Mathematical Model Coupled to Neural Networks Calculates the Extraction Recovery of Polycyclic Aromatic Hydrocarbons in Problematic Matrix

Tomas Drevinskas, Audrius Maruška,* Kristina Bimbiraitė-Survilienė, Gediminas Dūda,
Mantas Stankevičius, Nicola Tiso, Rūta Mickienė, Vilmantas Pedišius, Donatas Levišauskas,
Vilma Kaškonienė, Ona Ragažinskienė, Saulius Grigiškis, Enrica Donati, and Massimo Zacchini

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Unknown extraction recovery from solid matrix samples leads to meaningless chemical analysis results. It cannot always be determined, and it depends on the complexity of the matrix and properties of the extracted substances. This paper combines a mathematical model with the machine learning method—neural networks that predict liquid extraction recovery from solid matrices. The prediction of the three-stage extraction recovery of polycyclic aromatic hydrocarbons from a wooden railway sleeper matrix is demonstrated. Calculation of the extraction recovery requires the extract's volume to be measured and the polycyclic aromatic hydrocarbons' concentration to be determined for each stage. These data are used to calculate the input values for a neural network model. Lowest mean-squared error (0.014) and smallest retraining relative standard deviation (20.7%) were achieved with the neural network setup 6:5:5:4:1 (six inputs, three hidden layers with five, five, and four neurons in a layer, and one output). To train such a neural network, it took less than 8000 steps—less than a second——using an average-performance laptop. The relative standard deviation of the extraction recovery predictions ranged between 1.13 and 5.15%. The three-stage recovery of the extracted dry sample showed 104% of three different polycyclic aromatic hydrocarbons. The extracted wet sample recovery was 71, 98, and 55% for phenanthrene, anthracene, and pyrene, respectively. This method is applicable in the environmental, food processing, pharmaceutical, biochemical, biotechnology, and space research areas where extraction should be performed autonomously without human interference.
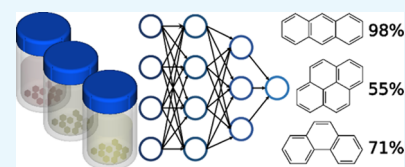
1. INTRODUCTION

The treatment and disposal of used railway sleepers is a problem worldwide.[1] The regulations do not allow the disposal of railway sleepers as a regular waste due to the presence of polycyclic aromatic hydrocarbons (PAHs), which can cause cancer to humans.[2] There are three main ways of disposing off the used railway sleepers: (a) storing in dedicated hazardous material storage areas, (b) high-temperature burning, and (c) bioremediation. Usually, high-temperature burning is avoided due to the high costs, and storing railway sleepers causes pollution of the nearby environment. The bioremediation technological process is still under research and development.[3,4] Bioremediation of railway sleepers on a miniaturized scale has already been demonstrated in a laboratory. However, the high-scale bioremediation technological process is still a challenge.[3,5] PAH concentration monitoring is necessary to develop a high-scale bioremediation technology, and this can only be done with the aid of analytical techniques. Mainly, high-performance liquid chromatography (HPLC), ultra-high-performance liquid chromatography (UPLC), or gas chromatography is used for determining PAH.[5−9] As demonstrated in previous studies, PAH distribution over the different parts of railway sleepers varies.[10]

The uneven distribution requires more sample volume or mass for extraction to achieve a repeatable extraction process. This requirement leads to higher solvent volumes needed for extending the costs of analytical procedures significantly. Besides, the extracts of railway sleepers also have to be disposed off properly. Another issue related to the extraction of PAH is the different and irreproducible recovery using other extraction methods. *Soxhlet* extraction is considered one of the etalons in the PAH extraction process. However, this method is practically inapplicable in the monitoring of multiple samples during the optimization process.[9,11] Maceration extraction using solvents is practiced more often because it is simple. Multiple samples can be extracted in parallel flasks or extraction bottles; however, the method's reproducibility fails dramatically if samples of different humidities and different constitutions are used for extraction.[5] Other methods such as

solid-phase micro-extraction or supercritical fluid extraction can achieve high reproducibility. On the other hand, some PAHs that are less volatile, such as pyrene and fluorene, are extracted in minute recoveries. Currently, machine learning, artificial intelligence, data ordering, and related methods are being applied for various chemical analysis cases. A technique utilizing the segmentation tree approach has been developed and applied to determine the peak in the chromatogram which is responsible for antiviral activity.[12] Tasks such as classification using neural networks of rapeseed oil have been performed.[13] Even improvement of peak properties in the electrophero-grams has been utilized.[14] A machine learning method—neural networks operating in multidimensional space——can be applied for various cases, including solving chemical analysis problems.[15,16] Extraction recovery, machine learning, or artificial intelligence methods are considered for the improve-ment of the process. Multivariate regression analysis and neural networks have been applied to predict the physicochemical properties of medicinal plant extracts.[17] Neural networks or response surface methodology have been used to improve the bioactive compounds' extraction process from medicinal plants.[18−21] On the other hand, none of these methods can predict the extraction recovery, which is demonstrated in this paper. Neural networks is a machine learning method, which belongs to the group of computerized methodologies which provide an opportunity to solve the problems which were not solved using common classical theories and means. Modern analytical chemistry cannot make any progress without the application of new technologies in integration, miniaturization of analysis, and analytical tools including also information technologies.

The failure to monitor PAH contents properly is probably the main reason why no high-efficiency PAH bioremediation technological process has been developed yet.

This work aimed to develop a mathematical multistage extraction model of PAHs that can be used with machine learning methods for determining the extraction recovery.

