Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Anal. Methods*, 2022, **14**, 3458

A field-deployable water quality monitoring with machine learning-based smartphone colorimetry†

Vakkas Doğan,[a] Tuğba Isık,[b] Volkan Kılıç *[a] and Nesrin Horzum[c]

Water quality monitoring is an increasing global concern as the pollution of water sources causes adverse effects on economic growth and human health. Traditional approaches to the detection of pollutants are time-consuming and labor-intensive due to the requirement of sophisticated equipment or laboratory settings. Therefore, portable devices featuring rapid response and easy operation are indispensable in water quality monitoring. Herein, smartphone-based colorimetric pollutant quantification is demonstrated in a machine learning (ML) framework. As a proof of concept, the presence of seven ions in water was analyzed using colorimetric strips. The color variation on the strip indicators was captured under eight lighting conditions with five smartphones, providing robustness against the illumination variation and camera optics for ML classifiers. Color and texture features were extracted from the images to train the classifiers. Among the twenty-three classifiers, K-Nearest Neighbors exhibits the best classification performance, leading to the integration with our custom-designed Android application called *Hydro Sens*. The proposed approach was also tested with real samples taken from local water sources. The results prove that incorporating color strips with ML with a smartphone application can be used for water quality monitoring, which offers promising alternatives for sophisticated equipment that is especially applicable in resource-limited settings.

Received 14th May 2022
Accepted 5th August 2022

DOI: 10.1039/d2ay00785a

rsc.li/methods

1 Introduction

The presence of a wide range of contaminants such as pesticides, pharmaceuticals, drugs, hospital effluents, and other industrial compounds has been threatening the aquatic environments at ng L^{-1} to μg L^{-1} levels. The uncontrolled growth of population and industrial activities cause the rapid consumption of natural resources, mainly water. The pollution of surface water and groundwater generally causes adverse effects on aquatic and human life with micro-pollutants entering the body through drinking water and thus threatening human health.[1,2] Each year, nearly 30 million people die because of the fatal diseases (diarrhea, cholera, dysentery, typhoid, and polio) caused by contaminated drinking water.[3] Moreover, polluted water destructs biodiversity triggering the proliferation of undesired organisms and contaminating the food chain by introducing toxic substances into foods. Therefore, the quality of water sources has become a serious concern for both environmental considerations and human health.

Water quality is specified by physical, chemical, and biological parameters. Some parameters such as electrical conductivity, pH, type and quantity of dissolved, suspended or colloidal substances can be measured onsite. However, chemical parameters such as heavy metal contamination (*e.g.*, the presence and balance of major/minor ions) require traditional monitoring systems.[4,5] On the other hand, monitoring the water quality in rural areas is difficult due to the lack of well-equipped laboratories requiring specialist labor. In addition, sample collection from different locations is another concern due to the transfer and storage problems.[6]

In water quality monitoring, there are several analytical techniques based on electrochemical, spectrometric, and colorimetric approaches.[7] A representative method in the electrochemical technique is capillary electrophoresis, and inductively coupled plasma mass spectrometry (ICP-MS) is the most widely used spectrometric technique that affords the measurement of multiple elements in water samples. However, these techniques require time-consuming sample preparation procedures and expensive instruments with expert users. Thus, they cannot be performed for on-site applications.[8]

For rapid and on-site pollutant analysis, colorimetric methods for naked-eye detection are the most convenient in terms of simplicity and portability. Basically, the concentration of analytes is determined with a reaction between the analyte

[a]*Department of Electrical and Electronics Engineering, Izmir Katip Celebi University, 35620 Turkey. E-mail: volkan.kilic@ikcu.edu.tr; Fax: +90 232 325 33 60; Tel: +90 232 329 35 35*

[b]*Department of Mineral Analysis and Technologies, General Directorate of Mineral Research and Exploration (MTA), Ankara, Turkey*

[c]*Department of Engineering Sciences, Izmir Katip Celebi University, 35620 Izmir, Turkey*

† Electronic supplementary information (ESI) available. See https://doi.org/10.1039/d2ay00785a

and color reagent. Rider *et al.* performed colorimetric nitrite analysis using 4-aminobenzenesulfonic acid and 1-aminonaphthalene as color reagents.[9] Adarsh *et al.* developed a novel colorimetric aza-BODIPY based probe for the detection of nitrite ions for on-site detection.[10] Plasmonic nanostructures (gold and silver nanoparticles) are sensitive biosensors due to surface plasmon resonance (LSPR), which leads to intense color measurement.[11] In addition, gold and silver colloids can be used for colorimetric sensor fabrication as their LSPRs are in the visible spectrum. However, the high fabrication cost of these sensors is still an issue for widespread application.

Lab-on-a-chip designs have also been proposed for the colorimetric detection of analytes in water.[12,13] However, extending the lab-on-a-chip technologies at the consumer level is limited because of the need for an external detection system to interpret the signal from the chip and provide the data for user. Portable smartphone-based detection systems provide a low-cost option for lab-on-a-chip design that smartphones are able to capture images with high resolution even under low-light conditions. Besides, they are portable, fast, and do not require special training, and their widespread usage makes this technology accessible to everyone. Smartphones can be used as a colorimeter,[14] spectrometer,[15] fluorometer,[16] and voltammeter.[17] Most often used smartphone applications rely on digital image colorimetry and image analysis.[18] In this way, the use of smartphones as chemical detectors is becoming widespread in various fields such as medicine,[19–22] food,[23,24] and environment.[25,26] Hossain *et al.* presented a "lab-in-a-phone" platform with a pH sensor concept to detect any possible contamination in drinking waters. They established an effective real-time pathogen mapping of a specific location using GPS coordinates.[27] Kılıç *et al.* proposed a local database to quantify the concentration of solutions, and they used a single image to determine four water quality parameters, which are nitrite, phosphate, chromium, and phenol concentrations. An Android application (*Chem Trainer SIR*) was tested, and the results show that the detection accuracy can reach up to 100% with the help of color matching algorithms.[28] Moreover, Gan *et al.* developed a smartphone-based colorimetric system with functionalized gold nanoparticles to detect cadmium species.[29]

