Wiley Online Library
Early View publication on www.wileyonlinelibrary.com
(issue and page numbers not yet assigned;
citable using Digital Object Identifier – DOI)

J. Biophotonics 1–6 (2016) / **DOI** 10.1002/jbio.201600169

Journal of BIOPHOTONICS

FULL ARTICLE

Raman spectroscopy and PCA-SVM as a non-invasive diagnostic tool to identify and classify qualitatively glycated hemoglobin levels *in vivo*

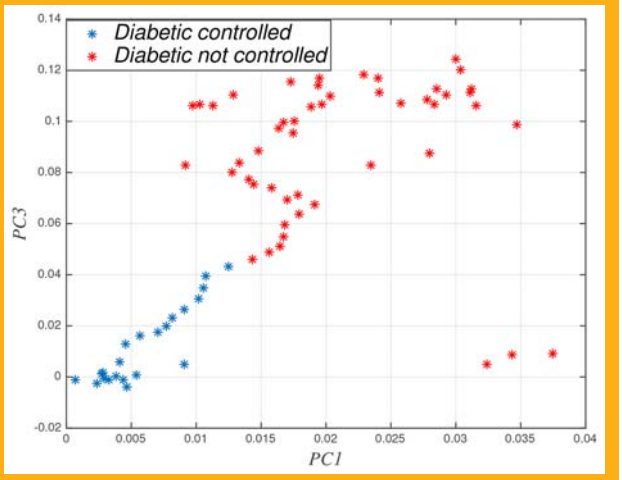
J. F. Villa-Manríquez[*,1], *J. Castro-Ramos*[1], *F. Gutiérrez-Delgado*[2], *M. A. Lopéz-Pacheco*[1], *and A. E. Villanueva-Luna*[3]

[1] Instituto Nacional de Astrofísica Óptica y Electrónica, apartado postal 51 y 216, Tonantzintla Puebla. CP 72000, México
[2] Centro de Estudios y Prevención del Cáncer, Bugambilias 30, Fraccionamiento la Rivera, Juchitan, Oaxaca, CP 70020, México
[3] Universidad Tecnológica de Campeche, Carretera Federal 180 S/N. San Antonio Cárdenas, Carmen, Campeche, CP 24100, México

Received 2 June 2016, revised 22 November 2016, accepted 24 November 2016
Published online 24 December 2016

Key words: medicine, spectroscopy, Biomedical application

In this study we identify and classify high and low levels of glycated hemoglobin (HbA1c) in healthy volunteers (HV) and diabetic patients (DP). Overall, 86 subjects were evaluated. The Raman spectrum was measured in three anatomical regions of the body: index fingertip, right ear lobe, and forehead. The measurements were performed to compare the difference between the HV and DP (22 well controlled diabetic patients (WCDP) (HbA1c <6.5%), and 49 not controlled diabetic patients (NCDP) (HbA1c ≥6.5%)). Multivariable methods such as principal components analysis (PCA) combined with support vector machine (SVM) were used to develop effective diagnostic algorithms for classification among these groups. The forehead of HV versus WCDP showed the highest sensitivity (100%) and specificity (100%). Sensitivity (100%) and specificity (60%), were highest in the forehead of WCDP, versus NCDP. In HV versus NCDP, the fingertip had the highest sensitivity (100%) and specificity (80%). The efficacy of the diagnostic algorithm by receiver operating characteristic (ROC) curve was confirmed. Overall, our study demonstrated that the combination of Raman spectroscopy and PCA-SVM are feasible non-invasive diagnostic tool in diabetes to classify qualitatively high and low levels of HbA1c *in vivo*.



