Spectral Band Subset Selection for Discrimination of Healthy Skin and Cutaneous Leishmanial Ulcers

Ricardo Franco-Ceballos[1](✉), Maria C. Torres-Madronero[1](✉),
July Galeano-Zea[2], Javier Murillo[3], Artur Zarzycki[2,4],
Johnson Garzon[5], and Sara M. Robledo[3]

[1] Research group on Automatic, Electronic and Computational Science,
Smart Machines and Pattern Recognition Laboratory,
Instituto Tecnologico Metropolitano, Medellin, Colombia
rfrancoce@gmail.com, mariatorres@itm.edu.co
[2] Research group on Advance Materials and Energy MatyEr,
Instituto Tecnologico Metropolitano, Medellin, Colombia
[3] Program for the Study and Control of Tropical Diseases-PECET-School
of Medicine, University of Antioquia, Medellín, Colombia
[4] Research group on Automatic, Electronic and Computational Science,
Robotics and Control System Laboratory, Instituto Tecnologico Metropolitano,
Medellin, Colombia
[5] Grupo de Optica y Espectroscopia, Universidad Pontificia Bolivariana,
Medellin, Colombia

Abstract. Leishmaniasis is a parasitic disease, transmitted by the bite of an insect that has previously fed on an infected host. One of its clinical forms is Cutaneous Leishmaniasis - CL and due to its increasing incidence, it is necessary to create effective and easy-use diagnostic methods. In this paper, we assess two unsupervised band-selection algorithms that allow the dimensional reduction of hyperspectral data taken from CL ulcers, maintaining a high classification accuracy. This is an important task for the development of an non-invasive system based on multispectral imaging, that support the diagnosis and treatment follow-up of cutaneous ulcer caused by Leishmaniasis. Spectral data was obtained in golden hamsters subjected to varying conditions of infection. Two algorithms, one based on similarity and the other based on singular values decomposition, are implemented using MATLAB functions and are applied to the spectral data. The selected subsets of bands are used to classify the spectra into healthy skin, border and ulcer centers using support vector machines - SVM and neural networks - NN. The obtained results are represented in precision tables and allow to observe that both methods achieve an appropriate dimensional reduction of multispectral data without losing key information for their subsequent classification.

At the end, we show that it is possible to obtain a subset of spectral bands to discriminate between healthy skin and cutaneous ulcers caused by Leishmaniasis.

1   Introduction

Leishmaniasis is a disease caused by protozoan parasites of the genus Leishmania, transmitted by the bite of an infected insect. There are two clinical presentations: Visceral Leishmaniasis (VL) and Cutaneous Leishmaniasis (CL). VL is the most serious and can be fatal. CL does not cause death, but it represents a large burden due to social stigma. Also, CL is related with psychological effects and decreasing of productivity of patients. Since the incidence of this disease is growing, it is necessary to develop new techniques for its diagnosis [1,2,10,11].

Some studies propose the use of spectral data for the diagnosis of skin diseases. Spectral data is refereed to spectral signatures obtained by spectrophotometer, as well as, multispectral or hyperspectral imagery collected by cameras. Spectral system measures the reflected and emitted energy by a surface along the electromagnetic spectrum. Spectral data from skin can provide accurate information to develop non-invasive techniques for the diagnosis of skin diseases. For example, Vyas et al. [14] proposed a non-invasive estimation of skin thickness from hyperspectral imaging; Attia et al. [4] developed a non-invasive real-time characterization of non-melanoma skin cancer; and [6] reviewed several non-invasive techniques for diagnosis of skin cancer, including some based on spectrophotometry data. Despite the advance in this field, more methodologies and techniques are necessary in order to characterize skin ulcers in their different phases of formation and treatment follow-up.

This paper presents results from a project that seeks to develop a portable non-invasive system based on multispectral imaging for the diagnosis and monitoring of skin ulcer treatments caused by Leishmaniasis. For the development of a new multispectectral system, we need to understand the spectral signature of both healthy skin and CL ulcers. An animal model for CL using golden hamsters was employed to build an spectral library. These include several spectra with nearly 2000 bands between 400 nm to 800 nm from healthy skin and ulcers in different phases. In this paper, we presents the evaluation of two unsupervised band selection algorithms, the first based on similarity [7] and the second based on singular value decomposition (SVD) [3]. These algorithms select the most relevant bands for the discrimination of healthy skin and leishmanial ulcers. The comparison of the unsupervised band selection algorithms is performed by using two classifiers: neural network (NN), and support vector machine (SVM).