For the colorimetric quantitative analysis, intelligent systems have emerged as a useful tool to process various independent variables due to their high computing performance on large amounts of data. In this regard, artificial intelligence (AI) methods have been used to enable the learning of computing devices without a need for human intervention.[30,31] The advances in AI help to improve the measurement techniques in various fields such as environmental data monitoring,[32–35] weather forecasting,[36,37] medical imaging,[38] diagnosis,[39–41] agricultural management,[42,43] and disaster prediction.[44,45] Tinelli *et al.* simulated the bio-contamination risk in the water distribution systems. They proposed an effective monitoring system that provides real-time information that reduces potential harm to citizens.[46] One of the methodologies in AI is machine learning (ML).[47] Li *et al.* successfully analyzed the wastewater samples from lagoons applying ML methods and estimated the water quality parameters such as nitrogen, phosphorus, bacteria, and total solids.[48]

In addition, smartphone-based colorimetric detection through ML allows researchers to employ handheld devices for a variety of colorimetric tests in limited-resource settings.[49] As a representative example, a smartphone platform that enables the determination of the pH levels using ML was proposed.[30] Solmaz *et al.* demonstrated a smartphone application-based colorimetric peroxide detection using ML classifiers trained in very diverse experimental settings such as different light sources and handsets.[50] In addition, colorimetric detection of enzymatic and non-enzymatic glucose using ML classifiers has been recently reported.[41,51]

In this study, a ML-based smartphone colorimetry approach is proposed to monitor water quality that can be deployed in fields and remote environments. The color change in the indicator due to the presence of chemicals (*i.e.*, ammonium, arsenic, carbonate, chloride, iron, nitrate, and sulfate) is captured with a smartphone camera to detect the concentration using a ML classifier. In this regard, the classifiers are trained separately for each ion, with a relevant dataset containing color and texture features extracted from the captured images. The K-Nearest Neighbors (KNN), which has the highest accuracy among all classifiers, is integrated into our custom-designed Android application *Hydro Sens*. The proposed approach is also tested with real samples taken from local water sources. The results prove that the proposed smartphone colorimetry has great potential for water quality monitoring in resource-limited settings and remote environments.

2 Experimental methods

2.1 Materials

All the chemicals were of reagent grade and used without further purification. Ultrapure water (Milli-Q Millipore 18.2 MΩ cm^{-1} at 25 °C) was used throughout the study.

Synthetic solutions were prepared using sodium sulfate (Merck, ≥99.0%), iron(II) chloride (Sigma-Aldrich, 98.0%), ammonium acetate (Merck, 98.0%), sodium nitrate (Supelco, ≥99.5%), calcium carbonate (Sigma-Aldrich, ≥99.0%), arsenic(V) oxide (Merck, 99.0%), and sodium chloride (Sigma-Aldrich, ≥99.0%) to identify the color codes for the system. The concentration of solutions was adjusted according to the defined range on each test strip (Fig. 1) (MQuant®) as follows: sulfate (200–400–800–1200–1600 mg L^{-1} SO_4^{2-}), iron (3–10–25–50–100–250–500 mg L^{-1} Fe^{2+}), ammonium (10–30–60–100–200–400 mg L^{-1} NH_4^+), arsenic test strips (0.005–0.010–0.025–0.05–0.10–0.25–0.50 mg L^{-1} As), nitrate test strips (10–25–50–100–250–500 mg L^{-1} NO_3^-), carbonate hardness (4–8–12–16–24 °d), and chloride (500–1000–1500–2000 ≥ 3000 mg L^{-1} Cl^-).

2.2 Experimental design and image capturing

An experimental setup was designed to create a dataset for training ML classifiers. The requirement of the classifiers was to be robust to any illumination change, leading to more accurate results regardless of the smartphone brand. Therefore, the

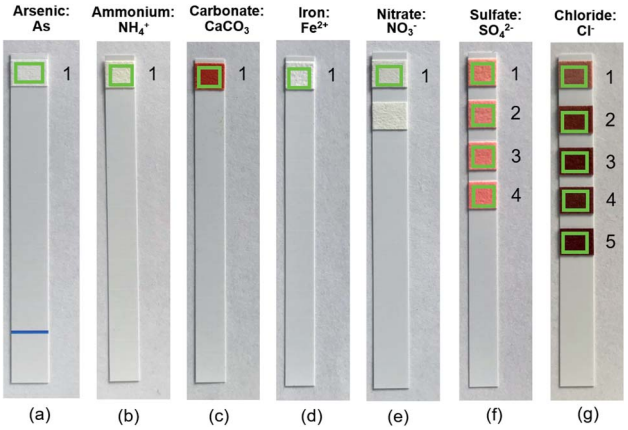

Fig. 1 The number of indicators used in the project of seven different ions.

dataset needed to be enlarged with images captured with several camera optics and various illumination sources to meet this requirement. In this sense, halogen (H), fluorescent (F), and sunlight (S) bulb sources were used to imitate the indoor lighting conditions. Besides, images were also taken in daylight (D) under the clear sky to include the outdoor conditions in the dataset.

