EasyChair Preprint

№ 9097

Determination of Optical Properties of Skin Tissues Using Spatial Domain Frequency Imaging and Random Forests

Bruno Garbim Silva, Murillo Rodrigues Gonçalves, Guilherme Henrique Sousa Alves, Adamo Ferreira Gomes Monte and Diego Merigue Cunha

October 24, 2022

Determination of Optical Properties of Skin Tissues using Spatial Domain Frequency Imaging and Random Forests

B. G. Silva[1], M. R Gonçalves[1], G. H. S. Alves[2], Á. F. G. Monte[2] and D. M. Cunha[2]

[1]School of Electrical Engineering, Federal University of Uberlândia, Uberlândia, Brasil
[2]Physics Institute, Federal University of Uberlândia, Uberlândia, Brasil

Abstract— **The evaluation of optical properties of biological tissues has been pointed as an important tool for detection and diagnosis of tissue alterations. The Spatial Frequency Domain Imaging (SFDI) is a fast and non-invasive technique that provides quantitative information about light absorption and scattering properties in tissues from measurements of light diffuse reflectance. A fundamental step in this imaging technique is the proper correlation between the measured values of diffuse reflectance of light by the tissue, R_d, at different spatial frequencies and the corresponding pair of absorption and reduced scattering coefficients μ_a and μ_s', respectively. In this work, the machine learning technique of Random Forests was applied to provide a regression model that efficiently computes μ_a and μ_s' from R_d values. The database employed consisted of values of R_d at different spatial frequencies for different combinations of μ_a and μ_s', obtained from Monte Carlo simulations. The database was split into training and testing groups, and a 3% Gaussian random noise was applied to the test group. Results showed that the correlation coefficient R^2 between predicted and expected values from the test group were 0.96 and 0.97, for μ_a and μ_s', respectively. The relative average errors for each coefficient were, respectively, 1% and 0.004%, with standard deviations of 11% and 7%. These results point to the good accuracy and precision of the models in predicting values of absorption and scattering coefficients. The developed models were applied to an *in vivo* study, where values of R_d from the dorsal region of the hand of a volunteer were obtained with SFDI equipment using light wavelength of 650 nm. The obtained images of μ_a and μ_s' showed enhanced contrast of blood vessels, pointing to the potential of the technique to identify vascular alterations that could be related to skin cancer.**

Keywords: **Spatial frequency domain imaging, optical properties, machine learning, random forests.**

I. Introduction

Nowadays, cancer is one of major concerns in public health all over the world. Recent projections estimate a worldwide incidence of 18 million new cases per year [1, 2]. In Brazil, for each year of the triennium 2020–2022, an occurrence of 625 thousand new cases of cancer is estimated. Of this total, non-melanoma skin cancer is the most incident, corresponding to an estimated total of 177,000 new cases per year [2]. Accurate and early diagnosis of the disease is essential to define the course of treatment, as well as its effectiveness [3]. Diagnostic techniques are of fundamental importance, for example, for the appropriate characterization of actinic keratosis-like lesions in cases of non-melanoma skin cancer, thus contributing to the reduction of mortality from the disease [4].

The Spatial Frequency Domain Imaging (SFDI) technique has emerged as a promising technique for the identification of morphological and functional changes in skin tissues in early stages. This imaging technique allows the evaluation of the properties of absorption and scattering of light by the tissues in a fast and non-invasive way [5]. Image acquisition in SFDI is performed by illuminating the tissue with an incident light beam of spatially modulated intensity at different spatial frequencies, and capturing the reflected beam with a camera. Then, the different obtained images are processed, thus reconstituting the optical properties of the tissues, which are quantitatively measured by the absorption and reduced scattering coefficients μ_a and μ_s', respectively. The evaluation of these parameters at different light wavelengths can enhance visualization of different skin structures, making eventual alterations in the tissue evident [5].

The proper correlation of light diffusion properties of a tissue with its optical properties is one of the main steps in SFDI imaging processing. Although pre-compiled look-up tables can be employed to this purpose [6, 7], it has been pointed out that machine learning techniques could accomplish this task more efficiently [8,9]. In particular, the use of different Artificial Neural Network models has been reported in the literature [10-14]. Recently, Panigraphi et al. [9] showed that the random forest method for regression can also provide optical properties from diffuse reflectance with high accuracy. In particular, the authors considered diffuse reflectance at two spatial frequencies for the determination of values of μ_a and μ_s' [9], and the performance of the model

was evaluated for noiseless test data. Alternatively, Zhao et al [15] recently showed that using more than two spatial frequencies can improve the performance of the machine learning models [15] in determining tissue chromophore concentration from noisy diffuse reflectance data.

