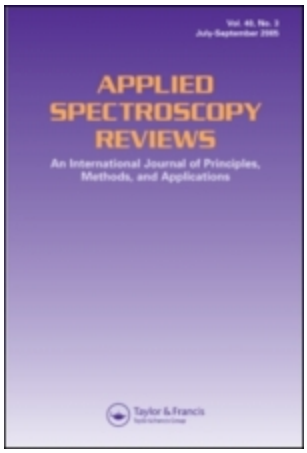


Applications of Machine Learning in Spectroscopy

Journal:	*Applied Spectroscopy Reviews*
Manuscript ID	LAPS-2020-0104.R1
Manuscript Type:	Reviews
Date Submitted by the Author:	n/a
Complete List of Authors:	Meza Ramirez, Carlos ; Lancaster University, Engineering Greenop, Michael ; Lancaster University, Engineering Ashton, Lorna ; Lancaster University, Chemistry Rehman, Ihtesham; Lancaster University, Engineering
Keywords:	analytical, Machine learning, chemometrics, artificial intelligence, data science, Infrared and Raman spectroscopy

Applications of Machine Learning in Spectroscopy

Carlos A. Meza Ramirez [a], Michael Greenop [a], Lorna Ashton [b], Ihtesham ur Rehman [a*].

[a]*Engineering Department, Lancaster University, Lancaster, United Kingdom;*

[b]*Department of Chemistry, Lancaster University, Lancaster, United Kingdom*

* Corresponding Author: i.u.rehman@lancaster.ac.uk

Abstract

The way to analyze data in spectroscopy has changed substantially. At the same time, data science has evolved to the point where spectroscopy can find space to be housed, adapted and be functional. The integration of the two sciences has introduced a knowledge gap between data scientists who know about advanced machine learning techniques and spectroscopists who have a solid background in chemometrics. To reach a symbiosis, the knowledge gap requires bridging. This review article focuses on introducing data science subjects to non-specialist spectroscopists, or those unfamiliar with the subject. The article will explain concepts that are covered in machine learning, such as supervised learning, unsupervised learning, deep learning, and most importantly, the difference between machine learning and artificial intelligence. This article also includes examples of published spectroscopy research, in which some of the concepts explained here are applied. Machine learning together with spectroscopy can provide a useful, fast, and efficient tool to analyze samples of interest both for industrial and research purposes.

Keywords

Machine learning, chemometrics, artificial intelligence, data science, Infrared and Raman spectroscopy

Aims of the study

The main objective of the article introduced here is to present to scientists involved in any spectroscopic field a brief background on machine learning and artificial intelligence, as well as to present the different concepts and methods commonly used in this field.

Introduction

Data science is the combination of statistics, computer science and domain knowledge [1]. Scientists in the spectroscopy field by definition have spectroscopy specific domain knowledge. Most spectroscopists have a background in the use of statistics specific to chemical analysis, chemometrics, but few have a broad computer science background. The use of machine learning (ML) has become increasingly popular in a wide range of scientific fields, including spectroscopy, as datasets increase in size and milestones in the wider field of artificial intelligence (AI) are publicized [2, 3].

The authors are spectroscopists, who have aimed to take advantage of the potential advantages provided by ML algorithms such as, reducing pre-processing [4 - 7] alongside increasing accuracy and / or computational efficiency [7 - 9]. During our research, we have identified a number of definitions, practices and concepts that are common between data science and spectroscopy, which if clear earlier would have made ML use quicker and easier. We have also found a need for comparison between different techniques, using the various strengths of different algorithms for different stages of research and kinds of data.

It is for this reason that we have written this review, with synonymous terms provided in brackets to aid translation. Different spectroscopic techniques, such as, dielectric, Infrared, Raman Spectroscopy together with ML will be discussed, demonstrating the breadth of ML in spectroscopy. Advantages of the techniques, such as the high-throughput, non-destructive nature of vibrational spectroscopy that provides the opportunity to repeatedly analyze and sample using for multiple methods are also highlighted. Our research focuses on bio-spectroscopy and we therefore focus on biomedical examples but the principles are expandable to other fields of spectroscopy.

The review is designed for focused reading, meaning that different sections do not rely on all previous sections. A reader wishing to know about clustering can read only the clustering sections, the same applies for classification, regression or a background in AI. The theory for each ML sub-type, how they relate or overlap with chemometric techniques is provided in the background. The reader can then determine the best group of algorithms for their application and then read about examples of these algorithms in later sections of the review.

Common examples of chemometric techniques include principle component analysis (PCA) [10, 11], linear discriminant analysis (LDA) [11 - 15] partial least squares (PLS) [17, 18], support vector machines (SVM) [18-20] and hierarchical cluster analysis (HCA) [16]. Most of these algorithms (LDA, PLS, SVM and HCA) may justifiably also be called ML, a potential cause of confusion elaborated on in further sections. A typical feature of traditional chemometric techniques is accuracy, when the number of features (measurements / wavenumbers) is larger than the number of observations (number of samples / collected spectra). Accuracy when the number of collected spectra is lower than the number of wavenumbers is an advantage in spectroscopy, where thousands of spectra may need collecting before observations (spectra from separate samples) exceeds the number of features. Collecting thousands of spectra may initially sound easy to a researcher collecting Raman or IR maps, but collecting multiple spectra from the same sample risks overfitting. Overfitting reduces the generalisation of the model, meaning that it will be highly accurate when analysing the original sample but may be inaccurate when a new sample is introduced (even if the sample is similar to the original), result in errors. Therefore, thousands of spectra maps may need to be collected (a significant time burden), making the production of large datasets unpractical, highlighting chemometrics advantage for accuracy in spectroscopy when features are lower than observations.

The disadvantage of some chemometric algorithms is their computational cost, an increasingly problem as dataset become larger. Machine learning (ML) is a branch of artificial intelligence (AI) [21] that uses advanced statistical methods to determine key features within a dataset [22]. Learning is the process of modified understanding that results from interactions with an environment, both physical and virtual [23]. In spectroscopy, ML algorithms identify features (wavenumbers) produced within the samples physical environment that produce labels (biomarkers) in classification or predictions in regression.

Vibrational spectroscopy has been used in cancer research [11, 24, 25], analysis of biofilms [26-28] and the development and monitoring of drugs and drug delivery [29-32]. In spite of the high accuracies published using vibrational spectroscopy, it has been relatively slow to be used in mainstream pathology labs. There may be several reasons for this but one cause may be the relatively small size of the typical vibrational spectroscopy dataset when compared to data collected by health services such as the NHS (National Health Service) in the UK or during phase III trials. The use of chemometrics, whilst providing relatively quick analysis during an

early stage of an investigation may be slowing the use of vibrational spectroscopy techniques on a larger scale. ML techniques potentially provide a solution to this problem.

Artificial Intelligence and Machine Learning (Background)

Artificial intelligence (AI) is being applied to numerous fields of research [33]. But what is artificial intelligence? AI is better defined by what it has done than what it aims to do, with in-depth discussions on how to define AI available [34]. Traditionally the focus has been on replicating different factors of human intelligence in computational form, from game playing algorithms [35] that have become world champions in games such as chess [2] and Go [3], grandmaster at Starcraft II [36] and played poker [2]. With applications in medical research such as speeding cancer drug development [37] and predicting useful protein structures [38]. The traditional goal of human-like general intelligence, considered impossible by some [39] has incrementally been supplemented by subfields such as machine vision, data mining, natural language processing, robotics and relevant for spectroscopy, machine learning [34]. AI is therefore a general term, with ML being the specific area of AI applicable to spectroscopy [22, 23].

ML is defined as an algorithm that "learns" associations within data [22, 40, 41]. Learning is a complex process to define [22]. A suggested definition of when a machine learns is its core structure changing, constituted by a program or asset of data, where the behaviour of the algorithm is estimated to improve [5]. There are several overlapping sub-sections of machine learning. These relate to the kind of learning they carry out (unsupervised, semi-supervised, unsupervised and reinforcement), the kind of problem they are used for (clustering, classification or regression) and its influence on its environment from the data provided (active or passive) [23, 41].

Active learning interacts with the environment, in other words, while the experiment is running the active learner questions and flag out queries. Passive learning needs the data from outside. Whatever happens within the environment, the passive learner observes that data without direct manipulation or influence. An algorithm may for example be an unsupervised, passive learning algorithm used for classification. In some cases, capable of both classification and regression depending on the algorithm's configuration and application. PLS for example is a regression algorithm [42-44], but partial least squares

discriminant analysis (PLS-DA) is a classification algorithm [45, 46]. Neural networks are also capable of both classification and regression [6, 47, 48, 49].

The algorithms that publicized the advances in ML such as AlphaGo [35], AlphaZero [3] and AlphaStar [36] are reinforcement-learning algorithms. Reinforcement learning aims to replicate the human brain, where dopamine "rewards" the brain for behaviours expected to be evolutionarily beneficial to humans. In ML, the algorithm investigates a virtual environment (game space) with the goal of maximising rewards (increased game play score) [3]. Self-play, where the algorithm plays itself at the game repeatedly, allows the algorithm to determine and retain (learn) reward maximising (game winning) strategies, which are stored in memory. The high number of games played (millions-billions) allows the algorithm to experience a far higher number of possible games than a human player, who is limited by life span, allowing the algorithm to accurately compute the probability of any given move leading to victory. Reinforcement learning and self-play have been used for vibrational spectroscopy investigations [50-52] but are not as widely applied as supervised learning or unsupervised learning algorithms. No evidence of semi-supervised learning algorithm use in spectroscopy is known to the authors.

Unsupervised learning methods are defined by their lack of sample labels. Because the algorithm has no way of knowing what a sample is, unsupervised methods have the advantage of objectivity. An example application of unsupervised learning in medicine may be analysis of patient clinical information to identify heart attack risk [53]. The disadvantage of unsupervised methods is that the algorithm has no way of knowing a typical sample, making them sensitive to outliers that can distort the relationships determined by the algorithm. Examples of unsupervised learning include clustering in chemometrics and autoencoders in neural networks [11, 54]. An application of unsupervised learning may be to determine the features (biomarkers) that distinguish the spectra cancerous and non-cancerous tissue with clustering. In supervised learning, sample labels direct the algorithm. The advantage of labelled data is potentially increased accuracy; the disadvantage is the introduction of subjectivity into the analysis. It is due to the reduction in objectivity that unsupervised algorithms are commonly paired with unsupervised algorithms for feature selection [13, 18, 49]. Selection of the best features to direct supervised algorithms is a key skill in data science.

It is predicted that being able to analyze big data sets, will bring productivity, efficiency, innovation and competitiveness to every field of industry. Three different attributes exist in process-systems engineering that need fulfilling so machine learning can be applied. Compatibility of process knowledge, effectiveness to deal with uncertainties, and to generate interpretable solutions. Since, these attributes are valid for engineering processes, they might be extrapolated to any scientific field [40]. ML can be applied to different fields, such as finance, marketing, IT, medicine, biology, physics, astronomy, chemistry, robotics, etc. Machine learning can also potentially be used to detect, voices and faces [22]. However, as essential objectives, ML will search accuracy of prediction and interpretation of data at every analysis [40].

Theory: Chemometrics, dimension reduction and unsupervised learning

Principle component analysis (PCA) is one of the most commonly used techniques in chemometrics. Dimension reduction is required to reduce the features being analyzed by a model, improving computational efficiency and highlighting regions of interest within a spectrum. Although there are also different forms of dimension reduction [55], PCA is the most common by far in spectroscopy. PCA is used in ML but is not widely considered a ML algorithm, as it highlights key features in the data through transformation rather than learning. PCA is typically used as a pre-processing step for machine learning algorithms such as LDA [12 - 15], SVM [18 - 20] or PLS [16, 17]. PCA provides an objective perspective during the exploration of vibrational spectroscopic data, as it is an unsupervised method. Unsupervised techniques analyze unlabelled data, allowing the data to be inspected without pre-existing bias.

It is important to note that in different situations, features, wavenumbers and biomarkers can be the same thing. To a spectroscopist, a spectrum contains three thousand wavenumbers, to a data scientist these are features. PCA reduces the dimensions, and aids feature selection (the process of determining significant features) by projecting the wavenumbers into hyperspace along new dimensions that are orthogonal and ordered to account for the maximum variance from the first principle component. Once features (wavenumbers / ratios or combinations of wavenumbers) have been determined that best label, diagnose or

separate a pathogen, it becomes a biomarker to a biomedical researcher. The different terms used in the different fields of research all ultimately refer to the individual molecules or bonds that the wavenumbers are linked to.

PCA orders the dimensions that provide the greatest information (variance) through eigen decomposition of the data matrix X into the score matrix W and the loadings matrix T [56]. The data matrix X is produced by stacking the collected spectra on top of each other in a n×m matrix, where n is the number of observations (spectra) and m are the number of measurements (wavenumbers). The scores matrix is calculated by multiplying the data matrix by the transpose of the data matrix ($W = XX^T$) and the loadings matrix is calculated by multiplying the data matrix by the score matrix ($T = XW$). Each column of scores matrix is an eigen vector, where each value is a score for the entire spectrum on that row of the data matrix.

