Research Article Vol. 60, No. 27 / 20 September 2021 / Applied Optics 8291

applied optics

Machine learning algorithms to control concentrations of carbon nanocomplexes in a biological medium via optical absorption spectroscopy: how to choose and what to expect?

Olga Sarmanova,[1,*] Kirill Laptinskiy,[2] Sergey Burikov,[1,2] Maria Khmeleva,[1,2] Anna Fedyanina,[1] Alexandra Tomskaya,[3,4] Aleksandr Efitorov,[2] Sergey Dolenko,[2] and Tatiana Dolenko[1,2]

[1]Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia
[2]Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow 119991, Russia
[3]A.M. Prokhorov General Physics Institute, Russian Academy of Science, Moscow 119991, Russia
[4]North-Eastern Federal University, Yakutsk 677007, Russia
*Corresponding author: oe.sarmanova@physics.msu.ru

Received 30 June 2021; revised 13 August 2021; accepted 19 August 2021; posted 19 August 2021 (Doc. ID 434984); published 13 September 2021

A solution of spectroscopic inverse problems, implying determination of target parameters of the research object via analysis of spectra of various origins, is an overly complex task, especially in case of strong variability of the research object. One of the most efficient approaches to solve such tasks is use of machine learning (ML) methods, which consider some unobvious information relevant to the problem that is present in the data. Here, we compare ML approaches to the problem of nanocomplex concentrations determination in human urine via optical absorption spectra, perform preliminary analysis of the data array, find optimal parameters for several of the most popular ML methods, and analyze the results.

https://doi.org/10.1364/AO.434984

1. INTRODUCTION

Nearly 20 commercial nano-based drugs had been approved by Food and Drug Administration (FDA) by 2019 [1], indicating bright prospects of nanotechnology application in medical purposes. FDA demands nanoagents injected into the human body to be cleared completely in reasonable time [2], which, in fact, restricts agents' exposure time as well as prohibits their accumulation in the body.

For small particles (less than 5 nm), elimination via the genitourinary system (urine) prevails, due to the effective pore size in normal intact endothelium being about 5–5.5 nm [2]. There are several works concerning nanoparticle clearance from the organism [2,3]. In [2], it was shown that quantum dots with hydrodynamic diameter less than 5.5 nm are quickly eliminated from the body. The authors [3] used confocal microscopy to demonstrate a significant difference between the excretion pathways of quantum dots with different zeta potentials. After intravenous administration, positively charged particles quickly left the organism, while negatively charged particles were accumulating in mouse kidneys for more than 30 days. To sum it all up, the problem of nanoparticle excretion from the organism is extremely relevant and important for effective theranostic nanoagent creation. It is crucial to perform express control over the nanoparticles' elimination from the organism to prevent their accumulation in the body.

Optical spectroscopy favorably stands out among the other techniques for monitoring the luminescent nanoparticles' elimination from the body for its noninvasiveness and the ability to register spectral signals in the on-line mode. The cornerstone of any spectroscopic method is the solution of inverse problems—extraction of the demanded information about the object from the spectral signal originating from the interaction between optical radiation and this object. A gradual increase in the number of problem parameters due to the complexity of the systems under study and the improvement of spectroscopic equipment leads to a complication of inverse problems. As a result, traditional methods to solve them often turn out to be ineffective, which is why so much effort is now channeled toward find alternatives that can be the key to solve a wide diversity of practical problems. One of the most promising alternatives is machine learning (ML) methods, primarily artificial neural networks [4].

ML algorithms are actively applied to spectroscopic data analysis to increase the efficiency of optical methods in many fields of science. For instance, Bayesian optimization efficiently

accelerates metamaterial design, allowing one to create novel structures such as ultranarrow-band wavelength-selective thermal radiators [5–7]. The main problem of spectral analysis in biomedicine is the complexity of biological media that hampers spectra analysis. It is not surprising that data-driven methods, which do not require detailed knowledge about the research object, are being widely integrated in biomedicine. ML algorithms are frequently used for qualitative research of biological samples or making diagnoses—that is, for solving classification problems. Huang *et al.* [8] introduced a method employing surface-enhanced Raman spectroscopy to determine spectral markers for recognition of heparin-induced thrombocytopenia. The developed platform is based on specific class of ML methods: chemometrics methods [principal components analysis (PCA), linear discriminant analysis, and partial least-squares (PLS)]. In [9], the application of Raman spectroscopy coupled with artificial neural networks (ANN) to discriminate between type 2 diabetes mellitus and healthy control groups provided this discrimination with 88.9%–90.9% accuracy, depending on the type of biotissue where the Raman signal was recorded. ML algorithms also represent a powerful tool for solving regression problems—numerical determination of parameters of optical spectroscopy inverse problems. Olaetxea *et al.* [10] employed Raman spectroscopy with PLS regression to monitor relevant variations of pH and lactate as known biomarkers of several pathologies.

