engineering proceedings

MDPI

Proceeding Paper

Alzheimer Disease Detection from Raman Spectroscopy of the Cerebrospinal Fluid via Topological Machine Learning [†]

Francesco Conti [1,2], Martina Banchelli [3], Valentina Bessi [4], Cristina Cecchi [5], Fabrizio Chiti [5], Sara Colantonio [1], Cristiano D'Andrea [3], Marella de Angelis [3], Davide Moroni [1], Benedetta Nacmias [4,6], Maria Antonietta Pascali [1,*], Sandro Sorbi [4,6] and Paolo Matteini [3,*]

[1] Institute of Information Science and Technologies "A. Faedo", National Research Council, Via G. Moruzzi 1, 56124 Pisa, Italy; francesco.conti@phd.unipi.it (F.C.)
[2] Department of Mathematics, University of Pisa, Largo B. Pontecorvo, 56127 Pisa, Italy
[3] Institute of Applied Physics "N. Carrara", National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy
[4] Department of Neuroscience, Psychology, Drug Research and Child Health, University of Florence, 50134 Florence, Italy
[5] Department of Clinical and Experimental Biomedical Sciences "Mario Serio", University of Florence, 50134 Florence, Italy
[6] IRCCS Fondazione Don Carlo Gnocchi, 50143 Florence, Italy
[*] Correspondence: maria.antonietta.pascali@isti.cnr.it (M.A.P.); p.matteini@ifac.cnr.it (P.M.)

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

Abstract: The cerebrospinal fluid (CSF) of 19 subjects who received a clinical diagnosis of Alzheimer's disease (AD) as well as of 5 pathological controls was collected and analyzed by Raman spectroscopy (RS). We investigated whether the raw and preprocessed Raman spectra could be used to distinguish AD from controls. First, we applied standard Machine Learning (ML) methods obtaining unsatisfactory results. Then, we applied ML to a set of topological descriptors extracted from raw spectra, achieving a very good classification accuracy (>87%). Although our results are preliminary, they indicate that RS and topological analysis may provide an effective combination to confirm or disprove a clinical diagnosis of AD. The next steps include enlarging the dataset of CSF samples to validate the proposed method better and, possibly, to investigate whether topological data analysis could support the characterization of AD subtypes.

Keywords: topological data analysis; machine learning; Raman spectroscopy; cerebrospinal fluid; Alzheimer disease

Citation: Conti, F.; Banchelli, M.; Bessi, V.; Cecchi, C.; Chiti, F.; Colantonio, S.; D'Andrea, C.; de Angelis, M.; Moroni, D.; Nacmias, B.; et al. Alzheimer Disease Detection from Raman Spectroscopy of the Cerebrospinal Fluid via Topological Machine Learning. *Eng. Proc.* **2023**, *51*, 14. https://doi.org/10.3390/engproc2023051014

Academic Editors: Gianluca Cadelano and Giovanni Ferrarini

Published: 27 October 2023

1. Introduction

Alzheimer's disease (AD) affects tens of millions of people worldwide, as it is the most common neurodegenerative disease. At present, the clinical diagnosis of AD requires a series of neurological examinations, while the definitive diagnosis is possible only after the patient's death. Therefore, there is a need to improve the accuracy of clinical diagnosis with innovative and cost-effective approaches. Raman spectroscopy (RS) represents a fast, efficient, non-invasive diagnostic tool [1]. The high-precision detection of RS is expected to reduce or replace other AD diagnostic tests. Recently, RS techniques demonstrated significant potential in identifying AD by detecting specific biomarkers in body fluids [2]. Given the increasing number of RS studies, a systematic evaluation of the accuracy of RS in the diagnosis of AD was already performed, showing that RS is an effective and accurate tool for diagnosing AD, though it cannot rule out the possibility of misdiagnosis [3]. Recently, RS of tissue samples has been coupled with Topological Machine Learning (TML) for bone cancer grading [4], showing the feasibility of a topological approach for multi-label classification. The detection of CSF biomarkers is one of the diagnostic criteria for AD [5] because CSF is more sensitive than blood or other biofluids in diagnosing AD. Therefore, RS can be used as an effective tool to analyze CSF samples, as shown previously [6,7].

Here, we propose a novel method based on the collection of the vibrational Raman fingerprint of the proteomic content of cerebrospinal fluid (CSF) and on the topological machine learning analysis of the Raman spectra in order to support the AD diagnosis. The achieved results encourage continuous investigation of topological machine learning tools to understand whether looking at Raman spectra of CSF with the topological lens could also help to characterize AD subtypes.