2. RESULTS AND DISCUSSION

2.1. Development of Mathematical Extraction Model.
In the typical machine learning methods, supervised and unsupervised approaches are used. The developed method is considered as a supervised, expert (chemical analysis special-ist)-assisted technique. The following mathematical model has been developed by carefully observing the multistage extraction procedure. Assume that the railway sleeper content that is being extracted contains x_1 amount of selected PAH. The amount here can be a dimensionless number or an actual dimension such as mass. After the extraction process, the amount of PAH can be expressed as eq 1

$$x_1 = r_1 x_1 + (1 - r_1)x_1 \tag{1}$$

where x_1 is the initial amount of selected PAH and r_1 is the recovery (between 0 and 1) of the first extraction stage. Here, $r_1 x_1$ can be defined as the amount of PAH in the solvent after the first stage extraction, and $(1 - r_1)x_1$ can be defined as that which is not extracted and left in the railway sleeper pieces and the rest of the matrix. After decanting and measuring the extract in the measurement cylinder, it is observed that some part of the solvent has soaked the extraction content (railway sleeper pieces and the rest of the matrix). Therefore, a less volume of the extract is decanted after the first extraction stage

than the solvent volume before extraction. The expression $r_1 x_1$ can be reformatted as eq 2

$$r_1 x_1 = v_1 r_1 x_1 + (1 - v_1)r_1 x_1 \tag{2}$$

where $r_1 x_1$ is the amount of PAH in the solvent after the first stage extraction, v_1 is the ratio of the volume between the collected extract and the added extrahent (a dimensionless number between 0 and 1), and $1 - v_1$ is the fraction of the volume that soaked the extraction material and was not retained. Such observations suggest that eq 1 must be refined. Assume that the first stage extract has been transported to the storage bottle, so that the initial content can be expressed as (eq 3)

$$x_1 = v_1 r_1 x_1 + (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \tag{3}$$

where x_1 is the amount of the selected PAH amount before the first stage extraction, r_1 is the first stage recovery (between 0 and 1), and v_1 is the ratio between the volume of the added solvent and the volume of the decanted extract. Therefore, $v_1 r_1 x_1$ is the amount of PAH stored in the bottle after decanting the first stage extract. The volume that soaked the extraction content but not decanted into the storage bottle is $1 - v_1$; therefore, $(1 - v_1)r_1 x_1$ is the amount of PAH extracted but not decanted into the storage bottle due to the soaking of the extraction content. The amount that is not extracted after the first stage and left in the sleeper pieces is $(1 - r_1)x_1$. The $v_1 r_1 x_1$ amount can be measured by analytical methods such as UPLC. Assume that C_1 is the concentration determined by UPLC in the first stage extract, and it relates to $v_1 r_1 x_1$ by eq 4

$$C_1 V_1 = u_1 = v_1 r_1 x_1 \tag{4}$$

where C_1 is the concentration determined by UPLC, V_1 is the volume of the extract, u_1 is the amount of PAH that has been determined, and $v_1 r_1 x_1$ is the amount of PAH stored in the bottle after decanting the first stage extract.

Proceeding to the second extraction stage, assume that the amount $(1 - v_1)r_1 x_1 + (1 - r_1)x_1$ left in the extraction bottle after decanting the extract into the storage bottle is assigned to x_2—the second stage amount that is to be extracted. The extraction at the second stage can be expressed by eq 5

$$x_2 = r_2 x_2 + (1 - r_2)x_2,$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \tag{5}$$

where x_2 is the total amount of PAH that is being extracted at the second stage. After decanting and measuring the extract, the volume of the extract is still lower than that of the solvent added. Assume that the second stage extract has been transported to the storage bottle; so, eq 4 requires adjustment, and the initial content can be expressed by eq 6

$$x_2 = v_2 r_2 x_2 + (1 - v_2)r_2 x_2 + (1 - r_2)x_2,$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \tag{6}$$

where x_2 is the total amount of PAH extracted at the second stage, v_2 is the ratio between the volume of the added solvent and the decanted extract volume at the second stage extraction, and r_2 is the recovery of the second stage. The concentration C_2 measured by UPLC in the second stage extract is related to the content $v_2 r_2 x_2$ by eq 7

$$C_2 V_2 = u_2 = v_2 r_2 x_2 \tag{7}$$

where C_2 is the concentration of the second stage extract measured by UPLC, V_2 is the volume of the decanted second stage extract, u_2 is the amount of PAH that has been determined in the second stage extract, and $v_2 r_2 x_2$ is the amount of PAH that is stored in the bottle after decanting the second stage extract.

Proceeding to the third extraction stage, assume that the amount which was neither decanted nor extracted and left in the extraction bottle, $(1 - v_2)r_2 x_2 + (1 - r_2)x_2$, after decanting the extract into the storage bottle is assigned to x_3—the third stage amount that is to be extracted. The extraction at the third stage can be expressed by eq 8

$$x_3 = v_3 r_3 x_3 + (1 - v_3)r_3 x_3 + (1 - r_3)x_3,$$
$$x_3 = (1 - v)r_2 x_2 + (1 - r_2)x_2,$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \quad (8)$$

where x_3 is the total amount of PAH that is being extracted in the third stage, v_3 is the ratio between the volume of the added solvent and the volume of the decanted extract in the third stage extraction, and r_3 is the recovery of the third stage. The concentration of PAH in the third stage extract can be determined and is related to the content $v_3 r_3 x_3$ by eq 9

$$C_3 V_3 = u_3 = v_3 r_3 x_3 \quad (9)$$

where C_3 is the concentration of the determined PAH in the third stage extract, V_3 is the volume of the decanted third stage extract, and u_3 is the amount of PAH that has been determined in the third stage extract.

As this research aims to find the total extraction recovery in a multistage extract, full recovery after the three stages of extraction—r_{3tot}—can be expressed by eq 10

$$r_{3tot} = \frac{v_1 r_1 x_1 + v_2 r_2 x_2 + v_3 r_3 x_3 + (1 - v_3)r_3 x_3}{x_1},$$
$$x_3 = (1 - v)r_2 x_2 + (1 - r_2)x_2,$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \quad (10)$$

where r_{3tot} is the total extraction recovery after the three stages, x_1 is the amount of selected PAH before the first stage extraction (x_2 and x_3—second and third stages, correspondingly), r_1 is the first stage (r_2 and r_3—second and third stages, correspondingly) recovery (between 0 and 1), and v_1 is the ratio between the volume of the added solvent and volume of the decanted extract in the first extraction stage (v_2 and v_3—second and third stages, correspondingly).

The component $(1 - v_3)r_3 x_3$ is not retained after decanting the third stage extract and is left in the extraction bottle. Therefore, it is meaningful to calculate the apparent recovery for the three-stage extraction (eq 11)

$$r_{3a} = \frac{v_1 r_1 x_1 + v_2 r_2 x_2 + v_3 r_3 x_3}{x_1},$$
$$x_3 = (1 - v)r_2 x_2 + (1 - r_2)x_2,$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \quad (11)$$

The expression for the two-stage apparent extraction recovery is represented in eq 12

$$r_{2a} = \frac{v_1 r_1 x_1 + v_2 r_2 x_2}{x_1},$$
$$x_2 = (1 - v_1)r_1 x_1 + (1 - r_1)x_1 \quad (12)$$

where r_{2a} is the two-stage and r_{3a} is the three-stage apparent extraction recovery, x_1 is the amount of selected PAH before the first stage extraction (x_2 and x_3—second and third stages, correspondingly), r_1 is the first stage (r_2 and r_3—second and third stages, correspondingly) recovery (between 0 and 1), and v_1 is the ratio between the volume of the added solvent and volume of the decanted extract in the first extraction stage (v_2 and v_3—second and third stages, correspondingly).