The authors [11] successfully applied artificial neural networks for both pattern recognition and regression to solve the inverse problem of fluorescence spectroscopy. They performed optical imaging and determined the concentration of nanoagents based on nanodiamonds and carbon dots in biological tissues from their fluorescence spectra. The authors [12] demonstrated a successful solution to the problem of nanocomposites' renal clearance monitoring. The nanocomposites were based on nanometer graphene oxides, covered with poly(ethylene imine)–poly(ethylene glycol) copolymer and folic acid using fluorescence spectroscopy and artificial neural networks. The authors managed to recognize not only nanocomposites but also their components by their fluorescence and also managed to determine their concentration.

Even though most of the drug carriers approved by FDA are either liposomes or polymers, great efforts are channeled to develop commercially available drug carriers based on carbon dots (CDs). Such interest lies in CDs' unique properties including low toxicity; superb photophysical properties; easy surface functionalization; remarkable water solubility; and diversity of simple, fast, and cheap synthesis routes [13,14]. Although there are problems hampering CDs' clinical applications [15], the perspectives of their application as drug carriers seem promising.

This study presents the approach to solve the task of CD-based nanoagents (CD + doxorubicin) renal clearance control employing optical absorption spectroscopy and ML algorithms. Application of the latter is conditioned by the absorption properties of biological medium, as urine demonstrates sufficiently intense broadband absorption. Thus, one needs to isolate the absorption signal of nanocomplexes and their components against the background of urine absorption. The shape of the urine spectrum depends on a variety of factors: age and gender of the donor, his nutrition and health status, the time of sampling, and so on. The practical impossibility of controlling all these factors hinders creation of a "model" spectrum of urine absorption. Therefore, overcoming the variability of urine absorption comes to the fore.

Despite the seeming versatility of ML methods, not every model is suitable for solving a specific problem. Moreover, setting the parameters of a particular model has a significant impact on the quality of the solution. In this paper, the three-parameter inverse problem of optical absorption spectroscopy—determination of CD and doxorubicin (Dox) concentrations in urine, as well as pH value of the suspension, via optical absorption spectra—was solved employing artificial neural networks, linear regression (LR), PLS regression, random forest (RF), and gradient boosting (GB). Based on the obtained results, general recommendations on the choice of machine learning methods to solve the multiparameter inverse problem of optical absorption spectroscopy were formulated.

2. MATERIALS AND METHODS

A. Objects of Research

CDs were synthesized via "green chemistry" hydrothermal method. Briefly, 1.098 g of citric acid powder was dissolved in 7 ml of aqueous ethylenediamine and 20 ml of deionized water. The solution was treated in an ultrasonic bath for 5–10 min. After the treatment, the solution was kept in an autoclave for 2 h at 220°C. The resulting suspension contained CDs and synthesis products. To isolate CDs nanoparticles, the suspension was filtered through a track membrane with a pore size of 100 nm and through silica gel with a pore size of 10 nm. The resulting CD suspension was placed in a dialysis bag with a pore size of 1 kDa and stirred in water for 8 h using a magnetic stirrer to get rid of residues of various chemical reaction products.

Urine samples were taken from four healthy donors (both male and female) in different age groups (20–45 years), with their consent. Doxorubicin hydrochloride (Sigma-Aldrich CAS Number 25316-40-9) was used in the experiment.

B. Measurements

The values of the zeta potential and size of nanoparticles in the suspension were determined using the Malvern ZetaSizer Nano ZS (Malvern, Worcestershire, UK).

Infrared (IR) spectra of the CDs were recorded with a Bruker IFS 66 Fourier transform IR (FTIR) spectrometer.

The optical absorption spectra of the nanoparticles' suspensions in urine were registered in the range from 190 to 800 nm via a double-beam Shimadzu UV-1800 spectrophotometer with a scanning increment of 1 nm. Standard photometric cuvettes with 10 mm optical path length were used.

The pH values of aqueous and urine suspensions of nanoparticles were measured via ionometric converter Acvilon I-500, equipped with pH electrode InLab Ultra-Micro-ISM (Mettler Toledo). The measured pH values of the suspensions varied from 5.25 to 6.21.

C. Model of CD-Doxorubicin Nanocomplexes' Excretion from the Organism

The study considers the renal clearance model of nanocomposite components—CDs and doxorubicin. Under the natural conditions existing in different parts of the body, Dox, which is located on the CD surface due to physical adsorption, is ablated from the CD surface. Thus, we consider the case when the CD-Dox nanocomplex administered to a patient completely breaks down into two components, CDs and Dox, being excreted with urine. The proposed model does not exclude secondary adsorption of the drug on the nanoparticle surface. However, since adaptive algorithms are trained on real experimental spectra and consider all passing physical and chemical processes, the nonlinear contribution to the change in intensity due to secondary adsorption at certain concentrations will be considered in the process of computational model training.

