
Faculty of Science

Faculty Publications

Raman microspectroscopy and machine learning for use in identifying radiation-induced lung toxicity

Ali-Adeeb, R. N., Shreeves, P., Deng, X., Milligan, K., Brolo, A. G., Lum, J. J., . . . Jirasek, A.

2022

© 2022 Ramie N. Ali-Adeeb et al. This is an open access article distributed under the terms of the Creative Commons Attribution License.

<http://creativecommons.org/licenses/by/4.0/>

This article was originally published at:

<https://doi.org/10.1371/journal.pone.0279739>

Citation for this paper:

Ali-Adeeb, R. N., Shreeves, P., Deng, X., Milligan, K., Brolo, A. G., Lum, J. J., . . . Jirasek, A. (2022). "Raman microspectroscopy and machine learning for use in identifying radiation-induced lung toxicity." *PLOS One*, 17(12), e0279739.
<https://doi.org/10.1371/journal.pone.0279739>

RESEARCH ARTICLE

Raman microspectroscopy and machine learning for use in identifying radiation-induced lung toxicity

Ramie N. Ali-Adeeb¹ , Phil Shreeves² , Xinchen Deng¹, Kirsty Milligan¹, Alex G. Brolo³, Jullian J. Lum^{4,5}, Christina Haston¹, Jeffrey L. Andrews^{2*}, Andrew Jirasek^{1*} 

1 Department of Physics, The University of British Columbia - Okanagan campus, Kelowna, BC, Canada, **2** Department of Statistics, The University of British Columbia - Okanagan campus, Kelowna, BC, Canada, **3** Department of Chemistry, University of Victoria, Victoria, BC, Canada, **4** Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada, **5** Trev and Joyce Deeley Research Centre, BC Cancer, Victoria, BC, Canada

 These authors contributed equally to this work.

* jeff.andrews@ubc.ca (JLA); andrew.jirasek@ubc.ca (AJ)



OPEN ACCESS

Citation: Ali-Adeeb RN, Shreeves P, Deng X, Milligan K, Brolo AG, Lum JJ, et al. (2022) Raman microspectroscopy and machine learning for use in identifying radiation-induced lung toxicity. PLoS ONE 17(12): e0279739. <https://doi.org/10.1371/journal.pone.0279739>

Editor: Shaiju S Nazeer, Indian Institute of Space Science and Technology, INDIA

Received: September 21, 2021

Accepted: December 14, 2022

Published: December 30, 2022

Copyright: © 2022 Ali-Adeeb et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data files will be available on OSF at the following: https://osf.io/vqw7/?view_only=f409d84fcc08458e835a7aa89c37eddb.

Funding: This research was supported by the Natural Sciences and Engineering Research Council of Canada through their Discovery Grants program (RGPIN-2014-06730 - Andrews, RGPIN-2020-07232 - Jirasek) and the Cluster of Research Excellence in Medical Physics and Data Analytics (PI: Jirasek) enabled via the University of British

Abstract

Objective

In this work, we explore and develop a method that uses Raman spectroscopy to measure and differentiate radiation induced toxicity in murine lungs with the goal of setting the foundation for a predictive disease model.

Methods

Analysis of Raman tissue data is achieved through a combination of techniques. We first distinguish between tissue measurements and air pockets in the lung by using group and basis restricted non-negative matrix factorization. We then analyze the tissue spectra using sparse multinomial logistic regression to discriminate between fibrotic gradings. Model validation is achieved by splitting the data into a training set containing 70% of the data and a test set with the remaining 30%; classification accuracy is used as the performance metric. We also explore several other potential classification tasks wherein the response considered is the grade of pneumonitis and fibrosis sickness.

Results

A classification accuracy of 91.6% is achieved on the test set of fibrotic gradings, illustrating the ability of Raman measurements to detect differing levels of fibrotic disease among the murine lungs. It is also shown via further modeling that coarser consideration of fibrotic grading via binning (ie. 'Low', 'Medium', 'High') does not degrade performance. Finally, we consider preliminary models for pneumonitis discrimination using the same methodologies.

Columbia's Eminence Fund. Infrastructure support was provided by the Canada Foundation for Innovation through their John R. Evans Leaders Fund program (#35578 - Andrews, #32859 - Jirasek PI).

Competing interests: The authors have declared that no competing interests exist.

Introduction

Cancers treated with radiation in the thoracic region are limited by the occurrence of radiation induced lung toxicity, which can occur in up to 15–30% [1–3] of patients after treatment. Generally, the pathologies can be characterized by excessive inflammation (pneumonitis) or the deposition of extracellular matrix/collagen in the interstitium of the lung (fibrosis). Furthermore, lung toxicity can occur months to years after exposure resulting in impaired lung function and, in some cases, respiratory failure [3].

To study the toxicity response observed clinically, the radiation-induced pneumonitis and fibrosis response has been modelled in mice which have strain and genotype dependent presentations of these traits [4–9]. It has been documented [7, 8, 10, 11] that certain inbred mice will present respiratory distress at various times after thoracic irradiation with some being more prone to pneumonitis while others are more likely to develop fibrosis. Specifically relevant to this work, lethal pneumonitis develops 10–14 weeks following thoracic irradiation in C3H/HeJ mice and pneumonitis with fibrosis develops at 22 weeks or later in C57BL/6J mice.

Raman spectroscopy (RS) is an optical technique based on the inelastic scattering of light upon interaction with matter. This results in a unique “fingerprint” which is dependent upon the chemical composition of the material under investigation. The main advantages of RS for use in biological applications centre around the non-invasive, label-free nature of the technique. Although a fairly weak phenomenon, RS has been used extensively in many studies to distinguish between cancerous and non-cancerous tissue [12] in breast cancers [13–16], skin cancers [17–20], lung cancers [21], brain cancers [22, 23], and many other types of cancerous tissue [24–29]. More recently, RS has proved to be a useful tool in surgical guidance technology, where Zuniga *et al.* [30] demonstrated the use of RS as an efficient method in measuring surgical margins in tumour excisions of breast cancer patients. This can be enormously beneficial in ensuring all cancerous tissue is removed during surgery [31, 32].

Radiation effects have also been studied using RS. It has been previously demonstrated that radiation induced metabolic response pathways can be identified using RS [33–37]. Identification of metabolic pathways that are potentially involved in radiation resistance is enormously beneficial in terms of delivering a more successful treatment plan as radio sensitising drugs can be identified and used in combination therapies [34, 38, 39]. Raman spectroscopy has also been used to create mapped profiles in bronchial and lung tissues [40–42] as well as evaluate tissue pathologies such as liver fibrosis, human skin pilomatrixoma, and oncogenesis in the cervix [43–45]. Taken together, the capabilities of Raman spectroscopy make it an ideal tool for use in biomedical analysis.