2.2. Investigation of Relations between the Variables of the Developed Model. The developed mathematical model is intuitive and straightforward. On the other hand, finding an apparent recovery from a determined concentration in different stage extracts in a multistage extraction process is difficult or even impossible. Fortunately, the existing machine learning and artificial intelligence methods such as neural networks can solve complex multifactor-based mathematical relations. In this case, two problems can arise: (i) dimensionality issue and (ii) data set size issue. For dimensionality, using UPLC, the concentrations can be determined in mass/volume units, moles, parts per million, or even expressed in peak area units.[12,22,23] It can be a problem trying to solve the equation if the dimensions do not complement. Such questions are solved by normalizing the data set or modifying the values to dimensionless factors such as ratios.[22] Another critical issue is related to the size of the data set. For separation methods, the data set size is usually small (classically, it takes from 10 to 60 min to perform a single analysis). In this work, several factors were expressed as ratios between the determined amounts of PAH in different extracts. The obtained ratios are dimensionless values. They do not exceed 1 and are not lower than 0. The first ratio—f_1—was expressed using eq 13

$$f_1 = \frac{u_2}{u_1} = \frac{v_2 r_2 x_2}{v_1 r_1 x_1} \quad (13)$$

where f_1 is the ratio between the determined amount of PAH in the second stage extract, and u_2 and u_1 are the amounts of the first stage extract. x_1 is the selected PAH amount before the first stage extraction (x_2—second stage), r_1 is the first stage (r_2—second stage) recovery, and v_1 is the ratio between the volume of the added solvent and volume of the decanted extract in the first extraction stage (v_2—second stage). The second ratio—f_2—was expressed using eq 14

$$f_2 = \frac{u_3}{u_2} = \frac{v_3 r_3 x_3}{v_2 r_2 x_2} \quad (14)$$

where f_2 is the ratio between the determined amount of PAH in the third stage extract, and u_3 and u_2 are the amounts of the second stage extract. x_2 is the PAH amount before the second stage extraction (x_3—before third stage), r_2 is the second stage (r_3—third stage) recovery, and v_2 is the ratio between the volume of the added solvent and volume of the decanted extract in the second extraction stage (v_3—third stage). The ratios f_1 and f_2 can be calculated from the determined concentrations in the different stage extracts. Thus, following previously expressed equations, a python script was programmed, which calculates the values for different stages and outputs them in a comma-separated value (csv) file format.
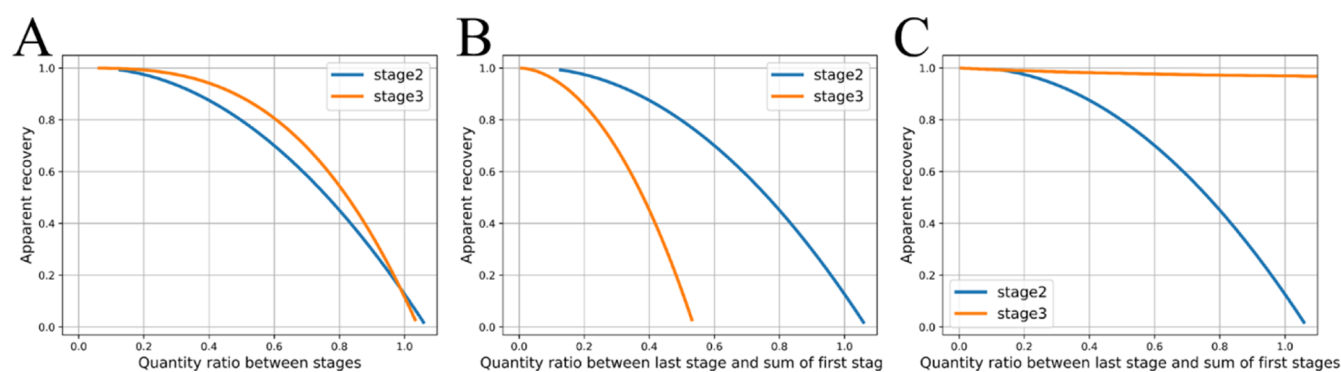
Figure 1. Apparent extraction recovery dependency on different stage recoveries (0.01−0.99). (A) Simulated quantity ratios (f_1, f_2) vs apparent recovery, (B) simulated quantity ratios (f_1, f_{2a}) vs apparent recovery, (C) simulated quantity ratios (f_1, f_{2a}) vs apparent recovery, when the third stage extraction recovery is constant. Settings: $v_1 = 0.89$, $v_2 = 0.95$, $v_3 = 0.99$; $r_{1min} = r_{2min} = r_{3min} = 0.01$ (A−C); $r_{1max} = r_{2max} = r_{3max} = 0.99$ (A,B); $r_3 = $ constant $= 0.95$ (C).

Table 1. Comparison of Different Polynomial Fittings for Different Variables and Conditions

no.	stage	x	y	r_{min}	r_{max}	equation	R^2
1	second	f_1	r_{2a}	0.30	0.99	$0.9368x2 + 0.0632x + 1.0000$	1.0000
2	third	f_2	r_{3a}	0.30	0.99	$0.9534x2 + 0.4588x + 0.9437$	0.9976
3	second	f_1	r_{2a}	0.01	0.99	$0.9368x2 + 0.0632x + 1.0000$	1.0000
4	third	f_2	r_{3a}	0.01	0.99	$1.4401x2 + 0.6529x + 0.9267$	0.9970
5	third	f_{2a}	r_{3a}	0.30	0.99	$-3.2344x2 - 0.0690x + 1.0008$	1.0000
6	third	f_{2a}	r_{3a}	0.01	0.99	$-3.3974x2 - 0.0153x + 0.9982$	1.0000
[a]7	third	f_{2a}	r_{3a}	0.30	0.99	$0.0227x2 - 0.0521x + 0.9995$	0.9987
[a]8	third	f_{2a}	r_{3a}	0.01	0.99	$0.0001x2 - 0.0055x + 0.9860$	0.6095

[a]$r_3 = 0.95$ (constant).

