

Published in final edited form as:

Proc SPIE. 2014 March 21; 9034: 90341W-. doi:10.1117/12.2043848.

A Minimum Spanning Forest Based Hyperspectral Image Classification Method for Cancerous Tissue Detection

Robert Pike¹, Samuel K. Patton¹, Guolan Lu², Luma V. Halig¹, Dongsheng Wang³, Zhuo Georgia Chen³, and Baowei Fei^{1,2,4,5,*}

¹Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA

²The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA

³Department of Hematology and Medical Oncology, Emory University, Atlanta, GA

⁴Department of Mathematics & Computer Science, Emory University, Atlanta, GA

⁵Winship Cancer Institute of Emory University, Atlanta, GA

Abstract

Hyperspectral imaging is a developing modality for cancer detection. The rich information associated with hyperspectral images allow for the examination between cancerous and healthy tissue. This study focuses on a new method that incorporates support vector machines into a minimum spanning forest algorithm for differentiating cancerous tissue from normal tissue. Spectral information was gathered to test the algorithm. Animal experiments were performed and hyperspectral images were acquired from tumor-bearing mice. In vivo imaging experimental results demonstrate the applicability of the proposed classification method for cancer tissue classification on hyperspectral images.

Keywords

Hyperspectral imaging; image classification; minimum spanning forest; support vector machine

1. INTRODUCTION

Surgical resection of cancerous tissue is a common treatment option. There is a need for verification of complete cancerous tissue removal within the tumor bed during the procedure. The use of hyperspectral imaging may provide a non-invasive detection method for marginal cancerous tissue left behind following the removal of the tumor. Current diagnostic imaging techniques require cumbersome equipment that cannot be easily accessible during surgery. Hyperspectral imaging can be performed using equipment of a much smaller size and greater mobility.

A hyperspectral image consists of a three-dimensional array representing a two-dimensional spatial image and one-dimensional spectral information. Currently used by NASA to differentiate between geographical features via satellite,¹ this evolving imaging modality shows promise in medical imaging. By capturing a spectrum of light intensities over a range of wavelengths, a hyperspectral image provides both spectral and spatial information. Hyperspectral cameras detect the intensity of light within hundreds of narrow spectral bands for each pixel. The spectral bands provide rich information for tissue classification. Previous studies have demonstrated the feasibility of support vector machines (SVM) to perform pixel-wise classification of hyperspectral images.² Studies have shown that support vector machines can be highly modified to work well with large scale data sets such as hyperspectral images.³ Other studies have produced effective results of combining support vector machines with other segmentation techniques.⁴ Linear support vector machines and new hybrid support vector selection have been shown to effectively classify hyperspectral images.⁵

Previous studies have also focused on methods of segmentation using contours of spectral information, but fail to perform distinctive classification of these regions.⁶ Minimum spanning forests have been shown to be an effective way of segmenting geographic landscapes from hyperspectral imagery.⁷ Two-dimensional segmentation of grayscale images using minimum spanning trees has been studied extensively⁸ but recently it has been adapted to work with hyperspectral images without the need for manual marker selection.⁹

Hyperspectral imaging has also been shown to be a potential imaging modality for cancerous tissue identification.¹⁰ This imaging modality has also been used to aid in the segmentation and classification of breast cancer regions of interest,¹¹ these studies have shown that the cancer spectra can be distinguished using hyperspectral imaging. This study focuses on expanding upon these techniques to increase accuracy by introducing a spectral-spatial segmentation component to the spectral classification methods. The novel part of this study comes from the new adaptive dissimilarity measures that are calculated while the minimum spanning forest is being grown.

The method to be introduced relies upon SVM classification and probability data to determine which pixels are most likely to be classified correctly. These pixels are then determined through an algorithm designed by⁹ to serve as markers for which spanning trees are grown. These trees are minimized within a minimum spanning forest to segment the hyperspectral image. The segments then use majority voting of the SVM classification to determine which value they should be given. This method is shown to increase the accuracy of the SVM by incorporating spatial information within the spanning forests along with the spectral information the SVM uses for classification.