2   Spectral Band Subset Selection Algorithms

In the literature, several algorithms for band subset selection - BSS can be found. These methods are known as dimensional reduction approaches, which select a set of bands according to a separability criteria. The difference between BSS algorithms with other dimensional reduction approach, such as principal component analysis, is that BSS selects bands from the measured spectrum, allowing the characterization of the materials, and opening the possibility to build low-cost sensing system using the selected bands. For this work, we select two unsupervised BSS algorithms with low computational complexity: similarity-based band selection [7] and singular value decomposition - SVD based band subset selection [3].

2.1   Similarity-Based Band Selection

Du and Yang [7] proposed two unsupervised methods: Linear Prediction - LP and Orthogonal Subspace Projection - OSP, whose basic idea is to look for the most distinctive bands, but ensuring that the selected bands also are the most informative ones. For this paper, we use the LP algorithm, since both algorithms offer the same results, but LP is computationally more efficient by operating relatively smaller matrices. For both, LP and OSP, the hyperspectral data must go through a pre-processing to eliminate water absorption and low signal-to-noise ration bands [7]. Once these bands are removed, a noise whitening is applied. This whitening is easily achieved thanks to the self-decomposition of the covariance matrix, using the method presented in [12].

   The algorithm begins with the combination of the two best bands, and this combination increases consecutively until the desired number of bands is selected. The authors suggest a random selection of the first band and then, a projection of the additional bands in the orthogonal subspace of the first band, this to select the bands most dissimilar to each other. However, we chose a different selection method for the first band seeking to improve the performance of this algorithm. Since the LP algorithm seeks also for the most informative ones, we choose the band with the highest variance as the first one. Then, the next band is selected such that it is the most distant from the first one using the euclidean distance [7].

   The LP algorithm assumes two bands, B_1 and B_2, belonging to the subset φ, which contains the selected bands, with N pixels each one. To find the band most dissimilar to B_1 and B_2, these bands are used to estimate a third band B using Eq. 1.

$$B' = a_0 + a_1 \ B_1 + a_2 \ B_2 \tag{1}$$

where B' is the linear estimation of B using B_1 and B_2, and a_0, a_1 and a_2 are the parameters that minimize the error of the linear prediction: $e = \parallel B - B' \parallel$. The parameter vector will be $a = (a_0, a_1, a_2)$, which can be determined using the least squares solution shown in 2.

$$a = (X^T \ X)^{-1} X^T y \tag{2}$$

In [2], X is a matrix N x 3 where the first column is one, the second column includes the N pixels of B_1 and the third column includes the pixels of B_2, and y is a vector of N x 1 with the pixels from the band that is being compared. The band B with the minimum error e is the most closely to the band B', and then it is chosen as B_3. This process is iteratively repeated until reaching the desired number of bands. A seudo-code for this procedure is presented in the Algorithm 1.

Algorithm 1. Similarity Band Selection Pseudo-code

1: **function** SUBSETBANDSELECTION (hsi,bands)
2:     $bandSubset(1) \leftarrow$ find max std band of hsi
3:     $bandSubset(2) \leftarrow$ most distant band to $bandSubset(1)$ with euclidean norm
4:     $n \leftarrow$ number of pixels of each hsi bands
5:     **for** $i = 3$ to $bands$ **do**
6:         $X \leftarrow$ Matrix with a ones $nx1$ column vector and each band of
                $bandSubset$ as column vector
7:         **for** each B band in hsi **do**
8:             **if** B is in $bandSubset$ **then**
9:                 $error(index of B) \leftarrow -\infty$
10:            **else**
11:                $A \leftarrow (X'X)^{-1}Xb$
12:                $B' \leftarrow XA$
13:                $error(index of B) \leftarrow \|B - B'\|$
14:            **end if**
15:        **end for**
16:        $bandSubset(i) \leftarrow$ band-index with max $error$
17:    **end for**
18:    **return** $bandSubset$
19: **end function**

2.2   SVD-Based Band Subset Selection

Velez and Jiménez [3] proposed an unsupervised method based on the singular value descomposition - SVD. This method combines the SVD with the revealing range QR factorization and allows to obtain a subset of bands that retain the data meaning without a transformation [3]. The method used the strongly restricted projection of a matrix A (see Eq. 3).

$$A = P \begin{bmatrix} I_p \\ 0 \end{bmatrix}$$ (3)

where A is a n x p matrix with $p < n$ and $A^T A = I_P$, and P is a permutation matrix. To compute the permutation matrix, first it is calculate the covariance Σ_{data} for the hyperspectral data. Then, the QR factorization with pivoting is used to compute the matrix $V_1{}^T$ where V_1 is formed by the first p eigenvectors of Σ_{data}. The pivot matrix P that results from this factorization is the permutation

matrix for the Eq. 3. Finally, the first p elements of $\overline{x}$ are the selected bands [3]. A seudo-code for this procedure is presented in the Algorithm 2.