The fragment of the generated data is provided in Table S1. In actual multiple extraction experiments ($n \geq 5$), the average decanted volume ratios after multistage extraction were: v_1−−0.89, v_2−−0.95, and v_3−−0.99. Subsequently, simulations were done with the mentioned numbers. Figure 1 represents plots showing how apparent recovery (r_{2a} and r_{3a}) changes for different ratios ($f_1, f_2,$ and f_{2a}) between the extracted amounts of PAH.

It is observed (Figure 1A) that at very low extraction recoveries (<0.1), the ratios f_1 and f_2 exceed unity. It means that more substance is extracted in later stages than in initial stages ($u_2 > u_1$ and $u_3 > u_2$). The soaking effect leaves a significant volume of the extract in the bottle without decanting it, and in a later stage, it is decanted more (0.89 vs 0.95 vs 0.99). Also, if the raw material contains gel-forming or extrahent-absorbing components, a multistage extraction might not achieve a higher extract content (Figure S1).

Investigating the relationship between the last stage's quantity ratio and the sum of ratios of previous stages (Figure 1B, stage 3 (orange)), it was observed that no ratios exceeding unity are present in the range of recoveries per stage of 0.01−0.99. Additionally, for the same ratios of quantity, apparent recovery yields a higher extraction rate.

Investigating the case (Figure 1C) where the initial stages (1 and 2) output changing recovery and stage 3 is a constant (0.95), it is observed that the third stage extraction efficiency predetermines the apparent recovery if initial stages yield low numbers. It is important for the cases where a multistep extraction process is used where some stages are not necessarily directed to the extraction. Initial stages can be used for sample cleanup or water removal. Even though multiple extraction conditions have been compared and used

for the determination of PAH in water and used railway sleepers, the recovery dramatically fails if bioremediated wet samples are extracted with a water-immiscible solvent such as dichloromethane.[5,9,24] In this case, it is better to use an extrahent combination that can mix with a small quantity of water; therefore, a dichloromethane−acetone mixture was used, as in a previous study.[5] We adopted the same strategy in the current work, as acetone in the first and second extraction stages can significantly lower the water content in the sample. We also decided to use a less toxic extrahent—ethyl acetate.

The polynomial equation was fit between the x and y values, and it was noticed that the coefficient of determination was very high for all fittings (Table 1. Other mathematical model fittings (linear, exponential, and logarithmic) were also tested, and polynomial fittings provided the highest coefficients of determination (R^2). The represented cases are different from the perspective of possible fitting and possible use cases.

It was observed that the prediction of the three-stage apparent recovery from the ratio of third and second stage extracted quantities is lower than that of the second stage apparent recovery from the ratio of second and first stage extracted quantities (R^2: 0.9970 vs 1.0000). Obviously, in the fitting and mathematics of the third stage quantities, no values from first stage quantities were included. Therefore, it was decided to add another ratio which has first stage quantity included in the calculations (eq 15)

$$f_{2a} = \frac{u_3}{u_2 + u_1} = \frac{v_3 r_3 x_3}{v_2 r_2 x_2 + v_1 r_1 x_1} \qquad (15)$$

where f_{2a} is the ratio between the determined amount of PAH in the three-stage extract—u_3−−and the sum of u_1 and u_2 amounts (first and second stage extract quantities). x_1 is the
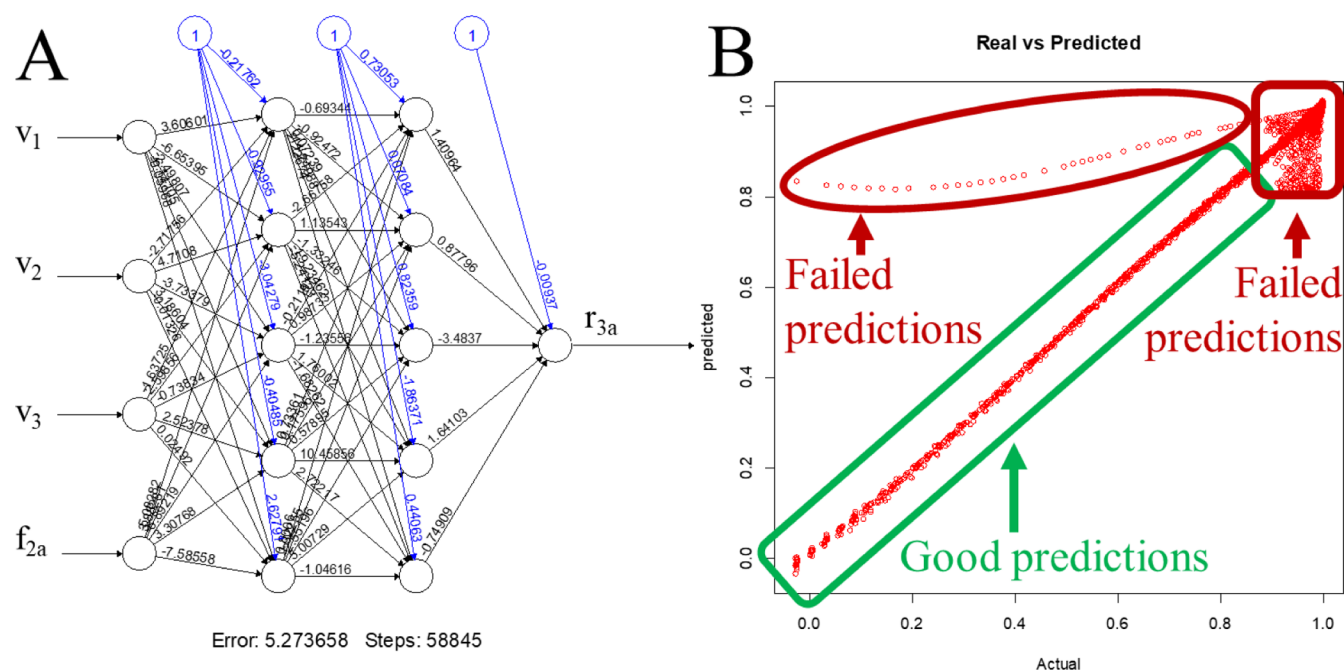
Figure 2. Neural network predictions of three-stage apparent recovery from the v_1, v_2, v_3, and f_{2a} values. (A) Neural network model and (B) plot representing actual and predicted apparent recovery values.

amount of selected PAH before the first stage extraction (x_2 and x_3—second and third stages, correspondingly), r_1 is the first stage (r_2 and r_3—second and third stages, correspondingly) recovery (between 0 and 1), and v_1 is the ratio between the volume of the added solvent and volume of the decanted extract in the first extraction stage (v_2 and v_3—second and third stages, correspondingly).