1. Introduction

Diabetes is a health public issue worldwide. Globally in 2012, there were 350 million diabetic patients in the world and 1.5 million deaths. More than 80% of deaths from diabetes occurred in low and middle-income countries. The World Health Organization (WHO) projects that diabetes will be the 7th leading cause of death by 2030 [1]. Glycated hemoglobin (HbA1c) is recommended by WHO as a diagnostic

* Corresponding author: e-mail: jfvillam@gmail.com

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

test for diabetes. A HbA1c of 6.5% is recommended as the cut point for diagnosing diabetes but values less than 6.5% does not exclude diabetes diagnosed using glucose test [2]. Glycated hemoglobin has been increasingly accepted as the gold standard for diabetes monitoring, due to their half-lives of red blood cell is about 120 days, indicating that the levels of glycosylated hemoglobin reflect the average blood glucose over 2–3 months [3]. Therefore, HbA1c can be used to monitor long-term glucose control in diabetes, making it possible to realize early diagnosis of diabetes. Vigneshwaran et al. [4] showed that HbA1c is the better index of integrated glycemia (blood glucose concentration) *in vivo* and *in vitro* using auto-fluorescence. Enejder et al. [5] and Shao et al. [6] established that Raman spectroscopy can be used for noninvasive measurement of blood analyses, because fasting plasma glucose test requires that patients have fast at least 8 hours. Lin et al. [7] showed that Raman spectroscopy combined with multivariate analysis (PCA-LDA) was able to distinguish the blood Hb in vitro of diabetic patients from that of healthy subjects. Due to weak signal of the plasma glucose, researches have worked with other technique, like Kiran et al. [8] found a characteristic band in the SERRS (Surface Enhanced Resonance Raman Scattering) spectrum of HbA1c, which was not present in the SERRS spectrum of HbA. In this study, we applied Raman spectroscopy and PCA-SVM *in vivo* to identify and classify glycated hemoglobin levels in healthy population and diabetic patients.

2. Experimental

2.1 Clinical measurements

Blood samples were collected from 15 healthy volunteers and 71 diabetic patients. Among diabetic patients, 49 of them had HbA1c high levels (≥6.5%), and 22 had HbA1c low levels (<6.5%). Patients and healthy volunteers were asked to have at least eight hours without food or liquid intake (except water). HbA1c levels were measured by using dry chemical technique. CEPREC's Institutional Review Board (IRB) approved this protocol and patients signed an informed consent.

2.2 Raman spectroscopy measurements

Raman spectroscopy measurents were made with the Raman spectrometer *QE*65000 from Oceans Optics, it has a resolution of 0.14–7.7 nm FWHM (~6 cm^{-1}), and enables integration time from 8 ms to 15 min. We selected the spectral range that varies from 360 cm^{-1} to 1700 cm^{-1}. The experimental arrangement covers a laser as excitation source of 785 nm, by using a power of 50 mW and 15 s of integration time in the measurements. *InPhotonics* Raman probe RIP-RPB with two fibers, one of them as a collector and the other as excitation source, the core diameters were 200 µm and 105 µm respectively. This Raman probe has a lens and band-pass after the output excitation fiber and dichroic filter, which only allows to transmit wavelenght of laser and reflects scattering of sample, after the a lens with focal length of 7.5 mm, with is optimal working distance of the test probe. Near to working distance we measurement all spectra, at that distance the spot size was around 4.8 mm. The experimental setup scheme is shown in Figure 1. It shows a computer, laser, and Raman spectrometer at the time of index fingertip of the right hand measurements.

Raman measurements were performed on the index fingertip right hand, ear lobe, and forehead was analized without any skin preconditioning because any other substance in the sample would affect our Raman spectra. The Raman probe was place carefully near to working distance with the aid of a mechanical probe; in the sample to be analyzed care was taken to do not radiate hair, to avoid burns in the patients. We choose anatomical regions, which had not either mole, scratch or another injury to measure; owing to Raman spectroscopy detects information by the skin lesion. All measurements was taken carefully to do not exposure the laser in the eye when we take Raman spectrum in the forehead. The acquired spectra were processed in two stages. In the first one, we used the algorithm developed by A. Cao et al. [9] that is based in polynomial fitting to remove the fluorescence background. In the second one, we used the algorithm developed by Villanueva-Luna et al. [10] that is based in wavelet transform to reduce the shot noise, finally we used normalization at 1486 cm^{-1} to reduce the intensity differences. To develop our algorithm we use PCA. It is a multivariate statistical technique that transforms

