



## Article

# Machine Learning for COVID-19 Determination Using Surface-Enhanced Raman Spectroscopy

Tomasz R. Szyborski <sup>1,\*</sup>, Sylwia M. Berus <sup>1,†</sup>, Ariadna B. Nowicka <sup>2</sup>, Grzegorz Słowiński <sup>3</sup> and Agnieszka Kamińska <sup>1,\*</sup>

<sup>1</sup> Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, 01-224 Warsaw, Poland; sberus@ichf.edu.pl

<sup>2</sup> Institute for Materials Research and Quantum Engineering, Poznan University of Technology, Piotrowo 3, 60-965 Poznan, Poland; anowicka@ichf.edu.pl

<sup>3</sup> Department of Software Engineering, Warsaw School of Computer Science, Lewartowskiego 17, 00-169 Warsaw, Poland; gslowinski@ms.wysi.edu.pl

\* Correspondence: tszyborski@ichf.edu.pl (T.R.S.); akaminska@ichf.edu.pl (A.K.)

† These authors contributed equally to this work.

**Abstract:** The rapid, low cost, and efficient detection of SARS-CoV-2 virus infection, especially in clinical samples, remains a major challenge. A promising solution to this problem is the combination of a spectroscopic technique: surface-enhanced Raman spectroscopy (SERS) with advanced chemometrics based on machine learning (ML) algorithms. In the present study, we conducted SERS investigations of saliva and nasopharyngeal swabs taken from a cohort of patients (saliva: 175; nasopharyngeal swabs: 114). Obtained SERS spectra were analyzed using a range of classifiers in which random forest (RF) achieved the best results, e.g., for saliva, the precision and recall equals 94.0% and 88.9%, respectively. The results demonstrate that even with a relatively small number of clinical samples, the combination of SERS and shallow machine learning can be used to identify SARS-CoV-2 virus in clinical practice.

**Keywords:** surface-enhanced Raman spectroscopy; SERS; SARS-CoV-2; machine learning; random forest classifier



**Citation:** Szyborski, T.R.; Berus, S.M.; Nowicka, A.B.; Słowiński, G.; Kamińska, A. Machine Learning for COVID-19 Determination Using Surface-Enhanced Raman Spectroscopy. *Biomedicines* **2024**, *12*, 167. <https://doi.org/10.3390/biomedicines12010167>

Academic Editor: Natalia Komarova

Received: 23 November 2023

Revised: 23 December 2023

Accepted: 3 January 2024

Published: 12 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid, accurate, and inexpensive detection of viruses in clinical samples took on new importance with the outbreak of the COVID-19 pandemic [1]. Two types of tests for SARS-CoV-2 were in widespread use: (i) molecular diagnostic test for the detection of viral genetic material (RNA) in patient's sample and (ii) serologic test for the detection of immune response against the SARS-CoV-2 virus. The former (e.g., RT-PCR [2]) possess very high sensitivity and specificity; however, a number of factors can lead to incorrect results (e.g., virus mutation, PRC inhibition, etc.) and the technique is time consuming and requires expensive reagents [3,4]. The latter techniques (e.g., chemiluminescent methods, ELISA, antibody tests) strongly depend on the kinetics of SARS-CoV-2 antibody, and thus the optimization of the tests and interpretation of the results can be challenging. Additional detection methods for SARS-CoV-2 virus include optical biosensors [5], electrochemical [6–8], FET biosensor [9], colorimetric detection [10], and others [11–13].

In recent years, spectroscopic techniques such as surface-enhanced Raman spectroscopy (SERS) are increasingly being used for the detection of viruses using indirect and direct methods [14–17]. Indirect detection uses an SERS tag: Raman reporter molecule (e.g., p-MBA) with the recognition element (specific antibody) and capture substrate, which is functionalized with a capture antibody. The measurement is based on tracking the Raman signal from the reporter molecule, which is a more sensitive and specific method than the direct method. SERS is well known for its incredibly high sensitivity, allowing

for the detection of even a single molecule. In addition, it is a fast, simple, reagent free, and non-destructive method for sample analysis. For these reasons, SERS recently gained attention as a prospective technique in biosensing [18], where SARS-CoV-2 is of special interest. Particularly promising is the combination of the highly efficient SERS technique with advanced chemometric methods, especially those based on artificial intelligence methods (e.g., machine learning or deep learning).

Machine learning was previously demonstrated for the identification of bacteria using a combination of SERS and deep learning algorithms [19–21], the detection of biomarkers for head and neck cancer samples [22], and the detection of Sjogren's syndrome and diabetic nephropathy [23]. SERS and machine learning techniques were also used to differentiate between samples containing influenza A and B viruses with an accuracy of 93% at a concentration of 200 µg/mL [24]. One of the cutting-edge trends in analytics is combining SERS with advanced chemometric, especially in machine learning methods [25,26]. This approach is especially attractive for the detection and identification of the SARS-CoV-2 virus [27–30], especially in clinical samples [31]. Ikponmwoba et al. [32] used SERS and machine learning for the detection of COVID-19 in biological samples obtained from 20 patients. The authors performed dimensionality reduction using PCA and nonlinear dimensionality reduction using UMAP (Uniform Manifold Approximation and Projection). Finally, they used Gaussian process (GP) classification to predict the occurrence of a negative or positive sample as a function of low-dimensional space variables. The GP classifier provided a probability, either 0 or 1, to classify whether a sample was COVID-negative or COVID-positive. Karunakaran et al. [33] demonstrated the use of the label-free SERS technique and machine learning for the analysis of saliva samples in healthy COVID-19-infected and COVID-19-recovered patients. The proposed method could also differentiate the three classes of corona virus spike protein, i.e., SARS-CoV-2, SARS-CoV, and MERS-CoV. The authors used trained support vector machine (VSM) and achieved high accuracy predictions for healthy and COVID-19-infected patients, respectively. Yang et al. [14] developed a label-free diagnostic platform, which combines SERS and machine learning algorithms for the detection of thirteen viruses, e.g., SARS-CoV-2, SARS-CoV-2 B1, CoV 229E, IBV, HMPV-A, and others. The samples were prepared in the laboratory, where the viruses were propagated in Vero E6 cells, and after 48 h, the viruses were harvested with the procedure described in the article. The authors used support vector machine (SVM), k-nearest neighbor, and random forest algorithms. Yang et al. [34] developed a sensor with a deep learning algorithm for the detection of SARS-CoV-2 in human nasopharyngeal swabs. The authors prepared the sensor using a silver nanorod array substrate by assembling DNA probes to capture SARS-CoV-2 RNA. For chemometric analysis, the authors used a recurrent neural network (RNN)-based deep learning model. They classified 40 positive and 120 negative samples with an accuracy of 98.9%. Hwang et al. [35] reported an SERS face mask for the label-free detection of the aerosolized SARS-CoV-2 virus using Au-TiO<sub>2</sub> nanocomposites. The SERS platform was placed on the inside of the face mask, where the Au-TiO<sub>2</sub> SERS face mask continuously preconcentrated and efficiently captured the oronasal aerosols. An autoencoder neural network was then employed for the accurate classification of the SARS-CoV-2 virus at various concentrations. Ansah et al. [36] presented the identification of viruses via a combination of SERS with pathogen-mediated composite materials on Au nanodimple electrodes and ML. Viruses were trapped in 3D plasmonic concave spaces via electrokinetic pre-concentration, leading to ultrasensitive SERS detection. The authors used two ML models, SVM and CNN, for the specific identification of eight virus species, including influenza A viruses, human rhinovirus, and human coronavirus.

The same set of spectral data presented in this manuscript, both saliva and nasopharyngeal swabs, have been analyzed by the authors through chemometric methods using commercial software Unscrambler® (CAMO software AS, version 10.3, Oslo, Norway). The results and the use of typical chemometrics in the analysis of clinical samples infected with SARS-CoV-2 virus was presented by Berus et al. [37] The analysis was divided into two steps: (i) preparation and optimization of calibration models and (ii) external validation

using samples with an unknown origin (CoV+ or CoV−) for checking the classification abilities of previously created calibration models. Based on the analysis with different methods (PLS-DA, SVMC, and PCA-LDA), we indicated that the best COVID-19 diagnosis can be delivered via the SVMC method over saliva samples. Such an analysis results in impressive diagnostic parameters both on the calibration (sensitivity 100%, specificity 100%, and accuracy 100%) as well as the validation step (sensitivity 100%, specificity 80%, and accuracy 90%). The PCA-LDA and PLS-DA methods also deal well with saliva samples as the accuracy equals 80% and the sensitivity 90% or 100% (validation). We also demonstrated that the nasopharyngeal swabs can be an equally convenient methodology as SVMC presents sensitivity and accuracy at the level of 88% and 75%, respectively. The relatively low specificity of 69% can be upgraded via PLS-DA and PCA-LDA methods, which would result in 75% for both cases.