Calculating the polynomial coefficients of determination for polynomial fittings between f_{2a} and r_{3a} values indicated significant improvements (from 0.9970 to 1.000). Therefore, it was decided to include the f_{2a} value in later calculations instead of f_1 and f_2. The decanted volume ratio was also changed between 0.8 and 0.99, and relations were obtained between f_{2a} and r_{3a} values (Figure S1). Similar tendencies have been observed. These findings suggest a strong relationship between the determined quantities' ratios in different stage extracts and apparent recoveries. However, the relation mechanism is not clear, except for only the obvious observations: (i) higher stage recovery yields higher apparent recovery and (ii) higher decanted extract ratio yields higher apparent recovery.

Knowing the factors which affect the outcome of apparent recovery, a dataset containing 4560 data points was generated. Each data point was a vector consisting of the following variables: v_1, v_2, v_3, f_{2a}, and calculated r_{3a}. In the dataset, different combinations of different stage recoveries and decanted volume ratios were simulated. The simulated/generated data were used for training the neural network model. The variables v_1, v_2, v_3, and f_{2a} were used as training inputs, and r_{3a} values were used as prediction values. Various neural network combinations were tested: (i) two to eight hidden layers, (ii) two to six neurons in a layer. For different combinations, the training lasted between 15,132 and 68,550 steps. The mean-squared errors (MSEs) obtained were relatively high—in the range of 3.97−5.27 (Figure 2).

It was noticed that the neural network model was not capable of predicting apparent recovery values. Therefore, it

was decided to add f_1 and f_2 ratios into the training data set. New data set contained 3710 data points consisting of v_1, v_2, v_3, f_1, f_2, and f_{2a} values for inputs and r_{3a} value for training or predictions. Various combinations of neural network models were tested with the newly generated data set: (i) two to three hidden layers, (ii) two to seven neurons in a hidden layer. The shortest training took 3118 steps, and the longest training took 34,961 steps. The lowest MSE was 0.009, and the highest MSE was 0.050 for different models. Table 2 shows the selected neural network models and their performance. An example of the performance of other models is provided in Figures S2 and S3.

Table 2. Performance of Selected Neural Network Models for Different Trainings ($n = 10$)

no.	model	average MSE	RSD (%)
1	5:5	0.018	27.0
2	6:6	0.015	25.7
3	7:6	0.016	52.2
4	4:4:4	0.021	39.6
5	5:5:4	0.014	20.7

It was noticed that some models showed high performance (low MSE) and were selected as potentially useful for the current application. Each of the selected neural network models was retrained 10 times, and means of MSE and relative standard deviations (%) were calculated for each model. The 4:4:4 NN model provided the highest MSE, and the 7:6 NN model provided the highest RSD. The neural network model of 5:5:4 configuration provided the lowest MSE and lowest RSD, suggesting that it can be trained, and the expected MSE should be within 0.014 ± 0.003. For further research, the 5:5:4 NN configuration was selected (Figure 3).

Neural networks can predict r_{3a}, and the initial quantity in the extraction material can be calculated following eq 16
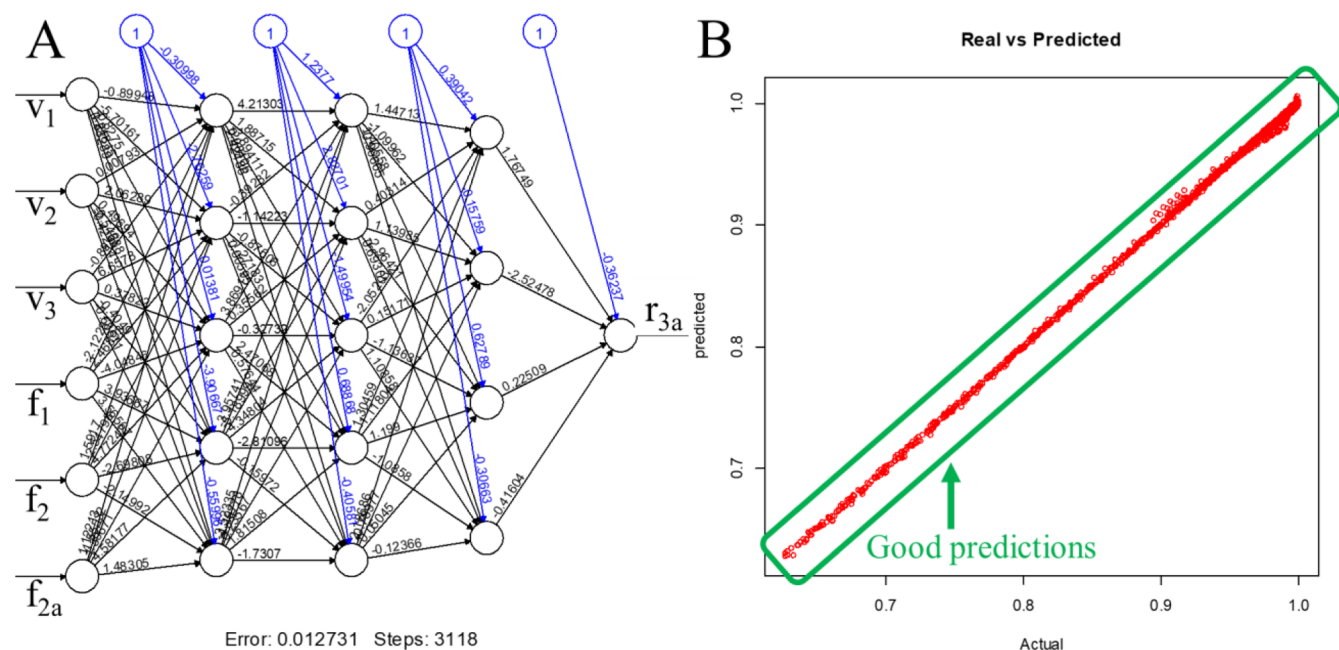
Figure 3. Highest performance neural network model. Input values v_1, v_2, v_3, f_1, f_2, and f_{2a}. (A) Neural network model and (B) plot representing actual and predicted apparent recovery values.

$$x_1 = \frac{u_1 + u_2 + u_3}{r_{3a}} \qquad (16)$$

where x_1 is the amount of selected PAH in the extraction material; r_{3a} is the three-stage apparent extraction recovery; u_1, u_2, and u_3 are the amounts of PAH determined in different stage (first, second, and third) extracts. The neural network model performance characteristics were evaluated. It was found that the accuracy, precision, sensitivity, specificity, and F1 score were 0.962, 0.961, 0.9648, 0.959, and 0.963, respectively.