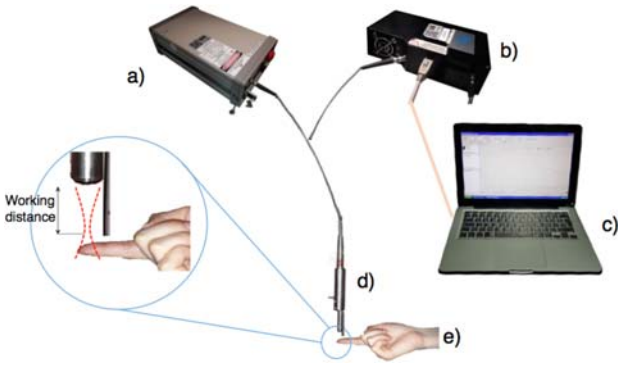


Figure 1 Experimental setup scheme. (**a**) laser of 785 nm, (**b**) spectrometer, (**c**) computer, (**d**) Raman probe, and (**e**) sample.

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

a set of original, correlated variables (spectra) to a new set of uncorrelated variables called principal components based on the maximum variance. PCA can be used to group spectra according to their similar variances since each principal component (PC) is orthogonal to each component. The first PC loading bring the most relevant spectral features presented in the dataset, whereas the PC scores bring the intensity of each PC loading in each spectrum of the dataset. Then one could use the scores to discriminate spectra from individuals of a population according to variation of the spectral characteristics, or use them to correlate the spectral information to a biomedical parameter of the sample [11]. In the classification part of our algorithm we use support vector machines. This is a method that used for binary classification. The basic idea is to make use of a nonlinear mapping function that transforms data in input space to data in feature space in such a way as to render a problem linearly separable. The SVM then automatically discovers the optimal separating hyperplane, however, since example data often not linearly separable, SVM's introduce the notion of a kernel induced feature space which casts the data into a higher dimensional space where the data is separable. There are different types of Kernels, the most simple Linear polynomial, Quadratic, high degree polynomial (order 3 or more), Gaussian Radial basic and Multilayer perceptron [11]. Sensitivity and specificity Eqs. (1–2) were calculated after PCA-SVM analysis of the entire spectral dataset acquired from the patients by using the Statistics Toolbox of MATLAB R2015b (Math Works, Natick, MA). Once sensitivity and specificity were calculated, the Receiver Operating Characteristic (ROC) curve was obtained. The curve is created by plotting the sensitivity versus 1 – *Specificity*, in other words the ROC curve is the probability distributions for detection and false alarm [11].

$$\text{Sensitivity} = \frac{\text{TruePositive}}{\text{FalseNegative} + \text{TruePositive}} \quad (1)$$

$$\text{Specificity} = \frac{\text{TrueNegative}}{\text{FalsePositive} + \text{TrueNegative}} \quad (2)$$

3. Results

3.1 Raman spectrum

After processing the data, we average patient groups spectra. In order to compare each of the Raman spectra profiles. There is a similar spectra profile between HV, WCDP and NCDP groups. After subtraction of them for WCDP and NCDP were found at 617 cm^{-1} and 1148 cm^{-1}, difference between HV and WCDP were found at 495, 964, and 1330 cm^{-1} while for HV with NCDP were found at 495, 529 and 626 cm^{-1}. Table 1 are shown assignments of the Raman bands according to De Gelder et al. [12], about substraction of it. According to the work of Barman et al. [13], we observed that the spectrum obtained have some peaks of HbA1c reported by us. Approximately these are in 1000, 1200, 1400 and 1600 cm^{-1}. These peaks are shown in the Figure 2.

Table 1 Raman bands and assignments.