2. METHODS

2.1 Hyperspectral Imaging System

To capture the hyperspectral image data, a CRi camera system (PerkinElmer, Hopkinton, MA) was used to acquire images from animals. The system is a light-tight apparatus that

uses a Cermax-type 300 Watt Xenon light source. This provides lights that span the electromagnetic spectrum from 450–950 nm. The CCD is a 16-bit, high-resolution, scientific-grade imaging sensor. Four, fiber-optic, adjustable illuminator arms yield an even light distribution to the subject. The light radiates from the excitation source and then illuminates the sample. Reflect lights pass through the camera lens to the solid-state liquid crystal tuning element and finally to the CCD. The excitation and emission filter sliders hold two, 50-mm diameter, long pass filters. The long pass filters remove the band light especially from the excitation source. These filters are color coded to indicate the wavelength they represent. The field of view (length \times width) is from 3.4 \times 2.5 cm to 10.2 \times 7.6 cm with variable zoom. The resolution is from 25 to 75 μ m based on the zoom lens position. The fluence rate, i.e., the flow of photons over a set angle, is from 4 to 20 mW/cm² based on the light position. The scan time is from 5 sec to 1 min.¹⁰ The images were then normalized using the method as previously reported by us.¹⁰

2.2 Hyperspectral Imaging Experiment in Animals

We used tumor-bearing mice for the HSI experiments. A head and neck tumor xenograft model using HNSCC cell line M4E was adopted in the experiment. The HNSCC cells (M4E) were maintained as a monolayer culture in Dulbecco's modified Eagle's medium (DMEM)/F12 medium (1:1) supplemented with 10% fetal bovine serum (FBS). M4E-GFP cells which are generated by transfection of pLVTHM vector into M4E cells were maintained in the same condition as M4E cells. Animal experiments were approved by the Animal Care and Use Committee of Emory University. Four female mice aged 4–6 weeks were injected with 2×10^6 M4E cells with green fluorescence protein (GFP) on the lower back. During the image acquisition, each mouse was anesthetized with a continuous supply of 2% isoflurane in oxygen. First, both the interior infrared and the white excitation were opened for reflectance image acquisition with 50 ms exposure time. Reflectance images contain 251 spectral bands from 450 to 950 nm with 2 nm increments. Then, blue excitation and 50 ms exposure time were selected for the fluorescence image acquisition. Tumors show green signals in fluorescence images due to tumor cells with GFP. In this study, we used GFP fluorescence images as the gold standard for the evaluation of cancer detection by HSI.

2.3 Support Vector Machine (SVM) based Classification

The proposed algorithm calls for an initial classification to be completed using a support vector machine. For this study a least squares SVM (LS-SVM) is proposed to conduct the classification.¹⁰ This SVM model has been used in previous cancer detection methods.² The SVM was trained using manually selected pixels contained within the primary tumor as well as pixels contained within the normal tissue. This specific SVM model was then used to predict the classification of each pixel to detect tumor masses. The classification also yielded probability estimates for each classified pixel and was implemented using the LIBSVM¹² library following the specifications as previously described.¹³

2.4 Marker Selection

Markers are selected from the SVM classification and probability maps using the method as previously reported;⁹ the SVM classification map is first segmented into regions of

connected components. These regions are then classified as large or small based upon their area. The following rules taken from ⁹ govern marker selection for each region type:

1. In a large region, the top P% of pixels with the highest probability estimates are selected to be used as markers,
2. Pixels contained within small regions will only be selected as a marker if their probability exceeds a predetermined threshold S.

The spectrums of these markers are then averaged for each band to create a single characteristic spectrum $S_r(sp) = (I_1, \dots, I_B)$ for each region type where B is the number of spectral bands in each pixels and I is the corresponding intensity, where sp is the set of all pixels p within that region type.

2.5 Region Growing Using a Minimum Spanning Forest (MSF) Approach

The proposed method has thus far created a classification map for specific markers meeting the above criteria. The next step involves the expanding of these classifications to the nearest neighbors of the markers in order to classify all pixels within the image. These classifications are determined by treating each marker as the root of tree and classifying each pixel within a minimum spanning forest.¹⁴ This forest is grown according to the prim's algorithm ¹⁵ with the use of a binary heap to efficiently store the queue.

