ISSN 1060-992X, Optical Memory and Neural Networks, 2024, Vol. 33, Suppl. 3, pp. S475–S482. © Allerton Press, Inc., 2024.

Solution of an Inverse Problem of Optical Spectroscopy Using Kolmogorov-Arnold Networks

G. Kupriyanov[a, b, *], I. Isaev[b, c], K. Laptinskiy[a, b], T. Dolenko[a, b], and S. Dolenko[a, b, **]

[a] *Physics Department, Moscow State University, Moscow, 119991 Russia*

[b] *Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119991 Russia*

[c] *Kotelnikov Institute of Radioengineering and Electronics, Russian Academy of Sciences, Moscow, 125009 Russia*

**e-mail: kupriyanovga@my.msu.ru*

***e-mail: dolenko@srd.sinp.msu.ru*

Received August 16, 2024; revised September 13, 2024; accepted September 30, 2024

Abstract—Kolmogorov-Arnold Networks (KAN), introduced in May 2024, are a novel type of artificial neural networks, whose abilities and properties are now being actively investigated by the machine learning community. In this study, we test application of KAN to solve an inverse problem for development of multimodal carbon luminescent nanosensors of ions dissolved in water, including heavy metal cations. We compare the results of solving this problem with four various machine learning methods—random forest, gradient boosting over decision trees, multi-layer perceptron neural networks, and KAN. Advantages and disadvantages of KAN are discussed, and it is demonstrated that KAN has high chance to become one of the algorithms most recommended for use in solving highly non-linear regression problems with moderate number of input features.

Keywords: Kolmogorov-Arnold networks, carbon nanosensors, inverse problems, fluorescence spectroscopy

DOI: 10.3103/S1060992X24700747

1. INTRODUCTION

One of the most common machine learning approaches today is an artificial neural network (NN). In the most used type of NN, multi-layer perceptron (MLP), the universal building block is taking a weighted sum of the input features with subsequent applying of a nonlinear activation function to this sum. The theoretical basis of this approach is the universal approximation theorem (1989), which proves the ability of an MLP (a fully connected NN) with a single hidden layer to approximate any continuous function of many variables with a given accuracy [1–3].

However, there is an alternative approach based on the Kolmogorov-Arnold theorem (1956). According to it, any continuous function of many variables can be represented with a given accuracy in the form of superpositions and sums of functions of a single variable [4]. Moreover, the number of such functions of one variable that is sufficient is known, and it depends on the number of the input features (1):

$$f(\mathbf{x}) = f(x_1,\ldots,x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right). \quad (1)$$

Despite the similarity of the models of MLP and KAN, there is a significant difference between them: while the trained parameters of a perceptron are the continuous values of the weights, in KAN it is necessary to select continuous activation functions of one variable [5]. KAN are described in more detail in Section 2.

Despite the recent emergence of KAN, its authors have already published an open source software in Python for working with KAN [6], which resulted in a set of studies using KAN in various applied problems.

Based on KAN, Abueidda et al. [7] improved the algorithm for solving differential equations using DeepONET. The improved algorithm (DeepOKAN) turned out to be more effective than DeepONET with an equal number of trainable parameters. In addition, DeepOKAN surpassed the asymptotic accu-

racy of DeepONET, which is a great success and makes the approach to solving differential equations using KAN a promising direction.

The authors of [8] proposed a method combining KAN and recurrent NN for time series prediction. The new method outperformed traditional LSTM and GRU models in predicting long-term periods and gave more stable results on historical market data. The authors see great potential for the further development of recurrent NN models based on KAN.

Another study comparing the performance of KAN and MLP on the task of predicting satellite traffic also indicates the superiority of KAN in predictive ability [9].

Bozorgasi and Chen explore the ability of KAN to identify features and patterns in problems with a large space of input data [10]. Using a combination of KAN and wavelet transform with activation functions in the form of the second derivative of the Gaussian form and the Mexican hat, the advantage of the proposed model compared to standard wavelet NN was demonstrated on the MNIST problem.

Also, KAN can be used in large language models [11]. However, this task requires large computational and time resources, and so far there has been little meaningful research in this direction.

