



Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools

EDGAR GUEVARA,^{1,2,*} JUAN CARLOS TORRES-GALVÁN,² MIGUEL G. RAMÍREZ-ELÍAS,³ CLAUDIA LUEVANO-CONTRERAS,⁴ AND FRANCISCO JAVIER GONZÁLEZ²

¹*CONACYT-Universidad Autónoma de San Luis Potosí, Mexico*

²*Terahertz Science and Technology Center (C2T2) and Science and Technology National Lab (LANCyTT), Universidad Autónoma de San Luis Potosí, Mexico*

³*Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, Mexico*

⁴*Department of Medical Sciences, University of Guanajuato, Leon, Mexico*

*eguevara@conacyt.mx

Abstract: Type 2 diabetes mellitus (DM2) is one of the most widely prevalent diseases worldwide and is currently screened by invasive techniques based on enzymatic assays that measure plasma glucose concentration in a laboratory setting. A promising plan of action for screening DM2 is to identify molecular signatures in a non-invasive fashion. This work describes the application of portable Raman spectroscopy coupled with several supervised machine-learning techniques, to discern between diabetic patients and healthy controls (Ctrl), with a high degree of accuracy. Using artificial neural networks (ANN), we accurately discriminated between DM2 and Ctrl groups with 88.9–90.9% accuracy, depending on the sampling site. In order to compare the ANN performance to more traditional methods used in spectroscopy, principal component analysis (PCA) was carried out. A subset of features from PCA was used to generate a support vector machine (SVM) model, albeit with decreased accuracy (76.0–82.5%). The 10-fold cross-validation model was performed to validate both classifiers. This technique is relatively low-cost, harmless, simple and comfortable for the patient, yielding rapid diagnosis. Furthermore, the performance of the ANN-based method was better than the typical performance of the invasive measurement of capillary blood glucose. These characteristics make our method a promising screening tool for identifying DM2 in a non-invasive and automated fashion.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (170.5660) Raman spectroscopy; (170.4580) Optical diagnostics for medicine; (070.5010) Pattern recognition

References and links

1. World Health Organization, “Global Report on Diabetes,” www.who.int/diabetes/global-report.
2. IQWiG (Institute for Quality and Efficiency in Health Care), “Defunciones por diabetes mellitus por entidad federativa y grupo quinquenal de edad según sexo, 2010 a 2015,” http://www.beta.inegi.org.mx/app/tabulados/pxweb/inicio.html?rxid=75ada3fe-1e52-41b3-bf27-4cda26e957a7&db=Mortalidad&px=Mortalidad_4.
3. IQWiG (Institute for Quality and Efficiency in Health Care), “Type 2 diabetes: Overview” (PubMed Health, 2014).
4. R. T. Demmer, A. M. Zuk, M. Rosenbaum, and M. Desvarieux, “Prevalence of diagnosed and undiagnosed type 2 diabetes mellitus among US adolescents: results from the continuous NHANES, 1999–2010,” *Am. J. Epidemiol.* **178**(7), 1106–1113 (2013).
5. J. Beagley, L. Guariguata, C. Weil, and A. A. Motala, “Global estimates of undiagnosed diabetes in adults,” *Diabetes Res. Clin. Pract.* **103**(2), 150–160 (2014).
6. M. I. Harris and R. C. Eastman, “Early detection of undiagnosed diabetes mellitus: a US perspective,” *Diabetes Metab. Res. Rev.* **16**(4), 230–236 (2000).
7. J. F. Villa-Manríquez, J. Castro-Ramos, F. Gutiérrez-Delgado, M. A. López-Pacheco, and A. E. Villanueva-Luna, “Raman spectroscopy and PCA-SVM as a non-invasive diagnostic tool to identify and classify qualitatively glycated hemoglobin levels in vivo,” *J. Biophotonics* **10**(8), 1074–1079 (2016).