For this method dissimilarity measures between a pixel and its 8 nearest neighbors must be constructed. Many dissimilarity measures have been suggested for image segmentation.¹⁶ The proposed measures in a previous work ⁹ consisted of a vector norm (L1), spectral angle mapper (SAM), and spectral information divergence (SID). For this study, the L1 and SAM measures were evaluated, and a vector norm taken with the first derivative (D1) was tested and the dissimilarity measurement with the best results was used.

The L1 dissimilarity measure is calculated between two vector pixels $p_i = (p_{i1}, \dots, p_{iB})$ and $p_j = (p_{j1}, \dots, p_{jB})$ where B is the number of bands for each pixel. The L1 measure is then calculated by

$$L1(p_i, p_j) = \sum_{b=1}^B |p_{ib} - p_{jb}| \quad (1)$$

The SAM is used to determine the spectral angle between the two pixel vectors and is defined as:

$$SAM(p_i, p_j) = \cos^{-1} \frac{\sum_{b=1}^B p_{ib} p_{jb}}{\sqrt{\sum_{b=1}^B p_{ib}^2 \sum_{b=1}^B p_{jb}^2}} \quad (2)$$

The D1 vector norm measure compares the difference in spectral shape instead of intensity as the L1 vector norm does. The D1 vector norm first converts the spectral information for each pixel from an intensity measurement to a measurement of the spectral shape. This resulting vector can then be compared using an L1 vector norm comparison and is given by:

$$D1(p_i, p_j) = \sum_{b=1}^{B-1} \left| \frac{(p_{ib+1} - p_{ib})}{n} - \frac{(p_{jb+1} - p_{jb})}{n} \right| \quad (3)$$

Previous algorithms⁹ calculate these weightings for the 8 nearest neighbors and store them in a binary heap. The proposed algorithm calculates the dissimilarity weighting $w_{i,j}$ between a pixel's 8 neighbors but includes the dissimilarity between the adjacent pixel p_j and S_r of the region identifying the pixel p_i . The weighting

$$L2_{i,j} = \begin{cases} L1(p_i, p_j) + L1(p_j, S_r) \\ SAM(p_i, p_j) + SAM(p_j, S_r) \\ D1(p_i, p_j) + D1(p_j, S_r) \end{cases} \quad (4)$$

thus provides additional grouping measurements beyond nearest neighbor dissimilarities that increases accuracy along spectral gradients. This additional dissimilarity comparison is helpful in distinguishing blurred regions as were observed.

A minimum spanning forest M is constructed from a connected graph $G = T(V, E)$ consisting of a set of separated unconnected trees $T = (V, E)$ rooted on m separated vertices $t_i, i = 1, \dots, m$, where V represents pixels within the image and E represents the associated dissimilarity measure between those pixels. The forest M consists of a spanning forest F within the set of all spanning forest SF of unconnected trees T such that the sum of the weights $w_{i,j}$ of edges E in graph M is minimal.

$$F \in \arg \min_{F \in SF} \left\{ \sum_{e_{i,n} \in E_F} w_{i,j} \right\} \quad (5)$$

The algorithm implemented to construct this forest in $O(|E| \log |V|)$ time complexity is Prim's Algorithm, with the edge weightings stored within a binary minimum heap.¹⁷ Each pixel is then classified by giving it the same value as the marker with which it was originally grown.

2.6 Majority Voting for Each Tree

The classification map was constructed under the assumption that the selected markers were properly classified by the SVM. Majority voting is then performed to verify these classifications by assigning to each pixel within a given tree T , the classification held by the majority of those pixels in the original SVM classification map. This has shown good results in previous works to increase the accuracy of the classification.¹

2.7 Evaluation

Accuracy, sensitivity, specificity are commonly used performance metrics.^{18–26} In this study, we used GFP fluorescence images as the gold standard for the evaluation of cancer detection by HSI. Sensitivity measures the proportion of actual positives which are correctly identified as positive, i.e. the percentage of tumor pixels which are correctly identified as tumor tissue. Specificity measures the proportion of negatives which are correctly identified,

i.e. the percentage of healthy tissue correctly identified as not having cancer. Accuracy is defined as the sum of true positive and true negative divided by the total number of pixels of both tumor and normal tissue.