Due to the complexity of Raman spectral data, typical methods for analysis involve the application of dimensionality reduction techniques such as principal component analysis (PCA) [33–35, 46, 47], allowing for the decomposition and interpretation of biochemical changes to be measured by RS. As an alternative approach, the successful application of non-negative matrix factorization (NMF) to spectral data sets has been shown to afford advantages over other dimensionality reduction techniques like PCA [36, 48–50]. One of the main advantages of NMF is the linearly additive, factor based representations of a non-negative data matrix which results in increased interpretability, particularly in the context of spectral data as it is representative of spectral superposition [51–53].

In this manuscript, we investigate the utility of RS for use in discrimination between healthy and fibrotic lung tissue. Currently, fibrosis is most commonly diagnosed through pathological examinations and immuno-histochemical (IHC) staining [54, 55]. Although these techniques prove successful in identifying fibrotic tissue, they can be both time-consuming and subjective with respect to different persons carrying out the analysis of tissue sections. RS can provide

large multiplexed biological information concerning multiple regions of lung tissue, which is often difficult to achieve using single IHC analyses, resulting in more efficient use of patient samples. Combining RS with both supervised and unsupervised machine-learning techniques, such as NMF, has the potential to create a rapid and robust method of diagnosing and distinguishing fibrotic tissue from healthy lung tissue [43].

The current aim of this work is to utilize RS to identify the formation of radiation-induced murine lung toxicity post radiation therapy. Throughout the following sections, we utilize RS in combination with machine learning techniques to, ultimately, generate models to differentiate between healthy and diseased lung tissue.

Materials and methods

Murine model

Mice: C57BL/6J and C3H/HeJ mice were purchased from the Jackson Laboratory (Bar Harbor, ME) and housed in the animal facility of the University of British Columbia—Okanagan. All mice were handled according to protocol A18–0140 approved by the Animal Care Committee at the University of British Columbia, in accordance with regulations set by the Canadian Council on Animal Use and Care.

Murine irradiation

Eight week old mice were treated with 14 Gy whole thorax irradiation at the BC Cancer Agency—Sindi Ahluwalia Hawkins Centre for the Southern Interior. Mice were anesthetized with intraperitoneal (i.p.) injections of sodium pentobarbitol (30 mg/kg) and xylazine (5 mg/kg). Upon reaching surgical level of anesthesia, mice were placed in a perspex box with a 13 mm thick lid and were irradiated with a flattening filter free, 6 MV Varian Linear Accelerator (Varian, Palo Alto, CA). Field was accurately positioned over thoracic cavity using a kV x-ray.

Tissue harvesting

After irradiation, mice were monitored weekly until ~ 10 weeks post treatment for C3H/HeJ and ~ 20 weeks post-treatment for C57BL/6J mice and subsequently were monitored semi weekly or daily for indications of physical distress as a result of the lung disease development. The experimental endpoint used to define the parameters for humane termination for the mice were $\geq 20\%$ body weight loss (from greatest observed weight) and/or shallow and rapid breathing, hunching, slow and tip-toe movement, ungroomed fur, and decreased physical responsiveness. The control mice (N = 2 for each strain), were sacrificed with lethal dose of sodium pentobarbitol using standard operating procedures to reduce suffering at the experiment end date (~ 15 weeks for C3H/HeJ and ~ 25 weeks for C57BL/6J).

Lung cryopreservation

Once a mouse qualified for humane endpoint, it was sacrificed with a lethal intraperitoneal injection of sodium pentobarbitol. After breathing ceased, cervical dislocation was performed and the lungs were removed. The left lobe was filled with a 1:1 dilution of optimal cutting temperature solution (TissueTek OCT; Sakura Finetek USA, Inc.) and 1X phosphate buffer solution (PBS; pH 7.4) to expand the lung to maintain anatomical morphology. The expanded lung was then placed in a plastic mold and surrounded by undiluted OCT. The mold was placed on crushed dry ice to freeze.

Sample processing

Tissue blocks were sectioned using a rotary cryostat (MICROM HM 550) and cuts were made (~ 50 microns) until the center of the frozen lung was reached. One $20\mu\text{m}$ section was placed on magnesium fluoride slides (MgF_2) and an adjacent $5\mu\text{m}$ section was taken for histology. Histology sections were placed on glass slides and stored in a dry ice cooler for transport and storage.

Pathological grading

Adjacent lung slices were sent to the University of British Columbia Vancouver Histology Lab where they were thawed and chemically fixed with formalin for staining. Following standard histology protocols, one section was stained with haematoxylin and eosin (H&E) to visualize microscopic anatomy while the other section was stained with Mason's TriChrome to highlight connective tissue, specifically collagen. To quantify degree of fibrosis phenotype expression, Trichrome stained sections were assessed in accordance with previously established protocols [56]. To summarize, areas determined to be fibrotic, characterized by condensed tissue and collagen build up, were identified, summed, and divided by the total area of the lung.

The H&E stained sections were used to assess the degree of pneumonitis using a semi-quantitative method established in previous work [6, 8, 56]. Sections were inspected and subjectively scored on a scale from 0 to 6 where 0 was no pathology characteristics observed and 6 was extreme pathology development was observed in the form of acute thickening of alveolar walls, high levels of infiltration from immune system cells, and excessive amounts of cellular exudate filling the alveolar spaces throughout the entire lung [8]. Assessment of stained slides was completed on a Motic BA210 microscope at 50x and 100x magnification.

In order to increase the resolution of the pathology grades from whole lung to specific areas in the lung, Raman map locations (Fig 1A) were matched as accurately as possible by eye to images taken of the histology stains of adjacent lung sections (Fig 1B). In trying to match features in the lung, arbitrary areas were drawn and tissue contained by area was graded by percent of disease present excluding air pockets. Ultimately, the more localized pathology grades were used in the final analysis.