2. Population Study and Data Acquisition

The study population is made up of 24 patients, enrolled in the framework of the Bando Salute 2018 PRAMA project ("Proteomics, RAdiomics and Machine learning-integrated strategy for precision medicine for Alzheimer's"), co-funded by the Tuscany Region, with the approval of the Institutional Ethics Committee of the Careggi University Hospital Area Vasta Centro (ref. number 17918_bio). All of them showed pathological symptoms: the majority of them, 19 subjects, were diagnosed with AD, while the others were considered as controls (noAD), even if diagnosed with other neurological conditions: one with vascular dementia, three with hydrocephalus and one with multiple sclerosis. The CSF samples were collected by lumbar puncture, then immediately centrifuged at $200\times g$ for 1 min, 20 °C and stored at −80 °C until analysis [8,9]. On the day of analysis, CSF samples were thawed and centrifuged again at $4000\times g$ for 10 min at 4 °C. The pellet was separated from the supernatant and further used for the analyses. A 2 µL drop of the pellet was deposited onto a gold mirror support (ME1S-M01; Thorlabs, Inc., Newton, NJ, USA), followed by air drying for 30 min and acquisition of Raman spectra from the outer ring of the dried drop. A set of five spectra were collected for each drop-casted sample by using a micro-Raman spectrometer (Horiba, France) in back-scattering configuration, equipped with a laser excitation source tuned at 785 nm (40 mW power, 20 s integration time, 10 accumulations) and a Peltier cooled CCD detector. In some cases, the same procedure was replicated two or three times; it resulted in a dataset of 30 acquisitions of RS, 22 belonging to the AD class and 8 to the noAD class.

3. Methods

After the spectra are acquired, the data enter the following pipeline to return the final predictive model with classification accuracy. For each patient, the average of the five acquisitions of the raw Raman spectrum is computed. The following transformations are then applied to the RS: Fourier Transform (FT), Welch Transform and AutoCorrelation. The pipeline is applied individually on the original spectra and on each of the transformations listed above. These computations are performed using the Python package SciPy [10]. Then, the spectra are processed by TML (see [11] for more information). The pipeline performs a lower star filtration to extract the Persistence Diagrams (PDs). Since the data is a 1D spectrum, the only non-trivial homology group is H_0. The PD is vectorized using the following vectorization methods: Persistence Image [12] with parameters $\sigma \in \{0.1, 1, 10\}$, $n \in \{5, 10, 25\}$, Persistence Landscape [13], Persistence Silhouette [14] and Betti curve [15] all with parameters $n \in \{25, 50, 75, 100\}$. Finally, these vectors enter one of the following Machine Learning (ML) classifiers: Support Vector Classifier, Random Forest Classifierand Ridge Classifier. The validation scheme of the pipeline is the Leave One Patient Out cross-validation (LOPO). This scheme is a generalization of the leave one out cross-validation [16], with the difference that all data from the same patient are recursively left in the validation set, instead of a single piece of data. This avoids biased high accuracy due to the similarity of the data from the same patient that may otherwise be found both in the training and validation sets.

Figure 1 shows eight Persistence images, two for each combination AD-noAD and RS-FT. It is interesting to note that in the PIs coming from the spectra, the pattern seems more chaotic between the two classes, while in the PIs coming from the FT there is a clearer division. More in detail, the lit pixels in the PIs of class noAD have a more elongated shape than those of the AD class. This corroborates the results achieved in Section 4. In Figure 2, eight Persistence silhouettes are shown in the same fashion as in Figure 1. Again, there is a clearer division for the PSs coming from the FT. A peak at the tail end of the signal is present for the noAD class.