3. RESULTS

3.1 Simulation Images

This algorithm was applied to simulation images in which a primary circle was surrounded by smaller secondary circles representing primary and secondary tumors as seen in Figure 3A. The intensity values for the primary circle were taken from random pixels contained near the center known tumor section of the mouse image. The average spectral values for the pixels selected for cancerous and normal tissues are shown in Figure 2.

The values at the secondary circles were taken from random pixels within an outlying region still contained within a known tumor section of the mouse image. The background intensity values on the simulation image were taken from randomly selected pixels contained within a healthy tissue section of the same mouse image. The simulation image was implemented to test the algorithm. The simulation was run based on four separate mouse images and the results are shown below. As shown in Figure 3A, the central and secondary circles are almost invisible on the simulation image. The classification algorithm was still able to segment the circles and achieved a high accuracy of more than 97%.

3.2 Spectral Band and Dissimilarity Measure Selection

Figure 4 shows the spectra of normal and cancer tissue, which were used to decide the appropriate selection of spectral bands to segment cancerous tissue from normal tissue. From these spectra it was seen that the range of 550 to 850 was a region of great dissimilarity that could be used to distinguish these tissue types. It was noted that while the intensities were different the spectral shape was of no great distinction; therefore the choice to use the L1 and L2 methods over the SAM or D1 dissimilarity measures was made.

3.3 Mouse Image Data

The classification algorithm was also applied to the in vivo hyperspectral images that were acquired from tumor-bearing mice. Figure 5 shows an example of the hyperspectral images and the classification results. The method was able to detect the tumor region. As shown in Table 2, the quantitative assessment of the classification method is accurate for detecting the tumor in four animals.

As seen in Figure 6 the greatest challenge in segmenting the tumor is on the shadowed boundaries. From Figure 6.D it is seen that the misclassified regions are along the boundaries that are not only changes in tissue type but also in elevation. The gradient shift along the elevation of the tumor region could have effects on the spectra leading to errors.

4. DISCUSSION

To the best of our knowledge this is the first time the minimum spanning forest approach has been applied to classify cancerous tissue in hyperspectral images. The algorithm allows

for the improved dissimilarity measure to be introduced emphasizing the spectral information of select tissue types while including the spatial information in the minimum spanning forest approach. The challenges in this imaging modality and segmentation are finding the specific bands that are best used for image classification. The minimum spanning forest approach can be computationally intensive and the implementation may need to be optimized in order to perform rapid, quantitative evaluation of many different spectral regions.

In this study the cancerous tissue was only present under a thin layer of skin. This layer may have acted as a barrier, shielding spectral information that can be used to better classify different types of tissue. This method demonstrates that even through this skin barrier the dissimilarity measure was still able to distinguish the cancerous from normal tissue with a high accuracy and provide useful boundaries for cancer detection.

5. CONCLUSIONS AND FURTHER STUDIES

The minimum spanning forest algorithm combined with the support vector machine is able to more accurately classify cancerous tissues compared to the previous region growing algorithms or support vector machines alone. The hyperspectral imaging modality may provide the necessary verification that surgeons need when accessing the surrounding tumor bed for cancerous tissue. Further studies will examine the ability for this algorithm to classify hyperspectral images in which the tumor bed is fully exposed unlike the images above which required the light to penetrate the skin. Higher wavelengths should also be captured as greater cancer discrimination has been observed in the infrared and far infrared regions. Further work will include the incorporation of multiple dissimilarity measures to obtain a spectral shape and intensity dissimilarity measurement.

Furthermore, the adaptive dissimilarity measure explored in this study will be further investigated. In this study the dissimilarity was measured against the marker roots of which class it was attempting to join, further studies will examine the measurement of a pixel not only to its neighbors, but include all pixels included in the branches from which its neighbors are joined. This may provide greater segmentation accuracy along slow moving gradients and better distinguish blurred boundaries.

Acknowledgments

This research is supported in part by NIH grants (R01CA156775, R21CA176684, and P50CA128301) and Georgia Cancer Coalition Distinguished Clinicians and Scientists Award.