The bulbs for indoor imaging have unique properties. The color temperature of the halogen light source (Osram 60 W) is 2700 K (warm), and the color rendering index (CRI) is 80, while the fluorescent (Klite 6 W) color temperature is 4000 K (neutral), and the CRI value is 80. The sunlight (Philips 5.5 W) bulb has a 6500 K (cold) color temperature and a CRI of 90. Seven lighting conditions (H, F, S, HF, HS, FS and HFS) were obtained with both individual and combinations of these bulbs. The distances of the bulb sources to the smartphones during imaging were kept constant at 40, 44 and 48 cm for H, F and S, respectively. In addition, images were taken at 37° angle of incidence with a distance of 15 cm between the smartphone and strips. Images were captured with a smartphone camera after the strips were dipped into the prepared synthetic solutions as illustrated in Fig. 2.

In this study, five Android (Asus Zenfone 3, HTC One A9, LG G4, Samsung Galaxy A5, and Sony Xperia T2 Ultra) smartphones with different camera properties (Table 1) were used to offer inter-phone repeatability. Furthermore, the camera settings of the smartphones were used in the automatic mode to capture forty images (5 smartphones under eight lighting conditions) for each ion (ammonium, arsenic, carbonate, chloride, iron, nitrate, and sulfate).

2.3 Feature extraction and machine learning analysis

Feature extraction is the process of obtaining distinctive features representing an object based on color, texture, size, shape, and location and is an essential step in training ML classifiers.[34,51] Here, image features are extracted based on color and texture information only, as it has been found to be adequate based on extensive experimentation. After the region of interest (ROI) was cropped, it was converted from RGB (red-green-blue) to HSV (hue-saturation-value) and L*a*b* (lightness, green-red, blue-yellow) for each concentration to obtain in R, G, B, H, S, V, L*, a*, and b* color channels separately as HSV is more robust towards external lighting changes, and L*a*b* is particularly useful for boosting colors in images due to the way it handles colors, which offers more distinguish features for the training of machine learning classifiers. Theample converted images from RGB to HSV and L*a*b* are given in Fig. S1, ESI.† Then, the mean, standard deviation, skewness, and kurtosis values were calculated for each color channel. Moreover, contrast, correlation, homogeneity, and energy were extracted as texture properties, leading to forty features being used in training. The equations for the feature extraction are given in Section S1, ESI.† In this study, all these features have been employed in training. However, the number of the features can

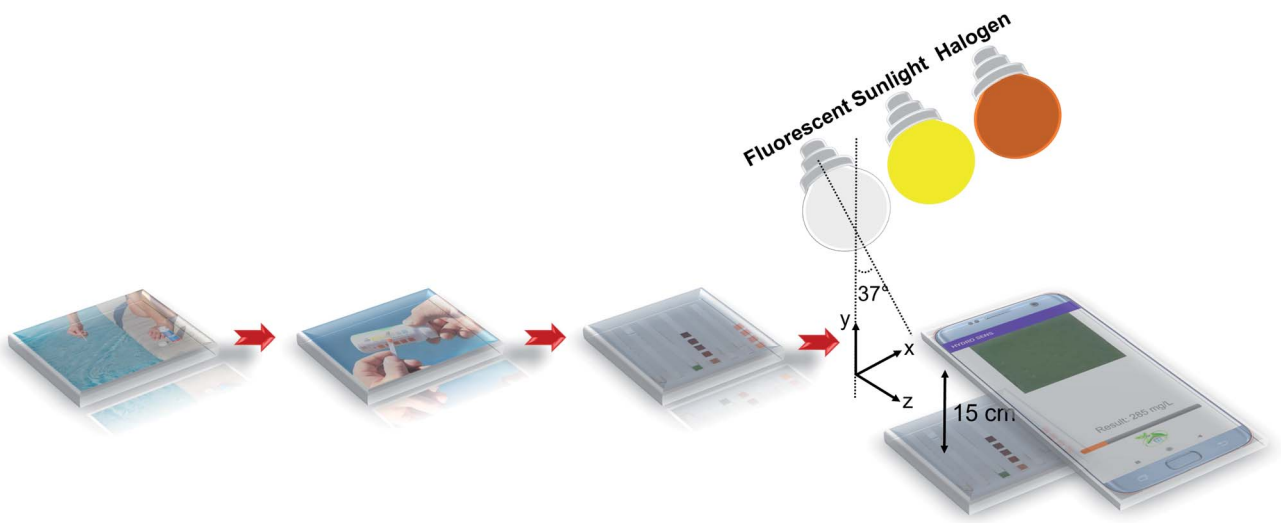

Fig. 2 Schematic illustration of the water quality monitoring approach. The color change in the indicators of the strip was imaged using a smartphone camera under various combinations of fluorescent, sunlight, and halogen sources.

Table 1 Camera properties of the smartphones used in capturing

Smartphone brand	Image resolution	Optics	Camera resolution
Asus Zenfone 3	4032 × 3024	*f*/2	16 MP
HTC One A9	4160 × 3120	*f*/2	13 MP
LG G4	5312 × 2988	*f*/1.8	16 MP
Samsung Galaxy A5	4608 × 3456	*f*/1.9	13 MP
Sony Xperia T2 Ultra	4128 × 3096	*f*/2.4	13 MP

be reduced by applying feature selection or dimensionality reduction methods.

In this study, twenty-three ML classifiers were trained in MATLAB (MathWorks, MA, USA) to determine the quality of water based on the color change caused by the ions in the solution. The *k*-fold cross-validation technique was used in training where the *k* value was chosen as 10. The performance of classifiers was measured with evaluation metrics including classification accuracy, precision, recall and the F1-score to determine the outstanding classifier. The KNN, a sample-based classification algorithm, outperformed other classifiers and therefore was integrated into our custom-designed smartphone application called *Hydro Sens*. In KNN, K closest (most similar) samples are determined from the dataset for the input sample. Then, the input sample is assigned to the class label based on the majority voting of its nearest neighbors.

