our findings, also those authors observed analogous patterns within curves of fatty tissue, muscle and nerve and they were able to find discriminative characteristics for these tissue types only by using an additional DRS sensor able to analyze the spectrum beyond 1000 nm. However, the hyperspectral imager we employed reached a maximal spectral range until 1000 nm, but still we were able to successfully discriminate those ambiguous tissue types. This was achieved with the fine adjustment of the CNN parameters and it demonstrates that CNN are particularly suitable to analyze with high precision the large hypercubes.

In conclusion, this work demonstrates that an *in vivo* recognition of anatomical key structures (in particular blood vessels and nerves) using the hyperspectral technology in association with deep learning, is possible and highly accurate. Future efforts will be allocated in developing an advanced HSI-based navigation tool to support the mental anatomical map of the surgeon, as a tool to avoid inadvertent injuries to key anatomical structures. Furthermore, as the era of autonomous robotic surgery emerges, automatic tissue classification will be an essential tool to facilitate safe autonomous procedures, and HSI combined with deep learning may prove a key enabling technology. As future follow-up work, we aim to enlarge the study to more subjects and to repeat the experimental protocol with human data captured in an observational clinical trial.

Materials And Methods

4.1 Overview

In this section we discuss the HSI image dataset, the machine learning task and training process, and our performance evaluation setup. The models are trained to recognize the tissue type at any spatial location within HSI images, among a set of pre-defined tissue types. We solve this problem by training state-of-the-art models using supervised learning. This involves two stages: First is the training stage, where the

model parameters are automatically optimized to maximize recognition performance on a training dataset with known tissue classes provided by an expert surgeon. Second is the inference stage, where the predictive performance is evaluated with HSIs of subjects that are not present in the training dataset with LOPOCV. This evaluation measures the performance and generalization capability of the model.

4.2 Animals characteristics

In the present study 8 adult pigs (Large White) were included. This experiment was part of the ELIOS project (Endoscopic Luminescent Imaging for Oncology Surgery), approved by the local Ethical Committee on Animal Experimentation (ICOMETH No. 38.2016.01.085), and by the French Ministry of Superior Education and Research (MESR) (APAFIS#8721-2017013010316298-v2). All animals were managed according to French laws for animal use and care, and according to the directives of the European Community Council (2010/63/EU) and ARRIVE guidelines[33].

A 24 hours preoperative fasting with free access to water was observed. Premedication was administered 10 minutes preoperatively, with an intramuscular injection of ketamine (20mg/kg) and azaperone (2mg/kg) (Stresnil, Janssen-Cilag, Belgium). Intravenous propofol (3mg/kg) combined with rocuronium (0.8mg/kg) was used for induction. Anesthesia was maintained with 2% isoflurane. At the end of the procedures, pigs were sacrificed with an intravenous injection of Pentobarbital Sodium (40mg/kg) (Exagon®, AXIENCE, France), under 5% isoflurane anesthesia.

4.3 Surgical Procedure and Hyperspectral Data Acquisition

A mid-line neck incision was performed, and the neurovascular bundle of the neck was carefully dissected. Successively, the vagal nerve, common carotid artery and the common jugular vein were isolated bilaterally, and the nerves were marked using a 3-0 polypropylene suture thread. The surgical scene was stabilized using self-retaining retractors. The hyperspectral imager employed during the experiment was a push-broom scanning complementary metal–oxide–semiconductor (CMOS) compact system (TIVITA®, Diaspective Vision GmbH, Germany), with a 640x476 pixel spatial resolution, acting within a spectral range from 500 to 1000nm (5 nm spectral resolution). The hyperspectral camera was placed at approximately 45 cm from the surgical field and an acquisition was taken. In order to avoid any potential bias, environmental lights were switched off and the ventilation was paused for the few seconds required for each acquisition (<10 seconds).

4.4 Imaging Postprocessing and Data Annotation

Immediately after each acquisition, in the OR and with the surgical scene still available, the operating surgeon (MB) by means of an image manipulator software (GIMP, GNU Image Manipulation Program, open source) annotated manually the resulting RGB image. The annotated classes were vagal nerve, jugular vein, carotid artery, subcutaneous fat, muscle, skin and metal of the retractor. The RGB images for each subject are shown in **Figure 2 (a)**, which are provided by the camera device and synthesized from the HSI. As visible in the pictures the nerves and the vessels are rather small and difficult to distinguish

from the RGB images alone. For this reason, the annotation was carried on by the operating surgeon, still looking at the surgical field, in which the structures were well distinguishable, given the precise anatomical position of the dissected elements of the neck neurovascular bundle (carotid artery, jugular vein and vagal nerve). These images are visualized in greyscale with annotations overlaid as colored regions in **Figure 2 (b)**. These figures are cropped and oriented similarly for better visualization. It is not feasible to correctly annotate every pixel in each image. Consequently, the annotation of each class is a subset of the pixels of the class for which the human annotator is certain of the class type. The machine learning models are trained and evaluated using only the annotated regions. One of the primary challenges is that the HSI camera has a limited resolution of 640 x 476 pixels, and consequently the thin structures (in particular nerves) are challenging to recognize.

4.5 Image Pre-processing and Spectral Curve Distributions

The HSI images are calibrated to account for the illumination source, dark current and the current of the Charged Coupled Device (CCD) camera. Therefore, intensity calibration was not required as a pre-processing step. The images have 100 wavelength bands in the range 500 nm (visible green light) to 1000 nm (near infra-red). Normalization techniques are commonly applied as a pre-processing stage to reduce spectral variability caused by tissue morphology effects (changes in illumination angle, and non-flat structures) [34, 35]. The most popular normalization technique has been tested: Standard Normal Variate (SNV), where each individual curve is transformed to have a zero mean and standard deviation of one. SNV makes a spectral curve invariant to linear transforms caused by e.g. surface orientation variability. However, SNV may adversely affect recognition performance because of the potential loss of discriminative information. Models have been trained by us both with and without SNV normalization to assess its benefits and limits.

Before applying machine learning, a qualitative analysis of the spectral curves of each tissue class was performed to understand the intrinsic difficulty of the recognition problem. These curves are plotted in **Figure 5** and **Figure 6**, where we show spectral curve distributions for each class with and without the use of SNV normalization. Classes that are easier to recognize by the machine have two general characteristics: *i*) low intra-class variation (spectral distribution is similar for the same class) and *ii*) high inter-class variation (spectral distribution is different for different classes). Intra-class variation is primarily caused by differences in the same tissue class in different subjects. In **Figure 5** and **Figure 6**, intra-class variation is represented by the spectral curve spread (one standard deviation from the mean, illustrated by the grey zone). Inter-class variation is visualized by differences in the mean spectral curves (solid black line). Inter-class variability is tissue specific, where vein and skin are most dissimilar tissues. Muscle, nerve and fat are relatively similar, indicating that recognizing them with machine learning is not trivial. The metal class has a strongly different profile compared to the tissue classes. The large intra-class variation of metal can be explained by strong specular reflection.

