Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Anal. Methods*, 2024, **16**, 1836

Mineral oil emulsion species and concentration prediction using multi-output neural network based on fluorescence spectra in the solar-blind UV band

Bowen Gong,[ab] Shilei Mao,[ab] Xinkai Li[a] and Bo Chen*[a]

Received 14th October 2023
Accepted 28th February 2024

DOI: 10.1039/d3ay01820b

rsc.li/methods

The accurate monitoring of oil spills is crucial for effective oil spill recovery, volume determination, and cleanup. Oil slicks become emulsified under the effects of wind and waves, which increases the consistency of the oil spills. This phenomenon makes oil spills more challenging to handle and exacerbates environmental pollution. In this study, the variation of the solar-blind ultraviolet (UV) fluorescence spectra obtained from simulated oil spills with different oil types and oil–water ratios was investigated. By designing and constructing a multi-angle excitation and detection system, an apparent fluorescence peak of the oil emulsions was observed at around 290 nm under 220 nm excitation. By utilizing competitive adaptive reweighted sampling (CARS) and multi-output neural network algorithms, both the types and concentrations of the emulsified oils were obtained simultaneously. The classification accuracy for identifying the oil type exceeds 98%, and the mean absolute percentage error (MAPE) for concentration regression is around 2%. The results indicate that active solar-blind UV fluorescence could become a supplementary method for on-site oil spill detection to achieve comprehensive monitoring of oil spills. This study provides potential applications for UV-induced fluorescence spectrometry in oil spill on-site monitoring during the daytime.

1. Introduction

Oil spills typically occur during the extraction and transportation processes, and due to ship accidents.[1,2] According to statistics, over 6 million tons of oil enter the marine environment globally every year. One ton of oil can form an oil film covering 12 km^2 at sea, which blocks gas exchange and causes massive fish kills.[3] Toxic substances in oil dissolve lipids, thereby destroying the cellular structure of organisms.[4] They also affect the nervous system and organs such as the liver, lungs, and kidneys. The polycyclic aromatic hydrocarbons (PAHs) in oil can be absorbed by living organisms and eventually enter the human body, posing significant carcinogenic risks to human health. Considering the importance of protecting the environment and reducing the risk of oil spill disasters, accurate and timely information about the location, type, and volume of an oil spill is crucial for efficient oil spill recovery and the prevention of future oil spill accidents.

The thickness of oil spills on the sea surface is not uniform due to various factors, such as wind patterns, tidal forces, and seafloor activities. Additionally, the spills can evaporate, dissolve, and emulsify, resulting in changes to the physical and chemical properties of the spilled oil. Depending on the degree of emulsification, there are two types of emulsions: "oil-in-water" and "water-in-oil".[5] Oil-in-water emulsions consist of oil droplets dispersed in a water phase, while water-in-oil emulsions consist of water droplets dispersed in an oil phase. Emulsification increases the volume, viscosity, and density of the oil spill, which results in more significant damage to the environment and greater challenges in handling and recovery.

At present, sensors that can be used for sea surface oil spill detection include microwave sensors, optical sensors, laser sensors, and photoacoustic spectroscopy[6–8] sensors, among others. The ultraviolet (UV)-induced fluorescence method has been widely used in many fields for object identification monitoring and quantitative assessment, as it has the advantages of high sensitivity, simple operation, and low false alarm rates.[9] Different types of oils contain different types and contents of aromatic hydrocarbons,[10] leading to different fluorescence spectral characteristics.[11] The fluorescence spectra of oil spills also vary with the degree of emulsification. In the detection of offshore oil spills using UV-induced fluorescence, long-wave UV excitation (such as 308 nm and 355 nm) is typically employed due to the high cost and weak intensity of short-wave UV light sources.[12,13] The visible fluorescence excited by long-wave UV light may be drowned out by the background in sunlight,[14,15] making it suitable for nighttime monitoring only.

Oxygen in the atmosphere strongly absorbs radiation energy in the UV band below 200 nm, while ozone significantly absorbs

---

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin Province, 130033, China. E-mail: ciomp@ciomp.ac.cn
[b] University of Chinese Academy of Sciences, Beijing, 100049, China

the UV radiation ranging from 200–280 nm.[16] In the near-Earth atmosphere, the UV radiation is evenly distributed due to strong scattering by the atmosphere. The UV radiation below 300 nm near the ground is weak, forming a region known as the solar-blind UV band.[17] The feasibility of solar-blind fluorescence detection of offshore oil was investigated.

The composition of crude oil is complex and varies greatly depending on its source. The fluorescence composition and ratios of oil products produced using different boiling points also vary. For example, the main fluorescent substances in gasoline are toluene and *m*-xylene with fluorescence peaks near 270 nm.[11] The fluorescent substances in diesel oil consist of monocyclic aromatic hydrocarbons with longer side chains, bicyclic aromatic hydrocarbons (such as naphthalene, acenaphthene, fluorene, and biphenyl), tricyclic aromatic hydrocarbons (phenanthrenes, anthracene, and their derivatives), PAHs, and cycloalkane-aromatic mixtures.[18] In addition to the aromatic hydrocarbons in the middle fraction, crude oil contains more cycloalkyl-aromatic hydrocarbons with high ring numbers. The short-wave fluorescence intensity decreases with an increase in the number of aromatic rings.[19]

When fluorescent molecules come together, the electronic coupling between them creates new excited states, and fluorescence quenching occurs. In a broad sense, fluorescence quenching refers to any effect that reduces the quantum yield of fluorescence.[20] The interaction between different PAH compounds is strong, and there is a nonlinear summation caused by fluorescence quenching, resulting in different quenching concentrations of each component and wide differences in the range of influence.[21] Spatial isolation and electronic isolation prevent quenching and allow the molecule to recover fluorescence. The higher water content in oil-in-water emulsions leads to an increase in intermolecular distance, which reduces the probability of collision energy transfer. Specifically, the long-wave fluorescence of heavy PAHs dominates when the concentration is high, and the short-wave fluorescence spectra only gradually appears as the concentration decreases. In other words, the fluorescence peak shows a certain degree of blue-shift in wavelength as the concentration decreases.