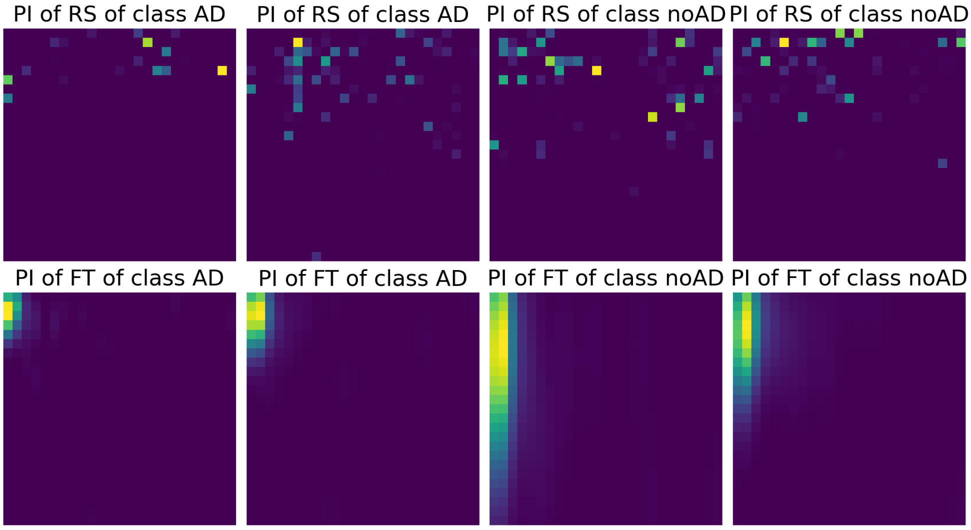

Figure 1. First row: Two persistence images (PI) of the spectra for class AD and two for class noAD. Second row: two PI of the FT for class AD and for class noAD.

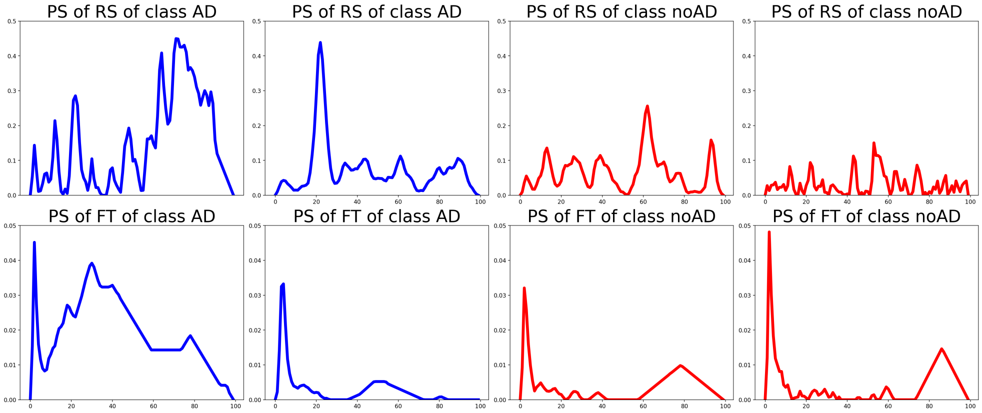

Figure 2. First row: Two persistence silhouettes (PS) of the spectra for class AD and for class noAD. Second row: two PS of the FT for class AD and for class noAD.

4. Results

The results obtained from the pipeline on each of the transformations are shown in Table 1. The FT is the one that obtains, clearly, the better results. We used as baseline accuracy the value of 0.733 due to the imbalance in classes (i.e., the accuracy achieved by the classifier assigning always the most frequent label to any sample).

Table 1. Accuracy result of the TML pipeline for different inputs.

Method	Accuracy	Vectorization and Classifier
H_0	0.833	PI and Ridge
Fourier Transform (FT)	0.875	PS and SVC
Welch Transform	0.763	PI and SVC
AutoCorrelation	0.667	PI and Ridge

It is worth pointing out that even standard preprocessing applied to Raman spectra could lead to a classification accuracy below the baseline accuracy. This is probably due to the fact that in our dataset, the signal-to-noise ratio is quite low. On the other hand, the accuracy value (>83%) achieved by extracting H_0 features from raw spectra is in line with results of [6], while results achieved by extracting topological features after performing the FT are even better (87.5%).

5. Discussion

The results described above support strongly that RS and topological analysis together may provide an effective combination to confirm or disprove a clinical diagnosis of AD. Also, the training of the classification ML model trained on the topological features extracted from the spectra acquired on an CSF sample does not need the choice or set of any parameters; hence, the proposed methodology may evolve in automatic support to AD diagnosis, which could be easily embedded in a commercial platform of RS. The above considerations are preliminary and require further confirmation from the statistical viewpoint. From this perspective, the next steps include enlarging the dataset of CSF samples to validate the proposed method better and, possibly, to understand whether topological machine learning could support the characterization of AD subtypes.