It is clear that intra-class variability has been reduced by SNV normalization, but inter-class variability has also been reduced. The mean spectral curves of all classes are plotted together in **Figure 6**, right-most

graphs, showing without and with SNV normalization (top and bottom right graphs of **Figure 6**, respectively).

4.6 Machine Learning Recognition Problem

Given a hyperspectral image, for each spatial coordinate (pixel), a sub-volume centered on the pixel is extracted, which is then input to a predictive machine learning model. This model produces a score for each tissue class, where a higher score indicates more likelihood of the tissue class being present at the spatial coordinate. Finally, the pixel is assigned to the class with the highest predictive score. This process repeats for each pixel of interest. This machine learning recognition problem is illustrated in **Figure 7**.

4.7 Machine Learning Models

Various machine learning models have been applied to other HSI image classification problems including medical imaging and remote sensing, and there no single model that performs best on all datasets[36, 37]. Consequently, the two most successful models for HSI image segmentation are studied in this work: Support Vector Machines (SVMs)[37-39] and Convolutional Neural Networks (CNNs)[36, 40].

4.8 Support Vector Machine (SVM)

SVMs have been shown to work well for HSI classification problems, in particular remote sensing[41-43]. They are simple to train and they offer good performance when the training set size is limited[37] and when the input data is high-dimensional, as is the case for our dataset. A SVM classifier fits a decision boundary (hyperplane) in multidimensional feature space. The feature space is constructed by transforming the data points using a kernel, such as radial basis functions, to feature space. Training is then performed to determine a hyperplane that separates the classes in feature space. Once trained, a new spectral curve at a given pixel is first transformed to feature space and then classified using the hyperplane. SVMs generalize naturally to multi-class classification problems such as ours, and they have been shown to outperform other classical machine learning models in HSI classification problems such as k-nearest neighbor (KNN) and decision trees[44]. We construct the feature space using the relative absorption of each wavelength in the HSI sub-volume centered at a given pixel. Consequently, the feature space has 100 dimensions (one dimension per wavelength).

4.9 Convolutional Neural Network (CNN)

The second model is a CNN, shown to work well with many kinds of image data including HS[36, 40, 45, 46]. A CNN is a special kind of feed-forward neural network where neurons are arranged in a series of hidden layers. The image is fed into the first layer, and the purpose of the layer is to extract meaningful features to aid classification. These features are fed into the second layer, where higher-level features are computed from the inputted features, and the process repeats until the last layer. Finally, a classification decision is made using the features at the last layer. CNNs are specially designed so that the features are extracted using convolutional filters. The filter weights are trainable parameters of the mode, and during

learning they automatically adapt to extract features that are most pertinent to the specific classification task.

4.10 Implementation and Training

For each HS image in the dataset, we sample sub-volume centered on every annotated pixels. Spatial patches of 5x5 pixels are extracted to form 5x5x100 sub-volumes. The third dimension corresponds to the wavelength dimension with the 100 bands. The dataset is then composed of 213,203 samples.

Those sub-volumes are successively split into training and testing sets respecting the Leave-One-Patient-Out Cross-Validation (LOPOCV) process to train and evaluate model performance. LOPOCV is standard practice in medical image classification with small datasets, and it ensures that a model is not trained and tested on data from the same subject. Performance statistics with LOPOCV measure the generalization performance of a classifier on new subjects. LOPOCV was performed by partitioning the subjects into 8 subsets, $S_1, S_2 \dots S_8$, where each subset S_i is formed of every sub-volumes extracted from all subject HS images excluding the i^{th} subject. The models are trained on subset S_i , and then performance was tested on the i^{th} subject HS image. The process is repeated 8 times for each subject.

The SVM classified was implemented with Python's Scikit-learn library[47]. Auto-scaling was applied to each spectral band (unit variance and mean centering)[48] to eliminate the influence of variable dimension Default SVM parameters were used from Python's `sklearn.svm.SVC` class (Radial Basis Kernels (RBFs of degree 3 and regularization parameter $C=1$)). The CNN was implemented in Pytorch v1.2 using an established neural network architecture[45] that has been implemented in a public code. The CNN processes a sub-volume centered around a given pixel using a 5x5 spatial window) and it has 32,628 trainable parameters with 7 hidden layers. The first 6 layers are convolutional and the last hidden layer is fully connected. The final output layer has 7 neurons, where 7 is the number of classes (6 tissue classes and the metal class). Each output layer computes the predictive score of each class. Finally, the class with the highest score is selected. The architecture of the CNN is provided in **Figure 8**. for full reproducibility. The CNN is trained using Pytorch's backpropagation implementation with Adam optimization (0.001)[49]. Because the dataset is highly imbalanced with some classes significantly more represented than others, training was implemented with a weighted cross entropy loss with weight inversely proportional to the number of training samples per class. This is standard practice to prevent the classifier attempting to maximize performance for the most represented class (in our case, the skin class).

Training was performed on an DGX V1 (Nvidia corp.) equivalent deep learning server (Cysco inc.) taking approximately 15 hours for the CNN and 25 hours for the SVM. Training was performed 8 times in order to implement LOPOCV. Each training session has seven training images and one test image, and training was repeated eight times so that every image is used once as a test image. .

4.11 Performance Metrics and Statistical Methods

Performance was evaluated with standard metrics used in machine learning, implemented by Python scikit-learn (version 0.23, <https://scikit-learn.org>). Multi-class Type 1 and Type 2 errors are evaluated and presented as confusion matrices. Area Under Receiver Operator Curve (AU-ROC), sensitivity, specificity and Sørensen–Dice coefficient score (DCS) were used to evaluate class-specific performance using ‘one versus all’ and macro averaging. Statistically significant differences were computed with a two-tailed paired t-test ($\alpha=0.05$) similarly to previous works in HSI classification analysis[30] using Excel (Microsoft Office 365, Microsoft, U.S.A.).

Declarations

Funding, acknowledgments, and disclosures 5.1 Funding This work was funded by the ARC Foundation through the ELIOS (Endoscopic Luminescent Imaging for precision Oncologic Surgery) grant. 5.2 Disclosures J.M. is the President of IRCAD, which is partly funded by KARL STORZ, and Medtronic. M.D. is member of the Scientific Board of Diagnostic Green GmbH and is the recipient of the ELIOS grant. M.B., T.C., V.B., R.N., E.F.; M.V. and A.H. have nothing to disclose.