Emilia Baszanowsk measured the fluorescence of dissolved oil in seawater by measuring the excitation emission matrix (EEM) with a fluorescence spectrometer.[22] Rather than measuring the fluorescence spectra after sampling using a fluorescence spectrophotometer,[23] a simulated field test of emulsified oil on the sea surface was employed to measure the fluorescence. The fluorescence of the oil emulsion was measured using a shorter excitation light source wavelength of 222 nm. The fluorescence peak appeared at 290 nm, which falls within the solar-blind UV range, in which the background noise is nearly negligible near the ground. By utilizing competitive adaptive reweighted sampling (CARS) and a multi-output neural network algorithm, both the type and concentration of emulsified oil can be obtained simultaneously. The results show that this algorithm can obtain accurate classification and concentration prediction results at the same time. This method can attenuate the background interference caused by sunlight and provides a new idea and method for the all-day, all-weather application of UV-induced fluorescence spectrometry.

2. Materials and methods

2.1 Samples and experimental principles

2.1.1 Oil sample preparation. Three refined oil products and one crude oil sample were used in this paper. The refined oil products were −20# D diesel (−20# D), −35# D diesel (−35# D), and 95# gasoline (95# G). The above refined oil samples were purchased from local gas stations. The crude oil was Russian light crude oil (LC).[24]

The physical parameters of the experimental sample oils are shown in Table 1. The oil samples are shown in Fig. 1a. The seawater was collected from Liaodong Bay in China. API is a parameter developed by the American Petroleum Institute to indicate the density of petroleum.

In order to obtain the trend in the variation of the fluorescence with concentration,[25] the fluorescence spectra of oil-in-water emulsions with ten concentrations from 20 ppm to 20 000 ppm were measured, as shown in Table 2. The "ppm" here represents parts per million of a volume ratio. The prepared oil emulsion samples are shown in Fig. 1b.

The preparation of an oil-in-water emulsion of −20# D at 100 ppm is presented as an example:

1.0.02 mL of −20# D was transferred into a 200 mL beaker using a sampling bottle,[26] and set aside for later use. 100 mL of seawater was measured and added to the previously prepared sampling bottle.

2. The mixture liquid was shaken for two days to increase the solubility to simulate the state of oil spills at sea.

3. The sampling bottle containing the mixture of oil and seawater was placed into an ultrasonic emulsifying machine for emulsification.[27]

4. After 30 minutes of ultrasonic treatment, a uniformly distributed oil-in-water emulsion was obtained.

5.60 mL of the middle layer emulsion was transferred into a 90 mm diameter Petri dish using a pipette for use as the test sample.

6. The above steps were repeated to prepare oil-in-water emulsions of different concentrations and types.

2.1.2 Experimental process. The solar radiation reaching the ground is influenced by the area and weather conditions, so the solar radiation spectrum near the ground in Changchun on a sunny day was first collected. In addition to sunlight, space light scattered by the atmosphere and reflected light from the nearby environment were also received. The solar spectrum was collected at 13:00 BST with the fiber optic receiver head of the portable spectrometer pointed at the sun. The sky light was measured by changing the receiving angle to near the zenith angle equal to 0°.

Table 1 Physical parameters of the experimental sample oils

	Density (20 °C, g mL^{-1})	API (°)	Viscosity (40 °C, mm^2 s^{-1})
95# G	0.74	60.3	0.72
−35# D	0.82	40.9	2.10
−20# D	0.83	38.8	2.80
LC	0.81	42.3	4.30

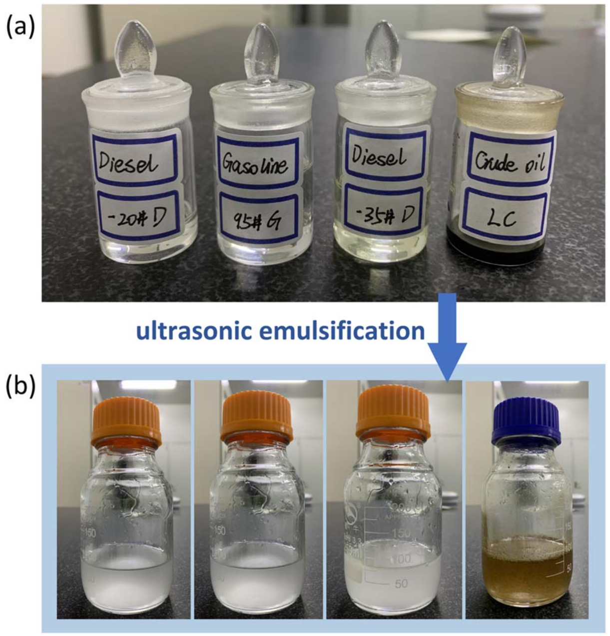

Fig. 1 Oil samples. (a) Samples of the four pure oils. (b) Samples of four oil-in-water emulsions.

Table 2 Oil-in-water emulsion concentration ratios

Oil volume (mL)	Seawater volume (mL)	Oil content (ppm)
0.02	1000	20
0.02	200	100
0.02	100	200
0.04	100	400
0.10	100	1000
0.20	100	2000
0.40	100	4000
0.80	100	8000
1.60	100	16 000
2.00	100	20 000

Subtracting the sky light from the sunlight gives the radiation spectrum of the sun (Fig. 2). As shown in Fig. 2, the light intensity in the range of 200–300 nm was almost 0, so the whole 200–300 nm region can be regarded as the solar-blind area.

Fluorescent substances with fluorescence peaks below 300 nm mainly include benzene and its derivatives. Due to the Stokes shift, the fluorescence emission wavelength is slightly greater than the excitation wavelength.

The absorption bands of benzene, toluene, and xylene are observed between 200 nm and 220 nm, and the absorption peaks shift to longer wavelengths as substituents are introduced. Therefore, a light source with a center wavelength of 220 nm was chose as the excitation light.

The experimental schematic diagram is shown in Fig. 3. A 20 W excimer lamp was used as the excitation light source, and light was collimated with a lens set due to the strong scattering and weak energy of solar-blind UV light. The collimated light was vertically irradiated to the emulsified oil sample in a Petri dish. Finally, the fiber optic head of the spectrometer captured the emitted fluorescence at a 45° angle to obtain the maximum fluorescence signal while minimizing the influence of interference signals such as reflected and scattered light. Interference signals would be eliminated during the data preprocessing.