This article demonstrates the SERS technique for a fast and simple measurement of the clinical samples: saliva and nasopharyngeal swabs. In the next step, we applied machine learning algorithms, namely gaussian naive Bayes (GNB), random forest (RF), support vector classifier (SVC), and logistic regression (LR). We tested them regarding classification abilities expressed by diagnostic parameters (accuracy, precision, and recall). The results show that ML, even for sets of data considered as small, can classify samples with reasonable precision and accuracy. Lastly, we compared the results with the analysis performed using typical chemometric methods. This study is an extension of the previous one by Berus et al. [37], where chemometric methods (via Unscrambler software, CAMO software AS, version 10.3, Oslo, Norway) were tested on the same spectral data. With support machine classification, we reached the sensitivity and accuracy of 100% and 90%. This research is also an ideal complement as the RF algorithm can improve the precision to 90%.

## 2. Materials and Methods

### 2.1. Clinical Samples

Clinical samples from 289 patients of saliva and nasopharyngeal swabs were collected from the Department of Clinical Genetics, Medical University of Łódź (Łódź, Poland) and stored in frozen form at  $-80\text{ }^{\circ}\text{C}$ . The samples, prior to storage at low temperature, were tested for SARS-CoV-2 via the qRT-PCR method, according to the guidelines of the Department of Clinical Genetics, Medical University of Łódź (Łódź, Poland). The extraction of viral RNA of SARS-CoV-2 was performed with the use of chemagic 360 automated extraction platform (PerkinElmer, Naperville, IL, USA). To verify the presence of SARS-CoV-2 in the samples, the following methods were used: qRT-PCR amplification of open reading frame 1ab (ORF1ab), nucleocapsid protein (NP) gene fragments, and positive reference gene using DiaPlexQ Novel Coronavirus (2019-nCoV) Detection Kit (SolGent Co, Ltd., Daejeon, Republic of Korea). Detailed information on the extraction of RNA and qRT-PCR measurements can be found in our previous article by Berus et al. [37].

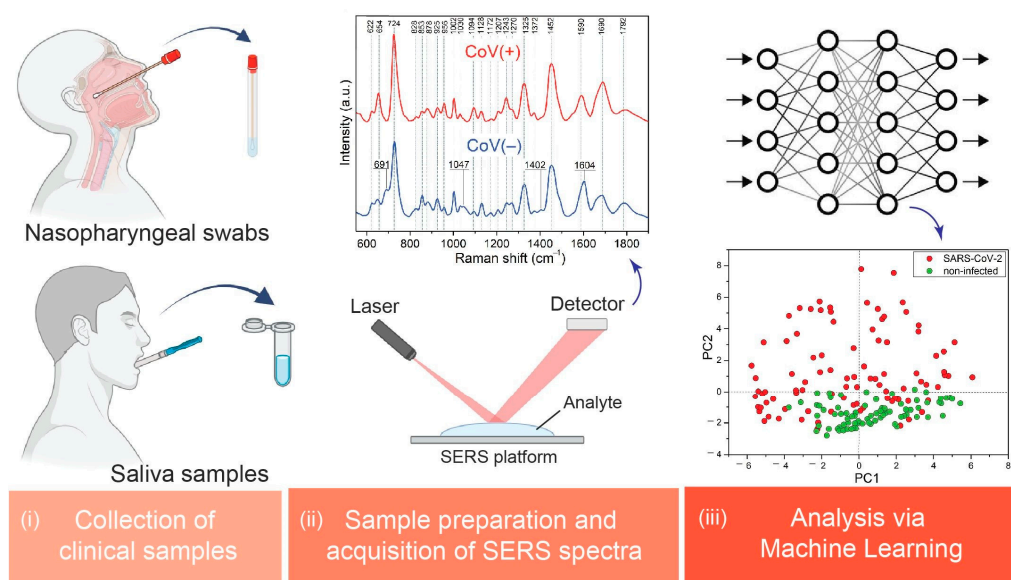
### 2.2. Measurements of the Clinical Samples

SERS measurements require using cost-effective SERS platforms in large quantities, which provide high enhancement factor (EF), reproducibility, and stability over time. For this purpose, we have used SERS platforms based on femtosecond laser-modified silicon [38]. The silicon was modified using a femtosecond laser ( $\lambda = 1030\text{ nm}$ ) with a repetition rate of 300 kHz and a pulse of 300 femtoseconds. The modification occurred on mechanically pre-cut silicon squares ( $3\text{ mm} \times 3\text{ mm}$ ), thus obtaining a reproducible and large number of SERS platforms for later use. The final step of the procedure was the sputtering of 100 nm of silver using the PVD device (Quorum, Q150T ES, Laughton, UK) using 25 mA current. SERS-active platforms were stored in an inert gas atmosphere.

The samples were conditioned at room temperature, and then ca.  $2\text{ }\mu\text{L}$  of liquid from each sample was pipetted onto the SERS-active platform. The platform with the liquid sample was attached to a glass slide and placed under a laminar flow cabinet, typically for

2–3 min, to evaporate the water. Then, the glass slide was placed under a spectrometer so the laser beam was in the very center of the SERS platform. The measurements were performed using a BRAVO (Bruker, Rosenheim, Germany) spectrometer equipped with a Duo Laser system (700–1100 nm, 100 mW) and a CCD camera. The spectral resolution was 2–4  $\text{cm}^{-1}$ . Typically, 30 SERS spectra were recorded for a single sample, and the time of acquisition for a single spectrum was 30 s.

SERS spectra were pre-processed using OPUS software (Bruker Optic GmbH, ver. 2012). The raw spectra were subjected to smoothing using a Savitzky–Golay filter (five points), baseline concave rubber band correction with six iterations (five baseline points), cutting to the range between 600  $\text{cm}^{-1}$  and 1700  $\text{cm}^{-1}$ , and finally, Min–Max normalization. Such pre-processed spectra were subjected to machine learning analysis. The whole procedure of measurement and analysis is presented in Figure 1.



**Figure 1.** Procedure consists of three main steps: (i) collection of clinical sample, (ii) sample preparation and spectra acquisition, and (iii) machine learning analysis. Figure created with BioRender.com.

### 2.3. Machine Learning Analysis

#### 2.3.1. Principal Component Analysis (PCA)

In the Principal Component Analysis (PCA) method, the correlated data are reduced into uncorrelated data presented in a dimension described by so-called principal components (PCs). The method is based on bilinear decomposition mathematically described as:

$$X = TP^T + E$$

where:

X—Initial matrix of data;

T—Scores matrix;

P—Loading matrix;

E—Error matrix.

PC scores are related to the linear combination of the original variables and describe the differences and similarities between samples. The first principal component (PC-1) accounts for the most significant variance in the data. The loadings describe the data structure concerning the correlation of the variables and show how well a PC takes the variation of these variables into account. By analyzing the plot of PC loadings as a function of variables (i.e., Raman shifts), one can indicate the main diagnostic variables or regions related to the differences in the dataset [39].

### 2.3.2. Machine Learning Classification

Machine learning (ML) classification experiments were performed with selected ML methods: gaussian naive Bayes (GNB) classifier [40], random forest (RF) [41], support vector classifier (SVC) [42], and logistic regression (LR) [43]. Default parameters were used for all classifiers. The only exception was performed for the logistic regression classifier that needed more than 100 default iterations to converge, and the number of maximal iterations was increased to 500. Data were analyzed via shallow learning techniques using Python and its frameworks for data manipulation, visualization and machine learning: numpy, pandas, matplotlib, seaborn, and scikit-learn.