References

1. C. Wallner, M. M. Lange, B. A. Bonsing, C. P. Maas, C. N. Wallace, N. F. Dabhoiwala, H. J. Rutten, W. H. Lamers, M. C. DeRuiter, and C. J. van de Velde, "Causes of fecal and urinary incontinence after total mesorectal excision for rectal cancer based on cadaveric surgery: a study from the Cooperative Clinical Investigators of the Dutch total mesorectal excision trial," *Journal of clinical oncology* **26**, 4466-4472 (2008).
2. J. H. Lefevre, C. Tresallet, L. Leenhardt, C. Jublanc, J.-P. Chigot, and F. Menegaux, "Reoperative surgery for thyroid disease," *Langenbeck's archives of surgery* **392**, 685-691 (2007).
3. H. Köhler, A. Kulcke, M. Maktabi, Y. Moulla, B. Jansen-Winkel, M. Barberio, M. Diana, I. Gockel, T. Neumuth, and C. Chalopin, "Laparoscopic system for simultaneous high-resolution video and rapid hyperspectral imaging in the visible and near-infrared spectral range," *Journal of Biomedical Optics* **25**, 086004 (2020).
4. J. Yoon, J. Joseph, D. J. Waterhouse, A. S. Luthman, G. S. Gordon, M. Di Pietro, W. Januszewicz, R. C. Fitzgerald, and S. E. Bohndiek, "A clinically translatable hyperspectral endoscopy (HySE) system for imaging the gastrointestinal tract," *Nature communications* **10**, 1-13 (2019).
5. G. Lu, and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics* **19**, 010901 (2014).
6. B. Jansen-Winkel, M. Maktabi, J. Takoh, S. Rabe, M. Barberio, H. Köhler, T. Neumuth, A. Melzer, C. Chalopin, and I. Gockel, "Hyperspektral-imaging bei gastrointestinalen anastomosen," *Der Chirurg* **89**, 717-725 (2018).
7. H. Köhler, B. Jansen-Winkel, M. Maktabi, M. Barberio, J. Takoh, N. Holfert, Y. Moulla, S. Niebisch, M. Diana, and T. Neumuth, "Evaluation of hyperspectral imaging (HSI) for the measurement of ischemic

- conditioning effects of the gastric conduit during esophagectomy," *Surgical endoscopy* **33**, 3775-3782 (2019).
8. B. Jansen-Winkeln, N. Holfert, H. Köhler, Y. Moulla, J. Takoh, S. Rabe, M. Mehdorn, M. Barberio, C. Chalopin, and T. Neumuth, "Determination of the transection margin during colorectal resection with hyperspectral imaging (HSI)," *International journal of colorectal disease* **34**, 731-739 (2019).
 9. M. Barberio, E. Felli, M. Pizzicannella, V. Agnus, M. Al-TaHER, E. Seyller, Y. Moulla, B. Jansen-Winkeln, I. Gockel, J. Marescaux, and M. Diana, "Quantitative serosal and mucosal optical imaging perfusion assessment in gastric conduits for esophageal surgery: an experimental study in enhanced reality," *Surg Endosc* (2020).
 10. H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Transactions on Biomedical Engineering* **57**, 2011-2017 (2010).
 11. M. Barberio, F. Longo, C. Fiorillo, B. Seeliger, P. Mascagni, V. Agnus, V. Lindner, B. Geny, A.-L. Charles, and I. Gockel, "HYPERspectral Enhanced Reality (HYPER): a physiology-based surgical guidance tool," *Surgical endoscopy*, 1-9 (2019).
 12. E. Felli, M. Al-TaHER, T. Collins, A. Baiocchini, E. Felli, M. Barberio, G. M. Ettore, D. Mutter, V. Lindner, A. Hostettler, S. Gioux, C. Schuster, J. Marescaux, and M. Diana, "Hyperspectral evaluation of hepatic oxygenation in a model of total vs. arterial liver ischaemia," *Sci Rep* **10**, 15441 (2020).
 13. H. Akbari, L. Halig, D. M. Schuster, B. Fei, A. Osunkoya, V. Master, P. Nieh, and G. Chen, "Hyperspectral imaging and quantitative analysis for prostate cancer detection," *Journal of biomedical optics* **17**, 076005 (2012).
 14. E. J. Baltussen, E. N. Kok, S. G. B. de Koning, J. Sanders, A. G. Aalbers, N. F. Kok, G. L. Beets, C. C. Flohil, S. C. Bruin, and K. F. Kuhlmann, "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," *Journal of biomedical optics* **24**, 016002 (2019).
 15. Z. Han, A. Zhang, X. Wang, Z. Sun, M. D. Wang, and T. Xie, "In vivo use of hyperspectral imaging to develop a noncontact endoscopic diagnosis support system for malignant colorectal tumors," *Journal of biomedical optics* **21**, 016001 (2016).
 16. Y. Li, L. Deng, X. Yang, Z. Liu, X. Zhao, F. Huang, S. Zhu, X. Chen, Z. Chen, and W. Zhang, "Early diagnosis of gastric cancer based on deep learning combined with the spectral-spatial classification method," *Biomedical optics express* **10**, 4999-5014 (2019).
 17. B. Hu, J. Du, Z. Zhang, and Q. Wang, "Tumor tissue classification based on micro-hyperspectral technology and deep learning," *Biomedical Optics Express* **10**, 6370-6389 (2019).
 18. H. Fabelo, S. Ortega, D. Ravi, B. R. Kiran, C. Sosa, D. Bulters, G. M. Callicó, H. Bulstrode, A. Szolna, and J. F. Piñeiro, "Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations," *PLoS One* **13** (2018).
 19. L. Ma, G. Lu, D. Wang, X. Wang, Z. G. Chen, S. Muller, A. Chen, and B. Fei, "Deep learning based classification for head and neck cancer detection with hyperspectral imaging in an animal model," in

20. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Diery, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *Journal of biomedical optics* **22**, 086009 (2017).
21. M. Halicek, G. Lu, J. V. Little, X. Wang, M. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *Journal of biomedical optics* **22**, 060503 (2017).
22. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, and B. Fei, "Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning," *Biomedical Optics Express* **11**, 1383-1400 (2020).
23. K. J. Zuzak, S. C. Naik, G. Alexandrakis, D. Hawkins, K. Behbehani, and E. Livingston, "Intraoperative bile duct visualization using near-infrared hyperspectral video imaging," *The American Journal of Surgery* **195**, 491-497 (2008).
24. C. D. Nawn, B. Souhan, R. Carter, C. Kneapler, N. Fell, and J. Y. Ye, "Distinguishing tracheal and esophageal tissues with hyperspectral imaging and fiber-optic sensing," *Journal of biomedical optics* **21**, 117004 (2016).
25. M. Barberio, M. Maktabi, I. Gockel, N. Rayes, B. Jansen-Winkel, H. Köhler, S. M. Rabe, L. Seidemann, J. P. Takoh, and M. Diana, "Hyperspectral based discrimination of thyroid and parathyroid during surgery," *Current Directions in Biomedical Engineering* **4**, 399-402 (2018).
26. E. L. Wisotzky, F. C. Uecker, P. Arens, S. Dommerich, A. Hilsmann, and P. Eisert, "Intraoperative hyperspectral determination of human tissue properties," *Journal of biomedical optics* **23**, 091409 (2018).
27. D. Nouri, Y. Lucas, and S. Treuillet, "Hyperspectral interventional imaging for enhanced tissue visualization and discrimination combining band selection methods," *International journal of computer assisted radiology and surgery* **11**, 2185-2197 (2016).
28. D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression* (John Wiley & Sons, 2013).
29. L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**, 297-302 (1945).
30. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning," *Cancers* **11**, 1367 (2019).
31. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," (2017).
32. R. M. Schols, M. Ter Laan, L. P. Stassen, N. D. Bouvy, A. Amelink, F. P. Wieringa, and L. Alic, "Differentiation between nerve and adipose tissue using wide-band (350–1,830 nm) in vivo diffuse reflectance spectroscopy," *Lasers in surgery and medicine* **46**, 538-545 (2014).

33. C. Kilkenny, W. J. Browne, I. C. Cuthill, M. Emerson, and D. G. Altman, "Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research," *Osteoarthritis and cartilage* **20**, 256-260 (2012).
34. M. Vidal, and J. M. Amigo, "Pre-processing of hyperspectral images. Essential steps before image analysis," *Chemometrics and Intelligent Laboratory Systems* **117**, 138-148 (2012).
35. R. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied spectroscopy* **43**, 772-777 (1989).
36. M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing* **158**, 279-317 (2019).
37. G. Camps-Valls, and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing* **43**, 1351-1362 (2005).
38. F. Melgani, and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on geoscience and remote sensing* **42**, 1778-1790 (2004).
39. L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE transactions on geoscience and remote sensing* **53**, 6663-6674 (2015).
40. W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep Convolutional Neural Networks for Hyperspectral Image Classification," *Journal of Sensors* **2015**, 258619 (2015).
41. T. Qiao, J. Ren, Z. Wang, J. Zabalza, M. Sun, H. Zhao, S. Li, J. A. Benediktsson, Q. Dai, and S. Marshall, "Effective denoising and classification of hyperspectral images using curvelet transform and singular spectrum analysis," *IEEE transactions on geoscience and remote sensing* **55**, 119-133 (2016).
42. Y. Ji, L. Sun, Y. Li, J. Li, S. Liu, X. Xie, and Y. Xu, "Non-destructive classification of defective potatoes based on hyperspectral imaging and support vector machine," *Infrared Physics & Technology* **99**, 71-79 (2019).
43. V. Lukin, S. Abramov, S. Krivenko, A. Kurekin, and O. Pogrebnyak, "Analysis of classification accuracy for pre-filtered multichannel remote sensing data," *Expert Systems with Applications* **40**, 6400-6411 (2013).
44. L. Ma, M. M. Crawford, and J. Tian, "Local Manifold Learning-Based k -Nearest-Neighbor for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing* **48**, 4099-4109 (2010).
45. A. B. Hamida, A. Benoit, P. Lambert, and C. Ben-Amar, "Deep learning approach for remote sensing image analysis," (2016).
46. B. Liu, X. Yu, P. Zhang, X. Tan, A. Yu, and Z. Xue, "A semi-supervised convolutional neural network for hyperspectral image classification," *Remote Sensing Letters* **8**, 839-848 (2017).
47. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research* **12**, 2825-2830 (2011).

