



HAL
open science

Application of Machine Learning Methods to Raman Spectroscopy Technique in Dentistry

Iulian Otel, J. M. Silveira, V. Vassilenko, A. Mata, S. Pessanha

► **To cite this version:**

Iulian Otel, J. M. Silveira, V. Vassilenko, A. Mata, S. Pessanha. Application of Machine Learning Methods to Raman Spectroscopy Technique in Dentistry. 12th Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS), Jul 2021, Costa de Caparica, Portugal. pp.339-348, 10.1007/978-3-030-78288-7_33 . hal-03685921

HAL Id: hal-03685921

<https://inria.hal.science/hal-03685921v1>

Submitted on 2 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Application of Machine Learning Methods to Raman Spectroscopy Technique in Dentistry

Iulian Otel^{1,2}, J. M. Silveira^{1,3}, V. Vassilenko^{1,2}, A. Mata^{1,3}, S. Pessanha^{1,2}

¹ LIBPhys – Laboratory of Instrumentation, Biomedical Engineering and Radiation Physics, Portugal.

² Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2829-516, Monte Caparica, Portugal.

³ Faculdade de Medicina Dentária da Universidade de Lisboa, 1649-003 Lisboa, Portugal.

Abstract. Raman spectroscopy is nowadays regarded as a practical optical method and non-destructive photonic tool, which can be applied in several biomedical fields for analyzing the molecular composition. This technique is considered appropriate for human dental tissues characterization, from caries detection to evaluation of demineralization caused by acidic external agents. Discrimination techniques (linear regression), and classification techniques (neural networks) are often used for spectroscopic data analysis in disease detection and identification. Usually, Raman raw spectra obtained from teeth are **processed** using baseline correction, smoothing, **normalized** for noise, fluorescence, shot noise removal, and subsequently **analysed** using principal component analysis, to reduce the variable dimensionality. Raman chemical images can be constructed with another simple and uncomplicated unsupervised machine learning method – represented by k-means clustering, enabling the identification of similar areas/features, for classifying different acquired spectral data. The Machine Learning methods choice depends always on type and amount of information provided by Raman spectra. In this paper, was applied Principal Component Analysis methods to the analysis and interpretation of several parameters extracted from Raman spectra acquired before and after a simulated acid challenge of human enamel. These parameters and their correlation allow to assess the protective effect of a fluoride-based dental varnish.

1 Research Question and its Motivation

Raman spectroscopy, an emerging technique based on the functioning principle of vibrational spectroscopy, is presently nowadays regarded as a viable, non-destructive and practical optical method for different biomedical applications, which is mainly employed for the analysis of molecular biochemical and structural composition. This method is becoming progressively relevant in scientific biomedical research especially for its main features, such as high biochemical and structural specificity, contactless, non-invasiveness, non-destructiveness, and extremely low water sensitivity, as well as the possibility to be employed in the near-infrared (NIR) region

and remotely by using with fiber-optic probes and devices [1, 2]. When light arrives to an examined sample, a small part of it is scattered in several different directions.