Algorithm 2. SVD-Based Band Subset Selection Pseudo-code

1: **function** SUBSETBANDSELECTION ($hsi, bands$)
2:     $\Sigma_{data} \leftarrow$ covariance of hsi
3:     $p \leftarrow$ desired length of band-subset given in $bands$
4:     $V_1 \leftarrow$ first p eigenvectors of Σ_{data}
5:     $Q, R, P \leftarrow$ QR Factorization wit Pivoting of V_1^T
6:     $\overline{x} \leftarrow P * hsi$
7:     $bandSubset \leftarrow$ first p elements of $\overline{x}$
8:     **return** $bandSubset$
9: **end function**

---

3   Spectral Classification

Classification is a process during which each sample is labeled as a class [8], by applied decision rules, either in the multispectral or spatial domain. Classification process can be done through supervised or unsupervised approaches. Supervised classification uses a prior information to learn the decisions rules. Instead, unsupervised approaches seek for patterns in the data using some similarity criterion. In this paper, we used two supervised classification methods: support vector machines - SVM and neural networks - NN. Both methods are selected for their high performance documented in the literature with spectral data.

3.1   Support Vector Machines - SVM

SVMs are a useful technique for data classification. The objective of using SVM for classification is to find a optimal decision hyperplane to separate unknown data in two or several classes. A kernel can be used to solve the problem for non-linear separable data. Most used kernels for hyperspectral data are polynomial and radial basis function kernel [9].

3.2   Neural Networks - NN

Neural networks are a learning paradigm based on the human brain. These networks are composed of individual units that process information through highly interconnected individual nodes. NN models are useful algorithms for cognitive tasks, such as classification [8]. In this document, an NN classification was implemented with a network formed by a hidden layer of five neurons (nodes).

4   Experimental Procedure

4.1   Data Set

Animal models are widely used to analyze new drugs and treatments. For CL studies, golden hamsters are recommended due to the similarity of their skin structure with human skin [5,13]. Diffuse reflectance spectral from healthy and CL ulcers were acquired using a spectrometer Ocean Optics HR4C3337. The acquired spectra were calibrated using white and black diffuse reflectance standards. A total of 39 golden hamsters, distributed in 18 females and 21 males, were used. Hamsters are subject to several conditions of infection and treatment. For this paper, we used only spectral signatures acquired before treatment. From the 39 golden hamsters, 27 were infected with Leishmaniasis Braziliensis (LB), while 4 were hamsters infected with Leishmaniasis Panamensis (LP), and 8 hamsters were in the control group (i.e. without CL).

Spectral signatures of each hamster's skin are obtained each fifteen days. The first measure is taken before the inoculation of CL, then two more measures are taken during the development of the ulcers. In each date, up to 12 spectra are measured for each area: healthy skin, border and ulcer center. This data collection allows an exhaustive analysis of the evolution of the disease, from the inoculation process followed by the analysis of ulcer development. This protocol had the approval from the Universidad de Antioquia animal ethics committee.

Figure 1 presents the average signatures from healthy skin, ulcer border and ulcer center between 400 nm and 800 nm. After 750 nm, the signature noise increased. We can also see that the spectral response from the ulcer center is lower that from healthy skin; but, the spectral signature from the ulcer border is very similar to healthy skin.
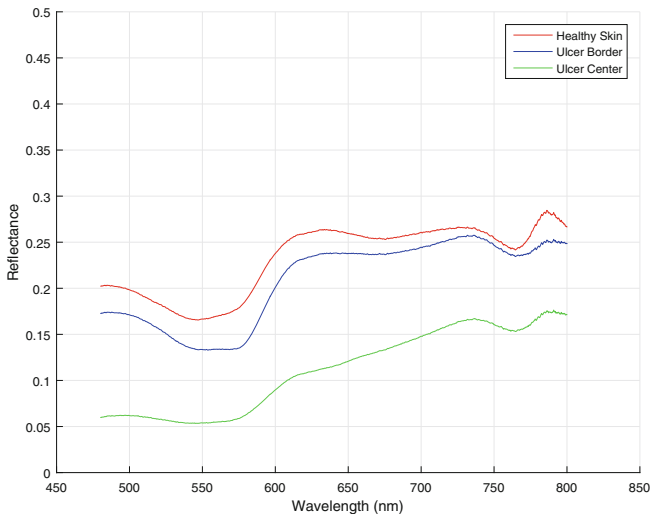