48. C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).

49. D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

Figures

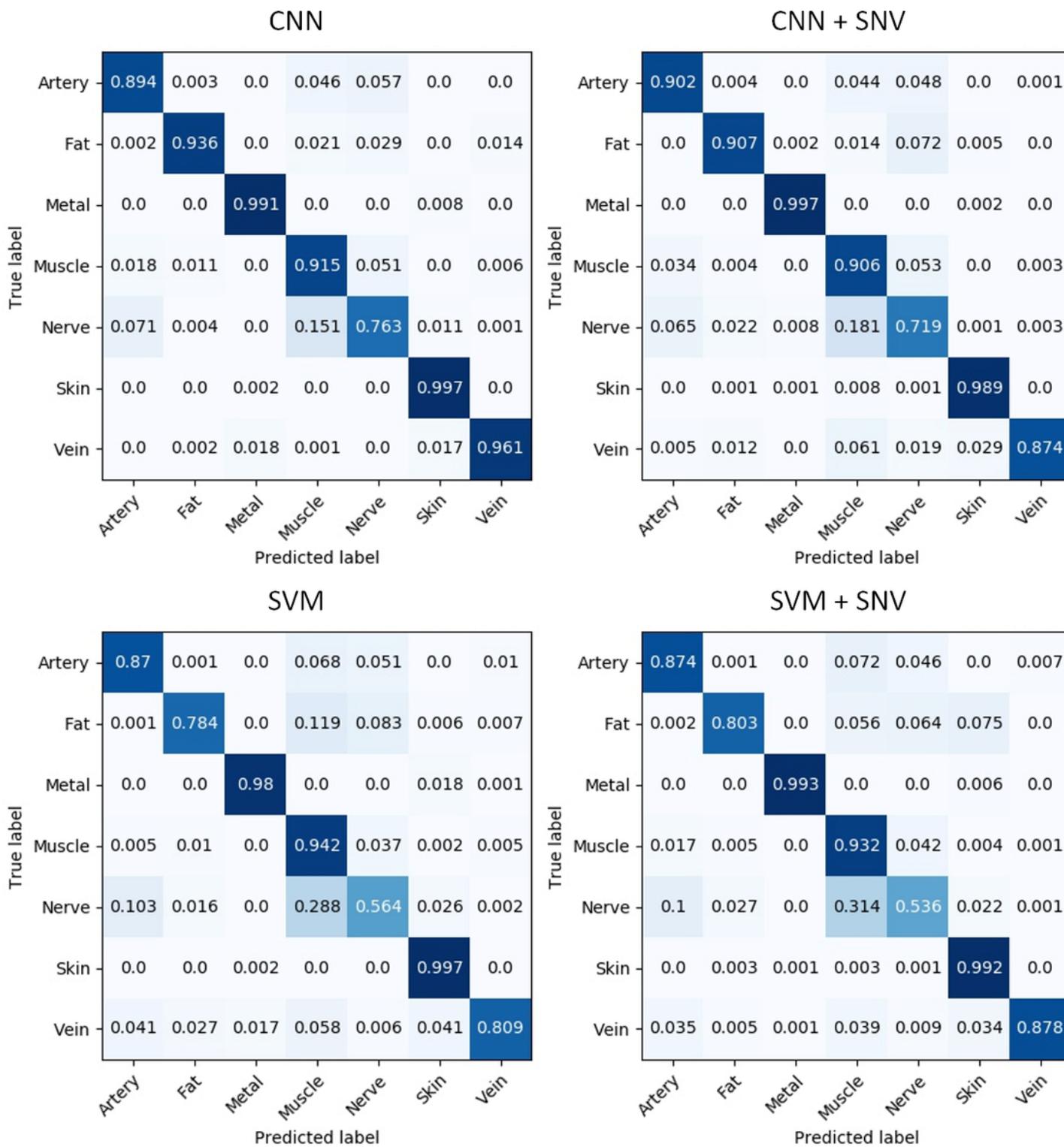


Figure 1

Confusion matrices: Predictive performance of 4 different models visualized as confusion matrices.

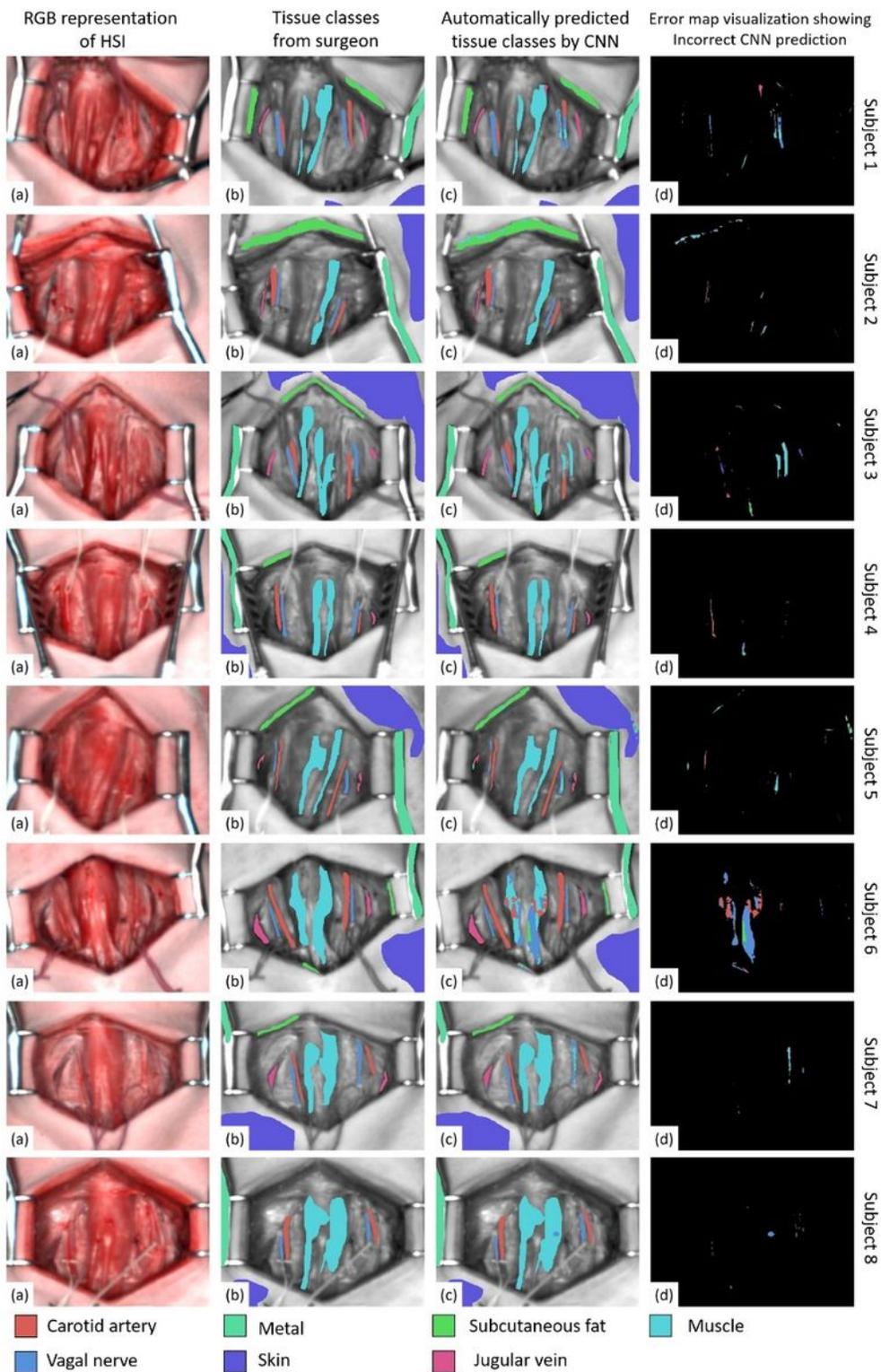


Figure 2

Automatic Tissue Recognition Visualization: Visualization of automatic tissue recognition results from the CNN model. In each row we show images from subjects 1-8. From left to right in columns we show

the RGB image simulated from HS (a), the ground-truth classifications from the surgeon (b), the predicted classifications from the CNN (c) and the error map (d). Black pixels in the error map indicates perfect classification (no error). Colored pixels indicate an incorrect classification, and the color provides the incorrect prediction.