2.3. Recovery of PAHs. Two samples were extracted and analyzed. Both samples contained visually similar contents. Mass did not differ more than 2% in the cotton bags. The first bag (I) was dried and had humidity not more than 10%, and the second bag (II) that was soaked in water contained > 50% humidity. Both samples were extracted using the same conditions, and after performing the extraction procedure and chemical analysis, it was determined that the extracts of sample I (dry) contained more PAHs than the extracts of sample II (wet) (Table 3). It was also observed that water in

Table 3. Determined PAHs in Measured Extracts

stage	sample	phenanthrene (mg/L)	anthracene (mg/L)	pyrene (mg/L)	V (L)
first	I	45.48	12.96	22.34	0.086
second	I	6.42	1.83	3.33	0.095
third	I	1.11	0.30	0.81	0.099
first	II	6.93	4.64	4.55	0.091
second	II	3.05	0.86	2.35	0.096
third	II	0.69	0.10	0.56	0.099

the extraction matrix influenced the decanted extract volume. After the first and second stages of extraction, the wet sample decanted more extract compared to the dry sample.

To determine phenanthrene, anthracene, and pyrene, a previously developed method demonstrated in laboratory-scale bioremediation experiments was used for chemical analysis.[5]

The obtained chemical analysis values and volume measurements were used to calculate the necessary variables for training the neural networks. The solvent volume added to the extraction bottle was 0.1 L, and together with the measured extract volumes, v_1, v_2, and v_3 values were calculated. Quantities u_1, u_2, and u_3 were calculated from the concentrations, and later, these quantities were used for calculating f_1, f_2, and f_{2a} ratios (Table 4). The calculated variables were used in the trained neural network model, so that the apparent three-stage extraction recovery (r_{3a}) can be predicted. Twenty iterations were performed for retraining and repredicting the r_{3a} value—the apparent three-stage recovery. The average values are presented in Table 4, and RSD (%) is calculated for the predictions. It was noticed that there were two predictions (out of 120 predictions) providing negative values, and they were excluded as the outliers. Other outliers that significantly differed were identified using the Thompson Tau test calculations and excluded like it was done in the previous study.[5]

It was observed that in sample I (dry), the extracts showed recoveries around 1. Recoveries slightly exceeding unity can be explained by the fact that chromatographic methods usually determine concentrations within 5% accuracy and precision. Additionally, the errors introduced in preparing the sample are higher than the instrumentation errors and cannot always be tracked. Furthermore, the predicted recoveries exceeding unity suggest that minor error has been introduced in either measuring the extract volume or determining the extract concentration. Therefore, the initial quantity x_1 should be adjusted.

The RSD of predictions did not exceed 3% except for sample II extracts where pyrene was determined, and the predicted apparent recovery was 0.55 (55%). It was observed that sample I contained more PAH than sample II. It was found that sample I contained 4.47 mg phenanthrene, 1.26 mg anthracene, and 2.21 mg pyrene and sample II contained

Table 4. Data Used for Predictions and the Recovery and Determined Amounts of PAH

substance	sample	v_1	v_2	v_3	f_1	f_2	f_{2a}	r_{3a}	RSD (%)	$u1 + u_2 + u_3$ (mg)	x_1 (mg)
phenanthrene	I	0.86	0.95	0.99	0.16	0.18	0.02	1.04	1.79	4.63	4.47
anthracene	I	0.86	0.95	0.99	0.16	0.17	0.02	1.04	2.67	1.32	1.26
pyrene	I	0.86	0.95	0.99	0.16	0.25	0.04	1.05	2.38	2.32	2.21
phenanthrene	II	0.91	0.96	0.99	0.46	0.23	0.07	0.71	1.13	0.92	1.30
anthracene	II	0.91	0.96	0.99	0.20	0.12	0.02	0.98	0.65	0.50	0.51
pyrene	II	0.91	0.96	0.99	0.54	0.24	0.09	0.55	5.15	0.64	1.15

1.30 mg phenanthrene, 0.51 mg anthracene, and 1.15 mg pyrene.

This method will be useful in food industry, pharmaceutical manufacturing, biotechnology, and chemical industry. The authors are of the opinion that the developed method will be of importance in unmanned autonomous investigations such as planetary explorations as well. In such research, spiking cannot be performed, and human assistance is impossible; fortunately, a multistage extraction process is achievable.[25]

3. CONCLUSIONS

An unprecedented method for determining the recovery of extracted substances in a multistage extraction process has been developed. The method is a combination of a mathematical model and a machine learning method—neural networks. The method does not require an external standard, spiking procedure, or any reference. The method is expected to be useful in any multistage extraction, even for different substances other than PAHs.

4. MATERIALS AND METHODS

4.1. Chemicals and Instrumentation. Acetone (99.8%) and methanol (MeOH) (99.9%) were purchased from Macron (Poland). Acetonitrile (ACN) (99.9%), ethylacetate (99.9%), and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Germany). Bidistilled water was produced in our laboratory using Fistreem Cyclon bidistillator (United Kingdom). Cotton tea bags 8 × 12 cm (PRC) were purchased from the local store. An Acquity UPLC system equipped with a fluorescence detector was purchased from Waters (USA).

4.2. Crushing of Railway Sleepers. Used railway sleepers were collected from Lithuanian railway company *Lietuvos Geležinkeliai* (coordinates: 54.881187, 23.934914). Three standard-sized intact softwood railway sleepers stored for 10 years after usage were ground on September 07, 2018, using a Jensen A530 wood chipper (Jensen, United Kingdom). The pieces of the crushed material occupied not less than 0.75 m³. After the crushing, the pieces were stored in three 0.3 m³ volume high-density polyethylene bags near the bioremediation site.

4.3. Analytical Procedure. Chemical analysis was performed using previously optimized conditions.[5] A gradient acidified with TFA water (0.05%) and ACN was used for separation in a Waters Acquity UPLC HSS T3 polar embedded 2.1 × 150 mm column. The gradient initially used 15% ACN which raised to 85% in 20 min. The column was thermostated at 35 °C. 5 μL of samples thermostated at 5 °C was injected before separation.