8. N. C. Dingari, G. L. Horowitz, J. W. Kang, R. R. Dasari, and I. Barman, "Raman spectroscopy provides a powerful diagnostic tool for accurate determination of albumin glycation," *PLoS One* **7**(2), e32406 (2012).
9. A. A. Martin, L. Pereira, S. M. Ali, C. D. Pizzol, C. A. Tellez, P. P. Favero, L. Santos, V. V. da Silva, and C. E. O. Praes, "Detection of advanced glycation end products (AGEs) on human skin by *in vivo* confocal Raman spectroscopy," in *Biomedical Vibrational Spectroscopy 2016: Advances in Research and Industry*, A. Mahadevan-Jansen and W. Petrich, eds. (2016), p. 97040S.
10. C. Koushik, S. Anuj, S. Neeraj, and S. Shiru, "Estimation of fasting Blood glucose levels by invasive and indigenously developed noninvasive technology and its correlation with the glycated hemoglobin (HbA1c) biomarker in healthy and diabetic subjects," *Res. J. Biotechnol.* **9**, 61–68 (2014).
11. E. Guevara and F. J. González, "Prediction of Glucose Concentration by Impedance Phase Measurements," in *MEDICAL PHYSICS: Tenth Mexican Symposium on Medical Physics* (AIP, 2008), Vol. **1032**, pp. 259–261.
12. E. Guevara and F. J. González, "Joint optical-electrical technique for noninvasive glucose monitoring," *Rev. Mex. Fis.* **56**, 430–434 (2010).
13. R. A. Shaw and H. H. Mantsch, "Infrared spectroscopy in clinical and diagnostic analysis," in *Encyclopedia of Analytical Chemistry* (John Wiley & Sons, Ltd, 2006).
14. M. Jermyn, J. Desroches, J. Mercier, M.-A. Tremblay, K. St-Arnaud, M.-C. Guiot, K. Petrecca, and F. Leblond, "Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts," *J. Biomed. Opt.* **21**(9), 094002 (2016).
15. M. Gniadecka, P. A. Philipsen, S. Sigurdsson, S. Wessel, O. F. Nielsen, D. H. Christensen, J. Hercogova, K. Rossen, H. K. Thomsen, R. Gniadecki, L. K. Hansen, and H. C. Wulf, "Melanoma diagnosis by Raman spectroscopy and neural networks: structure alterations in proteins and lipids in intact cancer tissue," *J. Invest. Dermatol.* **122**(2), 443–449 (2004).
16. J. Cheng, L. Xu, G. Lü, J. Tang, J. Mo, X. Lü, and Z. Gao, "Study on the echinococcosis blood serum detection based on Raman spectroscopy combined with neural network," *Optoelectron. Lett.* **13**(1), 77–80 (2017).
17. M. G. Shim, L. M. Song, N. E. Marcon, and B. C. Wilson, "In vivo near-infrared Raman spectroscopy: demonstration of feasibility during clinical gastrointestinal endoscopy," *Photochem. Photobiol.* **72**(1), 146–150 (2000).
18. M. Sattlecker, C. Bessant, J. Smith, and N. Stone, "Investigation of support vector machines and Raman spectroscopy for lymph node diagnostics," *Analyst (Lond.)* **135**(5), 895–901 (2010).
19. S. Li, Y. Zhang, J. Xu, L. Li, Q. Zeng, L. Lin, Z. Guo, Z. Liu, H. Xiong, and S. Liu, "Noninvasive prostate cancer screening based on serum surface-enhanced Raman spectroscopy and support vector machine," *Appl. Phys. Lett.* **105**(9), 091104 (2014).
20. B. Yan, B. Li, Z. Wen, X. Luo, L. Xue, and L. Li, "Label-free blood serum detection by using surface-enhanced Raman spectroscopy and support vector machine for the preoperative diagnosis of parotid gland tumors," *BMC Cancer* **15**(1), 650 (2015).
21. S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomed. Opt. Express* **7**(6), 2249–2256 (2016).
22. J.-O. Jeppsson, U. Kobold, J. Barr, A. Finke, W. Hoelzel, T. Hoshino, K. Miedema, A. Mosca, P. Mauri, R. Paroni, L. Thienpont, M. Umemoto, and C. Weykamp, International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), "Approved IFCC reference method for the measurement of HbA1c in human blood," *Clin. Chem. Lab. Med.* **40**(1), 78–89 (2002).
23. I. J. Pence, E. Vargas, and A. Mahadevan-Jansen, "Assessing variability of *in vivo* tissue Raman spectra," *Appl. Spectrosc.* **67**(7), 789–800 (2013).
24. J. Zhao, H. Lui, D. I. McLean, and H. Zeng, "Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy," *Appl. Spectrosc.* **61**(11), 1225–1232 (2007).
25. J. R. Beattie, A. M. Pawlak, M. E. Boulton, J. Zhang, V. M. Monnier, J. J. McGarvey, and A. W. Stitt, "Multiplex analysis of age-related protein and lipid modifications in human Bruch's membrane," *FASEB J.* **24**(12), 4816–4824 (2010).
26. A. M. Pawlak, J. V. Glenn, J. R. Beattie, J. J. McGarvey, and A. W. Stitt, "Advanced glycation as a basis for understanding retinal aging and noninvasive risk prediction," *Ann. N. Y. Acad. Sci.* **1126**(1), 59–65 (2008).
27. G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory* **14**(1), 55–63 (1968).
28. M. S. Bartlett, "Tests of significance in factor analysis," *Br. J. Stat. Psychol.* **3**(2), 77–85 (1950).
29. J. M. López-Alonso, J. Alda, and E. Bernabéu, "Principal-component characterization of noise for infrared images," *Appl. Opt.* **41**(2), 320–331 (2002).
30. F. J. González, J. Alda, B. Moreno-Cruz, M. Martínez-Escanamé, M. G. Ramírez-Elías, B. Torres-Álvarez, and B. Moncada, "Use of Raman spectroscopy for the early detection of filaggrin-related atopic dermatitis," *Skin Res. Technol.* **17**(1), 45–50 (2011).
31. J. Alda, C. Castillo-Martínez, R. Valdes-Rodriguez, D. Hernández-Blanco, B. Moncada, and F. J. González, "Use of Raman spectroscopy in the analysis of nickel allergy," *J. Biomed. Opt.* **18**(6), 061206 (2012).
32. I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer, 2002).
33. V. P. Singh, A. Bali, N. Singh, and A. S. Jaggi, "Advanced Glycation End Products and Diabetic Complications," *Korean J. Physiol. Pharmacol.* **18**(1), 1–14 (2014).
34. R. Meerwaldt, T. Links, C. Zeebregts, R. Tio, J.-L. Hillebrands, and A. Smit, "The clinical relevance of assessing advanced glycation endproducts accumulation in diabetes," *Cardiovasc. Diabetol.* **7**(1), 29 (2008).

35. L. Pereira, C. A. T. Soto, L. D. Santos, P. P. Favero, and A. A. Martin, "Confocal Raman Spectroscopy as an Optical Sensor to Detect Advanced Glycation End Products of the Skin Dermis," *Sens. Lett.* **13**(9), 791–801 (2015).
36. J. M. Ashraf, S. Ahmad, G. Rabbani, Q. Hasan, A. T. Jan, E. J. Lee, R. H. Khan, K. Alam, and I. Choi, "3-Deoxyglucosone: A Potential Glycating Agent Accountable for Structural Alteration in H3 Histone Protein Through Generation of Different AGEs," *PLoS One* **10**(2), e0116804 (2015).
37. N. Shangari and P. J. O'Brien, "The cytotoxic mechanism of glyoxal involves oxidative stress," *Biochem. Pharmacol.* **68**(7), 1433–1442 (2004).
38. I. Allaman, M. Bélanger, and P. J. Magistretti, "Methylglyoxal, the dark side of glycolysis," *Front. Neurosci.* **9**, 23 (2015).
39. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
40. S. Thorvaldsen, T. Flå, and N. P. Willlassen, "DeltaProt: a software toolbox for comparative genomics," *BMC Bioinformatics* **11**(1), 573 (2010).
41. G.-B. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Trans. Neural Netw.* **14**(2), 274–281 (2003).
42. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
43. P. Gaspar, J. Carbonell, and J. L. Oliveira, "On the parameter optimization of Support Vector Machines for binary classification," *J. Integr. Bioinform.* **9**(3), 201 (2012).
44. N. Chiles Shaffer, L. Ferrucci, M. Shardell, E. M. Simonsick, and S. Studenski, "Agreement and Predictive Validity Using Less-Conservative Foundation for the National Institutes of Health Sarcopenia Project Weakness Cutpoints," *J. Am. Geriatr. Soc.* **65**(3), 574–579 (2017).
45. E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *J. Neurosci. Methods* **250**, 126–136 (2015).
46. F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain* **130**(2), 314–333 (2007).
47. J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology* **148**(3), 839–843 (1983).
48. L. Franzen and M. Windbergs, "Applications of Raman spectroscopy in skin research--From skin physiology and diagnosis up to risk assessment and dermal drug delivery," *Adv. Drug Deliv. Rev.* **89**, 91–104 (2015).
49. X. Zhao, W. Zhao, H. Zhang, J. Li, Y. Shu, S. Li, L. Cai, J. Zhou, Y. Li, and R. Hu, "Fasting capillary blood glucose: an appropriate measurement in screening for diabetes and pre-diabetes in low-resource rural settings," *J. Endocrinol. Invest.* **36**(1), 33–37 (2013).
50. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.* **115**(3), 211–252 (2015).
51. A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. J. Minshew, C. S. Monk, S. Mueller, R.-A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn, K. A. Pelpfrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thivoux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, and M. P. Milham, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry* **19**(6), 659–667 (2014).
52. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," *ArXiv170502315 Cs* (2017).