In this study, KAN were used to develop multimodal carbon luminescent nanosensors of ions dissolved in water, including heavy metal cations. The operating principle of luminescent nanosensors is based on the dependence of the luminescent properties of nanoparticles on environmental parameters or on the type and concentration of certain substances in the environment [12]. One of the most promising nanoparticles for nanosensing are carbon dots (CDs), which have stable, intense luminescence with a maximum in the range from UV to IR [13]. The quantum yield of luminescence of these nanoparticles can reach 90% or more [14, 15]. In addition, carbon nanomaterials are easy to synthesize, their surface can be modified in accordance with the specific problem being solved, and they are the most biocompatible and non-toxic even among other carbon nanomaterials [16, 17]. This set of properties of CDs makes it possible for them to be widely used in various fields, including nanosensors [12, 14, 16].

According to the results of many authors, the luminescence of carbon dots is very sensitive to the type and content of certain ions. This manifests itself in changes in the luminescence intensity and other characteristics of the luminescence spectrum in the presence of these ions [18–21]. This influence of the ions surrounding the nanoparticle on its luminescence is used to create a nanosensor. Thus, in [22] authors showed that the luminescence of carbon dots synthesized by the hydrothermal method and doped with boron and phosphorus atoms selectively "reacts" to the presence of Cu^{2+} and Fe^{3+} ions in water, but does not "feel" the presence of other ions. The authors suggested that it is the doping of nanoparticles with boron and phosphorus that changes the composition of chromophores on the surface of the CDs accordingly. Having obtained calibration dependencies of the luminescence intensity of CDs on the ion concentration, the authors showed that CDs can determine the concentration of Cu^{2+} or Fe^{3+} ions in living cells with an accuracy of 0.18 and 0.27 μM, respectively. A nanosensor based on carbon dots synthesized by the hydrothermal method and doped with nitrogen, sulfur and oxygen atoms for measuring the concentration of Cu^{2+} ions in water was proposed in the article [23]. The linear dependence of the luminescence intensity of CDs on the concentration of Cu^{2+} ions in the range of its variation from 10 to 33.3 μM, obtained by the authors, provided a detection limit of Cu^{2+} in water of 2 μM. The authors of [24] discovered the high sensitivity of the luminescence of carbon dots obtained by hydrothermal synthesis and doped with nitrogen, boron and sulfur atoms to the changes in the concentration of Fe^{3+} ions in water. The developed nanosensor is based on the quenching of luminescence by these ions makes it possible to detect Fe^{3+} ions in the range of changes of their concentration of 0.3–546 μM with an accuracy of 90 nM.

It should be noted that all articles on the development of luminescent carbon nanosensors propose single modal sensors, i.e. sensors capable of determining the concentration of one, maximum, two dissolved ions using calibration dependencies. However, at present, many problems in biomedicine, ecology, and the food industry require simultaneous monitoring of a number of ions with high accuracy in determining the concentration of each of them in real time. Artificial NN are successfully used to solve such multiparameter inverse problems, including in the field of optical spectroscopy. For example, in [25], the content of nutrients, metabolites and viable cell concentration in culture media of BHK-21 cell line was simultaneously determined by MLP using UV-VIS absorption spectra of phenol red. In order to increase the accuracy of diagnosing melanoma and brain cancer, the authors of [26] processed the Raman spectra of biological tissue using MLP.

The authors of this manuscript solved the problem of controlling the removal of theranostic fluorescent nanocomposites and their components from the body [27, 28], and developed CD-based nanosensors for simultaneous measurement of environmental parameters such as pH and temperature [29], to