## 3. Results

### 3.1. SERS Measurements and Band Assignments

In the current study, we examined clinical samples taken from 289 patients, including 175 samples of saliva and 114 samples of nasopharyngeal swabs. All patients were diagnosed with the PCR method and, depending on the result, divided into COVID-19(+) (infected with SARS-CoV-2 virus) and COVID-19(−) (non-infected with SARS-CoV-2 virus) samples. For clarity, we have labeled these samples as CoV(+) and CoV(−), respectively. The summary of the used samples is demonstrated in Table 1.

**Table 1.** Sets of saliva and nasopharyngeal samples.

Type	Total Number of Samples	CoV(+)	CoV(−)
saliva	175	81	94
nasopharyngeal swab	114	51	63
total:	289	132	157

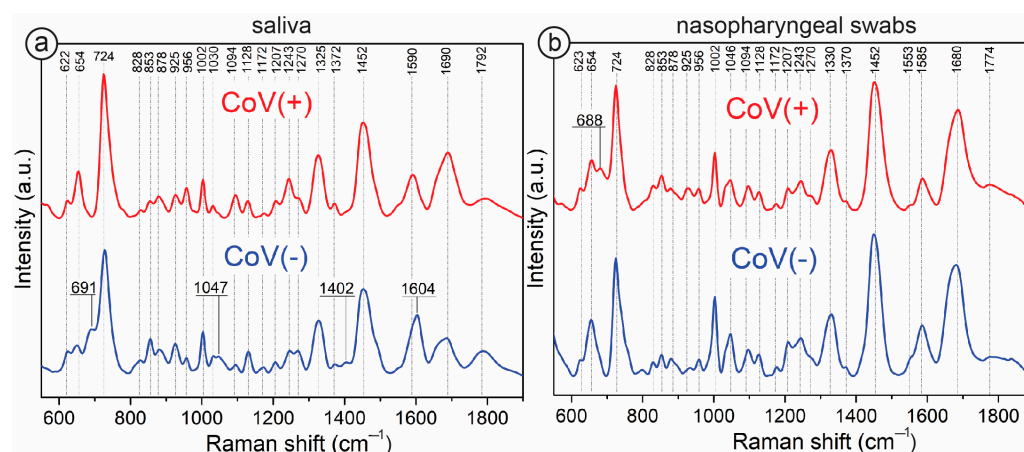
Figure 2 demonstrates averaged SERS spectra of saliva (a) and nasopharyngeal swabs (b). The figure contains a spectrum of infected samples labeled as CoV(+) and non-infected samples labeled as CoV(−). Dashed lines mark assigned bands, whereas the continuous lines with the band value show the bands that are present in the sample (e.g., CoV(−)) and simultaneously do not exist in the other sample type, i.e., CoV(+). When comparing the averaged spectra of samples infected with SARS-CoV-2 and non-infected samples (CoV(−)), both saliva and nasopharyngeal swabs, we observe apparent differences between them. For CoV(−) samples of saliva (Figure 2a: CoV(−)), the most characteristic bands are at 691, 724, 853, 878, 1002, 1047, 1128, 1270, 1325, 1452, 1590, 1690, and 1792  $\text{cm}^{-1}$ . All bands were identified and their origin was described in Table S1 (see Supplementary Materials). Here, we present the origins of the most intense bands:

- (i) An origin of 724  $\text{cm}^{-1}$  corresponds to O–O stretching vibration in oxygenated proteins, glycoproteins (e.g., mucin), and to the ring breathing mode of tryptophan.
- (ii) An origin of 1325  $\text{cm}^{-1}$  corresponds to amide III band in proteins and DNA.
- (iii) An origin of 1452  $\text{cm}^{-1}$  corresponds to C–H stretching of glycoproteins, including mucin.
- (iv) An origin of 1585  $\text{cm}^{-1}$  corresponds to ring and C=C vibrations in tyrosine and phenylalanine.

Spectral changes between CoV(+) and CoV(−) reflect the differences in biochemical composition and provide information between components in saliva (e.g., the intensity ratio of 853/828  $\text{cm}^{-1}$  can be explained by the interaction between tyrosine residues with viral proteins, other expressed molecules, and immunity proteins) [44,45]. The SERS fingerprint of saliva infected with the SARS-CoV-2 virus is characterized by bands at 654, 720, 1320, and 1443  $\text{cm}^{-1}$ , which can be assigned to specific oscillations in methionine and methionine adenosyl transferase [46]. We observed an increased intensity of these four bands for CoV(+) saliva. In relation to the band 1002  $\text{cm}^{-1}$  with fixed intensities, the ratios are as follow: CoV(+)  $I_{654}/I_{1002} = 1.3$ ;  $I_{720}/I_{1002} = 4.17$ ;  $I_{1320}/I_{1002} = 1.72$ ;  $I_{1445}/I_{1002} = 2.72$  and



CoV(−)  $I_{654}/I_{1002} = 0.61$ ;  $I_{720}/I_{1002} = 3.3$ ;  $I_{1320}/I_{1002} = 1.32$ ;  $I_{1445}/I_{1002} = 2.25$ . This can be explained by the increased requirement for methionine during infection [47]. Also, recent studies demonstrated that the level of ferritin in saliva can rise during infection [48]. These increased levels of ferritin and specific immunoglobulins in saliva lead to intensified bands in region  $1200\text{--}1300\text{ cm}^{-1}$ , as well as bands at  $1325$ ,  $1450$ , and  $1690\text{ cm}^{-1}$  from amide III and amide I of proteins. Some band assignments can have multiple origins. Bands at  $1094$ ,  $1242$ , and  $1325\text{ cm}^{-1}$  originate from the phosphodiester group and purine bases of nucleic acids. Their increased intensity in CoV(+) samples of saliva can be explained by the multiplication of genetic material during infection.



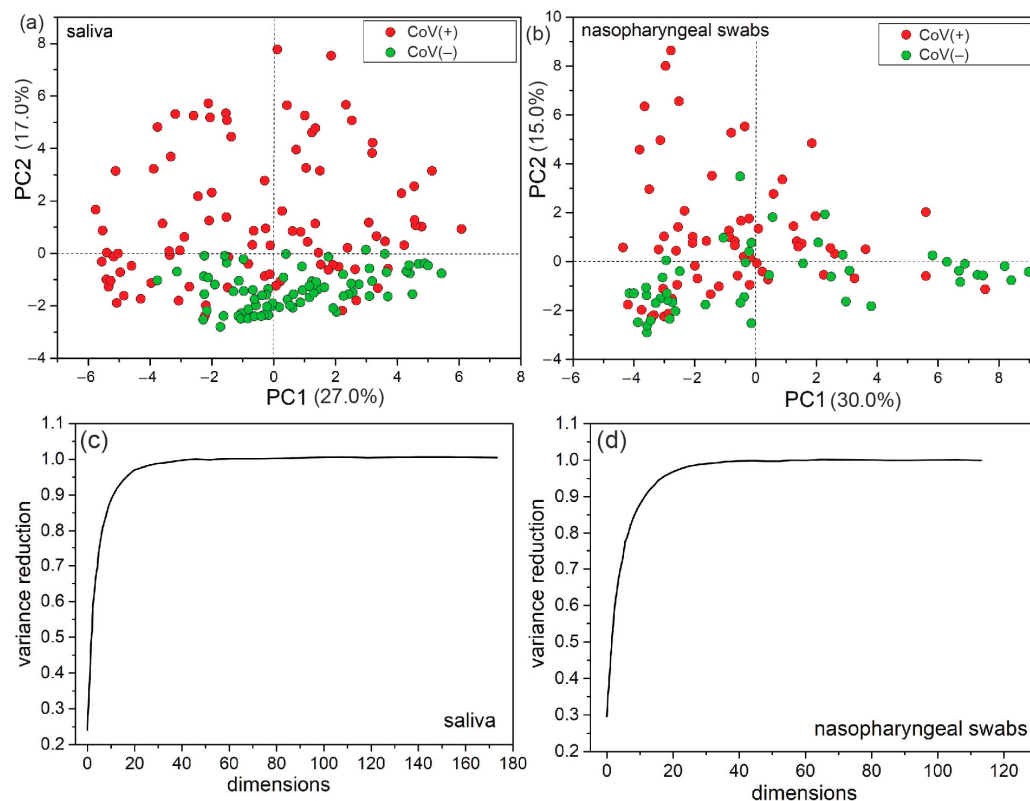
**Figure 2.** SERS spectra of COVID-19 infected (CoV(+)) and non-infected samples (CoV(−)) from saliva (a) and nasopharyngeal swabs (b).

