Trends in Analytical Chemistry 169 (2023) 117311

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Will data analytics revolution finally bring SERS to the clinic?

Ahmed Y.F. Mahmoud [a], Alexandra Teixeira [a,b,c], Marta Aranda [a,d], Maria S. Relvas [a,e], Sergio Quintero [a,f], Maria Sousa-Silva [a,n], Alexandre Chícharo [a], Mengkun Chen [g], Mohadeseh Hashemi [g], Jason B. King [g], James W. Tunnell [g], Carlo Morasso [h], Francesca Piccotti [h], Fabio Corsi [h,i], Malou Henriksen-Lacey [j,k], Dorleta Jimenez de Aberasturi [j,k,l], David Méndez-Merino [m], Alfonso Rodríguez-Patón [m], Sara Abalde-Cela [a,n,*], Lorena Diéguez [a,n,**]

[a] International Iberian Nanotechnology Laboratory (INL), Avenida Mestre José Veiga, 4715-310, Braga, Portugal
[b] Life and Health Sciences Research Institute (ICVS), Escola de Medicina, Universidade do Minho, Campus Gualtar, 4710-057, Braga, Portugal
[c] ICVS - PT Government Associate Laboratory, Braga, Guimarães, Portugal
[d] Colloid Chemistry Group, Centro de Investigaciones Biomédicas (CINBIO), University of Vigo, Spain
[e] Center for Research in Biological Chemistry and Molecular Materials (CiQUS, University of Santiago de Compostela, Spain
[f] Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus de Montegancedo, Pozuelo de Alarcón, 28223, Madrid, Spain
[g] Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, USA
[h] Laboratory of Nanomedicine, Istituti Clinici Scientifici Maugeri IRCCS, Pavia, Italy
[i] Department of Biomedical and Clinical Sciences, Università degli studi di Milano, Milano, Italy
[j] CIC biomaGUNE, Basque Research and Technology Alliance (BRTA), 20014, Donostia-San Sebastián, Spain
[k] Centro de Investigación Biomédica en Red, Bioingeniería, Biomateriales, y Nanomedicina (CIBER-BBN), 20014, Donostia-San Sebastián, Spain
[l] Ikerbasque, Basque Foundation for Science, 48009, Bilbao, Spain
[m] Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Spain
[n] RUBYnanomed LDA, Praça Conde de Agrolongo, 4700-312, Braga, Portugal

ARTICLE INFO

Keywords:
Surface-enhanced Raman scattering
SERS
Data analytics
Machine learning
Artificial intelligence
Healthcare
Clinic

ABSTRACT

Since its discovery, surface-enhanced Raman scattering has shown to be an ultrasensitive analytical tool. The technique was revolutionized with advances in micro/nanofabrication techniques. This resulted in the expansion of SERS applications in biosensing. However, SERS is still facing many challenges to be translated into the clinic. In this review, we will discuss how data analytics algorithms may be used to democratize the applications of SERS in translational medicine. This combination has the potential to create a new dimension of SERS in translational medicine to unlock the SERS access to the clinic, offering high precision and accuracy in evaluating or monitoring health and disease states. We believe that AI/ML may open new avenues towards the benchmarking of SERS (bio)sensors and their applications in the medical field.

1. Introduction

The development of sensitive, robust, fast, and inexpensive molecular biosensors is an active ongoing research field, particularly during and after the era of the global pandemic of COVID-19. Traditional diagnostic tools are complex, expensive and fail to address patient heterogeneity. Current needs demand cheap, fast, early and personalised diagnosis, and decentralized testing. In particular, regarding patient diversity and diagnostic personalisation, current efforts are focused on the identification of quantitative biomarkers and their application as part of larger panels that ensure specificity and sensitivity, and that lead to improved patient stratification and continuous minimally-invasive monitoring [1].

Surface-enhanced Raman scattering (SERS) offers many advantages over other analytical techniques, such as improved detection limits and multiplexing capabilities. Some artificial intelligence (AI) and machine learning (ML) methods have been widely applied to promote a more detailed and accurate classification of samples by analysing their data

* Corresponding author. International Iberian Nanotechnology Laboratory (INL), Avenida Mestre José Veiga, 4715-310, Braga, Portugal.
** Corresponding author. International Iberian Nanotechnology Laboratory (INL), Avenida Mestre José Veiga, 4715-310, Braga, Portugal.
E-mail addresses: sara.abalde@inl.int (S. Abalde-Cela), lorena.dieguez@inl.int (L. Diéguez).

https://doi.org/10.1016/j.trac.2023.117311
Received 14 July 2022; Received in revised form 19 September 2023; Accepted 20 September 2023
Available online 12 October 2023
0165-9936/© 2023 Published by Elsevier B.V.

previously obtained via SERS [2]. Such methods enabled the classification of fungal cells [3], small molecules [4], infectious disease diagnosis [5], antiviral drugs and protein antigens [6,7] and bacterial cells [8], mammalian cells [9], and extracellular vesicles [10].

SERS is a stable, non-destructive technique and, not only improves on sensitivity and specificity of current tools but, due to the latest advances in manufacturing and nanotechnology, also improves on reliability and robustness. Also, the multiplexing capabilities of SERS enables quantitative analysis of a large panel of biomarkers simultaneously, despite using a single excitation wavelength. Yet, to obtain meaningful information, it is necessary to analyse a large amount of complex data. This is particularly challenging when analysing clinical samples, with high complexity and variability [11].

Despite all their advantages, SERS biosensors are still confined to research labs with sophisticated Raman systems and are not widespread among clinicians [12]. The struggle of SERS in translational medicine and its transfer into clinics may be based on many factors, including, but not limited to, the cost and reproducibility of SERS substrates and the lack of standardization and benchmarking of SERS biosensing platforms [13]. In any SERS biosensing platform, there are a lot of variables that may impact the biosensor performance due to variations in the SERS spectral intensity. These variables can be related to three major SERS components: the SERS substrate, the Raman system, and the analyte/s of interest [13]. The main focus of this review is to showcase some examples on how data analytics can bring Raman and SERS closer to clinical applications. The reader is advised to read these seminal reviews for better understanding about SERS in general, its enhancement mechanisms, different fabrication strategies of SERS substrates, and how SERS performance can be dramatically influenced by the nanostructures types, sizes, shapes, and the nanogaps between them [14–17]. Indeed, the field of designing and developing SERS substrates has been advanced in the past decade by innovative synthetic strategies of different nanostructures, nanocomposites, and well-engineered plasmonic nanogaps [14,15].

We summarized a "wishlist" of characteristics of an ideal SERS substrate based on the recommendations of Michael Natan [16] in a seminal Faraday Discussion article and a comprehensive review of Pilot et al. [18], as follows: 1) easy to fabricate and have a relative standard deviation of no more than 20% of signal intensity from spot-to-spot within the same substrate, and from substrate-to-substrate; 2) enhance the Raman spectral intensity to offer high sensitivity and low detection limit; 3) applicable to various kinds of samples whether they adsorb strongly or poorly to the substrate; 4) stable over a long period of time; and 5) inexpensive [16,18]. Additionally, the complexity of biological matrices has to be taken into account during the initial design of SERS substrates to avoid non-specific interactions and increase selectivity and specificity. However, it is well recognized that to find a SERS substrate with all of the aforementioned characteristics and to be a universal substrate for any application is very difficult [18]. Moreover, variations in Raman equipment have highlighted the complications that arise when attempting to compare the same SERS substrates between different labs [13]. Most SERS publications show the SERS performance of the substrate, using stable Raman probes (aromatic thiols and/or Raman dyes) with high adsorption affinity to the plasmonic nanostructures and with large Raman cross-sections [13,16]. Unfortunately, this is not the general case for clinical samples where molecules have poor adsorption affinity to the substrate and small Raman cross-sections. Furthermore, sample pre-treatment of real samples can be crucial in developing efficient SERS bioassays, like any other bioassays. Sample pre-treatment should take into consideration the matrix complexity of real biological samples (e.g blood, cells, tissues, etc), and how this complex matrix may have different components with higher concentration than the analyte (s) of interest. If the sample pre-treatment was not efficient to deal with this complexity, non-specific adsorption and interferences may result in from the matrix components, leading to a SERS assay with poor performance [19]. For example, Porter's group showed how serum sample pre-treatment can improve the detection limit by 2–3 orders of magnitude in the development of SERS immunoassay for tuberculosis [20]. However, the pre-treatment step should be fast, simple, and efficient to keep minimal steps involved in the SERS bioassay to keep it amenable to point-of-care applications [19].

Compared to well-established bioanalytical techniques, such as enzyme-linked immunosorbent assay, chromatographic, and mass spectroscopic techniques, there is a lack of standard operating procedures and good laboratory practices regarding how SERS can be implemented for label or label-free bioassays [13]. Commercial SERS substrate companies provide general guidelines on how to use their products, such as the sample volume to be used, laser wavelength, power, and acquisition time, which may or may not be suitable for actual clinical samples. These guidelines were primarily made by using analytes with large Raman cross sections and high-to-moderate adsorption affinities to their SERS substrates for qualitative or semi-quantitative purposes. In addition, there is no agreement in the SERS research community on one or two commercial "benchmark" SERS substrates, which different labs can use to compare and/or validate the performance of their SERS platform/bioassay. Subsequently, there is a need for global collaborations between different labs to start a mission for the standardisation and benchmarking of SERS assays [13]. We believe that merging AI algorithms and SERS can accelerate and optimize this global mission for harmonizing standardisation and benchmarking of SERS biosensing platforms.

2. Spectral data analysis approaches

Clinical decision-support systems based on spectral data require spectral analysis algorithms to interpret acquired spectra and provide a binary classification of a task (e.g., cancer vs. no cancer). Spectral data analysis algorithms generally fall into two main approaches: biophysical and machine learning (ML) approaches. Biophysical approaches directly extract information from spectra (e.g. biological, biochemical, optical information, etc.), then build classification models based on those extracted metrics. ML approaches build models directly on the spectra without extracting any intermediate biophysical information. The final task for both approaches is to build diagnostic models based on either the extracted information or the original spectra. Therefore, the spectral analysis methods vary mainly in their information extraction processes. Here, both biophysical spectral analysis methods and ML based methods will be briefly introduced.

2.1. *Biophysical approaches*

Biophysical approaches build models based upon the physical mechanisms of light propagation and seek to directly determine the underlying biological and physical composition of the tissue and/or sample. Common spectroscopic approaches such as diffuse reflectance spectroscopy (DRS) and Raman spectroscopy utilize combinations of analytical models (e.g. optical diffusion) or numerical techniques (e.g. Monte Carlo simulation). Early examples of these approaches focused on extracting optical properties and detecting single molecules from various spectroscopic modalities, including Raman spectroscopy [21], DRS [22], polarized light scattering spectroscopy [23], and fluorescence spectroscopy [24].

Most recently, inverse Monte Carlo methods remain the primary numerical technique used for spectroscopic analysis methods. Various forms of inverse models have been developed to efficiently extract optical properties (absorption and scattering) from acquired spectra. Those optical properties are then used to determine the biophysical features that distinguish them, such as blood content, blood oxygenation, melanin content, lipid concentration, and microstructural (scattering) density. For example, a Monte Carlo look-up-table inverse model was used on DRS both non-clinically [25] and preclinically [26] for applications in detecting skin cancer noninvasively. The Monte Carlo

approach allows the simulation of more complex tissue structures, such as multi-layer tissues [27] and thick tissues [28]. A notably powerful technique applied to Raman spectroscopy is to extract the biochemical composition of a tissue sample by breaking down the spectra into sub-spectra of different biochemical components [29]. Raman spectroscopy discriminates a plurality of soft tissue biochemical components, including structural proteins (collagen, elastin, keratin), DNA content, and lipids (triolein, ceramide) [29]. These examples highlight biophysical approaches applied to Raman and diffuse reflectance spectroscopy; however, biophysical approaches have also been applied to light scattering [23] (e.g. enhanced backscattering spectroscopy, low coherence interferometry) and fluorescence/life fluorescence spectroscopy [24]. These approaches can extract nanostructure metrics, cell nuclei density, and molecular composition, and they are often implemented in an imaging mode to map these features over surfaces.

Biophysical models are widely used because they provide biological and physical information that can be directly related to diseases. However, the models employed require an additional layer of computational complexity that can be a barrier to their use. Often, experimental artifacts cannot be properly accounted for or maybe unknown, complicating the extraction of the biophysical parameters. Thus, ML-based approaches offer a complementary approach to such tasks.

2.2. *Machine learning approaches*

ML approaches seek to categorize spectral data based upon their inherent statistical variances, without directly determining the underlying biological or physical mechanism for the variances. Spectral analysis using ML often uses dimension reduction because the spectral data is represented by thousands of points that are reduced to a few parameters that describe the primary variances of a spectrum. Classification models are based on the statistically extracted features just as used on the biophysically extracted features. One of the most widely used dimension reduction techniques when analyzing spectra is principal component analysis (PCA) [30,31]. Other dimension reduction algorithms like Partial Least Squares- Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA) are also used widely. Another direction of using ML approaches is to build the classification models directly on the original spectra. When the number of spectra is sufficient, this is often fulfilled by using artificial neural networks (ANN) because of their ability to create complex classification boundaries.

