Babeș-Bolyai University

Faculty of Physics

Doctoral School of Physics

Machine learning analysis of biofluid SERS data for clinical applications

Summary of the PhD Thesis

PhD student:

Ștefania-Dana Iancu

Scientific advisor:

Prof. dr. Nicolae Leopold

Cluj-Napoca

2023

Content

Keywords

Surface-enhanced Raman scattering (SERS); machine learning; cancer screening; support tool; liquid biopsy.

Introduction

Clinical spectroscopy is an emerging research field that aims to translate the use of optical spectroscopy into the clinical setting for the benefit of patients. This includes improving patient diagnosis and assessing prognosis. There is a need for new diagnostic tools in the clinical setting as alternatives to current techniques that are time-consuming and costly. This drives researchers to push the boundaries of this field, especially for diseases that are difficult or impossible to cure, such as cancer.

Raman spectroscopy, particularly the technique known as Surface-enhanced Raman spectroscopy, has become increasingly popular in the medical field because of its ability to detect small amounts of substances and its flexibility. However, its implementation in clinical settings has been hindered by a a lack of clear guidelines, regulatory challenges, and the fact that studies to test its effectiveness have typically been done with small numbers of samples.

The current PhD project's goal is not to discover new SERS bands in biofluid spectra or to create new methods for collecting SERS spectra in clinical settings, but rather to improve the way spectra are analyzed. It's important to note that the majority of data analyzed in this thesis come from SERS spectra collected in a clinical setting using a portable Raman spectrometer.

Chapter 1

The first chapter presents the theoretical background of SERS and a short overview of machine learning approaches for spectra analysis. Surface Enhanced Raman Scattering (SERS) is a powerful technique that is used to amplify the Raman scattering signal from molecules adsorbed on a surface. The signal enhancement is usually orders of magnitude larger than that obtained in conventional Raman scattering, making it possible to detect even very weak signals. The enhancement of the Raman scattering signal is caused either by the enhancement of the electric field at the surface of a metal, typically gold or silver or by the charge transfer between the metal and the molecule that is adsorbed onto the metallic structure. The electromagnetic theory explains the SERS enhancement as a result of the plasmons oscillations creating an enhanced electromagnetic field surrounding the metallic structure which interacts with the molecule found in the proximity of the metal. Plasmons are the eigenmodes of collective oscillations of the quasi-free electrons in metals.

The diameter of the nanoparticles is lower than the penetration depth of electromagnetic waves in metals. Thus, the electromagnetic wave oscillates 'through' metal and move the quasi-free electrons to a pole of the nanoparticle (Figure 1).

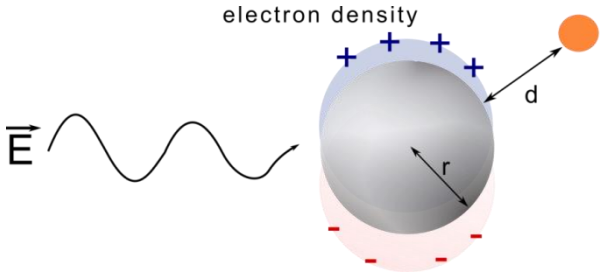


Figure 1. Dipole excitation in a nanometric metal sphere.

The lattice ions from the other pole of the nanoparticle, which have a partly positive charge action as a restoring force for the quasi-free electrons. Therefore, the electrons will oscillate between the two poles, as a result of the incident electromagnetic field and the restoring force from the lattice ions [1]. The resonance condition is met at the frequency where the transition between metal transparency and absorption of the electromagnetic wave happens. The plasmon oscillation at resonant condition creates an electromagnetic field surrounding the nanoparticle (E_{SPR}), which will cumulate with the incident electromagnetic field, leading to an enhancement of the electormagnetic field. The field intensity decreases with the distance to the surface of the metal and is dependent of the diameter of the nanoparticle (2r), the distance from nanoparticle (d) and the dielectric constant of the nanoparticle (ε) and the environment (ε_0).

$$E_{SPR} = r^3 \frac{\varepsilon-\varepsilon_0}{\varepsilon+2\varepsilon_0} E_0 \frac{1}{(r+d)^3} \qquad (1)$$

Hence, the molecules must be in the proximity of the nanoparticles in order to ,feel' the electromagnetic enhancement.

The chemical enhancement presumes that a charge transfer between the metal and the molecule occurs (Figure 2)[2]. The chemical mechanism involves two types of interactions between the molecule and the substrate: physisorption and chemisorption [3]. Physisorption involves a weak van der Waals interaction between the molecule and the metal substrate without changing the energy levels of the molecule. Chemisorption consists in the formation of a molecule–metal assembly through the hybridization of the metal and molecular orbitals of the adsorbed molecule. Bonding and nonbonding states are formed. The SERS effect consists of resonant electronic transfer between the highest occupied molecular orbitals (HOMO) and the lowest unoccupied molecular orbitals (LUMO) of the metal-adsorbed complex [4]. The chemical effect explains the preferential enhancement of antisymmetric modes of molecules adsorbed on the metal surface.

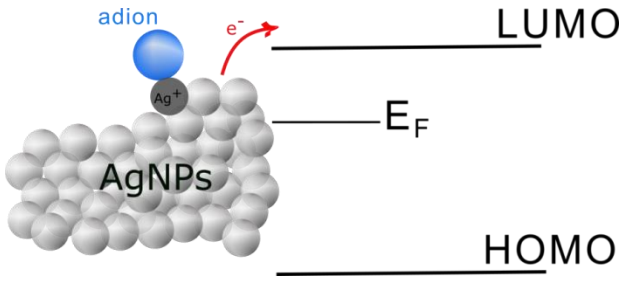


Figure 2. Charge transfer between silver nanoparticles and a molecule adsorbate onto the silver surface.

The adsorption of molecules onto metallic substrates can happen spontaneously or be promoted by specific ions. The spontaneous adsorption is specific for molecules with high affinity for the metal surface, such as molecule containing thiol groups that show a high affinity for the surface of noble metallic structures. However, not all molecules have a strong attraction to metal surfaces, and therefore will not adsorb under any conditions. In recent years, it has been discovered that certain ions can promote the adsorption of such molecules onto silver nanoparticle surfaces and lead to a significant increase in the Raman scattering signal [5]. For example, the SERS spectra of anionic molecule such as salicylic acid, fumaric acid is observed after the addition of Ca^{2+}, Mg^{2+} to a mixture of organic acids and silver nanoparticles [6]. However, the mechanism behind the influence of Ca^{2+} or Mg^{2+} to the SERS spectra of anionic molecules is not fully understood. In turn, Cl^-, Br^- or I^- promote the adsorption of cationic molecules such as Nile Blue, Methylene Blue, Crystal Violet etc. An analysis of the reduction of Methylene Blue in the presence of silver nanoparticles and halide ions showed that the variation of the Fermi level of silver nanoparticles in the presence of halide ions promotes charge transfer between the nanoparticles and the molecule adsorbed onto the metallic surface. [7]

The electromagnetic theory is unable to account for the absence of SERS enhancement in substances such as water, nanoparticles, and surfactants, while the chemical effect cannot account for the magnitude of the SERS enhancement factor. As a result, it is believed that the SERS effect is a combination of these two effects. The chemical theory describes the requirement for molecules to adsorb onto metallic structures, while the electromagnetic theory provides a quantitative measure of the SERS effect.

Moreover, the first chapter introduces the hallmarks of cancer detected by SERS: purine metabolites, proteins and carotenoids. The promising results of SERS liquid biopsy stem from the fact that the levels of purine metabolites, proteins and carotenoids are informative of the cellular turnover rate, inflammation, and oxidative stress, respectively. These processes are disrupted in virtually any disease, from cancer to autoimmune maladies. Purine metabolites are a class of

molecules resulting from the catabolysis of purine nitrogen bases (adenine and guanine), starting from inosine, hypoxanthine, xanthine and the end product uric acid (Figure 3). In contrast, the pyrimidine nitrogen bases (thymine, cytosine and uracil) are broken down to water and carbon dioxide [8].

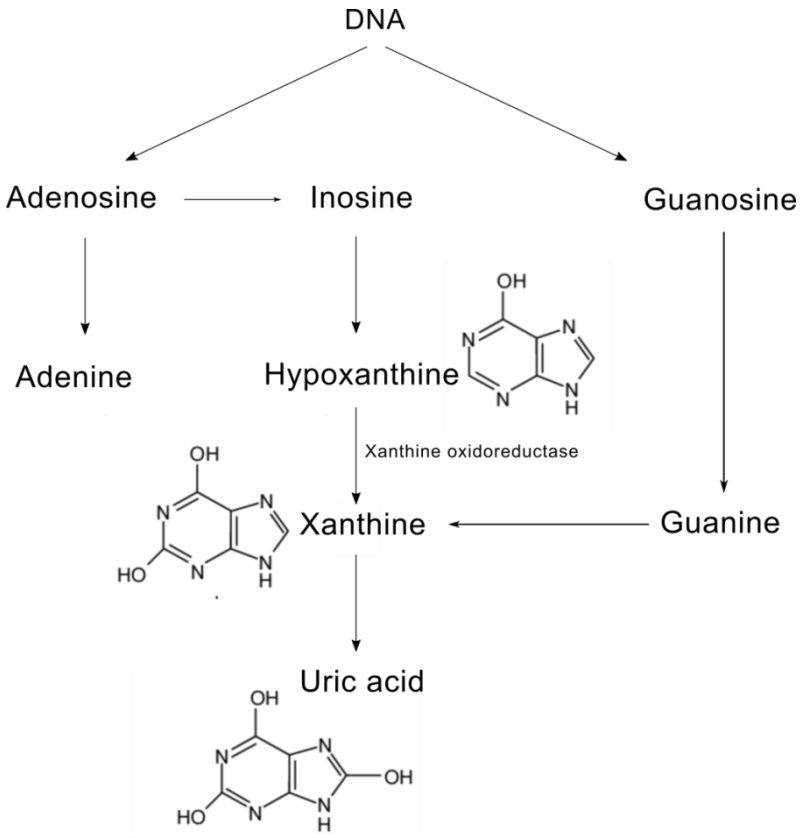


Figure 3. A simple representation of the purine metabolism pathway.