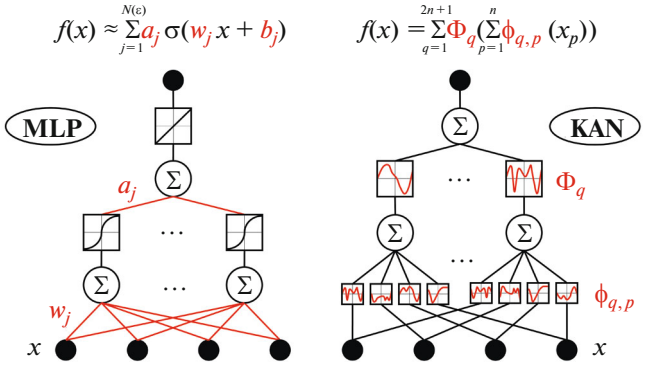


Fig. 1. Comparison of MLP and KAN. The entities that are subject to adjustment during NN training, are marked red (weights for MLP and activation functions for KAN).

determine dissolved ions [30] using the combination of fluorescent spectroscopy and NN. To solve all these problems, either MLP or convolutional NN were used.

In this study, KAN and other machine learning algorithms were used to develop a nanosensor based on CDs for simultaneous determination of the concentration of each of the 4 ions in water from the luminescence spectra of nanoparticles.

2. KOLMOGOROV-ARNOLD NETWORKS

As stated above, despite the similarity of the models of MLP and KAN, there is a significant difference between them: while the trained parameters of MLP are the continuous values of the weights, in KAN it is necessary to select continuous activation functions of one variable (Fig. 1).

The problem of parameterization of activation functions in KAN is solved by using various bases: trigonometric functions, polynomial functions, splines. Particularly successful was the B-spline approach implemented by Liu et al. [5]. Within this approach, the functions are not selected from some pre-defined set, but they are composed of splines, and their shapes are optimized based on the change of their amplitudes in some control points; the approximation by splines is necessary in order both to perform a smooth approximation and to obtain derivatives of the resulting functions for the implementation of gradient-based training methods.

Thus, the properties of B-splines are:

(1) Controlled smoothness of splines;

(2) Controlled number of control points (grids). Each activation function is parameterized by a linear combination of splines whose coefficients are trained;

(3) Locality of splines.

The first property allows B-splines to effectively approximate analytic functions. For example, KAN is effective in approximating physical functions of the Feynman set (Fig. 2). This can be an extremely useful property when solving both applied physical problems and when modeling various processes: from solving partial differential equations to modeling quantum mechanics processes.

The second and third properties allow seamlessly adding new degrees of freedom to the KAN model during the learning process, adaptively increasing its nonlinearity.

After KAN training, it is possible to approximate the obtained activation functions with analytical functions, which allows one to interpret the decision-making process in KAN.

3. PHYSICAL EXPERIMENT

The results of the physical laboratory experiment described in our previous publication [30] were taken as the dataset used in this study.

Aqueous suspensions of CD with fixed concentration (8 mg/L) containing heavy metal ions Cu^{2+}, Ni^{2+}, Cr^{3+} were prepared with concentrations of the ions varying from 0 to 4.95 mM with 0.55 mM increment. To obtain such suspensions, we added $Cu(NO_3)_2$, $Ni(NO_3)_2$, $Cr(NO_3)_3$ salts to the CD-suspension

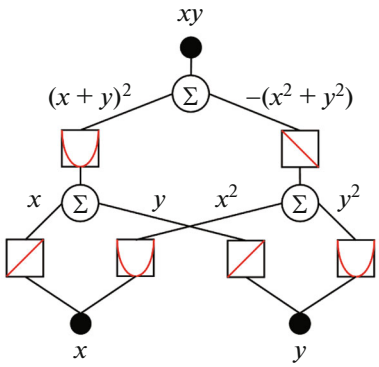


Fig. 2. An example of approximations of KAN functions from the Feynman set.