D. Machine Learning Algorithms Application

In this study, three-parameter inverse problem of optical absorption spectroscopy—determination of CD and Dox concentrations in urine, as well as pH value of the suspension—was solved. For the inverse problem solution, the following algorithms have been chosen as basic ones, implemented with optimal parameter selection: A) Artificial neural networks of multilayer perceptron (MLP) type [16]. We used the logistic activation function in the hidden layers and the linear activation function in the output layer. The MLP was trained via stochastic gradient descent with 500 epochs passing after minimum error on the validation dataset as a stopping criterion. To eliminate the influence of weights' random initialization, each MLP was trained 5 times followed by averaging the results. B) Linear regression in a linear basis with preliminary data scaling [17], C) partial least-squares regression [18], D) random forest [19], E) gradient boosting [4].

The inverse problem was solved simultaneously for all the parameters, i.e., one simultaneously determined the CD and Dox concentrations and the pH value of the suspension. The dataset was divided into training, validation, and test sets to work with neural networks in the ratio of 70:20:10, and for other models the division into training and test sets was carried out in the ratio of 80:20. One used the training set to train the models; the validation set was used to ensure that the training was stopped in a timely manner; and the test set was for testing the quality of trained models on independent data.

For each ML algorithm, a five-fold cross-validation procedure was used to prevent the influence of the method of data splitting into sets and to assess the quality of the constructed predictive models more correctly. Thus, the original dataset was randomly divided into a training/validation/test set 5 times. The results of applying each of the ML algorithms to these sets were averaged.

These ML methods were implemented using the Keras and Scikit Learn libraries in Python [4].

3. RESULTS AND DISCUSSION

A. Characterization of Nanoparticles

FTIR spectroscopy results showed the multifunctional content of surface groups of synthesized CDs (Fig. 1.): there are bands of stretching vibrations of O-H and N-H/O-H groups at 3285 and 3209−3309 cm^{-1} [20–22] and of C-H groups at 2936 and 2875 cm^{-1} [23]. The appearance of peaks at 1659, 1573–1575, and 1485 cm^{-1} indicates the presence of groups CC, CO, and CN in the conjugated structure [24]. The peaks at 1434 cm^{-1} and 1384 cm^{-1} correspond to the stretching vibrations of CN [22] and to the asymmetric stretching vibrations of COC in the carboxylate groups [25], respectively. The peak at 818−819 cm^{-1} corresponds to the out-of-plane bending vibrations of CH groups of the benzene ring.

The values of the nanoparticle sizes and their zeta potential in an aqueous suspension with a concentration of 0.01 mg/mL are presented in Table 1. The value of the zeta potential modulus of CDs indicates their high colloidal stability.

B. Optical Absorption Spectroscopy

Four series of one- and two-component suspensions of CD and Dox in urine were prepared, corresponding to four samples of urine from different donors. CD concentration varied from 0 to 1.2 mg/L with 0.05 mg/L increment; Dox varied from 0 to 1 mg/L with 0.042 mg/L increment. In total, 624 absorption spectra CD and Dox suspensions were obtained, as well as four urine spectra corresponding to different experimental series. The spectra of one-component suspensions in urine are shown in Fig. 2(A). The preprocessing of the optical absorption spectra of the samples consisted of subtracting the minimum value for each spectrum.

Dox has a broad absorption band in the region 400–550 nm and characteristic band in the UV region (See SI1.). Broad

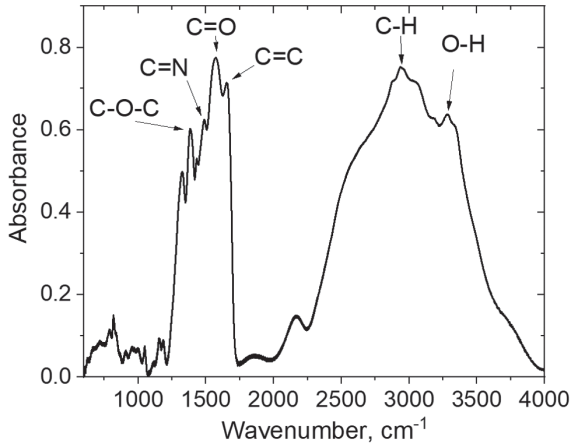

Fig. 1. FTIR spectrum of synthesized CDs.

Table 1. Characteristics of the Research Objects in an Aqueous Suspension (CD Concentration 0.01 mg/mL)