I also present some fundamental concepts regarding the machine learning (ML) algorithms used for analyzing the SERS spectra of biofluids. The central idea of ML is that an algorithm which gained experience for a specific task (by learning from a dataset named training set) should perform well (or at least similar compared to training set) on a new, previously unseen input called test set. The task is the classification problem which will be accomplished with a certain performance measured by performance metrics. To resolve a task, the ML classifiers learn from the training group by extracting the relevant features, thresholds or densities. The experience is applied onto a new dataset to be evaluated. The performance of the ML models can be tuned by changing the error thresholds, or the function involved in the discrimination, by changing the hyperparameters of the models. The hyperparameters are settings that can be tuned to control the ML models behaviour. The choice of the ML and of the hyperparameters is a trial and error process. The performance metrics of ML with different hyperparameters are determined on a validation set, different from the training set and test set. Why do we need three different data sets?

I mentioned that ML is defined by three important steps: experience gaining, which implies to learn a specific pattern from a training set; changing the hyperparameters we focus on a specific task; the accomplishment of the task is analysed in the validation group; and finally to analyse the performance of the trained ML inclined to a specific task (optimized for the problem raised) we analyse the accuracy of the model on a test group. The training and test sets should be independent of each other, but with identically distributions [9]. The big differences in the training and test errors lead to under fitting when the training error is high or overfitting when the test error is significantly lower than the training error. By changing the hyperparametres or choosing the optimal model we can reduce the risk of overfitting or under fitting.

The differences between the ML algorithms employed in this thesis are represented by the mathematical criteria utilized to categorize the data set into two groups. Prior to classification algorithms, data dimensionality is reduced using for example, Principal Component Analysis (PCA), an unsupervised algorithm which finds the vectors that stand for the highest variance within data set. Next, supervised methods that also consider the category of each sample extract the essential features to differentiate the classes/groups. Linear Discriminant Analysis (LDA) finds the vectors that minimize the variance within each group and maximize the distance between the groups, and represents the data on these vectors. In the new representation space, a line that best separates the groups is sought. Moving from a line that best separates the groups (for example LDA) to a more complex boundary such as a plan, we will talk about Support Vector Machine (SVM). SVM relies on the transformation of the features into points in an *n*-dimensional space and finding a hyperplane that would separate the classes. When the classes are not linearly separable in the initial space, data are transformed into a new space with higher dimensionality using a kernel function (linear, polynomial or more complex transformation). For any data set there is a kernel function which allows the data to be linearly separable [10]. However, adding new dimensions to the data set increases the risk of overfitting.

In the new space a line is searched, that separates the classes. The model looks for a boundary line, for which the distance between it and the nearest point in each class is maximal. Thus, SVM focuses on the samples at the margin of each class, named support vectors.

Logistic Regression (LR) is the most used statistical techniques for binary medical outcomes [11]. LR fits a Sigmoid function (Equation 2) to the data, actually to the weighted sum of the input features (a linear combination of the input features). The Sigmoid function (a S sharped curve) takes values between 0 and 1. For each sample we get a probability (ranging from 0 to 1) for a sample to correspond in class 0 or class 1. The prediction class is made based on the odds, a probability less than 0.5 will target the samples as being class 0, while a probability greater than 0.5 will predict class

1. One of the most important aspects in LR is that you can see the importance of each feature in the classification of the outcome.

$$sigmoid\ (x) = \frac{1}{1+e^{-x}} \qquad (2)$$

The training of LR model implies getting the right weights/coefficients for each feature. Multiple interactions are implied to test the coefficients until the best fit of odds is obtained. Once the optimal coefficients are found, the conditional probabilities for each observation can be calculated, logged, and summed together to yield a predicted probability. The relationship between the variables and the outcome can be linearly related via logarithmic odds, which is a bit more flexible than a linear relationship.

Random Forest (RF) contains multiple uncorrelated decision trees (DTs), which are simple classifiers, selecting a set of if-then-else conditions to categorize. Each tree analyse different randomly selected sub-sets from the data set and extracts the features that maximize the discrimination of the groups. To be more explicit, a random vector is sampled independently on each tree. The final decision is a vote based on the output of each decision tree. Bagging is the ensemble technique used by Random Forest that chooses a random sample from the data set. Hence, each model is generated from the samples provided by the original data with replacement, known as row sampling. This row sampling step with replacement is called bootstrap with each model being trained independently. The hyperparameters that can be tuned in RF are the number of trees, meaning that the number of different decision paths can be changed. With increasing the number of trees, the performance of the model on the training group is increased but with the cost of overfitting.

The main advantage of RF is that is it robust to outliers and noisy data [12] and the samples need no prior pre-processing. Moreover, the risk of overfitting is reduced [13]. The use of ML models in the Raman spectroscopy was developed from the early 90s, when ANNs were employed to determine the composition of materials based on their vibrational spectra. The authors highlighted that the hyperparameters and spectra pre-processing play an important role in the ML result [14]. ML algorithm such as RF is not influence by the pre-processing methods, reducing the direct implications of the people [15]. The minimal user intervention is favourable for the translation of SERS to clinical application. However, RF need a large dataset to be trained [11] and may not always be adapted to spectral data due to 'the topology of its constituent classification trees'[16].

Gaussian Naïve Bayes (GNB) algorithms are based on the Bayes theorem, assuming strong independence between features, and provide a simple, yet highly functional, joint condition of

probabilistic classification method. Bayes theorem calculates the probability of something to happen, knowing that something else has already occurred.

GNB calculates the probability of an event (feature) to happen for each hypothesis (belong to a target/class). A product of all the probabilities is calculated and the hypothesis is accepted or rejected based on the final probability value. The target/class with the highest probability will be the predicted output. The assumption that GNB makes is that the features are uncorrelated and the data are normally distributed. GNB have no hyperparameters to be tuned and are simple and rapid classifiers. kNN is a nonparametric learning model, that assumes that the similar objects are in close proximity in the initial space [17]. The distances between k nearest neighbours are used to determine which class is the closest to the unknown sample. The distance between two samples is not necessarily to be the physical distance, a Euclidian distance. Different equations for the calculation of the distance between two neighbours can be employed to fit the task. Together with the equation, the number of neighbours which are evaluated for each sample can be tuned. Since all the samples are mapped when the distances between the neighbours are calculated, outliers influence the kNN model.

Chapter 2

The common analysis methods of the studies included in this thesis are discussed in chapter 2.

Chapter 3

The chapter 3 highlights the results of this thesis, being divided in four main subchapters that focuses onto the choice of the optimal biofluid for SERS analysis, the applications of SERS urine biopsy in clinical setting and the improvement of the machine learning approaches for spectral analysis.

The subchapter 3.1 focuses on the assignment of SERS signals in serum and urine samples and a comparison of the two biofluids in terms of SERS spectral characteristics and performance for cancer screening. SERS spectroscopy is commonly used to analyze biofluids, with urine being a particularly useful sample. When comparing the performance of serum and urine in breast cancer screening, I found that urine SERS biopsy had a slightly higher accuracy in distinguishing between patients and controls. Both urine and serum have similar SERS spectra of purine metabolites, and these similarities suggest that using urine samples analyzed by SERS as a tool for cancer screening or diagnosis could be effective.

Urine and serum SERS spectra show striking similarities, the major contribution in the both spectra being attributed to purine metabolites (Figure 4).

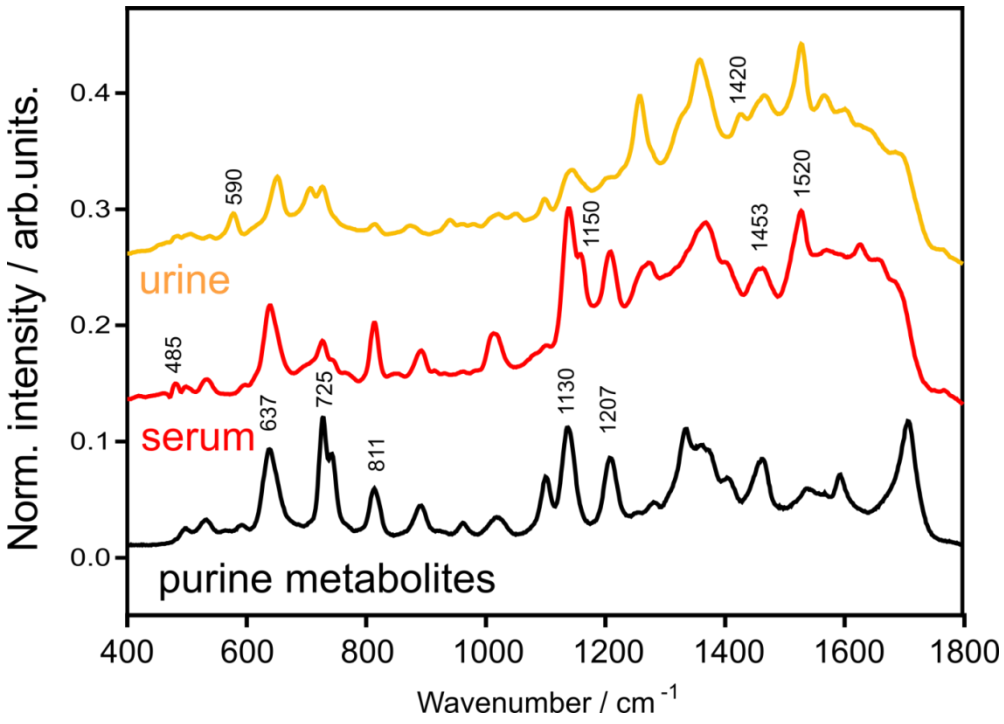


Figure 4. Similarities between SERS spectra of urine, serum and a mixture of uric acid, hypoxanthine and xanthine.[18]

For the assignment of the SERS bands, SERS spectra of pure solution of uric acid, hypoxanthine, xanthine, creatinine and urea were acquired (Figure 5).