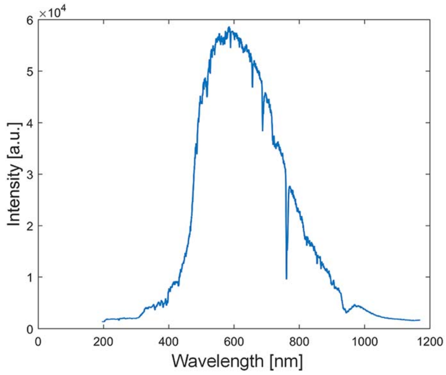

Fig. 2 Spectrum of solar radiation received at ground level.

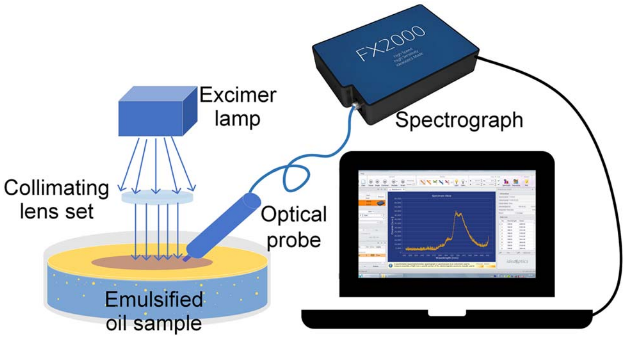

Fig. 3 Schematic diagram for simulated field monitoring of emulsified oil fluorescence spectra.

The FX2000+ optical fiber spectrometer was selected to collect the fluorescence spectrum; its detection range is 197–419 nm. The spectrometer operates with a slit width of 100 μm, yielding a spectral resolution of 0.59 nm.

In this study, fluorescence spectra were collected for ten different concentrations of four types of oils. Each sample was measured twenty times to increase veracity and reliability. The emulsified oil sample was replaced after measuring the current sample, and this process was repeated until all samples had been measured.

2.2 Data preprocessing

The spectral data preprocessing in this work consisted of three parts: denoising, smoothing, and feature selection. Denoising was achieved by subtracting the background noise spectrum from the measured fluorescence spectrum.

The Savitzky–Golay (S–G) filter was proposed in 1964 and has been widely used for data smoothing and noise reduction. It is a filtering method based on local polynomial least-squares fitting. The advantage of the S–G filter is that it removes noise while preserving the shape and width of the signal. In this

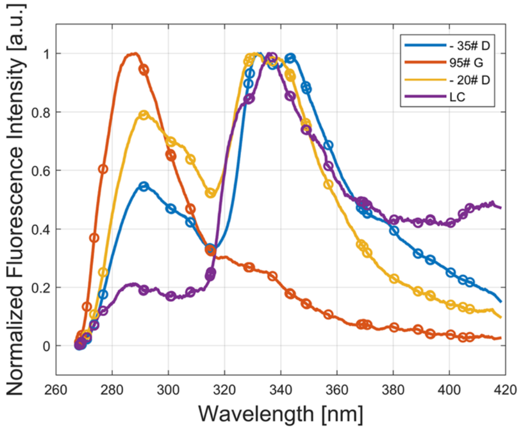

Fig. 4 Competitive adaptive reweighted sampling (CARS) feature processing results.

study, the fluorescence signals obtained were weak and had a low signal-to-noise ratio.[28] Therefore, the S–G filter was chosen as the smoothing and processing algorithm, which is superior to the moving average smoothing algorithm.

In the present work, a variable selection algorithm based on iterative statistical information, Competitive Adaptive Reweighted Sampling (CARS), was used to extract features. CARS is a feature variable selection method that combines Monte Carlo sampling with partial least-squares (PLS) model regression coefficients. Each time, the algorithm retains a subset of points with relatively high absolute weight values in the regression coefficients of the PLS model through adaptive reweighted sampling (ARS) and removes points with relatively low weight values. Then, a new PLS model is built based on the new subset, and after multiple calculations, the wavelengths in the subset with the smallest root mean square error of cross-validation (RMSECV) are selected as the characteristic wavelengths.[29]

By eliminating redundant information variables in the spectrum and selecting representative variables that represent sample properties instead of using the full spectrum to establish a quantitative model, this algorithm can improve the accuracy of the analysis results while reducing the time required for data processing.

The feature extraction results are shown in Fig. 4. The CARS algorithm selected 41 feature points from 2048 spectral points. The circles indicate the selected feature points, and the lines indicate the original spectral data of the four oil samples.

3. Results and discussion

3.1 Trends in fluorescence spectrum variation

To simulate the state of emulsified oil on the sea surface as accurately as possible, the oil-in-water emulsion used in this article is an emulsified oil suspension that has only undergone ultrasonic treatment without any emulsifiers. As the concentration increases, some droplets float to the seawater surface under the buoyancy of seawater and re-diffuse into an oil film due to the instability of ultrasonically emulsified oil. The fluorescence of the emulsified oil cannot be obtained when the excitation light and emitted fluorescence cannot penetrate the oil film, and the fluorescence spectrum at this time is basically the same as that of the pure oil.[30]

The initial spectrum of −35# D is shown in Fig. 5a. The fluorescence of emulsified −35# D at 20 ppm was virtually undetectable by our spectrometer. The short-wave UV fluorescence intensity shows a gradual increase as the concentration of the oil-in-water emulsion increases, until it reaches its highest point at 4000 ppm.

The short-wave UV fluorescence intensity does not increase but instead decreases when the concentration exceeds 8000 ppm. When the concentration increases further, the emulsion breaks down to form an oil film on the seawater surface, and the fluorescence spectrum becomes identical to the fluorescence of the oil film. Fig. 5b shows the normalized fluorescence spectrum of −35# D. The intensity ratio of the short-wave UV fluorescence and long-wave UV first increases and then decreases with increasing concentration. The changes in the peak fluorescence of the emulsified oil in gasoline 95# G are different from those in −35# D diesel.

As shown in Fig. 5c, the fluorescence of emulsified 95# D at 20 ppm is virtually undetectable. When the concentration of 95# G in seawater is only 100 ppm, the fluorescence spectrum has only one peak at 290 nm. As the concentration of the sample gradually increases, the intensity of the fluorescence peak at 290 nm in the fluorescence spectrum also increases. Additionally, peaks at around 320 nm and 335 nm emerge, and the intensities of these two fluorescence peaks also increase (Fig. 5c). When the concentration exceeds 20 000 ppm, the fluorescence intensity at 290 nm decreases, and the fluorescence intensity at 320 nm and 335 nm increases dramatically, approaching the shape of the oil film fluorescence spectrum. Fig. 5d shows the normalized fluorescence spectrum of 95# G.