1. Introduction

Diabetes mellitus affects more than 8.5% of the adult population; it is associated with over 3.7 million deaths worldwide per year [1] and in Mexico alone it is responsible for approximately 15% of deaths every year (98,521 out of 655,688) [2]. Diabetes mellitus type 2 (DM2), the most common type of diabetes, is characterized by inadequate insulin secretion or insulin resistance by the cells of the body. DM2 often develops later in life, although it may appear in younger people [3]. Unfortunately, about a third of the cases in younger people [4] and up to 45% of adult cases remain undiagnosed [5], leading to increased healthcare costs and further micro and macrovascular complications for the patient, such as cardiovascular disease, nephropathy, neuropathy, and retinopathy [6]. Therefore, early screening of DM2 would be very valuable to improve the management of the disease and the outcomes for the patients. Moreover, it could be even more useful if such screening were non-invasive, unlike current techniques that diagnose diabetes with a blood or interstitial fluid sample.

Raman spectroscopy is a promising tool for non-invasive biomedical applications, such as screening for metabolic conditions. This all-optical technique would eliminate the painfulness and invasiveness of both clinical measurements and self-monitoring of glucose, that rely on finger-prick blood sampling. Raman spectroscopy relies on the elastic scattering of photons by some molecules present in the sample, therefore the shift in energy of scattered photons depends on the specific bonds of the interacting molecule, yielding a molecular fingerprint. Furthermore, it has considerable benefits in biomedical diagnosis, such as non-invasiveness, relatively short acquisition time and the ability to provide molecular information. It has been shown that Raman spectroscopy and principal component analysis (PCA) combined with support vector machine (SVM) can classify glycated hemoglobin levels in vivo [7]. Dingari et al. have shown that *in vitro* quantification of glycated albumin in serum samples is possible through Raman spectroscopy and partial least squares (PLS) [8]. The group of A. Martin has established that confocal Raman spectroscopy can be used to assess the presence of advanced glycation end products (AGEs) in the dermis of DM2 patients and healthy volunteers [9].

Some complementary techniques to Raman spectroscopy such as near-infrared (NIR) spectroscopy have also been used to detect glucose in a non-invasive fashion, by a number of research groups [10], as well as our own laboratory [11,12]. However, the human body's major component, water (H_2O) presents some strong absorption bands in the NIR (e.g. 970nm and 1450nm) and thus may interfere with the spectra of other biological compounds [13]. On the other hand, H_2O does not interfere with Raman spectroscopy, which is especially advantageous for *in vivo* measurements.

Numerous techniques from the field of machine learning have been implemented in Raman spectroscopy analysis for classification purposes. Artificial neural networks (ANN) have been proposed by a number of researchers in biomedical applications, such as brain cancer detection [14], melanoma diagnosis [15], echinococcosis [16] and gastrointestinal tract diseases [17]. Support vector machines have also been widely utilized for lymph node diagnostics [18], prostate cancer screening [19], pro-operative diagnosis of parotid gland tumors [20] and analysis of dengue infection [21].

In this work, we perform *in vivo* Raman measurements at different anatomical locations and apply ANN and SVM to automatically classify each subject either as diabetic patients or as healthy.

2. Methods

2.1 Subjects and data acquisition

Eleven patients with type 2 diabetes (DM2, 7 females, Age: 49.5 ± 6.7 years) and 9 controls (healthy subjects) (Ctrl, 7 females, Age: 33.2 ± 4.9 years) were studied in the University of Guanajuato, Mexico. All subjects provided informed consent and the Institutional Review Board approved the study before subjects enrollment. In this study, all DM2 patients have been previously diagnosed by their physicians, through standard methods, such as fasting plasma glucose test. Monitoring glycated hemoglobin (HbA1c) in human blood is considered the gold standard to control the glycemic state of patients with diabetes in the long-term [22]. A sample volume of 5 μ l of whole blood was extracted from each subject and analyzed by boronic acid affinity chromatography (LabonaCheck MH-200, Ceragem Medisys Inc.) to determine HbA1c levels.

A portable Raman spectrometer (PEK-785, Agiltron Inc.) with a 785 nm, 90 mW laser beam, focused to a spot size of 200 μ m, was used. Five scans at 12 cm^{-1} resolution were collected at different skin sites, such as the left earlobe, left inner arm, left thumbnail and left median cubital vein, with approximately 15 s total exposure time. These settings allowed laser incidence well below the maximum exposure limits for skin, set by ANSI Z136.1. Therefore, for each subject, there were four data sets, each one corresponding to a particular skin site. All acquisitions were performed in sequential order, during a period of \sim 3 minutes

for each individual patient. The images corresponding to in vivo measurements in various skin sites are displayed in Fig. 1(A-D). The greatest care was taken to hold the probe at 90° angle against the skin, without exerting additional pressure, to minimize the effect of sources of variability on the acquired spectra [23].