The analysis of nasopharyngeal swabs demonstrated that in CoV(+) and CoV(−) samples, the strong SERS bands are located at  $724$ ,  $1002$ ,  $1045$ ,  $1330$ ,  $1452$ ,  $1590$ , and  $1680\text{ cm}^{-1}$ . A distinct difference is observed between  $680\text{ cm}^{-1}$  and  $950\text{ cm}^{-1}$ . For CoV(+) spectra, a new band appears at  $688\text{ cm}^{-1}$ , which is characteristic of neopterin. An increase in the  $925\text{ cm}^{-1}$  band, assigned to carboxylates and proline rings compounds, is also observed [49,50]. The relative spectral intensity of the bands  $745\text{ cm}^{-1}$  to  $1455\text{ cm}^{-1}$  demonstrates efficacy for the classification of the samples. CoV(+) samples demonstrate an intensity ratio of  $I_{724}/I_{1455}$  at the level  $1.03 \pm 0.05$ , whereas for CoV(−), this ratio is  $1.23 \pm 0.04$ .

The above analysis demonstrates that the analyzed biological samples are characterized by biochemical complexity and variability from patient to patient. SERS spectra, complex to analyze empirically, are excellent research material for chemometric analyses, especially those based on machine learning and artificial intelligence.

### 3.2. Principal Component Analysis (PCA)

In the first step, Principal Component Analysis (PCA) was used to reduce the dimensionality of many uncorrelated spectral data into correlated ones. In a new dimension described by principal components (PC-1, PC-2), every single spectrum is represented by a single point, and the dependencies between them are more visible. In a plot score (Figure 3), we observe that saliva samples are characterized by a more symmetrical distribution than nasopharyngeal swabs. Moreover, the spectral information explained by PC-1 and PC-2 is higher and equals  $44\%$ , while for nasopharyngeal swabs, it is  $45\%$ . Thus, we can conclude that PCA works better for saliva samples regarding CoV(+) and CoV(−) differentiation.



**Figure 3.** Visualization of saliva (a) and nasopharyngeal swab (b) datasets after PCA transformation to 2D space. Red dots stated as samples infected with the SARS-CoV-2 virus (CoV(+)), and green dots represent non-infected samples (CoV(-)). (c) Visualization of saliva and (d) nasopharyngeal swab variance reduction as a function of dimensions. Variance in reduction increases linearly with an increase in dimensions and achieved a plateau at dimensions of ca. 25–30.

Dataset dimensionality can be reduced from over 500 to a smaller number with insignificant variance reduction (see Figure 3c,d). For both types of samples, the dimensionality of 20 has been chosen as the optimal value, providing the variance perseverance at the level of 96.3% and 96.3% for saliva and nasopharyngeal swabs, respectively. We considered a value of 20 to be optimal as further size increases do not lead to a significant improvement in variance reduction. A change from 20 to 40, i.e., 100%, results in an improvement of only 3%. The variance perseverance ratio as a function of a number of dimensions for both saliva and nasopharyngeal swabs are demonstrated in Table 2.

**Table 2.** Variance perseverance ratio for selected dimensions during PCA transformation for saliva and nasopharyngeal swabs. For both types of samples, the optimal number of dimensions was set to 20.

Number of Dimensions	Variance Perseverance Ratio (%)	
	Saliva	Nasopharyngeal Swab
1	24.5	30.0
2	41.3	44.5
5	70.0	70.0
10	86.8	86.3
20	96.3	96.3
30	98.6	98.8
40	99.4	99.5

### 3.3. Classification of the Samples Using Machine Learning Algorithms

The second step in the classification of the clinical samples was the testing of selected methods of ML algorithms. For preliminary assessment, we selected four methods: Gaussian naive Bayes (GNB), random forest (RF), support vector classifier (SVC), and logistic regression (LR).

Single ML experiment results can be influenced by dataset random division to train and test subsets. To mitigate this effect, five-fold cross-validation has been applied with a train/test division proportion equal to 80/20. As performance metrics, we used precision, recall, and balanced accuracy. The datasets are slightly imbalanced; positive and negative class sizes are not equal, and for that reason, balanced accuracy is a better metric in such cases than ordinary accuracy. The precision, recall, and adjusted balanced accuracy values of all tested classifiers are displayed in Table 3, whereas Figure 4 presents their graphical representation as bar graphs. Figure 4 also consists of a dashed line representing the mean value for the SVMC method, which was the best chemometric method of analysis in our previous article [37].

**Table 3.** Results of cross-validation of different ML techniques for saliva and nasopharyngeal swabs. Minimal, maximal, and average (bold) results are presented.

Classifier Type	Saliva			Nasopharyngeal Swabs		
	Precision (%)	Recall (%)	Adjusted * Balanced Accuracy (%)	Precision (%)	Recall (%)	Adjusted ** Balanced Accuracy (%)
gaussian naive Bayes (GNB)	71.4	68.7	79.1	50	63.6	58.1
	85.7	100	92.1	71.4	100	84.6
	<b>81.4</b>	<b>83</b>	<b>83</b>	<b>62.6</b>	<b>82.7</b>	<b>71.5</b>
random forest (RF)	85.7	62.5	78.6	38.4	45.4	39.4
	93.7	93.7	94.2	90	90	90.8
	<b>90.4</b>	<b>81.6</b>	<b>87.1</b>	<b>65.7</b>	<b>61.1</b>	<b>66.1</b>
support vector classifier (SVC)	75	68.7	74.3	46.1	60	53.1
	93.7	100	91.1	81.8	90.9	86.7
	<b>85</b>	<b>87.7</b>	<b>85.8</b>	<b>69</b>	<b>78.2</b>	<b>74</b>
logistic regression (LR)	80	62.5	78.6	41.2	50	40.1
	92.8	100	89.5	77.8	70	77.3
	<b>87.1</b>	<b>85.1</b>	<b>86.7</b>	<b>61.2</b>	<b>62.7</b>	<b>62.8</b>

\* The result is adjusted for chance, so that random performance would score 0, while keeping perfect performance at a score of 1. \*\* The result is adjusted for chance, so that random performance would score 0%, while keeping perfect performance at a score of 100%.