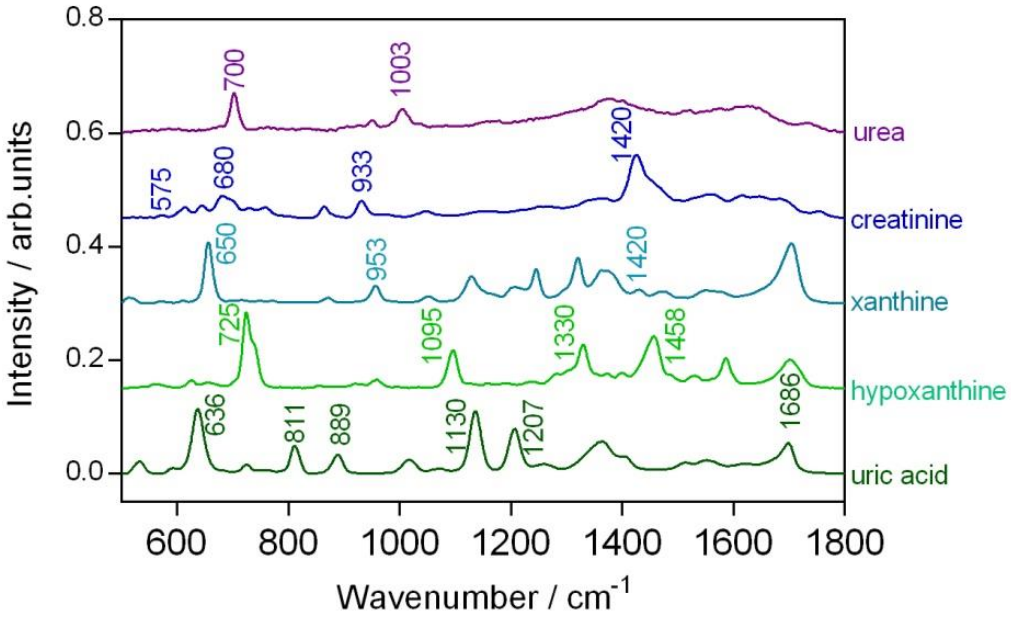


Figure 5. The SERS spectra of pure solution of 10^{-4} M uric acid, 10^{-4} M hypoxanthine, 10^{-4} M xanthine, 10^{-4} M creatinine and 10^{-2} M urea.

The discussion is based on the normal concentrations of the metabolites in the serum and urine and the SERS band assignment presented in Table 1.

Table 1. Normal metabolite concentrations and tentative assignment of the SERS bands of serum and urine samples.[18]

Metabolite	**Serum level (μM) [19]**	**Urine level (μmol/mmol creatinine) [19]**	**SERS bands assignment (cm^{-1})**		
			Serum	**Urine**	**Serum and urine**
Ergothioneine	0.84 – 1.52		482 [20]		
Uric acid	200 – 420	100 – 800	592, 637, 1260, 1368, 1686		499, 533, 725 811, 886, 1136, 1206
Urea	4000– 8000	6000– 30000		700	
Creatinine	25 – 100			575, 1420, 1459 [21]	
Hypoxanthine	8 – 14	1.8 – 24.1	740, 1450, 1686	958	725, 1093
Xanthine	0.5 – 2	0 – 30	1368, 1686	1250	
Carotenoids			1150, 1520 [22]		

Among purine metabolites, uric acid has the highest concentration in the biofluids (Table 1. Normal metabolite concentrations and tentative assignment of the SERS bands of serum and urine samples.Table 1) and therefore uric acid is responsible for the SERS spectral backbone of the biofluids (SERS band at 637, 811, 889, 1130 and 1207 cm^{-1}). However, its contribution is more pronounced in the serum SERS spectra compared to urine SERS spectra. Although the normal concentration of uric acid in serum is lower than in urine, the SERS intensities of uric acid bands in serum are higher than in urine. We explain this negative correlation by the pH dependency of the SERS signal of uric acid (pKa value 5.4). For this purpose, the SERS spectra of a mixture of three purine metabolites (uric acid, hypoxanthine, xanthine at 10^{-4} M concentrations) were measured at pH 5 and 7, pH values that are close to the pH of urine and serum (Figure 6). At acidic pH the SERS bands characteristic to uric acid are less intense compared to neutral pH, most probably because protonate uric acid shows a lower affinity for the metal surface.

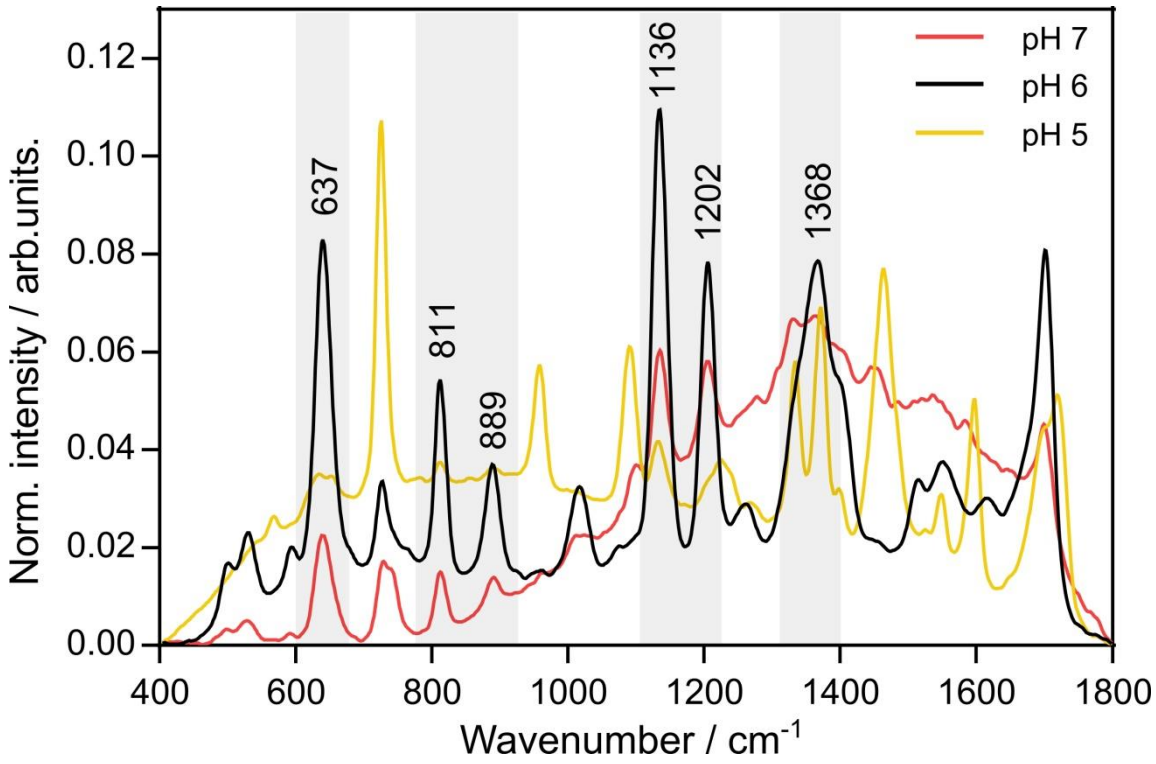


Figure 6. SERS spectra of a mixture of 0.1 mM uric acid, 0.1 mM xanthine, 0.1 mM hypoxanthine at pH 5, pH 6 and pH 7. [23]

The doublet at 724 and 740 cm^{-1} is usually attributed to hypoxanthine (Figure 5), another purine metabolite, although the 724 SERS band is also present in the SERS spectrum of uric acid (albeit contributing weakly) [24]. The SERS bands at 1368 and 1686 cm^{-1} could not be assigned to only one metabolite since these are observed in the SERS spectra of uric acid, xanthine and hypoxanthine as well.

Besides similarities assigned to uric acid and hypoxanthine SERS band, there are three main differences between serum and urine SERS spectra. SERS spectra of serum display usually a Raman shoulder around 1150 cm^{-1} and a specific band at 1520 cm^{-1}, attributed to carotenoids, but only when the spectra are acquired under resonant conditions [25]. In our case, the SERS spectra were acquired with the 532 nm excitation laser, which fulfils the Raman pre-resonance conditions for carotenoids. As in the case of purine metabolites, carotenoids have been suggested as possible cancer markers due to their antioxidant character, their level in serum reflecting the antioxidant capacity of the organism. However, carotenoids are not specific to one type of disorder, and can be easily influenced by the nutrition. Thus, its detection in serum does not play an important role in the cancer screening based on SERS liquid biopsy.

Moreover, the SERS spectra of serum also contained a band at 482 cm^{-1}, characteristic of ergothioneine [20], a dietary amino acid. Ergothioneine could only be detected in serum because of the low excretion in urine [26].