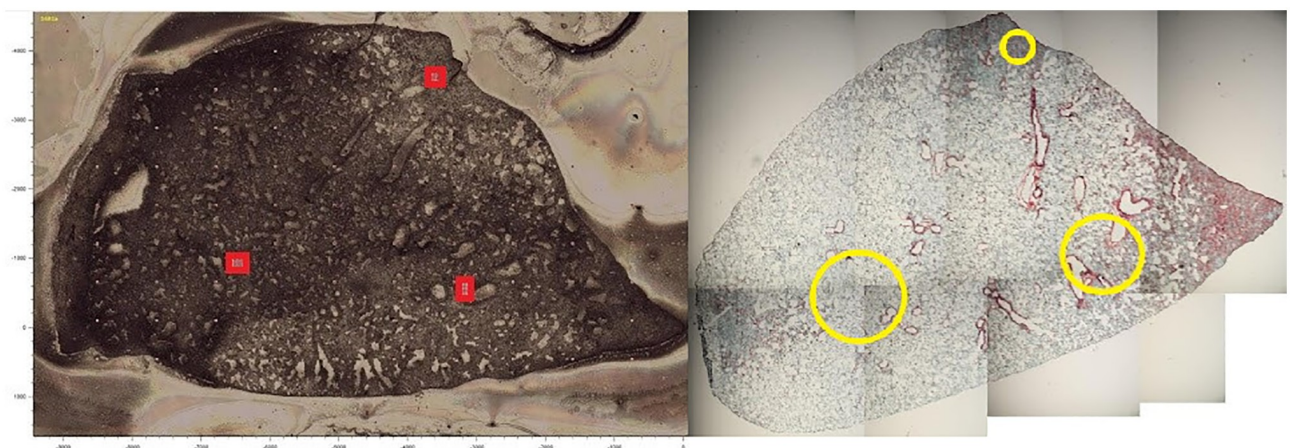


Fig 1. Region matching between Raman bright field images and IHC images. Left: Image of murine lung showing region of Raman acquisition maps (red squares). Right: Mouse lung showing corresponding regions (yellow circles) where pathological grades were assessed.

<https://doi.org/10.1371/journal.pone.0279739.g001>

Raman spectral acquisition

Re-frozen samples were allowed to thaw for 15–20 minutes before being gently rinsed with PBS to remove any excess OCT remaining in and around the sample. After rinsing with PBS, the samples were allowed to air dry for 15–20 minutes before spectral acquisition. Total time for Raman maps across two tissue samples to be acquired ranged from 6–8 hours. RS measurements were performed on an inVia Raman microscope (Renishaw Inc., Illinois, IL USA) with a 100X dry objective (N.A. = 0.9) (Leica Microsystems, Wetzlar, Germany) and a 830 lines/mm or 1200 lines/mm diffraction grating. A 785 nm continuous wave diode laser (300mW max power, $\sim 35\text{mW}$ power at sample, Renishaw Inc) was used for sample excitation with a laser power density of $\sim 0.5\text{ mW}/\mu\text{m}^3$ at the sample and a sampling volume of $\sim 100\mu\text{m}^3$.

From each lung section, spectra was acquired from ~ 3 regions of interest (ROI) with a mapping grid size of $15 \times 15\ \mu\text{m}^2$, the location of which were chosen at random across the lung. Generally, each region would consist of 30 – 80 points per map with the goal of ~ 160 points per lung section. Each point in the grid is a single spectral observation collected for 30 seconds. A 830 lines/mm grating was used for the majority of the data collection (Data set A). A small portion of the data was collected with a 1200 lines/mm grating with the same acquisition parameters. To merge the two data sets, the spectral window of both data sets were reduced to $[461\text{--}1606\ \text{cm}^{-1}]$ and Data set B was interpolated using a shape-preserving piece wise cubic interpolation method to match Data set A (MatLab r2018b).

Data processing

Cosmic rays were removed within the Renishaw instrument software (WiRE, Renishaw Inc). High frequency noise was removed from the spectra using a Savitzky-Golay (SG) filter [57] (window size of $n = 3$, polynomial order $p = 1$).

The baseline estimation uses an algorithm based on the Schulze signal removal method [58]. The algorithm first does a rough estimate of the baseline by applying a first order SG filter with a window size of 7% of the total spectra range. The baseline estimation process was repeated for 20 iterations as any further iterations were found to make no significant difference in the estimation. The data was normalized with respect to the area under the curve (AUC) of each spectrum. A Gaussian curve was fitted and centered on the most prominent peak (phenylalanine (Phe); $\sim 1005\ \text{cm}^{-1}$) for the first spectrum in a trial set and subsequent data was shifted accordingly to match to address any minor x-axis calibration drift. This was applied across the entire data set for two full iterations. This is the last step in the processing of the data in preparation for statistical analysis.

Non-negative matrix factorization

Non-negative matrix factorization (NMF) [51] is the process of decomposing a non-negative data matrix of interest into two lower rank non-negative matrices. This approach leads to each observation being represented as an additive linear combination of a common factor set. Here, the lung data of interest is decomposed into matrices H and W ($X \approx WH$), which have been shown to represent the chemical bases that the molecules are composed of [36] and the amount these bases contribute to each spectrum, respectively. As such, each individual spectrum can be written as a linear combination of the chemical bases acquired. A constrained form of NMF, known as group and basis restricted non-negative matrix factorization (GBR-NMF) [59], was used in order to allow for constraining of the factors in the model. The factors specified include both a raw spectrum of the OCT-PBS solution, as well as both averages of 10 quality spectra (i.e. regions uncontaminated by air pockets) and 10 undesirable (i.e.

air-pocket or OCT:PBS-contaminated) spectra. The following sections show that the GBR-NMF method is a useful way to remove observations from the data set that may be masked by the OCT-PBS solution.

Sparse logistic regression

Least absolute shrinkage and selection operator (LASSO) [60] regression models are commonly used in high dimensional analyses as they provide a form of variable selection. This is done by adding a penalty term to the optimization criterion, which tends to shrink the size of the coefficients in the model; even forcing some of them to zero. The severity of the penalization term can be adjusted by the user, allowing the number of coefficients to increase or decrease with respect to said term. This is effectively equivalent to constraining $\sum_{i=1}^j |\beta_i|$ such that it must be less than or equal to some specified value s [61]. Herein, we use the logistic regression implementation of LASSO in order to predict disease in observations using only the Raman spectra acquired throughout, without any other form of data transformation other than the post-processing of the spectral data set described above.

Model validation and evaluation measures

We first applied the GBR-NMF method to the 4068×678 data set using a rank of $q = 3$. These 3 factors include the average quality spectrum, the average undesirable spectrum, and a raw spectrum of OCT-PBS solution described in Results: Air pocket removal. Observations were then removed from the data set according to a pre-determined spectrum score threshold of ≤ 1 for the quality spectrum and ≥ 1 for the media score. The remaining “valid” spectra were then used to predict toxicity grade with respect to the spectra using multinomial LASSO logistic regression. Model performance is evaluated using classification rate, which is expressed as

$$CR = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{g}_i = g_i), \quad (1)$$

where $\mathbf{1}()$ is the indicator function and \hat{g}_i and g_i are the predicted and true groups of observation i . The LASSO penalty term was chosen using 10-fold cross validation with the largest term yielding a multinomial deviance within 1 standard error of the minimum as the objective; the formula and plots for which can be found in Supporting Information. Our main results for fibrotic grading prediction are found through a traditional training-testing setup. Therein, 70% of the observations were randomly retained to train the model and the remaining 30% were used to evaluate the expected classification rate. Additional models, utilizing the full data, were created with respect to binned pathology grading methods and separate mouse strains.

