Synthesis : Interpreting Raman spectroscopy towards diagnostic purposes : an explainable deep-learning based approach

Nina Singlan

1 General context of the thesis

This work is part of a collaboration between the LABION (Laboratorio di Nanomedicina e Biofotonica Clinica) research group of the Don Gnocchi Foundation and the MIND (Models in decision making and data analysis) research laboratory of the University of Milan-Bicocca. Aims of this collaboration are the study and design of Machine Learning and Deep Learning algorithms for the analysis, characterization and classification of Raman spectra from saliva samples. The ultimate goal of this project is to allow the development of an innovative, rapid and non-invasive screening methodology for the diagnosis of neurodegenerative and COVID diseases.

The first thing that was done to achieve this goal was to create a deep learning model, called a classifier, which is able to tell whether a patient is infected with the disease or not, to classify him or her as healthy or sick. This classifier is the starting point for the work presented here. Indeed, having a model capable of telling whether a patient is ill or not, even with a very good accuracy, is not enough. Taking the Covid as a case study, the question this work aims to answer is: is the model reliable? To answer this question, the key idea is to understand the classifier: how does it classify? What part of the input data does it focus on to give its answer? These questions are at the heart of Explicable Artificial Intelligence. Thus, in simple terms, the aim of this thesis is to propose an Explicable Artificial Intelligence approach to understanding the classification, in the case of Covid, of Raman spectra.

2 Biological and medical context

2.1 Introduction to Raman spectroscopy

Raman Spectroscopy is a non-destructive chemical analysis technique which provides detailed information about chemical structure, phase and polymorphy, crystallinity and molecular interactions of materials. It is based upon the interaction of light with the chemical bonds within a material. More precisely it is based on Raman Effect, that is to say frequency of a small fraction of scattered radiation is different from frequency of monochromatic incident radiation. This effect is also called Raman scattering. Absorption of a photon excites the molecule to the imaginary state and re-emission leads to Raman or Rayleigh scattering. As it is shown in Fig. 1, there are three different case, in all of them the final state has the same electronic energy as the starting state but is higher in

vibrational energy in the case of Stokes Raman scattering (shown in orange in the figure), lower in the case of anti-Stokes Raman scattering (shown in blue in the figure) or the same in the case of Rayleigh scattering (shown in green in the figure).

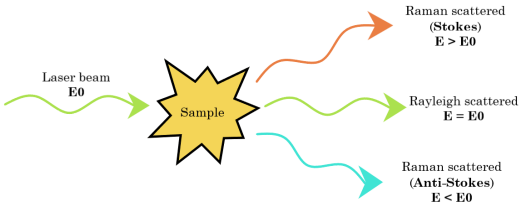

Figure 1: Basic scheme of the Raman scattering. The original vibrational energy is denoted by E0 while the resulting one is denoted by E.

The spectrum of the Raman-scattered light depends on the molecular constituents present and their state, allowing the spectrum to be used for material identification and analysis. Thus, Raman spectroscopy is a versatile method used to analyze a wide range of materials, including gases, liquids, and solids. Highly complex materials such as biological organisms and human tissue can also be analyzed by Raman spectroscopy.

A Raman spectrum features a number of peaks, showing the intensity and wavelength position of the Raman scattered light. Each peak corresponds to a specific molecular bond vibration.
The key features of a Raman spectra are :

- **The Raman shifts and relative intensities of all of the Raman bands of the material:** Basically, the Raman shift is the energy difference between the incident (laser) light and the scattered (detected) light. With these, the material can be identified.
- **Individual band changes:** A band may shift, narrow or broaden, or vary in intensity. These changes can reveal information about stresses in the sample, variations in crystallinity, and the amount of material respectively.
- **Variations in spectra with position on the sample:** This will reveal changes in the uniformity (homogeneity) of the material. You can analyse at several arbitrary points, or systematically measure an array of points (enabling the production of images of composition, stress, crystallinity, etc.)

In practice, the spectra in this study are acquired using the Surface-enhanced Raman Spectroscopy (SERS) version of Raman spectroscopy which is characterised by the addition of a particular nano-structure (in this case, an aluminium substrate) which can enhance the Raman signal.