The value of unsupervised analysis is highlighted in Figure 1, where the H&E stained sample is shown in Fig. 1A and PC 1-4 is shown in Fig 1. B-E. To many people the word “cancer” is emotive, the reason the sample is being analyzed is because it is cancerous. To the researcher collecting the data, the cancer is “important”. PC1 disagrees, the highest scores are shown in the extracellular matrix (ECM) regions of the sample, shaded light pink in Fig. 1A. Potentially surprising to the researcher, at least initially. PC2 and 3 appear to highlight the tumorous cell regions, shaded dark purple in Fig. 1A. But the greatest detail is captured in PC4. As PC4 has been measured to represent only 99% of the variance, the PCA analysis seems not to have considered the cancer as important as the ECM. But if complete ignorance of cancer is assumed, the results make more sense. The majority of the sampled area is ECM, if no pre-existing bias is present, the larger number of ECM spectra becomes more prominent. It is this capacity to remove pre-existing bias that makes unsupervised exploration of data prior to supervised analysis so valuable.

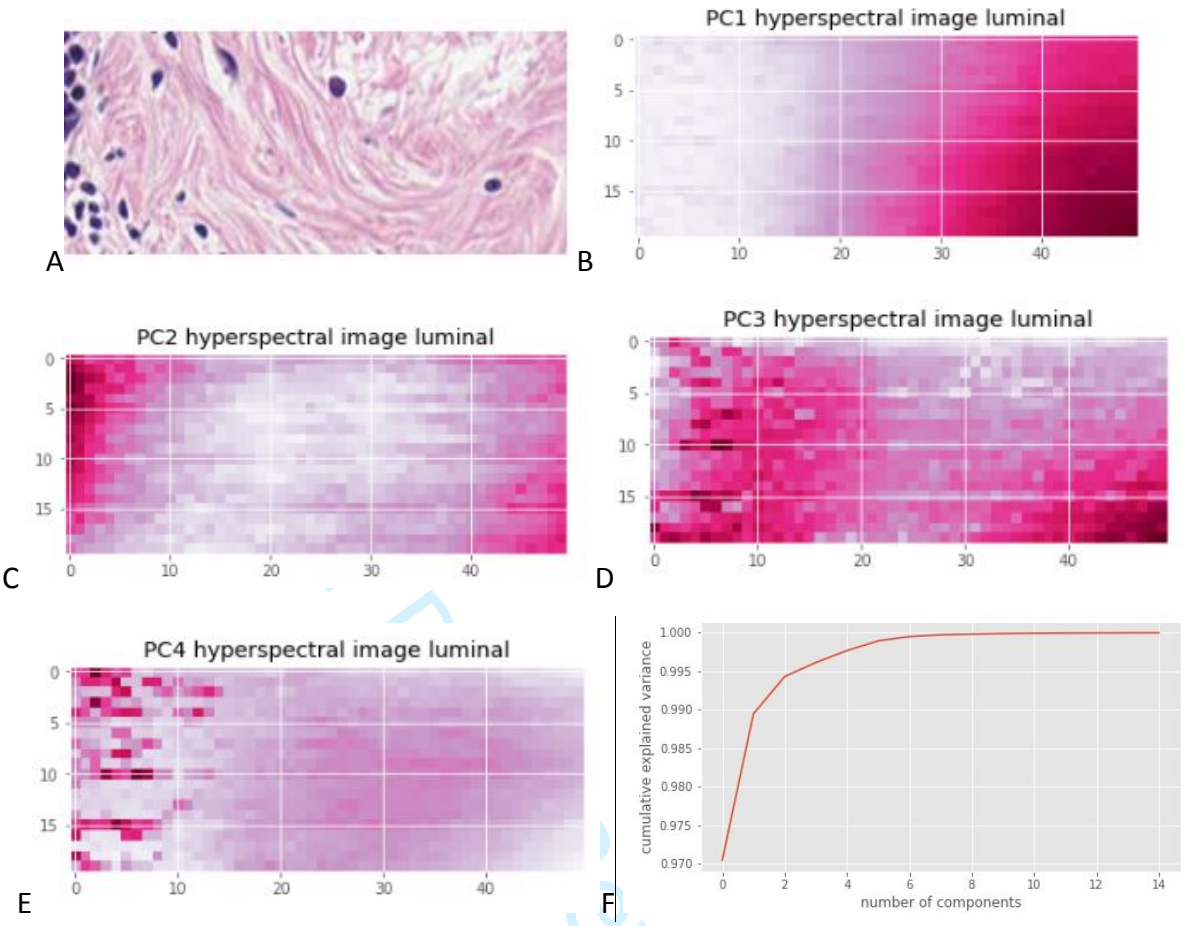


Figure 1 – A) H&E stained luminal breast cancer sample showing the cancerous cells (purple) and the extracellular matrix (pink). B-E) False-colour shading based on principle components 1, 2, 3 and 4 score respectively (white low / dark purple high). F) Cumulative explained variance plot showing the majority of variance is explained in the first six principle components. This figure was made with data owned by the authors.

It is possible to classify samples in two or three-dimensional score plots, the bigger the distinction between clusters letting the researcher know how well the samples are classified. Repeatability is indicated through the tightness of a cluster, the tighter the cluster, the greater the repeatability. The loading plot indicates the wavenumbers that contribute to the spread of the clusters, the further from zero the loading for a given wavenumber, the more "statistically interesting" it is. It is a common misconception that a specific wavenumber with a high loading is individually responsible for the distribution of the score plots, as a number of wavenumbers may be responsible in combination.

Another kind of unsupervised technique is clustering. There are numerous kinds of clustering algorithms, with some of the most popular including, K-means [57, 58] and hierarchical cluster analysis (HCA) [16]. Clustering labels samples based on features that group it with similar samples. The ability to label features without the analysis being influenced by previous

knowledge of the samples has the advantages of providing objectivity to a study, reducing potential bias in the analysis.

Unsupervised learning is a key tool when determining relationships within new data. An example is shown in Figure 2, where HCA is used to form a dendrogram (Fig. 2A) that splits the samples into descending, hierarchical clusters and sub- clusters. Each cluster is given a label and by rearranging the labels column back into the original shape of the collected image, the distribution of the groups can be plotted (Fig. 2B) and compared to an imaging technique such as H&E staining (Fig. 2C).

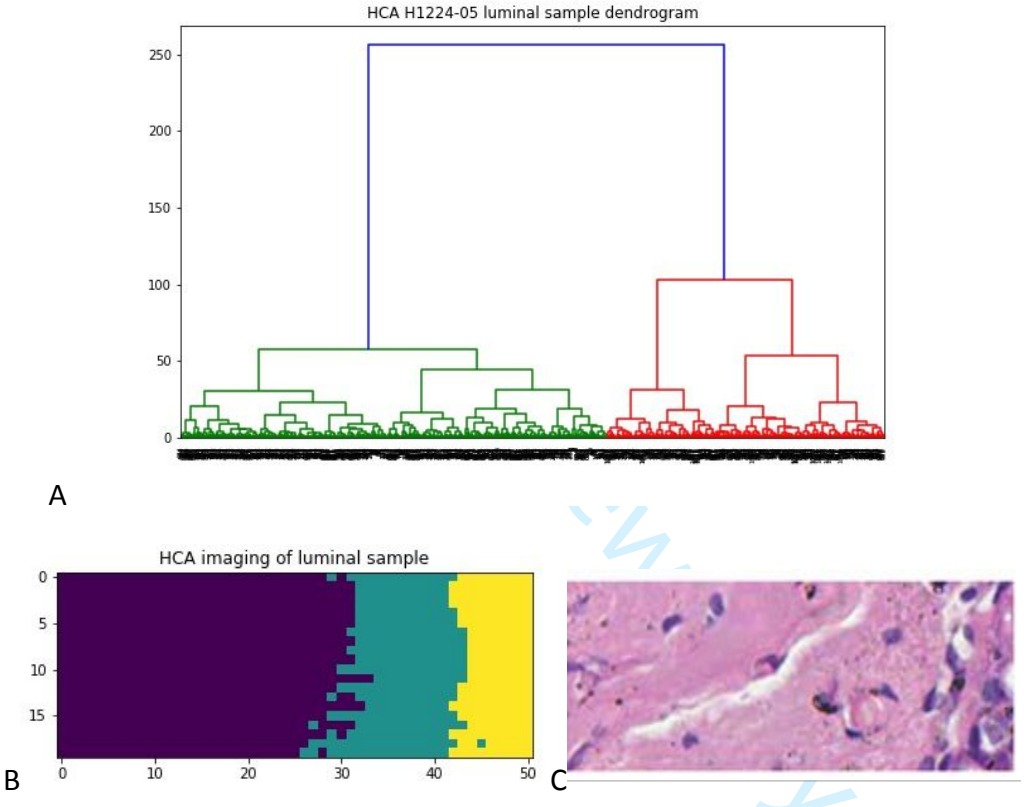


Figure 2 – A) HCA dendrogram, diving the spectra into clusters relating to their spectral similarity when each feature is projected into multidimensional space. B) Each pixel is coloured depending on its cluster, relating to either tumorous (yellow), healthy (blue) or intermediate tissue (green). C) To verify the HCA shaded image in B, a H&E stained image is provided, with the dark purple staining the nuclei. The high density of nuclei at the right of the image showing the location of the tumour and the light pink stains the collagen fibres of the ECM to the left of the image. This figure was made with data owned by the authors.

The value of unsupervised clustering is that it confirms and visualizes relationships within the data. In Figure 2, a question such as “can FTIR determine the difference between cancerous, intermediate and healthy tissue?” is quickly and simply answered, yes. The tumorous tissue shown in purple in the H&E stained image (Figure 2), intermediate and healthy tissues, stained

pink are shown as yellow, green and blue respectively in Figure 2. It separates the regions by labelling samples based on how similar they are to other samples. The dendrogram is used to inform the number of clusters chosen, each branch representing a cluster. It can be seen in Figure 2, two and three clusters are clearly defined but the clusters become similar after that point.

The speed and objectivity of unsupervised methods make them ideal for exploratory data analysis, an early phase of analysing a data set. Algorithms such as HCA allows for quick and easily visualisation of relationships within data. PCA can then be used for dimension reduction, reducing the number of features that need to be considered and highlighting potential regions of interest within the loadings plot. Once a relationship within the data has been confirmed through unsupervised techniques, supervised methods can be used for more specific questions. It is common to carry out dimension reduction with PCA before carrying out LDA, SVM or PLS analysis. The reduced dimensions lowers the computational cost and loadings plots can be used to highlight potential regions interest in the spectrum. Plotting principle components, from the covariance matrix plots.

PCA and HCA are two of the most commonly used unsupervised methods used in spectroscopy but there are several other that were out the remit of this paper to discuss (Fig 3). Each has advantages for different applications but the main advantage of all is the objectivity that they provide. The main disadvantage of unsupervised techniques is the lack of direction. For a more targeted approach, supervised learning is required.

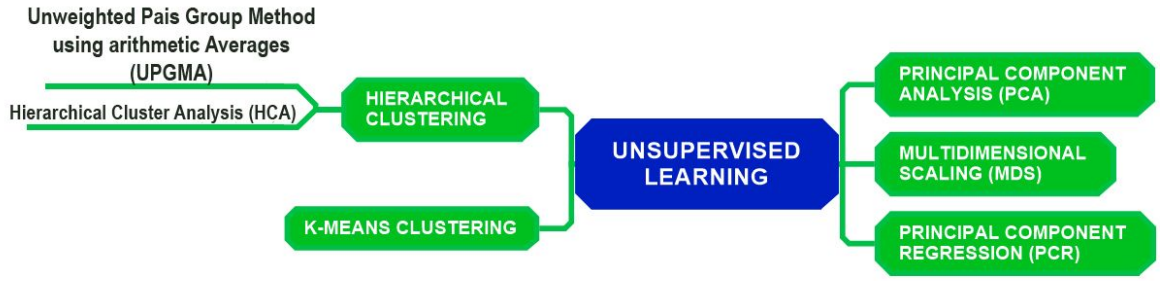


Figure 3. Unsupervised methods commonly employed when analysing spectral data.

Theory: Supervised learning (Regression, classification and Neural networks)

Supervised learning algorithms can be created either for classification predictive modelling, or as regression predictive modelling. Nonetheless, the difference lies in predicting classes (labels), or quantities. In spectroscopy both approaches can be achieved depending on the nature of the study, we have identified that the classification predictive models tend to be the most common [59, 60].

Support Vector Machines

Support Vector Machines (SVM) is a learning method used for classification and regression purposes. Based on pattern recognition method the SVM use hyperplanes (lines, planes) as decision boundaries to clearly separate classes in a multi-dimensional space. As shown on Figure 4 the hyperplane is separating the classes, however to find the ideal hyperplane it is necessary to calculate the distance between the nearest support vector (nearest data point) and the hyperplane. These distances are called margins [59, 61]. Normally, margins and the optimum hyperplane is calculated by the algorithms previously designed in the software used such as The Unscrambler and Python Scikit learn library [62]. Support vectors influence the hyperplane position, in other words the hyperplane location will depend on the layout of the data.

