Received: 20 September 2019 | Revised: 12 November 2019 | Accepted: 5 December 2019

DOI: 10.1002/jbio.201960125

FULL ARTICLE

JOURNAL OF BIOPHOTONICS

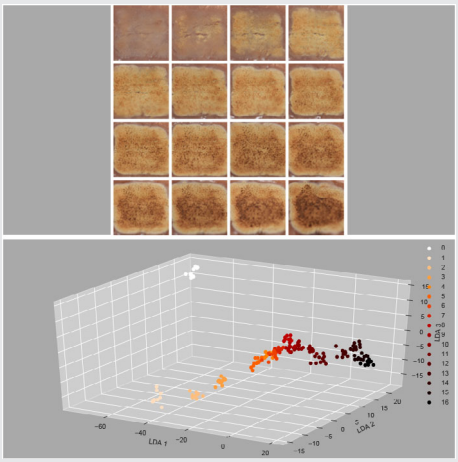
Diffuse reflectance spectroscopy as a monitoring tool for gastric mucosal devitalization treatments with argon plasma coagulation

Juan D. Osorio | **Sergio Vilches*** | **Hans Zappe**

Gisela and Erwin Sick Chair of Micro-optics, University of Freiburg, Germany

***Correspondence**
Sergio Vilches, Gisela and Erwin Sick Chair of Micro-optics, University of Freiburg, Germany.
Email: sergio.vilches@imtek.uni-freiburg.de

Abstract



Electrosurgery with argon plasma coagulation is a widespread technique used in various medical fields for applications which range from hemostasis to devitalization processes. Developing tools which provide feedback concerning tissue condition during these surgeries is fundamental for improving the safety and success of this treatment. We present here a method based on diffuse reflectance spectroscopy to monitor gastric mucosal devitalization treatments. The analysis of the diffusely reflected spectra of the tissue allows us to differentiate between ablation states by using linear discriminant analysis (LDA) as a classification algorithm. An ex vivo pilot study on several swine stomachs showed promising results, with 97.8% of correctly classified ablation states on a new unseen stomach, encouraging further tests with human tissue.

KEYWORDS

argon plasma coagulation, diffuse reflectance spectroscopy, endoscopy, gastric mucosa devitalization, linear discriminant analysis, thermal ablation

1 | INTRODUCTION

Argon plasma coagulation (APC) is a technique used in electrosurgery to induce thermal effects on the superficial layers of organic tissue in a contact-less manner. This technology is widely used in open surgery and in endoscopy across many medical specialist fields, such as bronchology, otolaryngology, gastroenterology, with primary applications in tissue shrinkage to treat swollen tissue; devitalization processes such as malignant tumor destruction; and minimization of blood loss through hemostasis [1]. Surgeons control the effects of APC

Abbreviations: APC, argon plasma coagulation; DRS, diffuse reflectance spectroscopy; GMD, gastric mucosa devitalization; LDA, linear discriminant analysis.

visually, observing the change in color and texture of the treated tissue until the desired effect is reached. This method loses efficacy when the visual information is impaired, for example, in the case of endoscopic surgery. Furthermore, the plasma itself, intrinsic to APC, temporarily blinds the surgeon's view or saturates endoscopic imaging systems, requiring periodic interruptions of the procedure to assess the ablation state of the tissue. In this context, we propose diffuse reflectance spectroscopy (DRS) as a complementary aid for monitoring APC treatments continuously and in real time. More specifically, this work addresses gastric mucosal devitalization (GMD) using APC techniques, which have been proven technically feasible for reducing body weight and visceral adiposity in a swine ex vivo model in its preliminary investigative stages [2].

The proposed approach is based on illuminating the tissue under treatment with a broadband, high intensity light source; collecting the photons which are diffusely reflected (DR) from the superficial layers of the tissue; and finally determining the spectrum of this light signal with a spectrometer.

In the following, we first show that APC induces a continuous change in the diffuse reflectance spectrum of porcine stomach tissue. With these results, three ablation states are defined and a classification problem among these states is formulated. We then undertake the classification of these ablation states using machine learning techniques.

2 | EXPERIMENTAL SETUP

2.1 | Tissue

The ablation experiments were realized on ex vivo swine stomach tissue, specifically on the proper gastric region of the pig, corresponding to the *body* and *fundus* regions of the human stomach. A total of five stomachs were obtained from a slaughterhouse and stored frozen until thawed before the experimental procedures. The age, gender, or race from the subjects could not be determined.