In addition, SERS spectra of urine are also dominated by bands attributed to creatinine at 575, 1421 and 1453 cm^{-1} [21]. The SERS bands of creatinine are even more pronounced when the urine sample is dried. Knowing the creatinine concentration of urine is highly valuable in quantifying protein concentration in urine samples (e.g. patients with renal dysfunction). Since the SERS-based quantification of proteins in urine samples was proved before, for concentrations as low as 3 µg/mL [27], theoretically it might be possible to also compute the creatinine to protein ratio only from SERS spectra, this ratio is more valuable for clinicians since creatinine is used to normalize the proteinuria.

Since the spectral information is similar in urine and serum, the choice of the biofluid is made based on the accessibility. Moreover, urine, which is collected through a non-invasive method, shows slightly higher classification accuracy for breast cancer screening based on SERS spectroscopy and ML models.

Subchapters 3.2 and 3.3 show the performance of urine SERS analysis combined with ML classifiers in screening and support tool for cancer by presenting two application of SERS urine biopsy in bladder and prostate cancer diagnosis. By tuning the parameter of the ML classifiers, SERS urine biopsy can be used either as an emerging test for cancer screening or support tool diagnosis.

Within these subchapters, I highlight the potential of urine SERS biopsy as screening method for bladder cancer, reaching classification accuracies ranging between 78-87% depending of the classifier involved. SERS urine biopsy has previously been shown to be an effective screening tool for a variety of cancers. However, my research has also shown that by incorporating complementary information such as microRNA, the accuracy of classifying bladder cancer and controls can be improved. By combining SERS with a complementary technique such as microRNA the classification accuracy is increased to 86-97% (Figure 7).

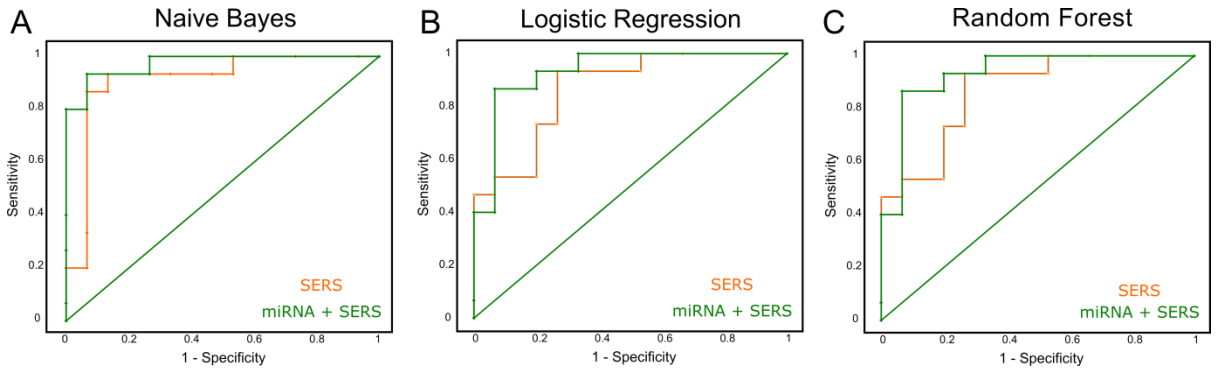


Figure 7. Receiver operating characteristic (ROC) curves for the classification accuracy yielded by SERS alone and the combination of the SERS and miRNA using three supervised classification algorithms (naïve Bayes (A), logistic regression (B), and random forest (C)).

As for the prostate cancer, SERS reached a classification accuracy of 80.95%, with sensitivity of 80% and a specificity of 81.81% (Table 2).

Table 2. The performance metrics for the machine learning classifiers applied to PC1 and PC2 on the validation group consisting on the SERS spectra of urine from 21 patients with PCa and BPH.

Classifier	AUC	CA	F1	Precision	Recall	Specificity
SVM	0.945	0.762	0.759	0.788	0.762	0.774
Random Forest	0.904	0.857	0.857	0.861	0.857	0.861
Naïve Bayes	0.840	0.714	0.714	0.714	0.714	0.713
kNN	0.927	0.762	0.755	0.782	0.762	0.747
Logistic Regression	0.945	0.857	0.856	0.859	0.857	0.852

When transitioning from using SERS urine biopsy as a screening tool, which aims for balanced specificity and sensitivity, to a support tool, where the goal is to ensure no true positive patients are missed, SERS urine biopsy has been found to be useful as a support tool to reduce the number of unnecessary biopsies in patients with clinically suspected prostate cancer. I have demonstrated that by fine-tuning a machine learning classifier, 100% sensitivity can be achieved for patients with Prostate Imaging-Reporting and Data System (PI-RADS) 3 lesions (patients for whom the clinical recommendation for biopsy is equivocal) at the cost of reduced specificity from 81.81% for the screening tool to 41.67% for the support tool. This result suggests that using this setup, 10 out of 24 patients could have been saved from unnecessary prostate biopsies.

Table 3. The threshold and the Specificity of the ML algorithms applied on training set such that 100% Sensitivity is achieved.

ML	Threshold	Sensitivity	Specificity
SVM	0.22	1.00	0.630
Random Forest	0.42	1.00	0.815
Naïve Bayes	0.23	1.00	0.481
kNN	0.51	1.00	0.510
Logistic Regression	0.08	1.00	0.556

I want to stress out that in this study external validation was used, the reliability of the results being sustained by the similarity classification accuracies being reached in both training (Table 3) and validation steps (Table 4).

Table 4. Performance of the ML ensemble based on the majority-voting rule. Confusion matrix of the test data set for the classification of patients with PI-RADS 3, with histologically proven prostate adenocarcinoma (PCa) and patients with benign prostate hyperplasia (BPH).

	Sample	**Predicted**	
		PCa	BPH
Actual	PCa	10	0
	BPH	14	10

SERS serum biopsy combined with an ensemble of ML classifiers assessed a sensitivity of 100% and a specificity of 41.67% for the stratification of PCa and BPH samples.

The subchapter 3.4 focuses onto the models employed on the SERS spectra of biofluids. The need of Principal Component Analysis (the most widely used ML tool in the clinical applications of SERS) is questioned in **subchapters 3.4.1 and 3.4.2**.

Instead of focusing on the accuracy of classifying samples, in **subchapter 3.4** the research shifts to refining the algorithms used for analyzing SERS spectra. Specifically, it examines the need for using Principal Component Analysis (PCA) as a preprocessing step before applying machine learning classification models. The majority of ML classifiers rely on a dimensionality reduction technique that provides orthogonal features, regardless of whether it significantly improves classification accuracy. Additionally, it is also revealed that by applying a t-test feature selection algorithm on the principal components and feeding only the most discriminant information to the machine learning classifiers, the accuracy of classification can be improved. Moreover, not all the Principal Components (PCs) contain discriminant information, since PCA is an unsupervised technique. I show

that the classification accuracy can be increased by extracting the PCs that stand for the discrimination of classes out of the PCs that contain maximum of the initial information without adding redundant data by analysing a SERS data set consisting of SERS spectra of serum from n=23 renal cancer (RCC) and n=27 control (CTRL) patients.

PCA was applied to the SERS spectra of serum of RCC and CTRL patients to reduce data dimensionality and analyse the significant features that cluster the RCC and CTRL groups. The first 9 PCs were kept for further analysis, which explained 98% of the variance in the initial dataset (Figure 8).

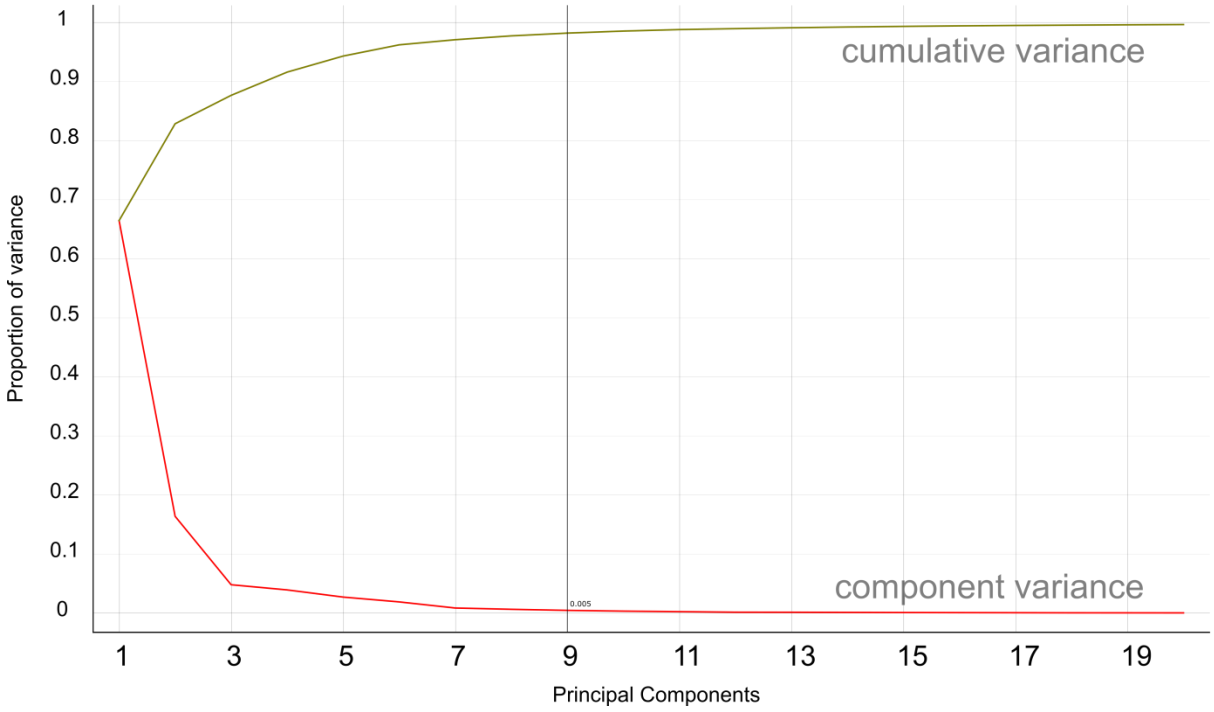


Figure 8. The cumulative (dark yellow) and component (red) explained variance for the first 19 PCs.[28]

The differences of score values of the 9 PCs between the RCC and CTRL groups were tested by Student's t-test. The scores of PC2 and PC6 showed statistically significant differences between the two groups ($p<0.05$). The score plot for PC2 and PC6 showed a clear tendency of clustering of the two groups.