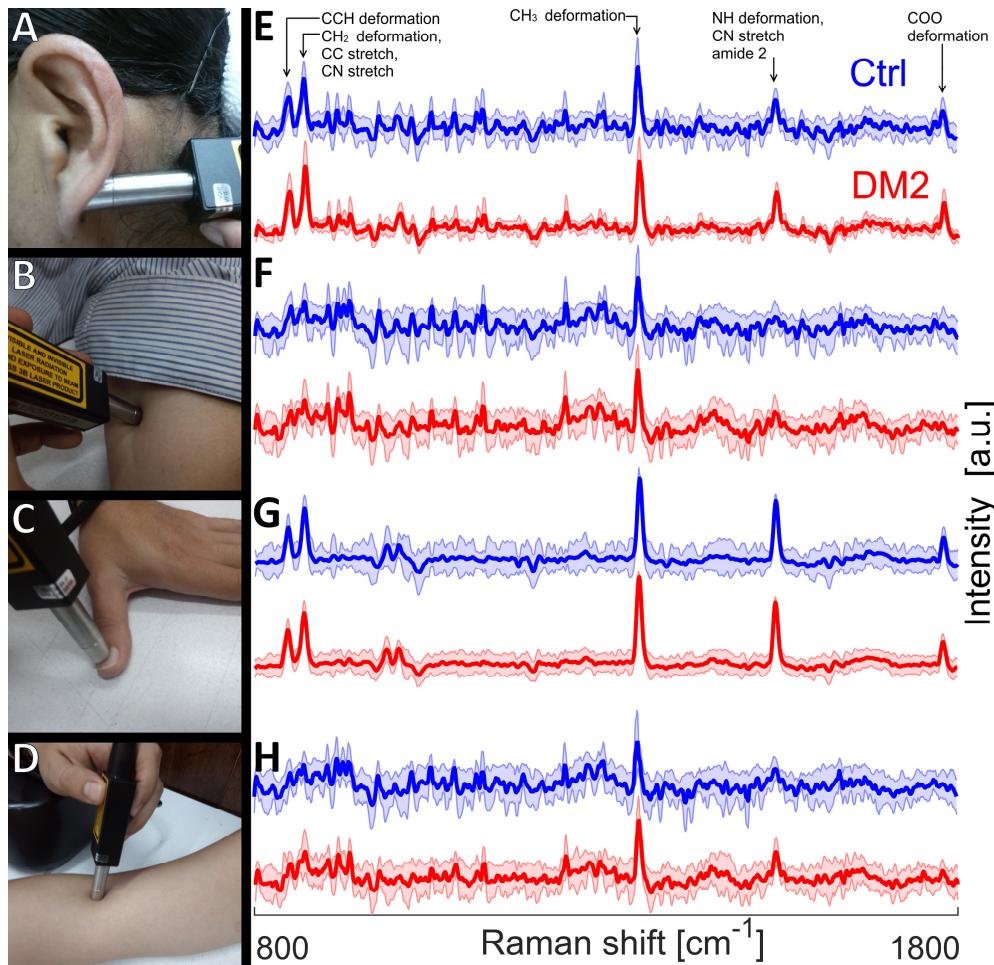


Fig. 1. Images of skin sites for in vivo Raman spectra acquisition: (A) ear lobe, (B) inner arm (C) thumb nail (D) median cubital vein. Also shown at the right side are the corresponding Raman measurements (mean \pm standard deviation) acquired at an excitation wavelength of 785 nm (E-H), where control spectra are displayed in blue, whereas DM2 spectra are shown in red.

2.2 Pre-processing of Raman spectra

All spectra were pre-processed to remove background fluorescence using an iterative polynomial fitting algorithm [24], then cropped to the region $800\text{-}1800\text{ cm}^{-1}$; this range was selected in order to match the spectral region of the advanced glycation end products (AGEs) reported in the literature [25,26]. This polynomial fitting algorithm was proposed by Zhao et al. [24] and is also known as the Vancouver Raman algorithm (VRA); it is widely used for fluorescence background removing in biomedical applications due to effectiveness and simplicity. The main advantage of this method is that it account for noise effects and Raman signal contribution. First, VRA applies a smoothing to the Raman spectrum, which reduces high frequency noise produced by the acquisition system or measurement procedure. Later, an iterative process is applied to adjust the polynomial function, which models the fluorescence,

to Raman spectrum. The Raman spectrum without fluorescence is computed by subtracting the adjusted polynomial from the original Raman spectrum.

Each measured Raman spectrum was standardized to have a unitary area under the curve, in order to avoid dependence on a single band. Before feeding the spectra to the classification algorithms, they were centered to zero-mean.

2.3 Principal component analysis and feature selection

Principal component analysis (PCA) can be described as an orthogonal transformation between the original data set onto a reduced subspace of features, with associated eigenvalues that describe their importance, obtained in decreasing order of explained variance. PCA has a two-fold purpose: first, reduce the computational complexity and second, mitigate the Hughes phenomenon, i.e. the reduction of predictive power with increased dimensionality [27]. Briefly, our data set can be described as a matrix X of order ($n \times d$) containing n spectra of length d, where d is the number of spectral points measured between 800 and 1800 cm⁻¹. PCA is based on a decomposition of the data matrix X into two matrices Z and V using a linear transformation of the form $Z = XV$. The matrix Z of order ($n \times d$) contains the original data in a rotated coordinate system or principal component space. The matrix V of order ($d \times d$) contains the eigenvectors and eigenvalues of the covariance matrix of X. The matrix V are also known as Loadings matrix, and the matrix Z as Scores matrix.

The appropriate number of principal components to retain was determined by Bartlett's chi-square test [28–31], where eigenvalues are left out sequentially until the test for equality fails to be rejected; those first excluded components are then retained. Although Bartlett's test is an objective procedure, an empirical rule based on the cumulative variance explained by the principal components is usually used to determine the number of components to retain. This empirical model is illustrated in panel (C) of Fig. 2. However Jolliffe [32] states that there is no clear advantage of a specific method over the rest. Visualizing features in more than two dimensions becomes increasingly hard, therefore, for the sake of illustrating the separating capabilities of our SVM model a simple search model was implemented: the algorithm searched all subsets of 2 principal components that maximized the objective function, in this case, classification accuracy.