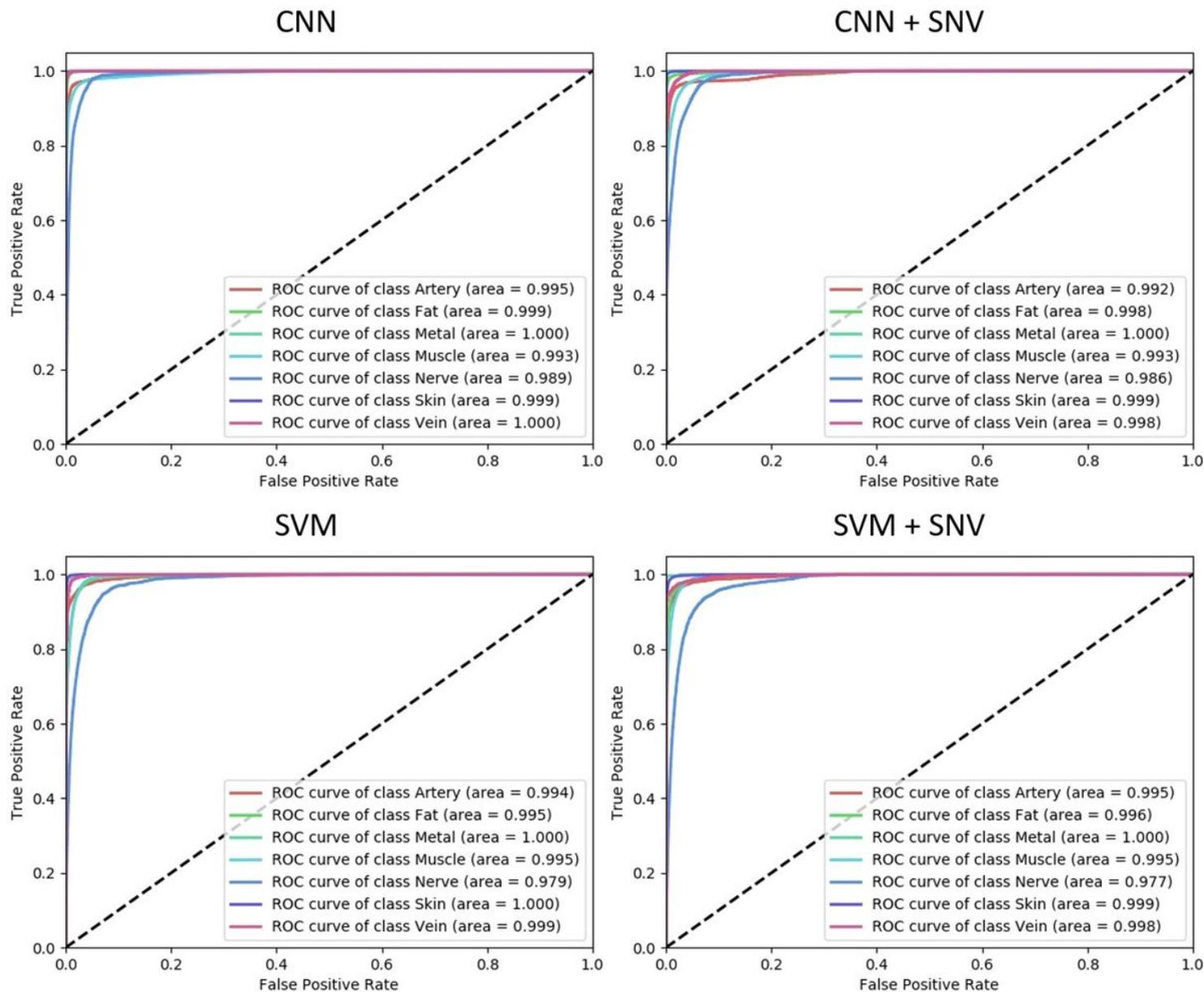
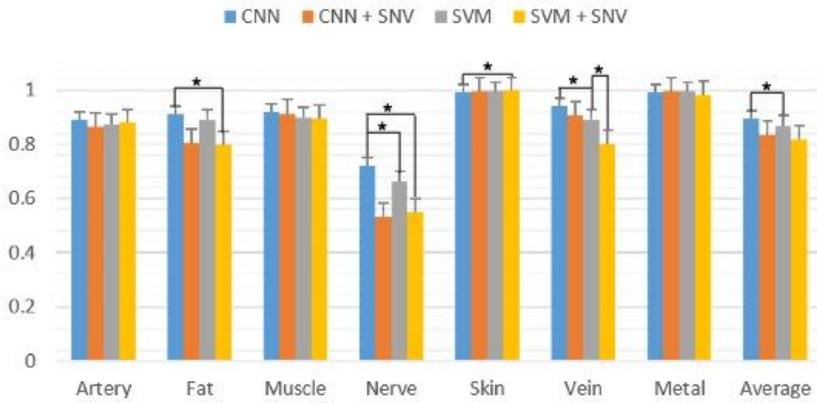


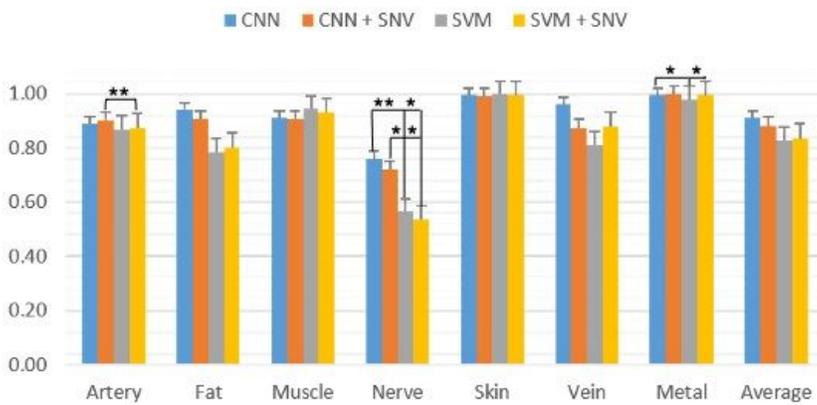
Figure 3

Receiver Operator Characteristic curves: Receiver Operator Characteristic curves for the four model. In each figure we plot the ROC curves for one model for all classes.

Dice Similarity Coefficient (DSC)



Sensitivity



Specificity

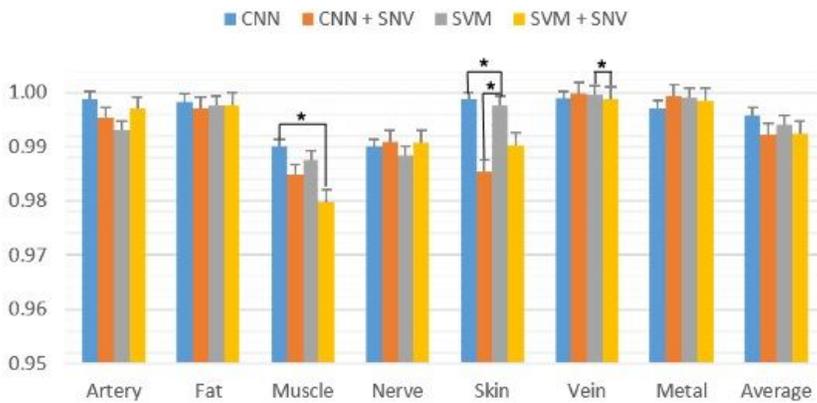


Figure 4

Dice Similarity Coefficient (DSC): Performance of the four classification models in terms of DSC (top), sensitivity (middle) and specificity (bottom). Error bars represent 1 Standard Error.