Results

Air pocket removal

Due to the nature of spectral mapping anatomical sections of the lung, acquiring spectra with little to no tissue present was unavoidable, due to the presence of air pockets. These air pockets contain residual OCT:PBS media which was used to fill the lungs during cryopreservation. To remove these spectral observations from the data set, constrained GBR-NMF was performed using three basis spectra representing RS acquired from lung tissue, lung tissue contaminated with OCT:PBS media (mixed spectrum), and RS of dried OCT:PBS buffer solution, as shown in Fig 2.

As the media spectra were distributed throughout the dataset, it was not feasible to remove all of them by hand. Additionally, there were spectra that resembled both the media and tissue

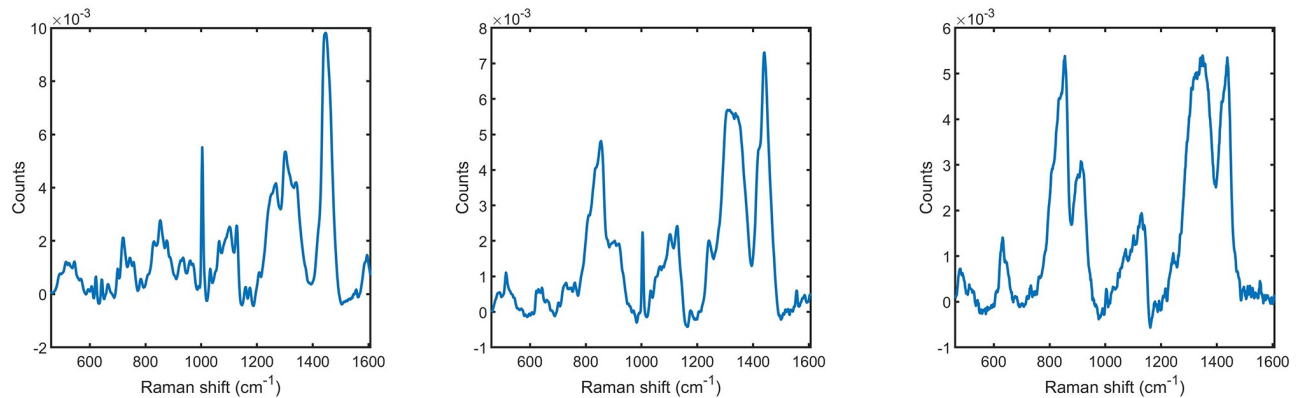


Fig 2. (a) Averaged lung tissue spectrum. (b) Averaged contaminated tissue spectrum. (c) OCT:PBS spectrum ('media').

<https://doi.org/10.1371/journal.pone.0279739.g002>

spectrum (example displayed in Fig 2(b)). As such, manual removal would introduce bias into the data as a result of qualitative assessment.

To filter out the air pocket and low quality spectral data, three basis spectra were generated (displayed in Fig 2) and used in GBR-NMF to analyze each spectral acquisition and output a score with respect to each basis. The data was filtered using a threshold of a lung tissue basis spectrum score ≥ 1 and a media basis spectrum score of ≤ 1 , which are the mean scores of each basis. Spectra falling inside this threshold were kept for further analysis, resulting in 59.4% data retention.

Prediction of fibrotic grade using Raman spectroscopy and LASSO

LASSO logistic regression was applied to the RS data. Based on the LASSO logistic regression model, certain wavenumbers displaying strong effects are retained for disease prediction, while other wavenumbers (coefficients) are reduced to zero in order to simplify the model as a whole; this is illustrated in Fig 3.

Within the LASSO model, disease is not given a binary classification but, rather, each region of observations is taken as a specific grade (probability) between zero and one, representing the proportion of the region considered as diseased. This proportion causes conventional logistic and LASSO logistic regression models to be invalid, as they use response

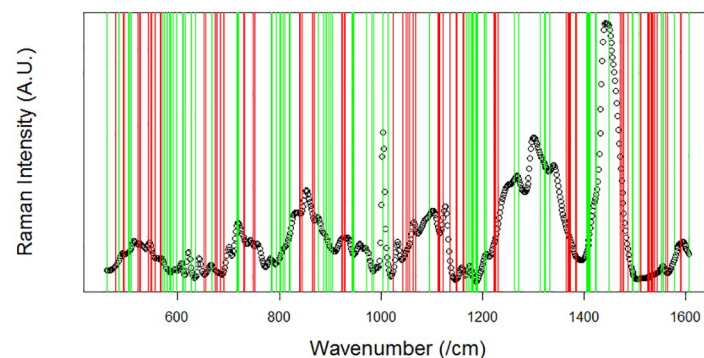


Fig 3. Example of wavenumbers with strong effects in predicting fibrotic grade. These wavenumbers are shown using both red and green lines, with red/green displaying that an increase in intensity results in a decrease/increase in disease likelihood, respectively. White areas are wavenumbers whose coefficients have been reduced to zero.

<https://doi.org/10.1371/journal.pone.0279739.g003>

Table 1. Multinomial logistic regression on full data set without separating into training and testing sets. Data was classified with respect to pathology gradings of the region the spectra were acquired from.

	Predicted							
	0	0.05	0.1	0.3	0.5	0.6	0.9	1
0	1905	1	6	0	0	0	0	0
0.05	13	30	3	0	0	0	0	0
0.1	17	1	84	7	0	0	0	0
0.3	4	1	1	132	0	0	0	0
0.5	5	0	0	6	13	0	0	0
0.6	2	0	1	1	0	59	0	0
0.9	2	0	0	0	0	0	59	0
1	0	0	0	0	0	0	0	59

<https://doi.org/10.1371/journal.pone.0279739.t001>

variables of zero and one. While it is possible to create a threshold in the grades, thereby narrowing the classification of observations into either diseased or not, doing so is arguably inappropriate. It cannot be assumed that the individual mouse itself is fully diseased or not diseased at all. For example, suppose a region has a grading of 0.6. This would then mean that roughly 40% of the observations sampled from this region would not have the disease. Therefore, to classify all observations in this region as infected would be misleading. Instead, we suggest that a classification process that has classes other than zero and one should be considered. Herein, multinomial logistic regression is used to classify observations in a multi-class scenario that is very similar to standard logistic regression. We provide the results of applying multinomial LASSO logistic regression for predicting fibrotic gradings to all our tissue spectra in [Table 1](#).