When these light photons are scattered from a certain molecule or existent crystal, the greatest part of all these photons is elastically dispersed. The largest part of the scattered photons is of the same frequency when compared to the incident photons. Nevertheless, a little fraction of light (which is approximately 1 in 10 million photons) is scattered at optical frequencies distinct, and typically lower, than the frequency of the incident photons. The process causing the described inelastic scatter is mainly known as the Raman scattering. And when a given sample, compound or mixture that is under examination is irradiated or illuminated by a monochromatic light source, such as laser, the obtained spectrum resulting from the scattered light involves a strong main line, called the exciting line, being of the same frequency as the incident light together with weaker lines on each side shifted from the main line [3-7]. A typical Raman spectroscopic measurement usually consists in directing a focused laser beam onto the sample and recording the energy profile of the light that is scattered [8-10]. Raman spectroscopy has shown repeatedly to be an appropriate method for the characterization of dental tissues, from caries detection to the evaluation of demineralization caused by acidic external agents [1, 2, 11]. The sensitivity of this technique for alterations in the symmetric stretching band that is characteristic for the phosphate ions in hydroxyapatite matrix can be used as a powerful tool for early diagnostics, even before signs of demineralization are detected by conventional methods [8, 9]. Human dental enamel represents the highest mineralized, as well as the toughest and hardest tissue in the human body. Its main function consists in covering and protecting the anatomic crown of the tooth. It resists to chewing forces and protects the internal layers, dentin, and pulp [3, 4, 11]. The provided information and evidence on the tooth's enamel crystalline state is provided by the determining the depolarization ratio (DR) of the ν_1 symmetric stretching band of phosphate (located approximately at 959 cm^{-1} wavenumber), while the determination of the DR of phosphate characteristic band is achieved by acquiring and comparing two distinct Raman measurements (Figure 1) of the same spectral region. These measures are usually obtained with parallel and cross-polarized (orthogonal) configurations of the employed spectrometer [1, 2, 5, 11, 12].

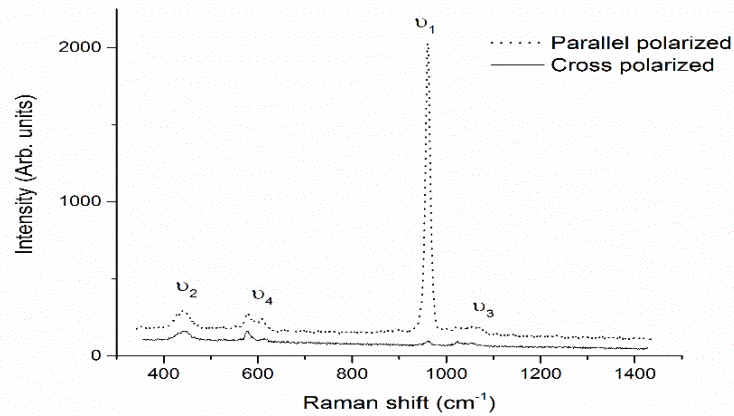


Fig. 1. Representation of two distinct Raman spectra, obtained with parallel and orthogonal polarized configurations of the used spectrometer, acquired from dental sound enamel. Raman peaks ν_2 at approximately 430 cm^{-1} , and ν_4 at 590 cm^{-1} , attributed to PO_4^{3-} groups in symmetric and asymmetric bending vibrations, respectively. Peak ν_1 is assigned to symmetric stretching band of phosphate at $\sim 959\text{ cm}^{-1}$, while ν_3 corresponds to a band with two distinct peaks at 1042 and 1070 cm^{-1} which is specific for PO_4^{3-} groups in the asymmetric stretching vibration of surface enamel's hydroxyapatite.

A machine learning (ML) model that combines principal component analysis (PCA) as well as support vector machine (SVM) optimized algorithms can be successfully implemented for an enhanced distinction and classification of spectral data, especially to detect and identify specific characteristics. Raman chemical images can be constructed by using another simple and unsophisticated, however useful unsupervised machine learning algorithm method, known as k-means clustering, that enables the identification of similar areas with the same features, to classify the acquired spectral data based on several Raman peak intensity ratios. The choice of ML methods to apply depends always on the type and amount of information contained in the Raman spectra [5].

2 Raman Spectroscopy Combined with PCA Method

In the last decades, the potential and usefulness of using multivariate techniques (indicated in Figure 2) for spectroscopic data analysis in disease detection, identification, comparison, as well as classification, has been successfully explored, providing very good results [1-5, 8, 11-12]. All existent multivariate classification techniques can be mainly separated into unsupervised methods and supervised learning procedures. Data compression method such as principal component analysis (PCA) is often employed to contribute for variability in the investigated data. PCA, considered a projection method, is a pre-processing statistical procedure that transforms several correlated variables into a lesser number of variables that are uncorrelated (called principal components) that still contains most of the information