The fluorescence spectra and normalized fluorescence spectra of −20# D are shown in Fig. 6a and b, respectively. Similar to the spectra of −35# D, the fluorescence at 290 nm is gradually enhanced as the concentration increases from 100 to 2000 ppm. When the concentration changes from 4000 to 20 000 ppm, the short-wave UV fluorescence intensity decreases, and the long-wave UV fluorescence gradually increases. The normalized fluorescence graph reveals that the ratio of long-wave to short-wave fluorescence peaks tends to decrease as the concentration increases. The number of fluorescence peaks changes from two to four.

The fluorescence of LC is shown in Fig. 6c. The fluorescence peak is located near 340 nm at low concentrations, and the fluorescence gradually increases as the concentration increases. Additionally, the fluorescence peak at 340 nm gradually disappears as the concentration increases, and the fluorescence peak maximum shifts towards visible wavelengths. Switching to a spectrometer with a wider measuring range (FX 2000), the fluorescence peak was observed near 520 nm. In this case, the fluorescence is blue-green to the naked eye. Fig. 6d shows the normalized fluorescence spectra of LC.

The oil-in-water emulsions with low concentrations (according to the experimental results, concentrations less than

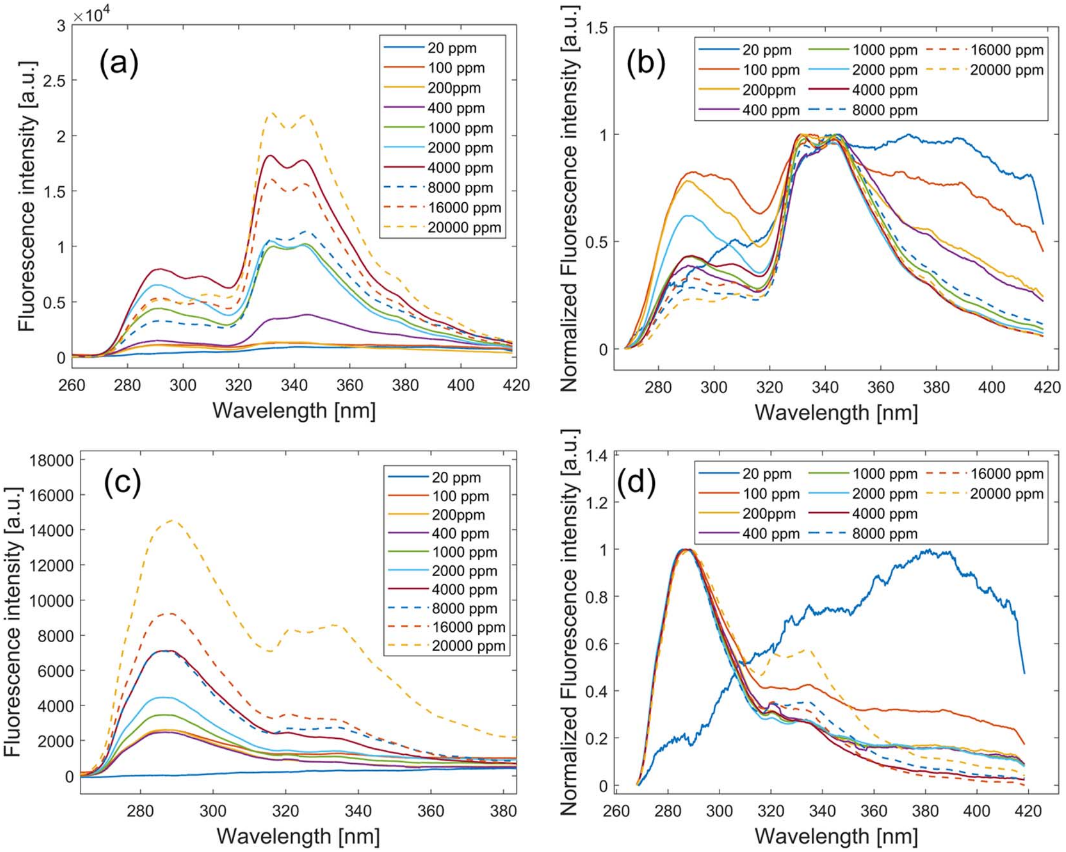

Fig. 5 Fluorescence spectra of −35# D and 95# G for different emulsion concentrations. (a) Initial fluorescence spectra of −35# D. (b) Normalized fluorescence spectra of −35# D. (c) Initial fluorescence spectra of 95# G. (d) Normalized fluorescence spectra of 95# G.

or equal to 2000 ppm) have a fluorescence peak at 290 nm in the solar-blind UV band, and the fluorescence intensity peaks of emulsions with concentrations of 1000–2000 ppm exceed the fluorescence intensity of the oil film at this wavelength. The experimental results demonstrate that solar-blind UV fluorescence can provide new ideas and application prospects for outdoor, all-day, and all-weather fluorescence monitoring.[25,31]

The comparison of selected long-wave fluorescence peaks and day-blind UV fluorescence peaks can clearly show the nonlinear trend in the fluorescence changes. As shown in Fig. 7, the ratios of the long-wave to short-wave fluorescence of the four emulsified oil samples gradually increase with increasing concentration. When the oil film on the water surface was formed, the thickness of the oil film increased to the point that the excitation light could not penetrate it, the fluorescence peak ratio tended to stabilize.

There are four main aspects of spectral changes in general:

1. Change in the fluorescence intensity;
2. Change in the number of fluorescence peaks;
3. Change in the ratio of characteristic peaks;
4. A slight shift in the fluorescence peak position.