2.4 Advanced glycation end products (AGEs)

In order to explore the relationship between the acquired spectra and the molecules related to diabetic complications, such as AGEs [33], Raman spectra from several AGEs (glyoxal-lysine dimer GOLD, methylglyoxal-derived hydroimidazolone MG-H2, and pentosidine) and AGEs precursors (3-deoxyglucosone, glyoxal, and methylglyoxal) were digitized from the literature [25,26]. GOLD and MG-H2 have been found to accumulate in tissue in diabetes, through oxidation of polyunsaturated fatty acids [34], while pentosidine has been identified with Raman microscopy in the progression of DM2 [35]. All the precursors investigated are intermediate products in the formation of AGEs, more specifically, 3-deoxyglucosone is one of the Amadori products [36], glyoxal is a metabolite formed by oxidative degradation of glucose [37], and methylglyoxal is a by-product of glycolysis [38]. A correlation analysis was performed on their spectra and the principal components of each data set. In order to control for type I errors in the resulting correlations, a false-discovery rate procedure was implemented [39].

2.5 Statistical analysis

Fisher's exact test was carried out to find mid-P value for gender difference [40]. Wilcoxon-Mann-Whitney test was used to evaluate statistical significance in HbA1c measures. Unpaired two-sample Student's t-test was applied to find out the statistical level of significance of classifiers vs. random-chance. All results are presented as an arithmetic mean \pm standard deviation, except where noted.

2.6 Artificial neural network

A feed-forward artificial neural network (ANN) classifier was implemented with a single hidden layer; this hidden layer had 14 neurons with sigmoid activation, and one output neuron (Fig. 2(A)). The number of hidden neurons was selected by following the method devised by Huang [41]. The ANN was trained to a maximum of 1000 epochs using scaled conjugate gradient backpropagation as the network learning algorithm. The stopping criterion chosen was cross-entropy error, which was set to 1×10^{-5} . All weights and biases were randomly initialized.

2.7 Support vector machine

A support vector machine (SVM) is a robust classifier, easy to implement and performs well when applied to the unseen data set [42], while optimizing the decision boundary. After choosing the appropriate number of components to retain in our model, according to Bartlett's test, described in section 2.3, these principal components were used as inputs to the SVM model. In the proposed automated system for DM2 screening, a linear kernel is utilized. This kernel assists in deriving complex relationships between the possible output classes and the Raman spectra. A radial basis function (RBF) kernel was also investigated, although its results were not substantially better than those obtained with a linear kernel. Both the factor for the soft-margin function and the RBF kernel sigma were fine-tuned following a grid-search methodology over the validation sub-set [43].

2.8 Training and testing data sets

Each one of the four data sets is randomly divided into 10 partitions of equal size, i.e. two subjects per partition. This process leaves eight sub-sets for training of the classifier with 80% of the samples, one sub-set for validation of the classifier with 10% of the subjects, and one sub-set for testing of the classifier with 10% of the subjects. This process, of randomly choosing one sub-set for testing, another for validation, and the rest for training, was repeated 10 times, i.e. until each of the sub-sets is used once for testing, in a 10-fold cross-validation. Since this process does not use all the subjects' data for building the classifier models, it prevents overfitting during training [44].

In order to account for variability in the random initial conditions of the ANN, and in the partition sets of both the ANN and SVM, all metrics of performance reported in this paper were computed by averaging the results of one thousand 10-fold cross-validation runs.

2.9 Assessment of performance

The success of classification of various pairwise combinations of sampling sites and classification methods was assessed with the following metrics were used: sensitivity (Se), specificity (Sp), Geometric mean of sensitivity and specificity (G-m), precision (also known as positive predictive value PPV), negative predictive value (NPV), F-measure (F-m) and accuracy (Acc), defined in the following equations as function of false positives (FP), true positives (TP), false and true negatives (FN and TN, respectively):

$$Se = \frac{TP}{TP + FN}. \quad (1)$$

$$Sp = \frac{TN}{TN + FP}. \quad (2)$$

$$G - m = \sqrt{Se \cdot Sp}. \quad (3)$$

$$PPV = \frac{TP}{TP + FP}. \quad (4)$$

$$NPV = \frac{TN}{TN + FN}. \quad (5)$$

$$F - m = \frac{2 \cdot PPV \cdot Se}{PPV + Se}. \quad (6)$$

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}. \quad (7)$$

To assess the capabilities of a classifier with a limited number of samples, it is recommended to compare its performance against random chance [45,46]. Therefore, our four data sets were randomly labeled (either DM2 or Ctrl) before each cross-validation run and then both the SVM and the ANN were re-run on all data sets, as described above. Significance level of the classification was computed for the given sample size ($n = 20$) and number of classes ($c = 2$) at $\alpha = 0.05$. A receiver-operating characteristic (ROC) curve and its corresponding area under the curve (AUC) were also used as evaluation criteria for comparing both classifiers, following the methodology proposed by Hanley et al. [47].

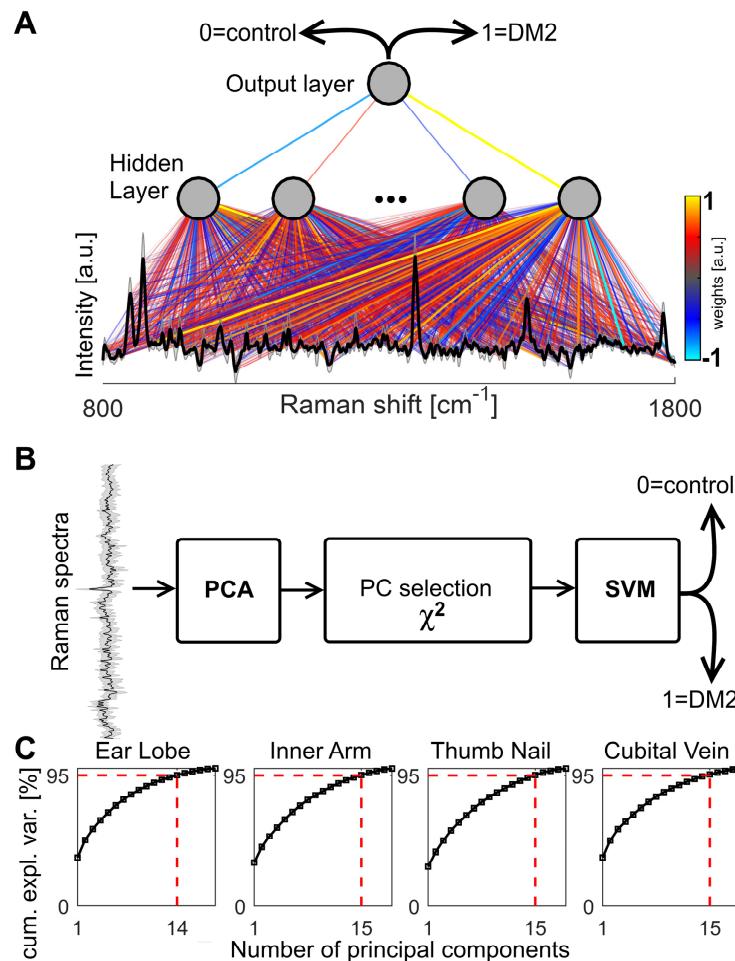


Fig. 2. Schematic diagrams of the supervised classifiers: (A) artificial neural network (ANN) architecture (B) support vector machine (SVM) structure (C) Cumulative percentage of total variation as explained by each principal component.

3. Results and discussion

The Raman spectra obtained at different sampling sites (ear lobe, inner arm, thumbnail and median cubital vein) from DM2 and Ctrl groups are shown in Fig. 1, panels E-H. Mean HbA1c levels in the DM2 group were $6.9 \pm 1.6\%$ while the Ctrl group showed mean HbA1c $4.5 \pm 0.3\%$ ($p = 8.4 \times 10^{-4}$).

Figure 1, panels E-H show some of the most recognizable Raman bands identified in skin: at 850 cm^{-1} there is deformation (CCH) aromatic, at 883 cm^{-1} deformation (CH2), stretch (CC), stretch (CN), at 1385 cm^{-1} deformation (CH3) symmetric, at 1552 cm^{-1} deformation (NH), stretch (CN) amide 2 and at 1768 cm^{-1} deformation (COO) [48].

Since no prominent bands showed a distinctive enhancement in DM2 patients, a machine learning approach allowed us to recognize patterns difficult to unravel with a simple band intensity or shape analysis. Figure 3(A) shows the ROC curve of the ANN classifier, including 95% confidence intervals, while Fig. 3(C) depicts the corresponding ROC curve of the SVM with linear kernel, while Fig. 6(A) shows the ROC curve for a RBF kernel. Table 1 summarizes the average values of the area under the curve (AUC). Figure 3(B) displays the different metrics used to assess the performance of our proposed ANN. The median values obtained for each different metric are the following: Se, Sp, PPV and Acc values ranged from 88.9% to 90.9%, G-m values spanned from 88.9% to 92.0%, F-m showed values between 88.2% and 90.0%, while NPV varied from 90.0% to 90.9%. When compared to fasting capillary blood glucose testing, a monitoring method, proposed for screening in remote areas [49], our method based in ANN showed improved sensitivity, specificity and AUC. Furthermore, when compared to the gold standard in screening DM2, the fasting venous plasma glucose measurement, our method showed improved sensitivity, a comparable AUC, but poorer specificity. These results suggest that ANN and Raman spectroscopy can be used for diabetic screening. The performance metrics for the SVM classifier with a linear kernel are shown in Fig. 3(D) and their median values are as follows: Se ranged from 69.6% to 81.9%, Sp varied between 76.5% and 100%, G-m values were between 76.4% and 83.4%, F-m ranged between 73.5% and 82.0%, PPV varied between 72.8% and 100% while NPV ranged from 63.6% to 86.4%. Median Acc values varied from 76.0% to 82.5%. The performance of the SVM classifier with RBF kernel is summarized in Fig. 6(B). These outcomes indicate that the SVM classifier depends largely on the probing location to yield accurate results, while ANN provides accurate results independently on the acquisition site.

Table 1. Comparison of the average area under the curve (AUC \pm 95% confidence intervals) of both classifiers

Skin site	Classifier AUC	
	ANN	SVM (linear kernel)
Ear Lobe	0.93 ± 0.01	0.82 ± 0.02
Inner Arm	0.96 ± 0.01	0.86 ± 0.02
Thumb Nail	0.92 ± 0.01	0.95 ± 0.01
Cubital Vein	0.95 ± 0.01	0.93 ± 0.01

Fisher's exact test showed no difference in gender ($p = 0.4892$). However, a statistically significant age difference between groups ($p < 0.001$) was found, therefore we tested the hypothesis that this factor could be driving the classification, by grouping the subjects according to their age: one group where they were older than the median age of 44.5 years ($n = 10$), and another group where they were younger ($n = 10$). The classification accuracy of all classifiers diminished to ~50%, thus suggesting that the age difference was not an influencing factor for classification accuracy of DM2 group vs. Ctrl group. Furthermore, the implemented ANN classifier showed a much higher accuracy than the minimum rate to assert statistical significance ($\text{Acc} = 70$, $*p < 0.05$) [45] in all data sets, as shown in Fig. 3(B). Statistical

significance was achieved for all the data sets using an ANN classifier, albeit not for the SVM classifier, which showed a performance significantly better than random chance when applied to the cubital vein and thumbnail data sets, but not for the earlobe and inner arm data, as depicted in Fig. 3(D). Although 10-fold cross validation ensures that each of the 10 sub-set goes 9 times in the training group and 1 in the test group, it does not ensure that each test will have the same accuracy, hence the results presented here are the average of multiple runs.