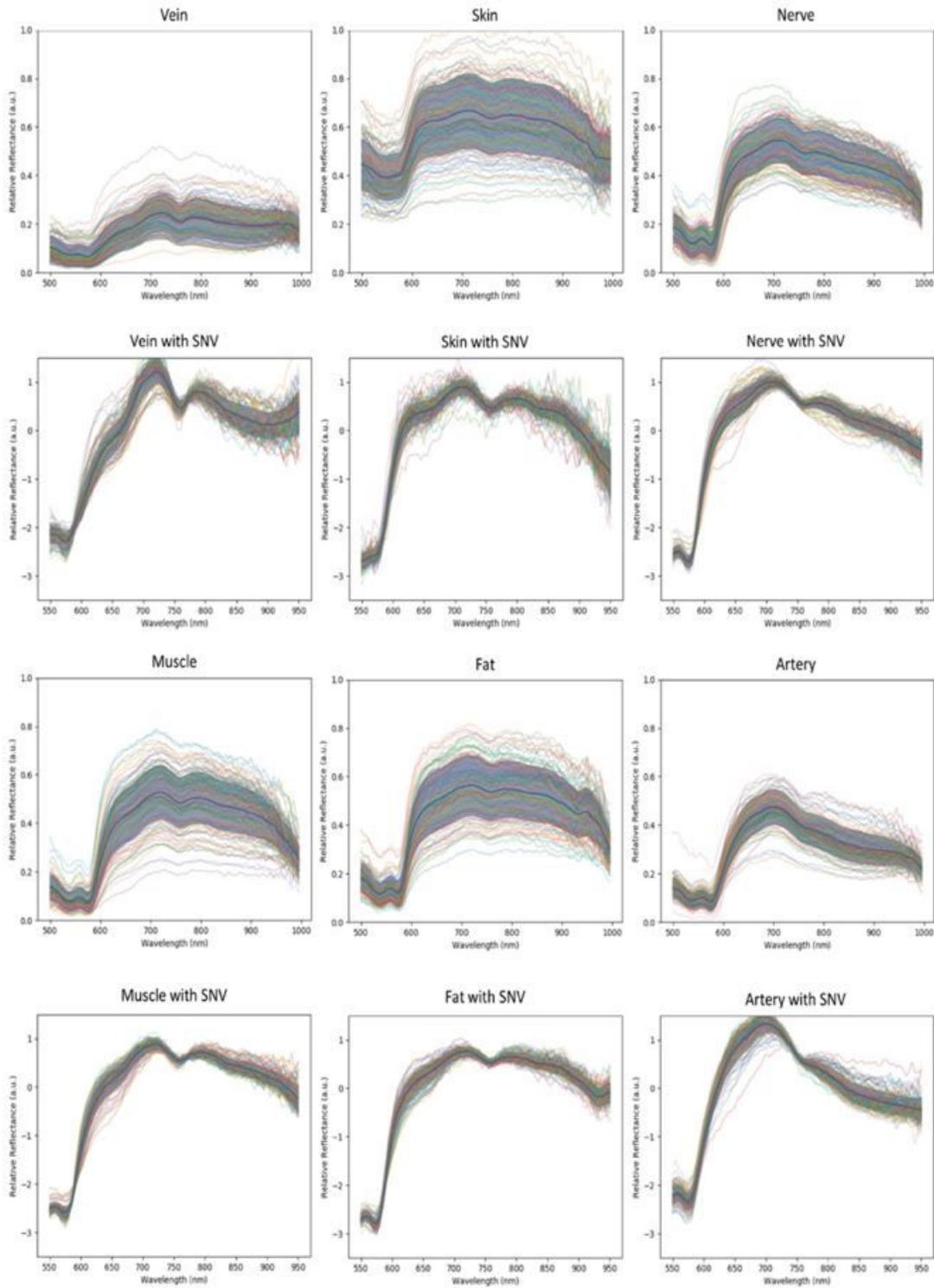


Figure 5

Standard Normal Variate (SNV) within the tissue classes: Spectral curve profiles of the tissue classes without and with SNV normalization. Relative absorption is plotted as a function of wavelength for each class. The mean spectral curve is shown as a black solid line, and the curve spread is shown with a grey band of one standard deviation from the mean curve.