3.2 Simultaneous prediction of oil spill types and concentrations

3.2.1 Multi-output neural network models. Regression refers to predictive modeling problems that involve predicting numerical values for a given input, while classification involves predicting the probability or label of class categories for a given input. Deep learning neural networks are commonly utilized to address both regression and classification tasks.[32] In certain predictive modeling scenarios, it is necessary to develop models for both regression and classification.[33]

The generally simple approach is to develop regression and categorical prediction models on the same data and use them sequentially. The problem with this approach is that different models can make different predictions. The other, more efficient approach is to develop a single neural network model that can predict both numbers and category labels from the same inputs at the same time, which is known as a multi-output model. The benefit of this type of model is that only one model needs to be developed and maintained instead of two, and training and updating the model for both output types at the same time may provide greater consistency in the predictions between the two output types.

Back propagation (BP), probabilistic neural network (PNN), radial basis function (RBF), and generalized regression neural network (GRNN) were selected for comparative analysis to evaluate the effectiveness of multi-output models. The structure of the multi-output GRNN is shown in Fig. 8. The number of nodes in the input layer is equal to the number of features in the input data. The pattern layer computes the Euclidean distance between the input vector and each sample, with the number of

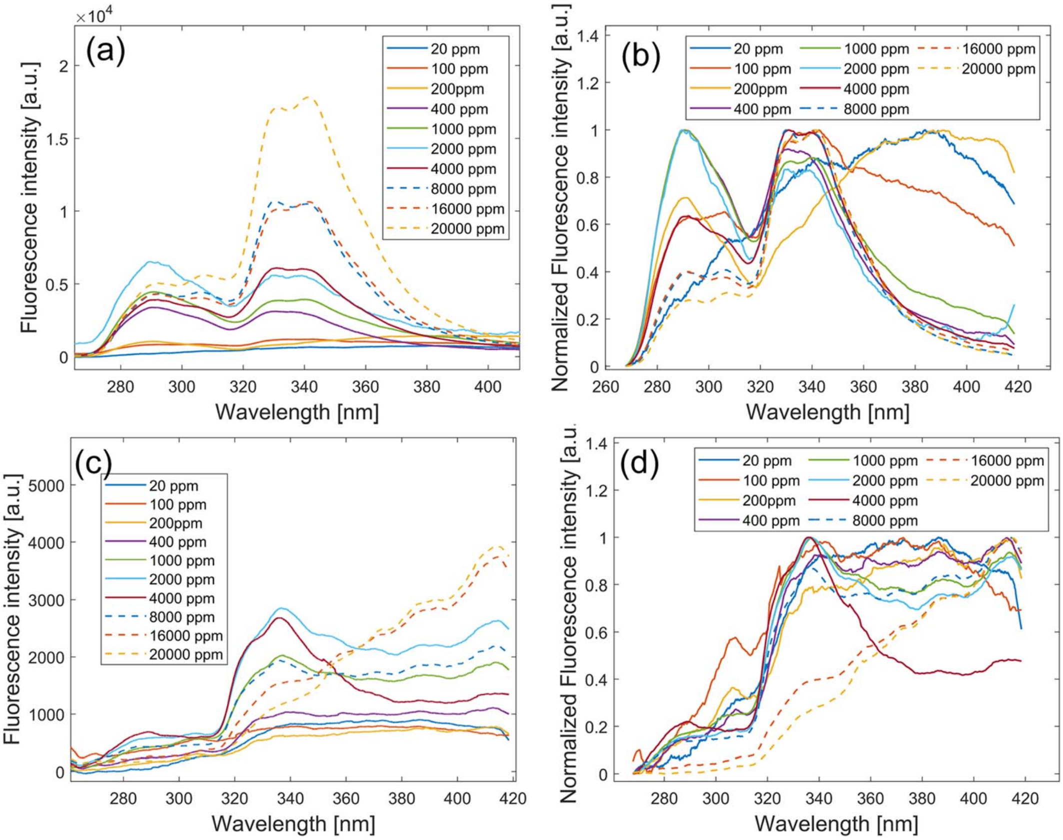

Fig. 6 Fluorescence spectra of −20# D and LC with different emulsion concentrations. (a) Initial fluorescence spectra of −20# D. (b) Normalized fluorescence spectra of −20# D. (c) Initial fluorescence spectra of LC. (d) Normalized fluorescence spectra of LC.

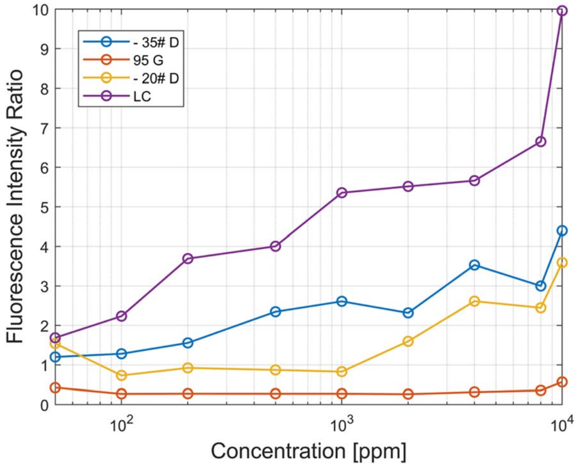

Fig. 7 Long-wave to short-wave fluorescence ratios of the four samples.

nodes being equal to the number of samples in the training data.[34] The summation layer performs the weight calculation and weighted average of the weights and the input of the pattern layer.[35] The output layer transforms the output values of the summation layer and outputs them, where Y_1 is the classification result and Y_2 is the concentration prediction result.

Four oils were measured with ten concentration samples for each oil, and 20 sets of data were collected for each concentration. The data was preprocessed to obtain a two-dimensional matrix of 41 × 800. 41 Represents the data features and 800 is the amount of data. 50% of the fluorescence data was used as the training set, and the rest was the testing set.

In the multi-output GRNN, the input layer has 41 nodes, the distribution density of the radial basis function is 0.01, and the pattern layer has 400 nodes.