of the primary larger set. Consequently, analysis methods, for instance, hierarchical cluster analysis (HCA), linear discriminant analysis (LDA) and other types, have also been frequently used to create classification algorithms for normal healthy condition and disease differentiation [6, 9]. Machine learning (ML) can include multivariate methods, which encompass several multivariate classification methods (also known as pattern recognition techniques) and multivariate regression. In unsupervised pattern recognition techniques, for example, PCA and cluster analysis (CA), there is not required any prior knowledge about the training set samples (in this case, represented by the acquired Raman spectra). Both referred statistical analysis methods are frequently applied to consider differences and/or similarities between spectra or image areas with the same features. Alternatively, supervised pattern recognition techniques, for instance, artificial neural networks (ANNS) and LDA necessitate minimal previous information from preceding experiences, for instance, undeniably detecting and identifying examined samples from disease or differentiating samples from healthy normal cases. In this manner, supervised learning methods are resulting in a more precise and accurate identification and classification, while unsupervised methods are considered more useful for investigative studies and exploratory analysis of data [9]. Generally, statistical methods applied to spectroscopic techniques, such as PCA, are based on the analysis of an entire given spectrum (comprising characteristic bands known as fingerprints) that is under examination. However, this approach does not provide valuable interpretations or explanations into the origins of the detected/identified changes. Additionally, statistical analysis methods usually presume a linear relationship between the components belonging to a certain Raman spectrum. Therefore, an adequate selection of quantitative information about the source of the Raman specific spectral peak data can be often problematic because due to the disorganized or, sometimes, unclear tissue type identification/classification or interference/overlay between different classes or varieties of spectroscopic characteristics (such as fluorescence, scattering, and absorption effects). Consequently, there is a tremendous need for developing quantitative models and quantifiable approaches that consider the investigated optical, structural, and biochemical properties of the examined tissue [6, 9].

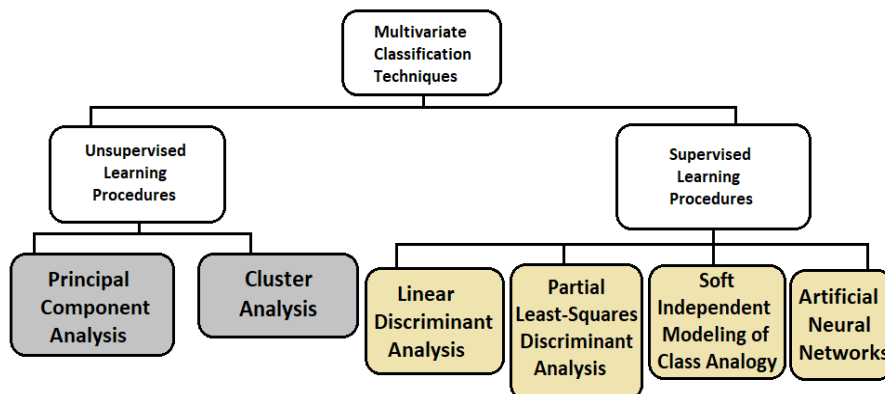


Fig. 2. Schematic representation for some of the most frequently applied multivariate data analysis (MVA) methods to spectroscopic techniques and spectral data processing and analysis [6].

So far, the principal component analysis method represents the unsupervised multivariate statistical tool most extensively used, which is widely applied especially when vast spectral data must be transformed into a smaller number of independent variations known as principal components (PCs), while contributions of these components are known as scores, which are extensively used as parameters for classification. PCA algorithms are frequently applied directly to collected (Raman) spectra within a certain data set to reduce the original dimensionality into a linear combination of orthogonal basis spectra known as principal components (PCs), resulting afterwards in derived loading plots represented as plots of first three PCs as function of the wave number. Generally, the first PC is the most important and accounts for the spectral features with highest variation, while subsequent secondary and tertiary PCs represent features with progressively lower variance. PCA method is typically used to characterize and distinguish the Raman spectra of a specific “calibration set” of samples that contain at least two distinct features/chemical components, as determined by pathology under study. As already described, in the case of Raman measurements, PCs represent an orthogonal set of spectral data (with varying spectral parameters), a linear numerical combination of which can accurately provide valuable insights on each of the existent spectra in the sample set. By correlating the mentioned fitting coefficients “scores” of the analyzed spectra in the data set with their established classifications, an accurate diagnostic algorithm can be easily obtained. However, its main disadvantage lies on the fact that PCs are pure mathematical concepts, without any direct physical significance or meaning [6, 9].