As mentioned earlier, the main aim of this thesis is to understand the classification of Raman spectra of salivary fingerprints. This classification is done according to three different classes: the patient is either Covid-positive, Covid-negative (meaning that he/she has already been infected, but is cured), or a control patient (he/she does not have Covid and has never been infected).Thus, for the sake of clarity, the Raman spectra shown as examples in Fig. 2 are taken directly from the Covid dataset which will be used throughout this work.

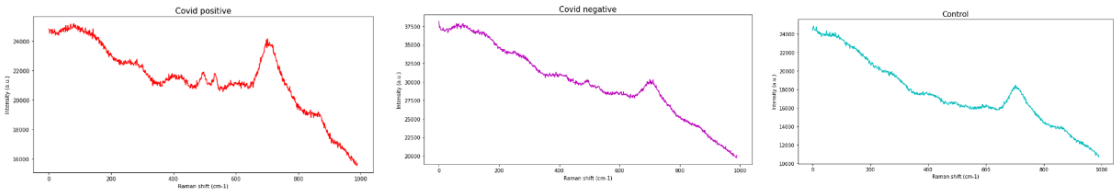

Figure 2: Example of Raman spectra. Each one is taken from a different classes. The first one come from the Covid Positive (Covid+) class, the second one came from the Covid Negative (Covid-) one and the third one is from the Control (CTRL) one.

2.2 Medical application

The Raman spectroscopy technique is capable of providing details on chemical composition, molecular structure, and molecular interactions in cells and tissues. Thus, changes in the tissues leaded by a disease could be reflected in the spectra. If theses changes are typical of a certain disease, the Raman spectra can be used for diagnostic purposes.

In the scientific literature, there are already many studies concerning the application of Raman spectroscopy as a clinical diagnostic tool. This technique can be based on tissue analysis or on bio-fluid analysis. Thus, Raman spectroscopy offers a new method of non-invasive diagnosis. In this study, samples are taken from saliva, which is a complex biofluid, being composed by several different molecules, among which proteins, metabolites, carbohydrates, nucleic acids, and hormones, in an overall aqueous solution. Not only presence of these molecules can lead to potential hints for the identification of a pathological state, but also their concentrations and variations in concentration, as well as their modifications, interactions and environments, can give insights for the disease progression and response to specific therapies. Moreover, these molecules are represented in pathological state in specific patterns that involve several immunological and physiological biomarkers, with the clear difficulty of the overall detection of all the associated levels.

3 Computer science context

3.1 Motivations

As can be noticed in Fig. 2, it is not easy to distinguish between a Covid-infected patient, a cured patient and a patient who has never been infected. Data analysis and computational techniques can be applied to the spectral data to unveil new insights in the sample characterization as well as to enable the discrimination of different classes and the identification of patterns and structures in the analyzed targets. Machine Learning (ML) techniques have already shown its ability to decode the Raman Spectra. Thus, Raman spectroscopy allied with ML techniques promises a new, rapid and non-invasive way to diagnose patients. This ability has led researchers to highlight that Convolutional Neural Networks (CNNs) outperform other ML methods.

The collaboration in which this work take place has already allow the construction of an efficient model. More precisely, a CNN which obtain an accuracy in the range 89-92%. In order to really benefit from this classifier, it is necessary to focus on understanding it.

3.2 Problem statement

In the past decades, Artificial Intelligence (AI), and more specifically, ML researchers have focus their effort on the results. The point was to obtain always better accuracy in record time. However, these days, obtaining good results is not enough anymore. This observation is even more true when ML is used for medical purposes. Indeed, a diagnostic method must be reliable, false positives and false negatives could lead to medical errors which, depending on the case, can be serious for the patient.

Even if the CNN constructed by the laboratory achieves good results, it does not allow the understanding and the interpretation of the classification. Indeed, one of the critical drawbacks of the CNN approach is the lack of interpretation, it is regarded as a black box. Thus, the purpose of this study is to suggest an interpretation of the CNN learning mechanism proposed in the previous work. This interpretation will be performed using Explainable Artificial Intelligence techniques and more specifically a technique originally designed for convolutional neural networks classifying images, called Class Activation Mapping.