To further ensure the validity of the models, the data set was divided into training and testing sets. The randomly selected training set held 70% of the observations and the testing set held the other 30%. Classification results of the model can be found in [Table 2](#).

The multinomial logistic LASSO model performs quite well when distinguishing fibrotic grades, yielding a testing set classification rate of approximately 91.6%. The training classification rate of the model that uses all available data (results shown in [Table 1](#)) is, as expected, higher than the testing set results at approximately 96.9%. This is to be expected as the full training model is not making predictions on new data, meaning the model has already seen the spectra it is classifying. The training/testing model, on the other hand, is predicting the class of observations that are not included in the model creation and provides a more believable estimate of the long-run classification rate.

Table 2. Multinomial logistic regression prediction on the testing data set (70% training, 30% testing). Data was classified with respect to pathology gradings of the region the spectra were acquired from.

	Predicted							
	0	0.05	0.1	0.3	0.5	0.6	0.9	1
0	565	2	7	3	0	0	0	0
0.05	8	4	2	0	0	0	1	0
0.1	6	2	20	5	0	1	0	0
0.3	2	1	0	36	1	0	1	1
0.5	3	0	1	1	2	0	0	0
0.6	5	0	2	0	0	10	0	3
0.9	2	0	0	0	0	0	12	1
1	0	0	0	0	0	0	1	13

<https://doi.org/10.1371/journal.pone.0279739.t002>

Table 3. Multinomial logistic regression binned with respect to fibrotic pathology grades: Low (0–0.1), medium (0.3–0.6), high (0.9–1).

	Predicted		
	Low	Medium	High
Low: 0–0.1	2050	18	1
Medium: 0.3–0.6	21	198	7
High: 0.9–1	3	2	115

<https://doi.org/10.1371/journal.pone.0279739.t003>

It should also be noted that the pathology grades, either whole lung or regional, are vulnerable to human error. To combat this problem, the regional pathology grades were partitioned into different risk levels, creating a more generalizable approach. The grades at hand (0,0.05,0.1,0.3,0.5,0.6,0.9,1) can be separated using the two different binning methods outlined below:

1. Low (0–0.1), Medium (0.3–0.6), High (0.9–1).
2. Zero (0), Low (0.1–0.3), Medium (0.5–0.6), High (0.9–1).

As expected, both binning methods still yielded excellent results. This is merely a simplification of the models shown in Tables 1 and 2, reducing the number of groups from eight to three and four. Since binned groups also follow a certain order, observations being binned together do have stronger similarities to each other than those in the other groups. Prediction tables of said grouping methods are shown in Tables 3 and 4, which held classification rates of 97.8% and 96.9% respectively. These values are similar to those without the binning process, demonstrating that binning the grades does not degrade the value of the results.

Mouse strain specific analyses using Raman spectroscopy and LASSO

Models were also created with respect to the two individual strains to test how the reduction of variability across strains in the data set may aid the modelling of fibrosis, and even allow for the modelling of pneumonitis. As shown in Tables 5 and 6, classification accuracy improved for the fibrotic gradings (99.1% for C57BL/6 mice only and 99.4% for C3H/HeJ mice only). Although it is only a preliminary model, Table 7 also provides excellent differentiation among the pneumonitis gradings (89.2% for C3H/HeJ mice only). For all three scenarios, given the small sample sizes for some non-zero grades, proper validation of these results via training/testing is left for future work.

Discussion

A variety of LASSO logistic regression models were created in order to ensure robustness and validity of said models. This was done to predict fibrotic disease grade with respect to Raman

Table 4. Multinomial logistic regression binned with respect to fibrotic pathology grades: Zero (0), low (0.1–0.3), medium (0.5–0.6), high (0.9–1).

	Predicted			
	Zero	Low	Medium	High
Zero: 0	1898	15	0	0
Low: 0.1–0.3	31	264	1	0
Medium: 0.5–0.6	6	18	62	1
High: 0.9–1	2	0	2	116

<https://doi.org/10.1371/journal.pone.0279739.t004>

Table 5. Multinomial logistic regression on spectral data obtained from C57BL/6 mice only. Data was classified based on degree of fibrosis (regional pathology grades).

	Predicted						
	0	0.05	0.1	0.3	0.6	0.9	1
0	1311	1	2	0	0	0	0
0.05	11	35	0	0	0	0	0
0.1	0	0	70	0	0	0	0
0.3	0	0	0	57	0	0	0
0.6	1	0	0	0	62	0	0
0.9	0	0	0	0	0	61	0
1	0	0	0	0	0	0	59

<https://doi.org/10.1371/journal.pone.0279739.t005>

Table 6. Multinomial logistic regression on spectral data obtained from C3H/HeJ mice only. Data was classified based on degree of fibrosis (regional pathology grades).

	Predicted			
	0	0.1	0.3	0.5
0	599	0	0	0
0.1	1	37	1	1
0.3	0	0	81	1
0.5	0	0	0	24

<https://doi.org/10.1371/journal.pone.0279739.t006>

Table 7. Multinomial logistic regression on spectral data obtained from C3H/HeJ mice only. Data was classified based on degree of pneumonitis (regional pathology grades).

	Predicted							
	0	0.05	0.1	0.15	0.2	0.3	0.5	0.8
0	30	6	6	0	0	0	0	0
0.05	1	223	5	0	0	0	1	0
0.1	4	10	172	6	0	0	0	0
0.15	0	0	6	93	0	0	0	0
0.2	0	0	0	0	35	1	0	0
0.3	0	0	2	2	1	24	11	0
0.5	0	1	1	1	0	1	76	2
0.8	0	0	1	0	1	2	8	12

<https://doi.org/10.1371/journal.pone.0279739.t007>

spectra and was shown to have excellent classification accuracy (96.9% on full training data & 91.6% on the testing set when separated into 70% training and 30% testing).

When further binning the grades, training classification rates (100% of data included in model) of 97.8% and 96.9% were found for binning process 1 and 2 respectively. This result agreed with the initial training classification of 96.9% using the original grading scheme. The first binning process has a slightly higher classification rate than the others, which could likely be attributed to the greater simplicity of the model.

Models were also created for the individual strains of mice. This was to further test how the variability affected the prediction process with respect to disease grades. As shown in Tables 5 and 6, classification rates improved when the strains were separated. Furthermore, modeling C3H/HeJ data alone permitted additional preliminary modelling of pneumonitis grades, which proved challenging when considering both strains together due to added complexities. The modelling of which has been left for future work.