3 State of the Art - MVA Applied to Raman Spectral Data Acquired from Dental Samples

According to several research studies [3-5, 11-22], Raman spectroscopy can be employed to obtain precise results, showing a very good sensitivity and specificity. When this spectroscopic technique is used in combination with PCA [12, 15-19, 21] or other MVA [9, 18-21] methods, it allows us to obtain unique spectral characteristics extracted from raw acquired spectra, which can be used for identification of similar areas, specific features, or certain types of samples, along with differentiation and classification of different sample or tissue condition, according to the case. Gonzalez-Solis et al. [15] demonstrated that Raman spectroscopic technique and PCA method can be used simultaneously to discriminate healthy tooth with sound enamel from severe dental fluorosis samples with high levels of sensitivity and specificity. Furthermore, the Raman data of affected enamel by fluorosis “allowed to identify the molecule of calcium fluoride and the vibrational modes of phosphate and carbonate ions” [15]. Another research group – Barrera – Ortega [16] performed a study showing also that Raman-PCA combination can be used to differentiate sound from demineralized enamel, as well as identify and assess enamel remineralization level by using different fluoride solutions. The authors stated that Raman-PCA technique represent “an excellent non-invasive method to demonstrate qualitatively the early detection and monitoring of incipient lesions due to caries”. In a study performed by Mihaly et al. [17] FT-Raman spectroscopy-PCA

methods were also used for the analysis of enamel surfaces of healthy unaffected, affected (cavitated), and highly affected tooth samples. This work revealed the existence of a susceptibility and tendency in teeth to caries that can be directly correlated with the structural/compositional enamel alterations that occur mainly in the principal c-axis of hydroxyapatite crystals. Thus, MVA unsupervised methods, such as PCA, applied to Raman technique have a great utility for data analysis in several relevant fields of dentistry diagnostics and research, as follows: (1) early detection and recognition of carious lesions; (2) identification and differentiation between normal and affected samples (by tooth decay, hypo-mineralization, fluorosis, Amelogenesis Imperfecta); (3) bleaching products performance; (4) assessment of dental pharmaceutical products; (5) demineralizing attack from acidic challenges, foods, and beverages; (6) proteomic differences between different types of human teeth; (7) remineralization agents and other related treatments. Raman raw spectra obtained from dental samples are usually processed by carrying out initially a baseline correction, smoothing, and normalization for noise, fluorescence, and/or shot noise removal, and then analysed using PCA, to reduce the variable dimensionality, to facilitate and simplify the data processing and differentiation [12]. The Raman spectral changes are mainly due to demineralization-induced and decay alterations of enamel crystallite morphology, structure, and orientation. Any small change in the enamel tissue that is produced by caries could end in a higher value of depolarization ratio of ν_1 symmetric stretch band, while the other bands of regions ν_3 , ν_2 and ν_4 do not show significant alterations [3, 21].

4 Research Contribution and Innovation - Evaluation of Protective Effect of Dental Varnish using Raman Microscopy and PCA Approach

4.1 Dental Samples

Two experimental groups of human dental enamel samples, control (C) and treatment (T), each containing 10 tooth samples, were subjected to an acid challenge, in order to test the erosion effect of the acid on the enamel. Previously to the acid attack, the T group was cleaned with Couto toothpaste and treated with a commercial dental fluorinated varnish - VOCO Profluoride, while group C was only cleaned with Couto toothpaste.