Raman shift (cm^{-1})	Assignments
495	D-(+)-Manose
529	D-(+)-Fucose
617	D-Fructose-6-Phosphate
626	D-(–)-Fructose
964	D-(+)-Fucose
1148	D-(+)-Xylose
1330	α-D-Glucose, D-(+)-Threalosa

3.2 PCA-SVM analysis for Raman spectrum

There are many studies that have shown that PCA and Raman spectroscopy are a powerful technique in several diseases as a diagnostic tool [14–19]. PCA was used to find the main principal components that maximize the variance of the most important features of Raman spectra of diabetics and healthy patients. This gives us a good separation in which we can see a worthy classification between different levels of HbA1c. Obtained the principal components that separate better our groups, combined with SVM, changing the input space and maximizing the separation between the different levels of HbA1c and thereby increase the sensitivity and specificity of our analysis.

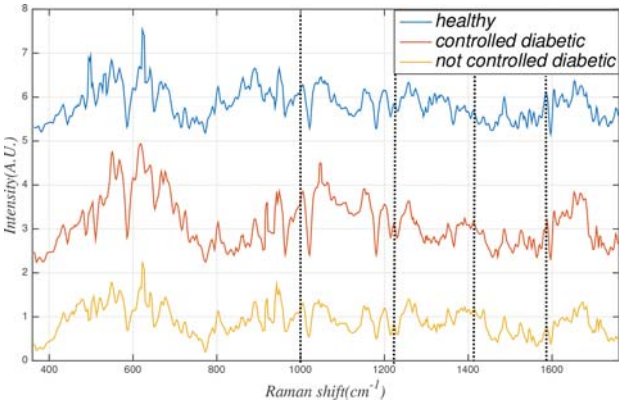


Figure 2 Peaks obtained of HbA1c, according to reported by Barman et al.

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

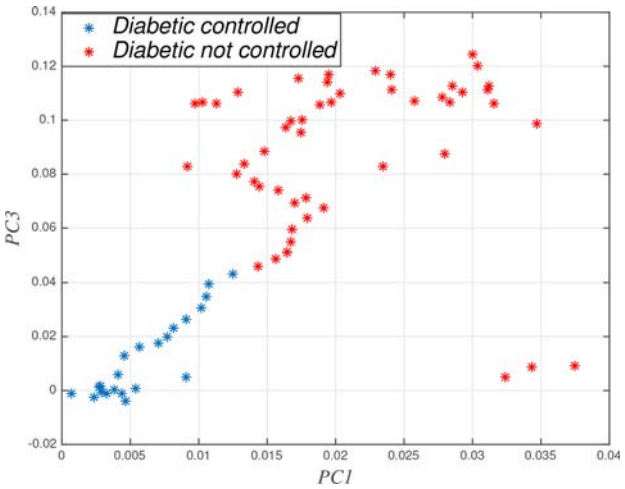


Figure 3 PCs scores 1 and 3 for the forehead as better region in controlled diabetic patients (blue dots) and not controlled diabetic patients (red dots).

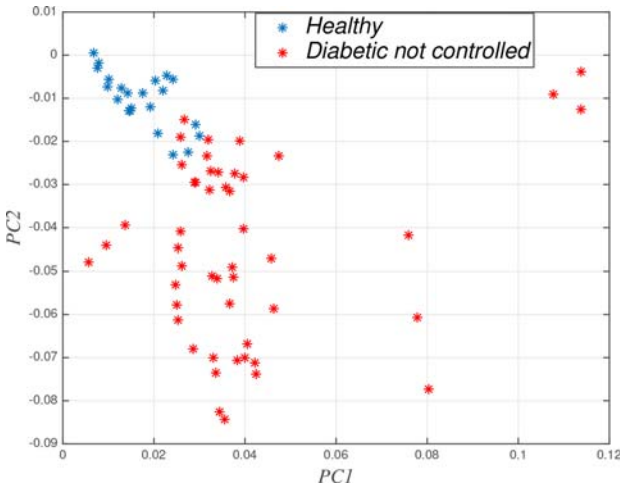


Figure 5 PCs scores 1 and 2 for the healthy patients (blue dots) and diabetic with not controlled diabetic patients (red dots) for the better region, fingertip.