Conclusion

Using RS in tandem with supervised machine learning techniques, radiation induced lung disease (fibrosis in particular) was effectively measured in a murine model. Looking forward, the methods developed and tested in this work can be applied to explore application in early detection of radiation induced lung toxicity. Ultimately, the goal would be to determine if Raman spectroscopy can predict the development of lung toxicity in murine lung tissue on time frames shorter than current standard of practice.

Supporting information

S1 Fig. Cross validation curve for general data set.

(TIFF)

S2 Fig. Cross validation curve for train/test data set.

(TIFF)

S3 Fig. Cross validation curve for binned data—Pathology grades: Low, medium, high.

(TIFF)

S4 Fig. Cross validation curve for binned data—Pathology grades: Zero, low, medium, high.

(TIFF)

S5 Fig. Cross validation curve for strain separated data—C57BL/6.

(TIFF)

S6 Fig. Cross validation curve for strain separated data—C3H/HeJ.

(TIFF)

S7 Fig. Cross validation curve for strain separated data—C3H/HeJ: Pneumonitis grades.

(TIFF)

S1 Appendix. Formula for multinomial deviance.

(PDF)

Author Contributions

Conceptualization: Ramie N. Ali-Adeeb, Christina Haston, Andrew Jirasek.

Data curation: Ramie N. Ali-Adeeb, Phil Shreeves, Jeffrey L. Andrews.

Formal analysis: Ramie N. Ali-Adeeb, Phil Shreeves, Jeffrey L. Andrews.

Investigation: Ramie N. Ali-Adeeb.

Methodology: Ramie N. Ali-Adeeb.

Project administration: Andrew Jirasek.

Resources: Christina Haston, Andrew Jirasek.

Supervision: Alex G. Brolo, Jullian J. Lum, Christina Haston, Jeffrey L. Andrews, Andrew Jirasek.

Validation: Phil Shreeves, Jeffrey L. Andrews.

Visualization: Ramie N. Ali-Adeeb, Phil Shreeves.

Writing – original draft: Ramie N. Ali-Adeeb, Andrew Jirasek.

Writing – review & editing: Ramie N. Ali-Adeeb, Phil Shreeves, Xincheng Deng, Kirsty Milligan, Alex G. Brolo, Jullian J. Lum, Christina Haston, Jeffrey L. Andrews.

References

1. Carver JR, Shapiro CL, Ng A, Jacobs L, Schwartz C, Virgo KS, et al. American society of clinical oncology clinical evidence review on the ongoing care of adult cancer survivors: Cardiac and pulmonary late effects. *Journal of Clinical Oncology*. 2007; 25(25):3991–4008. <https://doi.org/10.1200/JCO.2007.10.9777> PMID: 17577017
2. Barriger RB, Fakiris AJ, Hanna N, Yu M, Mantravadi P, McGarry RC. Dose-volume analysis of radiation pneumonitis in non-small-cell lung cancer patients treated with concurrent cisplatin and etoposide with or without consolidation docetaxel. *International Journal of Radiation Oncology Biology Physics*. 2010; 78(5):1381–1386. <https://doi.org/10.1016/j.ijrobp.2009.09.030> PMID: 20231061
3. Dr Emami B. Tolerance of Normal tissue to Therapeutic radiation. *Reports of Radiotherapy and Oncology*. 2013; 1(1):35–48.
4. Paun A, Haston CK. Genomic and genome-wide association of susceptibility to radiation-induced fibrotic lung disease in mice. *Radiotherapy and Oncology*. 2012; 105(3):350–357. <https://doi.org/10.1016/j.radonc.2012.08.004> PMID: 22954495
5. Sharplin J, Franko AJ. A quantitative histological study of strain-dependent differences in the effects of irradiation on mouse lung during the intermediate and late phases. *Radiation Research*. 1989; 119(1):15–31. <https://doi.org/10.2307/3577363> PMID: 2756106
6. Haston CK, Mckerlie C, Newbigging S, Corey M, Rozmahel R, Tsui Lc. Detection of modifier loci influencing the lung phenotype of cystic fibrosis knockout mice. *Mammalian Genome*. 2002; 13:605–613. <https://doi.org/10.1007/s00335-002-2190-7> PMID: 12461645
7. Chiang CS, Liu WC, Jung SM, Chen FH, Wu CR, McBride WH, et al. Compartmental responses after thoracic irradiation of mice: Strain differences. *International Journal of Radiation Oncology Biology Physics*. 2005; 62(3):862–871. <https://doi.org/10.1016/j.ijrobp.2005.02.037> PMID: 15936571
8. Haston CK, Begin M, Dorion G, Cory SM. Distinct Loci Influence Radiation-Induced Alveolitis from Fibrosing Alveolitis in the Mouse. *Cancer Research*. 2007; 67(22):10796–10803. <https://doi.org/10.1158/0008-5472.CAN-07-2733> PMID: 18006824
9. Iwakawa M, Noda S, Ohta T, Oohira C, Tanaka H, Tsuji A, et al. Strain dependent differences in a histological study of CD44 and collagen fibers with an expression analysis of inflammatory response-related genes in irradiated murine lung. *Journal of Radiation Research*. 2004; 45(3):423–433. <https://doi.org/10.1269/jrr.45.423> PMID: 15613788
10. Hong JH, Jung SM, Tsao TCY, Wu CJ, Lee CY, Chen FH, et al. Bronchoalveolar lavage and interstitial cells have different roles in radiation-induced lung injury. *International Journal of Radiation Biology*. 2003; 79(3):159–167. <https://doi.org/10.1080/0955300031000076894> PMID: 12745880
11. Paun A, Bergeron ME, Haston CK. The Th1/Th17 balance dictates the fibrosis response in murine radiation-induced lung disease. *Scientific Reports*. 2017; 7(1):1–12. <https://doi.org/10.1038/s41598-017-11656-5> PMID: 28912510
12. Cui S, Zhang S, Yue S. Raman Spectroscopy and Imaging for Cancer Diagnosis. *J Healthc Eng*. 2018; 2018:1–11. <https://doi.org/10.1155/2018/8619342> PMID: 29977484
13. Frank CJ, McCreery RL, Redd DCB, Gansler TS. Characterization of Human Breast Biopsy Specimens with Near-IR Raman Spectroscopy. *Anal Chem*. 1994; 66(3):319–326. <https://doi.org/10.1021/ac00075a002> PMID: 8135372
14. Frank CJ, McCreery RL, Redd DCB. Raman Spectroscopy of Normal and Diseased Human Breast Tissues. *Anal Chem*. 1995; 67(5):777–783. <https://doi.org/10.1021/ac00101a001> PMID: 7762814
15. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using raman spectroscopy. *Cancer Res*. 2002; 62(18):5375–5380. PMID: 12235010
16. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci U S A*. 2005; 102(35):12371–12376. <https://doi.org/10.1073/pnas.0501390102> PMID: 16116095
17. Nijssen A, Bakker Schut TC, Heule F, Caspers PJ, Hayes DP, Neumann MHA, et al. Discriminating basal cell carcinoma from its surrounding tissue by raman spectroscopy. *J Invest Dermatol*. 2002; 119(1):64–69. <https://doi.org/10.1046/j.1523-1747.2002.01807.x> PMID: 12164926
18. Gniadecka M, Philipsen PA, Sigurdsson S, Wessel S, Nielsen OF, Christensen DH, et al. Melanoma Diagnosis by Raman Spectroscopy and Neural Networks: Structure Alterations in Proteins and Lipids in