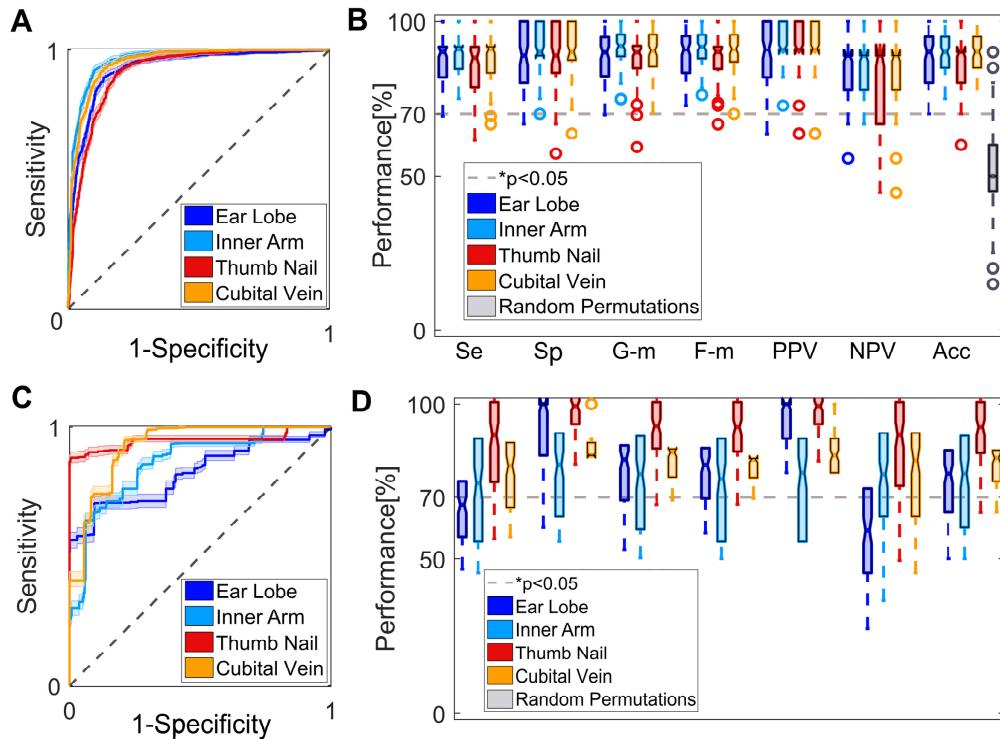


Fig. 3. (A) ROC curve of the proposed ANN (B) Performance metrics of the ANN for various sampling sites, dotted line at 70% represents the minimum accuracy to assert statistical significance (C) ROC curve of the SVM classifier with linear kernel (D) Performance metrics of the SVM classifier with linear kernel, dotted line at 70% represents the minimum accuracy to assert statistical significance

The correlation analysis between the scores of the principal components of our data set and the Raman spectra of AGEs are shown in Fig. 4. For the data acquired on the earlobe there was a significant correlation between the first PC and glyoxal and GOLD, as well as MG-H2, the third component shows a correlation to MG-H2 and pentosidine, the fourth component is correlated to MG-H2 and the fifth component to pentosidine. Inner arm data also showed a significant correlation between the first PC and GOLD. Data from the thumbnail also displays a significant correlation between the first component and GOLD, as well as between the third component and MG-H2. Panel D of Fig. 4 displays a significant correlation between GOLD and the first PC and between pentosidine and the fourth component from cubital vein data. It is to be noted that the correlation between the first PC and GOLD is always present regardless the sampling site.

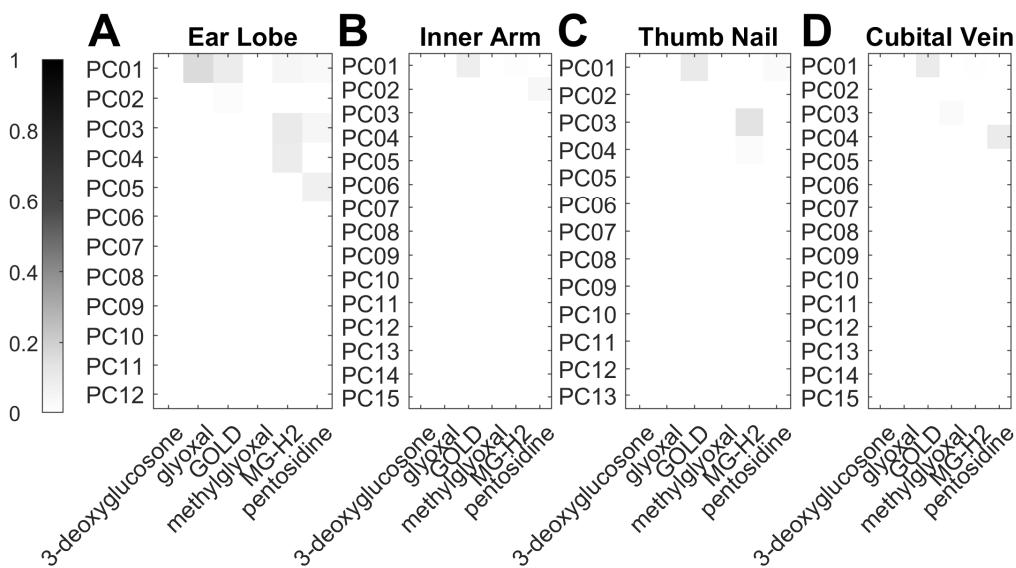


Fig. 4. Heatmaps of the Pearson's correlation mapping (absolute value) between the AGEs spectra (bottom) and the scores of principal components of the patients' data(left axis) at different sampling sites: A) Ear lobe, B) Inner arm, C) Thumbnail and D) Cubital vein. Correlations have been FDR-corrected.

Figure 5 shows the decision boundary of the linear RBF kernel of one of the runs using 2 PC components. A sequential feature selection process was carried out, where the sub-set of two PC components that minimized the misclassification rate was chosen. Decision values of the SVM model are also overlaid as a false-color image, where cold colors represent the decision values for which “Ctrl” label was assigned, meanwhile warm colors correspond to label “DM2”. It can be observed that these examples show an acceptable separation of the subjects by optimizing a decision boundary between the DM2 and Ctrl populations, despite a few misclassifications.