4.2 Research Design and Methodology

Both sets of samples were analysed using Raman during three distinct stages: 1 - before any treatment, 2 - after application of dental varnish and/or toothpaste and 3 - after acid challenge. Raman spectra were acquired using a Horiba XploRA Raman Confocal Microscope employing a NIR laser diode source operating at a wavelength of 785 nm, using a 1200 lines/mm grating, with a spectral range between 300 cm^{-1} and 1800 cm^{-1} , resulting in a spectral resolution of 4 cm^{-1} . An entrance slit of 200 μm together with a confocal hole of 300 μm were used to disperse the scattered radiation

collected by an 100x objective (N. A. = 0.9) onto the air-cooled CCD array of an Andor iDus detector. It was also employed a 50% neutral density filter resulting in an incident power on the dental sample of 5.0 ± 0.4 mW (lasercheck®, Edmund optics).

For each enamel specimen, an average of 20 measurements (10 different points with parallel polarization, always followed by other 10 with orthogonal polarization) were performed, with an exposure time for each measurement of 20 s with 4 accumulations. The following parameters were obtained and compared in each step of the study: depolarization ratio (DR), anisotropy, band position and full width at half maximum (FWHM) of the symmetric stretching band of phosphate (~ 959 cm^{-1}), according to Pessanha et al. [22]. PCA was performed using the mean values of all spectral parameters for each sample.

4.3 Results and Discussion

In this work, the first three principal components were extracted from the spectral parameters of all enamel samples and were used to identify, distinguish, and categorize two small clusters: control and varnish treated sample groups (Figure 3) and to show likewise the correlation degree between these 4 different peak spectral features, namely the depolarization ratio, anisotropy, peak position, and FWHM, which yield useful structural and chemical information. PC1 scores account for the majority of variance within all analysed data sets, with proportions of 50.55 %, 57.09 %, and 55.18 % for 1st, 2nd, and 3rd treatment stage, respectively. As can be seen in Figure 3, after 1st treatment stage the samples are randomly dispersed, revealing that there is no significant difference between the samples' overall condition within and between groups initially. After the 2nd stage, the treated samples tend to form a group in the centre, starting to separate from control samples that continue dispersed. After the 3rd stage, there is a clear separation between the two groups, with the treated samples getting spatially very close to each other, indicating that most of the varnish treated samples (8 out of 10) share almost the same biochemical characteristics, mineral content, and structural arrangement after undergoing the treatment and acid challenges, remaining equally affected. The key meaning of these biplots representing the sample distribution is that all analysed spectral parameters, especially band position and DR, are strongly influenced by the chemical environment, which has a great impact on hydroxyapatite chemical bonds, and inter- and intramolecular forces of outer enamel layers.