4.4. Extraction and Sample Preparation. The samples were extracted using a three-stage procedure. For the first stage, a tea bag containing ground railway sleeper and soil was placed in a 100 mL bottle, and 100 mL of acetone was added into the bottle. The bottle with the contents was shaken

overnight at ambient temperature (22 °C) at 200 rpm. After 24 h, the bottle was opened, and the acetone extract was decanted into a measurement cylinder. The first stage acetone extract was added into a storage bottle and kept in a fridge at 4 °C. For the second stage, 100 mL of acetone was added into a bottle with the partially extracted content and shaken overnight. After 24 h, the acetone extract was decanted into a measurement cylinder, and the second stage acetone-extract was added into another bottle. For the third stage, 100 mL of ethyl acetate was added into a bottle with the partially extracted content and shaken overnight. After 24 h, the ethyl acetate extract was decanted into a measurement cylinder, and after measurement, it was added into a bottle. Before use, the extracts were taken out of the fridge, filtered via a 0.47 μm membrane filter, diluted 20−100 times with MeOH, and used for direct injection and separation in the UPLC system.

4.5. Data Analysis and Modeling. The development of mathematical statements is described in the section Development of Mathematical Extraction Model. Following the mathematical statements, the data were generated using *Python* programming language with *Pycharm* software and exported as a csv file format.

Neural network modeling was performed in R environment, using Rstudio software (version 1.1.442) and a *neuralnet* package (version 1.33).[26,27] The generated data points were separated into two equal-sized random groups for training the neural network model: one for training and another to validate predictions. The default settings of the *neuralnet* package were used. To train the neural networks, a *resilient backpropagation with weight backtracking* algorithm was utilized. As the ANN model's output was a number and not a class, the differentiable activation function is bypassed. A *sum of squared errors* was used as a differentiable error function. Training and predictions were repeated 10 times, and MSEs were recorded. From the MSEs, the means and relative standard deviations [RSD (%)] were calculated. Two main types of neural network models were trained: (i) four inputs and (ii) six inputs. Different models were trained: (a) containing from two to eight hidden layers, (b) having from two to six neurons in a hidden layer (this included different combinations of neuron numbers in a hidden layer). All models were used for predicting only one parameter—the three-stage apparent recovery.

The predicted cases were classified to calculate the developed model performance characteristics. If the predicted value is equal to or higher than the actual value from the dataset, then it is classified as *positive*. If the predicted value is lower than the true value, then it is classified as *negative*. A 0.5% error criterion (the value that the chemical analytical high-performance instrumentation operates at) was used to classify true and false cases. If the predicted value differs by less than 0.5% from the actual value, then the data point is classified as *true*. If the predicted value differs by more than 0.5% from the actual value, then the data point is classified as

false. The obtained numbers from the dataset were used to calculate the accuracy, precision, sensitivity, specificity, and F1 score. Ten retrainings were performed, and mean values were reported.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c01737.

Example of generated data for training; change of apparent recovery and quantity ratios between the last and sum of first stages; trained neural network model with four inputs, one output, and three hidden layers (failing example); trained neural network model with six inputs, one output, and three hidden layers (one of potential examples) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Audrius Maruška − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania;* ⓞ orcid.org/0000-0002-9267-3805; Email: audrius.maruska@vdu.lt

Authors

Tomas Drevinskas − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Kristina Bimbiraitė-Survilienė − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Gediminas Dūda − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Mantas Stankevičius − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Nicola Tiso − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Rūta Mickienė − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Vilmantas Pedišius − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Donatas Levišauskas − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania; Process Control Department, Kaunas University of Technology, LT-51368 Kaunas, Lithuania*

Vilma Kaškonienė − *Instrumental Analysis Open Access Centre, Faculty of Natural Sciences, Vytautas Magnus University, LT-44404 Kaunas, Lithuania*

Ona Ragažinskienė − *Kaunas Botanical Garden of Vytautas Magnus University, LT-46324 Kaunas, Lithuania*

Saulius Grigiškis − *JSC Biocentras, LT-14117 Pagiriai, Lithuania*

Enrica Donati − *National Research Council, 300-00015 Monterotondo (Rome), Italy*

Massimo Zacchini − *National Research Council, 300-00015 Monterotondo (Rome), Italy*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.1c01737

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Wang, Y.-B.; Liu, C.-W.; Kao, Y.-H.; Jang, C.-S. Characterization and Risk Assessment of PAH-Contaminated River Sediment by Using Advanced Multivariate Methods. *Sci. Total Environ.* **2015**, *524-525*, 63−73.

(2) Bosetti, C.; Boffetta, P.; La Vecchia, C. Occupational Exposures to Polycyclic Aromatic Hydrocarbons, and Respiratory and Urinary Tract Cancers: A Quantitative Review to 2005. *Ann. Oncol.* **2007**, *18*, 431−446.

(3) Winquist, E.; Björklöf, K.; Schultz, E.; Räsänen, M.; Salonen, K.; Anasonye, F.; Cajthaml, T.; Steffen, K. T.; Jørgensen, K. S.; Tuomela, M. Bioremediation of PAH-Contaminated Soil with Fungi - From Laboratory to Field Scale. *Int. Biodeterior. Biodegrad.* **2014**, *86*, 238−247.

(4) Vila, J.; Tauler, M.; Grifoll, M. Bacterial PAH Degradation in Marine and Terrestrial Habitats. *Curr. Opin. Biotechnol.* **2015**, *33*, 95−102.

(5) Drevinskas, T.; Mickienė, R.; Maruška, A.; Stankevičius, M.; Tiso, N.; Mikašauskaitė, J.; Ragažinskienė, O.; Levišauskas, D.; Bartkuvienė, V.; Snieškienė, V.; et al. Downscaling the in Vitro Test of Fungal Bioremediation of Polycyclic Aromatic Hydrocarbons: Methodological Approach. *Anal. Bioanal. Chem.* **2016**, *408*, 1043−1053.

(6) Serpe, F. P.; Esposito, M.; Gallo, P.; Serpe, L. Optimisation and Validation of an HPLC Method for Determination of Polycyclic Aromatic Hydrocarbons (PAHs) in Mussels. *Food Chem.* **2010**, *122*, 920−925.

(7) Janoszka, B. HPLC-Fluorescence Analysis of Polycyclic Aromatic Hydrocarbons (PAHs) in Pork Meat and Its Gravy Fried without Additives and in the Presence of Onion and Garlic. *Food Chem.* **2011**, *126*, 1344−1353.