The first step was trained the PCA with 10 HV, 10 WCDP, and 12 NCDP. Once the algorithm was validated, the remaining patients were evaluated. Figures 3 to 5 shown the two-dimensional scatter plots of the two PCs score between the WCDP and NCDP, HV and WCDP, and HV and NCDP respectively. These figures show differences between patients groups when PCs were compared. PC1 and PC3 for WCDP and NCDP (Figure 3), PC2 and PC3 for HV and WCDP (Figure 4) and PC1 and PC2 for HV and NCDP (Figure 5).

We choose these PCs due to spread of the score in the figures, these let us to classify fast and easy between different groups of patients. The best PCs are used as new data. After calculated it, we used the first three components and these were used as new input data of the healthy volunteers and diabetic patients and we trained with these values the SVM to classify between these groups. Sensitivity and specificity values obtained for HV and diabetic patients are shown in Table 2. Healthy volunteers's forehead showed the highest sensitivity (100%) and specificity (100%) compared to the forehead in WCDP using multilayer perceptron kernel function. Positive predictive value (PPV) and Negative predictive value (NPV) were calculated with 5 out of 15 HV and 10 out of 22 WCDP. PPV and NCPV

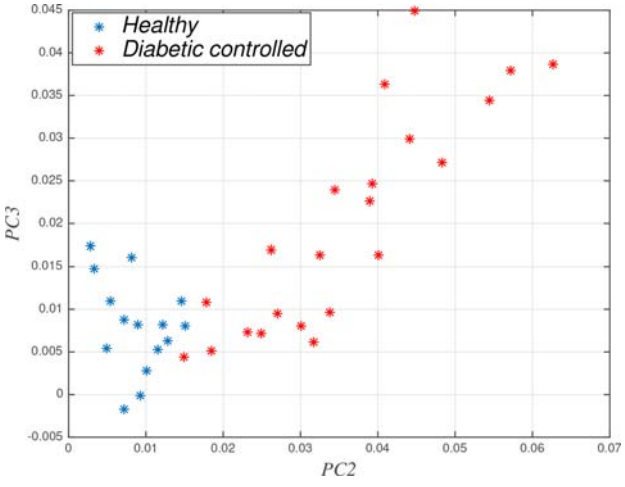


Figure 4 PCs scores 2 and 3 for the forehead as better region in healthy volunteers (blue dots) and controlled diabetic patients (red dots).

Table 2 Sensitivity and specificity by anatomical regions.

	Sensitivity (%)	Specificity (%)	PPV	NPV
Fingertip				
WCDP vs NCDP	94.59	60	0.89	0.75
HV vs WCDP	100	40	0.76	1
HV vs NCDP	100	80	0.97	1
Ear lobe				
WCDP vs NCDP	100	50	0.88	1
HV vs WCDP	100	40	0.76	1
HV vs NCDP	100	60	0.94	1
Forehead				
WCDP vs NCDP	100	60	0.9	1
HV vs WCDP	100	100	1	1
HV vs NCDP	100	40	0.92	1

* WCDP: well controlled diabetic patients, NCDP: Not controlled diabetic patients, HV: Healthy volunteers.
* These values were calculated using 37 NCDP, 10 WCDP, and 5 HV as a test.
** PPV: Positive predictive value, NPV: Negative predictive value.