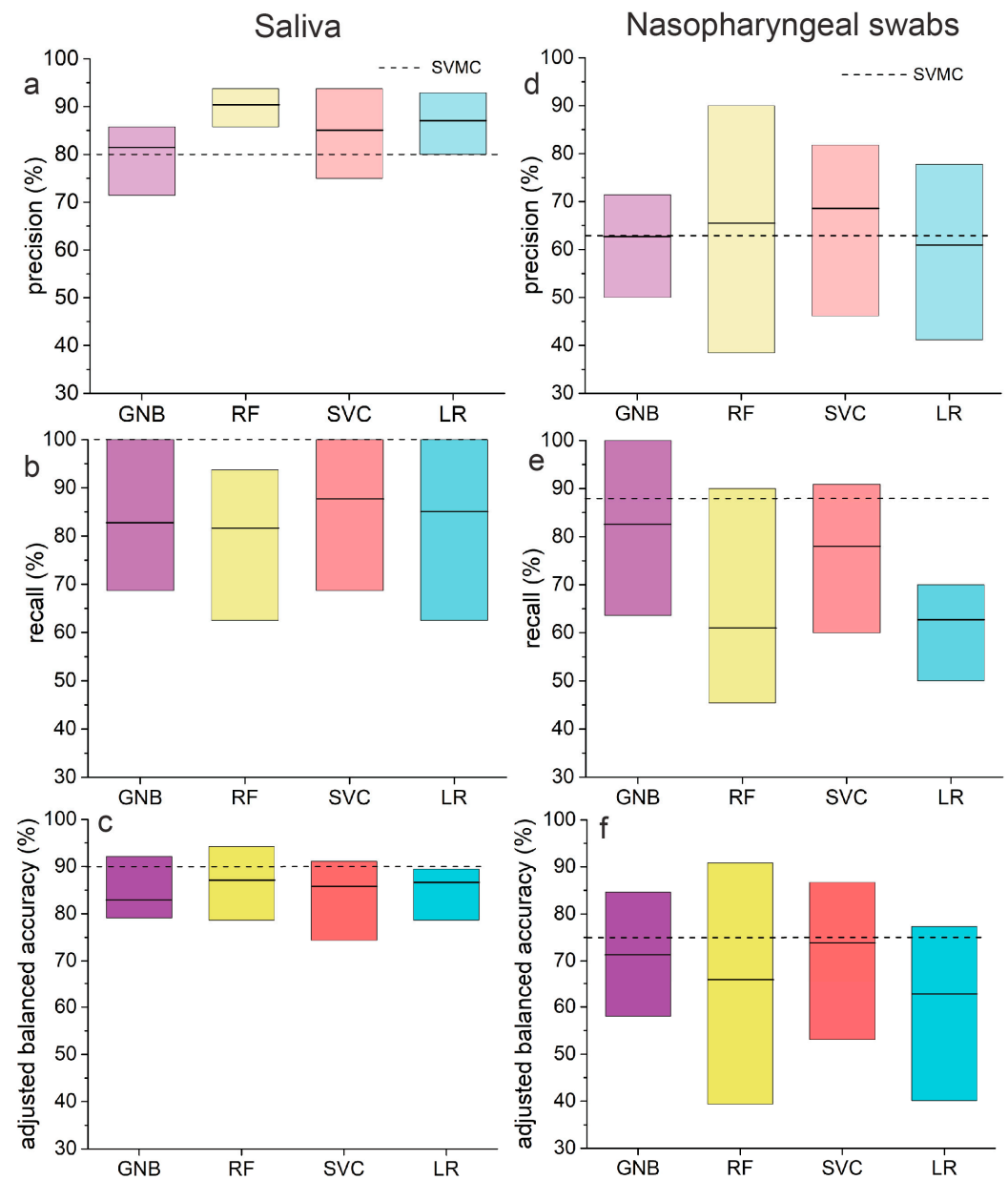
In the case of saliva, the highest averaged parameters of precision and adjusted balanced accuracy are obtained for the RF algorithm. The value of the recall parameter reaches 81.6%, similar to GNB (82.9%). LR algorithm provides high averaged values for all parameters, precision 87.1%, recall 85.1%, and adjusted balance accuracy 86.7%, which are very similar to the one obtained from the SCV. GNB algorithm works least effectively among all considered methods as it provides the lowest averaged values of precision and adjusts balanced accuracy that equals 81.4% and 83.0%, respectively.

Considering the range of values that all tested algorithms can reach, we conclude that RF offers the highest maximal value of precision with the smallest spread of values (85.7–93.7%) and a maximum balanced accuracy of 94.2%. Although the maximum recall is 93.7%, which is lower than other methods, RF performs within 78.6–94.2% of adjusted balanced accuracy with an average value of 87.1%, which is the best result.

For nasopharyngeal swabs, SVC can identify the highest CoV(−) cases and ensure proper diagnosis as the averaged values of precisions and accuracy equals 69.0% and 74.0%, respectively. In turn, GNB can identify CoV(+) cases with the recall of 82.7%, and this is the best working algorithm for the determination of the presence of SARS-CoV-2 in



nasopharyngeal swabs. The accuracy is also relatively high (71.6%). With LR and FR, the averaged values of all parameters are in the range of 61.0% and 66.0%.



**Figure 4.** Precision (a,d), recall (b,e) and adjusted balanced accuracy (c,f) for the saliva and nasopharyngeal swabs samples. The figure presents: gaussian naive Bayes (GNB), random forest (RF), support vector classifier (SVC) and logistic regression (LR). This dataset was previously analyzed using standard chemometric methods [37], where SVMC gave the best results. The mean value obtained with SVMC is presented as a dashed line.

For random forest, we noted the highest spread of the values of all parameters in the range between 40.0% and 90.0%. RF is the only algorithm capable of reaching the maximal values of precision and adjusted balance. GNB is characterized by the highest values of recall 63.6–100.0%, whereas LR has the lowest range: 50.0–70.0%.

To compare the results obtained using Unscrambler software (Berus et al. [37]) with the results presented here, we superimposed the data of the best-performing method—SVMC (support vector machine classification) in Figure 4 and marked them with a dashed line. The values of precision, recall, and accuracy were calculated using previously created

and optimized calibration models for saliva and nasopharyngeal swabs. The number of external samples used for testing the predictive abilities of models was 20 and 16 for saliva and nasopharyngeal swabs, respectively. Considering the averaged values, all algorithms operated via machine learning can upgrade the precision provided by the SVMC technique. It is advantageous while analyzing saliva samples with the RF method (precision 90.4%). For nasopharyngeal swabs, these differences are not so substantial as all values oscillate in the range between 61.0% and 68.0%. The recall offered by SVMC is 100.0% and 88.0% and is higher than all tested methods. The values of adjusted balanced accuracy offered by all methods (machine learning; SVMC) are comparable but still higher for SVMC.

All tested algorithms perform better in recognizing and identifying the SARS-CoV-2 virus in saliva samples than in nasopharyngeal swabs, as the precision, recall, and accuracy values are higher. This makes saliva a more reliable material for COVID-19 diagnosis. Since it is minimally invasive and does not cause damage during intake, this method of analysis would be preferred.

### 3.4. Classification of Saliva Samples via Random Forest

According to cross-validation results (see Section 3.3), where we tested different methods, the best classification results were achieved for saliva samples and random forest classifier. Herein, we present detailed information about using saliva samples and RFC to detect SARS-CoV-2 virus efficiently. When analyzing these samples, we used the classic approach of dividing the samples into training and testing sections.

The single training results and performance achieved depend to some extent on the random samples split to train or test sets. The experience shows that this effect is more substantial with smaller sets. Figure 5a,b show a confusion matrix for single training. The full dataset contained 175 samples, and 35 samples (17 from healthy patients and 18 infected with SARS-CoV-2 virus) were put into the test set. Diagonal fields in the matrix show correctly classified samples: true negatives (TNs) and true positives (TPs). There could also be two types of mistakes. Positive samples (infected with SARS-CoV-2) could, by mistake, be classified as healthy, which are called false negatives (FNs), and truly negative samples could be recognized as positive—and are thus considered false positives (FPs).

Precision and recall were calculated according to the following equations:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

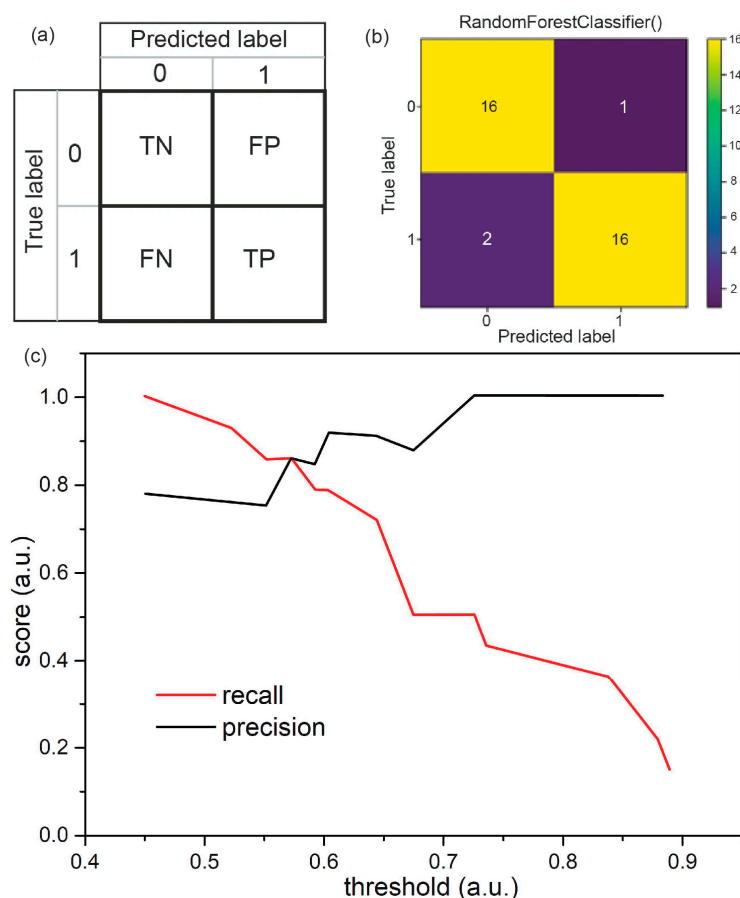
where TP is true positive; FN is false negative; and FP is false positive.