	CDs
Size, nm	11
Zeta potential, mV	−33.7

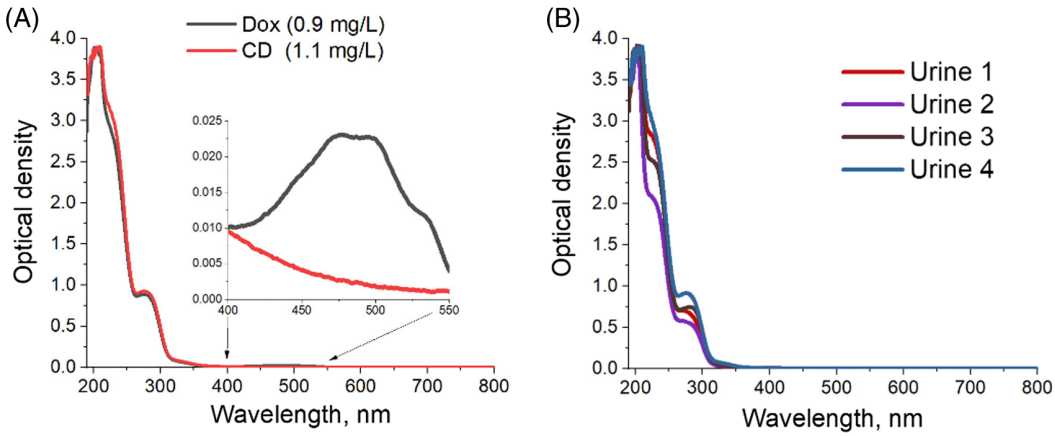

Fig. 2. (A) Optical absorption spectra of one-component suspensions of CD and Dox in urine; (B) optical absorption spectra of urine in different experimental series.

absorption bands in the UV and 300–400 nm regions correspond to CDs. The urine itself possess specific absorption spectra in the UV region. Moreover, the absorption spectra of the urine vary depending on the series (Fig. 2(B)). This is due to the complex chemical composition of urine and its variability depending on the processes in the human body.

Thus, absorption bands corresponding to the different components of the suspensions overlap, which, together with the variability of urine, allow us to assume a nonlinear nature of the optical density dependence on the concentration of carbon dots and doxorubicin.

C. Preliminary Data Analysis

Three numbers—three parameters of the inverse problem to be determined using ML algorithms—correspond to each absorption spectrum of CD and Dox suspensions. These are the measured pH value of the suspension, the CD concentration, and the Dox concentration in the suspension (mg/L). Before the direct application of machine learning algorithms, a preliminary analysis of the resulting dataset of absorption spectra was performed.

Statistics of the dataset—the minimum (Min), maximum (Max), average (Average) values of the optical density and its standard deviation in each channel of the spectrum are shown in Fig. 3(A). These data show that the most informative bands are in the region of 200–300 nm, corresponding to the overlap of CDs, Dox, and urine absorption bands, as well as in the spectral range from 400 to 550 nm, corresponding to the absorption band of Dox.

Analyzing the covariance matrix (Fig. 3(B)), composed of pairwise covariances of the vector elements corresponding to the data array, i.e., pairwise covariances of the intensity values in the channels of the studied spectra (features), one can conclude that the features used for training are strongly correlated in a wide spectral range.

For preliminary data analysis, one performed linear PCA—the unsupervised learning algorithm for data linear transformation [4]. PCA allows us to detect orthogonal directions of variance maximum (principal components, PC) in high-dimensional data and project the data into a new subspace—the space of PC; its dimension is determined by the number of PCs. PCA is widely used as a method of reducing the dimensionality of the source data to preserve most of the information or eliminate noise.

Compressing the original dataset into a new feature subspace, it is necessary to select a subset of the eigenvectors (PC) that contain most of the information variance. In our case, the first three PCs provide 81.9% of the explained variance, which is sufficient for a preliminary analysis.

Figure 4 shows the initial set of spectra projected into the space of the first three PCs—the so-called score plot. The color indicates the examples corresponding to certain values of the target variables—parameters of the inverse problem, which have to be determined in the framework of ML algorithm application. Score plots contain information that is useful for understanding how the data is arranged. The proximity of two points means their similarity. By highlighting the value of some external parameter (A—the number of the experimental sample, B—pH, etc.), one can understand which factors most strongly affect the spectral characteristics of suspensions.

Figure 4(A) shows that the samples corresponding to the different urine solution composition can be visually divided into four groups, which indicates that this factor strongly affects the spread of properties. Milder clusterization can be carried out for cases of changing pH values (Fig. 4(B)) and Dox concentration (Fig. 4(D)). The effect of pH on the variability of the data is most likely due to changes in the composition of the urine, but the effect of the Dox concentration can be attributed to the presence of a characteristic Dox band in the absorption spectra in the region of 400–550 nm, which does not overlap with the absorption bands of CDs and urine. In the case of CD concentration influence (Fig. 4(C)) on the suspension spectra's variability, it is impossible to distinguish distinct clusters. Based on the obtained data, it can be assumed that the inverse problem will be solved more successfully for determining the Dox concentration than the CD concentration.