References

1. Tarabalka Y, Chanussot J, Benediktsson JA. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*. 2010; 43(7):2367–2379.
2. Akbari H, Halig LV, Schuster DM, Osunkoya AO, Master VA, Nieh PT, Chen GZ, Fei B. Hyperspectral imaging and quantitative analysis for prostate cancer detection. *J Biomedical Optics*. 2012; 17:076005.
3. Li XR, Zhu JE, Wang J, Zhao LY. Hyperspectral image classification based on composite kernels support vector machine. *Journal of Zhejiang University (Engineering Science)*. 2013; 47(8):1403–1410.

4. Priego B, Souto D, Bellas F, Duro RJ. Hyperspectral image segmentation through evolved cellular automata. *Pattern Recognition Letters*. 2013; 34(14):1648–1658.
5. Kaya GT. A Hybrid Model for Classification of Remote Sensing Images With Linear SVM and Support Vector Selection and Adaptation. *IEEE J Selected Topics in Applied Earth Observations & Remote Sensing*. 2013; 6(4):1988–1997.
6. de la Vega SH, Manian V. Object segmentation in hyperspectral images using active contours and graph cuts. *International Journal of Remote Sensing*. 2012; 33(24):1246–1263.
7. Bernard K, Tarabalka Y, Angulo J, Chanussot J, Benediktsson JA. Spectral–Spatial Classification of Hyperspectral Data Based on a Stochastic Minimum Spanning Forest Approach. *IEEE Transactions on Image Processing*. 2012; 21(4):2008–2021. [PubMed: 22086502]
8. Xu Y, Oberbacher EC. 2D image segmentation using minimum spanning trees. *Image and Vision Computing*. 1997; 15(1):47–57.
9. Tarabalka Y, Chanussot J, Benediktsson JA. Segmentation and Classification of Hyperspectral Images Using Minimum Spanning Forest Grown From Automatically Selected Markers. *IEEE Trans Syst Man Cybern Part B-Cybern*. 2010; 40(5):1267–1279.
10. Fei, B.; Akbari, H.; Halig, LV. Hyperspectral imaging and spectral-spatial classification for cancer detection. *The 5th IEEE International Congress on Image and Signal Processing (CISP 2012) and the 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012)*; 2012. p. 62-64.
11. Kim, B.; Kehtarnavaz, N.; LeBoulluec, P.; Liu, H.; Peng, Y.; Euhus, D. Automation of ROI extraction in hyperspectral breast images. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*; 2013. p. 3658-3661.
12. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*. 2011; 2(3):1–27.
13. Lv, G.; Yan, G.; Wang, Z. Bleeding detection in wireless capsule endoscopy images based on color invariants and spatial pyramids using support vector machines. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*; 2011. p. 6643-6646.
14. Stawiaski, J. PhD dissertation. Paris School Mines; Paris, France: 2008. *Mathematical morphology and graphs: Application to interactive medical image segmentation*.
15. Prim RC. Shortest connection networks and some generalizations. *Bell System Technical Journal*. 1957; 36(6):1389–1401.
16. Makrogiannis S, Economou G, Fotopoulos S. A region dissimilarity relation that combines feature-space and spatial information for color image segmentation. *IEEE Trans Syst Man Cybern Part B-Cybern*. 2005; 35(1):44–53.
17. Cormen, TH.; Leiserson, CE.; Rivest, RL.; Stein, C. *Intro to Algorithms*. MIT Press; 2009.
18. Qin X, Cong Z, Fei B. Automatic segmentation of right ventricular ultrasound images using sparse matrix transform and a level set. *Phys Med Biol*. 2013; 58(21):7609–24. [PubMed: 24107618]
19. Yang X, Fei B. Multiscale segmentation of the skull in MR images for MRI-based attenuation correction of combined MR/PET. *J Am Med Inform Assoc*. 2013; 20(6):1037–45. [PubMed: 23761683]
20. Wang H, Fei B. Nonrigid point registration for 2D curves and 3D surfaces and its various applications. *Phys Med Biol*. 2013; 58(12):4315–30. [PubMed: 23732538]
21. Fei B, Yang X, Nye JA, Aarsvold JN, Raghunath N, Cervo M, Stark R, Meltzer CC, Votaw JR. MR PET quantification tools: registration, segmentation, classification, and MR-based attenuation correction. *Med Phys*. 2012; 39(10):6443–54. [PubMed: 23039679]
22. Yang X, Wu S, Sechopoulos I, Fei B. Cupping artifact correction and automated classification for high-resolution dedicated breast CT images. *Med Phys*. 2012; 39(10):6397–406. [PubMed: 23039675]
23. Akbari H, Fei B. 3D ultrasound image segmentation using wavelet support vector machines. *Med Phys*. 2012; 39(6):2972–84. [PubMed: 22755682]
24. Yang X, Fei B. A multiscale and multiblock fuzzy C-means classification method for brain MR images. *Med Phys*. 2011; 38(6):2879–91. [PubMed: 21815363]
25. Wang H, Fei B. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. *Med Image Anal*. 2009; 13(2):193–202. [PubMed: 18684658]