The F-score, which is the harmonic mean of the precision and recall, can be calculated as follows:

$$F = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

In the case of results presented in Figure 4, the precision is 94.1%, the recall (sensitivity) is equal to 88.9 %, and the F-score is 0.914 (i.e., 91.4%). Random forest classifier (as well as other classification models) generates the probability of a given sample belonging to each class for each sample (these probabilities sum to 1). By default, a sample is classified into the class for which the probability is higher. There may be a situation when we care about greater precision or greater recall. In such a case, the following should be assessed: the costs for a false positive sample (claiming that a healthy person is infected with the virus) as well as for a false negative sample (considering the infected person as healthy). And then, the probability level obtained from a learned model should be selected, by which we classify the samples in such a way that the overall cost of mistakes was as low as possible. Figure 5c demonstrates how recall and precision change depending on the threshold value for classifying a sample into class 1 (SARS-CoV-2 infected). We also

compared the results we obtained for random forest with other ML algorithms (see Table S2 in Supplementary Materials).



**Figure 5.** Confusion matrix scheme (a) and confusion matrix for RFC training process (b). Precision and recall as a function of the threshold for classifying a sample into class 1 (infected of the SARS-CoV-2) (c).

#### 4. Conclusions

In the present work, we used clinical samples of saliva and nasopharyngeal swabs from healthy and SARS-CoV-2-infected patients to demonstrate SERS techniques and machine learning algorithms as fast and efficient methods of detecting the SARS-CoV-2 virus. Several types of shallow machine learning were tested: gaussian naive Bayes, random forest, support vector, and logistic regression. Finally, we used a random forest classifier for the analysis, and the best results were obtained for RFC and the saliva samples (averaged precision of 90.4%, averaged recall of 81.6%, and averaged adjusted balanced accuracy of 87.1%). The data were compared with the SVMC method, which was recognized in our previous work as the best chemometric method. In the case of saliva samples, RF showed greater precision and comparable adjusted balanced accuracy compared to the SVMC method, while also showing a lower recall.

The results demonstrate that ML, even for sets of data considered as small, can classify samples with reasonable precision and accuracy. The precision for random forest was 94.1%, whereas recall (sensitivity) was 88.9%, which demonstrates the potential for the practical use of combined SERS and ML methods for the detection of the SARS-CoV-2 virus in clinical samples. Hence, the RF algorithm perfectly complements the SVMC method (analyzed in our previous work), especially in terms of precision, which can be raised up to 93.7%, making this approach more accurate for diagnostic purposes. For this reason, in the

future, we can create and test the multi-stage analysis of spectral data involving several methods that complement each other, i.e., SVMC and RF.

One of the most important challenges in diagnosis is detecting infected individuals at an early stage of infection. Such a task is essential for easily transmissible viruses such as SARS-CoV-2. Thus, one of the future directions for the use of SERS and machine learning would be to identify specific features of SERS spectra of saliva or nasopharyngeal swabs at an early stage of infection.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines12010167/s1>, Figure S1: Surface of the SERS platform (silicon micromachined via femtosecond laser and covered with 100 nm of silver) acquired by Scanning Electron Microscopy (SEM); Table S1: Tentative assignments for the main bands observed in the SERS spectra of SARS-CoV-2-infected CoV(+) and healthy CoV(−) subjects [51–58]; Table S2: The comparison of accuracy, sensitivity, and specificity for the analysis of body fluids in terms of COVID-19 diagnosis in a label-free manner [33,34,59–61].

**Author Contributions:** Conceptualization, T.R.S.; Data curation, S.M.B.; Formal analysis, A.B.N. and G.S.; Investigation, A.B.N. and A.K.; Methodology, S.M.B. and A.K.; Software, G.S.; Supervision, A.K.; Validation, T.R.S. and S.M.B.; Writing—original draft, T.R.S. and G.S.; Writing—review and editing, A.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Foundation for Polish Science under grant TEAM-TECH/2017-4/23 (POIR.04.04.00-00-4210/17-00).

**Institutional Review Board Statement:** All experiments were performed in compliance with the relevant laws and institutional guidelines. The protocol of study was approved by the Ethics and Bioethics Committee of Cardinal Stefan Wyszyński University (UKSW) in Warsaw, Poland; opinion number 13/2019.

**Informed Consent Statement:** Informed consent was obtained from all patients.

**Data Availability Statement:** Dataset available on request from the authors.

**Acknowledgments:** Authors thank Izabela Drózdź and Maciej Borowiec from Department of Clinical Genetics, Medical University of Łódź, Pomorska 251, 92-213 Łódź, Poland for saliva and nasopharyngeal swab samples.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
2. Oliveira, M.C.; Scharan, K.O.; Thomés, B.I.; Bernardelli, R.S.; Reese, F.B.; Kozesinski-Nakatani, A.C.; Martins, C.C.; Lobo, S.M.A.; Réa-Neto, Á. Diagnostic accuracy of a set of clinical and radiological criteria for screening of COVID-19 using RT-PCR as the reference standard. *BMC Pulm. Med.* **2023**, *23*, 81. [CrossRef] [PubMed]
3. Tahamtan, A.; Ardebili, A. Real-time RT-PCR in COVID-19 detection: Issues affecting the results. *Expert Rev. Mol. Diagn.* **2020**, *20*, 453–454. [CrossRef]
4. van Kasteren, P.B.; van der Veer, B.; van den Brink, S.; Wijsman, L.; de Jonge, J.; van den Brandt, A.; Molenkamp, R.; Reusken, C.B.E.M.; Meijer, A. Comparison of seven commercial RT-PCR diagnostic kits for COVID-19. *J. Clin. Virol.* **2020**, *128*, 104412. [CrossRef] [PubMed]
5. Xu, M.; Li, Y.; Lin, C.; Peng, Y.; Zhao, S.; Yang, X.; Yang, Y. Recent Advances of Representative Optical Biosensors for Rapid and Sensitive Diagnostics of SARS-CoV-2. *Biosensors* **2022**, *12*, 862. [CrossRef] [PubMed]
6. Hussein, H.A.; Hanora, A.; Solyman, S.M.; Hassan, Y.A. Designing and fabrication of electrochemical nano-biosensor for the fast detection of SARS-CoV-2-RNA. *Sci. Rep.* **2023**, *13*, 5139. [CrossRef]
7. Yakoh, A.; Pimpitak, U.; Rengpipat, S.; Hirankarn, N.; Chailapakul, O.; Chaiyo, S. Paper-based electrochemical biosensor for diagnosing COVID-19: Detection of SARS-CoV-2 antibodies and antigen. *Biosens. Bioelectron.* **2021**, *176*, 112912. [CrossRef]
8. Li, Z.; Luo, Y.; Song, Y.; Zhu, Q.; Xu, T.; Zhang, X. One-click investigation of shape influence of silver nanostructures on SERS performance for sensitive detection of COVID-19. *Anal. Chim. Acta* **2022**, *1234*, 340523. [CrossRef]
9. Alnaji, N.; Wasfi, A.; Awwad, F. The design of a point of care FET biosensor to detect and screen COVID-19. *Sci. Rep.* **2023**, *13*, 4485. [CrossRef]