Note that the percentage of variance not considered by the first few PCs may contain some of the information that is essential for the considered problem. Therefore, to solve the problem, we used the intensity values normalized in a certain way in all

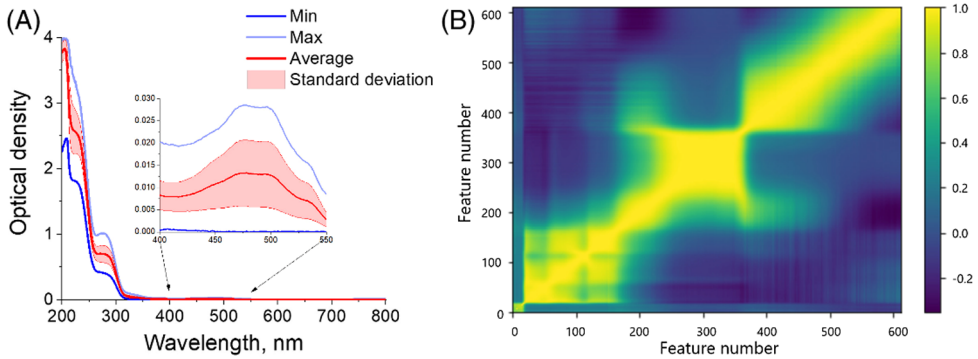

Fig. 3. (A) Channel-by-channel statistics of the dataset. (B) The covariance matrix of the dataset.

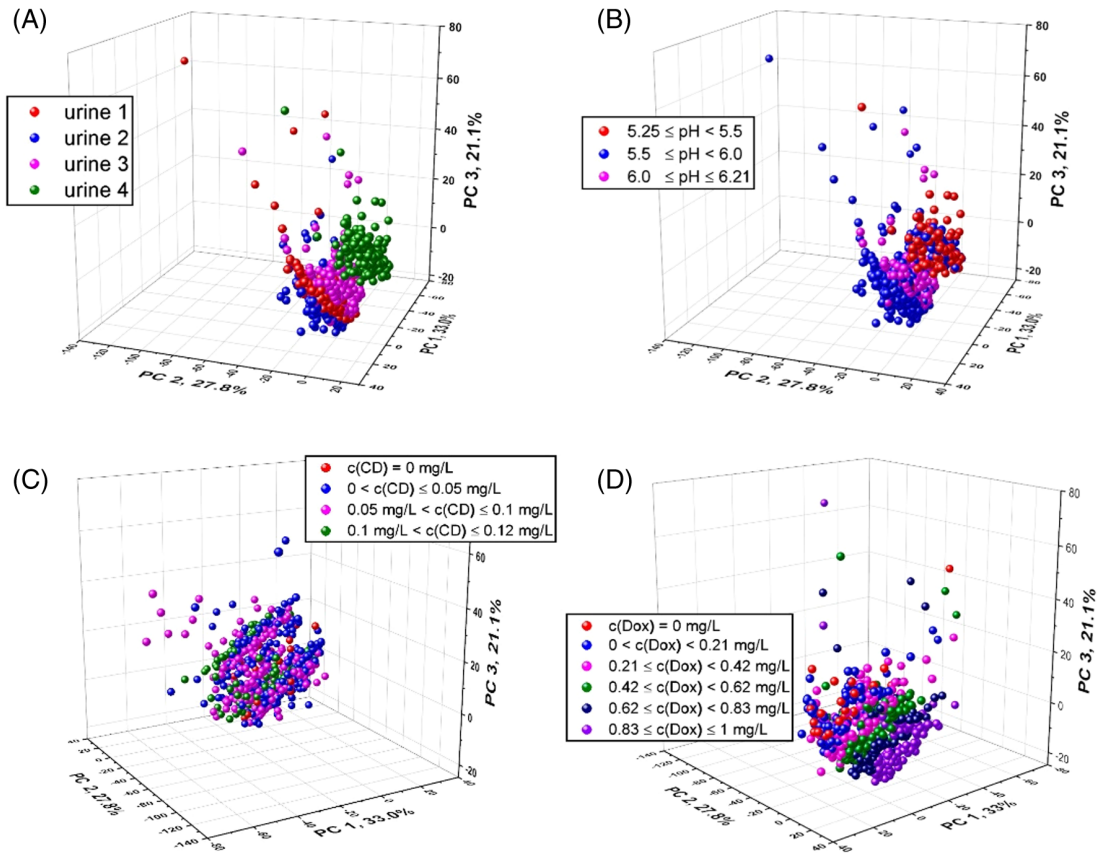

Fig. 4. PCA score plot for the first three PC: (A) urine composition; (B) pH; (C) concentration of CDs; and (D) concentration of Dox.

channels of the obtained spectra as input features, but not the values of any number of PCs.

D. Results of Machine Learning Method Application

We selected the optimal parameters of the models for all the used algorithms, except for the LR, to ensure a high-quality solution to the problem.

MLP: MLPs with different numbers of hidden layers and neurons in them were trained. Single-layer perceptrons (N01) with 8, 16, 32, 64, and 128 neurons in the hidden layer; two-layer perceptrons (N02) with $128+64$, $64+32$, $32+16$, and $16+8$; and three-layer configurations (N03) with $128+64+32$, $64+32+16$, and $32+16+8$ were considered.