2.4 Smartphone application: *Hydro Sens*

The KNN was integrated into our custom-designed Android-based *Hydro Sens* application due to its performance. *Hydro Sens*, with a simple and user-friendly interface, transfers images from a smartphone (Android) to a server (MATLAB), which runs a ML classifier to classify the relevant ion in water. The Firebase Cloud System has been used to enable communication between the smartphone and remote server as it supports both Android and MATLAB.

When the user opens the *Hydro Sens* application to determine the concentration of an ion in the water (Fig. 4(a)), the user is asked to capture a new image using the smartphone camera or select an image from the gallery. After an image is selected from the gallery (Fig. 4(b)), it is displayed on the screen to double-check as shown in Fig. 4(c). The user must tap the crop button to define the ROI on the image. The ROI is cropped using an adjustable crop box, as shown in Fig. 4(d) and (e). The cropped patch is displayed on the application screen, as shown in Fig. 4(f), and is sent to MATLAB *via* Firebase with the upload button. The KNN classifies ions based on color and texture features. Fig. 4(g) shows that the ion information is sent after tapping the "APPLY NOW" button, which initiates the classification process in the remote server. The result is sent back to *Hydro Sens via* Firebase, as shown in Fig. 4(h).

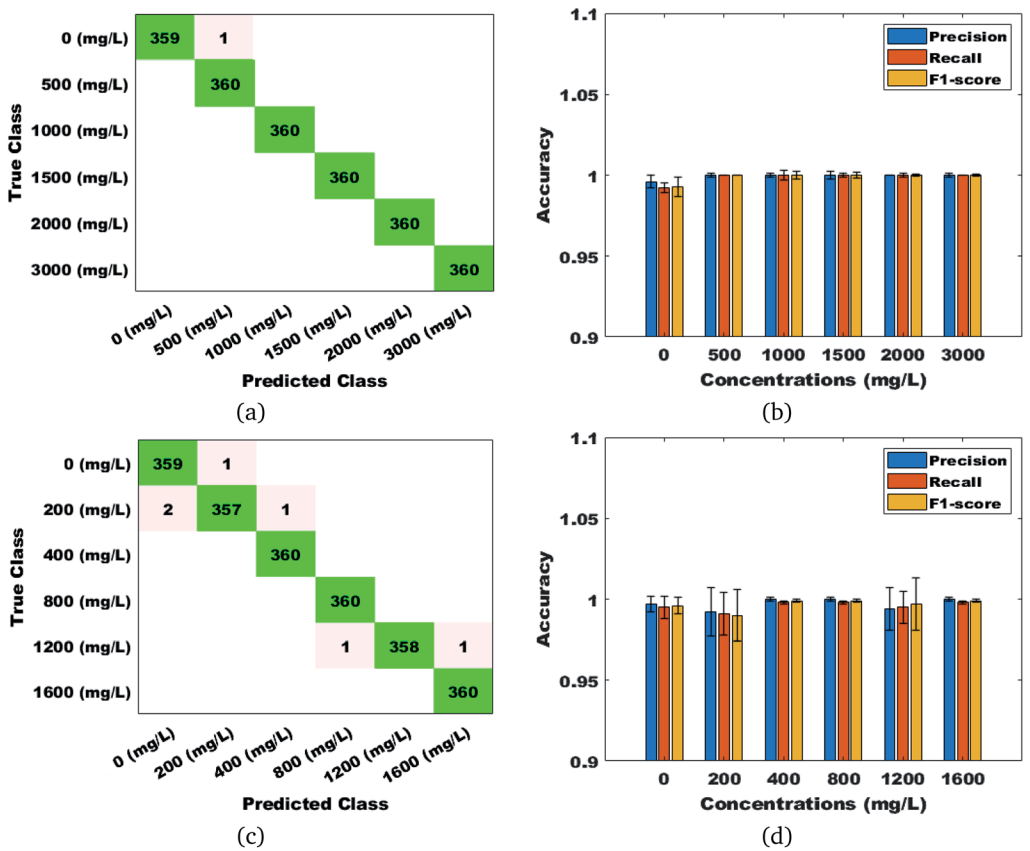

Fig. 3 Confusion matrices and error bars of chloride (a and b) and sulfate (c and d) at different concentrations.