- Intact Cancer Tissue. *J Invest Dermatol*. 2004; 122(2):443–449. <https://doi.org/10.1046/j.0022-202X.2004.22208.x> PMID: 15009728
19. Silveira L, Bodanese B, Albertini R, Zângaro RA, Pacheco MTT. Differentiating normal and basal cell carcinoma human skin tissues in vitro using dispersive Raman spectroscopy: A comparison between principal components analysis and simplified biochemical models. *Photomed Laser Surg*. 2010; 28 (SUPPL. 1):119–127. PMID: 20649423
 20. Bodanese B, Silveira FL, Zângaro RA, Pacheco MTT, Pasqualucci CA, Silveira L. Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through raman spectroscopy and principal component analysis. *Photomed Laser Surg*. 2012; 30(7):381–387. <https://doi.org/10.1089/pho.2011.3191> PMID: 22693951
 21. Lui H, Zhao J, McLean D, Zeng H. Real-time raman spectroscopy for in vivo skin cancer diagnosis. *Cancer Res*. 2012; 72(10):2491–2500. <https://doi.org/10.1158/0008-5472.CAN-11-4061> PMID: 22434431
 22. Desroches J, Jermyn M, Mok K, Lemieux-Leduc C, Mercier J, St-Arnaud K, et al. Characterization of a Raman spectroscopy probe system for intraoperative brain tissue classification. *Biomed Opt Express*. 2015; 6(7):2380. <https://doi.org/10.1364/BOE.6.002380> PMID: 26203368
 23. Jermyn M, Mok K, Mercier J, Desroches J, Pichette J, Saint-Arnaud K, et al. Intraoperative brain cancer detection with Raman spectroscopy in humans. *Sci Transl Med*. 2015; 7(274):1–10. <https://doi.org/10.1126/scitranslmed.aaa2384> PMID: 25673764
 24. Huang Z, Bergholt MS, Zheng W, Lin K, Ho KY, Teh M, et al. In vivo early diagnosis of gastric dysplasia using narrow-band image-guided Raman endoscopy. *J Biomed Opt*. 2010; 15(3):037017. <https://doi.org/10.1117/1.3420115> PMID: 20615046
 25. Huang Z, Teh SK, Zheng W, Lin K, Ho KY, Teh M, et al. In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy. *Biosens Bioelectron*. 2010; 26(2):383–389. <https://doi.org/10.1016/j.bios.2010.07.125> PMID: 20729057
 26. Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, So JBY, et al. Fiber-optic Raman spectroscopy probes gastric carcinogenesis in vivo at endoscopy. *J Biophotonics*. 2013; 6(1):49–59. <https://doi.org/10.1002/jbio.201200138> PMID: 23288709
 27. Bergholt MS, Zheng W, Ho KY, Teh M, Yeoh KG, Yan So JB, et al. Fiberoptic confocal raman spectroscopy for real-time in vivo diagnosis of dysplasia in Barrett's esophagus. *Gastroenterology*. 2014; 146 (1):27–32. <https://doi.org/10.1053/j.gastro.2013.11.002> PMID: 24216327
 28. Duraipandian S. Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination. *Journal of Biomedical Optics*. 2012; 17(8):81418. <https://doi.org/10.1117/1.JBO.17.8.081418> PMID: 23224179
 29. Shaikh R, Dora TK, Chopra S, Maheshwari A, Kedar KD, Bharat R, et al. In vivo Raman spectroscopy of human uterine cervix: exploring the utility of vagina as an internal control. *J Biomed Opt*. 2014; 19 (8):087001. <https://doi.org/10.1117/1.JBO.19.8.087001> PMID: 25104415
 30. Zúñiga WC, Jones V, Anderson SM, Echevarria A, Miller NL, Stashko C, et al. Raman Spectroscopy for Rapid Evaluation of Surgical Margins during Breast Cancer Lumpectomy. *Sci Rep*. 2019; 9(1):1–16. <https://doi.org/10.1038/s41598-019-51112-0> PMID: 31601985
 31. Santos IP, Barroso EM, Bakker Schut TC, Caspers PJ, Van Lanschot CGF, Choi DH, et al. Raman spectroscopy for cancer detection and cancer surgery guidance: Translation to the clinics. *Analyst*. 2017; 142(17):3025–3047. <https://doi.org/10.1039/C7AN00957G> PMID: 28726868
 32. Hollon T, Lewis S, Freudiger CW, Xie XS, Orringer DA. Improving the accuracy of brain tumor surgery via Raman-based technology. *Neurosurg Focus*. 2016; 40(3):1–9. <https://doi.org/10.3171/2015.12.FOCUS15557> PMID: 26926067
 33. Matthews Q, Jirasek A, Lum JJ, Brolo AG. Biochemical signatures of in vitro radiation response in human lung, breast and prostate tumour cells observed with Raman spectroscopy. *Phys Med Biol*. 2011; 56(21):6839–6855. <https://doi.org/10.1088/0031-9155/56/21/006> PMID: 21971286
 34. Matthews Q, Isabelle M, Harder SJ, Smazynski J, Beckham W, Brolo AG, et al. Radiation-induced glycogen accumulation detected by single cell raman spectroscopy is associated with radioresistance that can be reversed by metformin. *PLoS One*. 2015; 10(8):1–15. <https://doi.org/10.1371/journal.pone.0135356> PMID: 26280348
 35. Harder SJ, Matthews Q, Isabelle M, Brolo AG, Lum JJ, Jirasek A. A Raman spectroscopic study of cell response to clinical doses of ionizing radiation. *Applied spectroscopy*. 2015; 69(2):193–204. <https://doi.org/10.1366/14-07561> PMID: 25588147
 36. Deng X, Ali-Adeeb R, Andrews JL, Shreeves P, Lum JJ, Brolo A, et al. Monitor Ionizing Radiation-Induced Cellular Responses with Raman Spectroscopy, Non-Negative Matrix Factorization, and Non-Negative Least Squares. *Appl Spectrosc*. 2020; 0(0):1–11. PMID: 32098482