According to the obtained mean absolute errors presented in Fig. SI2, the simplest MLP architecture with one hidden layer and eight neurons in it demonstrates the best result, while the group of MLPs with three hidden layers demonstrates the worst ones. Consequently, the complication of the model (MLP)—an increase of the number of network hidden layers and the number of neurons in them—leads to reduction of the inverse problem solution quality in case of CD and Dox concentration determination in the suspension. Most likely, this is due to the unfavorable ratio of the number of examples in the training set and the number of adjusted MLP weight coefficients. The values of mean absolute errors in determination of the concentration

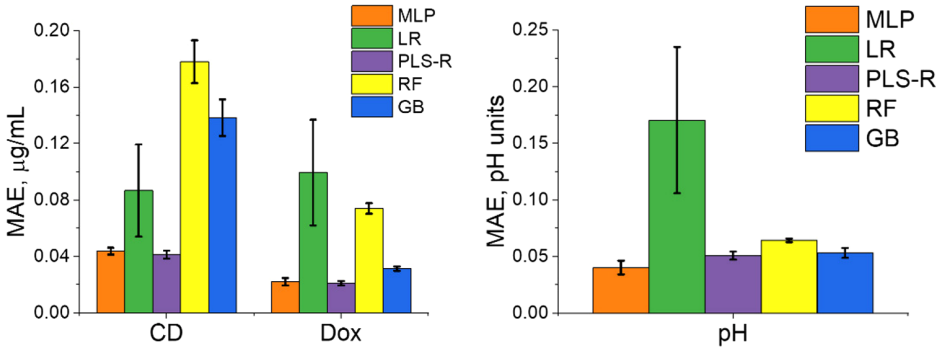

Fig. 5. Results of solving the inverse problem of optical absorption spectroscopy by different methods.

of components and the pH of suspensions do not differ much when using different architectures: the value of the standard deviation of the mean absolute error for different partitioning of data into sets is comparable to the corresponding value for different architectures. However, in absolute terms, the best result is again demonstrated by the perceptron with one hidden layer and eight neurons in it. It can be used to determine the concentration of CD and Dox with an accuracy of 43 ng/mL (3.6% of the maximum value) and 24 ng/mL (2.4% of the maximum value), respectively, and measurement of pH value with an accuracy of 0.04.

It should be noted that the error of determination of the concentration of doxorubicin is almost 2 times (1.95) less than the error of determination of the concentration of CDs. As we assumed above analyzing the PCA results, this may be since doxorubicin has an absorption band in the region of 400–550 nm, which does not overlap with the absorption bands of CDs and urine.

To compare the quality of the solution of the inverse problem by MLP with other methods, we used the architecture of the MLP N01 (8) with one hidden layer with eight neurons in it.

PLS: The optimal number of used principal components of the problem was found. For each target parameter of the problem, the number of components corresponding to the minimal mean absolute error (MAE) values for this parameter was determined. Next, the average value of the number of these components was calculated, which was subsequently used to compare the PLS method with others. The results of using PLS regression are shown in Fig. SI3. As a result, a PLS regression with 33 components was used to solve the problem.

Just like in case of ANN, the MAE of the Dox concentration determination via PLS method is 2 times less than for CDs. With PLS regression, CD and Dox concentrations can be determined with an accuracy of 41 ng/mL (3.4% of maximum value) and 21 ng/mL (2.1% of maximum value), respectively, and pH can be measured with an accuracy of 0.05.

RF: A committee of 10 trees was used. The maximum tree depth was selected as a regularization parameter—a constraint imposed on the complexity of the model to prevent its overfitting. Based on the data presented on Fig. SI4, a regressor with a maximum tree depth of 7 was chosen as a reference regressor. It can be used to determine the CD and Dox concentrations with an accuracy of 178 ng/mL (14.8% of the maximum value) and 74 ng/mL (7.4% of the maximum value), respectively, and measure the pH with an accuracy of 0.06.

GB: Since the sklearn library's GradientBoostingRegressor method does not allow us to determine multiple target parameters simultaneously, each model had single output, unlike the rest of the models used.

A model with 100 trees was used. One varied the maximum tree depth (3, 7, 15) varied to find the optimal parameters of the model. A regressor with a maximum depth of trees equal to 3 was chosen as an optimal one (Fig. SI5). The GB method allowed us to determine the CD and Dox concentrations with an accuracy of 138 ng/mL (11.5% of the maximum value) and 31 ng/mL (3.1% of the maximum value), respectively, and the pH value with an accuracy of 0.05 units.

E. Comparative Analysis of Various Machine Learning Methods' Application Efficacy

The results of the machine learning algorithms' application for solution of the three-parameter inverse problem of absorption spectroscopy to identify and determine the concentration of the CD-Dox nanocomplex components in urine and the pH value of urine with nanoparticles are shown in Fig. 5.

As follows from the obtained results, MLPs and PLS demonstrated the best results coping with the inverse problem solution. These algorithms showed almost the same results, considering the error caused by the method of splitting the data into sets. Due to the high variability of spectral data, the simplest MLPs with a single hidden layer of eight neurons turned out to be optimal. The efficiency of PLS clearly follows from Fig. 4. It is because PLS is, in fact, linear regression in the space of eigenvectors of the problem—the main components, and at the stage of application of PCA we found out that the data is well clustered in the space of the PC.

LR works poorly in problems where the dependence of responses on features is complex, nonlinear, and also with a high initial dimension of the data. Due to the high variability of the spectral data, introduced by the variability of the chemical composition of urine, and the significant overlap of the CD, Dox, and urine bands in the absorption spectra of suspensions, LR does not cope well with the solution of the inverse problem. This method makes it possible to determine the concentration

of CDs and Dox with a MAE of 87 ng/mL (7.3% of the maximum value) and 100 ng/mL (10% of the maximum value), respectively, and a pH of –0.17.

RF and GB differ from previous models in that they represent not a single approximating model but a committee of models. Since 50% of randomly selected features were used in the construction of each tree, the worse result shown by these models is not associated with cross-correlations of features or with the shape of the bands in the spectra. Decision tree models work well in situations where examples in the original feature space form isolated clusters, or "local islands." In our work, this is not the case, which is confirmed by the behavior of the MAE of determination of the concentrations of CDs and Dox. Determination errors of the concentration of CDs, provided by RF and GB, are the largest among those for all the algorithms used. At the same time, when determining the concentration of Dox, GB and RF demonstrate better results in comparison with LR. Moreover, the dependence of the MAE on the method of dividing the data into sets is less pronounced when determining the concentration of Dox, the absorption band of which does not overlap with any other band in the region of 400–550 nm (Fig. 3(A)).

4. CONCLUSION

Machine learning algorithms are a powerful tool for solving applied inverse problems of optical spectroscopy in the field of biomedicine. Their application allowed one to isolate the nanoparticles' and drugs' contribution to the absorption of optical radiation in biological medium (urine) from the background of the biological medium absorption itself. Despite the urine variability, ANN and PLS coped well with the solution of the problem of simultaneous determination of the CD and Dox concentrations, as well as pH value of the medium from the optical absorption spectra of the suspensions. Thus, using PLS, it is possible to determine the concentrations of CDs and Dox with an accuracy of 41 ng/mL (3.4% of the maximum value) and 21 ng/mL (2.1% of the maximum value), respectively, and measure pH with an accuracy of 0.05 units.

There are many different ML algorithms. To choose the model that is optimal for a specific task, it is necessary to conduct a preliminary data analysis. If the data demonstrates high variability (for example, spectra of biological media), and the spectral bands of the research objects overlap, then linear models (LR in particular) will cope poorly with the solution of this problem. Together with the plotting the spectra of the research objects, the construction of the covariance matrix for the data array allows us to make assumptions about how effectively the committees of piecewise approximation models (for example, RF and GB) will solve this problem.

PCA with the allocation of the first few PCs allows us to highlight signs that are significant for analysis, to determine the data structure—to highlight clusters of similar examples. The PCA also allows one to estimate in advance how well PLS will cope with the solution of the problem, which is determined by the common nature of both methods (PLS is the LR in the coordinated input and output spaces of the PC).

One should also pay attention to the need to increase the ratio of the number of examples in the training set and the number of input features. This can be done both by increasing the number of examples (which is not always possible) and by selecting or transforming input features. In addition, for complex models, such as ANN, the situation can be partially improved by simplifying the approximating model (reducing the number of neurons in the first hidden layer).

Funding. Russian Foundation for Basic Research (19-01-00738, 19-02-00859A); Ministry of Education and Science of the Russian Federation (FSRG 2020-0017); Foundation for the Advancement of Theoretical Physics and Mathematics (19-2-6-6-1); Interdisciplinary Scientific and Educational School of Lomonosov Moscow State University "Photonic and Quantum technologies. Digital medicine".