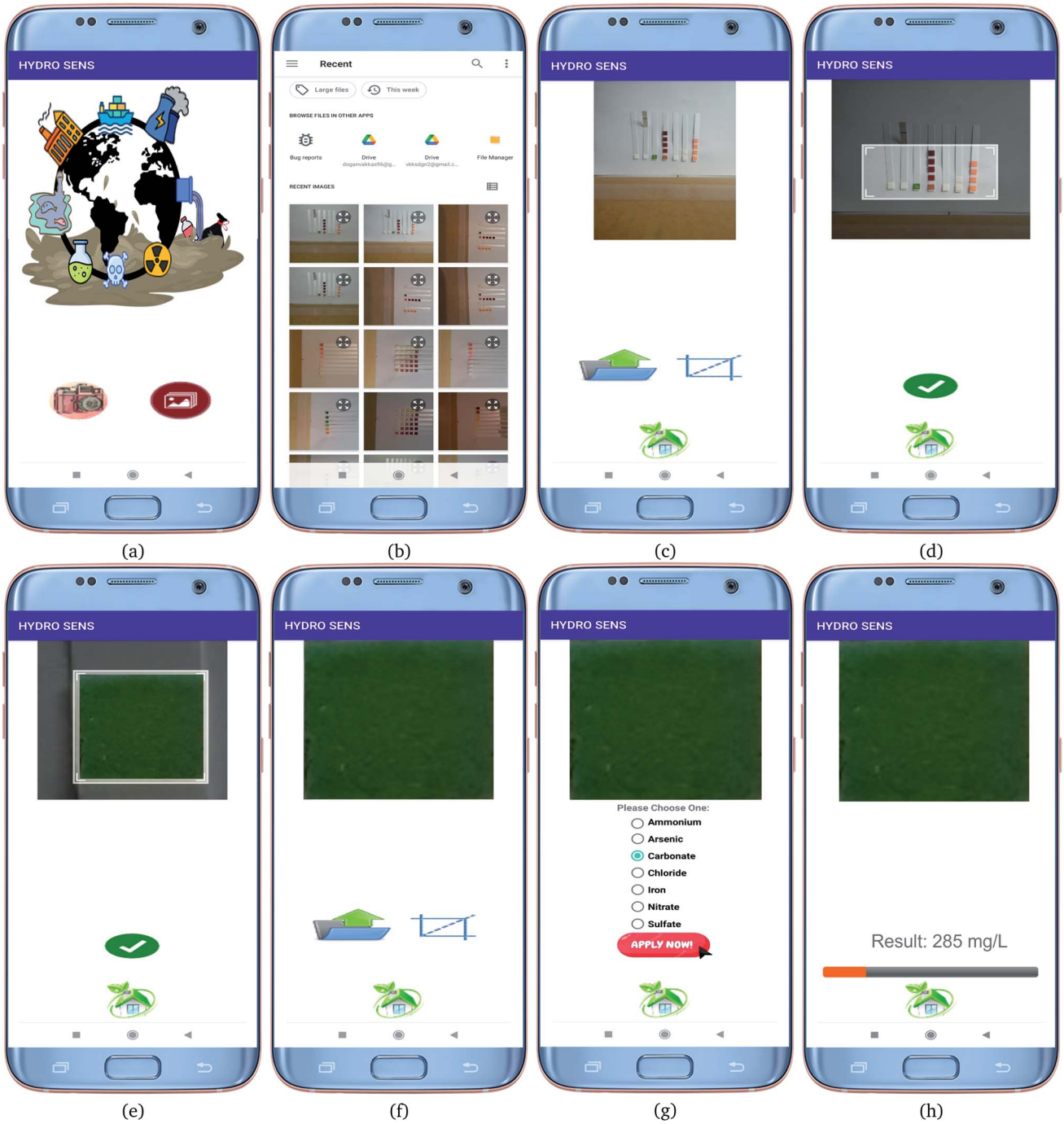

Fig. 4 Demonstration of colorimetric detection steps on the *Hydro Sens* application is presented. The homepage of the application is shown in (a). An image is selected from the gallery or captured with a smartphone camera as shown in (b). The selected image is displayed in (c). An adjustable crop box is used to crop the image in (d), and (e) shows the cropped patch, which is uploaded in (f) to be analyzed for the ion in (g). Finally, the result is given in (h).

2.5 Instrumentation and analysis

The concentrations of As and Fe ions were determined using an Inductively Coupled Plasma-Mass Spectrometer (ICP-MS) (Agilent 7500ce Series, Japan). The ICP-MS operation parameters are given in Table 2. Four main stock solutions (10 mg L^{-1}, 1 mg L^{-1}, 100 μg L^{-1}, and 10 μg L^{-1}) were prepared by dilution of 1000 mg L^{-1} As(v) and Fe(II) stock solutions. Then, standard solutions of varying concentrations between 0.05 μg L^{-1} and 1000.0 μg L^{-1} were prepared by appropriate dilution of the main stock solutions. The germanium internal standard was used in ICP-MS analysis for the samples and standard solutions at the same concentration to improve the precision of analysis. Afterwards, all solutions were acidified with HNO_3 to achieve 1.0% (v/v) acid in the final solution.

Table 2 ICP-MS operation parameters

Operation parameters	Value
RF power	1550 W
Frequency	27 MHz
RF matching	1.78 V
Makeup glass flow rate	0.15 L min^{-1}
Argon carrier gas flow rate	0.95 L min^{-1}
Chamber temperature	2 °C
Sample depth	8 mm
Nebulizer	Concentric
Nebulizer pump	0.1 rps
Tune setting (m/z)	7/89/205
Isotopes monitored (m/z)	^{75}As and ^{56}Fe

Ion chromatography (IC) with a conductivity detector was used to determine ammonium, carbonate, chloride and nitrate concentrations (Dionex ICS-5000+ Ion Chromatography System, Thermo Scientific). The samples were filtered through PTFE filters with 0.45 μm pore size, stored in polyethylene bottles and analysed directly without further processing. A Dionex IonPac AG19 guard column (4 × 50 mm) and AS19 analytical column (4 × 250 mm) were used for anion analysis, whilst a Dionex IonPac CG16 guard column (5 × 50 mm) and CS16 analytical column (5 × 250 mm) were used for cation analysis. The chromatographic run was performed with 1 mL min^{-1} flow rate and 20 μL injection volume. The column temperatures were 30 °C and 40 °C, and the eluents were methanesulfonic acid (30 mM) and potassium hydroxide (10 mM) for anion and cation analyses, respectively. The natural water samples contain various acid–base pairs that can accept/donate protons. Total alkalinity is the ability of an aqueous sample to neutralize an acid. The alkalinity of most waters comes from carbonates, bicarbonates, and hydroxides.[52] The samples were collected in polyethylene bottles. For the standard solution, 0.1 N sulphuric acid (H_2SO_4) was prepared against 0.05 N sodium carbonate (Na_2CO_3). The analysis was performed using an automatic potentiometric titrator (AT-510, Kyoto Electronics Manufacturing, Japan) and an automatic piston burette (APB-510, Kyoto Electronics Manufacturing, Japan). 100 mL of each sample were taken and titrated with 0.2 N standard (H_2SO_4), and the results were calculated from the endpoints of the samples.

