



Grading of Melanoma Tissues by Raman MicroSpectroscopy [†]

Gianmarco Lazzini  and Mario D'Acunto  *

CNR-IBF Istituto di Biofisica, Via Moruzzi 1, I-56124 Pisa, Italy; gianmarco.lazzini@ibf.cnr.it

* Correspondence: mario.dacunto@ibf.cnr.it; Tel.: +39-0506212763

[†] Presented at the 17th International Workshop on Advanced Infrared Technology and Applications, Venice, Italy, 10–13 September 2023.

Abstract: Melanoma is one of the most aggressive forms of cancer. Early-stage diagnosis is therefore a landmark for the success of the therapies and to improve the prognosis. Raman spectroscopy represents a powerful and label-free approach for the molecular characterization of biological samples. Due to its level of detail, when applied to cancer tissues, Raman spectroscopy can help the classification of cancer-related malignant degrees. However, there is a high similarity between Raman spectra related to different cancerous tissues, which requires the use of sophisticated techniques for the treatment of Raman data. In this work, we coupled Confocal Raman Microscopy and Machine Learning techniques for the automatic classification of ex vivo melanoma tissues. In particular, we compared the performance of a PCA+LDA routine with a Random Forest Classifier. The work demonstrated excellent Machine Learning performances in classifying the tissues under investigation.

Keywords: Confocal Raman Microscopy; Machine Learning; melanoma; cancer diagnosis

1. Introduction

Melanoma is the leading cause of skin-cancer-related deaths in individuals under 30. The dramatic statistical data related to the evolution of melanoma are essentially attributable to its rapid evolution, resulting in a high frequency of cases detected at an advanced stage [1].

The gold standard for the melanoma diagnosis involves a visual examination followed by a biopsy and histopathological analysis. The weakness of this protocol is mainly represented by the qualitative inspection, which is strongly dependent on the expertise of the dermatologist. Furthermore, melanoma often shows similarities to other types of skin lesions, e.g., basal cell carcinoma or pigmented nevi. These drawbacks result in a high occurrence of false positive cases after the visual examination and high frequency of invasive and unnecessary tissue excisions [2]. For these reasons, the introduction of non-invasive, powerful and cost-effective diagnostic protocols could represent an important landmark to significantly reduce the melanoma-related morbidity.

Due to its high potential, Raman spectroscopy (RS) turns out to be one of the most studied approaches for cancer detection [3–5]. In fact, RS is able to provide rapidly molecular information at an extraordinary level of detail, allowing to differentiate apparently identical samples on the basis of their molecular differences [6]. Furthermore, RS is a label-free and non-destructive technique that is potentially suitable for in vivo measurements.

However, the Raman signal is usually very weak. In particular, especially when the Raman scattering is probed with visible radiation, the sample fluorescence could partially mask the Raman component of the measured signal. In addition, the complexity and richness of the information of the Raman spectra imposes the introduction of automatic and sophisticated protocols to find the relevant differences among spectra. In this sense, Machine Learning (ML) offers a series of powerful tools for the treatment and the interpretation of “raw” Raman experimental data. For instance, Araujo et al. [2] developed a series of ML protocols capable of distinguishing cutaneous melanoma and melanocytic naevi



Citation: Lazzini, G.; D'Acunto, M. Grading of Melanoma Tissues by Raman MicroSpectroscopy. *Eng. Proc.* **2023**, *51*, 10. <https://doi.org/10.3390/engproc2023051010>

Academic Editors: Gianluca Cadelano, Giovanni Ferrarini and Davide Moroni

Published: 26 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

from fresh cutaneous tissues with high performances (AUC 0.95), allowing the detection of a restricted spectral interval within the fingerprint region, which is relevant for the process of detection. Baria et al. [7] tested machine learning routines, i.e., Linear Discriminant Analysis and Artificial Neural Network, to analyze in vitro cell cultures with the aim of detecting BRAF and NRAS gene mutations associated with the occurrence of melanoma. The methods allowed detecting the two genes with accuracies larger than 90%, testifying to the effectiveness in diagnosing the tumor.

In line with the aforementioned investigations, in this work, we coupled Confocal Raman Microscopy and ML with the aim of detecting melanoma in ex vivo skin lesions. We tested two ML approaches for the interpretation of the spectral data: Principal Component Analysis (PCA) followed by Linear Discriminant Analysis (LDA) (PCA+LDA) and Random Forest Classifier (RFC). The study resulted in discrimination accuracies larger than 97%.

2. Materials and Methods

The samples employed for the Raman analysis were ten 5 μm thick sections of formalin fixed skin lesions, deposited on common glass slides, half of which corresponded to melanoma (malignant tissue), and the other half corresponded to compound naevus (benign tissue). The RS setup employed in this work was an HoribaTM Xplora Plus Confocal Raman Microscope equipped with a 532 nm laser working at the power of 5.6 mW. The optical system included a 100 \times objective collecting the backward scattered light, a pinhole of diameter 100 μm and a 1200 gr/mm diffraction grating, resulting in a spectral resolution of $\sim 3 \text{ cm}^{-1}$. We performed the spectral acquisition in portions of the samples characterized by an abundance of melanocytes, which are considered promising sites for melanoma detection. Each spectrum was obtained by accumulating 15 repetitions and with a single spectrum acquisition time of 1 s. The final dataset included $n_b = 757$ spectra of compound naevi (benign lesions) and $n_m = 739$ spectra of melanoma (malignant lesions). We analyzed the spectral region of wave numbers between 400 and 1800 cm^{-1} (fingerprint region).

The raw dataset was subjected to a baseline subtraction according to the algorithm proposed by Zhao et al. [8]. This algorithm is an iterative procedure based on subtracting from the original spectra a polynomial fit in order to obtain a smoothed graph attributed to the baseline contribution. This algorithm required as input parameters the polynomial order, equal to 2, and the number of iterations, equal to 150. Finally, the spectra were normalized with the Min–Max method [9] and rescaled by setting the mean to 0 and the variance to 1.

In this work, we focused our attention on the problem of distinguishing two classes, i.e., “benign” and “malignant”. To this aim, we adopted two ML protocols: a first protocol was represented by Principal Component Analysis (PCA) followed by Linear Discriminant Analysis (LDA) (PCA+LDA). In particular, we excluded all the principal components except the first 25 to obtain a sum of the corresponding explained variance ratios larger than 0.95 [10]. To make predictions, we considered only the first LDA component. We then compared the performances of the PCA+LDA model to a second procedure based on the Random Forest Classifier (RFC). In this case, we set the number of trees in the forest to 1100 to minimize the Out-Of-Bag Error [11].

The performances of the aforementioned ML models were assessed through 10-Fold Cross Validation [12]. Each fold was generated by randomly splitting the dataset into a training set and a test set by following the proportion 80/20, respectively. We quantified the ML performances in terms of the accuracy, the sensitivity and the specificity averaged over the 10 folds, which were calculated by considering “malignant” as the positive class.

3. Results

Figure 1a shows the mean Raman spectra associated to the benign and the malignant skin tissues. A qualitative interpretation of the Raman spectra allowed identifying a spectral region located between 400 and $\sim 1200 \text{ cm}^{-1}$ and another spectral interval between ~ 1650 and $\sim 1800 \text{ cm}^{-1}$, in which the Raman signal associated to the malignant tissues tends

to be larger than the Raman signal of the benign tissues. We attributed these bands to chemical compounds, e.g., phospholipids ($427, 529, 862$ and 1710 cm^{-1}) and nucleic acids ($469, 571, 625, 685$ and 790 cm^{-1}) [13], whose higher concentration in malignant tissues is associated to the uncontrolled cellular proliferation characterizing the cancer. The remaining spectral region, located between ~ 1200 and $\sim 1650\text{ cm}^{-1}$, tends to show the opposite trend, i.e., a decreased signal of the malignant tissue with respect to the benign tissues. Despite vibrational modes associated to cell proteins and nucleic acids falling within these spectral bands, an attribution to this chemical compounds would contradict the aforementioned trend. Another possible interpretation of this Raman band can be found in melanin. In fact, former studies pointed out that despite melanoma being characterized by an uncontrolled proliferation of melanocytes, i.e., the cells deputize to the production of melanin, this form of cancer is not necessarily associated to an increased concentration of melanin [2]. This fact could explain the large standard deviations, represented by the shaded areas, associated to the mean spectra in this spectral interval.