3.2.2 Predicting results. The commonly used predictive evaluation indices include mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE),[36] and symmetric mean absolute percentage error (SMAPE). When the predicted value completely matches the true value, the above-mentioned evaluation indices are equal to 0, indicating a perfect model. The greater the error, the greater the corresponding value of the index.[37]

Suppose the predicted value is:

$$\hat{y} = \left\{\hat{y_1}, \hat{y_2}, \cdots, \hat{y_n}\right\} \tag{1}$$

The true value is:

$$y = \{y_1, y_2, \cdots, y_n\} \tag{2}$$

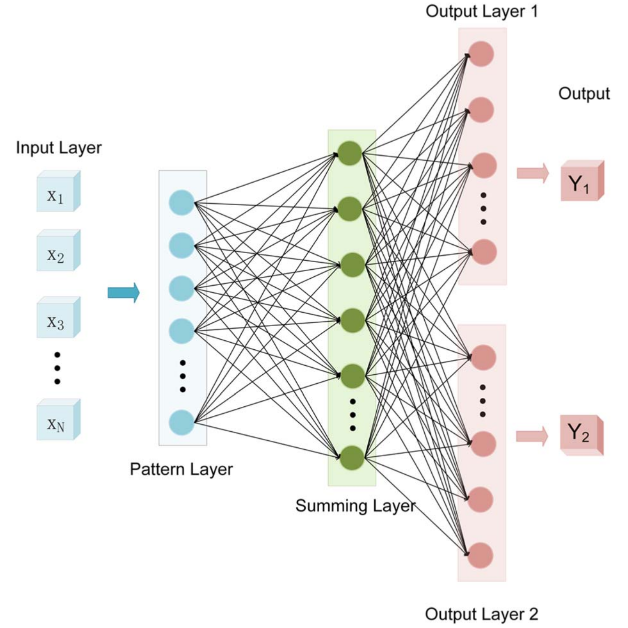

Fig. 8 Multi-output generalized regression neural network (GRNN) structure diagram.

Mean square error (MSE):

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y_i}-y_i\right)^2 \tag{3}$$

Root mean square error (RMSE):

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y_i}-y_i\right)^2} \tag{4}$$

Mean absolute error (MAE):

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y_i}-y_i\right| \tag{5}$$

Mean absolute percentage error (MAPE):

$$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y_i}-y_i}{y_i}\right| \tag{6}$$

A MAPE equal to 0 indicates a perfect model, and a MAPE greater than 100% indicates a poor model. MAPE was chosen as a regression precision index because it is independent of the magnitude of the true values. MAE is associated with concentration values and given as an indicator for comparison in the results. The training and testing time of the model was also chosen as an evaluation criterion for model performance. The indicator for classification results is accuracy.

Based on the experimental measurements of the spectral data, BP, RBF, PNN, and GRNN combined classification and regression models were established. To obtain more accurate indices, each model was tested five times. The mean value was calculated as the result, and the standard deviation was calculated to evaluate the stability.

The visual prediction effect is shown below (Fig. 9). Since the concentration data are not uniformly varying, the direct representation has some overlap in the low-concentration area. For better representation, the coordinates use numbers instead of species and real concentration.

Fig. 10 shows the regression prediction results for concentration. In order to more clearly represent the fitting effect at low concentrations, the logarithm of the concentration values is taken as the horizontal coordinates. The "Target" in the horizontal coordinate represents the logarithmic value of the true concentration. Fig. 10a shows the fitting results for the training set, Fig. 10b shows the fitting results for the validation set, Fig. 10c shows the fitting results for the test set, and Fig. 10d shows the fitting results for all data. The *R* coefficients of the regression predictions are all greater than 0.98, indicating that the regression predictions achieve a good level of effectiveness.

The classification results are shown in Table 3. The results demonstrate that multi-output BP and multi-output RBF

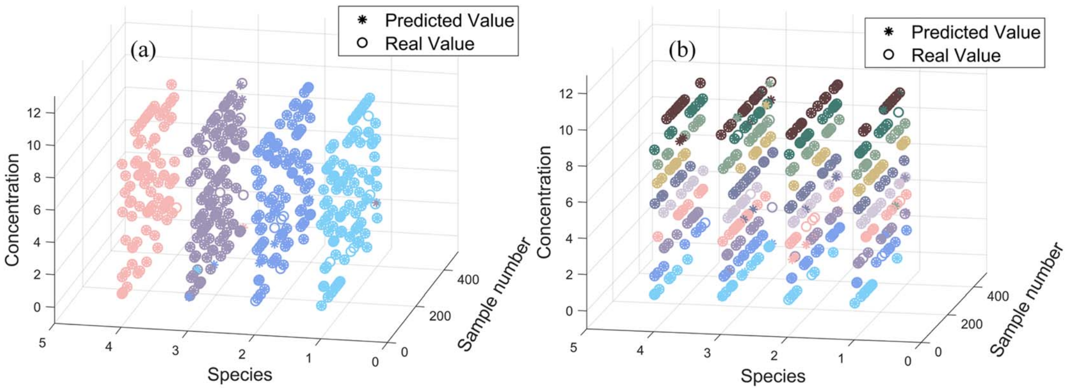

Fig. 9 Prediction of muti-output GRNN: (a) prediction of oil types, with each color representing a distinct category of oil; (b) prediction of oil concentrations, with each color representing a distinct concentration of oil.