To test the performance of SERS profiling as a screening tool for RCC, three ML algorithms were used (GNB, LR and RF). First we used as input the first 9 PCs and calculated the performance metrics of the ML algorithm to classify RCC and CTRL (Table 5), using random sampling cross-validation. The ML algorithm applied to the first 9 PCs yielded classification accuracies in the range of 60-67%. The number of PCs was chosen based on the cumulative explained variance; so that we kept most of the

initial information (9 PCs explained 98% from the initial variance-Figure 8), without adding redundant data.

Table 5. The diagnostic ability to distinguish renal cell carcinoma and control group patients with three classification algorithms (naïve Bayes, logistic regression, and random forest) run on the first 9 PCs, which explain 98% of initial variance.

ML model	AUC	CA	F1	Precision	Recall
Naïve Bayes	0.74	0.63	0.63	0.63	0.63
Logistic regression	0.58	0.60	0.60	0.60	0.60
Random forest	0.71	0.67	0.66	0.67	0.67

The same ML were employed using as input features only the two PCs that are different distributed in the RCC and CTRL groups, PC2 and PC6. The performance metrics of the three classifiers after random-sampling cross-validation are presented in Table 6. By reducing the information from the first 9 PCs (98% of the variance in the data set), to PC2 and PC6 (17.2% of the variance in the data set), the classification accuracies increased from 60-67% to 76-86%. The trend was not specific to a certain ML model, but all the models performed better when an extraction feature algorithm was applied prior to the classification model. PC2 (explains 14,2% of information) and PC6 (explains 3% of information) were chosen based on t-test feature selection with Mann-Whitney correction (data did not pass the normality test).

Table 6. The diagnostic ability to distinguish renal cell carcinoma and control group patients with three classification algorithms (naïve Bayes, logistic regression, and random forest) using as input features PC2 and PC6.

ML model	AUC	CA	F1	Precision	Recall
Naïve Bayes	0.91	0.86	0.86	0.86	0.86
Logistic regression	0.83	0.76	0.76	0.76	0.76
Random forest	0.84	0.80	0.79	0.80	0.80

Figure 9 presents the ROC curves of the ML algorithms applied to classify RCC and CTRL samples based on the SERS spectra of serum. As observed, the area under the curve in all ML algorithms is grater when only PC2 and PC6 are used as input, compared to all the first 9PCs.

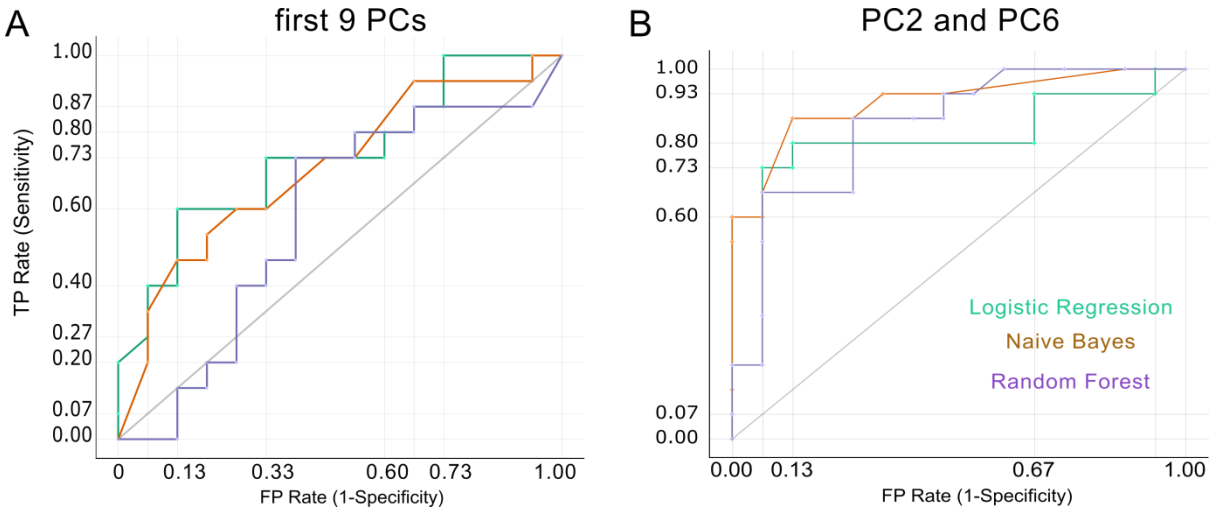


Figure 9. The receiver operating characteristic (ROC) curves for the classification accuracy yielded by SERS analysis of serum from renal cancer and control patients using three supervised classification algorithms (Naïve Bayes, logistic regression, and random forest). (A) The first 9 PCs were input variables and (B) PC2 and PC6 were input variables.

In **subchapter 3.4.3**, I investigated more thoroughly the contribution of purine metabolites, creatinine and urea to the SERS spectra. Knowing that most of the spectral information in biofluids SERS is coming from purine metabolites, creatinine, and urea. The SERS spectra of the biofluids were projected into a lower-dimensional subspace defined by the spectra of those pure metabolites. The sample's scores on each metabolite (uric acid, hypoxanthine, xanthine, creatinine, and urea) were then used to build classification models to differentiate cancer patients and controls, thus allows identifying the specific contribution of each metabolite to the results.

To be able to correlate the classification of cancer and controls based on the SERS spectra of biofluid with a medical outcome, I calculated the similarities between the SERS spectra of biofluids and the SERS spectra of pure solution of uric acid, hypoxanthine, xanthine, creatinine and urea. The score of each sample onto the new vectors representing the uric acid, hypoxanthine, xanthine, and creatinine and urea contribution was used to classify the patients.

We applied the model on a dataset of SERS spectra from serum from patients with prostate cancer (n=30) and controls (n=20). The normality of the data distribution was tested using D'Agostino Pearson test and data did not pass the normality test. Thus, t-test with Mann-Whitney correction was applied to test de differences in the distribution of metabolites vectors in cancer and control groups.

The scores on uric acid (p=0.001), hypoxanthine (p=0.017) and urea (0.003) were different between cancer and controls, while for xanthine (p=0.89) and creatinine (0.73) no significance was found.

The overall classification accuracy of prostate and control patients reached an AUC of the ROC curve of 69% (p=0.02) (Figure 10). A linear combination of the three significant components (uric acid, hypoxanthine and urea) was calculated with equal weights and a ROC curve analysis was applied to test the overall classification accuracy of the model for the classification of prostate and control patients.

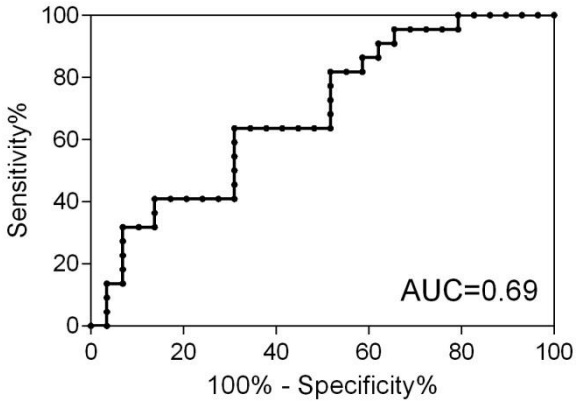


Figure 10. Receiver operating characteristic curve of the classification of prostate cancer and control patients based on the projection of the SERS serum samples on uric acid, hypoxanthine and urea vectors.

The classification accuracy is similar with the classification accuracy of ML models when PCA was applied. The ML classifiers applied to the first 19PCs which contained al 96.8% of the variance in the data set reached a classification accuracy of 64-71%. Only by applying feature selection technique (PC1 and PC2 had discrimination power) the accuracy was slightly increased to 63-77%, which surpasses the result presented above. However, in this case we know exactly the contributors to the classification.

The multivariate analysis based on machine learning analysis and statistical test have the risk of overfitting, especially when the number of samples is small. The excessive use of ML in every decision in the spectra processing steps is not always justified. Misconceptions about statistics and data analysis and their misinterpretation and application lead to the lack of reproducibility of many findings [29]. When the differences between the groups are evident and univariate, simple methods can be used to classify the models, why do we have to resort to complex algorithms that risk overfitting?

In **subchapter 3.4.4**, I aim to compare the performance of multivariate classification based on SERS spectra of serum to a univariate statistical model for cancer screening based on the 740/725 cm^{-1} SERS bands ratio, which was correlated with the variation of the uric acid and hypoxanthine. The

univariate model was validated on a variety of cancer types, from previously published data, measured on different Raman spectrometers across a time span of about two years.

Visually analysing the SERS spectra and loading plots resulted from the PCs with discriminant power it can be observed that the major discrepancy between cancer and controls is the doublet at 725 and 740 cm^{-1}. Thus, I tested a univariate analysis tool based on the ratio between the contribution of uric acid and hypoxanthine (740/725 cm^{-1} SERS bands) to classify the cancer and control patients on nine already published SERS serum biopsy data sets. Similar classification accuracies for six studies were obtained when using the univariate analysis compared to a complex ML classifier.