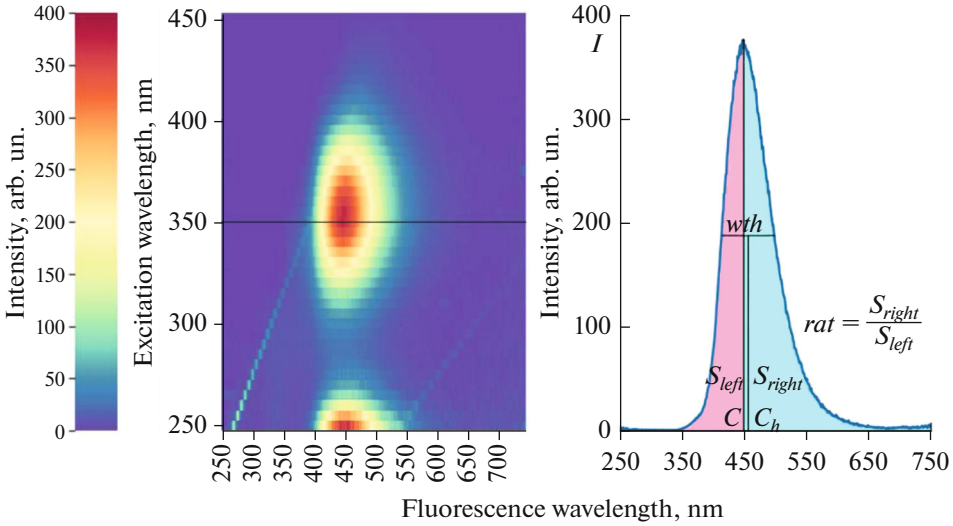


Fig. 3. Left: 2D excitation-emission matrix of a CD aqueous suspension in presence of heavy metal ions. Right—1D fluorescence spectrum at 350 nm excitation and its parameterization with 5 parameters.

in such quantities that Cu^{2+}, Ni^{2+}, Cr^{3+} cations were present in the suspensions in the desired concentrations. A total of 1000 suspensions were prepared.

Shimadzu RF-6000 spectrofluorimeter registered 2D excitation-emission matrices (EEM) of CD aqueous suspensions in presence of the heavy metal ions. The fluorescence signal was excited by the radiation with wavelengths ranging from 250 to 450 nm with 5 nm increment—at 41 excitation wavelengths. Each 1D fluorescence spectrum was registered in the range from 250 to 750 nm with 1 nm increment—in 500 spectral channels. An example of the EEM of a CD aqueous suspension is presented in Fig. 3 (left). At the right is the 1D fluorescence spectrum at 350 nm excitation wavelength, corresponding to the maximum of the EEM, and the demonstration of its parameterization with 5 parameters (explained below).

4. COMPUTATIONAL EXPERIMENT

4.1. Data

In this study, the 1D spectrum at the excitation wavelength of 350 nm was used to predict the concentrations of the metal cations Cu^{2+}, Ni^{2+}, Cr^{3+} and of the NO^{3-} anion. The total number of patterns was 1000.

Two representations of the input data were assumed (Fig. 3, right):

- Full representation—500 intensity values (the entire spectrum)
- Compressed representation—5 synthesized features:

Table 1. Hyperparameters of MLP and KAN used in experiment

Parameter	MLP	KAN
Hidden layer size	64	1
Activation functions	tanh	Tuning with B-splines + SiLU in res. connection
Optimizer	Adam	LBFGS
Learning rate	0.001	1.0
Patience	100	10
Tolerance	0.001	0.001
Number of learnable parameters:		
for 5 input features	449	48
for 500 input features	32 129	4008

(1) Maximum spectrum intensity (**I**);

(2) Coordinate of maximum intensity (**C**);

(3) Coordinate of the spectrum center at half-height from the maximum intensity (**Ch**);

(4) Spectral width at half maximum (**wth**);

(5) The ratio of the areas under the curve to the right of the maximum intensity and to the left of it (**rat**).

4.2. Statement of the Computational Experiment

MLP with 16 neurons in the single hidden layer, gradient boosting over decision trees (GB), and random forest (RF) were used as reference models, implemented in scikit-learn library [31]. KAN (pykan library, [6]) had one hidden layer with one neuron.

Data was randomly divided into train/validate/test subsets with ratios 0.7/0.2/0.1. Hyperparameters of MLP and KAN are presented in Table 1. Gradient boosting and random forest were used with default hyperparameters. The program code for the experiment is available at the link [32].

5. RESULTS

The results of solving the problem are presented in Fig. 4. The error bars represent the variance among the results of several runs of each method with various initializations.