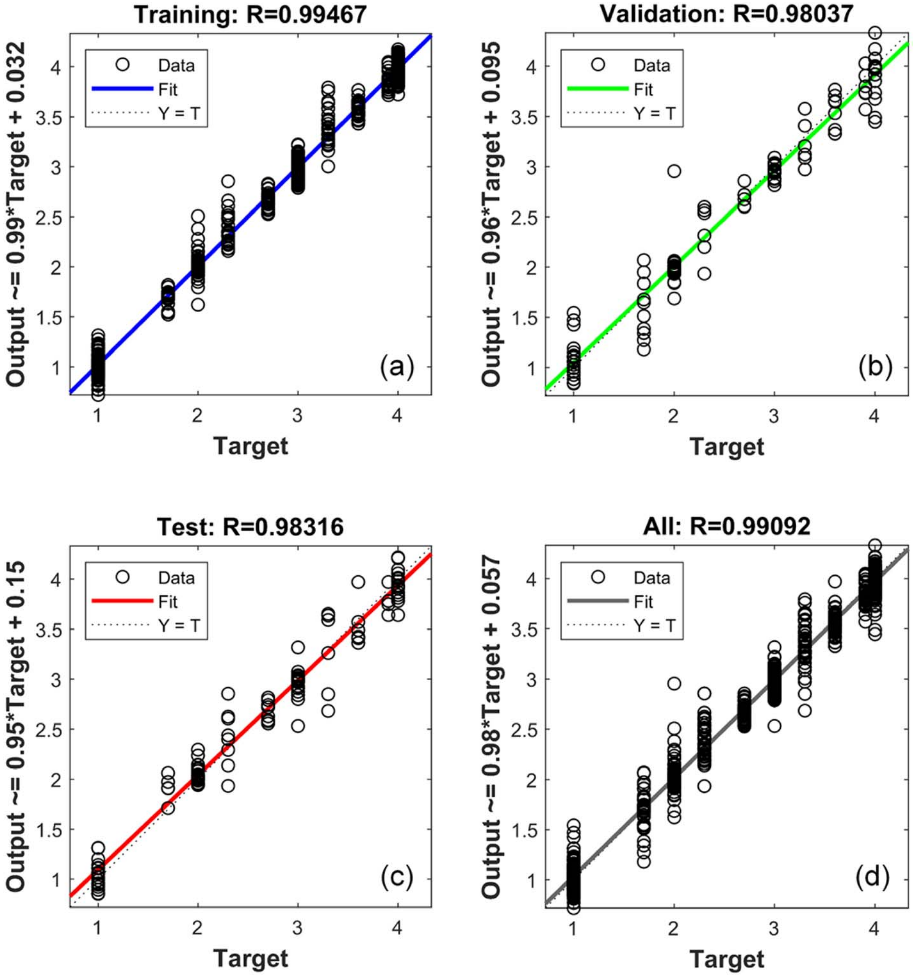

Fig. 10 GRNN regression prediction results: (a) regression results for training set; (b) regression results for validation set; (c) regression results for test set; (d) regression results for all data.

achieve an accuracy of 92% in species prediction, but the performance in regression is general. Both multi-output PNN and GRNN provide more satisfactory results in the oil species classification problem. Among them, multi-output GRNN performs best in both the classification and regression problems. Two independent models were constructed to evaluate the

Table 3 Processing results of the multi-output neural networks

	Model	Test 1	Test 2	Test 3	Test 4	Test 5	Mean value	Standard deviation
Accuracy	BP	91.00%	91.00%	95.46%	93.25%	91.01%	92.34%	1.79%
	RBF	93.50%	91.75%	92.00%	92.25%	94.25%	92.75%	0.96%
	PNN	98.25%	96.50%	98.75%	97.50%	97.50%	97.70%	0.76%
	GRNN	98.50%	98.75%	99.00%	97.75%	98.50%	98.50%	0.42%
MAE	BP	127.41	170.82	162.87	139.30	174.13	154.91	18.36
	RBF	345.83	249.51	192.40	175.79	171.46	226.99	65.62
	PNN	73.62	168.81	77.05	84.66	92.63	99.35	35.34
	GRNN	33.85	24.32	58.01	39.25	39.84	39.05	11.00
MAPE	BP	17.67%	17.19%	15.27%	20.42%	18.81%	17.87%	1.71%
	RBF	42.27%	40.27%	29.16%	25.03%	25.49%	32.44%	7.37%
	PNN	5.45%	9.32%	4.71%	6.69%	5.92%	6.42%	1.58%
	GRNN	2.16%	1.87%	1.89%	2.24%	1.92%	2.02%	0.15%
Time (s)	BP	0.866	0.902	1.25	0.878	1.086	0.996	0.150
	RBF	3.143	3.322	3.501	3.062	2.728	3.151	0.260
	PNN	0.326	0.240	0.269	0.221	0.265	0.264	0.036
	GRNN	0.038	0.029	0.029	0.033	0.027	0.031	0.004

Table 4 Processing results of normal neural networks

	Model	Test 1	Test 2	Test 3	Test 4	Test 5	Mean value	Standard deviation
Accuracy	BP	95.25%	98.50%	95.75%	93.50%	92.25%	95.05%	2.13%
	RBF	95.75%	96.00%	93.50%	95.75%	94.75%	95.15%	0.93%
	PNN	98.75%	99.00%	97.00%	98.50%	98.70%	98.39%	0.71%
	GRNN	98.25%	98.75%	98.00%	98.25%	98.50%	98.35%	0.26%
MAE	BP	158.03	147.55	212.61	170.94	188.23	175.47	23.01
	RBF	244.94	362.58	223.36	260.51	241.53	266.58	49.43
	PNN	94.70	93.42	93.06	93.24	93.99	93.68	0.60
	GRNN	71.37	69.82	35.05	43.64	42.99	52.57	15.00
MAPE	BP	15.88%	16.87%	18.49%	21.68%	22.59%	19.10%	2.63%
	RBF	38.18%	55.46%	37.29%	37.87%	38.53%	39.96%	7.22%
	PNN	4.13%	8.09%	5.59%	5.37%	5.22%	5.68%	1.31%
	GRNN	3.90%	2.86%	3.02%	2.62%	2.96%	3.07%	0.44%
Time (s)	BP	1.494	2.422	2.112	1.560	1.674	1.852	0.357
	RBF	4.412	4.393	4.822	4.605	4.127	4.472	0.232
	PNN	0.357	0.343	0.324	0.325	0.652	0.400	0.127
	GRNN	0.055	0.051	0.049	0.058	0.050	0.053	0.003

classification and regression performance for a comprehensive comparison, as indicated in Table 4. While the classification performance using the two independent models demonstrated similar results to the multi-output model, the fitting effect fell short compared to the multi-output neural network model. Furthermore, it is worth noting that these separate models required a longer training time.

4. Conclusions

An advanced multi-angle excitation and detection setup was designed and constructed for solar-blind UV fluorescence spectra analysis of oil-in-water emulsions. Using the setup, the fluorescence spectra of oil-in-water emulsions were measured by simulating the on-site detection method and the fluorescence characteristics were analyzed. The experimental results revealed that the fluorescence of the oil-in-water emulsion exhibits a blue-shift in wavelength and an increase in short-wave fluorescence intensity as the oil concentration decreases. A prediction method based on a multi-output neural network was employed to predict the emulsified oil type and concentration simultaneously. The prediction results show a high accuracy of 98.75% for oil type identification and a MAPE of 2.02% for concentration prediction. The concentration prediction yields better results within 100–16000 ppm. This method can serve as an auxiliary means for on-site fluorescence monitoring during the daytime, enabling comprehensive surveillance of oil spills.