In this work, we investigate the use of random forests for pattern recognition in the SFDI technique, to determine optical properties of tissues from their light reflection properties. The developed model considered values of diffuse reflectance at three spatial frequencies as input data, and the performance of the model was evaluated considering noisy input values.

II. Materials and methods

A. Experimental set-up for image acquisition in SFDI

Figure 1 shows the schematic representation of the experimental set-up employed for image acquisition in SFDI [5]. It comprises a digital projector as light source (model Vivitek® D555WH) a ThorLabs® filter wheel and CCD camera (DCC3240C). A white light beam is emitted from the source, reaching the sample and being reflected by the latter. The spatial distribution of the emitted beam is modulated according to a sinusoidal pattern, at different spatial frequencies f (0.05; 0.2 and 0.4 mm^{-1}) [5]. The reflected beam by the sample is filtered to a given wavelength by the filter, and it is then detected by the camera, which stores the collected signal.

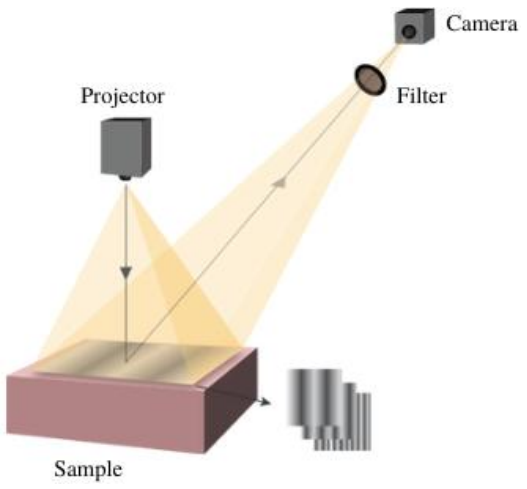

Fig. 1. Experimental set-up employed for image acquisition in SFDI.

The collected images by the camera must be processed since they carry dependence not only on the sample composition, but also on the emitted beam and camera properties [6]. To remove these dependencies on the imaging system, the obtained images were demodulated and calibrated with a phantom object, following the methodology proposed by Cuccia et al. [6]. The obtained signal therefore represents the diffuse reflectance, R_d, at the different spatial frequencies employed, which is the physical quantity that expresses the dependence of the reflected light on the absorption and scattering properties of the sample [5,6]. Determination of the sample absorption and reduced scattering coefficients μ_a and μ_s' from the obtained R_d values was performed using machine learning methods with random forests, as described in sections B and C.

B. Generating the database of diffuse reflectance values

Values of R_d as a function of the pair of coefficients μ_a and μ_s' were obtained using Monte Carlo (MC) simulations for the transport of light photons through matter. A MC code was developed to simulate the processes of absorption and scattering of light by a uniform material, according to the methodology proposed by Prahl et al. [16]. A material medium was characterized by its coefficients μ_a and μ_s' , and also by its refraction index and anisotropy factor. A pencil beam impinged perpendicularly on the material surface, and the spatial distribution of the backscattered photons was obtained as a point spread function. The values of R_d at different spatial frequencies were then obtained by applying a Fourier Transform on this spread function [6]. Different material compositions were simulated, by varying the values of μ_a between 0.001 and 1.8 mm^{-1}, and μ_s' between 0.3 and 6.3 mm^{-1} [6]. The values of the refraction index and anisotropy factor were fixed at 1.4 and 0.71, respectively [6,17].

Figure 2 illustrates the contour plot of R_d values obtained from the MC simulations, at a spatial frequency of 0.2 mm^{-1}, for different combinations of μ_a and μ_s'.

C. Random forest regression models

The random forest is an assembly method, based on the concept of decision trees, which can be employed for classification or regression problems [18]. It is a supervised machine learning technique, which combines a large number of decision trees for prediction purposes. For each tree, features and samples are randomly selected from the training data, with replacement, to determine the splitting of each node, in a process called bagging (bootstrap aggregation). For regression problems, the predicted output of the model is computed as the average of the value provided by each tree

[19]. The main hyperparameters of a random forest include the number of trees in the forest, the maximum depth of the trees, the minimum number of samples at leaf and minimum number of samples at split. A more detailed description about Random Forests, including mathematical aspects, can be found in the work of Breiman [19].