2.2 | APC

The electrosurgery APC equipment used in the experiments was the *ERBE VIO APC* system with a *VIO 300 D* generator module and an APC 2 argon plasma coagulator. It was configured with the following settings: *Pulsed mode*, *Effect 2* (16 Hz pulse signal), 0.8 L/min of argon flow and fixed maximum power of 40 W.

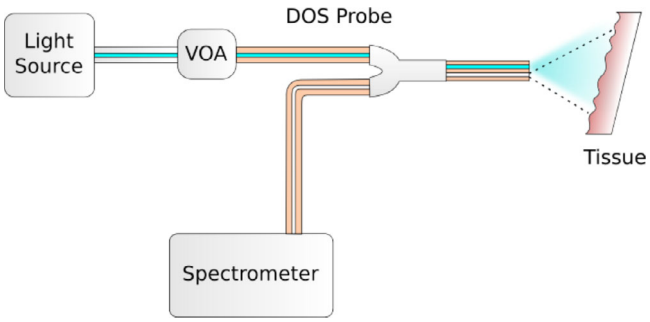


FIGURE 1 Schematic of the optical setup used to obtain the DRS measurements

2.3 | DRS measurements

The optical setup that was used to obtain the DRS measurements is presented schematically in Figure 1. An *Energetiq EQ-99FXC* high-intensity (60 μW/nm) continuum fiber-coupled light source (190-2100 nm) was used to illuminate the tissue. A mechanical variable optical attenuator was used to adjust the optical power delivered to the tissue. A DRS probe containing two multimode fibers with 400 μm core diameter and numerical aperture 0.5 was assembled to illuminate and collect the backscattered light from the tissue. The tip of the probe was placed at a fixed distance of 13 mm and at a fixed 30° angle from the tissue's surface. The *Ocean Optics HR2000* + high-resolution (0.2 nm) spectrometer with a wavelength range from 600 to 1020 nm was used to capture the spectral measurements. The signals were acquired using an integration time of 250 ms and averaged every four consecutive measurements.

3 | DRS VARIATION WITH ABLATION CYCLES

Using the previously described experimental setup, the variation of the DR spectra was analyzed with respect to the ablation process in a single swine stomach. To do so, 16 ablation cycles were executed on the tissue, followed by the corresponding DRS measurement. Each ablation cycle consisted of ablating a rectangular section of 1.8 cm by 1.5 cm while following a meander pattern that covered this area in approximately 6 seconds. The corresponding ablation effects on the tissue are shown in Figure 2, as they progressively change with each ablation cycle.

After each ablation cycle, 10 partially overlapping DRS measurements were taken across the previously ablated rectangular area. The mean spectral measurements corresponding to each ablation cycle are shown in Figure 3. All these spectral signals were previously preprocessed, as described in Section 4, but shown before