3 Results and discussion

Here, a ML-based colorimetric detection was employed using color variation in test strips caused by seven different ions. A comprehensive dataset was created as the capacity has a positive effect on the accuracy of the classifier. To create the dataset, test strips of seven different ions were sequentially immersed in water and then removed. The color changes in the strips were captured with five different brands of smartphones under eight different lighting conditions. The images were then transferred to the computer for preprocessing and training in the MATLAB environment. The strips may have a different number of indicators as given in Fig. 1. Test strips for ammonium, arsenic, carbonate, iron, and nitrate ions have one indicator, while sulfate and chloride have four and five indicators, respectively. The highlighted region on the indicator (Fig. 1) was cropped for feature extraction. Then, the dataset was created for each ion using color and texture information as described in Section 2.3. Twenty-three ML classifiers were trained with these datasets, and the KNN outperforms others for all ions. The comparative evaluation of KNN with other classifiers is given in Table S1, ESI.†

The performance of the classifiers was compared with classification accuracy (eqn (1)), precision (eqn (2)), recall (eqn (3)), and the F1-score (eqn (4)).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4)$$

True Positive (TP) is the positive class label correctly predicted by the classifier, while True Negative (TN) is the negative class label correctly predicted by the classifier. False Negative (FN) is defined as the negative class label that was incorrectly predicted by the classifier, while False Positive (FP) is referred to as the positive class label incorrectly predicted by the classifier.[53] The accuracy value is calculated from the ratio of the correctly predicted labels in the model to the total dataset. Precision refers to how many positively predicted values are actually positive.[54] Recall, on the other hand, is a measure of how many positively predicted labels are positive. The F1-score represents the harmonic mean of precision and recall values.

The classification accuracy and limit of detection (LOD) for each ion are given in Table 3 and the accuracy in chloride and sulfate is over 99%. The LOD was calculated to be $3.3\sigma/m$ where σ is the standard deviation of the intercept, and m is the slope of the calibration curve derived from the features. A confusion matrix shows the classification performance in terms of true–false predictions of the classifier as a table. The confusion matrix and error bar graphs of chloride and sulfate at different concentrations are given in Fig. 3(a–d), respectively. In addition,

Table 3 Results of the highest classification accuracy in KNN for seven different ions in the water

Ions	Classification accuracy (%)	LOD (mg L^{-1})
Ammonium	98.69	7.85
Arsenic	95.63	0.04
Carbonate	98.10	15.88
Chloride	99.95	479.86
Iron	97.88	2.27
Nitrate	98.45	6.41
Sulfate	99.72	100.16

Table 4 Classification accuracy of the KNN with respect to the visual test and IC device results for water samples taken from different districts of Izmir

	Ammonium		Carbonate		Chloride		Nitrate	
	Visual test (%)	IC (%)	Visual test (%)	IC (%)	Visual test (%)	IC (%)	Visual test (%)	IC (%)
Balcova	52.75	87.75	69.80	82.65	97.85	97.85	44.75	67.25
Buca	87.75	87.75	72.75	79.75	92.75	92.75	84.85	84.75
Kemalpasa	82.75	82.75	60.25	60.25	79.75	79.75	82.65	82.65

the chloride confusion matrix in Fig. 3(a) shows that our proposed approach only makes an error in the 0 mg L^{-1} concentration value, indicating the reliability of our approach. Evaluations of the KNN for other five ions (ammonium, carbonate, nitrate, arsenic and sulfate) in terms of precision, recall and the F1-score are given in Tables S2–S6 and Fig. S2–S6, ESI.† In addition, confusion matrices are given in Fig. S7–S11, ESI.†

In order to test the robustness, applicability, and adaptability of the proposed approach, samples were taken from water sources in 3 different districts (Balcova, Buca, and Kemalpasa) of Izmir province. To perform qualitative comparative analysis, the classification results of these samples had to be validated with the reference results. In this sense, these samples were also measured by a visual test using a reference card and IC device. The classification accuracy was calculated with respect to the visual test and IC device, and the results are given in Table 4. The raw concentration data with illumination conditions for each sample are given in Tables S7–S9, ESI.† The sample Images of ions given in Table 4 under halogen, sunlight and fluorescent illumination conditions are given in Table S10, ESI.†

The proposed approach showed high performance for four ions (ammonium, carbonate, chloride, and nitrate); however, arsenic, iron, and sulfate ions could not be detected due to their presence in low amounts in the water. Table 4 shows that the visual test scores are lower than those of the IC device, proving that the performance of the proposed classifier is closer to the IC device. Note that the visual test is more prone to error due to human-inference while the IC device gives more accurate results as expected. Therefore, it can be concluded that the proposed approach easily outperforms the visual test, although it may perform similarly to the IC device for some ions such as chloride.

Finally, the KNN classifier was integrated with *Hydro Sens*, a simple and user-friendly mobile application to monitor water quality. Screen-shots of the application are shown in Fig. 4. First, the image is selected from the gallery or captured using the camera. The ROI is cropped to transfer *via* Firebase to the remote server (MATLAB) running the ML classifier. The concentration level of the ions in water is classified, and the result is returned and displayed in *Hydro Sens* application. Finally, *Hydro Sens* correctly classified the carbonate ion concentration as 285 mg L^{-1}, as shown in Fig. 4(h). The flow chart of the proposed water monitoring method is also given in Fig. S12, ESI.†

One method to improve the sensitivity of machine learning is to expand the dataset. Images can be captured under more illumination sources with different brands of smartphones and operating systems. Data augmentation can then be used for further expansion. Next, feature selection algorithms can be utilized to reduce the number of features in training and testing. With these two methods, both sensitivity and robustness of the machine learning can be optimized.