(8) Olatunji, O. S.; Fatoki, O. S.; Opeolu, B. O.; Ximba, B. J. Determination of Polycyclic Aromatic Hydrocarbons [PAHs] in Processed Meat Products Using Gas Chromatography - Flame Ionization Detector. *Food Chem.* **2014**, *156*, 296−300.

(9) Stankevičius, M.; Maruška, A.; Tiso, N.; Mikašauskaite, J.; Bartkuviene, V.; Kornyšova, O.; Mickiene, R.; Bimbiraite-Surviliene, K.; Kaškoniene, V.; Kazlauskas, M.; et al. Gas Chromatographic Analysis of Polycyclic Aromatic Hydrocarbons in the Disposed Creosote Treated Wooden Railway Sleepers Collected from Several Storage Sites in Lithuania. *Chemija* **2015**, *26*, 198−207.

(10) Kohler, M.; Künniger, T. Emissions of Polycyclic Aromatic Hydrocarbons (PAH) from Creosoted Railroad Ties and Their Relevance for Life Cycle Assessment (LCA). *Holz Roh- Werkst.* **2003**, *61*, 117−124.

(11) Liu, Y.; Gao, Y.; Yu, N.; Zhang, C.; Wang, S.; Ma, L.; Zhao, J.; Lohmann, R. Particulate Matter, Gaseous and Particulate Polycyclic Aromatic Hydrocarbons (PAHs) in an Urban Traffic Tunnel of China: Emission from on-Road Vehicles and Gas-Particle Partitioning. *Chemosphere* **2015**, *134*, 52−59.

(12) Drevinskas, T.; Maruška, A.; Telksnys, L.; Hjerten, S.; Stankevičius, M.; Lelešius, R.; Mickienė, R.; Karpovaitė, A.; Šalomskas, A.; Tiso, N.; et al. Chromatographic Data Segmentation Method: A Hybrid Analytical Approach for the Investigation of Antiviral Substances in Medicinal Plant Extracts. *Anal. Chem.* **2019**, *91*, 1080−1088.

(13) Wesołowski, M.; Suchacz, B. Classification of Rapeseed and Soybean Oils by Use of Unsupervised Pattern-Recognition Methods and Neural Networks. *Anal. Bioanal. Chem.* **2001**, *371*, 323−330.

(14) Latorre, R. M.; Hernández-Cassou, S.; Saurina, J. Artificial Neural Networks for Quantification in Unresolved Capillary Electrophoresis Peaks. *J. Sep. Sci.* **2001**, *24*, 427−434.

(15) Alpaydın, E. *Introduction to Machine Learning*, 2nd ed.; Dietterich, T., Bishop, C., Heckerman, D., Jordan, M., Kearns, M., Eds.; The MIT Press: London, 2014; Vol. *1107*.

(16) Lavecchia, A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov. Today* **2015**, *20*, 318−331.

(17) Tušek, A. J.; Jurina, T.; Benković, M.; Valinger, D.; Belščak-Cvitanović, A.; Kljusurić, J. G. Application of Multivariate Regression and Artificial Neural Network Modelling for Prediction of Physical and Chemical Properties of Medicinal Plants Aqueous Extracts. *J. Appl. Res. Med. Aromat. Plants* **2019**, *16*, 100229.

(18) Ciric, A.; Krajnc, B.; Heath, D.; Ogrinc, N. Response Surface Methodology and Artificial Neural Network Approach for the Optimization of Ultrasound-Assisted Extraction of Polyphenols from Garlic. *Food Chem. Toxicol.* **2020**, *135*, 110976.

(19) Pavlić, B.; Kaplan, M.; Bera, O.; Oktem Olgun, E.; Canli, O.; Milosavljević, N.; Antić, B.; Zeković, Z. Microwave-Assisted Extraction of Peppermint Polyphenols − Artificial Neural Networks Approach. *Food Bioprod. Process.* **2019**, *118*, 258−269.

(20) Alara, O. R.; Abdurahman, N. H.; Afolabi, H. K.; Olalere, O. A. Efficient Extraction of Antioxidants from Vernonia Cinerea Leaves: Comparing Response Surface Methodology and Artificial Neural Network. *Beni-Suef Univ. J. Basic Appl. Sci.* **2018**, *7*, 276−285.

(21) Dahmoune, F.; Remini, H.; Dairi, S.; Aoun, O.; Moussi, K.; Bouaoudia-Madi, N.; Adjeroud, N.; Kadri, N.; Lefsih, K.; Boughani, L.; et al. Ultrasound Assisted Extraction of Phenolic Compounds from P. Lentiscus L. Leaves: Comparative Study of Artificial Neural Network (ANN) versus Degree of Experiment for Prediction Ability of Phenolic Compounds Recovery. *Ind. Crops Prod.* **2015**, *77*, 251−261.

(22) Drevinskas, T.; Mickienė, R.; Maruška, A.; Stankevičius, M.; Tiso, N.; Šalomskas, A.; Lelešius, R.; Karpovaitė, A.; Ragažinskienė, O.; Stankevicius, M.; et al. Confirmation of the Antiviral Properties of Medicinal Plants via Chemical Analysis, Machine Learning Methods and Antiviral Tests: A Methodological Approach. *Anal. Methods* **2018**, *10*, 1875−1885.

(23) Drevinskas, T.; Mickienė, R.; Maruška, A.; Stankevičius, M.; Tiso, N.; Šalomskas, A.; Lelešius, R.; Karpovaitė, A.; Ragažinskienė, O. Cytotoxic Attributes in Polyphenolic and Volatile Compounds Rich Antiviral Plants. *Chemija* **2018**, *29*, 124−131.

(24) Ikarashi, Y.; Kaniwa, M.-a.; Tsuchiya, T. Monitoring of Polycyclic Aromatic Hydrocarbons and Water-Extractable Phenols in Creosotes and Creosote-Treated Woods Made and Procurable in Japan. *Chemosphere* **2005**, *60*, 1279−1287.

(25) Kehl, F.; Kovarik, N. A.; Creamer, J. S.; Costa, E. T.; Willis, P. A.; da Costa, E. T.; Willis, P. A. A Subcritical Water Extractor Prototype for Potential Astrobiology Spaceflight Missions. *Earth Sp. Sci.* **2019**, *6*, 2443−2460.

(26) RStudio Team. *RSudio: Integrated Development*; R. RStudio, Inc.: Boston, MA, 2015.

(27) Günther, F.; Fritsch, S. Neuralnet: Training of Neural Networks. *R J* **2010**, *2*, 30−38.