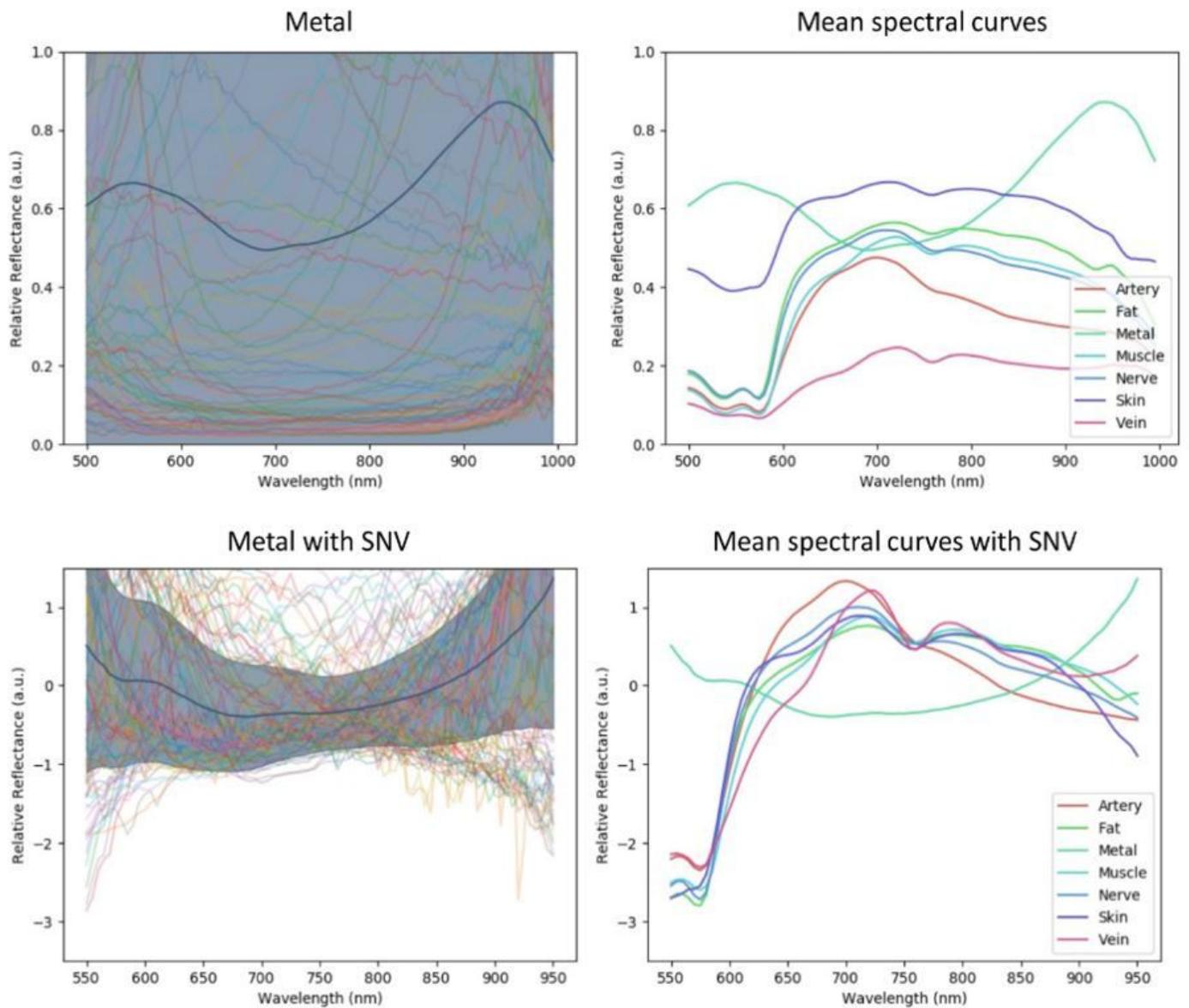


Figure 6

Standard Normal Variate (SNV) within the metal class: Spectral curve profiles of the metal classes without and with SNV normalization with similar visualization as Figure 5. The metal class has a very large spectral distribution because of specular reflections with a completely different profile to the tissue classes. Figures on the right show mean spectral curves of all classes overlaid in a single plot (top without SNV, bottom with SNV normalization).

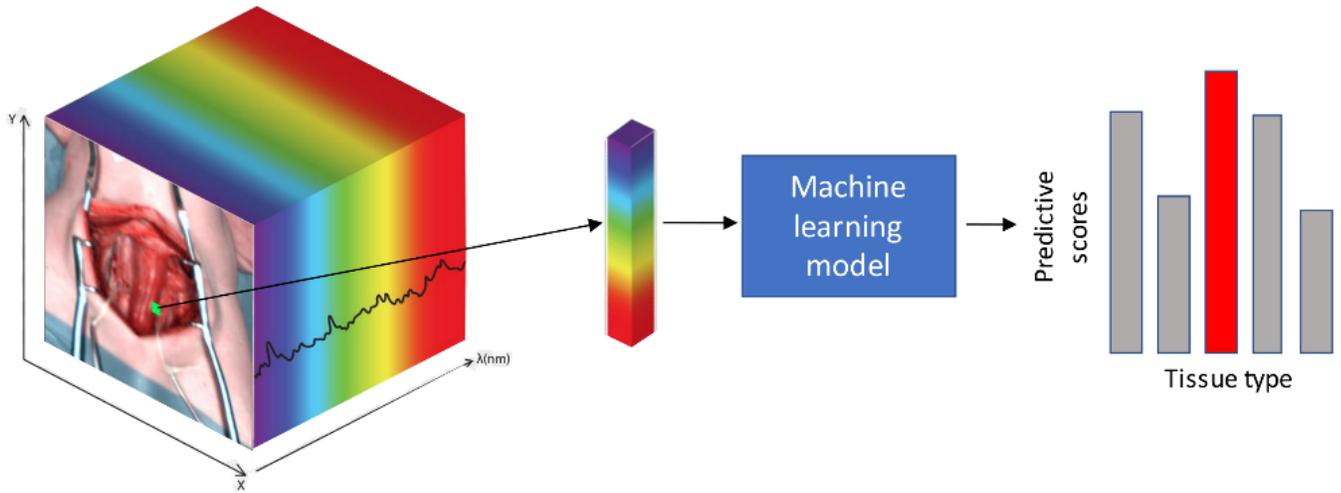


Figure 7

Schematic representation of hyperspectral (HS)-based tissue classes recognition: Illustration of the general tissue recognition problem with HS image data. The core task for is to recognize tissue type at a given spatial location within a HS image. A sub-volume is extracted at the spatial location with a small spatial window, which is passed to the predictive machine learning model. Predictive scores are then automatically generated for each tissue class. Finally, the class with highest predictive score is selected.

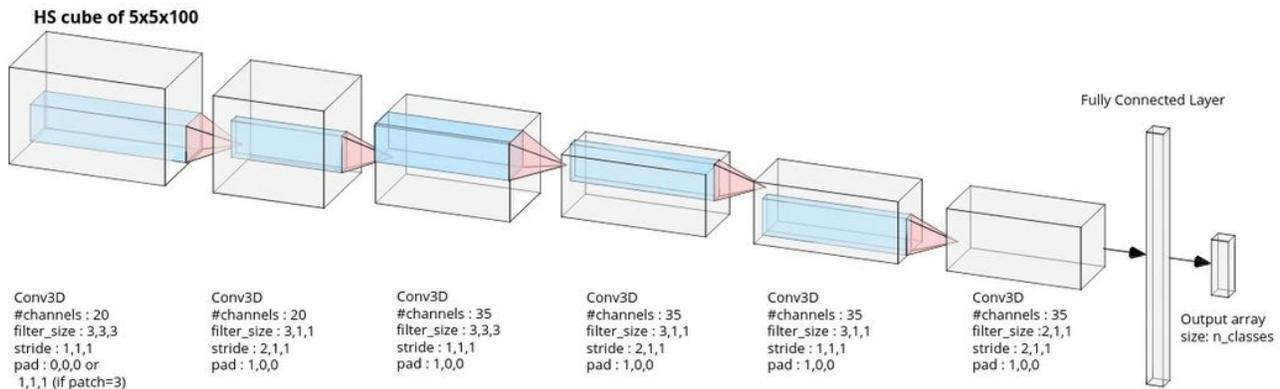


Figure 8

Convolutional Neural Network architecture: The input is a HSI sub-volume centered about a given pixel, with 5x5 spatial dimensions and 100 wavelength dimensions. Down-sampling convolutional operations

are applied to transform the input to a final 1D feature space vector, which is followed by a finally fully connected layer to produce the tissue prediction scores.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.pdf](#)