In this work, the Random Forest technique was employed for regression (RFR), using Python language with the scikit-learn library [20]. For each sample, input data consisted of the R_d values at three different spatial frequencies *f (*0.05; 0.2 e 0.4 mm^{-1}), obtained from the MC simulations, while the output data were the corresponding μ_a and μ_s' coefficients, as illustrated in figure 3.

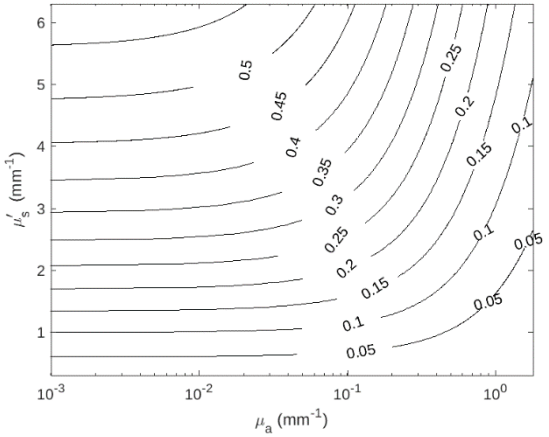

Fig 2. Contour plot of R_d values for different pair values of μ_a and μ_s' at spatial frequency of 0.2 mm^{-1}.

The dataset consisted of 108.360 samples. The samples were divided into training and testing data, at the proportions of 67% and 33%, respectively. Besides, for the testing data, a Gaussian noise of 3% was added, to simulate an experimental uncertainty, and also to evaluate the generalization capacity of the trained models [15]. Optimization of the model hyperparameters was performed employing K-fold cross-validation, with 5 subsets and 50 iterations, corresponding to 250 training loops. For the μ_a coefficient, the following optimum hyperparameters were obtained: the number of trees in the forest was 200, the maximum depth of the trees was set to 21, the minimum number of samples at leaf nodes was 4 and the minimum number of samples at split was 2. For the μ_s' coefficient, these values were, respectively, 250, 11, 4 and 2. In both cases, bootstrap sampling was used to train the decision trees.

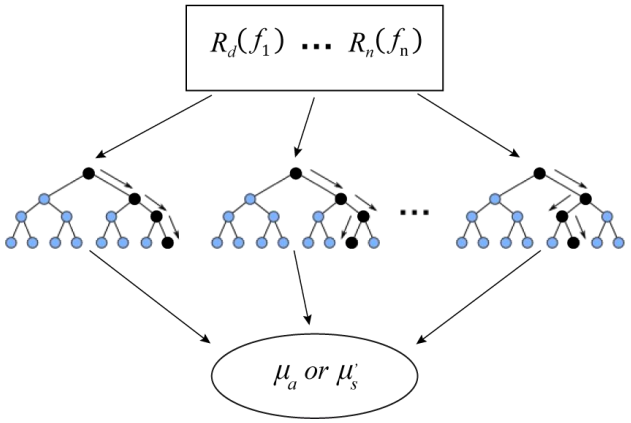

Fig. 3. Schematic representation of the RFR models for the determination of μ_a or μ_s' coefficients.

D. In vivo measurements

The experimental set-up for image acquisition, described in section A, and the RFR machine learning models, described in sections B and C, were applied to an *in vivo* measurement of the absorption and scattering coefficients of the dorsal region of the hand of a volunteer subject (figure 4). The rectangular region of interest (ROI) shown in figure indicates the area imaged. Values of R_d at the different spatial frequencies investigated were obtained for the ROI shown in figure 4, and the trained RFR model was then employed to compute the corresponding values of of μ_a and μ_s' at each pixel.

The present study in human beings was approved by the Research Ethics Committee of the Federal University of Uberlândia (process no. 85363417.9.0000.5152).

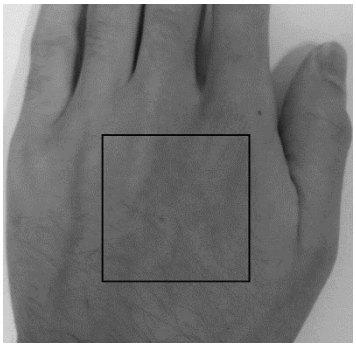

Fig. 4. Photography of the dorsal region of the volunteer's hand. The rectangular ROI indicates the area analyzed.

III. Results

The results of the application of the trained RFR on the testing data for determination of μ_a and μ_s' are shown in figure 5. The scatter plots represent the predicted values of μ_a and μ_s' *versus* the corresponding expected values.