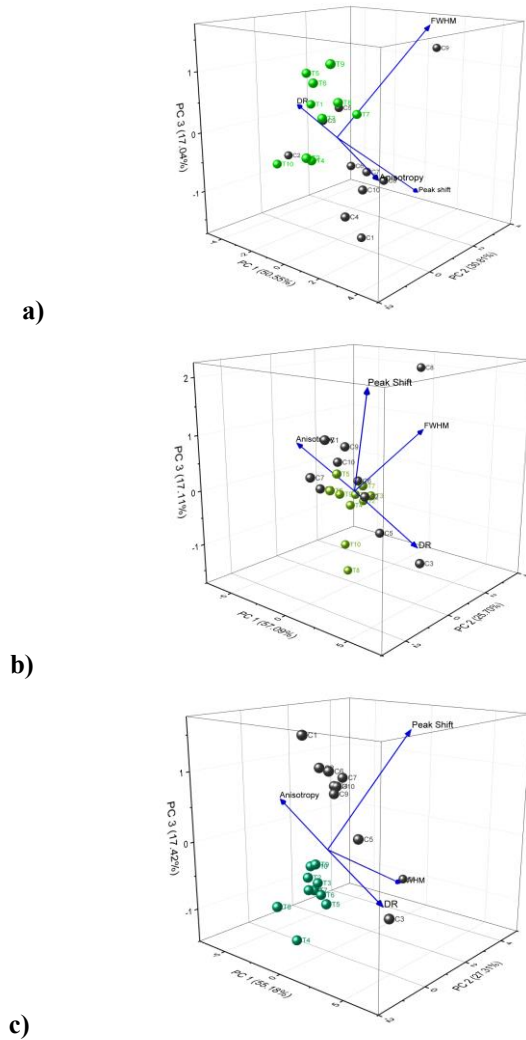


Fig. 3. 3D representation of PCA-correlation scatter biplots of analysed Raman parameter vectors (blue) and differentiation between examined control (black) and dental varnish treated (green) sample groups, corresponding to the following study phases: **a)** 1st – before any treatment; **b)** 2nd – after dental products application; **c)** 3rd – after acid exposure.

The exposed results show that PCA can distinguish and separate two different sample groups, according to the main three PC scores extracted from spectral parameter analysis, with the 1st PC score accounting the most. In this study, were chosen the first three most representative components since their summed scores represent all together the greatest majority, varying between 98% and 99,9 %, maintaining the most important information at a significant level. The exposed results confirm that Raman-PCA methods combination has a strong capability for an early

detection and evolution monitoring and assessment of dental incipient caries and demineralization/remineralization processes.

5 Conclusions and Further Work

The results obtained during this study show that after erosive acidic procedure, all dental enamel samples were strongly affected by the demineralization erosive process. The mineral loss was more obvious in the samples from the control group, meaning that the acid attack has a greater impact on the samples washed/treated only with toothpaste, opposite to the treated samples with dental fluorinated varnish (with surface adherent remineralizing capacity). This study proves that Raman spectroscopy technique combined with PCA have a strong potential to differentiate healthy sound enamel from demineralized ones, as well as control from treated samples with good sensitivity and specificity levels. Raman-PCA technique can be certainly employed to monitor evolution of dental samples during different dental product application and acid challenges, providing objective, precise, and quick results, significantly diminishing the subjectivity to human error. In future research the number of dental samples should be increased for a higher statistical significance. Moreover, PCA method could be applied directly to acquired Raman spectra to improve the extraction of the parameters to be evaluated. This way, a more precise examination of enamel de- and remineralization processes could be obtained.

Acknowledgements. The present research study is included in I4H Doctoral Program and financially supported by the Foundation of Science and Technology, identified with the Research Grant PD/BDE/143107/2018. Authors would also like to mention the contribution of the following professors: M. L. Carvalho^{1,2} and J. P. Santos^{1,2}.

References

- 1) Ko A. C.-T., Hewko M., Sowa M. G., Dong C. S., Cleghorn B. and Choo-Smith L.: Early dental caries detection using a fibre-optic coupled polarization-resolved Raman spectroscopic system, *Optics Express*, 16:9, 6274-6284 (2008).
- 2) Vargas-Koudriavtsev T., Durán-Sedó R., Sáenz-Bonilla P., Bonilla-Mora V., Guevara-Bertsch M., Jiménez-Corrales A. R., Herrera-Sancho O. A.: Effect of tooth-bleaching agents on phosphate concentration in dental enamel by means of Raman spectroscopy, *Revista Odontologica Mexicana*, 19:4, 228-235 (2015).
- 3) Ramakrishnaiah R., Rehman G., Basavarajappa S., Khuraif A. A., Durgesh B. H., Khan A. S. and Rehman I.: Applications of Raman Spectroscopy in Dentistry: Analysis of Tooth Structure, *Applied Spectroscopy Reviews*, 50:4, 332-350 (2015).
- 4) Buchwald T., Okulus Z. and Szybowicz M.: Raman spectroscopy as a tool of early dental caries detection—new insights, *Journal of Raman Spectroscopy* (2017).
- 5) Lussier F., Thibault V., Charron B., Wallace G. Q., Masson J.-F.: Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, *TrAC Trends in Analytical Chemistry*, 124 (2020).
- 6) Nunes A. and Magalhaes S.: Raman Spectroscopy Applied to Health Sciences, *IntechOpen*, 13, 275-291 (2018).

- 7) Vandenberghe P.: *Practical Raman Spectroscopy – An Introduction*, John Wiley & Sons, Ltd, ISBN 978-0-470-68319-4, Raman spectroscopy–Study and teaching (2013).
- 8) Ferraro J. R., Nakamoto K. and Brown C. W.: *Introductory Raman Spectroscopy*, Elsevier, Second Edition, ISBN: 978-0-12-254105-6 (2003).
- 9) Vo-Dinh T.: *Biomedical Photonics Handbook*, CRC Press, Boca Raton (2003).
- 10) Matousek P. and Morris M.: *Emerging Raman Applications and Techniques in Biomedical and Pharmaceutical Fields* (2010).
- 11) Monteiro M., Chasqueira F. and Pessanha S.: Raman spectroscopy in the characterisation of carious dental tissues, *Spectroscopy Europe*, 30:3, 11-14 (2018).
- 12) Kekkonen J., Finnilä M. A. J., Heikkilä J., Anttonen V., Nissinen I.: Chemical imaging of human teeth by a time-resolved Raman spectrometer based on a CMOS single-photon avalanche diode line sensor, *Analyst*, 144, 6089 (2019).
- 13) Akkus A., Akkus O., Roperto R., Lang L.: Investigation of Intra-and Inter-individual, Variations of Mineralisation in Healthy Permanent Human Enamel by Raman Spectroscopy (2016).
- 14) Castro J., Godinho J., Mata A., Silveira J. M and Pessanha S.: Study of the effects of unsupervised over-the counter whitening products on dental enamel using μ -Raman and μ -EDXRF spectroscopies (2015).
- 15) Gonzalez-Solis J. L., Martinez-Cano E., Magana-Lopez Y.: Early detection of dental fluorosis using Raman spectroscopy and principal component analysis (2014).
- 16) Barrera-Ortega C. C., Vazquez-Olmos A. R., Sato-Berru R. Y., Araiza-Tellez M. A.: Study of demineralized dental enamel treated with different fluorinated compounds by Raman spectroscopy, *J Biomed Phys Eng* 10:5, 635-644 (2020).
- 17) Mihály J., Gombás V., Afishah A., Mink J.: FT-Raman investigation of human dental enamel surfaces, 40(8), 898–902 (2009).
- 18) Bērziņš K., Sutton J. J., Loch C., et al.: Application of low-wavenumber Raman spectroscopy to the analysis of human teeth. *J Raman Spectrosc.* 1–13 (2019).
- 19) Sharma V., Rastogi S., Kumar B. K., Srinivasan A., Raychaudhuri A., Nikolajeff F., Kumar S.: Mapping the Inorganic and Proteomic Differences among Different Types of Human Teeth: A Preliminary Compositional Insight. *Biomolecules*, 10(11), 1540–(2020).
- 20) Das Gupta S., Killenberger M., Tanner T., Rieppo L., Saarakkala S., Heikkilä J. and Finnilä, M.: Mineralization of dental tissues and caries lesions detailed with Raman microspectroscopic imaging, *The Analyst* (2021).
- 21) Natarajan A. K., Fraser S. J., Swain M. V., Drummond B. K., Gordon K. C.: Raman spectroscopic characterisation of resin-infiltrated hypomineralised enamel. *Analytical and Bioanalytical Chemistry*, 407(19), 5661–5671 (2015).
- 22) Pessanha S., Silva S., Silveira J. M., Otel I., Luis H., Manteigas V., Jesus A. P., Mata A., Fonseca M.: Evaluation of the effect of fluorinated tooth bleaching products using polarized Raman microscopy and particle induced gamma-ray emission, *Spectrochim. Acta Part A: Biomol. Spectrosc.* 236 (2020) 118378.