26. Lu G, Fei B. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*. 2014; 19 (In Press).

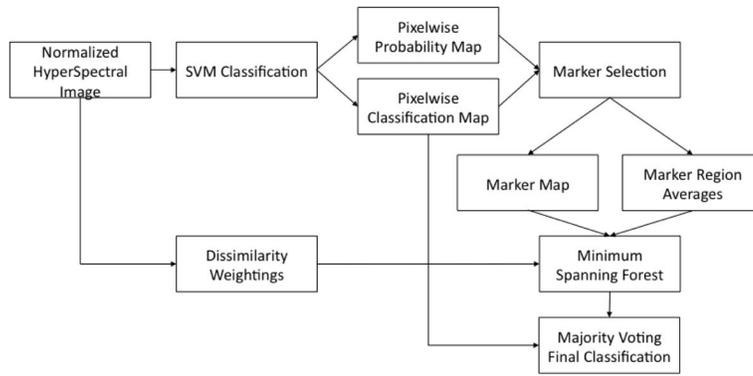


Figure 1. Algorithm flow chart showing the progression from a normalized hyperspectral image being first classified spectrally by a support vector machine and then being further classified spatially using a minimum spanning forest.

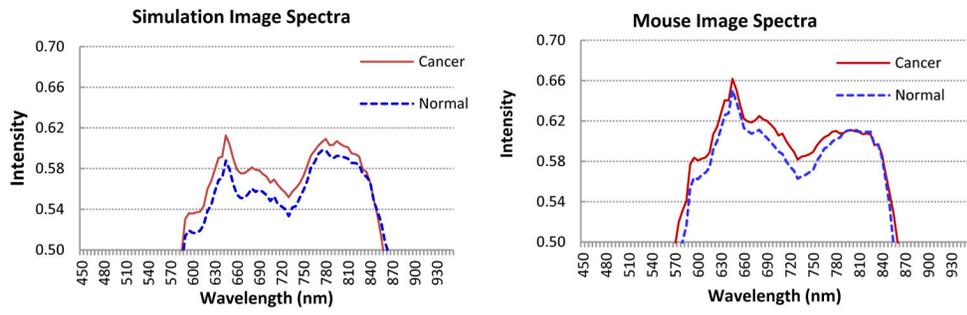


Figure 2. Average spectral values for pixels contained within the cancer and normal regions for the simulation (left) and for in vivo mouse images (right).

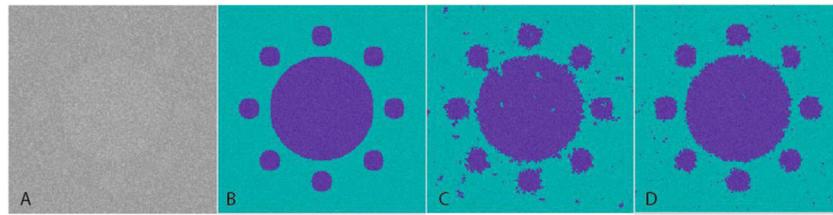


Figure 3. Classification results of the simulation image (A) compared to the gold standard (B), the results using the L1 method (C) and the L2 method (D).