The recent innovations in environmental monitoring, medical diagnosis, safety/quality control, and other applications are based on integrating smartphone technology with the Internet of Things, AI (*e.g.*, machine learning, deep learning, neural network, and natural language processing), sixth-generation networks, imaging algorithms, and wearable sensors.[14,18] A smartphone-based oblique incidence reflectometer employs image analysis to measure optical properties of tissues which is brand-dependent, caused to be limited for input variations.[19] A common approach in colorimetric analysis to use a calibration curve.[20–22,25] However, the calibration curve performs better only in a controlled environment. The change in conditions may cause a deviation in the results as the equation derived from the calibration curve is condition-dependent like illumination and camera sensors. To overcome this issue, statistical methods have been employed in discrimination of rice varieties.[23] However, statistical methods are not robust enough as AI based methods. Here, machine learning has been employed in the proposed water monitoring approach to offer robustness against the illumination variance and camera optics, leading to inter-phone repeatability.

4 Conclusion

This study proposes a smartphone-based machine learning approach to monitor the water quality with colorimetric strips. The classifiers were trained using features extracted from the images captured with five smartphones under eight lighting conditions, which improved the robustness against the illumination variance and camera optics, providing inter-phone repeatability. Among the tested classifiers, KNN outperformed the others in classification accuracy, with the lowest accuracy of 95.63% for arsenic and the highest accuracy of 99.95% for chloride. In addition, the proposed approach was evaluated with real water samples using a reference color card and an IC system as ground truth measurement. The results showed that the proposed approach showed comparative performance with

an IC system in Balcova district samples (97.85%, chloride). Moreover, the proposed approach was integrated with our custom-designed Android application (*Hydro Sens*), offering great potential for monitoring water quality in remote settings without advanced equipment.

Author contributions

Vakkas Doğan: methodology, investigation, and writing: original draft; Tuğba Isık: methodology, validation, and writing: original draft; Volkan Kılıç: conceptualization, supervision, writing: original draft and writing – review & editing; Nesrin Horzum: methodology, validation, supervision, writing: original draft and writing – review & editing.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgements

Dr M. Ertuğrul Solmaz is acknowledged for his initial and tentative ideas behind this research. The authors also thank the Izmir Institute of Technology Environmental Development, Application, and Research Center for the ICP-MS and IC analyses.