A schematic representation of the KAN model for the 5-input variant is shown in Fig. 5. After training the KAN with five input features, the model was reduced to an analytical formula presented at the bottom of Fig. 5.

As can be seen from Fig. 4, KAN performed significantly better than the MLP on the full feature set, which indicates the ability of KAN to effectively extract information from a large set of highly correlated features, possibly due to its ability to build flexible non-linear approximations. However, the computational cost of building a KAN model is much higher than that for models based on decision trees (RF and GB), with a very small difference with them in the results (within the statistical variance). Therefore, for a large input dimension of a problem of the kind considered here, KAN is not the model that should be preferred to achieve the minimal error.

On the compressed set of 5 inputs, KAN showed the best result among all the four machine learning methods solving the problem, with an acceptable computational cost; however, the advantage is not statistically significant.

At the same time, for a small number of input features, the KAN model is easily reduced to an analytical expression that allows one to estimate not only the relative significance of the input features, but also the dependence of the target on each of the input features. This is an obvious advantage of KAN against all the other machine learning methods considered in this study (MLP, RF, GB). Such analytical representation is easily implemented on any computing device, can take into account the properties and limitations of the solving problem while choosing approximation function and can be useful for further analysis.

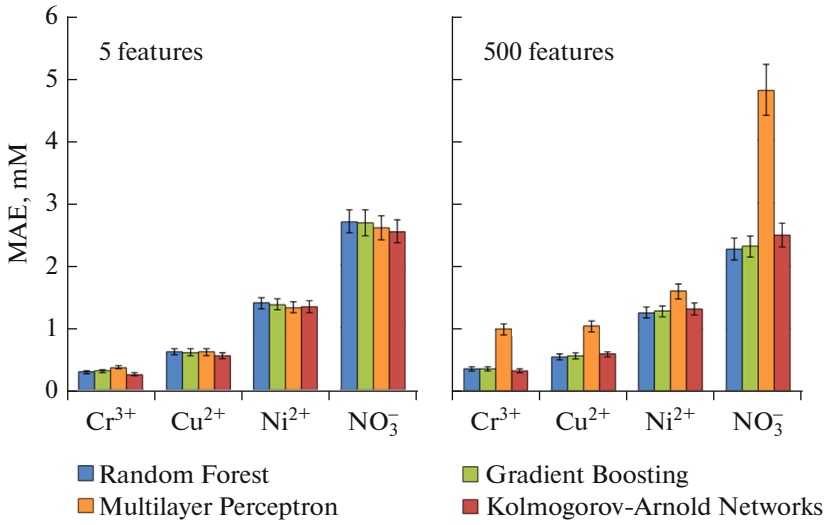
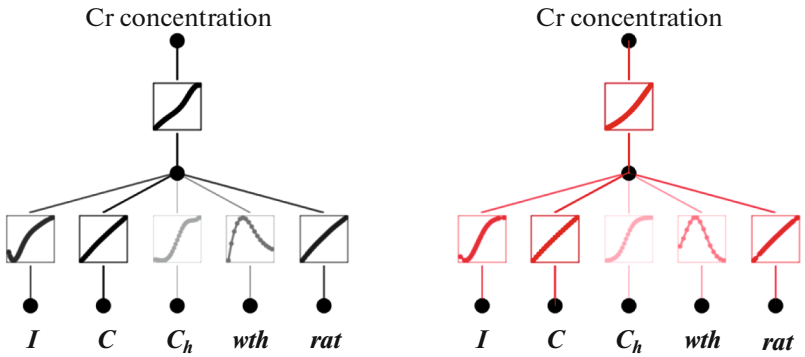


Fig. 4. Mean absolute error of determination of ion concentrations by the four methods based on the compressed representation (left) and on the full representation (right).



$$Conc(Cr^{3+}) = 5.02\ \mathrm{atan}(-4.51\ (1 - 0.033\ \boldsymbol{rat})^4 + 7.16\ (0.11\ \boldsymbol{C} - 9.74) + 7.16\ (0.11\ \boldsymbol{C} - 9.74) + 0.13\ \tanh(1.4\ \boldsymbol{C_h} + 0.21) + 6.82 + 0.12\ e^{-0.59(-\boldsymbol{wth} - 0.65)^2} - 0.91\ e^{-0.55(-0.75\boldsymbol{I} - 1)^2}) + 4.84$$

Fig. 5. KAN with five input features for determination of concentration of the Cr^{3+} ion: before reduction to analytical form (left) and after reduction (right). Below is the analytical formula of the reduced KAN.

6. CONCLUSIONS

Based on the results of the study, the following conclusions can be made:

• Kolmogorov-Arnold networks (KAN) are a promising competitor to fully connected neural networks. Its possible advantages are: high non-linearity; adjustable complexity of the model during the training process; ability to reduce the solution to an analytical formula. The main disadvantages of KAN are: complexity of learning, as a consequence of the nonlinearity of KAN, and its high computational cost. As the result, KAN may be recommended for use with a low or moderate number of input features.

• In the solution of the inverse problem for development of multimodal carbon luminescent nanosensors of ions dissolved in water (including heavy metal cations), based on the sample fluorescence spectrum, KAN significantly outperformed the multi-layer perceptron for the case with full input feature set. For the compressed representation of the input data, it slightly outperformed all the three competitor machine learning models (multi-layer perceptron, random forest, and gradient boosting). KAN also made it possible to formulate the model in the form of a relatively simple analytical formula.

FUNDING

This work was supported by the Russian Science Foundation, grant no. 22-12-00138, https://rscf.ru/en/project/22-12-00138/.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Hornick, K., Stinchcombe, M., and White, H., Multilayer feedforward networks are universal approximators, *Neural Networks,* 1989, vol. 2, no. 5, pp. 359–366. https://doi.org/10.1016/0893-6080(89)90020-8
2. Cybenko, G., Approximation by superpositions of a sigmoidal function, *Math. Control, Signals, Systems,* 1989, vol. 2, no. 4, pp. 303–314. https://doi.org/10.1007/BF02551274
3. Funahashi, K., On the approximate realization of continuous mappings by neural networks, *Neural Networks,* 1989, vol. 2, no. 3, pp. 183–191. https://doi.org/10.1016/0893-6080(89)90003-8
4. Kolmogorov, A.N., On representation of continuous functions of several variables by superpositions of continuous functions of a less number of variables, *Izv. Akad. Nauk SSSR,* 1956, vol. 108, pp. 179–182; *Am. Math. Soc. Transl.*, 1961, vol. 17, pp. 369–373.
5. Liu, Z. et al., KAN: Kolmogorov-Arnold Networks. arXiv:2404.19756v4 (2024). https://doi.org/10.48550/arXiv.2404.19756
6. Liu, Z., pykan: Kolmogorov Arnold Networks (KANs), https://github.com/KindXiaoming/pykan. Last accessed July 5, 2024.
7. Abueidda, D.W., Pantidis, P., and Mobasher, M.E., DeepOKAN: Deep Operator Network Based on Kolmogorov Arnold Networks for Mechanics Problems. arXiv:2405.19143 (2024) https://doi.org/10.48550/arXiv.2405.19143
8. Genet, R. and Inzirillo, H., *TKAN: Temporal Kolmogorov-Arnold Networks*, *SSRN Elec. J.,* 2024. https://doi.org/10.2139/ssrn.4825654
9. Vaca-Rubio, C.J. et al., Kolmogorov-Arnold Networks (KANs) for Time Series Analysis. arXiv:2405.08790 (2024). https://doi.org/10.48550/arXiv.2405.08790
10. Bozorgasi, Z. and Chen, H., Wav-KAN: Wavelet Kolmogorov-Arnold Networks. arXiv:2405.12832v2 (2024). https://doi.org/10.48550/arXiv.2405.12832
11. Galitsky, B.A., Kolmogorov-Arnold network for word-level explainable meaning representation. *Preprints*, 2024051981 (2024). https://doi.org/10.20944/preprints202405.1981.v1
12. Darwish, M.A. et al., *Advancements in nanomaterials for nanosensors: A comprehensive review*, *Nanoscale Adv.,* 2024. https://doi.org/10.1039/D4NA00214H
13. Tian, Z. et al: Full-color inorganic carbon dot phosphors for white-light-emitting diodes, *Adv. Opt. Mater.,* 2017, vol. 5, 1700416.
14. Mansuriya, B.D. and Altintas, Z., Carbon Dots: classification, properties, synthesis, characterization, and applications in health care—An updated review (2018–2021), *Nanomaterials,* 2021, vol. 11, no. 10.
15. Khmeleva, M.Yu., Laptinskiy, K.A., and Dolenko, T.A., The influence of pH on the properties of carbon dots with different surface functionalization: sizes and photoluminescence quantum yield, *Opt. Spectrosc.,* 2023, vol. 131, no. 6, pp. 752–759.
16. El-Shafey, A.M., Carbon dots: Discovery, structure, fluorescent properties, and applications, *Green Process. Synthesis,* 2021, vol. 10, no. 1, pp. 134–156.
17. Adeola, A.O. et al., Advances in the design and use of carbon dots for analytical and biomedical applications, *Nanotechnology,* 2023, vol. 35, 012001.
18. Batool, M. et al., Metal ion detection by carbon dots—A review, *Crit. Rev. Anal. Chem.,* 2020, vol. 52, pp. 756–767.
19. Vervald, A.M. et al., Quenching of photoluminescence of carbon dots by metal cations in water: Estimation of contributions of different mechanisms, *J. Phys. Chem. C,* 2023,vol. 127, pp. 21617–21628.
20. Omar, N.A.S. et al., A Review on carbon dots: Synthesis, characterization and its application in optical sensor for environmental monitoring, *Nanomaterials,* 2022, vol. 12, p. 2365.