A zoom-in into the 710-760 cm^{-1} spectral region of SERS spectra of serum is shown in Figure 11A-F for colorectal (CRC), prostate (PC), breast (BC), lung (LC), oral (OR) and ovarian (OV) cancer, together with the control group from each study. The negative correlation of the 740 and 725 cm^{-1} SERS band is observed for all the above-mentioned studies, regardless of the type of cancer, Raman equipment or the acquisition parameters. The results of the cancer screening by the ratio of 740 and 725 cm^{-1} SERS band are presented as AUC in FigureG. The classification accuracy of the univariate model is between 67% for lung cancer and 95% for ovarian cancer. Interestingly, in some cases the univariate analysis yielded a similar accuracy compared to the multivariate analysis, highlighting the fact that some SERS spectral regions are actually adding to the classification error instead of improving the classification power. For example, Moisoiu et. al. reported for ovarian cancer an accuracy of 94.8% and 92.15% for oral cancer when using multivariate analysis (PCA-LDA algorithm) [30]. The univariate analysis of the same spectra reaches a similar accuracy of 92% for oral cancer and 95% for ovarian cancer [30]. In contrast, for lung cancer [30], prostate cancer [31], breast cancer [18] and colorectal cancer (paper in preparation) the multivariate analysis yielded a better classification of the cancer and control groups. Lung cancer and controls were classified with an accuracy of 85.75% when PCA-LDA model 5-folds cross-validated was used and an accuracy of 67% when univariate analysis was applied [30]. The PCA-LDA model applied for prostate cancer reached an accuracy of 90%, while the univariate model had an accuracy of 76% [31]. In the case of breast cancer, univariate model yields lower accuracy compared to LDA (73% compared to 83%), but higher than other machine learning classifiers (decision tree, Naive Bayes, Support vector machine, random forest), which discriminate BC from CTRL with accuracies between 61 and 72% [18]. For colorectal cancer, LR showed a classification accuracy of 75.75% (data not shown). A similar performance of 70% was assessed by the univariate model.

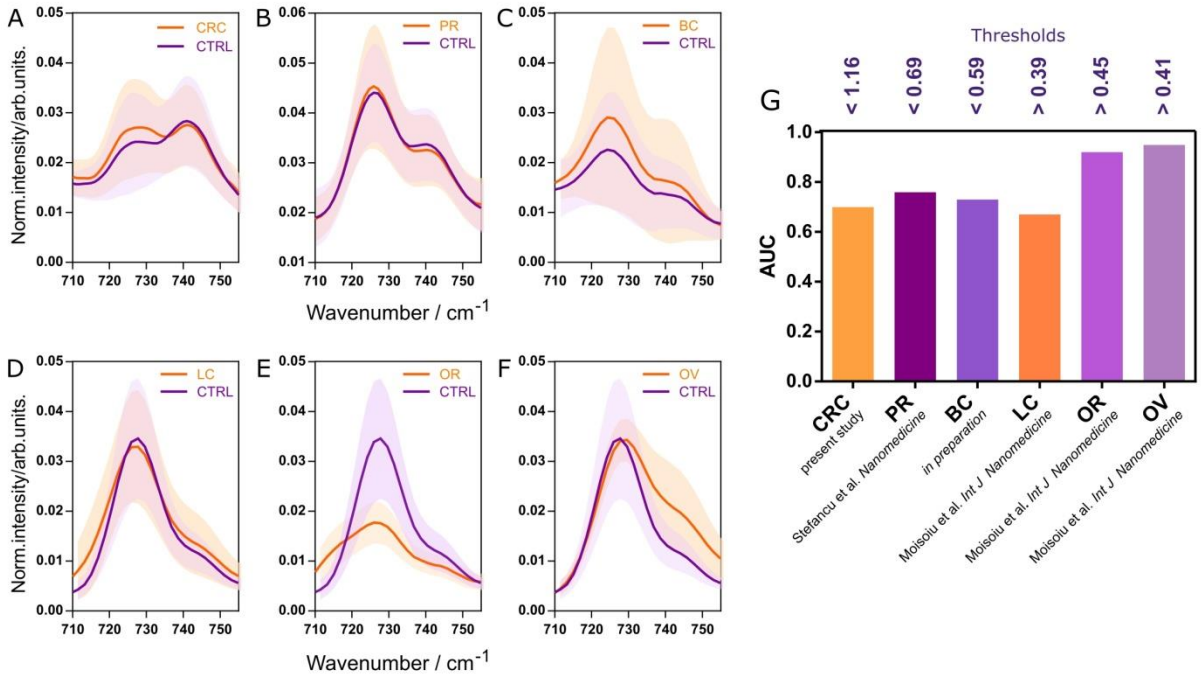
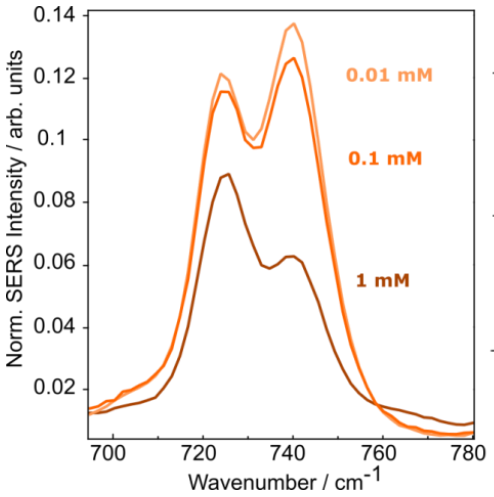


Figure 11. Univariate cancer diagnosis based on the SERS intensities at 725 cm^{-1} and 740 cm^{-1}. (A-F) Mean SERS spectra of cancer group (colorectal- CRC, prostate - PR, breast-BC, lung-LC, oral-OR and ovarian-OV cancer) and control group of each study. (G) AUC and optimal threshold for ROC curve built on 740/725 cm^{-1} SERS intensity ratio for each study.

To strength the idea that the ratio between the 740 and 725 cm^{-1} SERS band reflects an imbalance in uric acid and hypoxanthine concentration, we acquired SERS spectra of mixtures of hypoxanthine, xanthine and uric acid, which are known to be the major contributors to the SERS spectra of serum. For this, we kept the concentration of xanthine and hypoxanthine constant and vary the uric acid concentration. As observed in the Figure 12 the 740/725 cm^{-1} intensity is lower with higher uric acid concentration. A similar spectral change is observed when comparing controls with cancer mean SERS spectrum (Figure 11A-F).



Uric acid concentration (mM)	740/725 cm^{-1} intensity
1	0.70
0.1	1.09
0.01	1.13

Figure 12. SERS spectra of a mixture of 0.1 mM hypoxanthine, 0.1 mM xanthine and uric acid ranging from 0.01 to 1 mM and the ratio of 740/725 cm^{-1} SERS band.

These results are highly interesting since they could allow the detection of cancer samples simply based on the SERS intensity of two peaks. Moreover, as opposed to complex machine learning algorithms, the univariate discrimination is highly interpretable. In other words, we can pinpoint exactly the metabolic differences (hypoxanthine and uric acid) which give rise to the SERS spectral differences, and use those differences for classification. In addition, the validation of the model needs lower number of samples compared to the validation of machine learning models.

Final conclusions

The visual and mathematical analysis used in this thesis shows that the major contribution to SERS-based diagnosis is the imbalance between hypoxanthine and uric acid in biofluids. Since the translation of SERS into clinical settings lacks an SOP, this thesis briefly describes a workflow recommendation for SERS liquid biopsy. Starting from the acquisition of the SERS spectra, urine was found to be the optimal biofluid for SERS-based cancer screening; since the processing of spectra is an important factor in machine learning approaches, I showed that it is necessary to reduce the spectral data to orthogonal features/variables, out of which only the features with discriminating power should be kept; classifiers fed with the extracted features should be employed to predict previously unseen data for a rigours evaluation of SERS liquid biopsy performance. However, to be more confident in the component leading to class prediction, the data dimensionality reduction technique can be constrained to known metabolites (uric acid, hypoxanthine, xanthine, creatinine and urea) that are further used for sample classification. As an outlook, a combination of the projection of the SERS spectra on purine metabolites, creatinine and urea and the residues resulted from this reduction technique could help in the medical interpretation of cancer screening based on SERS liquid biopsy.

I analysed the performance of SERS urine biopsy in cancer screening and as clinical support tool where the decision based on the actual clinical techniques (such as blood biomarkers, MRI etc.) are equivocal. The classification accuracy over 80% attained for prostate cancer and benign prostate pathologies patients, based on urine SERS spectra, and of 71-84% for bladder cancer and control patients validated previous literature results and recommends SERS urine biopsy for cancer screening of various malignancies.

To increase the classification accuracy, SERS urine biopsy can be combined with complementary techniques, such as microRNA. I showed that by combining the two techniques the classification accuracy of bladder cancer and control patients was increased by up to 9%.

Further, I described a new support tool for biopsy decision management of equivocal PI-RADS 3 lesions, based on SERS urine biopsy, which provides 100% sensitivity when validated on an external data set (previously unseen data). However, I gained maximum sensitivity (a medical requirement for a cancer support tool) in the detriment of specificity, which dropped down from 81.81% to 41%. Yet, based on the specificity value, SERS urine biopsy can target 10 from 24 BPH patients, preventing them of unnecessary biopsies.