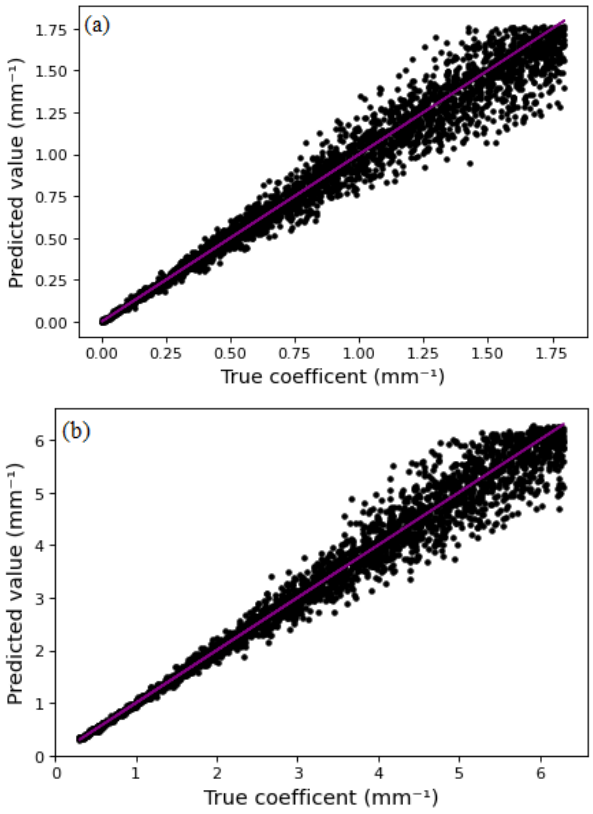

Fig. 5. Values of (a) μ_a and (b) μ_s' predicted by RFR models, as a function of the true values. Solid line indicates the expected values.

Figures 6a and 6b show the histograms of the distribution of relative percent errors for the predicted values of μ_a and μ_s', respectively.

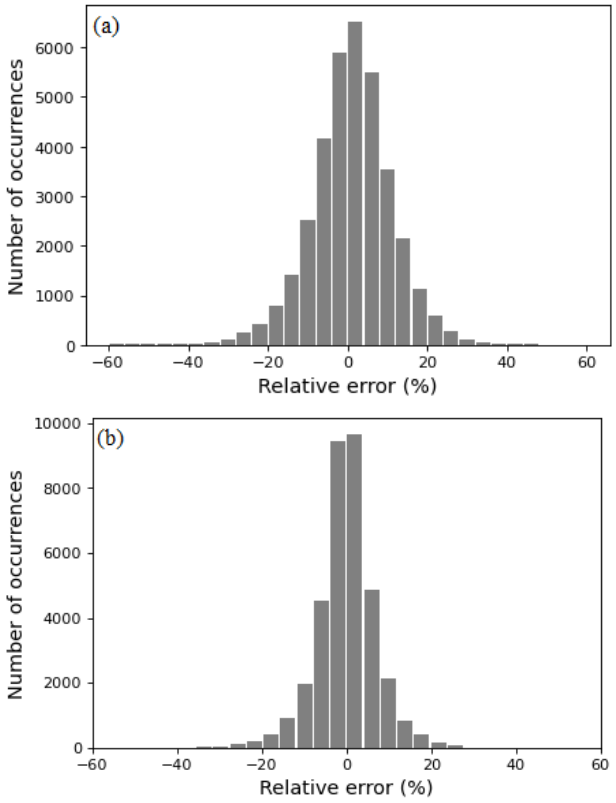

Fig. 6. Distribution of relative percent errors for the predicted values of (a) μ_a and (b) μ_s', respectively.

Figure 7 shows the results for the *in vivo* evaluation of μ_a and μ_s' for the back of the hand, for light wavelength of 650 nm.