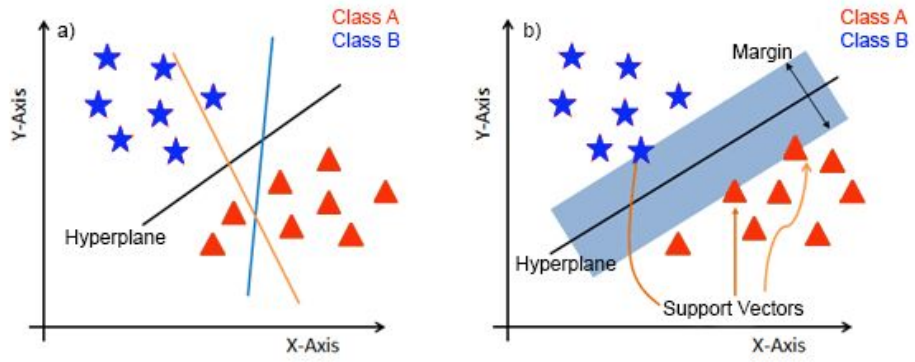


Figure 4. Support Vector machine data separation behaviour. A) Shows a cluster separation formed by a linear regression model fit. B) Shows a cluster separation by the hyperplane and the distance of the points to it.

An SVM algorithm works differently when it is used to classify than when a linear regression is performed. In the first case, binary or multiple classification is possible. In a binary classification, the algorithm decides which support vector belongs to a certain class by calculating the margin of each support vector based on the individual hyperplane of each

class. In other words, as Figure 5 illustrates, class A decision boundary is described by the formula w*x+b=-1 and class B decision boundary is described by w*x+b=+1, where w is weight vector (vector which helps to classify the training examples), x is the feature or input vector (wavelength, wavenumbers), and b the bias unit (useful to detect the best hyperplane between each class) [60, 61, 63]. Hence, each support vector with a margin less than -1 and greater than +1 will be assigned either to class A or B, respectively.

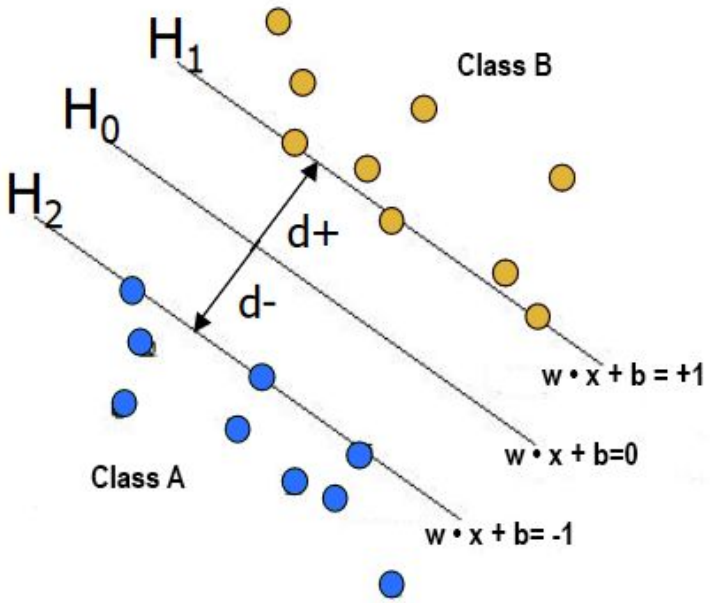


Figure 5. Illustrative approach from the decision boundary of a Support Vector Machine.

Nevertheless, data classified under SVM algorithms can be obtained by using non-linear decision boundaries. Decision boundaries are also known as kernels. Kernels offer a different approach to classify binary and multi-class datasets. Figure 6 shows different approaches of SVM decision boundaries using Scikit learn library for Python programming [62].

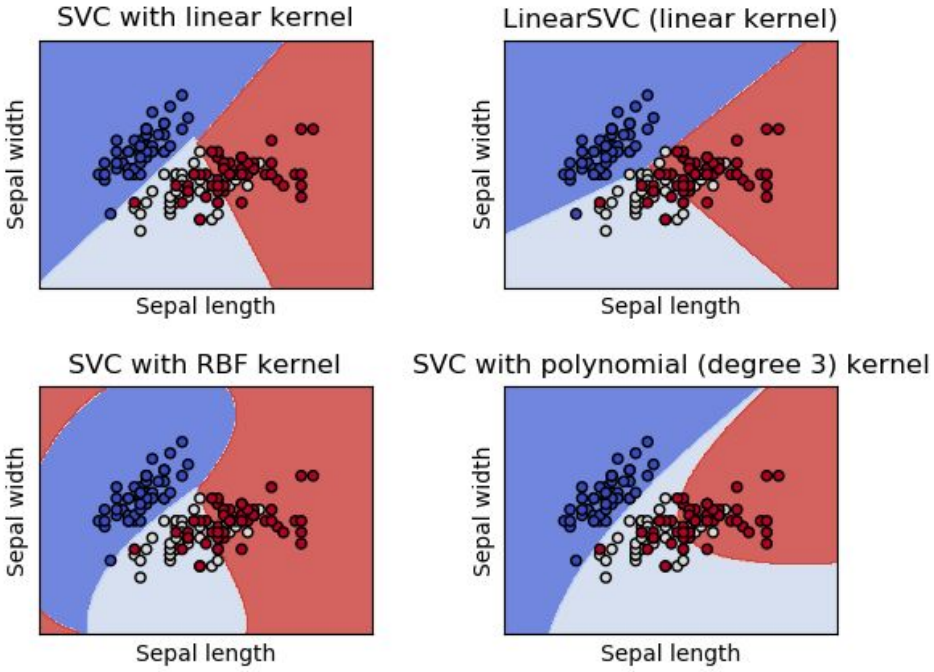


Figure 6. Different applications of decision boundaries applied to a SVM model of plants where sepal length and width are being classified.

As mentioned previously, SVM can be performed as a regression model. Differently from the classification method, this model depends only from a subset of training samples since the function who is dedicated to find the optimum decision boundary do not care on the error of predicting samples [62].

Linear Regression

Linear regression compares an independent variable *X* against a dependent variable Y in the equation $Y = \beta_0 + \beta_1 X + \varepsilon$ where β_0 is the y-axis intercept, β_1 is the slope of the regression line and ε is the error term [64]. If more than one feature is to be used to predict an outcome, multiple linear regression is used by including more terms into the equation e.g. $Y = \beta_0 + \beta_1 X + \beta_2 X ... \beta n X + \varepsilon$ [65, 66]. In this case, n denotes the number of features included in the analysis. Linear regression is rarely used in vibrational spectroscopy, where thousands of wavelengths (features) are compared when analysing samples. To reduce the number of features, principle component regression (PCR) may be used [66, 67]. Principle component regression is an unsupervised algorithm carried out over two steps. Firstly, PCA is used to reduce the dimensions (features / wavelengths) being analyzed, then a linear or multiple linear regression is carried out on the principle components, allowing predictions to be made that are compared to separate measurements. PCR is used in near-infrared spectroscopy to

calibrate quantitative measurements, where known substance concentrations are compared to predicted values and the error determined through measures such as the coefficient of determination R^2 [68-70]. The closer to one, the greater the accuracy of the model. PCR has the advantage of being relatively interpretable but has the disadvantage of not being directed by the known measurements (as it is an unsupervised technique).

Partial least squares (PLS) is a supervised technique, generally more accurate than PCR [69, 70] as it rotates its dimensions in such a way as to find the greatest variance with the known measurements accounted for [70]. As a result, PLS is more common in vibrational spectroscopy as a predictive method of calibration quantitative measurements and as a classifier when the discriminant analysis variant is applied (when Y is categorical) [71]. There are several computational methods to carry out PLS, they all generally aim to do the same thing. Compared to PCR, which uses eigenvalue decomposition (PCA) to rotate the data within the data matrix X onto a new axis that accounts for the maximum variance, PLS rotates the data so that it accounts for the maximum variance between the X and Y matrices [72]. When looking at a single feature (univariate analysis) PLS is known as type one PLS (PLS1) and when looking at more than one feature (multivariate analysis) type two PLS (PLS2). By accounting for both the observed data (X) and the known data (Y), PLS increases the accuracy of the model by including known measurements (Y) that direct prediction [69-71].

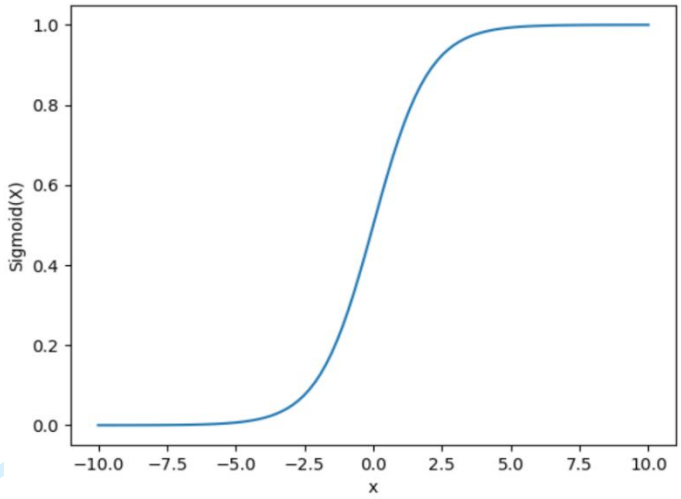
Logistic regression

Logistic regression (LR) has a similar behaviour as linear regression, however LR is widely used for classification thanks to the nature of the mathematical function of the algorithm [62]. LR is ruled by the sigmoid function [72]. The typical problems that LR approach are in the categorical form, i.e. spam or not spam, big or small, malignant or nor malignant .

The sigmoid function brings any real value between 0 and 1 and it is defined as:

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

The plotted form of ~~this~~ the previous function is shown in Figure 7a. ~~as follows:~~



~~*Figure 7. Sigmoid function plot.*~~

On the other hand, t within the function is a linear function, previously described.

$$t = \beta_0 + \beta_1 x$$

Hence, the logistic equation will become:

$$f(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Therefore, once the data is performed under the LR equation, every data point will fall between the values one and zero as shown in Figure 7b ~~8~~.

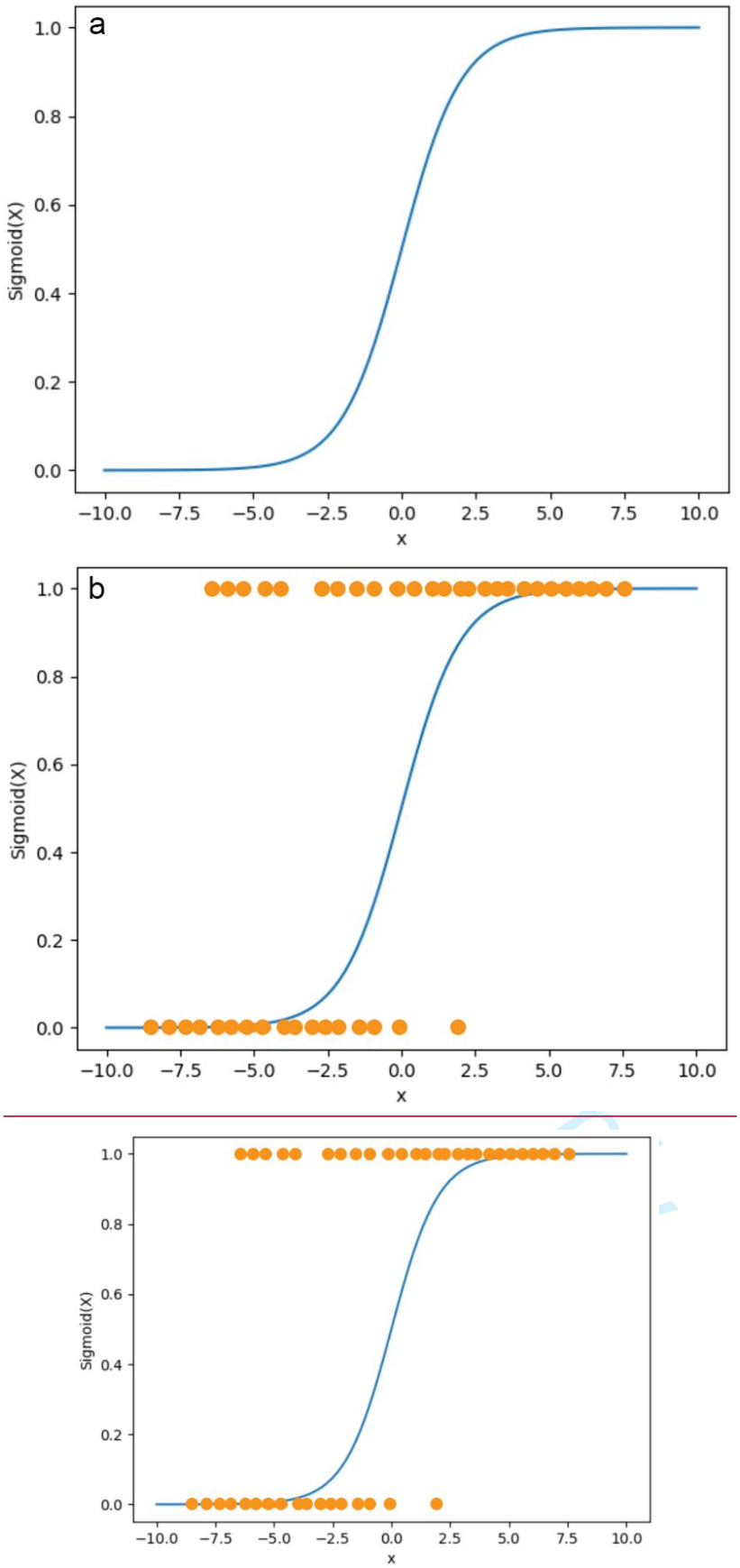


Figure 7&8. A) Sigmoid function plot. B) Illustrative example of data classification using logistic regression model. Logistic regression is typically used for binary classifications. For example, the X axis on the Figure may represent the tumor size

located at any specific region of the human body, whereas the Y axis represents the probability of the tumor to be malignant or not. Normally, the value of 1 is used to classify a TRUE outcome, in this case a positive malignant tumor, while 0 is a FALSE outcome or negative malignant tumor. In other words, and keeping with the tumor example, all those samples (orange dots) grouped on Y=0 will belong to a negative prediction of a malignant tumor, and all those samples (orange dots) on Y=1 will be considered as malignant .