Regarding spectra analysis, PCA is still necessary in the data mining approaches applied to SERS spectra, since most of the ML classifiers ask the input features to be orthogonal (which is not the

case for SERS spectral features). However, improvement can be made by extracting the Principal Components with discrimination power out of the PCs that cover the total variability in the SERS data sets. One way to extract the relevant PCs is using t-test feature extraction, which increases the classification accuracy of prostate cancer based on the SERS spectra of serum with 10-20%.

I developed a dimensionality reduction technique specific for SERS spectra of biofluids. To this regard, I projected the SERS spectra of serum onto vectors characteristic to the SERS spectra of uric acid, hypoxanthine, xanthine, creatinine and urea. I tested the differences between the metabolite scores of prostate cancer and control patients, and observed that uric acid, hypoxanthine and urea are responsible for the differences between the SERS spectra of prostate cancer and controls. The classification accuracy of prostate and control patients based on the equal contribution of uric acid, hypoxanthine and urea reached a classification accuracy of 69%, a value similar to the classification accuracy assessed by the ML classifiers applied on the PCs. Though, I observed that there is information in the spectra that is not correlated with uric acid, hypoxanthine, xanthine, creatinine and urea. This information highlights the minor contributors that could have importance in cancer screening based on SERS liquid biopsy.

Observing the recurrent pattern of the negative correlation of hypoxanthine and uric acid in the SERS spectra comparisons between cancer and control patients, we analysed the ratio between the hypoxanthine (725 and 740 cm^{-1} SERS band) and uric acid (725 cm^{-1} SERS band) expression. We highlighted that a univariate model based on the 725/740 cm^{-1} ratio can compete with the multivariate ML classifier.

I hope that these clinical spectroscopy studies contribute to the translation of SERS into the clinical setting, as support tool for diagnostics, and as rapid and cost effective screening method for cancer.

References

[1] N. Leopold, Surface-enhanced Raman spectroscopy, selected applications, Editura Napoca Star (2009).

[2] W.-H. Park, Z.H. Kim, Charge Transfer Enhancement in the SERS of a Single Molecule, Nano Letters 10(10) (2010) 4040-4048.

[3] F. Huber, J. Berwanger, S. Polesya, S. Mankovsky, H. Ebert, F.J. Giessibl, Chemical bond formation showing a transition from physisorption to chemisorption, Science 366(6462) (2019) 235-238.

[4] A. Otto, I. Mrozek, H. Grabhorn, W. Akemann, Surface-enhanced Raman scattering, Journal of Physics: Condensed Matter 4(5) (1992) 1143.

[5] A. Otto, A. Bruckbauer, Y.X. Chen, On the chloride activation in SERS and single molecule SERS, Journal of Molecular Structure 661-662 (2003) 501-514.

[6] S.D. Iancu, A. Stefancu, V. Moisoiu, L.F. Leopold, N. Leopold, The role of Ag+, Ca2+, Pb2+ and Al3+ adions in the SERS turn-on effect of anionic analytes, Beilstein journal of nanotechnology 10 (2019) 2338-2345.

[7] A. Stefancu, S. Lee, L. Zhu, M. Liu, R.C. Lucacel, E. Cortés, N. Leopold, Fermi Level Equilibration at the Metal–Molecule Interface in Plasmonic Systems, Nano Letters 21(15) (2021) 6592-6599.

[8] R. El Ridi, H. Tallima, Physiological functions and pathogenic potential of uric acid: A review, J Adv Res 8(5) (2017) 487-493.

[9] Y.B. Ian Goodfellow, Aaron Courville, Deep learning, MIT Press (2016).

[10] W.S. Noble, What is a support vector machine?, Nature biotechnology 24(12) (2006) 1565-7.

[11] E.W. Steyerberg, Clinical Prediction Models A Practical Approach to Development, Validation, and Updating, Springer New York, NY (2009).

[12] B.H. Menze, B.M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, F.A. Hamprecht, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data, BMC Bioinformatics 10(1) (2009) 213.

[13] L. Breiman, Random Forests, Machine Learning 45(1) (2001) 5-32.

[14] Y. Liu, B.R. Upadhyaya, M. Naghedolfeizi, Chemometric Data Analysis Using Artificial Neural Networks, Applied spectroscopy 47(1) (1993) 12-23.

[15] M. Poth, G. Magill, A. Filgertshofer, O. Popp, T. Großkopf, Extensive evaluation of machine learning models and data preprocessings for Raman modeling in bioprocessing, Journal of Raman Spectroscopy 53(9) (2022) 1580-1591.

[16] P.S. Gromski, H. Muhamadali, D.I. Ellis, Y. Xu, E. Correa, M.L. Turner, R. Goodacre, A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding, Anal Chim Acta 879 (2015) 10-23.

[17] 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, TrAC Trends in Analytical Chemistry 124 (2020) 115796.

[18] S.D. Iancu, R.G. Cozan, A. Stefancu, M. David, T. Moisoiu, C. Moroz-Dubenco, A. Bajcsi, C. Chira, A. Andreica, L.F. Leopold, D. Eniu, A. Staicu, I. Goidescu, C. Socaciu, D.T. Eniu, L. Diosan, N. Leopold, SERS liquid biopsy in breast cancer. What can we learn from SERS on serum and urine?, Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy 273 (2022).

[19] J. Lin, Z. Huang, X. Lin, Q. Wu, K. Quan, Y. Cheng, M. Zheng, J. Xu, Y. Dai, H. Qiu, D. Lin, S. Feng, Rapid and label-free urine test based on surface-enhanced Raman spectroscopy for the non-invasive detection of colorectal cancer at different stages, Biomedical optics express 11(12) (2020) 7109-7119.

[20] S. Fornasaro, E. Gurian, S. Pagarin, E. Genova, G. Stocco, G. Decorti, V. Sergo, A. Bonifacio, Ergothioneine, a dietary amino acid with a high relevance for the interpretation of label-free surface enhanced Raman scattering (SERS) spectra of many biological samples, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 246 (2021) 119024.

[21] M. Li, Y. Du, F. Zhao, J. Zeng, C. Mohan, W.-C. Shih, Reagent- and separation-free measurements of urine creatinine concentration using stamping surface enhanced Raman scattering (S-SERS), Biomedical Optics Express 6 (2015).

[22] M. Casella, A. Lucotti, M. Tommasini, M. Bedoni, E. Forvi, F. Gramatica, G. Zerbi, Raman and SERS recognition of β-carotene and haemoglobin fingerprints in human whole blood, Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy 79(5) (2011) 915-9.

[23] S.D. Iancu, R.G. Cozan, A. Stefancu, M. David, T. Moisoiu, C. Moroz-Dubenco, A. Bajcsi, C. Chira, A. Andreica, L.F. Leopold, D. Eniu, A. Staicu, I. Goidescu, C. Socaciu, D.T. Eniu, L. Diosan, N. Leopold, SERS liquid biopsy in breast cancer. What can we learn from SERS on serum and urine?, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy (2022) 120992.

[24] J. Chowdhury, K.M. Mukherjee, T.N. Misra, A pH dependent surface-enhanced Raman scattering study of hypoxanthine, Journal of Raman Spectroscopy 31(5) (2000) 427-431.

[25] M. Casella, A. Lucotti, M. Tommasini, M. Bedoni, E. Forvi, F. Gramatica, G. Zerbi, Raman and SERS recognition of β-carotene and haemoglobin fingerprints in human whole blood, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 79(5) (2011) 915-919.

[26] R.M.Y. Tang, I.K.-M. Cheah, T.S.K. Yew, B. Halliwell, Distribution and accumulation of dietary ergothioneine and its metabolites in mouse tissues, Scientific reports 8(1) (2018) 1601.

[27] A. Stefancu, V. Moisoiu, C. Bocsa, Z. Bálint, D.-T. Cosma, I.A. Veresiu, V. Chiş, N. Leopold, F. Elec, SERS-based quantification of albuminuria in the normal-to-mildly increased range, The Analyst 143(22) (2018) 5372-5379.

[28] T. Moisoiu, S.D. Iancu, D. Burghelea, M.P. Dragomir, G. Iacob, A. Stefancu, R.G. Cozan, O. Antal, Z. Bálint, V. Muntean, R.I. Badea, E. Licarete, N. Leopold, F.I. Elec, SERS Liquid Biopsy Profiling of Serum for the Diagnosis of Kidney Cancer, Biomedicines 10(2) (2022) 233.

[29] H.J. Motulsky, Common misconceptions about data analysis and statistics, Naunyn-Schmiedeberg's archives of pharmacology 387(11) (2014) 1017-23.

[30] V. Moisoiu, A. Stefancu, D. Gulei, R. Boitor, L. Magdo, L. Raduly, S. Pasca, P. Kubelac, N. Mehterov, V. Chiș, M. Simon, M. Muresan, A.I. Irimie, M. Baciut, R. Stiufiuc, I.E. Pavel, P. Achimas-Cadariu, C. Ionescu, V. Lazar, V. Sarafian, I. Notingher, N. Leopold, I. Berindan-Neagoe, SERS-based differential diagnosis between multiple solid malignancies: breast, colorectal, lung, ovarian and oral cancer, International journal of nanomedicine 14 (2019) 6165-6178.

[31] 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, Analytical and bioanalytical chemistry 411(22) (2019) 5877-5883.

List of publications on the subject of the thesis

SERS liquid biopsy: An emerging tool for medical diagnosis**,** V Moisoiu*, SD Iancu*, A Stefancu*, T Moisoiu, B Pardini, MP Dragomir, N Crisan, L Avram, D Crisan, I Andras, D Fodor, LF Leopold, C Socaciu, Z Bálint, C Tomuleasa, F Elec, N Leopold**,** Colloids and Surfaces B: Biointerfaces**,** 208, 2021, 112064.