21. Liu, M.L. et al., Carbon dots: synthesis, formation mechanism, fluorescence origin and sensing applications, *Green Chem.,* 2019, vol. 21, pp. 449–471.
22. Sonaimuthu, M. et al., Multiple heteroatom dopant carbon dots as a novel photoluminescent probe for the sensitive detection of Cu^{2+} and Fe^{3+} ions in living cells and environmental sample analysis, *Environ. Res.,* 2023, vol. 219, p. 115106.
23. Algethami, F.K. and Abdelhamid, H.N., Heteroatoms-doped carbon dots as dual probes for heavy metal detection, *Talanta,* 2024, vol. 273, 125893.
24. Liu, Y. et al., Red emission B, N, S-co-doped carbon dots for colorimetric and fluorescent dual mode detection of Fe^{3+} ions in complex biological fluids and living cells, *ACS Appl. Mater. Interfaces,* 2017, vol. 9, pp. 12663–12672.
25. Takahashi, M.B. et al., Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes, *Bioprocess Biosyst. Eng.,* 2015, vol. 38, pp. 1045–1054.
26. Jermyn, M. et al., Neural networks improve brain cancer detection with Raman spectroscopy in the presence of operating room light artifacts, *J. Biomed. Opt.,* 2016, vol. 21, 094002.
27. Sarmanova, O.E. et al., Applications of fluorescence spectroscopy and machine learning methods for monitoring of elimination of carbon nanoagents from the body, *Opt. Memory Neural Networks,* 2023, vol. 32, pp. 20–33.
28. Sarmanova, O.E. et al., A method for optical imaging and monitoring of the excretion of fluorescent nanocomposites from the body using artificial neural networks, *Nanomed.: Nanotechnol., Biol. Med.,* 2018, vol. 14, pp. 1371–1380.
29. Sarmanova, O.E. et al., Development of the fluorescent carbon nanosensor for pH and temperature of liquid media with artificial neural networks, *Spectrochim. Acta, Part A,* 2021, vol. 258, 119861.
30. Sarmanova, O.E. et al., Implementing neural network approach to create carbon-based optical nanosensor of heavy metal ions in liquid media, *Spectrochim. Acta, Part A,* 2023, vol. 286, 122003.
31. Scikit-learn—Machine Learning in Python: https://scikit-learn.org. Last accessed July 5, 2024.
32. https://github.com/Gavr101/KAN_spectroscopy, last accessed 2024/07/05.