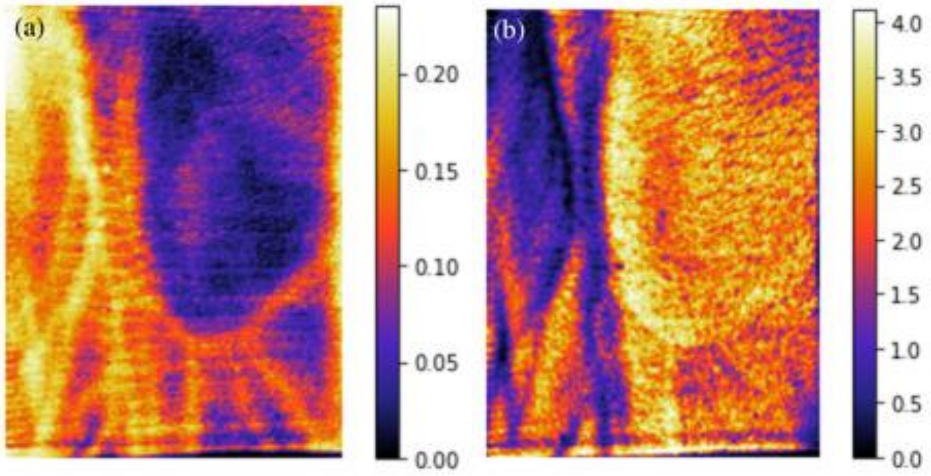

Fig. 7. Images of (a) μ_a e (b) μ_s' for the ROI analized. Values of both coefficients are given in units of mm^{-1}.

IV. Discussion

Evaluation of the performance of the RFR prediction models for μ_a and μ_s' values was performed through the determination of the R^2 correlation coefficient for the linear regression between expected and predicted values. A high degree of correlation between these quantities was observed, with values of R^2 of 0.96 e 0.97 for μ_a and μ_s', respectively.

A Levene's test was applied on the predicted values of μ_a and μ_s' showed in figure 5, to test for homogeneity of variances. For each coefficient, predicted values were divided in three subgroups, each one corresponding to a different interval within the whole range of values. The variances of these subgroups were then compared among each other. Results showed that, for each coefficient, there were no statistically significant differences between variances among subgroups, considering the significance level of 0.05. This result indicates that the increase in the errors observed in figure 5a and 5b is actually a graph scale effect, due to differences between the orders of magnitude of the values of each coefficient, and the performance of the RFR model is independent on the coefficient values.

From figure 6, it can be observed that the error distributions show symmetry around the null error value, indicating that the obtained models provided unbiased results. The average percent errors for these distributions were 1% and 0.004%, for μ_a and μ_s', respectively, with standard deviations of 11% and 7%. These results indicate that, even in the presence of noisy input data, the models show good accuracy and precision in the determination of μ_a and μ_s' .

Compared to the literature, values of average percent errors obtained in this work were higher than those reported by Panigraphi et al. [9]. Nevertheless, it should be pointed out that the authors did not considered noise in their input test data. By neglecting noise, our model provides performance comparable to those reported by the authors. Results were also comparable to those reported by Song et al. [14], who considered a deep neural network model for the determination of optical properties, for test data with 2% Gaussian noise.

Figure 7 shows that the developed models, combined with the SFDI system, are capable of providing values of μ_a and μ_s' for *in vivo* measurements. In particular, at a light wavelength of 650 nm, figure 7a shows that values of the absorption coefficient highlight blood vessels. This effect occurs due to the higher contribution of oxy- and deoxyhemoglobin to the absorption coefficients at this wavelength [21], and it could be useful to identify vascular changes associated with melanocytic or nonmelanocytic skin cancers [22].

V. Conclusions

In this work, the machine learning method of Random Forests for regression (RFR) was applied to the determination of light absorption and scattering properties of tissues from values of light diffuse reflectance in the SFDI technique. The developed models showed good performance in determining μ_a and μ_s' coefficients from R_d values at different spatial frequencies. The average percent errors were 1% and 0.004% for μ_a e μ_s', respectively, with standard

deviations of 11% and 7%. The RFR models were applied to an *in vivo* measurement for the determination of light absorption and scattering coefficients of the back of a hand. In particular, values of $\mu_a$obtained for light wavelength of 650 nm provided enhanced contrast of blood vessels. These results point to the potential of this imaging technique for detection and identification of vascular changes in skin tissues, which could be related to different types of skin tumors. Future works should be conducted, e.g. using tissue equivalent materials with known optical coefficients, in order to compare the estimated coefficients with true tissue values, and thus evaluating the accuracy of the technique for different skin tissue compositions.