Fig. 1. Average spectral signatures of healthy skin, ulcer border and ulcer center from golden hamsters infected infected with Leishmaniasis

4.2   Experiments

For the evaluation of both BSS algorithms, we used spectral signatures of healthy skin, border and ulcer center captured from Golden hamsters. First, a mean filter with a sliding window of 3 points is applied to each of the captured signatures, in order to reduce noise. Since the bands from 750 nm present higher noise than lower bands, we defined two experiments to analyze the spectral signatures. The first experiment applied the BSS algorithms to spectral signatures between 480 nm to 750 nm, eliminating upper bands for reducing the noise. The second experiment takes all bands between 750 nm to 800 nm. For both BSS algorithms, we select 10 bands. This number is chosen since the selected bands will be used in the development a portable system, and commercial filter wheels for 10 filters are very common. Both experiments applied the two BSS methods: SVD and Similarity-Based band selection. Bands subsets are converted into its respective commercial filter, to evaluate a real configuration for a multispectral system.

The evaluation of the selected bands is performed using supervised classification. SVM and NN are used to evaluated the capability of the selected bands to improve the discrimination of healthy skin, border and ulcer center. The parameters of both classifiers are optimized to obtain the highest overall accuracy. For SVM, a radial basis function kernel is used. For NN, a configuration with a hidden layer of 5 neurons provided the best performance. For training, 30 samples are randomly selected for each class. Since border signatures are close to healthy skin signatures, as shown in Fig. 1, we first classify only healthy skin and ulcer center. Then, we performed the classification process using the three class. Each experiment is repeated 100 times to obtain the general classification accuracy.

5   Results

The selected bands from the BSS algorithms using the signatures between 480 nm to 750 nm are presented in Fig. 2. The spectral signature (blue signal) presented in Fig. 2 is the average of the all spectra used in the experiment. We can note that the selected bands by both algorithms are very close. Then, when we identify the corresponding commercial filters, many spectral bands become the same from both BSS approaches. Values of the commercial filters are presented in the table inside Fig. 2.

The selected bands from the BSS algorithms using the signatures between 480 nm to 800 nm are presented in Fig. 3. Values of the commercial filters also are presented in the table inside Fig. 3. Comparing these results with the first experiment, we note that two bands are selected between 750 nm to 800 nn for both algorithms. In these bands (785 nm and 800 nm) we can see a interesting behavior of healthy skin, border and ulcer center (see Fig. 1), that can be helpful for the discrimination process.

Once the band subsets are experimentally obtained, these are classified using SVM and NN. First, a two-class classification is performed, using only healthy skin and ulcer center signatures. A classification baseline is obtained by using
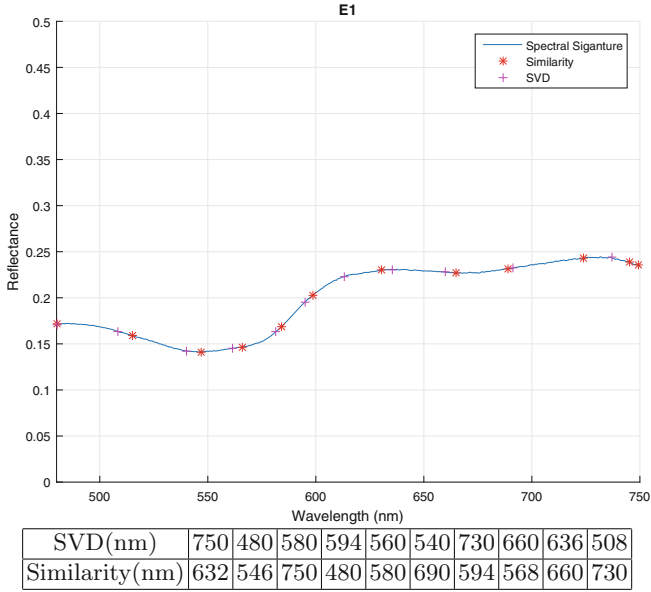
E1

SVD(nm)	750	480	580	594	560	540	730	660	636	508
Similarity(nm)	632	546	750	480	580	690	594	568	660	730

Fig. 2. Selected spectral bands for spectral signatures between 480 nm to 750 nm using SVD (∗) and similarity-based (+) Band Subset Selection. Table shows the equivalent commercial filters. (Color figure online)
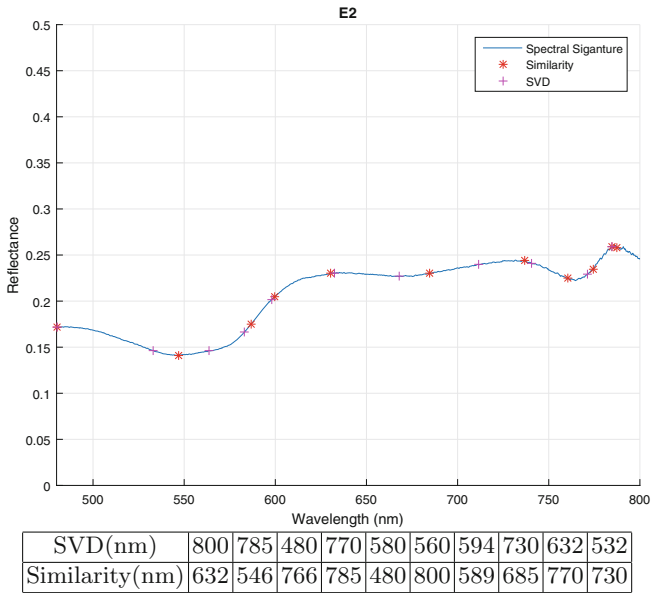

E2

SVD(nm)	800	785	480	770	580	560	594	730	632	532
Similarity(nm)	632	546	766	785	480	800	589	685	770	730

Fig. 3. Selected spectral bands for spectral signatures between 480 nm to 800 nm using SVD (∗) and similarity-based (+) Band Subset Selection. Table shows the equivalent commercial filters.

Table 1. Overall classification accuracy for two-class problem: healthy skin and ulcer center

	480 nm to 750 nm		480 nm to 800 nm	
	SVM	NN	SVM	NN
Similarity	95.89% ± 2.41	94.32% ± 3.80	95.25% ± 2.89	93.36% ± 5.47
SVD	95.74% ± 2.83	93.85% ± 4.81	95.94% ± 2.02	91.93% ± 5.98

all spectral bands (nearly 2000). For the two-class problem, we obtain an average accuracy of 44.66% (±30.43%) using SVM and 58.06% (±13.96%) by NN using all bands. Table 1 shows the overall classification accuracy for the two-class problem using the spectral band subsets. Using the selected bands from 480 nm to 750 nm, the best classification is obtained from the subset selected by similarity-based approach and using SVM classifier. This configuration obtained a overall accuracy of 95.89%. However, the result obtained using the band subset selected by the SVD approach is very similar (95.74%). The NN classifier obtained lower overall accuracies for both subset (similarity and SVD). Using the selected bands from 480 nm to 800 nm, the overall accuracies are very close to the first experiment. Also, best performance was obtained using SVM than NN.

For three-class problem, the baseline accuracy was so low as 26.63% (±20.84%) using SVM and 74.06% (±18.89%) using NN with all the spectral bands. Table 2 shows the overall classification accuracy for the three-class problem using the spectral band subsets. We can note that for the three-class problem, the overall classification accuracy decrease for all configuration in comparison with two-class results. The best performance in this case is obtained using the spectral signatures from 480 nm to 800 nm with the band subset selected by SVD approach and using NN (82.60%). Then, the two bands selected between 750 nm to 800 nm are relevants for the discrimiantion between border and healthy skin. This can also be noted in Fig. 1.

Table 2. Overall classification accuracy for three-class problem: healthy skin, ulcer center and ulcer border

	480 nm to 750 nm		480 nm to 800 nm	
	SVM	NN	SVM	NN
Similarity	77.58% ± 4.36	77.21% ± 14.25	75.13% ± 4.60	76.02% ± 14.03
SVD	77.15% ± 4.78	76.32% ± 12.60	75.49% ± 5.15	82.60% ± 4.81

Finally, Table 3 shows the confusion matrix for the best result from the three-class problem (band subset selected by SVD and NN classifier). This confusion matrix allows determining that the ulcer border zone is the most sensitive to

classification and tends to have a variability such that, depending on the location, it may have a reflectance like areas of healthy skin or ulcer center.