10. Vafabakhsh, M.; Dadmehr, M.; Kazemi Noureini, S.; Es'haghi, Z.; Malekkiani, M.; Hosseini, M. Paper-based colorimetric detection of COVID-19 using aptasensor based on biomimetic peroxidase like activity of ChF/ZnO/CNT nano-hybrid. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2023**, *301*, 122980. [[CrossRef](#)]
11. Huang, P.C.; Zhou, Y.; Porter, E.B.; Saxena, R.G.; Gomez, A.; Ykema, M.; Senehi, N.L.; Lee, D.; Tseng, C.P.; Alvarez, P.J.; et al. Organic Electrochemical Transistors functionalized with Protein Minibinders for Sensitive and Specific Detection of SARS-CoV-2. *Adv. Mater. Interfaces* **2023**, *10*, 2202409. [[CrossRef](#)]
12. Kim, H.E.; Schuck, A.; Park, H.; Huh, H.J.; Kang, M.; Kim, Y.-S. Gold nanostructures modified carbon-based electrode enhanced with methylene blue for point-of-care COVID-19 tests using isothermal amplification. *Talanta* **2023**, *265*, 124841. [[CrossRef](#)]
13. GhaderiShekhiAbadi, P.; Irani, M.; Noorisepehr, M.; Maleki, A. Magnetic biosensors for identification of SARS-CoV-2, Influenza, HIV, and Ebola viruses: A review. *Nanotechnology* **2023**, *34*, 272001. [[CrossRef](#)]
14. Yang, Y.; Xu, B.; Murray, J.; Haverstick, J.; Chen, X.; Tripp, R.A.; Zhao, Y. Rapid and quantitative detection of respiratory viruses using surface-enhanced Raman spectroscopy and machine learning. *Biosens. Bioelectron.* **2022**, *217*, 114721. [[CrossRef](#)]
15. Driskell, J.D.; Kwarta, K.M.; Lipert, R.J.; Porter, M.D.; Neill, J.D.; Ridpath, J.F. Low-level detection of viral pathogens by a surface-enhanced Raman scattering based immunoassay. *Anal. Chem.* **2005**, *77*, 6147–6154. [[CrossRef](#)]
16. Luo, S.C.; Sivashanmugan, K.; Der Liao, J.; Yao, C.K.; Peng, H.C. Nanofabricated SERS-active substrates for single-molecule to virus detection in vitro: A review. *Biosens. Bioelectron.* **2014**, *61*, 232–240. [[CrossRef](#)] [[PubMed](#)]
17. Saviñon-Flores, F.; Méndez, E.; López-Castaños, M.; Carabarin-Lima, A.; López-Castaños, K.A.; González-Fuentes, M.A.; Méndez-Albores, A. A Review on SERS-Based Detection of Human Virus Infections: Influenza and Coronavirus. *Biosensors* **2021**, *11*, 66. [[CrossRef](#)]
18. Lin, C.; Li, Y.; Peng, Y.; Zhao, S.; Xu, M.; Zhang, L.; Huang, Z.; Shi, J.; Yang, Y. REVIEW Open Access Recent development of surface-enhanced Raman scattering for biosensing. *J. Nanobiotechnol.* **2023**, *21*, 149. [[CrossRef](#)] [[PubMed](#)]
19. Wang, L.; Tang, J.-W.; Li, F.; Usman, M.; Wu, C.-Y.; Liu, Q.-H.; Kang, H.-Q.; Liu, W.; Gu, B. Identification of Bacterial Pathogens at Genus and Species Levels through Combination of Raman Spectrometry and Deep-Learning Algorithms. *Microbiol. Spectr.* **2022**, *10*, e02580-22. [[CrossRef](#)]
20. Tang, J.W.; Lyu, J.W.; Lai, J.X.; Zhang, X.D.; Du, Y.G.; Zhang, X.Q.; Zhang, Y.D.; Gu, B.; Zhang, X.; Gu, B.; et al. Determination of Shigella spp. via label-free SERS spectra coupled with deep learning. *Microchem. J.* **2023**, *189*, 108539. [[CrossRef](#)]
21. Zhao, Y.; Zhang, Z.; Ning, Y.; Miao, P.; Li, Z.; Wang, H. Simultaneous quantitative analysis of Escherichia coli, Staphylococcus aureus and Salmonella typhimurium using surface-enhanced Raman spectroscopy coupled with partial least squares regression and artificial neural networks. *Spectrochim. Acta—Part A Mol. Biomol. Spectrosc.* **2023**, *293*, 122510. [[CrossRef](#)] [[PubMed](#)]
22. Li, J.Q.; Dukes, P.V.; Lee, W.; Sarkis, M.; Vo-Dinh, T. Machine learning using convolutional neural networks for SERS analysis of biomarkers in medical diagnostics. *J. Raman Spectrosc.* **2022**, *53*, 2044–2057. [[CrossRef](#)] [[PubMed](#)]
23. Han, S.; Chen, C.; Chen, C.; Wu, L.; Wu, X.; Lu, C.; Zhang, X.; Chao, P.; Lv, X.; Jia, Z.; et al. Coupling annealed silver nanoparticles with a porous silicon Bragg mirror SERS substrate and machine learning for rapid non-invasive disease diagnosis. *Anal. Chim. Acta* **2023**, *1254*, 341116. [[CrossRef](#)] [[PubMed](#)]
24. Tabarov, A.; Vitkin, V.; Andreeva, O.; Shemanaeva, A.; Popov, E.; Dobrosлавин, A.; Kurikova, V.; Kuznetsova, O.; Grigorenko, K.; Tzibizov, I.; et al. Detection of A and B Influenza Viruses by Surface-Enhanced Raman Scattering Spectroscopy and Machine Learning. *Biosensors* **2022**, *12*, 1065. [[CrossRef](#)]
25. dos Santos, D.P.; Sena, M.M.; Almeida, M.R.; Mazali, I.O.; Olivieri, A.C.; Villa, J.E.L. Unraveling surface-enhanced Raman spectroscopy results through chemometrics and machine learning: Principles, progress, and trends. *Anal. Bioanal. Chem.* **2023**, *415*, 3945–3966. [[CrossRef](#)]
26. Ding, Y.; Sun, Y.; Liu, C.; Jiang, Q.Y.; Chen, F.; Cao, Y. SERS-Based Biosensors Combined with Machine Learning for Medical Application\*\*. *ChemistryOpen* **2023**, *12*, e202200192. [[CrossRef](#)]
27. Chen, H.; Park, S.G.; Choi, N.; Kwon, H.J.; Kang, T.; Lee, M.K.; Choo, J. Sensitive Detection of SARS-CoV-2 Using a SERS-Based Aptasensor. *ACS Sensors* **2021**, *6*, 2378–2385. [[CrossRef](#)]
28. Peng, Y.; Lin, C.; Long, L.; Masaki, T.; Tang, M.; Yang, L.; Liu, J.; Huang, Z.; Li, Z.; Luo, X.; et al. Charge-Transfer Resonance and Electromagnetic Enhancement Synergistically Enabling MXenes with Excellent SERS Sensitivity for SARS-CoV-2 S Protein Detection. *Nano-Micro Lett.* **2021**, *13*, 52. [[CrossRef](#)]
29. Yang, Y.; Peng, Y.; Lin, C.; Long, L.; Hu, J.; He, J.; Zeng, H.; Huang, Z.; Li, Z.Y.; Tanemura, M.; et al. Human ACE2-Functionalized Gold “Virus-Trap” Nanostructures for Accurate Capture of SARS-CoV-2 and Single-Virus SERS Detection. *Nano-Micro Lett.* **2021**, *13*, 109. [[CrossRef](#)]
30. Liu, H.; Dai, E.; Xiao, R.; Zhou, Z.; Zhang, M.; Bai, Z.; Shao, Y.; Qi, K.; Tu, J.; Wang, C.; et al. Development of a SERS-based lateral flow immunoassay for rapid and ultra-sensitive detection of anti-SARS-CoV-2 IgM/IgG in clinical samples. *Sens. Actuators B Chem.* **2021**, *329*, 129196. [[CrossRef](#)] [[PubMed](#)]
31. Mei, X.; Lee, H.C.; Diao, K.Y.; Huang, M.; Lin, B.; Liu, C.; Xie, Z.; Ma, Y.; Robson, P.M.; Chung, M.; et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat. Med.* **2020**, *26*, 1224–1228. [[CrossRef](#)] [[PubMed](#)]
32. Ikponmwoba, E.; Ukorigho, O.; Moitra, P.; Pan, D.; Gartia, M.R.; Owoyele, O. A Machine Learning Framework for Detecting COVID-19 Infection Using Surface-Enhanced Raman Scattering. *Biosensors* **2022**, *12*, 589. [[CrossRef](#)] [[PubMed](#)]