ML approaches have been used since the early uses of spectral diagnosis of tissues [32,33]. However, recent uses of ML for spectral diagnosis have been accelerated, demonstrating gains in classification accuracy. Commonly used ML models for spectral classification include techniques such as support vector machine (SVM), k-nearest neighbor (KNN), Random Forest (RF), Gradient Boosted Classifier (GBC), and LDA [34]. These classification models show outstanding performance when the feature numbers are small. Large numbers of features introduce more variance and increase the risk of overfitting to those models. ANN can be used for classification with large numbers of features and even with the original spectra without dimension reduction [35].

It is possible to find in the literature several examples in which by applying data analytics is possible to enhance the performance of chemical sensing systems. For example [36], Holgado and co-workers published a comparative study applying two different data processing methodologies over the same set of experimental data obtained from an optical interferometric sensor: one based only on monitoring the sensor resonance mode, and the other applying PCA for obtaining the sensor metrics. The same regression techniques were applied to the sensor metrics over the different methodologies obtaining an improvement on the limit of detection of 140 times with the ML approach. This case study application was in the agricultural field, but other examples can also be found in the biological field. Pandit et al. [37]presented a comparative study applied to protein detection using fluorescence with surface-functionalized carbon dots. The signal produced by the combination of the carbon dots and eight different proteins in different circumstances was analysed by LDA and different ML algorithms, obtaining better performance with the latest, especially SVM, LR, KNN, and GBC. Another example on how the value of data can be augmented by applying different forms of data analytics to a problem is the comparative study by Hong et al. [38]. Here the authors applied PCA and SVM over the same experimental data obtained from SERS signals from the serum of different patients with colorectal cancer. In this study, the authors obtained a superior performance in the classification of the serum applying SVM over PCA.

3. Analysis of Raman data in biomedicine

In order to analyse the advances in data analytics in SERS, it is necessary also to present the developments based on spontaneous Raman scattering (or simply Raman spectroscopy - RS). Thanks to the advancement in computational methods and easier-to-use software for their implementation, ML algorithms have become a common practice in analysing biomedical Raman data.

RS has been applied to analyse a wide variety of biological samples, including tissues, cells, and biofluids [39]. In addition, some groups also managed to bring RS directly in vivo by building Raman endoscopes or devices for the analysis of the skin [40].

Within this multitude of studies, the analysis of Raman spectra has been approached with different degrees of complexity. In some cases, the analysis of a single peak can already provide valuable information. For example, the Raman spectra of bones and breast cancer microcalcification are often based on the study of the width of phosphate peak [41]. The analysis of single peaks is, however, problematic as Raman spectra might be affected by many experimental factors related to the sample, its matrix, and the Raman instrumental configurations and measurement conditions (e.g. laser wavelength, power, irradiation time, confocal setup versus non-confocal, sample stability upon irradiation and with time, etc.). These factors may lead to variations in Raman shifts, peak width, and relative intensities [42].

To deal with the inherent complexity of biological Raman spectra, many groups working on the clinical implementation of RS applied multivariate statistical approaches. Principal Component Analysis - Linear Discriminant Analysis (PCA-LDA) techniques have been vastly exploited in the literature on clinical spectroscopy. For example, the Raman analysis of biofluids followed by a PCA-LDA analysis was proposed for the diagnosis of different kinds of cancer [43–45], neurological disorders [46,47], and metabolic diseases [48]. Similarly, this approach was used, among the other applications, for the classification of cancer cells in acute myeloid leukemia (AML) [49], for the differentiation between cancer cells in prostate cancer (PCa) [50], and the in vivo detection of gastric intestinal metaplasia [51].

Another commonly used multivariate approach, particularly useful when a binary classification is needed, is Partial Least-Squares regression (PLS). Similarly to what is done with PCA, PLS is followed by discriminant analysis and has been applied to classify cytological samples and biofluids of different kinds of cancers [52], and to identify patients with asthma [53].

As mentioned above, more complex ML methods are becoming more popular in the field. Masson et al. used a least absolute shrinkage and selection operator (LASSO) regularization of data to select the Raman features relevant to monitor the biochemical changes occurring *in situ* in the human cervix during spontaneous and induced labor [54]. Harris and co-workers used an RF algorithm in a large prospective study (408 patients) on RS of serum samples for the triage of patients referring to general practitioners for symptoms that might be associated with colorectal cancer [55]. SVM is a particularly popular algorithm used to classify Raman spectra [56], and has proven effective in discriminating Alzheimer's disease patients from a matched cohort of healthy controls [57], and in distinguishing gastric cancer from normal gastric tissues with an impressive reported sensitivity of 88.2% and a specificity of

91.9% [58].

Finally, different neural networks (NNs) have also been applied to the analysis of Raman spectra for the non-invasive screening of patients with type 2 diabetes mellitus [59]; for the identification of the extracellular vesicles released by prostate cancer cells [60]; and for the identification of breast cancer tissues samples [61]. It is also important to notice that ML and NNs are not only useful for spectral classification. NNs also show great promise for the pre-processing and exploration of Raman hyperspectral images, speeding up both the time required for their acquisition [62], and for the qualitative and quantitative analysis of the component present [63].

4. Advances in SERS data analysis

4.1. Tissues and 3D models

Over the past decade, the application of SERS in living biological samples, whether they are *in vitro* or in vivo systems, has advanced considerably. In both cases, for SERS to occur, light of a defined wavelength in resonance with the localized surface plasmons of the nanoparticle (NP) must be able to reach the NPs and the Raman probes or the analytes, and the subsequent inelastic scattering of photons must reach the detector. A significant limitation is the short penetration depth of the excitation laser light, which is shown to be ca. 2−5 mm for NIR light sources. One way to overcome this is via endoscopic methods. In this respect, Raman endoscopic probes have also gained attention to avoid light penetration problems, leading to real-time image-guided resection [64].

The biological environment itself presents many obstructions, both physical and molecular, which diminish both the incident and reflected light, as well as playing an important role in defining the physicochemical properties of the NPs *in situ* [65]. Due to this complicated environment, it is vital to have not only highly stable but also well-characterized NPs, which, in the case of SERS imaging, carry "bright" Raman active molecules [66,67]. The correlation of the obtained SERS signal (spectral intensity) with the number of such SERS tags is a crucial aspect [67].

Regarding the interaction of light with biological samples, there are various methods by which incident and reflected photons can interact with cells or tissues in question (Fig. 1) [68]. Reflectance from the skin surface accounts for a loss of ca. 4 - 7% of the incident light, with the remaining light penetrating the epidermis [69]. From here on, both absorbance and scattering play crucial roles and collectively determine the number of photons that will penetrate further [70]. In addition, changes in the refractive index at cellular and tissue boundaries and between healthy and diseased states can reduce the absolute amount of photons arriving and leaving the sample. On a cellular level, photons interacting with SERS tags in intracellular vesicles will be exposed to multiple changes in refractive index, varying from 1.46 at the plasma membrane to 1.36 in the cytosol [71]. When a laser passes through multi-cellular models and tissues, these changes are significantly augmented. Regardless of the various barriers, light can and does arrive to intracellular SERS substrates, in part thanks to the use of NIR laser irradiation sources located in the first and second Biological window (BW-I, 750–950 nm; BW-II, 1000–1350 nm). The extent of scattering and absorption is minimal in these BWs, and fluorescence interference can be considered negligible when compared to the intensity of SERS signal reaching the detector.

Concerning the 3D cell models or tissues themselves that are currently being explored for SERS measurements, improvements in control over production methods have allowed for reproducible and size-controlled spheroid formation, aiming to mimic simple tumours lacking vasculature. More complex 3D bioprinted models, composed of multiple cell types and with an increased overall size, can also now be routinely produced, and the final combination of such 3D cell models with microfluidics has allowed the "organ-on-a-chip" technology to grow. However, although the model itself may be highly advanced, without equally sophisticated SERS detection systems and data processing techniques, their use in SERS is trivial [72]. Yet, both aspects can lead to undesired cytotoxic effects if working with live cells. A further problem arises in that SERS, being a spectroscopic technique, is data rich, especially when high-resolution images and multiplexing of different molecules are desired. Therefore, methods to improve data acquisition and analysis are emerging to maximize the information

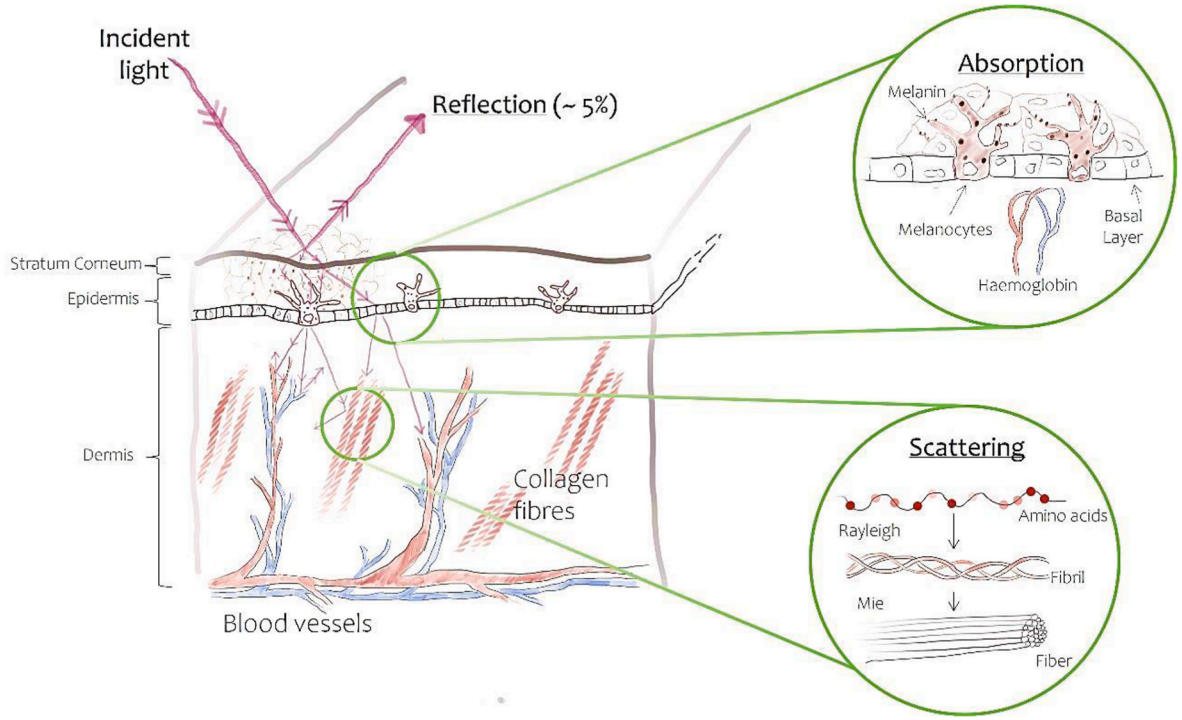


Fig. 1. Optical properties of the skin can reduce incident light intensity.

obtained from such complex images and ease its understanding. For example, ML which allows one to study the interactions of administered NPs in tissues, is gaining much attention as a next-generation data processing technique [73].

To conclude, SERS imaging and biosensing are highly interesting techniques for analysing 3D cell models and tissues in a non-invasive way. Whilst light scattering and absorption due to biological components can significantly reduce the intensity of incident light arriving to the plasmonic substrate, via the design of bright SERS NPs, sufficient data can be obtained to determine the molecular and cellular characteristics of the sample in 4D. Combining chemometrics for SERS data analysis and live-cell imaging makes biomolecule monitoring possible in real-time conditions.

4.2. Cells

In the clinic, the conventional techniques for a general diagnosis of cancer cells are based on the targeting of specific cancer-associated cellular proteins and biomarkers expression. However, for a precise diagnosis, such as differentiating between cancer subtypes and analysis of the cellular environment, more sensitive detection techniques are required. PCA paired with SERS has been implemented in many studies to interpret spectral data variations and classify diverse cell types and even stages of cancer. For example, Kaminska et al. developed a label-free SERS platform to analyse circulating tumour cells (CTC) from blood samples [74]. In their study, 300 SERS spectra from different cell lines were analysed to distinguish between the different cell populations, achieving a sensitivity of 82% [74]. Another example demonstrated the potential of a SERS substrate (fabricated as a Langmuir–Blodgett film) supported by the PCA approach to screen and