Notes and references

1 A. R. Ribeiro, O. C. Nunes, M. F. Pereira and A. M. Silva, *Environ. Int.*, 2015, **75**, 33–51.
2 A. Jurado, E. Vàzquez-Suñé, J. Carrera, M. L. de Alda, E. Pujades and D. Barceló, *Sci. Total Environ.*, 2012, **440**, 82–94.
3 W. Yan, J. Li and X. Bai, *Chemom. Intell. Lab. Syst.*, 2016, **155**, 26–35.
4 F. Edition, *WHO Chron.*, 2011, **38**, 104–108.
5 H. Z. Abyaneh, *J. Environ. Health Sci. Eng.*, 2014, **12**, 1–8.
6 S. Srivastava, S. Vaddadi and S. Sadistap, *Appl. Water Sci.*, 2018, **8**, 1–13.
7 S. N. Zulkifli, H. A. Rahim and W.-J. Lau, *Sens. Actuators, B*, 2018, **255**, 2657–2689.
8 H. Wei, S. M. H. Abtahi and P. J. Vikesland, *Environ. Sci.: Nano*, 2015, **2**, 120–135.
9 B. F. Rider and M. G. Mellon, *Ind. Eng. Chem.*, 1946, **18**, 96–99.
10 N. Adarsh, M. Shanmugasundaram and D. Ramaiah, *Anal. Chem.*, 2013, **85**, 10008–10012.
11 J. J. Mock, M. Barbic, D. R. Smith, D. A. Schultz and S. Schultz, *J. Chem. Phys.*, 2002, **112**, 6755.
12 M. F. Khanfar, W. Al-Faqheri, A. Al-Halhouli, *et al.*, *Sensors*, 2017, **17**, 2345.
13 A. Catini, R. Capuano, G. Tancredi, G. Dionisi, D. Di Giuseppe, J. Filippi, E. Martinelli and C. Di Natale, *Sensors*, 2022, **22**, 444.
14 S. Qian, Y. Cui, Z. Cai and L. Li, *Biosens. Bioelectron.: X*, 2022, 100173.
15 H. J. S. de Oliveira, P. L. de Almeida Jr, B. A. Sampaio, J. P. A. Fernandes, O. D. Pessoa-Neto, E. A. de Lima and L. F. de Almeida, *Sens. Actuators, B*, 2017, **238**, 1084–1091.
16 L. M. Azad, H. Ehtesabi and A. Rezaei, *Nano-Struct. Nano-Objects*, 2021, **26**, 100722.
17 D. Ji, Z. Shi, Z. Liu, S. Low, J. Zhu, T. Zhang, Z. Chen, X. Yu, Y. Lu, D. Lu, *et al.*, Smartphone-based square wave voltammetry system with screen-printed graphene electrodes for norepinephrine detection, *Smart Mater. Med.*, 2020, **1**, 1–9.
18 T. Alawsi and Z. Al-Bawi, *Eng. Rep.*, 2019, **1**, e12039.
19 Y. Cao, T. Zheng, Z. Wu, J. Tang, C. Yin and C. Dai, *Opt. Commun.*, 2021, **489**, 126885.
20 J. Li, T. Liu, R. A. Dahlgren, H. Ye, Q. Wang, Y. Ding, M. Gao, X. Wang and H. Wang, *Anal. Chim. Acta*, 2022, **1204**, 339703.
21 Y. Li, S. Liu, X. Yin, S. Wang, Y. Tian, R. Shu, C. Jia, Y. Chen, J. Sun, D. Zhang, *et al.*, *Biosens. Bioelectron.*, 2022, **210**, 114289.
22 T. Alawsi, G. P. Mattia, Z. Al-Bawi and R. Beraldi, *Sens. Bio-Sens. Res.*, 2021, **32**, 100404.
23 M. Arslan, M. Zareef, H. E. Tahir, Z. Guo, A. Rakha, H. Xuetao, J. Shi, L. Zhihua, Z. Xiaobo and M. R. Khan, *Food Chem.*, 2022, **368**, 130783.
24 J. Nelis, A. Tsagkaris, M. Dillon, J. Hajslova and C. Elliott, *TrAC, Trends Anal. Chem.*, 2020, **129**, 115934.
25 L. Zhang, D. Huang, P. Zhao, G. Yue, L. Yang and W. Dan, *Spectrochim. Acta, Part A*, 2022, **269**, 120748.
26 S. Satyam and V. Sharma, *Appl. Water Sci.*, 2021, **11**, 1–8.
27 M. A. Hossain, J. Canning, S. Ast, P. J. Rutledge and A. Jamalipour, *Photonic Sens.*, 2015, **5**, 289–297.
28 V. Kilic, G. Alankus, N. Horzum, A. Y. Mutlu, A. Bayram and M. E. Solmaz, *ACS Omega*, 2018, **3**, 5531–5536.
29 Y. Gan, T. Liang, Q. Hu, L. Zhong, X. Wang, H. Wan and P. Wang, *Talanta*, 2020, **208**, 120231.
30 A. Y. Mutlu, V. Kılıç, G. K. Özdemir, A. Bayram, N. Horzum and M. E. Solmaz, *Analyst*, 2017, **142**, 2434–2441.
31 Ö. B. Mercan and V. Kılıç, *International Conference on Intelligent and Fuzzy Systems*, 2020, pp. 1276–1283.
32 M. W. Gardner and S. Dorling, *Atmos. Environ.*, 1998, **32**, 2627–2636.
33 A. Thessen, *One Ecosyst.*, 2016, **1**, e8621.
34 V. Doğan, E. Yüzer, V. Kilic and M. Şen, *Analyst*, 2021, 7336–7344.
35 A. Y. Mutlu and V. Kılıç, *26th Signal Processing and Communications Applications Conference*, SIU, 2018, pp. 1–4.
36 S. E. Haupt, J. Cowie, S. Linden, T. McCandless, B. Kosovic and S. Alessandrini, *2018 IEEE 14th International Conference on e-Science (e-Science)*, 2018, pp. 276–277.
37 S. Suksri and W. Kimpan, *2016 International Computer Science and Engineering Conference (ICSEC)*, 2016, pp. 1–7.
38 H. Fujita, T. Katafuchi, T. Uehara and T. Nishimura, *J. Nucl. Med.*, 1992, **33**, 272–276.
39 G. Schwarzer, W. Vach and M. Schumacher, *Stat. Med.*, 2000, **19**, 541–561.
40 T. Golcez, V. Kilic and M. Sen, *Anal. Sci.*, 2021, **37**, 561–568.
41 V. Kılıç, O. B. Mercan, M. Tetik, O. Kap and N. Horzum, *Anal. Sci.*, 2021, 21P253.

42 S. Wolfert, L. Ge, C. Verdouw and M.-J. Bogaardt, *Agric. Syst.*, 2017, **153**, 69–80.
43 B. Kilic, V. Dogan, V. Kilic and L. N. Kahyaoglu, *Int. J. Biol. Macromol.*, 2022, **209**, 1562–1572.
44 R. I. Ogie, J. C. Rho and R. J. Clarke, *2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, 2018, pp. 1–8.
45 R. Jemli, N. Chtourou and R. Feki, *Panoeconomicus*, 2010, **57**, 43–60.
46 S. Tinelli and I. Juran, *Water Supply*, 2019, **19**, 1785–1792.
47 L. Dong, Q. Yang, R. H. Zhang and W. B. Wei, *eClinicalMedicine*, 2021, **35**, 100875.
48 Y. Li, X. Wang, Z. Zhao, S. Han and Z. Liu, *Water Res.*, 2020, **172**, 115471.
49 Ö. Kap, V. Kılıç, J. G. Hardy and N. Horzum, *Analyst*, 2021, **146**, 2784–2806.
50 M. E. Solmaz, A. Y. Mutlu, G. Alankus, V. Kilic, A. Bayram and N. Horzum, *Sens. Actuators, B*, 2018, **255**, 1967–1973.
51 Ö. B. Mercan, V. Kılıç and M. Şen, *Sens. Actuators, B*, 2021, **329**, 129037.
52 D. A. Wolf-Gladrow, R. E. Zeebe, C. Klaas, A. Körtzinger and A. G. Dickson, *Mar. Chem.*, 2007, **106**, 287–300.
53 D. Chicco and G. Jurman, *BMC Genom.*, 2020, **21**, 1–13.
54 V. Doğan and V. Kılıç, *J. Artif. Intell. Health Sci.*, 2021, **1**, 14–19.