33. Karunakaran, V.; Joseph, M.M.; Yadev, I.; Sharma, H.; Shamna, K.; Saurav, S.; Sreejith, R.P.; Anand, V.; Beegum, R.; Regi David, S.; et al. A non-invasive ultrasensitive diagnostic approach for COVID-19 infection using salivary label-free SERS fingerprinting and artificial intelligence. *J. Photochem. Photobiol. B Biol.* **2022**, *234*, 112545. [[CrossRef](#)]
34. Yang, Y.; Li, H.; Jones, L.; Murray, J.; Haverstick, J.; Naikare, H.K.; Mosley, Y.Y.C.; Tripp, R.A.; Ai, B.; Zhao, Y. Rapid Detection of SARS-CoV-2 RNA in Human Nasopharyngeal Specimens Using Surface-Enhanced Raman Spectroscopy and Deep Learning Algorithms. *ACS Sensors* **2023**, *8*, 297–307. [[CrossRef](#)] [[PubMed](#)]
35. Hwang, C.S.H.; Lee, S.; Lee, S.; Kim, H.; Kang, T.; Lee, D.; Jeong, K.H. Highly Adsorptive Au-TiO<sub>2</sub>Nanocomposites for the SERS Face Mask Allow the Machine-Learning-Based Quantitative Assay of SARS-CoV-2 in Artificial Breath Aerosols. *ACS Appl. Mater. Interfaces* **2022**, *14*, 54550–54557. [[CrossRef](#)] [[PubMed](#)]
36. Ansah, I.B.; Leming, M.; Lee, S.H.; Yang, J.-Y.; Mun, C.; Noh, K.; An, T.; Lee, S.; Kim, D.-H.; Kim, M.; et al. Label-free detection and discrimination of respiratory pathogens based on electrochemical synthesis of biomaterials-mediated plasmonic composites and machine learning analysis. *Biosens. Bioelectron.* **2023**, *227*, 115178. [[CrossRef](#)] [[PubMed](#)]
37. Berus, S.M.; Nowicka, A.B.; Wieruszewska, J.; Niciński, K.; Kowalska, A.A.; Szymborski, T.R.; Drózd, I.; Borowiec, M.; Waluk, J.; Kamińska, A. SERS Signature of SARS-CoV-2 in Saliva and Nasopharyngeal Swabs: Towards Perspective COVID-19 Point-of-Care Diagnostics. *Int. J. Mol. Sci.* **2023**, *24*, 9706. [[CrossRef](#)]
38. Szymborski, T.; Stepanenko, Y.; Niciński, K.; Piecyk, P.; Berus, S.M.; Adamczyk-Popławska, M.; Kamińska, A. Ultrasensitive SERS platform made via femtosecond laser micromachining for biomedical applications. *J. Mater. Res. Technol.* **2021**, *12*, 1496–1507. [[CrossRef](#)]
39. Ralbovsky, N.M.; Lednev, I.K. Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning. *Chem. Soc. Rev.* **2020**, *49*, 7428–7453. [[CrossRef](#)] [[PubMed](#)]
40. Ng, A.Y.; Jordan, M.I. On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. *Adv. Neural Inf. Process. Syst.* **2001**, *14*, 1–8.
41. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
42. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
43. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009. [[CrossRef](#)]
44. Isho, B.; Abe, K.T.; Zuo, M.; Jamal, A.J.; Rathod, B.; Wang, J.H.; Li, Z.; Chao, G.; Rojas, O.L.; Bang, Y.M.; et al. Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients. *Sci. Immunol.* **2020**, *5*, eabe5511. [[CrossRef](#)] [[PubMed](#)]
45. Baghizadeh Fini, M. Oral saliva and COVID-19. *Oral Oncol.* **2020**, *108*, 104821. [[CrossRef](#)] [[PubMed](#)]
46. Torreggiani, A.; Barata-Vallejo, S.; Chatgililoglu, C. Combined Raman and IR spectroscopic study on the radical-based modifications of methionine. *Anal. Bioanal. Chem.* **2011**, *401*, 1231–1239. [[CrossRef](#)]
47. Hoffman, R.M.; Han, Q. Oral methioninase for Covid-19 methionine-restriction therapy. *In Vivo* **2020**, *34*, 1593–1596. [[CrossRef](#)] [[PubMed](#)]
48. Franco-Martínez, L.; Cerón, J.J.; Vicente-Romero, M.R.; Bernal, E.; Cantero, A.T.; Tecles, F.; Resalt, C.S.; Martínez, M.; Tvarijonavičiute, A.; Martínez-Subiela, S. Salivary Ferritin Changes in Patients with COVID-19. *Int. J. Environ. Res. Public Health* **2021**, *19*, 41. [[CrossRef](#)] [[PubMed](#)]
49. Hailemichael, W.; Kiros, M.; Akelew, Y.; Getu, S.; Andualem, H. *Neopterin: A Promising Candidate Biomarker for Severe COVID-19*; Dove Press: Macclesfield, UK, 2021; Volume 14, p. 245. [[CrossRef](#)]
50. Kamińska, A.; Witkowska, E.; Kowalska, A.; Skoczyńska, A.; Gawryszewska, I.; Guziewicz, E.; Snigurenko, D.; Waluk, J. Highly efficient SERS-based detection of cerebrospinal fluid neopterin as a diagnostic marker of bacterial infection. *Anal. Bioanal. Chem.* **2016**, *408*, 4319–4327. [[CrossRef](#)]
51. Lin, X.; Lin, D.; Ge, X.; Qiu, S.; Feng, S.; Chen, R. Noninvasive Detection of Nasopharyngeal Carcinoma Based on Saliva Proteins Using Surface-Enhanced Raman Spectroscopy. *J. Biomed. Opt.* **2017**, *22*, 105004. [[CrossRef](#)]
52. Li, X.; Yang, T.; Lin, J. Spectral Analysis of Human Saliva for Detection of Lung Cancer Using Surface-Enhanced Raman Spectroscopy. *J. Biomed. Opt.* **2012**, *17*, 037003. [[CrossRef](#)]
53. Talari, A.C.S.; Movasaghi, Z.; Rehman, S.; Rehman, I.U. Raman Spectroscopy of Biological Tissues. *Appl. Spectrosc. Rev.* **2015**, *50*, 46–111. [[CrossRef](#)]
54. Austin, L.A.; Osseiran, S.; Evans, C.L. Raman Technologies in Cancer Diagnostics. *Analyst* **2016**, *141*, 476–503. [[CrossRef](#)]
55. Cao, G.; Chen, M.; Chen, Y.; Huang, Z.; Lin, J.; Lin, J.; Xu, Z.; Wu, S.; Huang, W.; Weng, G.; et al. A Potential Method for Non-Invasive Acute Myocardial Infarction Detection Based on Saliva Raman Spectroscopy and Multivariate Analysis. *Laser Phys. Lett.* **2015**, *12*, 125702. [[CrossRef](#)]
56. Oliveira, E.M.; Rogero, M.; Ferreira, E.C.; Gomes Neto, J.A. Simultaneous Determination of Phosphite and Phosphate in Fertilizers by Raman Spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *246*, 119025. [[CrossRef](#)] [[PubMed](#)]
57. Hu, P.; Zheng, X.S.; Zong, C.; Li, M.H.; Zhang, L.Y.; Li, W.; Ren, B. Drop-Coating Deposition and Surface-Enhanced Raman Spectroscopies (DCDRS and SERS) Provide Complementary Information of Whole Human Tears. *J. Raman Spectrosc.* **2014**, *45*, 565–573. [[CrossRef](#)]
58. Virkler, K.; Lednev, I.K. Forensic Body Fluid Identification: The Raman Spectroscopic Signature of Saliva. *Analyst* **2010**, *135*, 512–517. [[CrossRef](#)] [[PubMed](#)]

59. Carlomagno, C.; Bertazioli, D.; Gualerzi, A.; Picciolini, S.; Banfi, P.I.; Lax, A.; Messina, E.; Navarro, J.; Bianchi, L.; Caronni, A.; et al. COVID-19 salivary Raman fingerprint: Innovative approach for the detection of current and past SARS-CoV-2 infections. *Sci. Rep.* **2021**, *11*, 4943. [[CrossRef](#)]
60. Ceccon, D.M.; Amaral, P.H.R.; Andrade, L.M.; da Silva, M.I.; Andrade, L.A.; Moraes, T.F.; Bagno, F.F.; Rocha, R.P.; de Almeida Marques, D.P.; Ferreira, G.M.; et al. New, fast, and precise method of COVID-19 detection in nasopharyngeal and tracheal aspirate samples combining optical spectroscopy and machine learning. *Braz. J. Microbiol.* **2023**, *54*, 769–777. [[CrossRef](#)]
61. Goulart, A.C.C.; Zângaro, R.A.; Carvalho, H.C.; Lednev, I.K.; Silveira, L., Jr. Diagnosing COVID-19 in nasopharyngeal secretion through Raman spectroscopy: A feasibility study. *Lasers Med. Sci.* **2023**, *38*, 210. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.