The refined oil emulsions presented obvious solar-blind UV fluorescence under 220 nm excitation. In the case of crude oil, the inhibition of the solar-blind UV fluorescence is enhanced due to an increase in the number of aromatic hydrocarbon rings. Thus, both long-wave UV-A (320–400 nm) and medium-wave UV-B (280–320 nm) fluorescence are required for better identification. In subsequent work, a high-intensity laser light source and more oil samples will be used for further validation outdoors.

Author contributions

Bowen Gong: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, visualization. Shilei Mao: methodology, validation, review and editing, supervision. Xinkai Li: review. Bo Chen: supervision, review, project administration, funding acquisition.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the Joint Research Fund in Astronomy (grant numbers U2031122) under cooperative agreement between the National Natural Science Foundation of China (NSFC) and Chinese Academy of Science (CAS).

Notes and references

1 K. Li, H. Yu, J. Yan and J. Liao, *IOP Conf. Ser.*, 2020, **510**, 042011.

2 I. A. Silva, F. C. G. Almeida, T. C. Souza, K. G. O. Bezerra, I. J. B. Durval, A. Converti and L. A. Sarubbo, *Environ. Monit. Assess.*, 2022, **194**, 143.

3 K. Li, J. Ouyang, H. Yu, Y. Xu and J. Xu, *IOP Conf. Ser.*, 2021, **787**, 012078.

4 S. Bhattacharjee and T. Dutta, in *Advances in Oil-Water Separation*, Elsevier, 2022, pp. 3–15.

5 X. Zhang, B. Xie, M. Zhong and H. Hao, *Opt. Commun.*, 2022, **520**, 128492.

6 T. Liang, S. Qiao, Y. Chen, Y. He and Y. Ma, *Photoacoustics*, 2024, **36**, 100592.

7 Z. Lang, S. Qiao, T. Liang, Y. He, L. Qi and Y. Ma, *Opt. Express*, 2024, **32**, 379.

8 C. Fang, T. Liang, S. Qiao, Y. He, Z. Shen and Y. Ma, *Opt. Lett.*, 2024, **49**, 770.

9 Z. Asif, Z. Chen, C. An and J. Dong, *JMSE*, 2022, **10**, 762.

10 Y. Hou, Y. Li, Y. Liu, G. Li and Z. Zhang, *Mar. Pollut. Bull.*, 2019, **146**, 977–984.

11 I. B. Berlman, in *Handbook of Fluorescence Spectra of Aromatic Molecules*, Elsevier, 1971, pp. 107–415.

12 L. Sun, Y. Zhang, C. Ouyang, S. Yin, X. Ren and S. Fu, *Opt. Commun.*, 2023, **527**, 128914.

13 S. Yin, F. Sun, W. Liu, Z. Bi, Q. Liu and Z. Tian, *IEEE Sens. J.*, 2023, **23**, 13671–13679.

14 M. Jha, J. Levy and Y. Gao, *Sensors*, 2008, **8**, 236–255.

15 M. V. Bills, A. Loh, K. Sosnowski, B. T. Nguyen, S. Y. Ha, U. H. Yim and J.-Y. Yoon, *Biosens. Bioelectron.*, 2020, **159**, 112193.

16 R. Yuan and J. Ma, *China Commun.*, 2016, **13**, 63–75.

17 B. Peng, D. Jin, C. Ji, X. Chu and C. Pei, in *Ocean Optics and Information Technology*, ed. J. Yao, R. Zhang, X. Li, H. Yin, L. Wu and Z. Zhu, SPIE, Beijing, China, 2018, p. 26.

18 *Spectral Atlas of Polycyclic Aromatic Compounds*, ed. W. Karcher, Springer Netherlands, Dordrecht, 1988.

19 Y. Hou, Y. Li, G. Li, M. Xu and Y. Jia, *J. Spectrosc.*, 2021, **2021**, 1–10.

20 T. Geng, Y. Wang, X.-L. Yin, W. Chen and H.-W. Gu, *Crit. Rev. Anal. Chem.*, 2023, 1–23.

21 A. Pandey, A. Yadav, Bhawna and S. Pandey, *J. Lumin.*, 2017, **183**, 494–506.

22 E. Baszanowska and Z. Otremba, *Sensors*, 2022, **22**, 6014.

23 B. Gong, H. Zhang, X. Wang, K. Lian, X. Li, B. Chen, H. Wang and X. Niu, *Anal. Methods*, 2023, **15**, 1649–1660.

24 D. Lee, J. M. Seo, K. Kooistra and H. Lee, *Environ. Res.*, 2022, **212**, 113325.

25 A. Stelmaszewski, *Oceanologia*, 2011, **53**, 549–564.

26 *Standard Handbook Oil Spill Environmental Forensics: Fingerprinting and Source Identification*, ed. S. A. Stout and Z. Wang, Academic Press, London, 2nd edn., 2016.

27 Q. Lai, Y. Xie, C. Wang, M. Wang and J. Tan, *Appl. Opt.*, 2021, **60**, 6902.

28 J. Zhang and A. M. Mouazen, *Infrared Phys. Technol.*, 2023, **131**, 104720.

29 H. Li, Y. Liang, Q. Xu and D. Cao, *Anal. Chim. Acta*, 2009, **648**, 77–84.

30 Y. Li and D. Xiang, *PLoS One*, 2019, **14**, e0213189.

31 J. Jiao, Y. Lu and Y. Liu, *Mar. Pollut. Bull.*, 2022, **178**, 113640.

32 D. Xu, Y. Shi, I. W. Tsang, Y.-S. Ong, C. Gong and X. Shen, *IEEE Trans. Neural Netw. Learn. Syst.*, 2019, 1–21.

33 T. Son, D. Hong and B. Kim, *Sensors*, 2023, **23**, 3153.

34 B. Xu, J. Yi, X. Wan and F. Cheng, *IEEE Sens. J.*, 2023, **23**, 10776–10789.

35 S. R. Ghaffari-Razin, A. Rastbood and N. Hooshangi, *GPS Solut.*, 2023, **27**, 51.

36 K. Warneke, C.-M. Wagner, M. Keiner, M. Hillebrecht, S. Schiemann, D. G. Behm, S. Wallot and K. Wirth, *Front. Sports Act. Living*, 2023, **5**, 1105201.

37 J. A. Segovia, J. F. Toaquiza, J. R. Llanos and D. R. Rivas, *Electronics*, 2023, **12**, 1007.