37. Devpura S, Barton KN, Brown SL, Zhang Y, Sethi S, Klein MD, et al. Correlation of Raman Spectroscopic Analysis With Pathologic Scoring for Assessment of Radiation Induced Lung Tumor and Normal Tissue Response in a Preclinical Animal Model. *International Journal of Radiation Oncology*Biophysics*Physics*. 2015; 93(3):E597. <https://doi.org/10.1016/j.ijrobp.2015.07.2070>
38. Buckley AM, Lynam-Lennon N, O'Neill H, O'Sullivan J. Targeting hallmarks of cancer to enhance radio-sensitivity in gastrointestinal cancers. *Nat Rev Gastroenterol Hepatol*. 2020; 17(5):298–313. <https://doi.org/10.1038/s41575-019-0247-2> PMID: 32005946
39. Wang Y, Deng W, Li N, Neri S, Sharma A, Jiang W, et al. Combining immunotherapy and radiotherapy for cancer treatment: Current challenges and future directions. *Front Pharmacol*. 2018; 9(MAR):1–11. <https://doi.org/10.3389/fphar.2018.00185> PMID: 29556198
40. Koljenović S, Bakker Schut TC, van Meerbeeck JP, Maat APWM, Burgers Sa, Zondervan PE, et al. Raman microspectroscopic mapping studies of human bronchial tissue. *Journal of biomedical optics*. 2004; 9(6):1187–1197. <https://doi.org/10.1117/1.1805555> PMID: 15568939
41. Min YK, Yamamoto T, Kohda E, Ito T, Hamaguchi HO. 1064 nm near-infrared multichannel Raman spectroscopy of fresh human lung tissues. *Journal of Raman Spectroscopy*. 2005; 36(1):73–76. <https://doi.org/10.1002/jrs.1280>
42. Marzec KM, Kochan K, Fedorowicz A, Jaształ A, Chruszcz-Lipska K, Dobrowolski JC, et al. Raman microimaging of murine lungs: Insight into the vitamin A content. *Analyst*. 2015; 140(7):2171–2177. <https://doi.org/10.1039/C4AN01881H> PMID: 25535673
43. Ramírez-Elías MG, Kolosovas-Machuca ES, Kershennobich D, Guzmán C, Escobedo G, González FJ. Evaluation of liver fibrosis using Raman spectroscopy and infrared thermography: A pilot study. *Photodiagnosis and Photodynamic Therapy*. 2017; 19(February 2017):278–283. PMID: 28754542
44. Cheng WT, Liu MT, Liu HN, Lin SY. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. *Microscopy Research and Technique*. 2005; 68(2):75–79. <https://doi.org/10.1002/jemt.20229> PMID: 16228983
45. Daniel A, P A, Ganesan S, Joseph L. Biochemical assessment of human uterine cervix by micro-Raman mapping. *Photodiagnosis and Photodynamic Therapy*. 2017; 17:65–74. <https://doi.org/10.1016/j.pdpdt.2016.08.011> PMID: 27671518
46. Harder SJ, Matthews Q, Isabelle M, Brolo AG, Lum JJ, Jirasek A. A Raman spectroscopic study of cell response to clinical doses of ionizing radiation. *Applied Spectroscopy*. 2015; 69(2):193–204. <https://doi.org/10.1366/14-07561> PMID: 25588147
47. Meksiarun P, Aoki PHB, Van Nest SJ, Sobral-Filho RG, Lum JJ, Brolo AG, et al. Breast cancer subtype specific biochemical responses to radiation. *The Analyst*. 2018; 143:3850–3858. <https://doi.org/10.1039/C8AN00345A> PMID: 30004539
48. Sajda P, Du S, Parra LC. Recovery of constituent spectra using non-negative matrix factorization. *Wavelets: Applications in Signal and Image Processing X*. 2003; 5207(November 2003):321.
49. Li H, Adal T, Wang W, Emge D, Cichocki A. Non-negative matrix factorization with orthogonality constraints and its application to Raman spectroscopy. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*. 2007; 48(1-2):83–97. <https://doi.org/10.1007/s11265-006-0039-0>
50. Milligan K, Deng X, Shreeves P, Ali-Adeeb R, Matthews Q, Brolo A, et al. Raman spectroscopy and group and basis-restricted non negative matrix factorisation identifies radiation induced metabolic changes in human cancer cells. *Scientific reports*. 2021; 11(1):1–11. <https://doi.org/10.1038/s41598-021-83343-5> PMID: 33594122
51. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999; 401(6755):788–791. <https://doi.org/10.1038/44565> PMID: 10548103
52. Ding C, He X, Simon HD. On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering. *Society for Industrial and Applied Mathematics Proceedings of the SIAM International Conference on Data Mining*. 2005; p. 606–610.
53. Lin X, Boutros PC. Fast Nonnegative Matrix Factorization and Applications to Pattern Extraction, Deconvolution and Imputation. *bioRxiv*. 2018; p. 1–10.
54. Török NJ. Recent advances in the pathogenesis and diagnosis of liver fibrosis. *J Gastroenterol*. 2008; 43(5):315–321. <https://doi.org/10.1007/s00535-008-2181-x> PMID: 18592147
55. Valeyre D. Towards a better diagnosis of idiopathic pulmonary fibrosis. *Eur Respir Rev*. 2011; 20(120):108–113. <https://doi.org/10.1183/09059180.00001611> PMID: 21632798
56. Haston CK, Amos I, King M, Travis L. Inheritance of Susceptibility to Bleomycin-induced Pulmonary Fibrosis in the Mouse1. *Cancer Research*. 1996; 56:2596–2601. PMID: 8653703
57. Afseth NK, Segtnan VH, Wold JP. Raman spectra of biological samples: A study of preprocessing methods. *Applied Spectroscopy*. 2006; 60(12):1358–1367. <https://doi.org/10.1366/000370206779321454> PMID: 17217584

58. Schulze G, Jirasek A, Yu MML, Lim A, Turner RFB, Blades MW. Investigation of selected baseline removal techniques as candidates for automated implementation. *Applied Spectroscopy*. 2005; 59(5):545–574. <https://doi.org/10.1366/0003702053945985> PMID: 15969801
59. Shreeves P, Andrews J, Deng A, Ali-Adeeb R, Jirasek A. Nonnegative matrix factorization with group and basis restrictions. arXiv preprint arXiv:2107.00744 [Preprint]. 2021 [cited 2022 Dec 19]. Available from: <https://arxiv.org/abs/2107.00744>
60. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996; 58(1):267–288.
61. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2011; 73(3):273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>