Neural networks

Neural networks are a group of algorithms that simulate the architecture of a neuron. An artificial neural network (ANN) is a shallow neural network. A definitive distinction between a shallow and deep neural network is hard to define but around five to ten layers is the grey area between them, although with neural networks being produced that are hundreds of layers deep, the distinction may change in time. There are a range of software available to build neural networks such as R [73, 74], Matlab [75] and Python [6]. The most common method of building neural networks is in Python, using libraries such as PyTourch [73, 74] and TensorFlow/Keras [6, 9]. PyTorch was developed by Facebook and is one of the most popular package alongside Googles TensorFlow [77]. Google also developed Keras as a more approachable format that uses a Tensorflow backend. The libraries provide commands for different parameters, allowing the model to be customized for a specific application.

TensorFlow / Keras is regularly used in vibrational spectroscopy studies [4,5,78,79]. TensorFlow applications include improving diagnosis of bone metastasis of prostate cancer through the analysis of 1281 spectra from 427 patients [78]. The convolutional neural network had improved accuracy when compared to chemometric techniques such as PCA-LDA, PCA-LR and SVM, whilst excluding radiation exposure and reducing the cost of diagnosis induced during the typical detection method, radionuclide bone scan [78]. Another TensorFlow produced CNN was used to analyse Raman spectra collected from extracellular vesicles, spherical particles that are secreted by mammalian cells [4]. The aim of the project was to distinguish blood derived (healthy) and tumour derived (pathogenic) extracellular vesicles as a means of cancer diagnosis [4]. The CNNs maximum accuracy was 96.6% using the 400-1800cm^{-1} region of the spectrum [4]. The results were compared to a previous study, where quadratic discriminant analysis was used after pre-processing and PCA dimension reduction of the data, producing a maximum accuracy of 95% when the same region of the spectrum was analyzed [4]. The advantage of the CNN was that it did not require any data pre-processing or dimension reduction [4], an advantage also determined during the analysis of chemicals using a Keras-TensorFlow CNN [79]

Feed forward networks are made of three main kinds of layers, input, hidden and output layers. Input layers enter the data into the model; in spectroscopy, the initial input would typically be a spectrum or selected wavelengths with each node inputting one wavelength [48]. Each "neuron" (node) in the hidden layer (or layers) and output layer has an input (dendrite), a processing unit (soma) and an output (axon) (Figure 8 ~~9~~). Each processing unit multiplies inputs by a weight (a factor or parameter which has a substantial effect on the input), sums the multiples, adds a bias and classifies the summed inputs with an activation function. A range of activation functions exist [20], but a common one in beginner neural networks is the sigmoid function, found in logistic regressors as previously described.

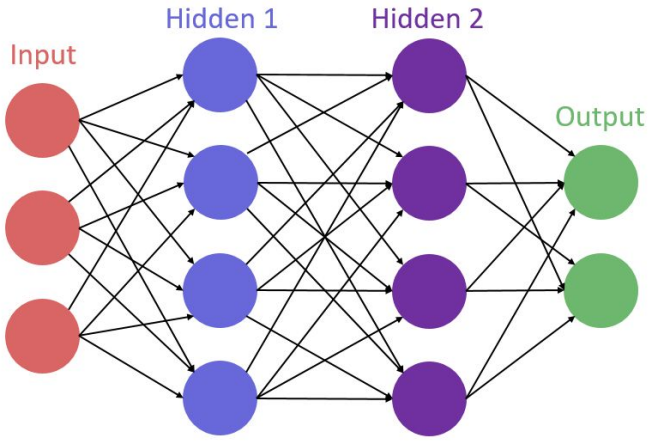


Figure ~~8~~9. Feed Forward Neural Network scheme. The flow of data goes forwards, from the input to the output as the arrows illustrate.

The exponential within the sigmoid function results in outputs tending to either one or zero (depending on where it is positive or negative), the same as the logistic regressor. The neurons that output close to zero result therefore diminish in statistical importance as the neuron fails to "fire", reducing its contribution to the next layer. The weight of each feature and bias of each neuron therefore influences how "important" the feature is within the model. The highlighted combinations are then passed onto the next layer in a process known as forward propagation. Each layer therefore acts as a filter, determining the statistical importance of each feature and combination of features towards the labelling of a sample.

The sigmoid function has remained popular in online courses and introductory lectures for neural networks, possibly as ML novices may find the near binary output they provide more intuitive. The disadvantage of sigmoid functions for deeper neural networks is the vanishing gradient problem [21]. Tanh, sigmoid and rectified linear unit (ReLU) activation functions are

other commonly used options [21]. ReLU is by far the most popular activation function, with variations such as the parametric ReLU (PReLU), leaky ReLU and randomized leaky ReLU (RReLU) [21, 78, 80], although these are outside the scope of this review. Softmax is the most commonly used activation function for the output layer because it provides a probability [21]. The need for experimentation is required to determine the correct activation function for a given application to maximize accuracy. The final (output) layer ends by outputting some result, the accuracy of which is determined by a loss function, such as cross-entropy for classification problems [81].

Loss functions determine the error by comparing the predicted label produced with test data, against the actual label. By quantifying the error, the loss function provides a metric to optimize through refinement of the model. The percentage of data held back for testing is typically 20-30%. If the loss of two configurations of weights, biases and neuron inputs are calculated, the difference between the two values reveals the difference in accuracy between the two configurations. Different factors, which could be (but are not limited to) weights for different inputs or different neuron biases, may have a greater or lesser influence on the accuracy. Put in mathematical terms, they have different rates of change. A factor with a larger rate of change has a greater influence on model accuracy.

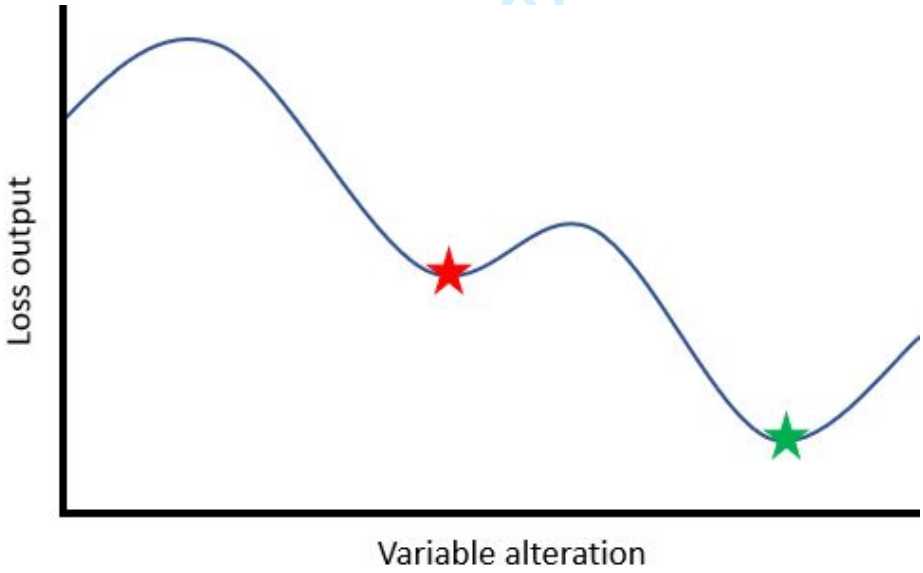


Figure 9 ~~10~~ – The output of the loss function (y-axis) for a single variable is plotted for a range of fabricated alterations (x-axis). A local minima is shown (red star) and the global minima (green star)

To visualize the rate of change, Figure 9~~10~~ shows the fabricated outputs of a loss function for a single factor (blue line), with the gradient of the line indicating the rate of change. The goal

of the optimisation function is to find the global minimum (the lowest output of all the potential losses – Figure 10 green star). Calculating the loss of all weights and biases for all combinations of inputs would be computationally inefficient. Strategies for efficiently determining the optimal configuration is a key topic in machine learning. Initially, the weights and biases were randomly selected and optimisation used to try to determine the global minimum but altering the weights and biases over steps, with step-size decided by the researcher. Care is required when selecting step-size, as local minima (Figure 10 red star) may result in a reduced optimisation. A key factor that effects a neurons loss is the input (feature) entering the neuron itself. In the input layer, the input values are fixed. In the hidden layers however, the weights and biases can be adapted to influence the input into the next level. In a process known as back-propagation, the optimisation of the previous layer (Figure 9 – hidden 2) takes the optimisation of the output layer into account, refining the input to the output layer. The same process happens to hidden layer 1, as the entire model is refined for maximum accuracy.

Different neural networks have been developed to suit different applications. Convolutional neural networks for example are typically for image analysis. Recurrent neural networks have been widely used in signal processing as the input data size is not required to be consistent, whereas unsupervised applications require autoencoders or Restricted Boltzmann machines (combinations of autoencoders). In spectroscopy, the most common forms of neural network are artificial and one-dimensional convolutional neural networks. Convolution uses a kernel to average regions of a spectrum or image. Different kernels may be used to highlight different features, or regions of the spectrums that identify a sample [80]. Convolution converts the image or spectrum into abstractions of the original data, that are typically smaller than the original dataset whilst retaining the key aspects of the data that defines a sample label. Convolutional layers therefore reduce the amount of information being processed in subsequent layers, reducing the computational cost and time [80]. The kind of model used is determined by the kind of data a researcher has. For example, images are typically analyzed with convolutional neural networks, sequences with recurrent neural networks and spectra most commonly are analyzed through feedforward neural networks.

There are numerous other supervised methods (Fig. 1011), with an algorithm for each major sub-section (regression, classification and neural networks) discussed. The main advantages

of directed learning and disadvantage of subjectivity were presented. With a background into how chemometrics and ML techniques are related and the theory into common algorithms, examples of ML algorithms use in life science and medical spectroscopic research will now be discussed.

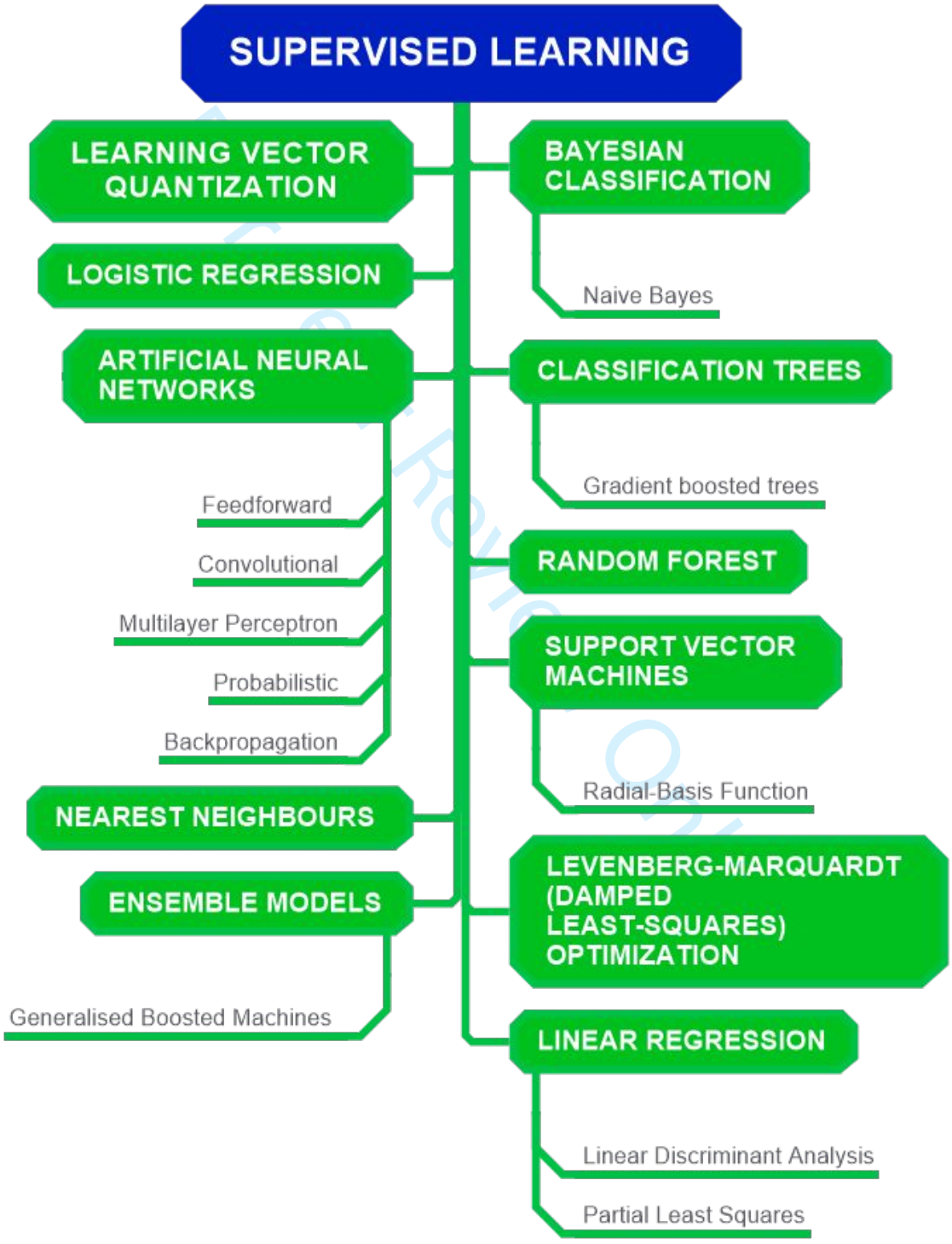


Figure 10 ~~11~~. Supervised learning methods used in spectroscopy techniques.