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. V. Bhardwaj, A. Kaushik, Z. M. Khatib, M. Nair, and A. J. McGoron, "Recalcitrant issues and new frontiers in nano-pharmacology," Front. Pharmacol. **10**, 1369 (2019).
2. H. S. Choi, W. Liu, P. Misra, E. Tanaka, J. P. Zimmer, B. I. Ipe, and J. V. Frangioni, "Renal clearance of nanoparticles," Nat. Biotechnol.**25**, 1165–1170 (2007).
3. X. Liang, H. Wang, Y. Zhu, R. Zhang, V. C. Cogger, X. Liu, and M. S. Roberts, "Short-and long-term tracking of anionic ultrasmall nanoparticles in kidney," ACS Nano **10**, 387–395 (2016).
4. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. (O'Reilly Media, 2019).
5. W. Zhang, B. Wang, and C. Zhao, "Selective thermophotovoltaic emitter with aperiodic multilayer structures designed by machine learning," ACS Appl. Energy Mater.**4**, 2004–2013 (2021).
6. A. Sakurai, K. Yada, T. Simomura, S. Ju, M. Kashiwagi, H. Okada, T. Nagao, K. Tsuda, and J. Shiomi, "Ultranarrow-band wavelength-selective thermal emission with aperiodic multilayered metamaterials designed by Bayesian optimization," ACS Cent. Sci. **5**, 319–326 (2019).
7. J. Guo, S. Ju, and J. Shiomi, "Design of a highly selective radiative cooling structure accelerated by materials informatics," Opt. Lett.**45**, 343–346 (2020).
8. Z. Huang, S. Siddhanta, G. Zheng, T. Kickler, and I. Barman, "Rapid, label-free optical spectroscopy platform for diagnosis of heparin-induced thrombocytopenia," Angew. Chem. (Int. Ed.) **59**, 5972–5978 (2020).
9. E. Guevara, J. C. Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, "Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools," Biomed. Opt. Express **9**, 4998–5010 (2018).
10. I. Olaetxea, A. Valero, E. Lopez, H. Lafuente, A. Izeta, I. Jaunarena, and A. Seifert, "Machine learning-assisted Raman spectroscopy for pH and lactate sensing in body fluids," Anal. Chem.**92**, 13888–13895 (2020).
11. K. A. Laptinskiy, S. A. Burikov, S. A. Dolenko, A. O. Efitorov, O. E. Sarmanova, O. A. Shenderova, I. I. Vlasov, and T. A. Dolenko, "Monitoring of nanodiamonds in human urine using artificial neural networks," Phys. Status Solidi A **213**, 2614–2622 (2016).
12. O. E. Sarmanova, S. A. Burikov, S. A. Dolenko, I. V. Isaev, K. A. Laptinskiy, N. Prabhakar, D. Ş. Karaman, J. M. Rosenholm, O. A. Shenderova, and T. A. Dolenko, "A method for optical imaging and monitoring of the excretion of fluorescent nanocomposites from the body using artificial neural networks," Nanomed. Nanotechnol. Biol. Med. **14**, 1371–1380 (2018).
13. P. Miao, K. Han, Y. Tang, B. Wang, T. Lin, and W. Cheng, "Recent advances in carbon nanodots: synthesis, properties and biomedical applications," Nanoscale**7**, 1586–1595 (2015).

14. N. Prabhakar, T. Nareoja, E. von Haartman, D. Ş. Karaman, S. A. Burikov, T. A. Dolenko, S. Jaikishan, V. Mamaeva, P. E. Hanninen, I. I. Vlasov, O. A. Shenderova, and J. M. Rosenholm, "Functionalization of graphene oxide nanostructures improves photoluminescence and facilitates their use as optical probes in preclinical imaging," Nanoscale **7**, 10410–10420 (2015).
15. K. Ghosal and A. Ghosh, "Carbon dots: the next generation platform for biomedical applications," Mater. Sci. Eng. C **96**, 887–903 (2019).
16. S. Haikin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. (McMaster University, 1999).
17. S. Weisberg, *Applied Linear Regression*, 3rd ed. (Wiley, 2005).
18. R. Wehrens and B. H. Mevik, "The PLS package: principal component and partial least squares regression in R," J. Stat. Softw. **18**, 1–23 (2007).
19. G. Biau and E. Scornet, "A random forest guided tour," Test **25**, 197–227 (2016).
20. V. Roshni and O. D. Praveen, "Fluorescent N-doped carbon dots from mustard seeds: one step green synthesis and its application as an effective Hg (II) sensor," Braz. J. Anal. Chem. **4**, 17–24 (2017).
21. X. Wang, J. Zhang, W. Zou, and R. Wang, "Facile synthesis of polyaniline/carbon dot nanocomposites and their application as a fluorescent probe to detect mercury," RSC Adv.**5**, 41914–41919 (2015).
22. L. Yang, W. Jiang, L. Qiu, X. Jiang, D. Zuo, D. Wang, and L. Yang, "One pot synthesis of highly luminescent polyethylene glycol anchored carbon dots functionalized with a nuclear localization signal peptide for cell nucleus imaging," Nanoscale**7**, 6104–6113 (2015).
23. Y. Sheng, J. Wei, J. Pan, P. Huang, S. Guo, J. Zhang, X. Zhang, and B. Feng, "The up-converted photoluminescence and cell imaging of water-soluble carbon dots," Chem. Phys. Lett.**638**, 196–200 (2015).
24. S. Liao, X. Zhao, F. Zhu, M. Chen, Z. Wu, X. Song, H. Yang, and X. Chen, "Novel S, N-doped carbon quantum dot-based "off-on" fluorescent sensor for silver ion and cysteine," Talanta**180**, 300–308 (2018).
25. Z. Ma, H. Ming, H. Huang, Y. Liu, and Z. Kang, "One-step ultrasonic synthesis of fluorescent N-doped carbon dots from glucose and their visible-light sensitive photocatalytic ability," New J. Chem. **36**, 861–864 (2012).