ACKNOWLEDGMENT

The authors acknowledge to the Brazilian agencies Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for their financial support.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I et al. (2018) Global cancer statistics: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394-424 DOI 10.3322/caac.21492
2. INCA (2021) at https://www.inca.gov.br/numeros-de-cancer
3. Travers J B, Poon C, Rohrbach D J, et al. (2017) Noninvasive mesoscopic imaging of actinic skin damage using spatial frequency domain imaging. Biomed Opt Express 8:3045–3052 DOI 10.1364/BOE.8.003045
4. Reinehr C P H, Bakos R M. (2019) Actinic keratoses: review of clinical, dermoscopic and therapeutic aspects. An Bras Dermatol 94:637-57 DOI 10.1016/j.abd.2019.10.004
5. Cuccia D J. (2012). Spatial Frequency Domain Imaging (SFDI): a technology overview and validation of an LED-based clinic-friendly device. Proc. SPIE 8254, Emerging Digital Micromirror Device Based Systems and Applications IV, 825405 DOI 10.1117/12.908860
6. Cuccia D J, Bevilacqua F, Durkin A J et al. (2009) Quantitation and mapping of tissue optical properties using modulated imaging. J. Biomed. Opt. 14:5-32 DOI 10.1023/A:1010933404324
7. Hennessy R J, Lim S L, Markey M K et al. (2013) Monte Carlo lookup table-based inverse model for extracting optical properties from tissue-simulating phantoms using diffuse reflectance spectroscopy. J. Biomed. Opt. 18(3): 037003 DOI 10.1117/1.JBO.18.3.037003
8. Jacques S L. (2013) Optical properties of biological tissues: a review. Phys Med Biol 58(11):R37-61 DOI 10.1088/0031-9155/58/11/R37
9. Panigrahi S, Gioux S. (2018) Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging. J. Biomed. Opt. 24:1-6 DOI 10.1117/1.JBO.24.7.071606
10. Yudovsky D, Nguyen J Q M, Durkin A J. (2012) In vivo spatial frequency domain spectroscopy of two layer media. J. Biomed. Opt. 17(10):107006 DOI: 10.1117/1.JBO.17.10.107006
11. Wang Q, Le D, Roman J R, Pfefer J. (2012) Broadband ultraviolet-visible optical property measurement in layered turbid media. Biomed Opt Express 3:1226-1240 DOI 10.1364/BOE.3.001226
12. Jäger M, Foschum F, Kienle A. (2013) Application of multiple artificial neural networks for the determination of the optical properties of turbid media. J. Biomed. Opt. 18(5): 057005 DOI 10.1117/1.JBO.18.5.057005
13. Tsui S Y, Wang C Y, Huang T H et al. (2018) Modelling spatially-resolved diffuse reflectance spectra of a multi-layered skin model by artificial neural networks trained with Monte Carlo simulations. Biomed Opt Express 9:1531-1544 DOI 10.1364/BOE.9.001531
14. Song B, Jia W, Zhao Y et al. (2022) Ultracompact Deep Neural Network for Ultrafast Optical Property Extraction in Spatial Frequency Domain Imaging (SFDI). Photonics 9:327 DOI 10.3390/photonics9050327
15. Zhao Y, Deng Y, Yue S et al. (2021) Direct mapping from diffuse reflectance to chromophore concentrations in multi-fx spatial frequency domain imaging (SFDI) with a deep residual network (DRN). Biomed Opt Express 12(1):433-443 DOI 10.1364/BOE.409654
16. Prahl S A, Keijzer M, Jacques S L et al. (1989) Monte Carlo Model of Light Propagation in Tissue. SPIE Institute Series 5:102-111
17. Yudovsky D, Durkin A J. (2011) Spatial frequency domain spectroscopy of two layer media. J Biomed Opt 16(10):107005 DOI 10.1117/1.3640814
18. Müller A C, Guido S. (2016) Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Sebastopol, CA.
19. Breiman, L. (2001) Random Forests. Machine Learning 45:5–32 DOI 10.1023/A:1010933404324
20. Pedregosa F, Varoquaux G, Gramfort A et al. (2011) Scikit-learn: Machine Learning in Python J Mach Learn Res 12:2825-2830.
21. Mazhar A, Dell S, Cuccia D J et al. (2010) Wavelength optimization for rapid chromophore mapping using spatial frequency domain imaging. J Biomed Opt 15(6):061716 DOI 10.1117/1.3523373
22. Argenziano G, Zalaudek I, Corona R et al. (2004) Vascular structures in skin tumors: a dermoscopy study. Arch Dermatol 140(12):1485-9 DOI 10.1001/archderm.140.12.1485

Enter the information of the corresponding author:

Author: Diego Merigue da Cunha
Institute: Instituto de Física
Street: Av. João Naves de Ávila – 2121, Campus Santa Mônica
City: Uberlândia
Country: Brazil
Email: dmerigue@ufu.br