Table 3. Confusion matrix for the best result using the three classes: band subset selected SVD-based algorithm from bands between 480 nm to 800 nm and NN classifier

	Healthy Skin	Center	Border	Accuracy(%)
Healthy Skin	241	1	23	91.67%
Center	2	51	6	86.44%
Border	7	2	26	74.24%
Accuracy(%)	96.70%	94.44%	47.27%	

6   Conclusions

In this article, we presented the evaluation of two band-selection algorithms: the first based on similarity measures and the second based on SVD. These algorithms were applied to spectral data captured from cutaneous ulcers caused by leishmaniasis on golden hamsters. The results shows that both algorithms allows to obtain an appropriate dimensional reduction of spectral signatures without losing key information for their subsequent classification. From the spectral range analyzed, best results are obtained using 480 nm to 800 nm for the discrimination of healthy skin, border and ulcer center. Ulcer border area is highly sensitive and represents a challenge for the classification, as this area tends to be confused with ulcer center and healthy skin.

Since, the band subset selected allows a suitable discrimination of healthy skin and cutaneous ulcers caused by leishmaniasis, this can be used to develop an portable multispectal imaging system, that support the diagnosis and follow-up of treatment of CL. As future work, the selected bands can be evaluated using images and combining spectral-spatial methods, helping to improve the overall classification accuracies.

References

1. Alvar, J., et al.: Leishmaniasis worldwide and global estimates of its incidence. PLoS ONE **7**(5), e35671 (2012). https://doi.org/10.1371/journal.pone.0035671
2. Alvar, J., Yactayo, S., Bern, C.: Leishmaniasis and poverty. Trends Parasitol. **22**(12), 552–557 (2006). https://doi.org/10.1016/j.pt.2006.09.004
3. Arzuaga-Cruz, E., Jimenez-Rodriguez, L.O., Vélez-Reyes, M.: Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis. Proc. SPIE-Int. Soc. Opt. Eng. **5093**(September), 462–473 (2003). https://doi.org/10.1117/12.485942

4. Attia, A.B.E., et al.: Noninvasive real-time characterization of non-melanoma skin cancers with handheld optoacoustic probes. Photoacoustics **7**, 20–26 (2017). https://doi.org/10.1016/j.pacs.2017.05.003

5. Avci, P., et al.: Animal models of skin disease for drug discovery. Expert Opin. Drug Discov. **8**(3), 331–355 (2013). https://doi.org/10.1517/17460441.2013.761202

6. Calin, M.A., Parasca, S.V., Savastru, R., Calin, M.R., Dontu, S.: Optical techniques for the noninvasive diagnosis of skin cancer. J. Cancer Res. Clin. Oncol. **139**(7), 1083–1104 (2013). https://doi.org/10.1007/s00432-013-1423-3

7. Du, Q., Yang, H.: Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geosci. Remote Sens. Lett. **5**(4), 564–568 (2008). https://doi.org/10.1109/LGRS.2008.2000619

8. Gao, J.: Digital Analysis of Remotely Sensed Imagery. McGraw Hill Professional, New York (2009)

9. Gholami, R., Fakhari, N.: Support vector machine: principles, parameters, and applications. In: Handbook of Neural Computation (1st edn.). Elsevier Inc. (2017). https://doi.org/10.1016/B978-0-12-811318-9.00027-2

10. Hotez, P.J., Bottazzi, M.E., Franco-Paredes, C., Ault, S.K., Periago, M.R.: The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. PLoS Negl. Trop. Dis. **2**(9) (2008). https://doi.org/10.1371/journal.pntd.0000300

11. Hotez, P.J., Remme, J.H.F., Buss, P., Alleyne, G., Morel, C., Breman, J.G.: Combating tropical infectious diseases: report of the disease control priorities in developing countries project. Clin. Infect. Dis. **38**(6), 871–878 (2004). https://doi.org/10.1086/382077

12. Ren, H., Chen, H.T.: Background whitened target detection algorithm for hyperspectral imagery. J. Mar. Sci. Technol. (Taiwan) **25**(1), 15–22 (2017). https://doi.org/10.6119/JMST-016-0630-1

13. Robledo, S. M., et al.: Cutaneous Leishmaniasis in the dorsal skin of hamsters: a useful model for the screening of Antileishmanial Drugs. J. Vis. Exp. (62) (2012). https://doi.org/10.3791/3533

14. Vyas, S., Meyerle, J., Burlina, P.: Non-invasive estimation of skin thickness from hyperspectral imaging and validation using echography. Comput. Biol. Med. **57**, 173–181 (2015). https://doi.org/10.1016/j.compbiomed.2014.12.010