IF 5.268, Rank Q1 (BIOPHYSICS) AIS 0.707

Combined miRNA and SERS urine liquid biopsy for the point-of-care diagnosis and molecular stratification of bladder cancer T Moisoiu*, MP Dragomir*, SD Iancu*, S Schallenberg, G Birolo, G Ferrero, D Burghelea, A Stefancu, RG Cozan, E Licarete, A Allione, G Matullo, G Iacob, Z Bálint, RI Badea, A Naccarati, D Horst, B Pardini, N Leopold, F Elec, Molecular Medicine 2022, 28 (1).

IF 6.354, Rank Q1 (BIOCHEMISTRY & MOLECULAR BIOLOGY) AIS 1.220

SERS liquid biopsy in breast cancer. What can we learn from SERS on serum and urine? SD Iancu, RG Cozan, A Stefancu, M David, T Moisoiu, C Moroz-Dubenco, A Bacjsi, C Chira, A Andreica, L Diosan, N Leopold, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2022;273:120992.

IF 4.098, Rank Q1 (SPECTROSCOPY) AIS 0.491

SERS Liquid Biopsy Profiling of Serum for the Diagnosis of Kidney Cancer T Moisoiu*, SD Iancu*, D Burghelea, MP Dragomir, G Iacob, A Stefancu, RG Cozan, O Antal, Z Bálint, V Muntean, RI Badea, E Licarete, N Leopold, FI Elec. Biomedicines 2022, 10 (2).

IF 6.081, Rank Q2 (BIOCHEMISTRY & MOLECULAR BIOLOGY) AIS 0.799

*these authors contributed equally as first authors

List of publications

SERS-based DNA methylation profiling allows the differential diagnosis of malignant lymphadenopathy Stefancu, A.; Moisoiu, V.; Desmirean, M.; Iancu, S.D.; Tigu, A.B.; Petrushev, B.; Jurj, A.; Cozan, R.G.; Budisan, L.; Fetica, B., et al. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy 2022, 264.

Gadolinium Accumulation and Toxicity on In Vitro Grown Stevia rebaudiana: A Case-Study on Gadobutrol Scurtu, V.F.; Clapa, D.; Leopold, L.F.; Ranga, F.; Iancu, Ș.D.; Cadiș, A.I.; Coman, V.; Socaci, S.A.; Moț, A.C.; Coman, C. International Journal of Molecular Sciences 2022, 23,.

The effect of 100–200 nm ZnO and TiO2 nanoparticles on the in vitro-grown soybean plants Leopold, L.F.; Coman, C.; Clapa, D.; Oprea, I.; Toma, A.; Iancu, Ș.D.; Barbu-Tudoran, L.; Suciu, M.; Ciorîță, A.; Cadiș, A.I., et al. Colloids and Surfaces B: Biointerfaces 2022, 216, doi:10.1016/j.colsurfb.2022.112536.

Antibacterial Effect of Eco-Friendly Silver Nanoparticles and Traditional Techniques on Aged Heritage Textile, Investigated by Dark-Field Microscopy Ilieș, A.; Hodor, N.; Pantea, E.; Ilieș, D.C.; Indrie, L.; Zdrîncă, M.; Iancu, S.; Caciora, T.; Chiriac, A.; Ghergheles, C., et al. Coatings 2022, 12.

The Use of Machine Learning Algorithms and the Mass Spectrometry Lipidomic Profile of Serum for the Evaluation of Tacrolimus Exposure and Toxicity in Kidney Transplant Recipients Burghelea, D.; Moisoiu, T.; Ivan, C.; Elec, A.; Munteanu, A.; Iancu, Ș.D.; Truta, A.; Kacso, T.P.; Antal, O.; Socaciu, C., et al. Biomedicines 2022.

Stage related metabolic profile of the synovial fluid in patients with acute flares of knee osteoarthritis Bocsa, D.C.; Socaciu, C.; Iancu, S.D.; Pelea, M.A.; Gutiu, R.I.; Leopold, N.; Fodor, D. Medicine and Pharmacy Reports 2022, 95, 438-445.

Selective Single Molecule SERRS of Cationic and Anionic Dyes by Cl-and Mg2+Adions: An Old New Idea Stefancu, A.; Iancu, S.D.; Leopold, N. Journal of Physical Chemistry C 2021, 125, 12802-12810.

Tuning the potential of nanoelectrodes to maximum: Ag and au nanoparticles dissolution by I– adsorption via Mg2+ adions Stefancu, A.; Iancu, S.D.; Coman, V.; Leopold, L.F.; Leopold, N. Romanian Reports in Physics 2021, 73.

SERS-Based Evaluation of the DNA Methylation Pattern Associated With Progression in Clonal Leukemogenesis of Down Syndrome Moisoiu, V.; Sas, V.; Stefancu, A.; Iancu, S.D.; Jurj, A.; Pasca, S.;

Iluta, S.; Zimta, A.A.; Tigu, A.B.; Teodorescu, P., et al. Frontiers in Bioengineering and Biotechnology 2021, 9.

Correlation between Volumes Determined by Echocardiography and Cardiac MRI in Controls and Atrial Fibrillation Patients Manole, S.; Budurea, C.; Pop, S.; Iliescu, A.M.; Ciortea, C.A.; Iancu, S.D.; Popa, L.; Coman, M.; Szabó, L.; Coman, V., et al. Life 2021, 11.

Adduct of Aquacobalamin with Hydrogen Peroxide Lehene, M.; Plesa, D.; Ionescu-Zinca, S.; Iancu, S.D.; Leopold, N.; Makarov, S.V.; Brânzanic, A.M.V.; Silaghi-Dumitrescu, R. Inorganic Chemistry 2021, 60, 12681-12684.

SERS-Based Assessment of MRD in Acute Promyelocytic Leukemia? Turcas, C.; Moisoiu, V.; Stefancu, A.; Jurj, A.; Iancu, S.D.; Teodorescu, P.; Pasca, S.; Bojan, A.; Trifa, A.; Iluta, S., et al. Frontiers in Oncology 2020, 10.

Photothermal property assessment of gold nanoparticle assemblies obtained by hydroxylamine reduction Tódor, I.S.; Marişca, O.T.; Rugină, D.; Diaconeasa, Z.; Leopold, L.F.; Coman, C.; Antonescu, E.; Szabó, L.; Iancu, S.D.; Bálint, Z., et al. Colloid and Polymer Science 2020, 298, 1369-1377.

Contribution of chemical interface damping to the shift of surface plasmon resonance energy of gold nanoparticles Stefancu, A.; Iancu, S.D.; Leopold, L.F.; Leopold, N. Romanian Reports in Physics 2020, 72.

Combining surface-enhanced Raman scattering (SERS) of saliva and two-dimensional shear wave elastography (2D-SWE) of the parotid glands in the diagnosis of Sjögren's syndrome Moisoiu, V.; Badarinza, M.; Stefancu, A.; Iancu, S.D.; Serban, O.; Leopold, N.; Fodor, D. Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy 2020, 235.

Assessment of gold-coated iron oxide nanoparticles as negative T2 contrast agent in small animal MRI studies Iancu, S.D.; Albu, C.; Chiriac, L.; Moldovan, R.; Stefancu, A.; Moisoiu, V.; Coman, V.; Szabo, L.; Leopold, N.; Bálint, Z. International journal of nanomedicine 2020, 15, 4811-4824.

SERS-based liquid biopsy of gastrointestinal tumors using a portable raman device operating in a clinical environment Avram, L.; Iancu, S.D.; Stefancu, A.; Moisoiu, V.; Colnita, A.; Marconi, D.; Donca, V.; Buzdugan, E.; Craciun, R.; Leopold, N., et al. Journal of Clinical Medicine 2020, 9.

SERS-based liquid biopsy of saliva and serum from patients with Sjögren's syndrome Stefancu, A.; Badarinza, M.; Moisoiu, V.; Iancu, S.D.; Serban, O.; Leopold, N.; Fodor, D. Analytical and bioanalytical chemistry 2019, 411, 5877-5883.

SERS assessment of the cancer-specific methylation pattern of genomic DNA: towards the detection of acute myeloid leukemia in patients undergoing hematopoietic stem cell transplantation Moisoiu, V.; Stefancu, A.; Iancu, S.D.; Moisoiu, T.; Loga, L.; Dican, L.; Alecsa, C.D.; Boros, I.; Jurj, A.; Dima, D., et al. Analytical and bioanalytical chemistry 2019, 411, 7907-7913.

Breast cancer diagnosis by surface-enhanced raman scattering (SERS) of urine Moisoiu, V.; Socaciu, A.; Stefancu, A.; Iancu, S.D.; Boros, I.; Alecsa, C.D.; Rachieriu, C.; Chiorean, A.R.; Eniu, D.; Leopold, N., et al. Applied Sciences (Switzerland) 2019, 9.

The role of Ag+, Ca2+, Pb2+ and Al3+ adions in the SERS turn-on effect of anionic analytes Iancu, S.D.; Stefancu, A.; Moisoiu, V.; Leopold, L.F.; Leopold, N. Beilstein journal of nanotechnology 2019, 10, 2338-2345.

Specific and selective sers active sites generation on silver nanoparticles by cationic and anionic adatoms Stefancu, A.; Iancu, S.D.; Moisoiu, V.; Leopold, N. Romanian Reports in Physics 2018, 70.

The role of adatoms in chloride-activated colloidal silver nanoparticles for surface-enhanced Raman scattering enhancement Leopold, N.; Stefancu, A.; Herman, K.; Tódor, I.S.; Iancu, S.D.; Moisoiu, V.; Leopold, L.F. Beilstein journal of nanotechnology 2018, 9, 2236-2247.