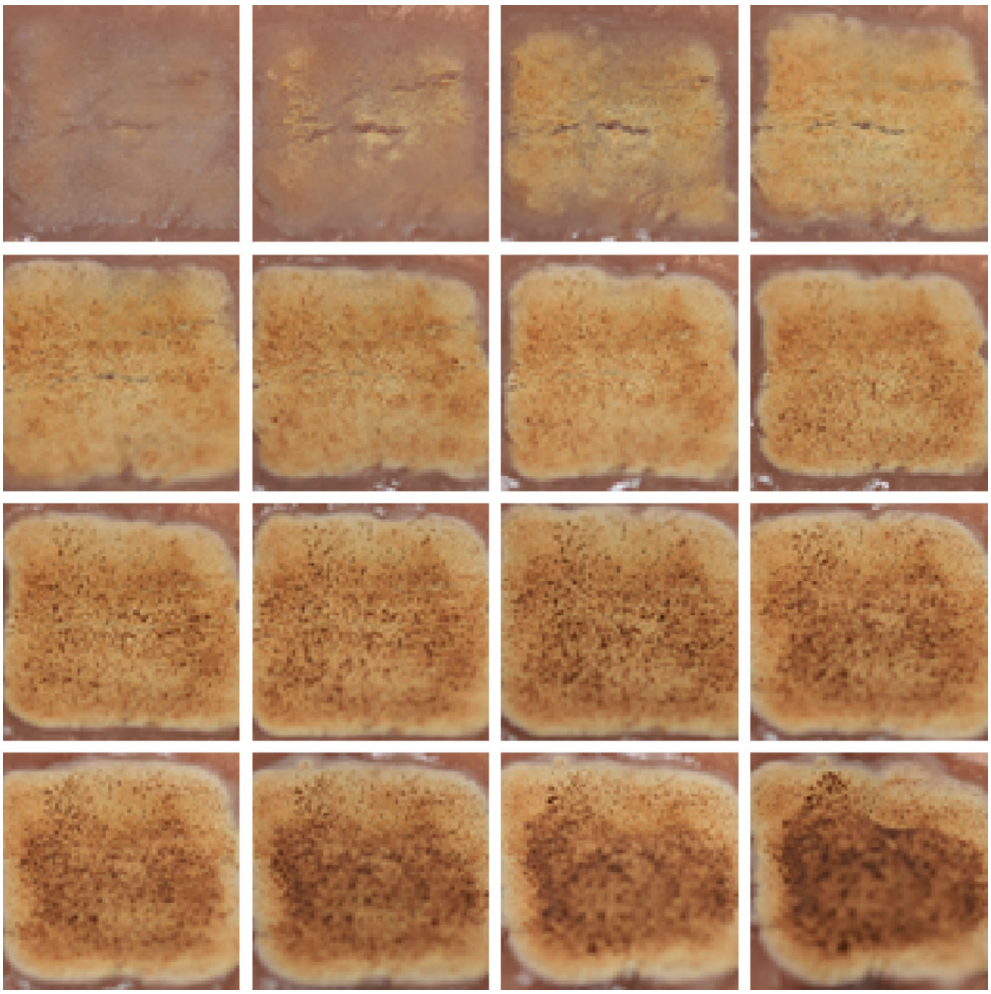


FIGURE 2 Ablation effects on tissue ranging from ablation cycle 1 (top-left corner) to ablation cycle 16 (bottom-right corner). Every photograph depicts an area of 1.8 by 1.5 cm

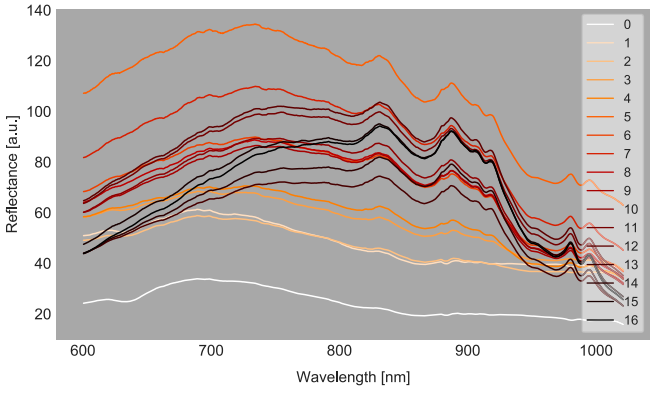


FIGURE 3 Variation of the DRS spectra corresponding to the number of ablation cycles shown in the legend

normalization. With this experiment, it was empirically demonstrated that the DR spectra vary gradually as the APC process is executed on the tissue, and this variation is determined by the number of ablation cycles inflicted on the tissue.

We picked three meaningful states for the GMD/APC procedure based on the number of ablation cycles: *normal*, *desiccated*, and *brown*. *Normal* corresponds to a state of no ablation, which indicates the surgeon an unoperated tissue area, ready to be treated. *Dessicated* is equivalent to three ablation cycles and may be interpreted as a state where the surgeon needs to exercise caution. *Brown* refers to seven ablation cycles, which represent the stop state, indicating to the surgeon that no further ablation is needed. These ablation states were used as classes in a classification problem, which correspond to the labels of each spectrum used in the posterior data generation process.

4 | ABLATION STATE CLASSIFICATION

4.1 | Data generation

A total of 1350 DR spectra were measured from five different ex vivo swine stomachs. Each stomach underwent the ablation procedure described in the previous section, using the same experimental setup, where 90 samples per

ablation state were collected and labeled accordingly, totaling 270 DR spectral samples per stomach. The 1080 samples from the first four stomachs were used for training and validation, while the 270 samples from the fifth stomach were used for the final test. The data analysis and ablation state classification process was developed under a **python** [3] based environment, using as main libraries: **scikit-learn** [4], **numpy** [5], **scipy** [6], **pandas** [7] and **matplotlib** [8].

4.2 | Noise filtering

In addition to the temporal averaging realized in the data acquisition phase, a moving average filter with a kernel of 25 wavelength points was applied to the data to further reduce the random noise implicit in the spectrum while preserving its shape. The noise level decreases by the square root of both the number of points used in the moving average and the number of scans temporally averaged, achieving an overall noise reduction by a factor of 10.

4.3 | Calibration

All collected signals were calibrated to obtain the wavelength dependent reflectance given by

$$R(\lambda) = \frac{I_c(\lambda) - I_b(\lambda)}{I_r(\lambda) - I_b(\lambda)} \times 100\%, \quad (1)$$

where I_c stands for every collected diffuse reflected spectral measurement and I_b and I_r correspond to the background and the reference light signals, respectively, both of which were assumed constant for all measurements. A diffuse reflector with silver coating was used to measure I_r.

4.4 | Data normalization

Normalization was undertaken to minimize the dependency of the measurements on the reflectance amplitude that varied considerably without normalization, as may be observed in the upper graph in Figure 4. It was also applied to minimize the influence of uncontrollable factors present in a medical APC procedure, such as the distance and angle from fiber to tissue; tip contamination or illumination changes.

After applying L2 normalization, an average reduction of 87.8% in the coefficient of variation was obtained among the different classes in the training data set, when compared with the unnormalized reflectance measurements. The Euclidian norm (L2) was also chosen as it resulted in the largest reduction when compared with other types of norm, such as the Manhattan norm (L1) or a maximum norm.

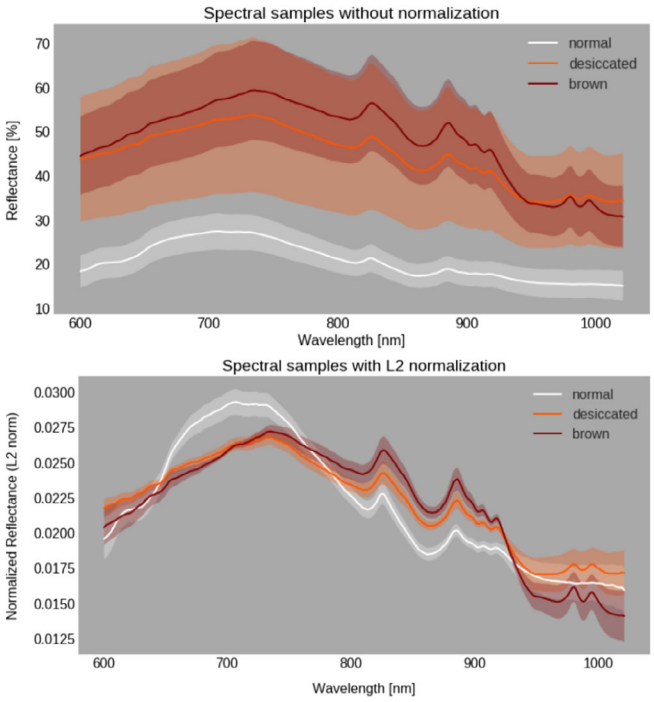


FIGURE 4 DRS samples with and without normalization. The SD around the mean is observed for each ablation state in both graphs ($\mu \pm \sigma$)

4.5 | Wavelength subsampling

Finally, wavelength subsampling was applied to extract meaningful features that represented the overall shape of the spectra characterizing each ablation state. The wavelength subsampling was realized every 25 wavelengths, to match the number of pixels that were averaged in the moving average filter; thus, 80 wavelengths were selected from the total 2000 wavelengths that were measured. In this way, feature colinearity was reduced by eliminating redundant information about neighboring pixels, as well as its sparsity, as it was transformed to a much lower dimensional space.