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

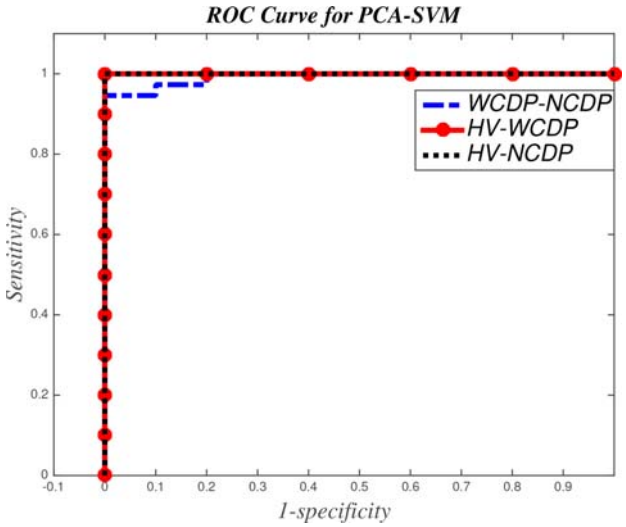


Figure 6 ROC curve of the discrimination result using the PCA-SVM based spectral classification of forehead in WCDP and NCDP, and patients HV and WCDP, and fingertip in patients HV and NCDP. The area under the ROC curve is 0.99, 1, and 1, respectively.

were 1 and 1, respectively. Sensitivity (100%) and specificity (60%), were highest in the forehead of WCDP, when compared to NCDP with linear kernel function. PPV and NPV were calculated with 10 out of 22 WCDP and 37 out 49 NCDP and values for PPV and NPV were 0.9 and 1, respectively. In HV, the fingertip had the highest sensitivity (100%) and specificity (80%), when compared to NCDP using multilayer perceptron kernel function. PPV and NPV were calculated with 5 out 15 HV and 37 out 49 NCDP. PPV and NPV were 0.97 and 1, respectively.

Finally to corroborate the efficacy of the diagnostic algorithm we calculated the area under the ROC curve (AUC) getting 0.99 for WCDP compared to NCDP, AUC 1 for HV compared to WCDP, and 1 for HV compared to NCDP (Figure 6).

The fingertip and lobe has a lot of vascularity, but in the forehead we have skull with high values of scattering in the near-infrared spectral range, it acts like a mirror. It is related with the fact the scattering coefficient (around $\mu_s' = 20\ \mathrm{cm}^{-1}$), on the other hand for ear lobe and fingertip has a low scattering coefficient in the near-infrared range, due that we have strong signals on forehead [20].

4. Conclusion

In this study Raman spectroscopy was evaluated as a non-invasive diagnostic tool to identify and classify HbA1c in healthy volunteers and diabetic patients, we do not measure HbA1c levels directly. We only identify and classify high and low levels of HbA1c in the diabetics and healthy patients. In our knowledge this is the first study *in vivo* to demonstrate the feasibility to identify and classify in high and low levels of HbA1c. PCA-SVM analysis added to Raman spectroscopy significantly increased the accuracy of this diagnostic tool. ROC curve confirmed the efficacy of the diagnostic algorithm proposed. The results of classification showed that forehead is the best anatomical region to differentiate between healthy volunteers (HV) vs well controlled diabetic patients (WCDP). Raman spectroscopy in combination with PCA-SVM analysis emerges as a promising non-invasive, accurate and qualitative diagnostic tool for HbA1c monitoring in healthy population and diabetic patients.

Acknowledgements The authors thank physicians, nurses, chemists and technicians from CEPREC and ARKAIA Laboratory for their support in this work, we thanks the physics Teresa Cerdà for her support with the images, and we thanks to CONACYT from the support received from the scholarship.

Author biographies Please see Supporting Information online.

References

[1] World Health Organization. Diabetes programme: World diabetes day 2015. http://www.who.int/diabetes/wdd_2015/en/http://www.who.int/diabetes/wdd_2015/en/

[2] World Health Organization. Diabetes programme: Use of glycated hemoglobin (HbA1c) in the diagnosis of diabetes mellitus. http://www.who.int/diabetes/publications/diagnosis_diabetes2011/en/http://www.who.int/diabetes/publications/diagnosis_diabetes2011/en/

[3] Konstantinos Makris and Loukia Spanous, J Diabetes Sci Technol. **5**(6), 1572–1583 (2011).