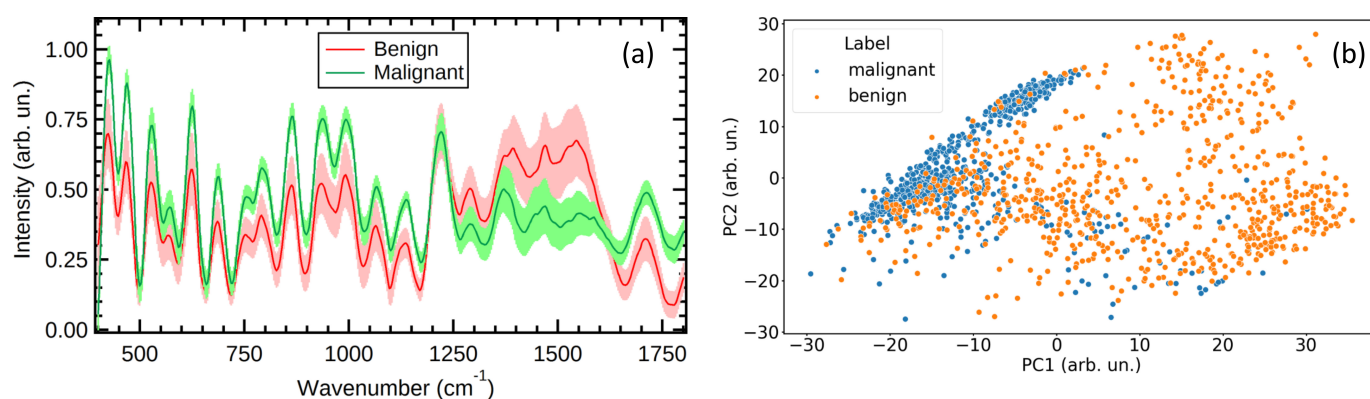


Figure 1. (a) Averaged Raman spectra of benign (red) and malignant skin lesions (green). The shaded areas represent the standard deviation; (b) score plot of the first two principal components.

In Figure 1b, the score plot, showing the first two principal components, highlights how the malignant tissues result in a dense cluster of points, while the benign counterparts appear as a more sparse cloud. In addition, except for the superposition of a relatively small number of points, the classes “benign” and “malignant” appear to be well separated and simply connected domains. Furthermore, the two classes can be roughly separated with a straight line. This trend reflects the high performances of the PCA+LDA routine, resulting in 0.96 ± 0.01 accuracy, 0.98 ± 0.01 sensitivity and 0.94 ± 0.02 specificity. This demonstrated the effectiveness of the PCA+LDA routine in melanoma detection.

The partial superposition between the “benign” and the “malignant” classes in the score plot pushed us to compare the performance of PCA+LDA to non-linear models. In this sense, the Random Forest Classifier (RFC) led to 0.97 ± 0.01 accuracy, 0.97 ± 0.01 sensitivity and 0.97 ± 0.01 specificity. The performances of the two ML protocols are comparable with a slight improvement of the sensitivity provided by RFC. The aforementioned performances are comparable with results of former investigations. For instance, Lima et al. [14] proposed a study performed either in in vivo or ex vivo skin tissues that aimed at distinguishing non-melanoma skin cancer from normal tissues from spectra obtained through a Raman spectrometer working in the infrared range. A discriminant analysis approach applied on the resulting experimental data allowed distinguishing normal tissues from lesions with accuracies reaching 100% for in situ measurements.

4. Conclusions

In this pilot study, we investigated the possibility of coupling Confocal Raman Microscopy and Machine Learning tools for the melanoma detection from ex vivo tissues. To this aim, we employed a commercial Confocal Raman Microscopy to obtain Raman spectra from micrometric sections of skin lesions deposited on common glass slides. We tested two

different Machine Learning protocols, consisting of PCA+LDA and Random Forest Classifier (RFC). Both the ML protocols resulted in excellent performances with values of 10-fold averaged accuracy, sensitivity and specificity larger than 95%. However, in order to have further proof about the effectiveness and the robustness of the Machine Learning discussed in this work, efforts have to be devoted to increasing the number of patients involved. In addition, the introduction of other types of skin cancer in the statistical sample could represent another interesting solution to test the performances of these models. Finally, the individuation of specific spectral intervals, which is considered relevant for the process of distinction, is a fundamental step in view of future technological implementations on devices suitable for non-invasive and in situ diagnosis.

Author Contributions: The authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Regione Toscana through the TELEMO Project under 146 Grant Ricerca Salute 2018.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Azienda USL Toscana Nord Ovest U.O.C. Dermatologia—Livorno Mod. PROTOCOLLO STUDIO PILOTA OSSERVAZIONALE B11 Vers_20160118.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are unavailable due to restrictions by the ethical committee.

Acknowledgments: The authors thanks Daniela Massi for supporting samples. The authors wish to thank Tiziana Amendola, Marco Carloni, Rosa Anna Cascone, Rosa Claps, Davide Nardi, Claudia Neri, Gianluca Presciuttini and Alessandro Puntoni for their useful support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Larson, A.R.; Konat, E.; Alani, R.M. Melanoma biomarkers: Current status and vision for the future. *Nat. Clin. Pract. Oncol.* **2009**, *6*, 105–117. [[CrossRef](#)] [[PubMed](#)]
2. Araújo, D.C.; Veloso, A.A.; de Oliveira Filho, R.S.; Giraud, M.N.; Raniero, L.J.; Ferreira, L.M.; Bitar, R.A. Finding reduced Raman spectroscopy fingerprint of skin samples for melanoma diagnosis through machine learning. *Artif. Intell. Med.* **2021**, *120*, 102161. [[CrossRef](#)] [[PubMed](#)]
3. Kim, S.; Kim, T.G.; Lee, S.H.; Kim, W.; Bang, A.; Moon, S.W.; Song, J.; Shin, J.H.; Yu, J.S.; Choi, S. Label-free surface-enhanced Raman spectroscopy biosensor for on-site breast cancer detection using human tears. *ACS Appl. Mater. Interfaces* **2020**, *12*, 7897–7904. [[CrossRef](#)] [[PubMed](#)]
4. Shaikh, R.; Daniel, A.; Lyng, F.M. Raman Spectroscopy for Early Detection of Cervical Cancer, a Global Women's Health Issue—A Review. *Molecules* **2023**, *28*, 2502. [[CrossRef](#)] [[PubMed](#)]
5. Guo, L.; Li, Y.; Huang, F.; Dong, J.; Li, F.; Yang, X.; Zhu, S.; Yang, M. Identification and analysis of serum samples by surface-enhanced Raman spectroscopy combined with characteristic ratio method and PCA for gastric cancer detection. *J. Innov. Opt. Health Sci.* **2019**, *12*, 1950003. [[CrossRef](#)]
6. Fox, S.A.; Shanblatt, A.A.; Beckman, H.; Strasswimmer, J.; Terentis, A.C. Raman spectroscopy differentiates squamous cell carcinoma (SCC) from normal skin following treatment with a high-powered CO₂ laser. *Lasers Surg. Med.* **2014**, *46*, 757–772. [[CrossRef](#)] [[PubMed](#)]
7. Baria, E.; Cicchi, R.; Malentacchi, F.; Mancini, I.; Pinzani, P.; Pazzagli, M.; Pavone, F.S. Supervised learning methods for the recognition of melanoma cell lines through the analysis of their Raman spectra. *J. Biophotonics* **2021**, *14*, 202000365. [[CrossRef](#)] [[PubMed](#)]
8. Zhao, J.; Lui, H.; McLean, D.I.; Zeng, H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Appl. Spectrosc.* **2007**, *61*, 1225–1232. [[CrossRef](#)] [[PubMed](#)]
9. Gautam, R.; Vanga, S.; Ariese, F.; Umopathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 1–38. [[CrossRef](#)]
10. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [[CrossRef](#)] [[PubMed](#)]
11. Fawagreh, K.; Gaber, M.M.; Elyan, E. Random forests: From early developments to recent advancements. *Syst. Sci. Control. Eng. Open Access J.* **2014**, *2*, 602–609. [[CrossRef](#)]
12. Stone, M. Cross-validation: A review. *Stat. J. Theor. Appl. Stat.* **1978**, *9*, 127–139.

13. Movasaghi, Z.; Rehman, S.; Rehman, I.U. Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **2007**, *42*, 493–541. [[CrossRef](#)]
14. Lima, A.M.F.; Daniel, C.R.; Navarro, R.S.; Bodanese, B.; Pasqualucci, C.A.; Pacheco, M.T.T.; Zângaro, R.A.; Silveira, L., Jr. Discrimination of non-melanoma skin cancer and keratosis from normal skin tissue in vivo and ex vivo by Raman spectroscopy. *Vib. Spectrosc.* **2019**, *100*, 131–141. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.