4.6 | Classification model

Linear discriminant analysis (LDA) was chosen as the classification algorithm used to discriminate among the three defined ablation states. The training data were used to find the weight parameters of a linear discriminant

TABLE 1 Leave-one-out crossvalidation evaluation

	Recall (%)			Precision (%)		
Accuracy (%)	N	D	B	N	D	B
99.7	100.0	99.4	99.7	100.0	99.7	99.4

function that maximizes separability among these three classes. This is realized by maximizing the Fisher Criterion for the generalized multiclass case [9]

$$J(\mathbf{W}) = \frac{\left|\mathbf{W}^\top \mathbf{S}_B \mathbf{W}\right|}{\left|\mathbf{W}^\top \mathbf{S}_W \mathbf{W}\right|}, \tag{2}$$

with respect to $\mathbf{W}$, which is the weight parameter matrix. $\mathbf{S_B}$ represents the between-class scattering matrix calculated as [10]

$$\mathbf{S}_B = \sum_{k=1}^{K} N_k (\mathbf{m}_k - \mathbf{m})(\mathbf{m}_k - \mathbf{m})^\top, \tag{3}$$

where N_k are the number of samples in each class; m_k is the mean of each class; and m is the overall mean. $\mathbf{S_W}$ is the within-class scattering matrix given by [10]

$$\mathbf{S}_W = \sum_{k=1}^{K} \mathbf{S}_k = \sum_{i=1}^{K} \sum_{\mathbf{x} \in C_k}^{n} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^\top. \tag{4}$$

The solution to this problem was found using the singular value decomposition solver implemented in the LDA module of the `Scikit-learn` [4] library. As a result, a discriminant function of the form

$$y(\mathbf{x}) = w_0^\top + w_1^\top x(\lambda_1) + w_2^\top x(\lambda_2) + \ldots + w_{80}^\top x(\lambda_{80}), \tag{5}$$

was obtained. This function corresponds to the prediction model, which may be applied on new DR spectral samples to obtain a prediction of its ablation state.

One of the biggest advantages of the using LDA as a classification algorithm is that the weight parameters can be computed analytically, since the solution towards finding them have a closed form. This means that the models are fast and simple to use both in training and recall time, which is a very desirable characteristic in medical applications as it may enable real-time ablation state assessment, or an eventual model retraining in a subject-specific calibration procedure.

TABLE 2 Intersubject crossvalidation evaluation

		Recall (%)			Precision (%)		
Hold-out	Accu. (%)	N	D	B	N	D	B
1	95.2	100	98.9	86.7	98.9	88.1	100
2	97.4	100	92.2	100	100	100	92.8
3	97.0	100	91.1	100	100	100	91.8
4	97.8	100	100	93.3	100	93.8	100
Avg.	96.9	100	95.6	95.0	99.7	95.5	96.2

5 | VALIDATION EVALUATIONS

The following two types of validation evaluations were realized on the 1080 spectral samples of the training data set to evaluate the performance of the model.

5.1 | Leave-One-Out Cross Validation

In this evaluation, the model was trained and validated 1080 times, corresponding to the number of measurements in the training set. In each iteration, a single different sample was left out for validation, while the rest of the 1079 data set samples were used for training. The accuracy, recall, and precision metrics were measured as shown in Table 1. The results for all metrics were above 99.4%, suggesting a strong model performance in this type of scenario.

5.2 | Intersubject Cross Validation

In this evaluation, the model was trained and validated four times, corresponding to the number of stomachs in the training set. In each iteration, the samples of one stomach were left out of the validation, while the samples from the rest of the stomachs were used for training. The same evaluation metrics were measured as shown in Table 2. The average metrics for all cases were above or equal to 95.0%, suggesting promising generalization performance for new subjects.

We observed discrepancies in the performance among the different iterations of the cross validation. It was

TABLE 3 Results on test set

	Recall (%)			Precision (%)		
Acc. (%)	N	D	B	N	D	B
97.8	100.0	98.90	94.4	100.0	94.7	98.8

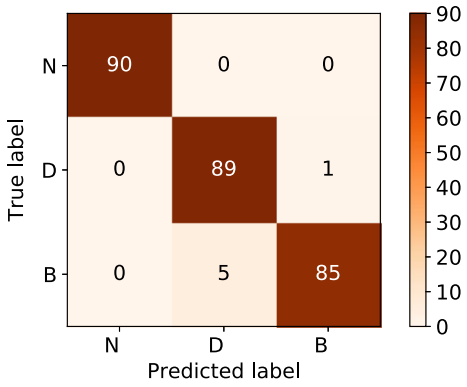


FIGURE 5 The confusion matrix for test results (illustration generated using adapted code from Reference [11])

observed that the accuracy of Holdout 1 significantly differs from the other iterations. Furthermore, errors are observed in both the desiccated and brown states, but no consistent trend is observed for all cases. Because of these deviations in homogeneity of these preliminary results, samples from a higher number of stomachs are suggested to further confirm generalization of the model on new subjects.