Applications

Spectroscopy, Machine Learning, and Life Sciences.

ML – Spectroscopy analysis has been extensively explored for a number of years. In the nineties dielectric spectroscopy was paired with neural networks to analyse the metabolite concentrations of different cell suspensions [82]. Dielectric spectroscopy is based on applying a static electric field with low frequency across the plasma membrane, causing amplification of the frequency signal. Employing neural networks to different cell cultures helped to differentiate and compare the metabolism of different cells. Additionally, it is described that artificial neural networks (ANN) analysis from the collected data allows qualitative prediction of the metabolic activity of unknown samples, specifically when cells are processing or metabolising glucose [82]. Different aspects of neural networks, such as node and layer numbers were explored to determine the maximum efficiency, finding that two or more layers actually reduced the accuracy [82]. The authors describe that coupling ANN with this spectroscopic technique can potentially be used to identify organisms as well as their main metabolic status [82].

Similarly, dielectric spectroscopy and electrochemical impedance spectroscopy identifies changes that occur between the interaction of the analyzed solution and the sensor, which has been applied to microorganism studies [83, 84]. Microorganisms suspended in an electrolyte solution makes the microbes, which have different external charges, attach to the electrode [83]. Results from a study of binary mixes of microorganisms studied suggested that basic statistical analysis does not identify and classify data efficiently [83]. However, when the ANN model was implemented to the impedance spectra, using Bayesian Regularization, the overall prediction was of 98.9%. Bayesian regularization is efficient with few data sets [83].

Dahlstrand et al. 2019 describe another significant aspect of ML – spectroscopy application study, using extended-wavelength diffuse reflectance spectroscopy (EWDRS). EWDRS uses a fibreoptic probe to analyse tissues by detecting the reflectance of visible light and near infrared (NIR) with a wavelength number between 400 to 1000 nm, and NIR to short-wave infrared range 1000 – 1700 nm [86, 87]. The study differentiated between five types of porcine

skin tissues by coupling EWDRS and ML methods, support vector machines (SVM) specifically. Implementing PCA-SVM as supervised methods to analyse the data resulted in a 98% overall accuracy, indicating the possibility to build efficient and accurate predictive models for further applications [85].

Both the ML and vibrational spectroscopy have also been used in molecular biology. Vibrational spectroscopy (Raman or Infrared), is characterized by raising the molecular energetic state, due to the absorption of external electromagnetic radiation causing specific vibrations of chemical bonds [88].

Using surface enhanced Raman spectroscopy (SERS) to analyse damaged DNA fixed onto a gold grate, the aim of this study was to classify, identify and predict photo-induced damaged DNA. The research showed that ANN presented positive results with 98% of accuracy in the prediction of damaged DNA. The application of SERS as a non-invasive tool indicates a promising applicability for nucleotide molecules even if small changes occur in the molecular structure [89].

Microbiology has also found ML applications in spectroscopy. Recently, Sharaha *et al.* 2019, analyzed *E. coli* strains resistant to antibiotics by FTIR in combination with machine learning methods as analytical tool. Antibiotics such as, Cotrimoxazole, Piperacillin/tazobactam, Ceftriaxone, and Ceftazidime were identified to not be susceptible against *E. coli* strains. FTIR spectra collection was performed directly from the identified resistant colonies. Support vector machines (SVM) as the side machine learning tool helped to identify and predict susceptible and resistant *E. coli* strains to antibiotics. The test was able to predict the best antibiotic choice above than 89% of sensitivity [90]. In the same year, bacteria responsible for severe food poisoning were studied by Bağcioğlu *et al.* 2019. *Bacillus cereus*, *Bacillus cytotoxicus*, *Bacillus thuringiensis*, *Bacillus mycoides* and *Bacillus weihenstephanensis*, microorganisms complicated to diagnose in ordinal clinical tests could efficiently be differentiated by FTIR spectroscopy and ML methods. methods did not provide a good differentiation of the data between the previous bacterial strains. The ANN model was formulated using the spectral regions between 3100 – 2800 cm^{-1} (fatty acids assignment) and 1800 – 700 cm^{-1} (Lipids, proteins, carbohydrates, assignments). The model identified 99.5%, overall, from the strains used [91].

Another example in which FTIR spectroscopy has been implemented together with ML, can be shown in the study done by Dziuba B., 2013, in which *Propionibacteria* species were analyzed. PCR was used to molecularly separate and identify bacterial strains. The previous allowed a correct identification. As complementary analysis and identification, the study used multilayer perceptrons (MLP) and probabilistic NN (PNN) to analyse the FTIR spectra from Propionibacteria strains. Three different analytical layers constituted the ANN. Absorbance values, wavenumbers selected by a genetic algorithm, and a hidden layer. The ANN algorithm could correctly identify *Propinobacterium* genus (93% of accuracy) at three specific wavelength regions 900 – 600 cm^{-1}, 1200- 900 cm^{-1}, and between 1500 – 1200 cm^{-1} [92]. Similarly, Rebuffo – Scheer *et al.* 2007, demonstrated the efficiency of using PCR-FTIR-ML to study and compare serovars of *Listeria monocytogenes*. The study showed that in the wavenumber region between 1200 – 900 cm^{-1} a carbohydrate peak assignment can be found, furthermore in this region peak absorbance changes, suggesting that serotype identification is linked to the carbohydrate structural changes. ANN analysis provides a bigger identification accuracy for sample the FTIR spectroscopy, giving a 98% of identification accuracy outcome compared to the 95% using PCR [93].

Moreover, it has been found that *Campylobacter coli* and *Campylobacter jejuni*, are bacteria responsible for causing severe acute gastroenterological diseases [94]. Four *Campylobacter* genotypes were isolated and analyzed under FTIR spectroscopy. As a supervised method, MLP and PNN, two types of neural network models, allows to classify and identify data. The analyzed spectral regions were 1200 – 900 cm^{-1} (Region W_4) and 900 – 700 cm^{-1} (region W_5). Four-layer ANN was build using these methods. The best prediction occurred by using MLP on W_4, giving 99.16% of correctly identified microorganisms. 94% for W_5. Whereas, the lowest identification percentage was seen with PNN method in the region W_5, showing 89% of identification [94]. Furthermore, in 2018 a study leaded by Lasch *et al.* also aimed to identify and differentiate pathogenic bacteria. FTIR spectral maps (spectra of some of the *Burkholderia* species is shown in Figure 11~~12~~) of different strains of *B. cenocepacia*, *B. thailandensis, B. caledonica, B. cepacia, B. gladioli, B. vietnamiensis, B. stabilis, and B. glathei* were processed using ANN analysis (Figure 12~~13~~). Every pixel from the maps is equals to a spectral point, hence the pixels giving information regarding the region of interest were subtracted to serve as an input for the ANN model. Resilient back propagation (BP) learning

algorithm was used to train data, as well as Covar feature as selective method. The detection of *Burkholderia* strains was variable, the image prediction accuracies substantially changed, since 90% and 75% of predictive accuracy was obtained, respectively [95].

Wenning *et al.* 2010, also studied microorganisms, lactic acid bacteria (LAB) specifically, using FTIR and ML to identify and recognize these non-pathogenic microorganisms. The analyzed spectral windows between 700 – 1800 cm^{-1} and 2800 – 3000 cm^{-1} serving for data training purposes. In order to validate the accuracy of the ANN test spectra from each of the known species were collected, giving 98% of accuracy. On the other side, to validate the model, 558 spectra of 85 unknown strains were collected and then processed with ANNs. The results gave 93% of accuracy, which is considered as a good prediction [96]. ANN - FTIR spectroscopy conjunction has been used to identify *S. aureus* serotypes of capsulated varieties. The principal aim of the study conducted by Grunert *et al*. 2013, was to differentiate the polysaccharide structure of the capsules from the different *S. aureus* variants. Within the spectral region between 1200 – 800 cm^{-1} polysaccharide peak assignments can be found, mainly C-O-C and C-O-P vibrational stretches. Additionally, between 845 – 810 cm^{-1} is described to be specific of alpha-anomeric composition of carbohydrates. NeuroDeveloper software was used to perform the ANN model. The model run under Rprop algorithm. The combination of ANNs – FTIR provided a 98.2% of accuracy, overall [97].

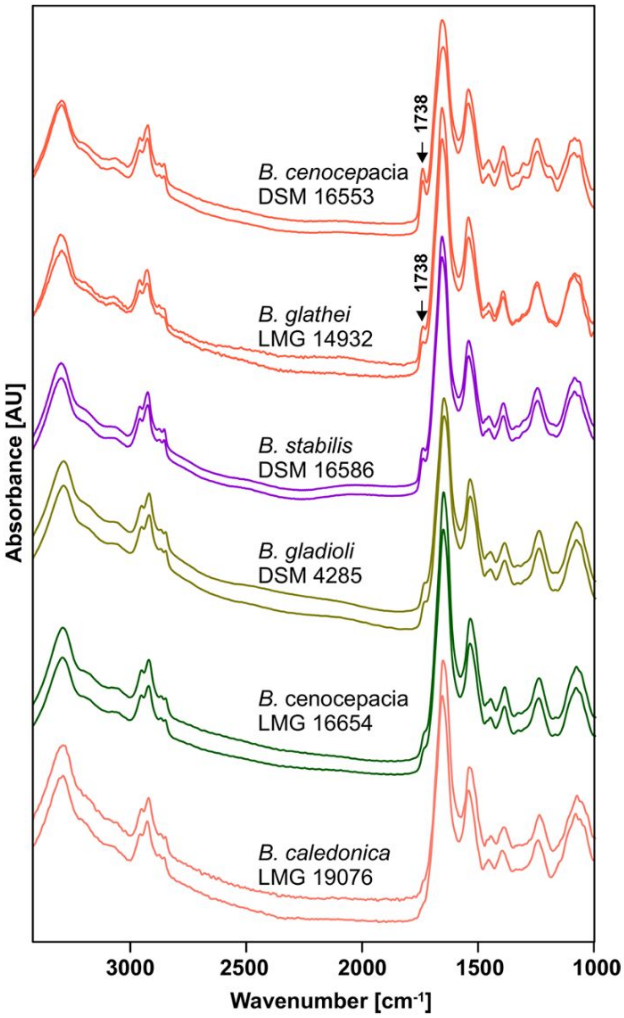


Figure 11~~12~~. FTIR spectrograms of Burkholderia species. A peak localized at 1738 cm^{-1} present in B. cenocepacia, and B. glathei indicates the presence of PHB (Poly beta-hydroxybutyrate), a bio-polyester. PHB interferes with the infrared identification of microorganisms based. Between the 3000-2800 cm^{-1} region its eidentified to belong to C-H stretching, additionally. 1490-1370 cm^{-1} possesses deformation mode of $=CH_2$, and 1200-900 cm^{-1} is characterized for having $-CH_3$ functional groups contributions. [17]

Adapted with permission from reference [95]. Copyright (2018), American Chemical Society

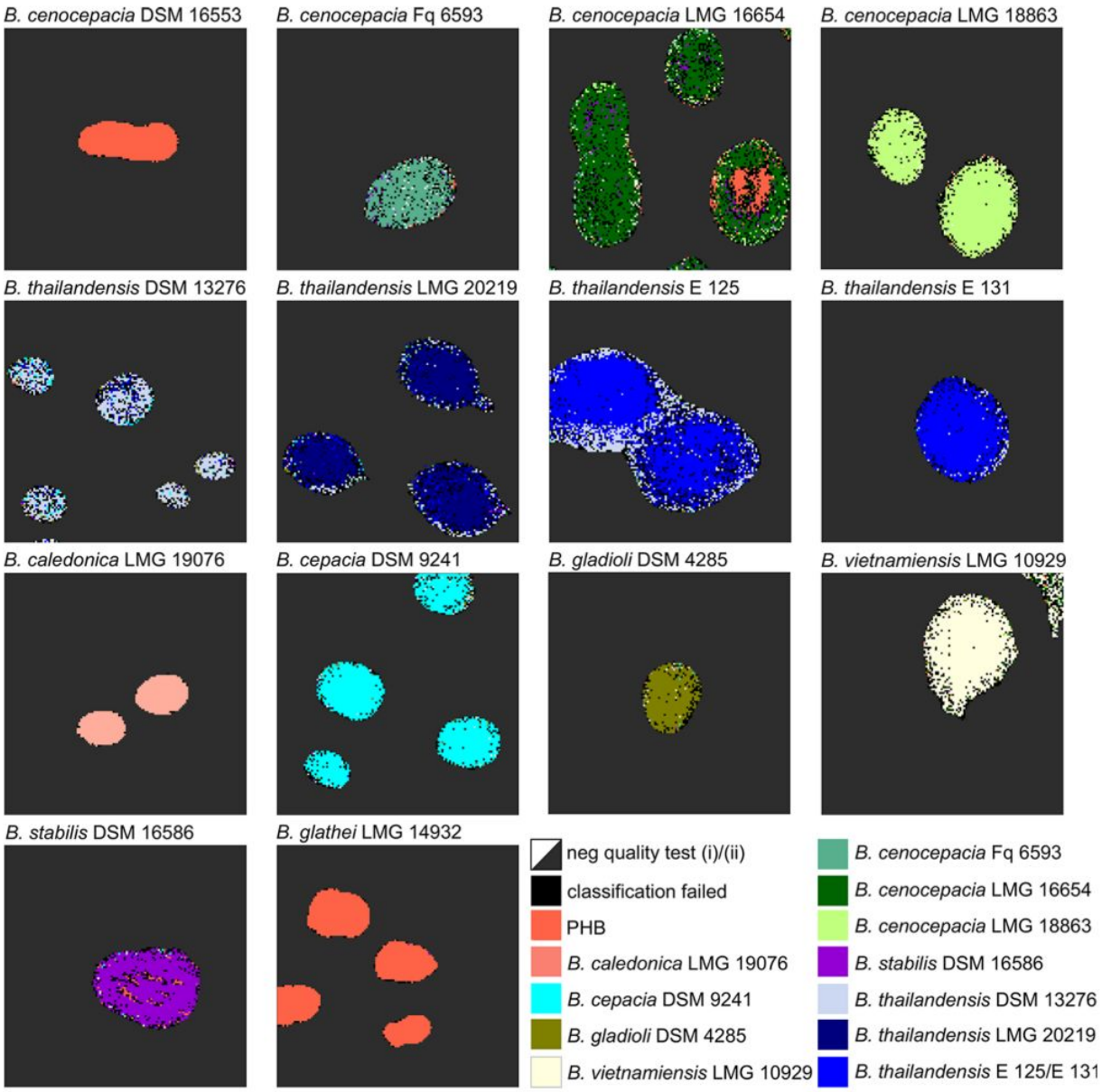


Figure 12 ~~13~~. Infrared imaging data from the ANN prediction of the different Burkholderia species. It can be seen that the ANN algorithm used to predict the microorganism gave 75% - 90% of accuracy. Indicating that spectral imaging together with machine learning algorithms such as ANN can be useful to identify and predict different bacterial species [17].

Adapted with permission from reference [95]. Copyright (2018), American Chemical Society.

Moving from the field of microbiology, but without leaving the life sciences, it is worth mentioning that in pharmacology the use of the spectroscopy-ML set has also been explored. Monoclonal antibodies [98], carbamazepine, nicotinamide, ibuprofen [99], Emtricitabine and Tenofovir alafenamide fumarate [100], active principle ingredients [101], and cephalexin [102]. However, cellular cultures that are used for production of biopharmaceutical drugs have been also evaluated using ANN and spectroscopy [103]. Le *et al.* 2018 analyzed by Raman spectroscopy and ML methods, bevacizumab, infliximab, ramucirumab, and Rituximab,

monoclonal antibodies (mABs), that help to treat cancer. In this study, it was found that the structure of the mABs can be identified along the 1650 – 1300 cm^{-1} region, peaks corresponding to amide I and amide III regions, respectively. Utilising ANN, compared to linear discriminant analysis, the identification error reduces up to 88.3%. Arabzadeh *et al.* 2019 and Takashi *et al.* 2015 utilized UV-vis – ML methods to evaluate different pharmaceutical aspects. In both cases the ANN methods used was feed forward back-propagation learning. Takashi *et al.* 2015 concluded that potential applications can be performed using UV-vis-ML, such as bio-sensors to quantify metabolites, and nutrients. Furthermore, comparing linear discriminant analysis, PLS, genetic algorithms methods to ANNs methods, results to be more efficient in predicting spectral data [98, 99, 101, 102, 103].

Mid-infrared spectroscopy and NIR- spectroscopy [104 - 108], coupled with ML methods, such as k nearest neighbours (kNN) [104, 107], logistic regression (LR) [104, 105], support vector machines (SVM) [103,107], random forests (RF) [104, 107, 108], gradient boosted trees (XGB) [103, 107], Levenberg-Marquardt (damped least-squares) optimization [105], Radial-Basis Function (RBF) [106], Learning Vector Quantization (LVQ) [109] , naïve Bayes (NB) [107], multilayer perceptron (MLP) [107], and Generalized boosted machines (GBM) [108], has been used in other life sciences fields. In the last 3 years, recent studies in fields like botany, zoology, and ecology, have demonstrated that using spectral data together with ML methods is better and more efficient to predict than using linear discriminant analysis [104 - 109].

Spectroscopy and Machine Learning in the Medical field.

Recently, ML methods also have been used in Medicine to enhance and aid the early detection of diseases. Chaber *et al.* 2019 employed FTIR spectroscopy together with ML methods, such as k-nearest neighbour (KNN), support vector machine (SVM), Random forest (RF), Linear discriminant Analysis (LDA), and Gradient Boosted Classifier (GBC), specifically, to predict Erwin Sarcoma. The authors found that predicting infrared spectra (Figure 136) and using SVM gives 92.3% of accuracy in relapsed patients before a chemotherapy. However, in death patients the best accuracy was predicted with Random Forest algorithm (92.3%) after chemotherapy sessions (Table 1) [109]. Similarly, in the prediction of tumours, FTIR micro-spectroscopy coupled with ANNs has been used to analyse and predict glial abnormal growth. ANN algorithm predicted 5% of error the changes that occur among nerve tissue sample, mainly in proteins (random coils, α-helices, β-sheets, and β-turns) [110].

Table 1. Machine learning methods used with their respective prediction accuracy of different sample groups. [106]
Adapted with permission from Molecular Diversity Preservation International.

Dead patients after chemotherapy		
Algorithm employed and accuracy obtained	KNN	69.2%
	SVM	88.5%
	RF	92.3%
	LDA	53.8%
	GBC	76.9%
Relapsed samples previous to chemotherapy		
Algorithm employed and accuracy obtained	KNN	61.5%
	SVM	92.3%
	RF	69.2%
	LDA	69.2%
	GBC	61.5%

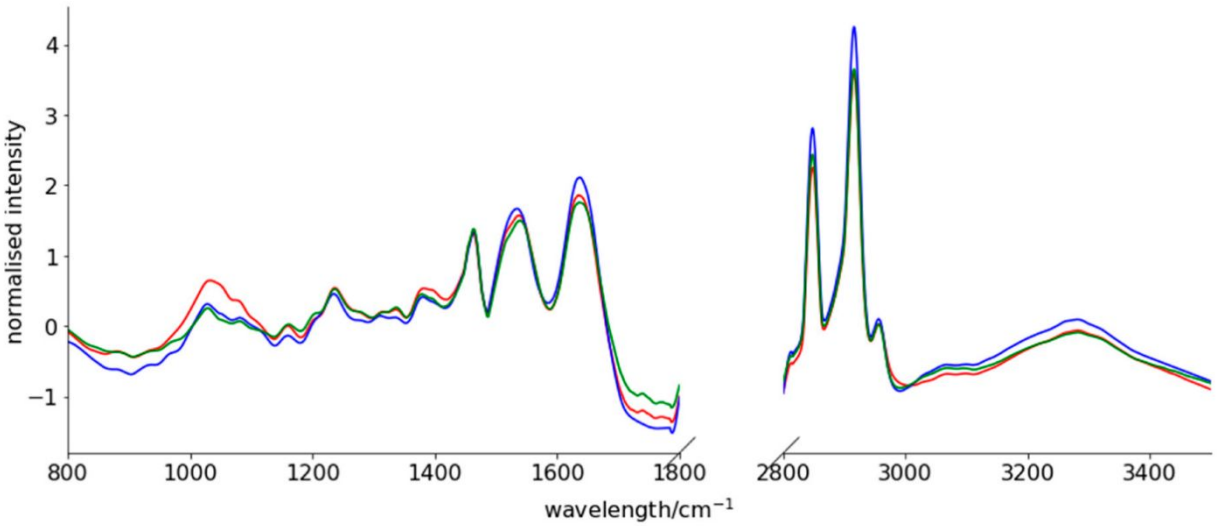


Figure 13 ~~14~~. Red line indicates normal bone tissue IR spectra. Blue line represents the IR spectra from the tissue diagnosed with Erwin Sarcoma (ES). In green, IR spectra of ES tissue after chemotherapy. [31]
Adapted with permission from reference [109], Copyright (2019). Molecular Diversity Preservation International (MDPI).

Other studies related to brain diseases and cancer have explored other spectroscopic techniques such as magnetic resonance [111] and Raman spectroscopy [112, 113], together with ML methods. Jermyn *et al*. 2016, used a portable Raman equipped with a 785 nm lase fibre optic probe to study healthy and cancerous human brain tissue from 177 subjects. To predict and classify the cancer prevalence in humans, ANNs together with boosted trees were

implemented and compared between each other. The best prediction accuracy was obtained using ANNs (90%) while boosted trees algorithm present poor predictive accuracy (71%). Erzina *et al*. 2020, predicted different cancer cell lines using SERS. The prediction was obtained using convolutional neural networks (CNN) of Raman spectroscopy data from the cancerous cell. After 400 iterations the model was able to predict regions of interest in the different substrates where SERS was performed. The predictive accuracy of prediction was of 100% indicating that this algorithm works perfectly for classification and prediction of abnormal samples.

However, Ralbovsky *et al.* 2019 analyzed biofluids (saliva), differently from tissues, saliva is easier to obtain since its acquisition can be classified as a non-invasive method. Saliva was collected from 39 patients and analyzed by Raman spectroscopy. ANNs was applied and validated using latin partition, the implementation of this algorithm helped to differentiate Alzheimer, mild cognitive impairment, and healthy donors. It was possible to differentiate Alzheimer patients with 99.33% of accuracy. The previous described, has shown that spectroscopy techniques coupled with ML methods can efficiently be applied to medicine.

Other medical research fields have also explored using the spectroscopy-machine learning combination to predict and classify different diseases and disorders. Raman spectroscopy and ANNs could classify and predict atherosclerosis with 5% of error. Sixty histological samples of coronary arteries from healthy and affected patients were analyzed [115]. Joint tissue, such as such as cartilage, subchondral bone, cancellous bone and meniscus were analyzed using diffuse reflectance spectroscopy coupled with FLDA (Fourier linear discriminant analysis) and LDA. Using these methods more than 99% of accuracy was achieve for all different tissues [116]. ATR-FTIR spectroscopy coupled with ML methods was applied to cervical cytology samples. Cervical cytology is used to detect abnormal cervix cells or potential cancerous cells, mainly used for HPV screening. The ML methods used, such as eClass, SVM, ANN, k-NN were compared. eClass for this study resulted more efficient as a predictive tool [117]. In addition, Hereditary Haemorrhagic Telangiectasia (HHT), a vascular disorder caused by gene mutations, was studied using mid-infrared spectroscopy and ANNs. The authors collected blood plasma from 202 healthy and diagnosed patients, concluding that the obtained 95 sensitivity and specificity results proved that the study could potentially be applied in a bigger scale [118].

NIR spectroscopy and ANNs was assessed estimation of glucose levels in blood. With a sample size of 50 samples, 94% of accuracy was achieved [119]. Likewise, Guevara *et al*. 2018, used a portable Raman probe (785 nm laser and 90 mW of power) to assess the early detection of diabetes mellitus type 2 (DM2). The Raman probe was positioned in different anatomical points, from healthy and DM2 diagnosed patients (ear lobe, inner arm, thumb nail, and median cubital vein), the spectra collection from the anatomical point were compared with blood samples. Feed-forward ANN and SVM were implemented as ML methods as supervised comparison methods. The best results were implementing ANN algorithm when collecting spectra from the inner arm site, since it was possible by this method to get 96% score of prediction accuracy [120]. On the other hand, blood hyper viscosity identification with 97% of accuracy, and using NIR spectroscopy was achieved by Liu *et al*. 2018.

In the last decade, applications of spectroscopy to medicine combined with machine learning methods have proven to be efficient in predicting diseases and medical disorders. The implementation of statistical algorithms along with analytical tools, such as spectroscopy, may in the not-too-distant future provide patients with a way to promptly diagnose critical medical conditions, thus physicians can prescribe an effective treatment to lessen the impact of the disease.

Conclusion

After taking a deep dive into this subject area, it can be confidently concluded that spectroscopy is an area that together with machine learning can have a significant use to develop better prediction and classification techniques and thus improve everyday processes. Although, it has been reported in multiple articles that artificial neural networks provide better efficiency to perform these tasks. However, the efficiency of prediction with other ML techniques could change according to the approach that the scientists decide to apply. Lastly, multiple software described in Figure 14~~15~~, aid the analysis, bringing a better overview of the multivariable analysis of the data collected. The combination of these techniques is booming, there is much to explore and exploit in this area, and will undoubtedly provide better prospects for possible future implementations and applications in multiple sectors.

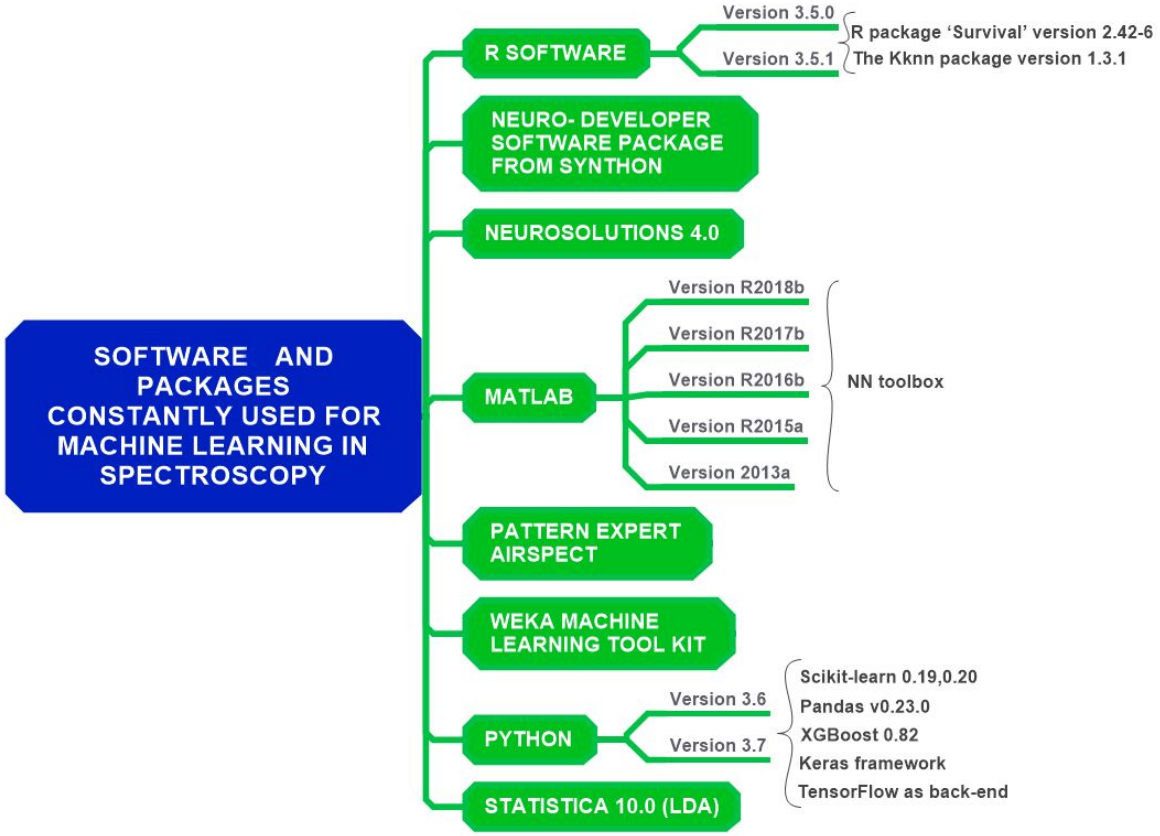


Figure 14 ~~15~~. In the last decade computational sciences has been a powerful ally on the analysis of, identification, classification and prediction of spectral data. This image describes some different software and side tools used for ML applications in spectroscopy.

References

1. D. Bleia, and P. Smyth. Science and data science. Proceedings of the National Academy of Sciences of the United States of America, 114, 8689–86922017 (2015)
2. Feng-hsiung Hsu. IBM'S DEEPBLUECHESS GRANDMASTERCHIPS. IEEE Mico 70-91, 1999.
3. Schrittwieser et al. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. arXiv, 1-21 (2020).
4. Lee et al. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection. Journal of Raman Spectroscopy, 51, 293-300 (2020).
5. Sohn et al. Single-layer multiple-kernel-based convolutional neural network for biological Raman spectral analysis. Journal of Raman Spectroscopy, 51, 414-421 (2020).
6. Liu et al. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. Analyst, 142, 4067–4074 (2017).
7. Aquarelli et al. Convolutional neural networks for vibrational spectroscopic data analysis. Analytica Chimica Acta, 954, 22-31 (2017).
8. Ho et al. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. NATURE COMMUNICATIONS. 10, 1-8 (2019).
9. Umehara et al. Analyzing machine learning models to accelerate generation of fundamental materials insights. Nature Computational Materials, 34, 1-9 (2019).
10. Baldock et al. Single-cell Raman microscopy of microengineered cell scaffolds. Journal of Raman Spectroscopy, 50, 371–379 (2019).
11. Talari et al. Raman spectroscopic analysis differentiates between breast cancer cell lines. Journal of Raman Spectroscopy, 46, 421–42 (2015).
12. Germond et al. Cell type discrimination based on image features of molecular component distribution. Scientific Reports, 8, 1-8 (2018).
13. Passos et al. Spectrochemical analysis in blood plasma combined with subsequent chemometrics for fibromyalgia detection. Scientific Reports, 10, 1-8 (2020).
14. Zúñiga et al. Raman Spectroscopy for Rapid evaluation of Surgical Margins during Breast cancer Lumpectomy. Scientific Reports, 9, 1-16 (2019).

15. Pablo et al. Biochemical fingerprint of colorectal cancer cell lines using label-free live single-cell Raman spectroscopy. Journal of Raman Spectroscopy, 49, 1323–1332 (2018).
16. Zhang et al. A multi-scale approach to study biochemical and biophysical aspects of resveratrol on diesel exhaust particle-human primary lung cell interaction. Scientific Reports. 9, 1-12 (2019) .
17. Kochan et al. Raman spectroscopy as a tool for tracking cyclopropane fatty acids in genetically engineered Saccharomyces cerevisiae. Analyst, 144, 901-912 (2019).
18. Parlatan et al. Raman spectroscopy as a noninvasive diagnostic technique for endometriosis. Scientific Reports, 9, 1-7 (2019).
19. Hunter and Anis. Genetic support vector machines as powerful tools for the analysis of biomedical Raman spectra. Journal of Raman Spectroscopy, 49, 1435–1444 (2018).
20. Saleem et al. Optical diagnosis of hepatitis B virus infection in blood plasma using Raman spectroscopy and chemometric techniques. Journal of Raman Spectroscopy, 51; 1067–1077 (2020).
21. Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. 1–20 (2018). Retrieved from http://arxiv.org/abs/1811.03378
22. El Naqa, I., & Murphy, M. J. What Is Machine Learning? Machine Learning in Radiation Oncology, 3–11 (2015).
23. Shalev-Shwartz, S., & Ben-David, S. Understanding machine learning: From theory to algorithms. In Understanding Machine Learning: From Theory to Algorithms (Vol. 9781107057) (2013).
24. Ren et al. Scalable nanolaminated SERS multiwell cell culture assay. Microsystems & Nanoengineering, 6, 1-11 (2020).
25. Sbroscia et al. Thyroid cancer diagnosis by Raman spectroscopy. Scientific Reports, 10, 1-10 (2020).
26. Horiue et al. Raman spectroscopic signatures of carotenoids and polyenes enable label-free visualization of microbial distributions within pink biofilms. Scientific Reports, 10, 1-10 (2020).
27. Daood et al. A quaternary ammonium silane antimicrobial triggers bacterial membrane and bioflm destruction. Scientific Reports, 10, 1-14 (2020).

28. Bodelón et al. Detection and imaging of quorum sensing in Pseudomonas aeruginosa biofilm communities by surface-enhanced resonance Raman scattering. NATURE MATERIALS, 15, 1203-1213 (2016).
29. Hammoud et al. Raman micro-spectroscopy monitors acquired resistance to targeted cancer therapy at the cellular level. Scientific Reports, 8, 1-11 (2018).
30. Han et al. Rapid antibiotic susceptibility testing of bacteria from patients' blood via assaying bacterial metabolic response with surface-enhanced Raman spectroscopy. Scientific Reports, 10, 1-18 (2020).
31. Vukosavljevic et al. Novel insights into controlled drug release from coated pellets by confocal Raman microscopy. Journal of Raman Spectroscopy, 47, 757–762 (2016).
32. Slipets et al. Volumetric Raman chemical imaging of drug delivery systems. Journal of Raman Spectroscopy, 51, 1153–1159 (2020).
33. Jennifer Sills. Artificial intelligence in research. Science, 357, 28-30 (2017).
34. Pei Wang. On Defining Artificial Intelligence. Journal of Artificial General Intelligence. 10, 1-37 (2019).
35. Silver et al. Mastering the game of Go without human knowledge. Nature, 550, 354-359 (2017).
36. Oriol Vinyals et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 575, 350-354 (2019).
37. Ho. Artificial intelligence in cancer therapy. Science, 367, 982-983 (2020).
38. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., … Hassabis, D. Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706–710 (2020).
39. Fjelland. Why general artificial intelligence will not be realized. Humanities & Social Sciences Communications, 7, 1-9 (2020).
40. Qin, S. J., & Chiang, L. H. Advances and opportunities in machine learning for process data analytics. Computers and Chemical Engineering, 126, 465–473 (2019).
41. Nilsson, N. J. (2005). Mlbook.Pdf (Predmet Application/Pdf). Retrieved from http://ai.stanford.edu/~nilsson/MLBOOK.pdf
42. Sanhueza et al. Raman microimaging as an analytical technique for simultaneous quantification and localization of active principles in pharmaceutical solid dosage forms. Journal of Raman Spectroscopy, 51, 649-659 (2020).

43. Zu et al. Real-time metabolite monitoring of glucose-fed Clostridium acetobutylicum fermentations using Raman assisted metabolomics. Journal of Raman Spectroscopy, 48, 1852–1862 (2020).
44. Marson et al. Simultaneous quantification of artesunate and mefloquine in fixed-dose combination tablets by multivariate calibration with middle infrared spectroscopy and partial least squares regression. Malaria Journal, 15, 1-8 (2016).
45. Liu et al. Raman Spectroscopy in Colorectal Cancer Diagnostics: Comparison of PCA-LDA and PLS-DA Models. Journal of Spectroscopy, 2016, 1-6, (2016).
46. Villa et al. Fast discriminationofbacteria usinga filterpaper–basedSERSplatform and PLS-DA with uncertainty estimation. Analytical and Bioanalytical Chemistry, 411, 705–713 (2019).
47. Rooki. Application of general regression neural network (GRNN) for indirect measuring pressure loss of Herschel–Bulkley drilling fluids in oil drilling. Measurement, 85, 184–191(2016)
48. Modaresi et al. A Comparative Assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for Monthly Streamflow Forecasting in Linear and Nonlinear Conditions. Water Resour Manage, 32, 243–258 (2018).
49. Goodacre, R., & Kel, D. B. Commentary on “rapid identification of streptococcus and enterococcus species using diffuse reflectance-absorbance fourier transform infrared spectroscopy and artificial neural networks.” FEMS Microbiology Letters, 364(10), 1–4 (2017).
50. Paparelle et al. Digitally stimulated Raman passage by deep reinforcement learning. Physics Letters, 384, 1-10 (2020).
51. Yu et al. Classification of pathogens by Raman spectroscopy combined with generative adversarial networks. Science of the Total Environment, 726, 1-9 (2020).
52. S. Yu et al. Classification of pathogens by Raman spectroscopy combined with generative adversarial networks. Science of the Total Environment. 726, 1-9, (2020).
53. Omar et al. Unsupervised clustering for phenotypic stratification of clinical, demographic, and stress attributes of cardiac risk in patients with nonischemic exercise stress echocardiography. Echocardiography, 37, 505–519 (2020).

54. B. Le. Application of deep learning and near infrared spectroscopy in cereal analysis. Vibrational Spectroscopy, 106(October), 1–7 (2019).
55. Chamber et al. Distinguishing Ewing sarcoma and osteomyelitis using FTIR spectroscopy. SCIENTIFIC REPORTS, 8, 1-8 (2018).
56. Shlens. A Tutorial on Principal Component Analysis. arXia, 1-12 (2014).
57. Alfonso-García et al. A machine learning framework to analyse hyperspectral stimulated Raman scattering microscopy images of expressed human meibum. Journal of Raman Spectroscopy, 48, 803–812 (2017).
58. Kopec. Monitoring glycosylation metabolism in brain and breast cancer by Raman imaging. SCIENTIFIC REPORTS, 9, 1-13 (2019) .
59. Chih-Wei Hsu., et al. A Practical Guide to Support Vector Classification, last updated: May 19, 2016, accessed July 22, 2020.
60. B. Schölkopf., et al. New support vector algorithms. Neural Computation, 12, 1207-1245 (2000).
61. Williams, C. Support Vector Machines. (October) (2008).
62. Pedregosa et al. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12(85):2825−2830 (2011).
63. Belousov, A. I., et al. Applicational aspects of support vector machines. Journal of Chemometrics, 16(8–10), 482–489 (2002).
64. Nussbaum, Daniel A ; Mislick, Gregory K. “Chapter 7, Linear Regression Analysis – section 7.3, Linear Regression Analysis”, Cost Estimation, Wiley Series in Operations Research and Management Science Ser, (1st ed.), John Wiley & Sons, p.126, ISBN: 9781118536131
65. Rencher, Alvin C.; Christensen, William F., "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 340, ISBN 9781118391679 (2012).
66. Vigneau et al. Application of latent root regression for calibration in near-infrared spectroscopy. Comparison with principal component regression and partial least squares. Chemometrics and Intelligent Laboratory Systems, 35, 231-238 (1996).
67. Gonzales et al. Two- and three-dimensional quantitative structure–permeability relationship of flavonoids in Caco-2 cells using stepwise multiple linear regression

(SMLR), partial least squares regression (PLSR), and pharmacophore (GALAHAD)-based comparative molecular similarity index analysis (COMSIA). Medicinal Chemical Research, 24, 1696–1706 (2015).

68. HU et al. Rapid determination of the texture properties of cooked cereals using nearinfrared reflectance spectroscopy. Infrared Physics and Technology, 94, 165–172 (2018).
69. Liu et al. Comparison of prediction power of three multivariate calibrations for estimation of leaf anthocyanin content with visible spectroscopy in Prunus cerasifera. PeerJ, 1-19 (2019).
70. Suryakala and Prince. Investigation of goodness of model data ft using PLSR and PCR regression models to determine informative wavelength band in NIR region for non-invasive blood glucose prediction. Optical and Quantum Electronics, 51:271, 1-20 (2019).
71. Wold et al. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58, 109–130 (2001).
72. Sperandei, S. Understanding logistic regression analysis. Biochemia Medica, 24(1), 12–18 (2014).
73. Laporte, F., Dambre, J., & Bienstman, P. Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework PyTorch. Scientific Reports, 9(1), 1–9 (2019).
74. L. Streun, G., P. Elmiger, M., Dobay, A., Ebert, L., & Kraemer, T. A machine learning approach for handling big data produced by high resolution mass spectrometry after data independent acquisition of small molecules - Proof of concept study using an artificial neural network for sample classification. Drug Test Anal, 12, 836–845 (2020).
75. García-Roselló, E., González-Dacosta, J., Lado, M. J., Méndez, A. J., Pérez-Schofield, B. G., & Ferrer, F. Visual NNet: An educational ANN's simulation environment reusing Matlab neural networks toolbox. Informatics in Education, 10(2), 225–232 (2011).
76. Steppa, C., & Holch, T. L. HexagDLy—Processing hexagonally sampled data with CNNs in PyTorch. SoftwareX, 9, 193–198 (2019).
77. Park, Y. J., Bae, J. H., Shin, M. H., Hyun, S. H., Cho, Y. S., Choe, Y. S., … Moon, S. H. Development of Predictive Models in Patients with Epiphora Using Lacrimal

Scintigraphy and Machine Learning. Nuclear Medicine and Molecular Imaging, 53(2), 125–135 (2019).

78. X. Shao et al. Deep convolutional neural networks combine Raman spectral signature of serum for prostate cancer bone metastases screening. Nanomedicine: Nanotechnology, Biology, and Medicine, 29, 1-7 (2020).
79. Liu et al. Deep Convolutional Neural Networks for Raman Spectrum Recognition : A Unified Solution. arXia. 1-14 (2017).
80. Xu et al. Empirical Evaluation of Rectified Activations in Convolution Network. arXiv, 1-5 (2015).
81. Gu et al. Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377 (2018) .
82. Das & Chaudhuri. On the Separability of Classes with the Cross-Entropy Loss Function. arXiv, 1-19 (2019).
83. Woodward, A. M., Jones, A., Zhang, X. Z., Rowland, J., & Kell, D. B. Rapid and non-invasive quantification of metabolic substrates in biological cell suspensions using non-linear dielectric spectroscopy with multivariate calibration and artificial neural networks. Principles and applications. Bioelectrochemistry and Bioenergetics, 40(2), 99–132 (1996).
84. Muñoz-Berbel, X., Vigués, N., Mas, J., del Valle, M., Muñoz, F. J., & Cortina-Puig, M. Resolution of binary mixtures of microorganisms using electrochemical impedance spectroscopy and artificial neural networks. Biosensors and Bioelectronics, 24(4), 958–962 (2008).
85. Randviir, E. P., & Banks, C. E. Electrochemical impedance spectroscopy: An overview of bioanalytical applications. Analytical Methods, 5(5), 1098–1115 (2013).
86. Dahlstrand, U., Sheikh, R., Dybelius Ansson, C., Memarzadeh, K., Reistad, N., & Malmsjö, M. Extended-wavelength diffuse reflectance spectroscopy with a machine-learning method for in vivo tissue classification. PloS One, 14(10), e0223682 (2019).
87. Frei, R. W. Diffuse Reflectance Spectroscopy; Applications, Standards, and Calibration (With Special Reference To Chromatography). J Res Natl Bur Stand Sect A Phys Chem, 80 A(4), 551–565 (1976).
88. Rehman, I. ur, Movasaghi, Z., & Rehman, S. (2012). Vibrational Spectroscopy for Tissue Analysis. Boca Raton: CRC Press, 2013.

89. Guselnikova, O., Trelin, A., Skvortsova, A., Ulbrich, P., Postnikov, P., Pershina, A., … Lyutakov, O. Label-free surface-enhanced Raman spectroscopy with artificial neural network technique for recognition photoinduced DNA damage. Biosensors and Bioelectronics, 145(June), 111718 (2019).
90. Sharaha, U., Rodriguez-Diaz, E., Sagi, O., Riesenberg, K., Salman, A., Bigio, I. J., & Huleihel, M. Fast and reliable determination of Escherichia coli susceptibility to antibiotics: Infrared microscopy in tandem with machine learning algorithms. Journal of Biophotonics, 12(7), 1–9 (2019).
91. Bağcıoğlu, M., Fricker, M., Johler, S., & Ehling-Schulz, M. Detection and identification of bacillus cereus, bacillus cytotoxicus, bacillus thuringiensis, bacillus mycoides and bacillus weihenstephanensis via machine learning based FTIR spectroscopy. Frontiers in Microbiology, 10(APR), 1–10 (2019).
92. Dziuba, B. Identification of Propionibacteria to the species level using Fourier transform infrared spectroscopy and artificial neural networks. Polish Journal of Veterinary Sciences, 16(2), 351–357 (2013).
93. Rebuffo-Scheer, C. A., Schmitt, J., & Scherer, S. Differentiation of Listeria monocytogenes serovars by using artificial neural network analysis of fourier-transformed infrared spectra. Applied and Environmental Microbiology, 73(3), 1036–1040 (2007).
94. Mouwen, D. J. M., Capita, R., Alonso-Calleja, C., Prieto-Gómez, J., & Prieto, M. Artificial neural network based identification of Campylobacter species by Fourier transform infrared spectroscopy. Journal of Microbiological Methods, 67(1), 131–140 (2006).
95. Lasch, P., Stämmler, M., Zhang, M., Baranska, M., Bosch, A., & Majzner, K. (2018). FT-IR Hyperspectral Imaging and Artificial Neural Network Analysis for Identification of Pathogenic Bacteria. Analytical Chemistry, 90(15), 8896–8904 (2018).
96. Wenning, M., Büchl, N. R., & Scherer, S. Species and strain identification of lactic acid bacteria using FTIR spectroscopy and artificial neural networks. Journal of Biophotonics, 3(8–9), 493–505 (2010).
97. Grunert, T., Wenning, M., Barbagelata, M. S., Fricker, M., Sordelli, D. O., Buzzola, F. R., & Ehling-Schulz, M. Rapid and reliable identification of Staphylococcus aureus capsular serotypes by means of artificial neural network-assisted fourier transform infrared spectroscopy. Journal of Clinical Microbiology, 51(7), 2261–2266 (2013).

98. Le, L. M. M., Kégl, B., Gramfort, A., Marini, C., Nguyen, D., Cherti, M., … Caudron, E. (2018). Optimization of classification and regression analysis of four monoclonal antibodies from Raman spectra using collaborative machine learning approach. Talanta, 184(October 2017), 260–265 (2018).

99. Barmpalexis, P., Karagianni, A., Nikolakakis, I., & Kachrimanis, K. Artificial neural networks (ANNs) and partial least squares (PLS) regression in the quantitative analysis of cocrystal formulations by Raman and ATR-FTIR spectroscopy. Journal of Pharmaceutical and Biomedical Analysis, 158, 214–224 (2018).

100. Arabzadeh, V., Sohrabi, M. R., Goudarzi, N., & Davallo, M. Using artificial neural network and multivariate calibration methods for simultaneous spectrophotometric analysis of Emtricitabine and Tenofovir alafenamide fumarate in pharmaceutical formulation of HIV drug. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 215, 266–275 (2019).

101. Takahashi, M. B. eatri., Leme, J., Caricati, C. P. ereir., Tonso, A., Fernández Núñez, E. G. ustav., & Rocha, J. C. Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes. Bioprocess and Biosystems Engineering, 38(6), 1045–1054 (2015).

102. Chalus, P., Walter, S., & Ulmschneider, M. Combined wavelet transform-artificial neural network use in tablet active content determination by near-infrared spectroscopy. Analytica Chimica Acta, 591(2), 219–224 (2007).

103. Huan, Y., Feng, G., Wang, B., Ren, Y., & Fei, Q. Quantitative analysis of cefalexin based on artificial neural networks combined with modified genetic algorithm using short near-infrared spectroscopy. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 109, 308–312 (2013).

104. González Jiménez, M., Babayan, S. A., Khazaeli, P., Doyle, M., Walton, F., Reedy, E., Wynne, K. Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. Wellcome Open Research, 4, 76 (2019).

105. Milali, M. P., Sikulu-Lord, M. T., Kiware, S. S., Dowell, F. E., Corliss, G. F., & Povinelli, R. J. Age grading An. gambiae and An. arabiensis using near infrared spectra and artificial neural networks. Plos One, 14(8), e0209451 (2019).

106. Golhani, K., Balasundram, S. K., Vadamalai, G., & Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. Information Processing in Agriculture, 5(3), 354–371 (2018).

107. Mwanga, E. P., Mapua, S. A., Siria, D. J., Ngowo, H. S., Nangacha, F., Mgando, J., … Okumu, F. O. Using mid-infrared spectroscopy and supervised machine-learning to identify vertebrate blood meals in the malaria vector, Anopheles arabiensis. Malaria Journal, 18(1), 1–9 (2019).

108. Nawar, S., & Mouazen, A. M. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. Sensors (Switzerland), 17(10), 1–22 (2017).

109. Chaber, R., Arthur, C. J., Łach, K., Raciborska, A., Michalak, E., Bilska, K., … Cebulski, J. Predicting Ewing sarcoma treatment outcome using infrared spectroscopy and machine learning. Molecules, 1–12 (2019).

110. Surowka, A. D., Adamek, D., & Szczerbowska-Boruchowska, M. The combination of artificial neural networks and synchrotron radiation-based infrared micro-spectroscopy for a study on the protein composition of human glial tumors. Analyst, 140(7), 2428–2438 (2015).

111. Arizmendi, C., Hernández-Tamames, J., Romero, E., Vellido, A., & Del Pozo, F. (2010). Diagnosis of brain tumours from magnetic resonance spectroscopy using wavelets and Neural Networks. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10, (1), 6074–6077 (2010)

112. Jermyn, M., Desroches, J., Mercier, J., Tremblay, M.-A., St-Arnaud, K., Guiot, M.-C., … Leblond, F. Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts. Journal of Biomedical Optics, 21(9), 094002 (2016).

113. Erzina, M., Trelin, A., Guselnikova, O., Dvorankova, B., Strnadova, K., Perminova, A., … Lyutakov, O. Precise cancer detection via the combination of functionalized SERS surfaces and convolutional neural network with independent inputs. Sensors and Actuators, B: Chemical (2020).

114. Ralbovsky, N. M., Halámková, L., Wall, K., Anderson-Hanley, C., & Lednev, I. K. Screening for Alzheimer's Disease Using Saliva: A New Approach Based on Machine

Learning and Raman Hyperspectroscopy. Journal of Alzheimer's Disease : JAD, 1–9 (2019).

115. De Paula, A. R., & Sathaiah, S. Raman spectroscopy for diagnosis of atherosclerosis: A rapid analysis using neural networks. Medical Engineering and Physics, 27(3), 237–244 (2005).
116. Gunaratne, Gonzalez Viejo, Gunaratne, Torrico, Dunshea, & Fuentes. Chocolate Quality Assessment Based on Chemical Fingerprinting Using Near Infra-red and Machine Learning Modeling. Foods, 8(10), 426 (2019).
117. Kelly, J. G., Angelov, P. P., Trevisan, J., Vlachopoulou, A., Paraskevaidis, E., Martin-Hirsch, P. L., & Martin, F. L. Robust classification of low-grade cervical cytology following analysis with ATR-FTIR spectroscopy and subsequent application of self-learning classifier eClass. Analytical and Bioanalytical Chemistry, 398(5), 2191–2201 (2010).
118. Lux, A., Müller, R., Tulk, M., Olivieri, C., Zarrabeita, R., Salonikios, T., & Wirnitzer, B. HHT diagnosis by Mid-infrared spectroscopy and artificial neural network analysis. Orphanet Journal of Rare Diseases, 8(1), 1–15 (2013).
119. Ramasahayam, S., Koppuravuri, S. H., Arora, L., & Chowdhury, S. R. Noninvasive Blood Glucose Sensing Using Near Infra-Red Spectroscopy and Artificial Neural Networks Based on Inverse Delayed Function Model of Neuron. Journal of Medical Systems, 39(1), 1–15 (2015).
120. Guevara, E., Torres-Galván, J. C., Ramírez-Elías, M. G., Luevano-Contreras, C., & González, F. J. (2018). Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools. Biomedical Optics Express, 9(10), 4998. https://doi.org/10.1364/boe.9.004998
121. Liu, M., Zhao, J., Lu, X. Z., Li, G., Wu, T., & Zhang, L. F. (2018). Blood hyperviscosity identification with reflective spectroscopy of tongue tip based on principal component analysis combining artificial neural network. BioMedical Engineering Online, 17(1), 1–12. https://doi.org/10.1186/s12938-018-0495-3