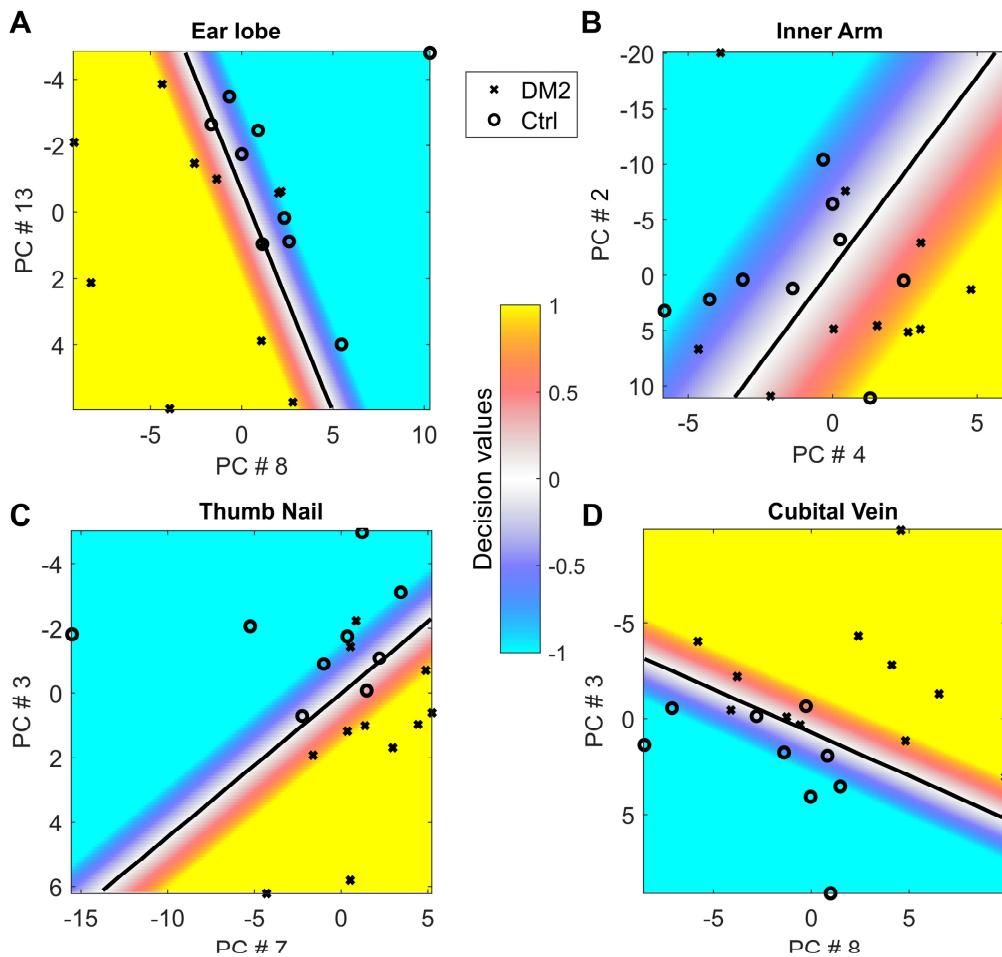


Fig. 5. Examples of SVM classification on a bi-dimensional space from different skin sites:
 (A) ear lobe (B) inner arm (C) thumbnail (D) cubital vein.

4. Conclusion

The results from this proof-of-concept study suggest that Raman spectra provide molecular signatures that, in conjunction with machine learning techniques, can be used to perform single-subject prediction of DM2, despite the small cohort size. This drawback may be addressed by data sharing models, such as those previously used in computer vision [50], neuroimaging [51] or thorax diseases [52]. The results also show that AGEs may explain the screening capabilities of our method. Further research is being carried out in our group to increase sensitivity as well as specificity. There is also the possibility of the concurrent use of Raman spectroscopy with other existing methods to increase their efficiency. The results presented in this work demonstrate an overall better performance of ANN in conjunction with Raman spectroscopy than the capillary blood glucose measurement and a comparable performance to the gold standard in screening, the venous plasma glucose test. Using ANN, the skin location with the highest classification accuracy is the inner arm, with 96%. In conclusion, Raman spectroscopy together with ANN has the potential for in-vivo, non-invasive, quick screening of diabetic condition in the general population.

Appendix

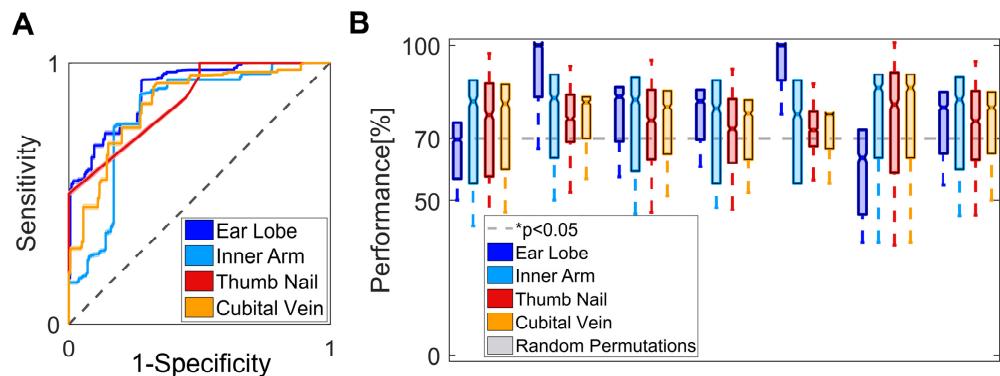


Fig. 6. (A) ROC curve of the SVM classifier with RBF kernel (B) Performance metrics of the SVM classifier with RBF kernel, dotted line at 70% represents the minimum accuracy to assert statistical significance.

Funding

Consejo Nacional de Ciencia y Tecnología.

Acknowledgments

This work was supported in part (E. G.) by the “Cátedras CONACYT” program, project 528. J.C.T-G. acknowledges support from CONACYT through scholarship No. 304501 and Beca Mixta de Movilidad Nacional 2016-291061. The authors also acknowledge support from CONACYT and the National Labs program through LANCYTT, the Terahertz Science and Technology National Lab. F. J. González would like to acknowledge support from Project 32 of “Centro Mexicano de Innovación en Energía Solar”.

Disclosure

The authors declare that there are no conflicts of interest related to this article.