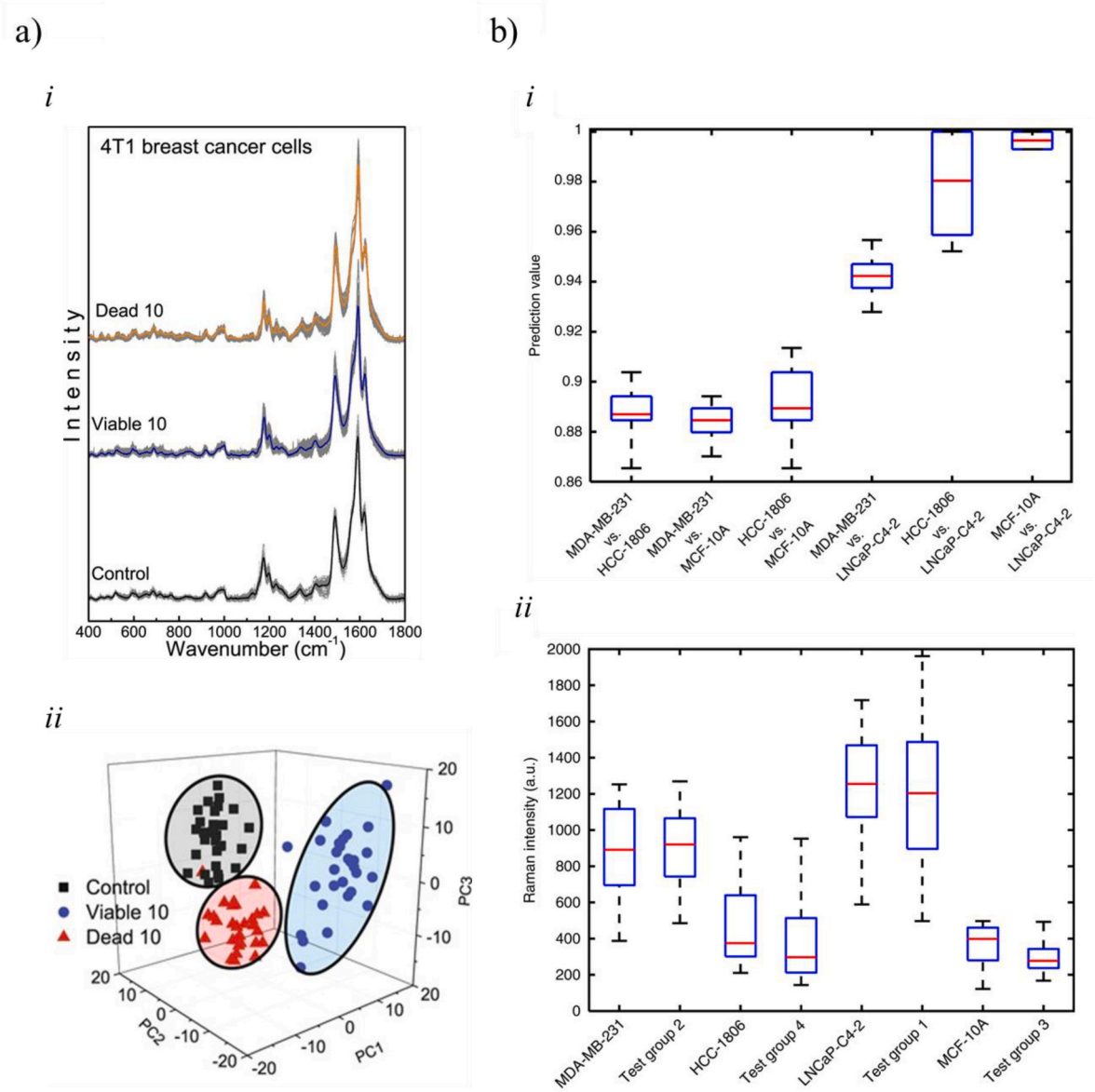


Fig. 2. Representation of different ML methods applied to SERS spectra from cells. a) Analysis of SERS spectra from cells treated with photodynamic therapy (PDT) using PCA: i) SERS spectra of the 4T1 breast cancer cells treated with PDT: dead and viable cells, and also the control group (cells without any treatment); ii) PCA scatter plot of the treated breast cancer cells and the distinction between the different groups: control and treated (viable and dead). Reprinted and adapted from Ref. [76]. Published by Springer Nature Limited. b) The spectra of 4 different cell lines (MDA-MB-231, HCC-1806, LNCaP-C4-2, and MCF-10A) examined by the NGK ML method: i) the predictions for every two sets of cells, and ii) comparison between the original experimental and test groups (represented by TGs: TG1, TG2, TG3, and TG4) Reprinted and adapted from Ref. [77]. Published by Springer Nature Limited.

identify human cancer cell lines from normal cells [75]. Additionally, the analysis of SERS results with PCA has identified and distinguished between non-treated controls, dead cells, and viable cancer cells after photodynamic therapy (Fig. 2a) [76].

Other approaches based on supervised ML have been successfully used to identify different cell types with high accuracy. For example, Ren et al. showed the application of the nonnegative garrote method on the kernel machine (NGK) algorithm to examine the SERS spectral data from four different cell lines (breast and prostate) and successfully identify the different cell types with a 100% match in the blind testing groups (Fig. 2b) [77]. Han and co-workers demonstrated the possibility of performing RF to classify breast cancer and normal single cell automatically [78]. The application of the algorithm resulted in results with a precision of 75.5%, an accuracy of 78.2%, and a recall of 66.7%. Another strategy selected by Hassoun and collaborators demonstrated the identification and classification of cells without any sample preparation by applying PCA and SVM to the collected SERS spectra [79]. Çulha's group has shown that combining PCA-LDA algorithms to analyse SERS spectral dataset can distinguish renal cancer cells from non-cancerous cells [80].

More recently, pairing convolutional neural networks (CNNs) with SERS has shown that SERS can reach a 100% prediction accuracy in distinguishing between cancer cells and normal cells [35]. Other methods, such as the characteristic peak ratio method, the PCA-KNN method, and residual network were also used to classify and identify the SERS spectra from different cell types. For example, Li's group showed the superiority of the deep learning approach of the residual network model over the feature peak ratio method and PCA combined with KNN models [81]. In 2021, it was also demonstrated the possibility of classifying hematological malignancies and discriminating them from two control groups: young–healthy, and aged–sick cohorts [82]. To this end, the spectra of 97 plasma samples from three different cohorts were collected, and the data analysis to distinguish them was performed using LDA and quadratic discriminant analysis (QDA) [82].

4.3. Advances in SERS data analysis applied to nucleic acids

4.3.1. DNA

The interpretation of Raman/SERS spectra from nucleic acids and the possibility of distinguishing the peaks of each analyte are challenging due to the high amount of data and minute differences among spectral features. For example, Guselnikova et al. combined SERS with artificial ANNs to identify and quantify light-induced DNA damage [83]. This combination enabled the discrimination of minimal key spectral features, including DNA damage invisible by standard analytical methods [83]. In addition, Fainman and co-workers have experimentally and numerically demonstrated that PCA and neural network models can detect the composition of a 200-base long single-stranded

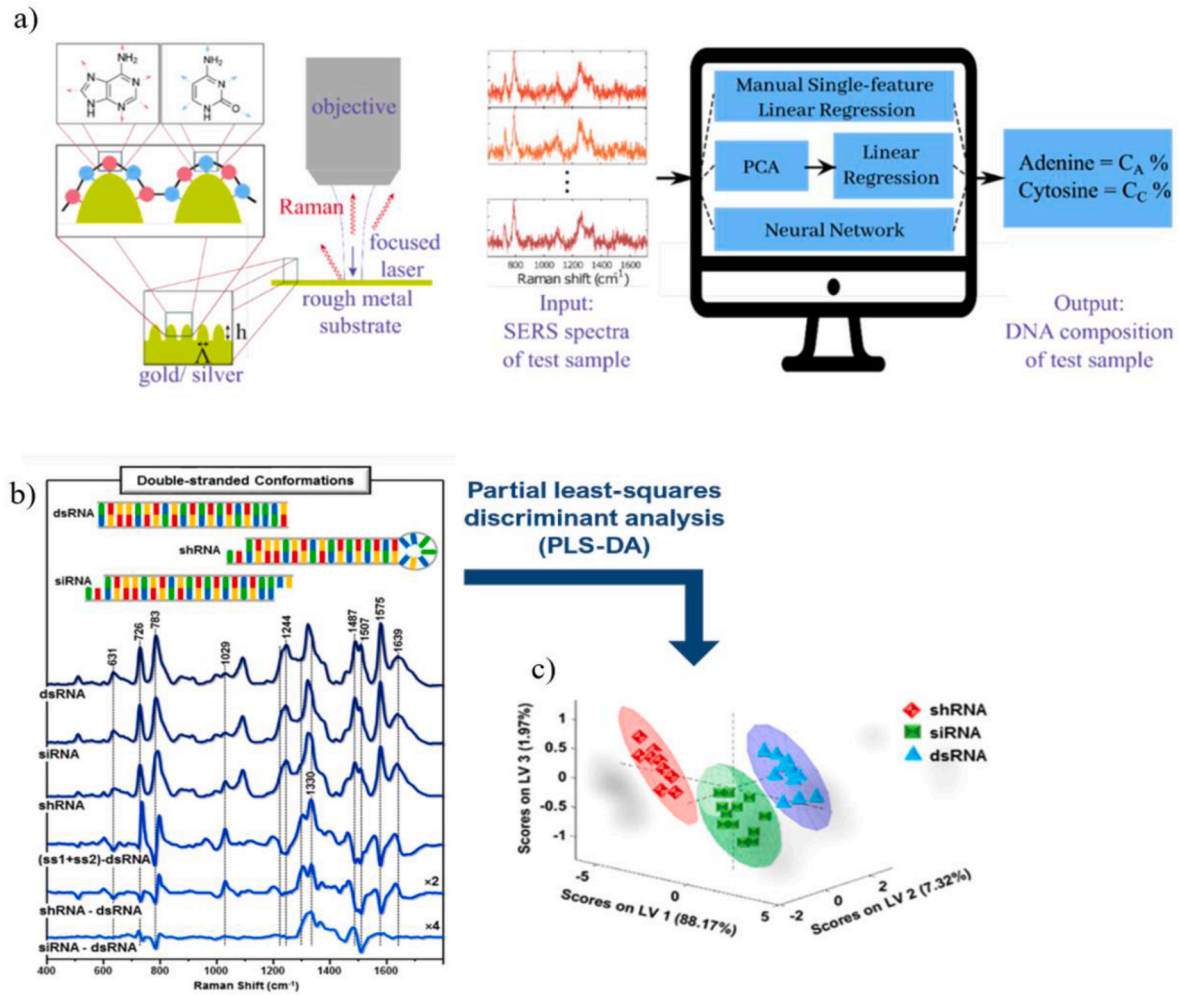


Fig. 3. Examples of AI-enabled SERS platform in nucleic acids measurements. a) treated Raman spectra from the different RNA samples [84]. Reproduced from Ref. [84]. Published by Optica Publishing Group. b) the clear differentiation between the samples using PLS-DA, and c) schematic description of key experimental and data processing components to DNA composition analysis resorting to PCA and linear regression and neural network methods to predict the percentage of adenine and cytosine bases in the ssDNA molecule [85]. Reprinted and adapted with permission from Ref. [85]. Copyrights 2016 American Chemical Society.

DNA (ssDNA) with high sensitivity (Fig. 3a) [84].

Moreover, Shi et al. have developed a label-free SERS biosensing platform to distinguish between DNA targets related to tumour suppressor genes (i.e., p16, p21, and p53 fragments) [86]. Their platform was based on a deep neural network (DNN) algorithm by extracting the main eigenvalue of peak number and height from DNA SERS spectra [86]. This AI-SERS platform was able to distinguish single DNA target and DNA mixtures with an accuracy of 90.28% and 74.83%, respectively [86].

4.3.2. RNA

Data treatment of Raman spectra has always been an important part of the analysis to extract the most reliable, clean, and useful information from measurements. That remains true for RNA analysis. So far, fairly simple analytical methods have been employed for a long time, such as vector normalization employed to validate the reliability of a SERS assay for micro RNA (miRNA) based on the different binding abilities of single-stranded RNA(ssRNA), thiolated ssDNA, and RNA:DNA complex [87]. Another example showed the possibility of detecting and quantifying the presence of the four bases of RNA (polyadenosine, polyuridine, polycytosine, and polyguanosine) in a mixture using the deconvolution of the mixed Raman spectrum into the elemental base pure spectra [88]. Another work also applied an LS regression approach to differentiate between hybridized and non-hybridized miRNA based on the Raman spectra obtained with silver nanorods SERS substrates [89]. This topic is taken even further in work by Morla-Folch et al. where vibrational features of different types of RNA (double-stranded RNA – dsRNA; small interfering RNA – siRNA; and short-hairpin RNA - shRNA) were distinguished by analysing SERS spectra obtained using silver nanoparticles coated with spermine (AgNP@Sp). In this work, the authors used PLS-DA to highlight the tenuous differences between the spectral features of dsRNA, shRNA, and siRNA (Fig. 3b and c) [85].

For clinical samples, Chen et al. have employed a combination of PCA-LDA to differentiate between gastric cancer patients and healthy controls by analyzing the SERS spectra of their RNA in serum samples [90]. Nasir et al. and Rafiq et al. [91,92] also showed the potential to use Ag colloidal solutions as SERS substrates to identify vibrational spectral features from RNA collected from healthy controls and Hepatitis C patients. Remarkably, the prediction of low, medium, and high viral loads was also studied by employing a partial least squares regression analysis (PLSR), with 99% accuracy and reliability (R^2) of 0.98 [91].

4.4. Metabolites and proteins

Metabolites are biomolecules involved in the complex machinery of metabolic processes of cells. They play crucial roles in life-sustaining chemical reactions, cellular physiology maintenance, or mediating cellular communications, such as activating cell receptors or feeding energy in productive cycles. The extended activity of metabolites is mainly mediated by proteins [93]. Monitoring of metabolomic and proteomic profiles allows the discrimination between normal and pathological pathways, which can help in diagnosis and prognosis [94]. SERS has shown to be a powerful tool for the label-free detection of these biomolecules [95].

ML methods, such as CNN and SVM, have been employed to categorize abnormal metabolism in different clinical samples such as urine or blood. For example, Chen et al. used CNN algorithms to classify patients with prostate cancer (PCa) and Bening prostatic hyperplasia (BPH) through Raman spectra of urine samples [96]. After training the CNN model, 75% of accuracy, 77.3% of sensitivity, and 72.4% of specificity were achieved for PCa diagnosis [96]. Alternatively, Zhou et al. employed SVM algorithms to classify PCa, BPH, and control subjects [97]. In this case, prostate specific antigen (PSA), prostate specific membrane antigen (PSMA), and human kallikrein 2 (hK2) were detected in serum samples. SVM method exhibited a clear classification of PCa, BPH, and control patients, with accuracy for PCa around 70%. Another example showed the use of PCA-LDA for the diagnosis of breast cancer, where urine samples from 53 patients and 22 control were analysed [98]. This study showed that PCA-LDA analysis can extract distinguishable spectral characteristics between breast cancer patients and healthy controls, with a precision of 88% based on the analysis of the most frequent metabolites (uric acid, xanthine, and hypoxanthine). Similarly, PCA-LDA analysis was employed in breast cancer diagnosis with serum samples of 11 patients and 12 healthy controls [45]. Most common metabolites detected correspond to polysaccharides, phospholipids and proteins in the later study.

Moreover, PCA-LDA, SVM, and PCA-QDA analyses have been applied to classify metabolites from gastrointestinal cancer patients. Hong et al. showed that the most common metabolites found in the serum of colorectal cancer patients were phenylalanine, tyrosine, hypoxanthine, thymine, and lipids [38]. The data was analysed and compared by PCA and SVM, and the later showed superior performance in the classification of the colorectal cancer group, with accuracy and sensitivity rates of over 80%. Another gastrointestinal tumour classification was performed with serum samples with an accuracy of 76.92% [99]. SERS spectra revealed the presence of carotenoids and purine metabolites. In another study, Lussier et al. showed cellular secretion and consumption of metabolites by combining SERS-optophysiology with ANN to monitor up to eight metabolites [100]. An interesting study was able to optimize conditions for investigating cervical cancer from serum by testing different lasers and dilutions of serum [92]. Li et al. showed a unified, systematic, and standard method to distinguish healthy from cancer patients by PCA [101]. A schematic representation of this method and results are presented in Fig. 4a, showing SERS spectra from normal and cervical cancer patients [101]. Another work demonstrates the ability to differentiate different sub-types of AML by blood plasma proteins and metabolites with Ag nanoparticles and PCA, as shown in Fig. 4b [102]. Wu et al. demonstrated the effect of radiotherapy on inducing SERS spectral changes in plasma samples from nasopharyngeal cancer (NPC) patients before and after treatment and compared the spectra to healthy volunteers [103]. A significant difference was observed between SERS spectra before and after radiotherapy. The changes in these SERS spectra are depicted in Fig. 4c, as well as the corresponding computer tomography images [94]. This study suggested the possibility of plasma SERS technology for predicting the effect of radiotherapy in NPC [103].

Moreover, other works have also developed diagnostic methods for diseases other than cancer. Stefancu et al. detected purine metabolites in saliva and serum samples to diagnose an autoimmune disease, Sjögren's syndrome [104]. The authors performed PCA-LDA analysis to classify the two groups of patients ($n = 29$) and controls ($n = 21$) based on SERS spectra of saliva and serum. An accuracy of 94% was obtained in the case of saliva and 98% for serum. Some examples from the above-mentioned sections are summarized in Table 1, showing a wide diversity in the SERS substrates, Raman systems, data analytics models, and applications. Table 1 reflects how complex it is to compare Raman/SERS biosensors from different labs with many variations. It also shows the need/potential for combining artificial intelligence in Raman/SERS data analysis for translational medicine applications.

5. Advances in SERS data analysis using deep learning

In the past few months, a novel platform, including a Raman spectrometer coupled with a neural network, was able to classify SERS spectra from a blood test of patients with hepatitis B virus, M5 leukemia, and breast cancer with an impressive accuracy of 98.6% [105]. For a similar purpose, a different group created a one-dimensional CNN for the identification of SERS from four cancer cell types (A549, HeLa, HepG2, and SW620) and normal cells (WRL68), reaching a mean accuracy of 97.56% [106]. Another project consisted of building a SERS-based sensor capable of detecting biomolecular information of liver diseases by direct analysis of SERS in serum. By integrating this sensor with a CNN classifier, the system performed direct serological

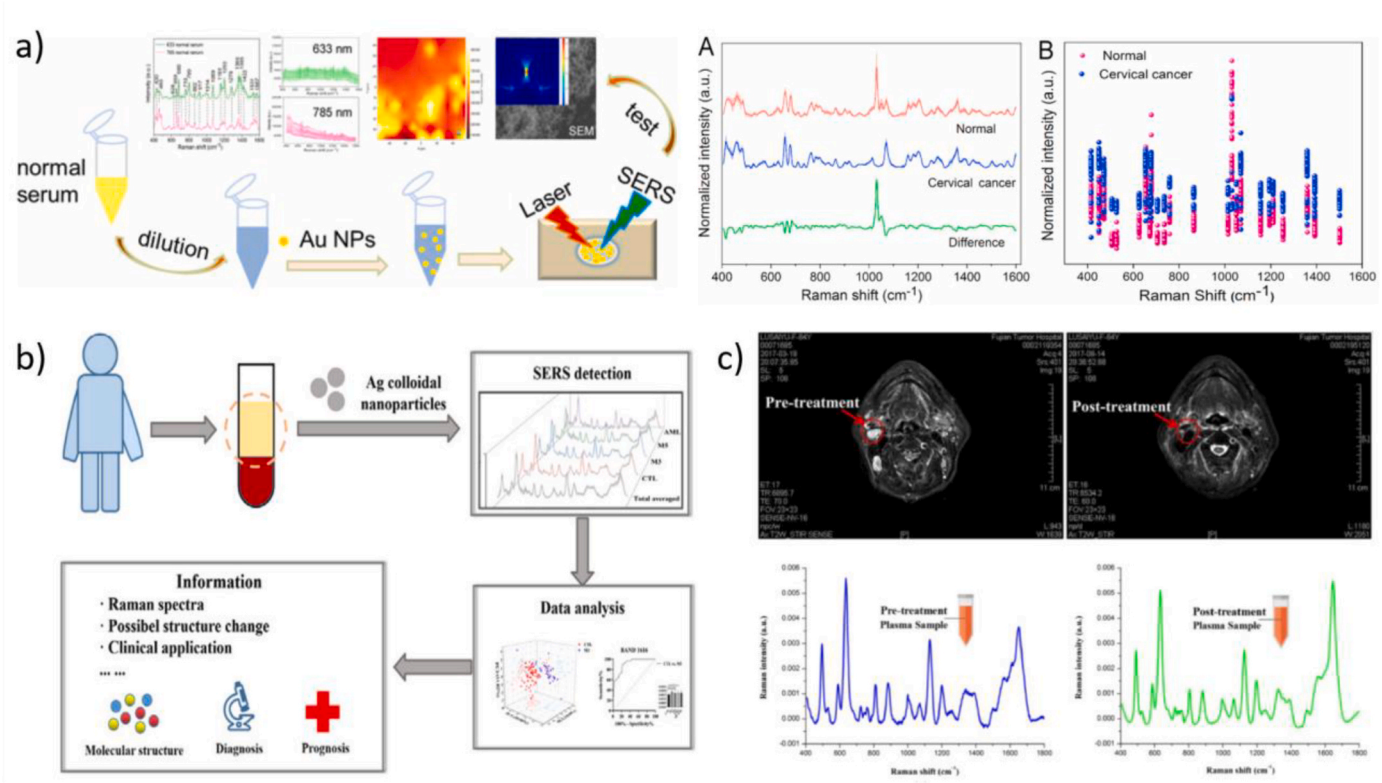


Fig. 4. Examples of SERS application in the detection of metabolites and proteins. a) Method to distinguish cervical cancer patients from healthy samples using blood serum, different lasers and PCA. A: the differences in serum SERS spectra between healthy volunteers and patients with cervical cancer. B: the scatter plot of different bands of SERS spectra between healthy volunteers and patients with cervical cancer [101]. Reprinted and adapted with permission from Ref. [101]. Copyrights 2021 Elsevier. b) Flow chart of detection methods for AML from blood plasma [102]. Reproduced with permission from Ref. [102]. Copyrights 2022 Elsevier. c) Computer tomography images and SERS spectrum of plasma of pre- and post-radiotherapy treatment [103]. Reproduced from Ref. [103]. Published by Optica Publishing Group.

detection of liver cancer within 1 min, achieving 97.78% predictive accuracy [107]. A different study showed how SERS processing through deep learning techniques was able to delineate the pH map of the tumour resection bed in animals. This method can substantially improve margin-guided surgery compared to the standard approach used in clinical practice, and it can potentially be applied to solid tumours [108]. In addition, a study has shown that it is possible to create screening systems without pre-processing of SERS spectra (denoising and baseline establishment), an operation that until now required human intervention. The proposed method, based on a Residual Neural Network (ResNet) combined with a data augmentation technique, has proven to be a robust system for the analysis of different grades of bladder cancer tissues and extracellular vesicles (EVs) of various strains of *E. coli*, especially when dealing with different Raman spectrometers [109].

Deep learning tools combined with Raman and SERS have also demonstrated to create new detection methods for SARS-CoV-2 with an accuracy of 87.7% [110], or to robustly discriminate the signal from patients with current COVID-19 infection, healthy subjects, and subjects with past infection (accuracy range of 89–92%) just using a saliva sample [111].

Antimicrobial resistance remains a threat and one of the biggest challenges for science today. The development of technology in this field is essential for diagnosing infectious diseases and analysing food safety. A few recent examples of how neural networks have contributed to this field over the past few months include a deep neural network capable of discriminating the antibiotic-resistant bacteria *methicillin-resistant Staphylococcus aureus* (MRSA) and *methicillin-sensitive S. aureus* (MSSA) using SERS with accuracies of 97.66% and 100% [112,113]. Another example is a deep learning model called dual-branch wide-kernel network (DualWKNet), capable of classifying the signals of two common bacteria (*E. coli*, *S. epidermidis*) and their resident media without any separation procedure, achieving up to 98% accuracy [114].

Other types of deep learning architectures have been applied to achieve a variety of objectives. For instance, a short-term memory neural network was implemented for the identification of pathogens [115] and pesticides [116].

6. Conclusions

Beyond improving on the robustness, efficiency, and reliability of the of technique/substrate, that is well underway, the main challenges to overcome in the field are: 1) speed of sample preparation and analysis to ensure faster sample-to-result; 2) scalability and commercial availability of enhancers production, yielding homogeneous and reproducible results; 3) availability of compact and simple-to-use commercial products compatible with clinical laboratories and established analytical protocols; 4) multiplexing expansion from current proof-of-concept studies, to enhanced analytical multidimensional chemometric tools; 5) reliability of results regardless of sample-to-sample variations (i.e. time since last meal) and patient-to-patient differences (gender, age, comorbidities, etc) [117]. Once the above-mentioned challenges are overcome, SERS could really face certification and be translated into the clinical setting enabling high-throughput, standardized and automated analysis of patient samples, yielding multidimensional results. The promising studies highlighted in this manuscript reveal the potential of the combination of SERS and advanced analytical techniques for creating new molecular diagnostic tools that are more accurate, simpler to use, require fewer resources, such as time and money, and are also substantially less invasive for patients. Biophysical and ML techniques

Table 1
Summary of the experimental parameters of the examples of some works merging SERS and advanced data analytics models to provide dimensionless opportunities in clinical research.

Biological entity	Sample/s and/or analytes	SERS substrate	Raman spectrometer and its wavelength	Data analytics algorithm/s	Key findings/applications	Ref
Cells	Prostate cancer (PC3), cervical carcinoma (HeLa), and leucocytes	Membrane-based SERS platforms with appropriate pore sizes and with a Ag:Au alloy layer of 40 nm deposited on the polymer	Renishaw inVia Raman system from Wotton-under-Edge Company (Laser:785 nm)	PCA	Direct detection and identification of different cancer cells and distinguishing them from normal cells.	[74]
Cells	Human lung fibroblast cells (WI-38), human breast cancer cells (MDA-MB-468) and Human prostate cancer cell lines (PC-3)	AuNCs with ~50 nm entrapped in a bilayer Langmuir–Blodgett film of stearic acid	LabRAM HR (Horiba) and using DPSS lasers with λ = 632.8 and 785 nm	PCA	Detection and differentiation of normal and cancerous cell based on their enhanced vibrational signatures of cancer cells	[75]
Cells	Mouse mammary glands cancer cells (4T1 cell line)	Electrodeposited films of silver	Triple spectrometer Jobin Yvon model T64000 (488 nm Argon laser line)	PCA	Detect differences between control without any treatment, live, and dead cells after therapy	[76]
Cells	Breast cancer cell lines (MDA-MB-231 and HCC-1806), non-malignant breast epithelial cells (MCF-10A), and prostate cancer cell line (LNCaP-C4-2)	Multilayered metal-insulator-metal nanolaminated SERS substrate integrated multiwell cell culture assay for label-free Raman profiling	Laser: 785 nm	NGK	Analysis of small amounts of samples and identification of specific cancer subtypes with high accuracy and reproducibility	[77]
Cells	15 fresh specimens of breast cancer tissue and normal breast tissue (from the same patient)	AgNPs	LabRAM Aramis from Horiba (Laser: 633 nm)	RF	Distinguish the SERS spectral characteristics of primary breast cancer cells from primary normal epithelial cells and improve the diagnostic technology for breast cancers	[78]
Cells	Cell lysates from two liver cancer cell lines (HepG2 and SK-Hep1) and breast cancer cell line (MCF-7) and human pancreatic ductal adenocarcinoma cell line (Capan-1)	AgNPs	Commercial Raman microscopy system (Holoprobe, Kaiser Optical system) (Laser: 785 nm)	PCA followed by SVM	Identification of cancer cells from normal ones	[79]
Cells	Human kidney adenocarcinoma (ACHN), human kidney carcinoma (A-498) and non-cancerous human kidney embryonic cells (HEK 293)	AuNPs (~50 nm)	Renishaw InVia Reflex Raman microscopy system (Laser: 830 nm diode laser)	PCA and LDA	Determination of nanotoxicity of living cells and classification of healthy and cancer cells	[80]
Cells	Cervical cancer cells (Hela), laryngeal cancer cells (Hep2), liver cancer cells (HepG2), ovarian cancer cells (HO-8910), liver cancer cells (SMMC7721), breast cancer cells (MCF-7), lung cancer cells (H446), human neuroblastoma cells (SH-SY5Y), clinical tissue samples, and healthy volunteer blood samples	Silver film substrate	Confocal Raman microscopy (Renishaw) (Laser: 785 nm)	PCA-KNN method, and residual network	Identification of tumor cells from other blood cells	[81]
Cells	Plasma samples from patients diagnosed with multiple myeloma, lymphoma or cardiovascular disease	AuNPs	Renishaw inVia Raman Spectrometer (helium-neon laser beam with 514 excitation wavelength)	PCA, LDA and QDA	Development of label-free SERS platform to monitor the status/disease progression of patients with hematological malignancies using two reference datasets	[82]
DNA	DNA molecules	Gold grating surface	Portable ProRaman-L spectrometer (785 nm)	ANN	Identification and quantification of light-induced DNA damage	[83]
DNA	200-base length single-stranded DNA	Gold or silver nanorod substrates	InVia Raman spectrometer from Renishaw (785 nm)	PCA, linear regression, and neural network	Prediction of the percentage of adenine and cytosine bases in the ssDNA molecule	[84]
DNA	Tumour suppressor genes (i.e., p16, p21, and p53 fragments)	Silver nanoparticles decorated silicon wafer	Raman microscope HR800 from Horiba Jobin Yvon (633 nm)	DNN	Distinguishing single DNA targets and DNA mixtures employing the developed AI-based sensing method	[86]
RNA	ssRNA, thiolated ssDNA, and RNA:DNA duplex	Silver nanorod array substrate	Confocal Raman microscope system from Bruker Senterra (785 nm)	Vector normalization spectra treatment	Different propensity to adsorption on the substrate leads to reliable RNA SERS based assay	[87]
RNA	Polyadenosine, polyuridine, polycytosine, and polyguanosine	Creighton silver colloid	Confocal Raman microspectrometer from	Spectra Deconvolution treatment	The deconvolution process allows estimation of the	[88]

(*continued on next page*)

Table 1 (*continued*)

Biological entity	Sample/s and/or analytes	SERS substrate	Raman spectrometer and its wavelength	Data analytics algorithm/s	Key findings/applications	Ref
			LabRam, Jobin Ivon (514 nm)		percentage of each base component of RNA strands	
RNA	Adenine, cytosine, guanine, uracil, and thymine	Silver nanorod substrate	Enwave system model HCT-10HT (785 nm) and Bruker Senterra system (532 nm)	Least-Squares (LS) Method	Quantify compositional ratios of the nucleotide components allowing for the differentiation between hybridized and non-hybridized sequences	[89]
RNA	microRNA, fully complementary duplex-RNA, small interfering-RNA, and short hairpin- RNA	Silver spheroidal nanoparticles coated with Spermine colloidal solution	In Via Reflex confocal microscope from Renishaw (532 nm)	PLS-DA	Differentiation amongst RNA with highly specific structural information and characterized by subtle chemical differences	[85]
RNA	RNA from serum samples collected from gastric cancer patients and healthy volunteers	Silver colloid solutions	In Via micro-Raman spectroscopy system from Renishaw (785 nm)	PCA-LDA	Differentiation between gastric cancer patients from healthy individuals	[90]
RNA	Viral RNA of Hepatitis C in HCV-infected patients with different viral loads	Silver nanoparticles solution	Raman microspectrometer ATR8300BS from Optosky, China (785 nm)	PCA and PLSR	Prediction of the viral loads of HCV of unknown samples	[91, 92]
Metabolites, proteins, DNA	Lipids, amino acids (cysteine, tyrosine, histidine or tryptophan) and nucleic acids (extracellular media from normal and cancer cells). Primary cell lines (dermal fibroblasts from healthy skin) and primary cancer cell lines (fibroblasts from skin metastasis of malignant melanoma)	Gold multibranched nanoparticles	ProRaman-L spectrometer (785 nm)	CNN	Development of an advanced approach capable of detecting compositional changes derived from the metabolic activity of the culture medium of normal and tumor cells and preparation of a reliable database.	[35]
Metabolites and Proteins	Nucleic acids, lipids and amino acids (urine samples from PCa and BPH patients)	Silver sol (obtained from a solution of $AgNO_3$)	Horiba HR Evolution LabRAM laser confocal micro-Raman spectrometer (785 nm)	CNN	Application of artificial intelligence in research based on clinical medicine.	[96]
Metabolites and Proteins	PSA, PSMA and hK2 (serum of patients with PCa, BPH and healthy volunteers).	Silver colloid Nanoparticle (50 ± 5 nm)	BWS415, B&W Tek Inc (785 nm)	SVM	To improve cancer diagnosis accuracy with potential applications in testing.	[97]
Metabolites and Proteins	Uric acid, xanthine, and hypoxanthine (urine of patients and controls)	Silver Nanoparticles	InVia Raman Spectrometer Renishaw. (532 nm)	PCA-LDA	Application in breast cancer screening.	[98]
Metabolites and Proteins	Polysaccharides, phospholipids and amino acids (serum of patients and healthy volunteers)	Raman analysis (no nanoparticles)	Raman system, Reni-Shaw 1000B. (830 nm)	PCA-LDA	Application in screening for breast cancer using Raman Spectroscopy of serum samples.	[45]
Metabolites and Proteins	Phenylalanine, tyrosine, hypoxanthine, thymine and lipids (serum of patients and healthy volunteers)	Gold Nanoparticles (70-75 nm)	Horiba spectrometer, iHR 550. (785 nm)	PCA-LDA and SVM	Application of SERS combined by blood analysis for diagnosis.	[38]
Metabolites and Proteins	Carotenoids and purine metabolites (serum of patients and healthy volunteers)	Silver Nanoparticles	Raman spectroscope (iRaman 532 nm, B&W Tek, Lübeck, Germany).	PCA and QDA	Diagnosis of oral cancer.	[99]
Metabolites and Proteins	Urea, adenosine diphosphate, adenosine triphosphate, pyruvate, glutamine, lactate, glucose and CO_2 (extracellular media from HeLa, fibroblasts, and HUVEC cells).	Gold Nanoraspberries	Alpha300 Raman microscope from WITec (Ulm), 632.8 nm.	ANN	Adenosine triphosphate surveillance on chronic inflammatory diseases, cancer, and other cellular functions.	[100]
Metabolites and Proteins	Nucleic acids, lipids, purine metabolites and amino acids (serum of patients and healthy volunteers)	Gold Nanoparticles	HORIBA Raman microscope (LabRAM HR Evolution, HORIBA Scientific). 532, 633, and 785	PCA	Potential advantages in disease screening and diagnosis.	[101]
Metabolites and proteins	Glucose, organic acids, lipids, nucleic acids, proteins, etc. (plasma from patients and control donors)	Silver nanoparticles	Micro-Raman system (Renishaw Corporation, Gloucestershire, London, England). 785 nm.	PCA	To detect tumour biomarkers by SERS in acute myeloid leukemia	[102]
Metabolites and proteins	Amino acids, tryptophan, nucleic acids, L-arginine, hypoxanthine, ascorbic acid and cholesterol (plasma from patients and controls)	Silver nanoparticles (35 nm)	Renishaw UK (785 nm)	PCA-LDA	Cancer diagnosis in non-invasive body fluid analysis	[103]
Metabolites and Proteins	Uric acid, xanthine and Hypoxanthine (saliva and serum of patients and controls).	Silver nanoparticles	InVia Raman Spectrometer (Renishaw). 532 nm	PCA-LDA	To acquire a novel point-of-care for the diagnosis of Sjögren's syndrome.	[104]

can complement one another for complex data analysis. Indeed, future solutions may require a mix of the two techniques, where biophysical models are used in preprocessing and accounting for instrumental and physical mechanisms, while ML is used for characterizing/discriminating between groups. With the development of AI, ML analytics becomes a powerful tool for making spectra-based diagnoses. Indeed, ML approaches can capture small changes between spectra and make complex decision boundaries and, as such, have demonstrated to make a significant impact on spectroscopic analysis and diagnosis. However, ML approaches require large data sets for proper training and to reduce the risk of overfitting. The choice of appropriate data analysis is a significant challenge to apply SERS to real biomedical applications, where large clinical data sets can be difficult to obtain. Nonetheless, without a doubt, the application of advanced data analysis techniques will bring SERS closer to the clinic.

Declaration of competing interest

All authors declare no conflict of interest.

Data availability

No data was used for the research described in the article.

Acknowledgements

S. A.-C., L.D., A. T., A. M., A. C., M. S. acknowledge the support from Fundação para a Ciência e Tecnologia e Programa Operacional Competitividade e Internacionalização for the Impact- L project (contract number 030782); S. A.-C., L.D., M. A., M. R., M. C., M. H., J. K., J. T. acknowledge the POFC – COMPETE; FCT for the support through the project SENTINEL (contract number NORTE-01-0247-FEDER-045914); A. T. acknowledges the FCT studentship SFRH/BD/148091/2019; The work of S. Q. is supported by European union Next Generation EU under grant RD 2289/2021; J. K. acknowledges support from the National Institutes of Health (T32EB007507). C. M., F. P., F. C., D., M., A.R.-P., A. C., L. D., S. A.-C. acknowledge financial support from the European Innovation Council (Horizon 2020 Project: 965018—BIOCELLPHE); S. A.-C. acknowledges the financial support of the 3DSecret project, funded by the EU under the program HORIZON-EIC-2022-PATHFINDEROPEN-01-01 (ga 101099066) and by the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (ga 10063360); L. D and S. A.-C. acknowledge funding by the Health From Portugal project (C630926586-00465198), supported by Component C5 – Capitalisation and Business Innovation, under the Portuguese Resilience and Recovery Plan, through the NextGenerationEU Fund; A.R.-P. was partially funded by project PID2019-106960 GB-I00 (AEI/FEDER, Spain) and by InGEMICS-CM, FSE/FEDER, Comunidad de Madrid Project B2017/BMD-3691. D.J.d.A. acknowledges financial support from the MCIN/AEI/ 10.13039/501100011033 through grant PID2019-108854RA-I00.

Abbreviations

Ag	Silver
AgNP@Sp	Silver Nanoparticles coated with spermine
AgNWs	Silver Nanowires
AI	Artificial Intelligence
AML	Acute Myeloid Leukemia
ANN	Artificial Neural Networks
AuNCs	Gold Nanocolloids
AuNPs	Gold nanoparticles
BPH	Bening Prostatic Hyperplasia
BW	Biological Window
CNNs	Convolutional Neural Networks
COVID-19	Coronavirus Disease 2019
CTC	Circulating Tumor Cell
CVD	Cardiovascular Disease
DRS	Diffuse Reflectance Spectroscopy
DNA	Deoxyribonucleic Acid
dsRNA	Double-Stranded RNA
DualWKNet	Dual-branch Wide-Kernel Network
EVs	Extracellular Vesicles
GBC	Gradient Boosted Classifier
hk2	Human Kallikrein 2
KNN	k-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
miRNA	Micro RNA
ML	Machine Learning
MRSA	Methicillin-Resistant Staphylococcus Aureus
MSSA	Methicillin-Sensitive Staphylococcus Aureus
NGK	Nonnegative Garrote method on the Kernel machine
NNs	Neural Networks
NP	Nanoparticle
NPC	Nasopharyngeal Cancer
NIR	Near Infrared
PCA	Principal Component Analysis
PCa	Prostate Cancer
PCA-KNN	Principal Component Analysis-k-Nearest Neighbor
PCA-LDA	Principal Component Analysis- Linear Discriminant Analysis
PDT	Photodynamic Therapy
PLS	Partial Least Squares
PLS-DA	Partial Least Squares- Discriminant Analysis
PLSR	Partial Least Squares Regression
PSA	Prostate Specific Antigen
PSMA	Prostate-Specific Membrane Antigen
QDA	Quadratic Discriminant Analysis
RF	Random Forest
RNA	Ribonucleic Acid
RS	Raman Scattering
SERS	Surface-enhanced Raman Scattering
shRNA	Short-Hairpin RNA
siRNA	Small Interfering RNA
ssDNA	Single-Stranded DNA
ssRNA	Single-Stranded RNA
SVM	Support Vector Machine

References

[1] S. Abalde-Cela, L. Wu, A. Teixeira, K. Oliveira, L. Diéguez, Multiplexing liquid biopsy with surface-enhanced Raman scattering spectroscopy, Adv. Opt. Mater. 9 (2021) 2001171, https://doi.org/10.1002/adom.202001171.

[2] S. Seifert, Application of random forest based approaches to surface-enhanced Raman scattering data, Sci. Rep. 10 (2020) 5436, https://doi.org/10.1038/s41598-020-62338-8.

[3] H. Li, Y. Cao, F. Lu, Differentiation of different antifungals with various mechanisms using dynamic surface-enhanced Raman spectroscopy combined with machine learning, J. Innov. Opt. Health Sci. 14 (2021) 2141002, https://doi.org/10.1142/S1793545821410029.

[4] W. Hu, S. Ye, Y. Zhang, T. Li, G. Zhang, Y. Luo, S. Mukamel, J. Jiang, Machine learning protocol for surface-enhanced Raman spectroscopy, J. Phys. Chem. Lett. 10 (2019) 6026–6031, https://doi.org/10.1021/acs.jpclett.9b02517.

[5] L.F. Tadesse, F. Safir, C.-S. Ho, X. Hasbach, B. (Pierre) Khuri-Yakub, S.S. Jeffrey, A.A.E. Saleh, J. Dionne, Toward rapid infectious disease diagnosis with advances in surface-enhanced Raman spectroscopy, J. Chem. Phys. 152 (2020) 240902, https://doi.org/10.1063/1.5142767.

[6] N.H. Othman, K.Y. Lee, A.R.M. Radzol, W. Mansor, U.R.M. Rashid, Detection of NS1 from SERS spectra using K-NN integrated with PCA, in: 2016 IEEE EMBS Conf. Biomed. Eng. Sci., 2016, pp. 91–95, https://doi.org/10.1109/IECBES.2016.7843421.

[7] B. Deng, X. Luo, M. Zhang, L. Ye, Y. Chen, Quantitative detection of acyclovir by surface enhanced Raman spectroscopy using a portable Raman spectrometer coupled with multivariate data analysis, Colloids Surf. B Biointerfaces 173 (2019) 286–294. https://doi.org/10.1016/j.colsurfb.2018.09.058.

[8] A.A. Moawad, A. Silge, T. Bocklitz, K. Fischer, P. Rösch, U. Roesler, M. C. Elschner, J. Popp, H. Neubauer, A machine learning-based Raman

spectroscopic assay for the identification of burkholderia mallei and related species, Molecules 24 (2019), https://doi.org/10.3390/molecules24244516.
[9] S. Managò, C. Valente, P. Mirabelli, D. Circolo, F. Basile, D. Corda, A.C. De Luca, A reliable Raman-spectroscopy-based approach for diagnosis, classification and follow-up of B-cell acute lymphoblastic leukemia, Sci. Rep. 6 (2016) 24821, https://doi.org/10.1038/srep24821.
[10] H.J. Koster, T. Rojalin, A. Powell, D. Pham, R.R. Mizenko, A.C. Birkeland, R. P. Carney, Surface enhanced Raman scattering of extracellular vesicles for cancer diagnostics despite isolation dependent lipoprotein contamination, Nanoscale 13 (2021) 14760–14776, https://doi.org/10.1039/D1NR03334D.
[11] S.X. Leong, Y.X. Leong, E.X. Tan, H.Y.F. Sim, C.S.L. Koh, Y.H. Lee, C. Chong, L. S. Ng, J.R.T. Chen, D.W.C. Pang, L.B.T. Nguyen, S.K. Boong, X. Han, Y.-C. Kao, Y. H. Chua, G.C. Phan-Quang, I.Y. Phang, H.K. Lee, M.Y. Abdad, N.S. Tan, X.Y. Ling, Noninvasive and point-of-care surface-enhanced Raman scattering (SERS)-Based breathalyzer for mass screening of coronavirus disease 2019 (COVID-19) under 5 min, ACS Nano 16 (2022) 2629–2639, https://doi.org/10.1021/acsnano.1c09371.
[12] J.H. Granger, N.E. Schlotter, A.C. Crawford, M.D. Porter, Prospects for point-of-care pathogen diagnostics using surface-enhanced Raman scattering (SERS), Chem. Soc. Rev. 45 (2016) 3865–3882, https://doi.org/10.1039/C5CS00828J.
[13] J.-F. Masson, The need for benchmarking surface-enhanced Raman scattering (SERS) sensors, ACS Sens. 6 (2021) 3822–3823, https://doi.org/10.1021/acssensors.1c02275.
[14] W.J. Thrift, C.Q. Nguyen, M. Darvishzadeh-Varcheie, S. Zare, N. Sharac, R. N. Sanderson, T.J. Dupper, A.I. Hochbaum, F. Capolino, M.J. Abdolhosseini Qomi, R. Ragan, Driving chemical reactions in plasmonic nanogaps with electrohydrodynamic Flow, ACS Nano 11 (2017) 11317–11329, https://doi.org/10.1021/acsnano.7b05815.
[15] J.-M. Kim, C. Lee, Y. Lee, J. Lee, S.-J. Park, S. Park, J.-M. Nam, Synthesis, assembly, optical properties, and sensing applications of plasmonic gap nanostructures, Adv. Mater. 33 (2021) 2006966, https://doi.org/10.1002/adma.202006966.
[16] M.J. Natan, Concluding remarks surface enhanced Raman scattering, Faraday Discuss 132 (2006) 321–328, https://doi.org/10.1039/B601494C.
[17] Y. Ying, Z. Tang, Y. Liu, Material design, development, and trend for surface-enhanced Raman scattering substrates, Nanoscale 15 (2023) 10860–10881, https://doi.org/10.1039/D3NR01456H.
[18] R. Pilot, C. Durante, L. Orian, M. Bhamidipati, L. Fabris, A Review on Surface-Enhanced Raman Scattering, 2019.
[19] A.I. Pérez-Jiménez, D. Lyu, Z. Lu, G. Liu, B. Ren, Surface-enhanced Raman spectroscopy: benefits, trade-offs and future developments, Chem. Sci. 11 (2020) 4563–4577, https://doi.org/10.1039/D0SC00809E.
[20] L.B. Laurentius, A.C. Crawford, T.S. Mulvihill, J.H. Granger, R. Robinson, J. S. Spencer, D. Chatterjee, K.E. Hanson, M.D. Porter, Importance of specimen pretreatment for the low-level detection of mycobacterial lipoarabinomannan in human serum, Analyst 142 (2017) 177–185, https://doi.org/10.1039/C6AN02109C.
[21] K. Kneipp, H. Kneipp, I. Itzkan, R.R. Dasari, M.S. Feld, Ultrasensitive chemical analysis by Raman spectroscopy, Chem. Rev. (1999), https://doi.org/10.1021/cr980133r.
[22] B.C. Wilson, S.L. Jacques, Optical reflectance and transmittance of tissues: principles and applications, IEEE J. Quant. Electron. (1990), https://doi.org/10.1109/3.64355.
[23] V. Backman, R. Gurjar, K. Badizadegan, I. Itzkan, R.R. Dasari, L.T. Perelman, M. S. Feld, Polarized light scattering spectroscopy for quantitative measurement of epithelial cellular structures in situ, IEEE J. Sel. Top. Quant. Electron. (1999), https://doi.org/10.1109/2944.796325.
[24] G.A. Wagnières, W.M. Star, B.C. Wilson, In vivo fluorescence spectroscopy and imaging for oncological applications, Photochem. Photobiol. (1998), https://doi.org/10.1111/j.1751-1097.1998.tb02521.x.
[25] M.H. Nguyen, Y. Zhang, F. Wang, J. De La Garza Evia Linan, M.K. Markey, J. W. Tunnell, Machine learning to extract physiological parameters from multispectral diffuse reflectance spectroscopy, J. Biomed. Opt. (2021), https://doi.org/10.1117/1.jbo.26.5.052912.
[26] Y. Zhang, X. Mi, X. Tan, R. Xiang, Recent progress on liquid biopsy analysis using surface-enhanced Raman spectroscopy, Theranostics 9 (2019) 491–525, https://doi.org/10.7150/thno.29875.
[27] S. Wang, J. Zhao, H. Lui, Q. He, J. Bai, H. Zeng, Monte Carlo simulation of in vivo Raman spectral measurements of human skin with a multi-layered tissue optical model, J. Biophot. (2014), https://doi.org/10.1002/jbio.201300045.
[28] C. Reble, I. Gersonde, C.A. Lieber, J. Helfmann, Influence of tissue absorption and scattering on the depth dependent sensitivity of Raman fiber probes investigated by Monte Carlo simulations, Biomed. Opt Express (2011), https://doi.org/10.1364/boe.2.000520.
[29] X. Feng, A.J. Moy, H.T.M. Nguyen, J. Zhang, M.C. Fox, K.R. Sebastian, J. S. Reichenberg, M.K. Markey, J.W. Tunnell, Raman active components of skin cancer, Biomed. Opt Express (2017), https://doi.org/10.1364/boe.8.002835.
[30] K. Hernández-Vidales, E. Guevara, V. Olivares-Illana, F.J. González, Characterization of wild-type and mutant p53 protein by Raman spectroscopy and multivariate methods, J. Raman Spectrosc. (2019), https://doi.org/10.1002/jrs.5655.
[31] D. Bovenkamp, R. Sentosa, E. Rank, M.T. Erkkilä, F. Placzek, J. Püls, W. Drexler, R.A. Leitgeb, N. Garstka, S.F. Shariat, C. Stiebing, I.W. Schie, J. Popp, M. Andreana, A. Unterhuber, Combination of high-resolution optical coherence tomography and Raman spectroscopy for improved staging and grading in bladder cancer, Appl. Sci. (2018), https://doi.org/10.3390/app8122372.
[32] I.J. Bigio, S.G. Bown, G. Briggs, C. Kelley, S. Lakhani, D. Pickard, P.M. Ripley, I. G. Rose, C. Saunders, Diagnosis of breast cancer using elastic-scattering spectroscopy: preliminary clinical results, J. Biomed. Opt. (2000), https://doi.org/10.1117/1.429990.
[33] B. Curry, D.E. Rumelhart, MSnet: a neural network which classifies mass spectra, Tetrahedron Comput. Methodol. (1990), https://doi.org/10.1016/0898-5529(90)90053-B.
[34] R. Chaber, C.J. Arthur, K. Łach, A. Raciborska, E. Michalak, K. Bilska, K. Drabko, J. Depciuch, E. Kaznowska, J. Cebulski, Predicting Ewing sarcoma treatment outcome using infrared spectroscopy and machine learning, Molecules (2019), https://doi.org/10.3390/molecules24061075.
[35] M. Erzina, A. Trelin, O. Guselnikova, B. Dvorankova, K. Strnadova, A. Perminova, P. Ulbrich, D. Mares, V. Jerabek, R. Elashnikov, V. Svorcik, O. Lyutakov, Precise cancer detection via the combination of functionalized SERS surfaces and convolutional neural network with independent inputs, Sensor. Actuator. B Chem. 308 (2020) 127660, https://doi.org/10.1016/j.snb.2020.127660.
[36] S. Quintero, R. Marino, J.M. Lanza-Gutierrez, F.J. Sanza, T. Riesgo, M. Holgado, A novel data processing technique for expert resonant nano-pillars transducers: a case study measuring ethanol in water and wine liquid matrices, IEEE Access 7 (2019) 129778–129788, https://doi.org/10.1109/ACCESS.2019.2939576.
[37] S. Pandit, T. Banerjee, I. Srivastava, S. Nie, D. Pan, Machine learning-assisted array-based biomolecular sensing using surface-functionalized carbon dots, ACS Sens. 4 (2019) 2730–2737, https://doi.org/10.1021/acssensors.9b01227.
[38] Y. Hong, Y. Li, L. Huang, W. He, S. Wang, C. Wang, G. Zhou, Y. Chen, X. Zhou, Y. Huang, W. Huang, T. Gong, Z. Zhou, Label-free diagnosis for colorectal cancer through coffee ring-assisted surface-enhanced Raman spectroscopy on blood serum, J. Biophot. (2020), https://doi.org/10.1002/jbio.201960176.
[39] H.J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N.J. Fullwood, B. Gardner, P.L. Martin-Hirsch, M.J. Walsh, M.R. McAinsh, N. Stone, F.L. Martin, Using Raman spectroscopy to characterize biological materials, Nat. Protoc. 11 (2016) 664–687, https://doi.org/10.1038/nprot.2016.036.
[40] P.J. Caspers, H.A. Bruining, G.J. Puppels, G.W. Lucassen, E.A. Carter, In vivo confocal Raman microspectroscopy of the skin: noninvasive determination of molecular concentration profiles, J. Invest. Dermatol. 116 (2001) 434–442, https://doi.org/10.1046/j.1523-1747.2001.01258.x.
[41] S. Gamsjaeger, P. Fratzl, E.P. Paschalis, Interplay between mineral crystallinity and mineral accumulation in health and postmenopausal osteoporosis, Acta Biomater. 124 (2021) 374–381, https://doi.org/10.1016/j.actbio.2021.02.011.
[42] S. Guo, C. Beleites, U. Neugebauer, S. Abalde-Cela, N.K. Afseth, F. Alsamad, S. Anand, C. Araujo-Andrade, S. Aškrabić, E. Avci, M. Baia, M. Baranska, E. Baria, L.A.E. Batista de Carvalho, P. de Bettignies, A. Bonifacio, F. Bonnier, E. M. Brauchle, H.J. Byrne, I. Chourpa, R. Cicchi, F. Cuisinier, M. Culha, M. Dahms, C. David, L. Duponchel, S. Duraipandian, S.F. El-Mashtoly, D.I. Ellis, G. Eppe, G. Falgayrac, O. Gamulin, B. Gardner, P. Gardner, K. Gerwert, E.J. Giamarellos-Bourboulis, S. Gizurarson, M. Gnyba, R. Goodacre, P. Grysan, O. Guntinas-Lichius, H. Helgadottir, V.M. Grošev, C. Kendall, R. Kiselev, M. Kölbach, C. Krafft, S. Krishnamoorthy, P. Kubryck, B. Lendl, P. Loza-Alvarez, F.M. Lyng, S. Machill, C. Malherbe, M. Marro, M.P.M. Marques, E. Matuszyk, C.F. Morasso, M. Moreau, H. Muhamadali, V. Mussi, I. Notingher, M.Z. Pacia, F.S. Pavone, G. Penel, D. Petersen, O. Piot, J. V Rau, M. Richter, M.K. Rybarczyk, H. Salehi, K. Schenke-Layland, S. Schlücker, M. Schosserer, K. Schütze, V. Sergo, F. Sinjab, J. Smulko, G. D. Sockalingum, C. Stiebing, N. Stone, V. Untereiner, R. Vanna, K. Wieland, J. Popp, T. Bocklitz, Comparability of Raman spectroscopic configurations: a large scale cross-laboratory study, Anal. Chem. (2020), https://doi.org/10.1021/acs.analchem.0c02696.
[43] D.K.R. Medipally, A. Maguire, J. Bryant, J. Armstrong, M. Dunne, M. Finn, F. M. Lyng, A.D. Meade, Development of a high throughput (HT) Raman spectroscopy method for rapid screening of liquid blood plasma from prostate cancer patients, Analyst 142 (2017) 1216–1226, https://doi.org/10.1039/C6AN02100J.
[44] X. Li, H. Chen, S. Zhang, H. Yang, S. Gao, H. Xu, L. Wang, R. Xu, F. Zhou, J. Hu, J. Zhao, H. Zeng, Blood plasma resonance Raman spectroscopy combined with multivariate analysis for esophageal cancer detection, J. Biophot. 14 (2021), e202100010, https://doi.org/10.1002/jbio.202100010.
[45] J.L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C.A. Ramírez-Alvarado, G. Gutiérrez-Juárez, C. Medina-Gutiérrez, Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients, Laser Med. Sci. 22 (2007) 229–236, https://doi.org/10.1007/s10103-006-0432-8.
[46] M. Paraskevaidi, C.L.M. Morais, D.E. Halliwell, D.M.A. Mann, D. Allsop, P. L. Martin-Hirsch, F.L. Martin, Raman spectroscopy to diagnose Alzheimer's disease and dementia with lewy bodies in blood, ACS Chem. Neurosci. 9 (2018) 2786–2794, https://doi.org/10.1021/acschemneuro.8b00198.
[47] A. Huefner, W.-L. Kuan, S.L. Mason, S. Mahajan, R.A. Barker, Serum Raman Spectroscopy as a Diagnostic Tool in Patients with Huntington's Disease †, 2020, https://doi.org/10.1039/c9sc03711j.
[48] C. Morasso, M. Truffi, R. Vanna, S. Albasini, S. Mazzucchelli, F. Colombo, L. Sorrentino, G. Sampietro, S. Ardizzone, F. Corsi, Raman analysis reveals biochemical differences in plasma of crohn's disease patients, J. Crohn's Colitis. 14 (2020) 1572–1580, https://doi.org/10.1093/ecco-jcc/jjaa080.
[49] R. Vanna, P. Ronchi, A.T.M. Lenferink, C. Tresoldi, C. Morasso, D. Mehn, M. Bedoni, S. Picciolini, L.W.M.M. Terstappen, F. Ciceri, C. Otto, F. Gramatica, Label-free imaging and identification of typical cells of acute myeloid leukaemia and myelodysplastic syndrome by Raman microspectroscopy, Analyst 140 (2015) 1054–1064, https://doi.org/10.1039/c4an02127d.

[50] P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, N. Stone, The use of Raman spectroscopy to differentiate between different prostatic adenocarcinoma cell lines, Br. J. Cancer 92 (2005) 2166–2170, https://doi.org/10.1038/sj.bjc.6602638.

[51] K. Lin, J. Wang, W. Zheng, K.Y. Ho, M. Teh, K.G. Yeoh, Z. Huang, Rapid fiber-optic Raman spectroscopy for real-time in vivo detection of gastric intestinal metaplasia during clinical gastroscopy, Cancer Prev. Res. 9 (2016) 476–483, https://doi.org/10.1158/1940-6207.CAPR-15-0213.

[52] I. Behl, G. Calado, A. Vishwakarma, D. Traynor, S. Flint, S. Galvin, C.M. Healy, M. L. Pimentel, A. Malkin, H.J. Byrne, F.M. Lyng, Classification of cytological samples from oral potentially malignant lesions through Raman spectroscopy: a pilot study, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 266 (2022) 120437, https://doi.org/10.1016/j.saa.2021.120437.

[53] R. Ullah, S. Khan, F. Farman, M. Bilal, C. Krafft, S. Shahzad, Demonstrating the application of Raman spectroscopy together with chemometric technique for screening of asthma disease, Biomed. Opt Express 10 (2019) 600–609, https://doi.org/10.1364/BOE.10.000600.

[54] L.E. Masson, C.M. O'Brien, R. Gautam, G. Thomas, J.C. Slaughter, M. Goldberg, K. Bennett, J. Herington, J. Reese, E. Elsamadicy, J.M. Newton, A. Mahadevan-Jansen, In vivo Raman spectroscopy monitors cervical change during labor, Am. J. Obstet. Gynecol. (2022), https://doi.org/10.1016/j.ajog.2022.02.019.

[55] C.A. Jenkins, S. Chandler, R.A. Jenkins, K. Thorne, F. Woods, A. Cunningham, K. Nelson, R. Still, J. Walters, N. Gwynn, W. Chea, R. Harford, C. O'Neill, J. Hepburn, I. Hill, H. Wilkes, G. Fegan, P.R. Dunstan, D.A. Harris, A new method to triage colorectal cancer referrals using serum Raman spectroscopy and machine learning, medRxiv (2020), https://doi.org/10.1101/2020.05.20.20108209, 2020.05.20.20108209.

[56] F. Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.-F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, Trends Anal. Chem. v. 124 (2020) 115796, https://doi.org/10.1016/j.trac.2019.115796, 2020 v.124.

[57] E. Ryzhikova, N.M. Ralbovsky, V. Sikirzhytski, O. Kazakov, L. Halamkova, J. Quinn, E.A. Zimmerman, I.K. Lednev, Raman spectroscopy and machine learning for biomedical applications: Alzheimer's disease diagnosis based on the analysis of cerebrospinal fluid, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 248 (2021) 119188, https://doi.org/10.1016/j.saa.2020.119188.

[58] M. Xu, J. Ma, Y. Qu, W. Mao, R. Zheng, Recognition of gastric cancer by Raman spectroscopy, Proc. SPIE (2009), https://doi.org/10.1117/12.845421.

[59] E. Guevara, J.C. Torres-Galván, M.G. Ramírez-Elías, C. Luevano-Contreras, F. J. González, Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools, Biomed. Opt Express 9 (2018) 4998–5010, https://doi.org/10.1364/BOE.9.004998.

[60] W. Lee, A.T.M. Lenferink, C. Otto, H.L. Offerhaus, Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection, J. Raman Spectrosc. 51 (2020) 293–300, https://doi.org/10.1002/jrs.5770.

[61] R. Kothari, V. Jones, D. Mena, V. Bermúdez Reyes, Y. Shon, J.P. Smith, D. Schmolze, P.D. Cha, L. Lai, Y. Fong, M.C. Storrie-Lombardi, Raman spectroscopy and artificial intelligence to predict the Bayesian probability of breast cancer, Sci. Rep. 11 (2021) 6482, https://doi.org/10.1038/s41598-021-85758-6.

[62] C.C. Horgan, M. Jensen, A. Nagelkerke, J.-P. St-Pierre, T. Vercauteren, M. M. Stevens, M.S. Bergholt, High-throughput molecular imaging via deep-learning-enabled Raman spectroscopy, Anal. Chem. 93 (2021) 15850–15860, https://doi.org/10.1021/acs.analchem.1c02178.

[63] A. Baliyan, H. Imai, A. Dager, O. Milikofu, T. Akiba, Automated hyperspectral 2D/3D Raman analysis using the learner-predictor strategy: machine learning-based inline Raman data analytics, Anal. Chem. 94 (2022) 637–649, https://doi.org/10.1021/acs.analchem.1c01966.

[64] E. Garai, S. Sensarn, C.L. Zavaleta, N.O. Loewke, S. Rogalla, M.J. Mandella, S. A. Felt, S. Friedland, J.T.C. Liu, S.S. Gambhir, C.H. Contag, A real-time clinical endoscopic system for intraluminal, multiplexed imaging of surface-enhanced Raman scattering nanoparticles, PLoS One 10 (2015), e0123185, https://doi.org/10.1371/JOURNAL.PONE.0123185.

[65] K.A. Dawson, Y. Yan, Current understanding of biological identity at the nanoscale and future prospects, Nat. Nanotechnol. 16 (2021) 229–242, https://doi.org/10.1038/s41565-021-00860-0, 2021 163.

[66] L. Fabris, SERS tags: the next promising tool for personalized cancer detection? ChemNanoMat 2 (2016) 249–258, https://doi.org/10.1002/cnma.201500221.

[67] E. Lenzi, L. Litti, D. Jimenez de Aberasturi, M. Henriksen-Lacey, L.M. Liz-Marzán, SERSTEM: an app for the statistical analysis of correlative SERS and TEM imaging and evaluation of SERS tags performance, J. Raman Spectrosc. 52 (2021) 355–365. https://doi.org/10.1002/jrs.6043.

[68] S.L. Jacques, Optical properties of biological tissues: a review, Phys. Med. Biol. 58 (2013) R37, https://doi.org/10.1088/0031-9155/58/11/R37.

[69] R.R. Anderson, J.A. Parrish, The optics of human skin, J. Invest. Dermatol. 77 (1981) 13–19, https://doi.org/10.1111/1523-1747.EP12479191.

[70] Z. Gajinov, M. Matić, S. Prćić, V. Đuran, Optical properties of the human skin/Optičke osobine ljudske kože, Serbian J. Dermatology Venerol. 2 (2013) 131–136, https://doi.org/10.2478/V10249-011-0029-5.

[71] P.Y. Liu, L.K. Chin, W. Ser, H.F. Chen, C.M. Hsieh, C.H. Lee, K.B. Sung, T.C. Ayi, P. H. Yap, B. Liedberg, K. Wang, T. Bourouina, Y. Leprince-Wang, Cell refractive index for cell biology and disease diagnosis: past, present and future, Lab Chip 16 (2016) 634–644, https://doi.org/10.1039/C5LC01445J.

[72] E. Lenzi, M. Henriksen-Lacey, B. Molina, J. Langer, C.D.L. de Albuquerque, D. Jimenez de Aberasturi, L.M. Liz-Marzán, ACS Sensors 7 (2022) 1747. https://doi.org/10.1021/acssensors.2c00610.

[73] E.D. SoRelle, O. Liba, J.L. Campbell, R. Dalal, C.L. Zavaleta, A. de la Zerda, Machine Learning-Assisted Hyperspectral Analysis of Plasmonic Contrast Agent Microbiodistribution with Single-Particle Sensitivity and Sub-cellular Resolution, 2017, pp. 73–82, https://doi.org/10.1117/12.2254117. Https://Doi.Org/10.1117/12.2254117. 10080.

[74] A. Kamińska, T. Szymborski, E. Witkowska, E. Kijeńska-Gawrońska, W. Świeszkowski, K. Niciński, J. Trzcińska-Danielewicz, A. Girstun, Detection of circulating tumor cells using membrane-based SERS platform: a new diagnostic approach for "liquid biopsy,", Nanomaterials 9 (2019) 366, https://doi.org/10.3390/nano9030366.

[75] S.K. Das, K. Pal, T.S. Bhattacharya, P. Karmakar, J. Chowdhury, Fabrication of SERS active Langmuir–Blodgett Film substrate for screening human cancer cell lines: experimental observations supported by multivariate data analyses, Sensor. Actuator. B Chem. 299 (2019) 126962, https://doi.org/10.1016/j.snb.2019.126962.

[76] A.B. Veloso, J.P.F. Longo, L.A. Muehlmann, B.F. Tollstadius, P.E.N. Souza, R. B. Azevedo, P.C. Morais, S.W. da Silva, SERS investigation of cancer cells treated with PDT: quantification of cell survival and follow-up, Sci. Rep. 7 (2017) 7175, https://doi.org/10.1038/s41598-017-07469-1.

[77] X. Ren, W. Nam, P. Ghassemi, J.S. Strobl, I. Kim, W. Zhou, M. Agah, Scalable nanolaminated SERS multiwell cell culture assay, Microsystems Nanoeng 6 (2020) 47, https://doi.org/10.1038/s41378-020-0145-3.

[78] L. Shen, Y. Du, N. Wei, Q. Li, S. Li, T. Sun, S. Xu, H. Wang, X. Man, B. Han, SERS studies on normal epithelial and cancer cells derived from clinical breast cancer specimens, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 237 (2020) 118364, https://doi.org/10.1016/j.saa.2020.118364.

[79] M. Hassoun, I.W. Schie, T. Tolstik, S.E. Stanca, C. Krafft, J. Popp, Surface-enhanced Raman spectroscopy of cell lysates mixed with silver nanoparticles for tumor classification, Beilstein J. Nanotechnol. 8 (2017) 1183–1190. https://doi.org/10.3762/bjnano.8.120.

[80] G. Kuku, M. Sarıçam, S. Mert, M. Çulha, Surface-enhanced Raman scattering from living cells: from differentiating healthy and cancerous cell to cytotoxicity assessment, Proc. SPIE (2015), https://doi.org/10.1117/12.2180466.

[81] X. Fang, Q. Zeng, X. Yan, Z. Zhao, N. Chen, Q. Deng, M. Zhu, Y. Zhang, S. Li, Fast discrimination of tumor and blood cells by label-free surface-enhanced Raman scattering spectra and deep learning, J. Appl. Phys. 129 (2021) 123103, https://doi.org/10.1063/5.0042662.

[82] S. Grieve, N. Puvvada, A. Phinyomark, K. Russell, A. Murugesan, E. Zed, A. Hassan, J.-F. Legare, P. C Kienesberger, T. Pulinilkunnil, T. Reiman, E. Scheme, K. R Brunt, Nanoparticle surface-enhanced Raman spectroscopy as a noninvasive, label-free tool to monitor hematological malignancy, Nanomedicine 16 (2021) 2175–2188, https://doi.org/10.2217/nnm-2021-0076.

[83] O. Guselnikova, A. Trelin, A. Skvortsova, P. Ulbrich, P. Postnikov, A. Pershina, D. Sykora, V. Svorcik, O. Lyutakov, Label-free surface-enhanced Raman spectroscopy with artificial neural network technique for recognition photoinduced DNA damage, Biosens. Bioelectron. 145 (2019) 111718, https://doi.org/10.1016/j.bios.2019.111718.

[84] P.H.L. Nguyen, B. Hong, S. Rubin, Y. Fainman, Machine learning for composition analysis of ssDNA using chemical enhancement in SERS, Biomed. Opt Express 11 (2020) 5092–5121, https://doi.org/10.1364/boe.397616.

[85] J. Morla-Folch, H.N. Xie, R.A. Alvarez-Puebla, L. Guerrini, Fast optical chemical and structural classification of RNA, ACS Nano 10 (2016) 2834–2842, https://doi.org/10.1021/acsnano.5b07966.

[86] H. Shi, H. Wang, X. Meng, R. Chen, Y. Zhang, Y. Su, Y. He, Setting up a surface-enhanced Raman scattering database for artificial-intelligence-based label-free discrimination of tumor suppressor genes, Anal. Chem. 90 (2018) 14216–14221, https://doi.org/10.1021/acs.analchem.8b03080.

[87] J.D. Driskell, R.A. Tripp, Label-free SERS detection of microRNA based on affinity for an unmodified silver nanorod array substrate, Chem. Commun. 46 (2010) 3298–3300, https://doi.org/10.1039/c002059a.

[88] E. Prado, N. Daugey, S. Plumet, L. Servant, S. Lecomte, Quantitative label-free RNA detection using surface-enhanced Raman spectroscopy, Chem. Commun. 47 (2011) 7425–7427, https://doi.org/10.1039/c1cc11925g.

[89] J.L. Abell, J.M. Garren, J.D. Driskell, R.A. Tripp, Y. Zhao, Label-free detection of micro-RNA hybridization using surface-enhanced Raman spectroscopy and least-squares analysis, J. Am. Chem. Soc. 134 (2012) 12889–12892, https://doi.org/10.1021/ja3043432.

[90] Y. Chen, Y. Chen, G. Chen, X. Zheng, C. He, S. Feng, X. Lin, R. Chen, H. Zeng, Discrimination of gastric cancer from normal by serum RNA based on surface-enhanced Raman spectroscopy (SERS) and multivariate analysis, Med. Phys. 39 (2012) 5664–5668, https://doi.org/10.1118/1.4747269.

[91] S. Nasir, M.I. Majeed, H. Nawaz, N. Rashid, S. Ali, S. Farooq, M. Kashif, S. Rafiq, S. Bano, M.N. Ashraf, M. Abubakar, S. Ahmad, A. Rehman, I. Amin, Surface enhanced Raman spectroscopy of RNA samples extracted from blood of hepatitis C patients for quantification of viral loads, Photodiagnosis Photodyn. Ther. 33 (2021) 102152, https://doi.org/10.1016/j.pdpdt.2020.102152, 1–11.

[92] S. Rafiq, M.I. Majeed, H. Nawaz, N. Rashid, U. Yaqoob, F. Batool, S. Bashir, S. Akbar, M. Abubakar, S. Ahmad, S. Ali, M. Kashif, I. Amin, Surface-enhanced Raman spectroscopy for analysis of PCR products of viral RNA of hepatitis C patients, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 259 (2021) 119908, https://doi.org/10.1016/j.saa.2021.119908, 1–8.

[93] W. Qin, F. Yang, C. Wang, Chemoproteomic profiling of protein-metabolite interactions, Curr. Opin. Chem. Biol. (2020), https://doi.org/10.1016/j.cbpa.2019.11.003.
[94] A.V. Aderemi, A.O. Ayeleso, O.O. Oyedapo, E. Mukwevho, Metabolomics: a scoping review of its role as a tool for disease biomarker discovery in selected non-communicable diseases, Metabolites 11 (2021), https://doi.org/10.3390/metabo11070418.
[95] J. Plou, I. García, M. Charconnet, I. Astobiza, C. García-Astrain, C. Matricardi, A. Mihi, A. Carracedo, L.M. Liz-Marzán, Multiplex SERS detection of metabolic alterations in tumor extracellular media, Adv. Funct. Mater. (2020), https://doi.org/10.1002/adfm.201910335.
[96] S. Chen, H. Zhang, X. Yang, X. Shao, T. Li, N. Chen, Z. Chen, W. Xue, J. Pan, S. Liu, Raman spectroscopy reveals abnormal changes in the urine composition of prostate cancer: an application of an intelligent diagnostic model with a deep learning algorithm, Adv. Intell. Syst. (2021), https://doi.org/10.1002/aisy.202000090.
[97] L. Zhou, Y. Liu, F. Wang, Z. Jia, J. Zhou, T. Jiang, L. Petti, Y. Chen, Q. Xiong, X. Wang, Classification analyses for prostate cancer, benign prostate hyperplasia and healthy subjects by SERS-based immunoassay of multiple tumour markers, Talanta 188 (2018) 238–244, https://doi.org/10.1016/j.talanta.2018.05.070.
[98] V. Moisoiu, A. Socaciu, A. Stefancu, S.D. Iancu, I. Boros, C.D. Alecsa, C. Rachieriu, A.R. Chiorean, D. Eniu, N. Leopold, C. Socaciu, D.T. Eniu, Breast cancer diagnosis by surface-enhanced Raman scattering (SERS) of urine, Appl. Sci. (2019), https://doi.org/10.3390/app9040806.
[99] L. Avram, S.D. Iancu, A. Stefancu, V. Moisoiu, A. Colnita, D. Marconi, V. Donca, E. Buzdugan, R. Craciun, N. Leopold, N. Crisan, I. Coman, D. Crisan, SERS-based liquid biopsy of gastrointestinal tumors using a portable Raman device operating in a clinical environment, J. Clin. Med. 9 (2020) 1–12, https://doi.org/10.3390/jcm9010212.
[100] F. Lussier, D. Missirlis, J.P. Spatz, J.F. Masson, Machine-learning-driven surface-enhanced Raman scattering optophysiology reveals multiplexed metabolite gradients near cells, ACS Nano (2019), https://doi.org/10.1021/acsnano.8b07024.
[101] H. Li, Q. Wang, J. Tang, N. Gao, X. Yue, F.R. Zhong, X. Lv, J. Fu, T. Wang, C. Ma, Establishment of a reliable scheme for obtaining highly stable SERS signal of biological serum, Biosens. Bioelectron. 189 (2021), https://doi.org/10.1016/j.bios.2021.113315.
[102] M. Ye, Y. Chen, Y. Wang, L. Xiao, Q. Lin, H. Lin, Z. Duan, S. Feng, Y. Cao, J. Zhang, J. Li, J. Hu, Subtype discrimination of acute myeloid leukemia based on plasma SERS technique, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 271 (2022), https://doi.org/10.1016/j.saa.2022.120865.
[103] Q. Wu, S. Qiu, Y. Yu, W. Chen, H. Lin, D. Lin, S. Feng, R. Chen, Assessment of the radiotherapy effect for nasopharyngeal cancer using plasma surface-enhanced Raman spectroscopy technology, Biomed. Opt Express 9 (2018) 3413, https://doi.org/10.1364/boe.9.003413.
[104] A. Stefancu, M. Badarinza, V. Moisoiu, S.D. Iancu, O. Serban, N. Leopold, D. Fodor, SERS-based liquid biopsy of saliva and serum from patients with Sjögren's syndrome, Anal. Bioanal. Chem. 411 (2019) 5877–5883, https://doi.org/10.1007/s00216-019-01969-x.
[105] X. Lin, D. Lin, Y. Chen, J. Lin, S. Weng, J. Song, S. Feng, High throughput blood analysis based on deep learning algorithm and self-positioning super-hydrophobic SERS platform for non-invasive multi-disease screening, Adv. Funct. Mater. 31 (2021) 2103382, https://doi.org/10.1002/adfm.202103382.
[106] Y. Chen, P. Yin, Z. Peng, Q. Lin, Y. Duan, Q. Fan, Z. Wei, High-throughput recognition of tumor cells using label-free elemental characteristics based on interpretable deep learning, Anal. Chem. 94 (2022) 3158–3164, https://doi.org/10.1021/acs.analchem.1c04553.
[107] N. Cheng, D. Chen, B. Lou, J. Fu, H. Wang, A biosensing method for the direct serological detection of liver diseases by integrating a SERS-based sensor and a CNN classifier, Biosens. Bioelectron. 186 (2021) 113246, https://doi.org/10.1016/j.bios.2021.113246.
[108] Z. Jin, Q. Yue, W. Duan, A. Sui, B. Zhao, Y. Deng, Y. Zhai, Y. Zhang, T. Sun, G.-P. Zhang, L. Han, Y. Mao, J. Yu, X.-Y. Zhang, C. Li, Intelligent SERS navigation system guiding brain tumor surgery by intraoperatively delineating the metabolic acidosis, Adv. Sci. (Weinheim, Baden-Wurttemberg, Ger. 9 (2022), e2104935, https://doi.org/10.1002/advs.202104935.
[109] M. Kazemzadeh, C.L. Hisey, K. Zargar-Shoshtari, W. Xu, N.G.R. Broderick, Deep convolutional neural networks as a unified solution for Raman spectroscopy-based classification in biomedical applications, Opt Commun. 510 (2022) 127977, https://doi.org/10.1016/j.optcom.2022.127977.
[110] J. Huang, J. Wen, M. Zhou, S. Ni, W. Le, G. Chen, L. Wei, Y. Zeng, D. Qi, M. Pan, J. Xu, Y. Wu, Z. Li, Y. Feng, Z. Zhao, Z. He, B. Li, S. Zhao, B. Zhang, P. Xue, S. He, K. Fang, Y. Zhao, K. Du, On-site detection of SARS-CoV-2 antigen by deep learning-based surface-enhanced Raman spectroscopy and its biochemical foundations, Anal. Chem. 93 (2021) 9174–9182, https://doi.org/10.1021/acs.analchem.1c01061.
[111] C. Carlomagno, D. Bertazioli, A. Gualerzi, S. Picciolini, P.I. Banfi, A. Lax, E. Messina, J. Navarro, L. Bianchi, A. Caronni, F. Marenco, S. Monteleone, C. Arienti, M. Bedoni, COVID-19 salivary Raman fingerprint: innovative approach for the detection of current and past SARS-CoV-2 infections, Sci. Rep. 11 (2021) 4943, https://doi.org/10.1038/s41598-021-84565-3.
[112] F.U. Ciloglu, A. Caliskan, A.M. Saridag, I.H. Kilic, M. Tokmakci, M. Kahraman, O. Aydin, Drug-resistant Staphylococcus aureus bacteria detection by combining surface-enhanced Raman spectroscopy (SERS) and deep learning techniques, Sci. Rep. 11 (2021) 18444, https://doi.org/10.1038/s41598-021-97882-4.
[113] S. Wang, H. Dong, W. Shen, Y. Yang, Z. Li, Y. Liu, C. Wang, B. Gu, L. Zhang, Rapid SERS identification of methicillin-susceptible and methicillin-resistant Staphylococcus aureus via aptamer recognition and deep learning, RSC Adv. 11 (2021) 34425–34431, https://doi.org/10.1039/D1RA05778B.
[114] E. Rho, M. Kim, S.H. Cho, B. Choi, H. Park, H. Jang, Y.S. Jung, S. Jo, Separation-free bacterial identification in arbitrary media via deep neural network-based SERS analysis, Biosens. Bioelectron. 202 (2022) 113991, https://doi.org/10.1016/j.bios.2022.113991.
[115] S. Yu, X. Li, W. Lu, H. Li, Y.V. Fu, F. Liu, Analysis of Raman spectra by using deep learning methods in the identification of marine pathogens, Anal. Chem. 93 (2021) 11089–11098, https://doi.org/10.1021/acs.analchem.1c00431.
[116] J. Hu, Y. Zou, B. Sun, X. Yu, Z. Shang, J. Huang, S. Jin, P. Liang, Raman spectrum classification based on transfer learning by a convolutional neural network: application to pesticide detection, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 265 (2022) 120366, https://doi.org/10.1016/j.saa.2021.120366.
[117] L. Guerrini, R.A. Alvarez-Puebla, Surface-enhanced Raman spectroscopy in cancer diagnosis, prognosis and monitoring, Cancers 11 (2019) 748, https://doi.org/10.3390/cancers11060748.