Author Contributions: Conceptualization, all authors; biological sample processing, C.C. and F.C. (Fabrizio Chiti); classification software, F.C. (Francesco Conti) and M.A.P.; validation, F.C. (Francesco Conti), M.A.P. and D.M.; resources, M.B., V.B., C.D., M.d.A., B.N., S.S. and P.M.; data curation, F.C. (Fabrizio Chiti), M.d.A. and P.M.; writing—original draft preparation, M.A.P. and F.C. (Francesco Conti); writing—review and editing, M.A.P., F.C. (Francesco Conti), D.M. and P.M.; visualization, F.C. (Francesco Conti); supervision, P.M., D.M., S.C. and M.A.P.; funding acquisition, S.C. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded in the framework of the Bando Salute 2018 PRAMA project co-funded by the Tuscany Region.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of the Careggi University Hospital Area Vasta Centro (protocol code 17918_bio, approved on 6 October 2020).

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

Data Availability Statement: Dataset used in this work is not public due to privacy restriction, as stated in the PRAMA PROJECT protocol (reference number 17918_bio) approved by the Institutional Ethics Committee of the Careggi University Hospital Area Vasta Centro.

Acknowledgments: M.B., C.D., M.d.A. and P.M. acknowledge Andrea Donati for his expert technical assistance.

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

References

1. Eberhardt, K.; Stiebing, C.; Matthäus, C.; Schmitt, M.; Popp, J. Advantages and limitations of Raman spectroscopy for molecular diagnostics: An update. *Expert Rev. Mol. Diagn.* **2015**, *15*, 773–787. [CrossRef] [PubMed]
2. Polykretis, P.; Banchelli, M.; D'Andrea, C.; de Angelis, M.; Matteini, P. Raman Spectroscopy Techniques for the Investigation and Diagnosis of Alzheimer's Disease. *FBS* **2022**, *14*, 22. [CrossRef]
3. Xu, Y.; Pan, X.; Li, H.; Cao, Q.; Xu, F.; Zhang, J. Accuracy of Raman spectroscopy in the diagnosis of Alzheimer's disease. *Front. Psychiatry* **2023**, *14*, 1112615. [CrossRef]
4. Conti, F.; D'Acunto, M.; Caudai, C.; Colantonio, S.; Gaeta, R.; Moroni, D.; Pascali, M.A. Raman spectroscopy and topological machine learning for cancer grading. *Sci. Rep.* **2023**, *13*, 7282. [CrossRef]
5. Blennow, K.; Zetterberg, H. Biomarkers for Alzheimer's disease: Current status and prospects for the future. *J. Intern. Med.* **2018**, *284*, 643–663. [CrossRef] [PubMed]
6. Ryzhikova, E.; Ralbovsky, N.M.; Sikirzhytski, V.; Kazakov, O.; Halamkova, L.; Quinn, J.; Zimmerman, E.A.; Lednev, I.K. Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid. *Spectrochim. Acta Part A* **2021**, *248*, 119188. [CrossRef]

7. Huang, C.C.; Isidoro, C. Raman Spectrometric Detection Methods for Early and Non-Invasive Diagnosis of Alzheimer's Disease. *J. Alzheimer's Dis.* **2017**, *57*, 1145–1156. [CrossRef]
8. Tashjian, R.S.; Vinters, H.V.; Yong, W.H. Biobanking of Cerebrospinal Fluid. In *Biobanking: Methods and Protocols*; Yong, W.H., Ed.; Springer: New York, NY, USA, 2019; pp. 107–114. [CrossRef]
9. Vanderstichele, H.; Bibl, M.; Engelborghs, S.; Le Bastard, N.; Lewczuk, P.; Molinuevo, J.L.; Parnetti, L.; Perret-Liaudet, A.; Shaw, L.M.; Teunissen, C.; et al. Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis. *Alzheimer's Dement.* **2012**, *8*, 65–73. [CrossRef]
10. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
11. Conti, F.; Moroni, D.; Pascali, M.A. A Topological Machine Learning Pipeline for Classification. *Mathematics* **2022**, *10*, 3086. [CrossRef]
12. Adams, H.; Emerson, T.; Kirby, M.; Neville, R.; Peterson, C.; Shipman, P.; Chepushtanova, S.; Hanson, E.; Motta, F.; Ziegelmeier, L. Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **2017**, *18*, 1–35.
13. Bubenik, P. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **2015**, *16*, 77–102.
14. Chazal, F.; Fasy, B.T.; Lecci, F.; Rinaldo, A.; Wasserman, L. Stochastic convergence of persistence landscapes and silhouettes. In Proceedings of the Thirtieth Annual Symposium on Computational Geometry, Kyoto, Japan, 8–11 June 2014; pp. 474–483.
15. Umeda, Y. Time series classification via topological data analysis. *Inf. Media Technol.* **2017**, *12*, 228–239. [CrossRef]
16. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2.

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.