4 Guideline

To be a good candidate, the chosen method must have certain characteristics. Indeed, considering that the overall project aims at developing an innovative, rapid and non-invasive screening methodology for the diagnosis of various diseases, the approach presented in this thesis must go beyond the understanding of the specific model created for the diagnosis of covid, it must be applicable to other models created for other diseases. In fact, the method chosen must be :

1. **Relevant for Convolutional Neural Networks :** The convolutional architecture is specific and, therefore, not all explicable artificial intelligence techniques are applicable to it, or at least not yet. Thus, in order to find a good candidate, the search was carried out in the restricted domain of the explicable convolutional neural network. But, there are many different techniques applicable to these networks. It is therefore necessary to restrict the search domain again to find a good approach.

2. **Usable on Raman spectra :** Of course, the model used is a Convolutional Neural Network but it is slightly different from the majority of them. Indeed, generally, these networks are two-dimensional and designed to process images. In the case of Raman spectra classification, the network is one-dimensional and works on spectral data. This is the second particularity that must be taken into account.

3. **Easily portable :** As mentioned a few lines above, the final aim of this work is to propose a method that can be used on other case studies. When one says other cases of study, one probably means slightly different models. These differences in models mean that the proposed method must be "portable". In computing, **portability** is defined as the ability to use the same software in different environments. Saying that the method must be portable means that the method must be usable on different models with a minimum of changes. Thus, a (very) good candidate must also satisfy this property.

4. **Preserve model :** In an ideal scenario, the method proposed in this thesis should preserve the performance of the model and be applicable on a pre-trained one.

Basically, this means that the approach should not require any modification of the model. Indeed, making a modification to make it more explainable is problematic, it induces a re-training, which is time consuming, and, moreover, it may be irrelevant. Indeed, the objective is to explain the results obtained by a model, but by modifying it, the risk is to obtain different results and, even if the results remain unchanged, the explanation may not be appropriate. If, for example, the modification involves the removal of certain layers, their roles will not be taken into account in the explanation. Yet they must have a role. Thus, to be a "perfect" method, it must satisfy this property.

These characteristics form the guideline for this thesis, and this is what needs to be kept in mind in order to find a good approach.

5 Class Activation Mapping

By restricting the research area to a method suitable for Convolutional Neural Networks and spectral data, several methods were found. However, the one that seemed to be the most interesting was the Class Activation Mapping. Indeed, this method has a number of advantages. First of all, it provides a method to visualise the most important variables for classification. This is particularly interesting because this visual representation can allow a specialist to analyse the results and thus provide a better level of understanding of the classification. To achieve this visualisation, this method uses the last convolutional layer, which is also an advantage because, although not all Convolutional Neural Networks have exactly the same architecture and layers, they all have convolutional layers by definition. But one of the main strengths of this method is that it has already been applied to spectral data, and more particularly to Raman spectra. This existing application confirms that this method is suitable for Raman spectra. In addition, the results of this application can be used as a basis to confirm the methodology used on Covid. However, even if this method satisfies a number of the characteristics specified above, it has a major flaw. Indeed, this technique requires a particular architecture : after the last convolutional layer, a Global Average Pooling layer is expected. Unfortunately, the model used in this study is not built with this architecture. Thus, to apply this method, the network must be modified and re-trained.

6 Gradient Class Activation Mapping

This weakness is the reason why a variant of Class Activation Mapping, called Gradient Class Activation Mapping has been introduced. Indeed, this method corrects this problem by being usable on all Convolutional Neural Networks. This variant retains most of the advantages of Class Activation Mapping. Unfortunately, however, it has never been applied to spectral data of any kind. To overcome this problem, Gradient Class Activation Mapping is first applied on the same dataset as the one used for Class Activation Mapping. By comparing the results obtained in each case, the method can be verified on spectral data. One of the main advantages of this method is that it can be applied to very different networks working on very different tasks. For example, the method is applied to image classification, but also to image captioning and visual question answering.