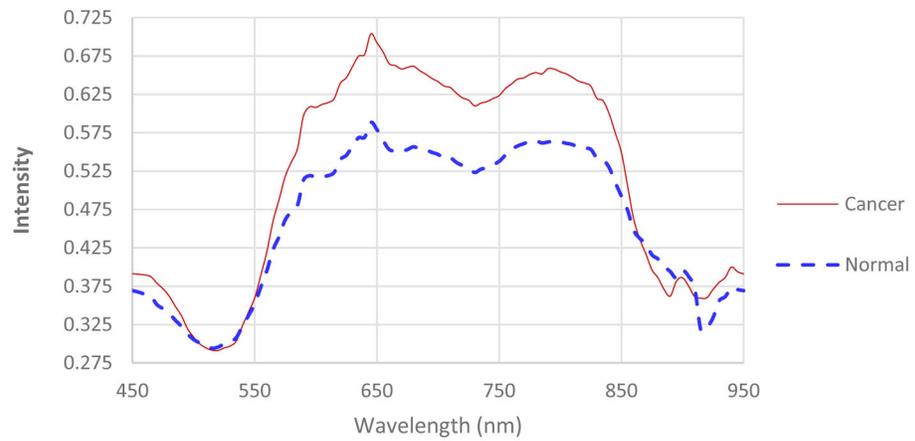


Figure 4. Typical spectral intensities of normal and cancer tissue. Gold-standard data from green fluorescence protein (GFP) imaging were used to separate cancerous and normal tissue. The average intensity values were used across all pixels of each tissue type.

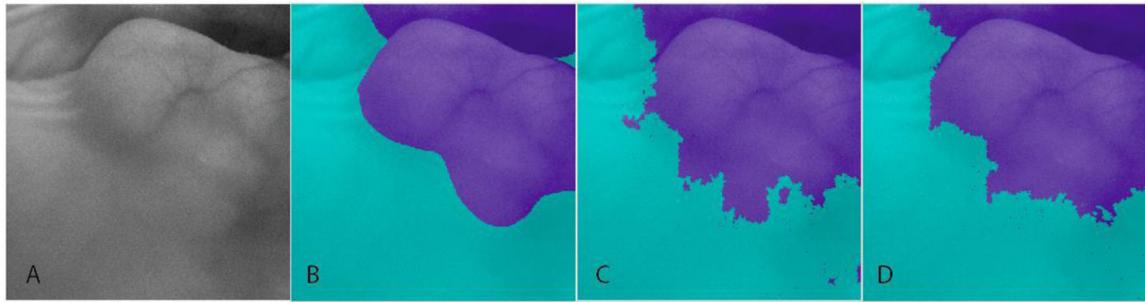


Figure 5. Classification results of the mouse image (A) compared to the gold standard data (B), and the results using the L1 method (C) and the L2 method (D).

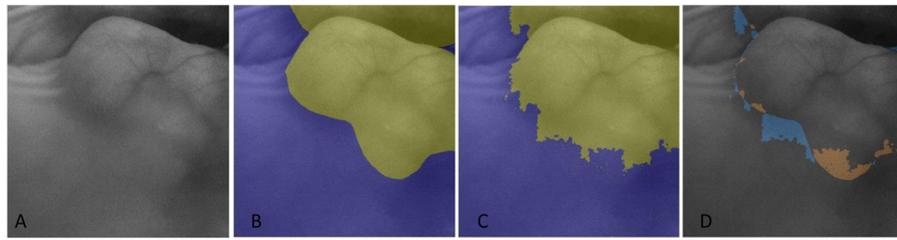


Figure 6.
Classification results of the mouse image (A) compared to the gold standard (B) using the L2 method (C) with the difference image (D)

Table 1

Results from the simulation images using the SVM, L1 and L2 methods

	SVM	L1	L2
Sensitivity (%)	95.9 ± 1.6	95.3 ± 2.1	97.8 ± 0.6
Specificity (%)	92.4 ± 4.9	97.2 ± 0.7	96.9 ± 0.4
Accuracy (%)	93.3 ± 3.8	96.7 ± 0.2	97.2 ± 0.1

Table 2

Results from the mouse images using the SVM, L1 and L2 methods

	SVM	L1	L2
Sensitivity (%)	89.9 ± 1.9	90.7 ± 2.6	93.6 ± 2.4
Specificity (%)	97.1 ± 2.2	98.0 ± 2.3	98.0 ± 2.2
Accuracy (%)	94.1 ± 0.6	95.3 ± 0.9	96.0 ± 0.8