6 | TEST RESULTS

The pretrained model was then tested on a new, unseen stomach to evaluate its generalization performance when used on a new subject. The results are summarized in Table 3, suggesting good performance generalization when the model is used on a new subject. These are also in accordance with the intersubject cross-validation evaluation, resembling most likely the performance of the best case scenario, although in this case, the model is trained with samples from four stomachs.

The confusion matrix observed in Figure 5 illustrates the test results in detail, where a total of six misclassification errors out of the total 270 samples are observed. Five of the errors correspond to the *brown* ablation state, one to the *dessicated* state and none to the *normal* state. These results confirm the behavior observed on the evaluation tests, where the performance on the normal state is nearly perfect, and a few errors are observed in both the desiccated and brown states.

7 | CONCLUSIONS AND OUTLOOK

The presented methodology to monitor GMD by analyzing the diffuse reflected spectra from the tissue under treatment shows promising results when tested in porcine tissue. The implemented model based on LDA and used to classify among three ablation states exhibited good and stable generalization performance on new subjects, with 97.8% accuracy in the test results, and an average accuracy of 96.9% in the intersubject cross validation.

The presented methodology and results could serve as a base for the design of further animal trials, leading ultimately to human trials. To further confirm the generalization of this methodology, it is essential to obtain samples from a higher number of stomachs. This would allow a more complete representation of the true underlying data distribution of the stomach population, considering meaningful types of variation in the test population, such as age, race, gender, size and among others. It would also allow exploring possible trade-offs between more complex classification algorithms that, with more data, may deliver better performance at the expense of higher computational time in both training and prediction and less interpretability.

Given that this pilot experiment was performed in ex vivo, thawed tissue, further in vivo tests would be required to assess the performance of this methodology under a clinical operation environment. Major factors to consider are the effect of blood perfusion, the presence of fluids on the superficial layers of the tissue and the possible contamination of the fiber due to aerosols emitted during APC.

CONFLICT OF INTEREST

No conflict of interest, financial or otherwise, is declared by the authors.

ORCID

Sergio Vilches https://orcid.org/0000-0001-6457-3084

REFERENCES

[1] J. Raiser, M. Zenker, *J. Phys. D Appl. Phys.* **2006**, *39*(16), 3520.
[2] V. Kumbhari, S. Lehmann, N. Schlichting, M. Heinrich, Y. Kullnick, U. Retschlag, M. Enderle, A. Dietrich, M. A. Khashab, A. N. Kalloo, A. Oberbach, *Gastrointest. Endosc.* **2018**, *88*(1), 175.
[3] Guido Van Rossum, Fred L. Drake, *Python 3 Reference Manual*, CreateSpace, Paramount, CA, **2009**.
[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.

[5] T. E. Oliphant, *Guide to NumPy*, 2nd ed., CreateSpace Independent Publishing Platform, USA **2015**.
[6] Travis Oliphant, *Comput Sci Eng*. **2007**, *9*, 10.
[7] Wes Mckinney, in Proceedings of the 9th Python in Science Conf. (Eds: Stéfan van der Walt and Jarrod Millman), **2010**, 51–56.
[8] J. D. Hunter, *Comput. Sci. Eng*. **2007**, *9*(3), 90.
[9] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification 2nd ed*., Wiley-Interscience, New Jersey, **2000**, chapter 3.8.2–3.8.3, pp. 117–124.
[10] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, **2006**, chapter 4.1.6, pp. 191–192.
[11] Scikit-learn, *Confusion Matrix — scikit-learn 0.20.3 documentation*, https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html. Accessed April 11, **2019**.

How to cite this article: Osorio JD, Vilches S, Zappe H. Diffuse reflectance spectroscopy as a monitoring tool for gastric mucosal devitalization treatments with argon plasma coagulation. *J. Biophotonics*. 2020;13:e201960125. https://doi.org/10.1002/jbio.201960125