[4] Nadanathangam Vigneshwaran, Gopalakrishnapillai Bijukumar, Nivedita Karmakar, Sneh Anand, and Anoop Misra, Spectrochimica Acta Part A. **61**, 163–170 (2005).

[5] Annika M. K. Enejder, Thomas G. Scecina, Jeankun Oh, Martin Hunter, Wei-Chuan Shih, Slobodan Sasic, Gary L. Horowitz, and Michael S. Feld, Journal of Biomedical optics **10**(13), 031114 (2005).

[6] Jingwei Shao, Manman Lin, Yongqing Li, Xue Li, Junxian Liu, Jianpin Liang, and Huilu Yao, PLos ONE **10**(7), e48127 (2012).

[7] Juqiang Lin, Jinyong Lin, Zufang Huang, Peng Lu, Jing Wang, Xuchao Wang, and Rong Chen, Journal of Innovative Optical Health Sciences **7**(1), 1350051 (2014).

[8] Manikantan Syamala Kiran, Tamitake Itoh, Ken-ichi Yoshida, Nagako Kawashima, Vasudevanpillai Biju, and Mitsuru Ishikawa, Anal. Chem. **82**, 1342–1348 (2010).

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[9] A. Cao, A. K. Pandya, G. K. Serhatkulu, R. E. Weber, H. Dai, J. S. Thakur, V. M. Naik, G. W. Auner, R. Rabah, and D. C. Freeman, Journal of Raman spectroscopy **38**(9), 1199–1205 (2007).

[10] A. E. Villanueva-Luna, J. Castro-Ramos, S. Vazquez-Montiel, A. Flores-Gil, J. A. Delgado-Atencio, and E. E. Orozco-Guillen, Optical memory and neural networks **19**(4), 310–317 (2010).

[11] Sergios Theodoris, Konstantinos Koutroumbas, Pattern Recognition Forth edition (Elservier, San Diego California, 2009), p. 119–142, 275–276, 326–334.

[12] De Gelder, J. De Gussem, K. Vandenabeele, and Moens, Journal Raman spectroscopy (**38**), 1133–1147 (2007).

[13] I. Barman, N. Ch. Dingari, J. W. Kang, G. L. Horowitz, R. R. Dasari, and M. S. Field, Anal. Chem. (**84**), 2474–2482 (2012).

[14] L. Wang, D. He, J. Zeng, Z. Guan, Q. Dang, X. Wang, J. Wang, L. Huang, P. Cao, G Zhang, J. Hsieh, and J. Fan, Journal of biomedical optics **18**(8), 087001 (2013).

[15] J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-Garca, R. Huerta-Franco, J. L. Gonzlez-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez, and C. Medina-Gutiérrez, Laser Med Sci. **22**(4), 229–236 (2007).

[16] S. Sigurdson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, and H. C. Wulf, IEEE Transactions on biomedical engineering **51**(10), 1784–1793 (2004).

[17] Zhengfei Zhuang, Ning Li, Zhouyi Guo, Meifang Zhu, Ke Xiong, and Sijin Chen, Journal of biomedical optics **18**(3), 031103 (2013).

[18] Y. Li, J. Pan, G. Chen, C. Li, S. Lin, Y. Shao, S. Feng, Z. Huang, S. Xie, H. Zeng, and R. Chen, Journal of biomedical optics **18**(2), 027003 (2013).

[19] A. A. Martin, R. A. Bitar Carter, L. de Oliveira Nunes, E. A. Loschiavo Arisawa, and L. Silveira, Proc. SPIE 5321, Biomedical Vibrational Spectroscopy and Biohazard Detection Technologies, Bellingham, WA (SPIE, Bellingham, WA, July 1, 2004), pp. 1605–7422.

[20] Valery Tuchin, Tissue optics light scattering methods and instruments for medical diagnosis, Second edition (The society of photo-optical instrumentation engineers, Bellingham, Washington, 2007), p. 162, 190 and 210.

© 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim