Journal of Water & Health

IWA PUBLISHING

Journal of Water and Health Vol 20 No 5, 803 doi: 10.2166/wh.2022.302

Determination of salt concentration in water using decision trees and electromagnetic waves

Ebru Efeoglu [a] and Gurkan Tuna [b,*]

[a] Department of Software Engineering, Kütahya Dumlupınar University, Kütahya, Turkey
[b] Department of Computer Technologies, Trakya University, Edirne, Turkey
*Corresponding author. E-mail: gurkantuna@trakya.edu.tr

EE, 0000-0001-5444-6647; GT, 0000-0002-6466-4696

ABSTRACT

Salt water adversely affects human health and plant growth. In parallel with the increasing interest in non-contact determination of salt concentration in water, a novel approach is proposed in this study. In the proposed approach, *S* parameter measurements, which show the scattering properties of electromagnetic waves, are used. First, the relationship between salt concentration in water and permittivity values, a distinguishing feature for liquids, is shown. Then, based on the derived correlations from a set of *S* parameter measurements, it is shown that the salt concentration in water can be predicted. Finally, after exactly determining the relations of permittivity, salt concentration and *S* parameter, a system that allows non-contact determination of salt concentration is proposed. Since the proposed system makes its prediction using a classifier, decision tree algorithms are employed for this purpose. In order to evaluate the appropriateness and success of the algorithms, a set of classification experiments were held using various water samples with different levels of salt concentration. The results of the classification experiments show that the Hoeffding tree algorithm achieved the best results and is the most suitable decision tree algorithm for determining the salt concentration of liquids. For this reason, the proposed non-contact approach can be used to determine the salt concentration in water reliably and quickly if its hardware and software components can be embedded into a prototype system.

Key words: accuracy, decision trees, electromagnetic waves, Hoeffding tree, *S* parameters, salt concentration

HIGHLIGHTS

- Showing the relationship between salt concentration in water and permittivity values.
- Predicting salt concentration.
- Proposing a system that allows non-contact determination of salt concentration.

1. INTRODUCTION

The effect of salt water on human health and life can be examined in two ways. First, the salt concentration in drinking water causes direct health problems, and second, the salt concentration in irrigation waters and water used in agriculture and animal husbandry causes damage to plants, soil, animals and ecosystems. In (Deniz & Kadioglu 2021) the geochemistry of salts and in the study by Jabed *et al.* (2020), the effects of water salinity on human health were investigated. Jabed *et al.* (2020) showed the relationship between drinking water salinity and hypertension and excessively diluted urine. It was shown that the high salinity level in drinking water affects human health and may pose a significant health risk, especially for diabetic and renal dialysis patients who need to control their daily salt intake (Ekinci *et al.* 2011). Evaluation of water quality and associated health risk was performed in the Terme River, Turkey, using physicochemical quality indices and multivariate analysis (Ustaoğlu *et al.* 2021).

Salt water has effects on agriculture and plants as well as human health. Salinity is one of the environmental factors affecting the growth of plants; because water with high salinity damages plants (Bartels & Sunkar 2005). Soils with high salinity are barren soils and inhibit the growth of plants; because salinity reduces the ability of plants to take up water (Fipps 2003). In

addition, excess salt in the transpiration stream damages the cells in the transpiration leaves (Munns & Tester 2008). In addition, excessive salinity damages cells in sweaty leaves (Munns & Tester 2008).

The amount of salt in water can be determined by ultrasonic waves. On the other hand, it is known that the speed of ultrasonic waves changes according to the salt concentration and temperature (Wong & Zhu 1995). The absorption and fluorescence properties of salts and their aqueous solutions were investigated using UV-vis spectroscopy and spectrofluorometry (Chai 2008). Since the refractive index of an environment increases with the salinity level, salinity sensors based on the refractive index was also developed (Quan & Fry 1995).

Optical techniques based on various fibre optic sensors were proposed for *in situ* monitoring of water salinity (Zhao & Liao 2002; Meng 2014). For this purpose, separate optoelectronic components such as laser source, photodetector, microcontroller and microprocessor are used, but such detection systems are costly. In addition, a separate communication setup is required for the transfer of on-site data to the laboratory. It is important to determine whether there is salt in water and if yes how much salt it contains with low-cost systems. Therefore, an effective optoelectronic system was proposed for remote sensing of salinity (Villarreal *et al.* 2018).

Microstrip antennas are widely used in many applications such as data communications, medicine (Daliri 2010) and agriculture (Yahaya 2012). However, in this study, a microstrip patch antenna was used during a process carried out to determine the salt concentration in water. To do this, first, the relationship between the salt concentration in water and the permeability values, which is a distinguishing feature for liquids, was investigated. This relationship led to the understanding that there may be a relationship between salt concentration and scattering (S) parameters. Therefore, a measurement system was designed to measure the S parameters. The measurement system consists of a microstrip patch antenna designed and manufactured in this study and a commercial Vector Network Analyser (VNA). The S parameters of the liquids, i.e. different brand waters and water with different salt concentration levels, classified in this study were measured with the VNA and the antenna. After deriving the relationship between the permeability value obtained from the measurements and the S parameter, the relationship between the S_{max} value and the salt concentration was determined using the curve fitting approach.

The approach proposed in this study was designed to allow a quick, automated and non-contact determination of salt concentration. Such an automated process requires a software application relying on a machine learning algorithm embedded to the designed system. In this study, decision tree algorithms were preferred as classifiers and their performances were compared using well-known performance metrics in order to make a fair comparison. The results obtained in the performance evaluation part of this study showed that the Hoeffding tree algorithm is the most suitable algorithm to determine the salt concentration in water. The remainder of this paper is as follows. Experimental setup and methodology used in this study are presented in Section 2. Classifiers used by the proposed system and metrics employed to evaluate the results are explained in Section 3. The results are reported in Section 4. Finally, this paper is concluded in Section 5.

2. EXPERIMENTAL SETUP AND METHODOLOGY

From a hardware point of view, the experimental setup consists of a VNA and an antenna (Figure 1). A VNA consists of a signal source, a receiver and a display. In order to make a measurement, a signal is sent from the source and the circuit is triggered with this signal. The response signal consists of the reflected signal and the transfer signal received from the circuit

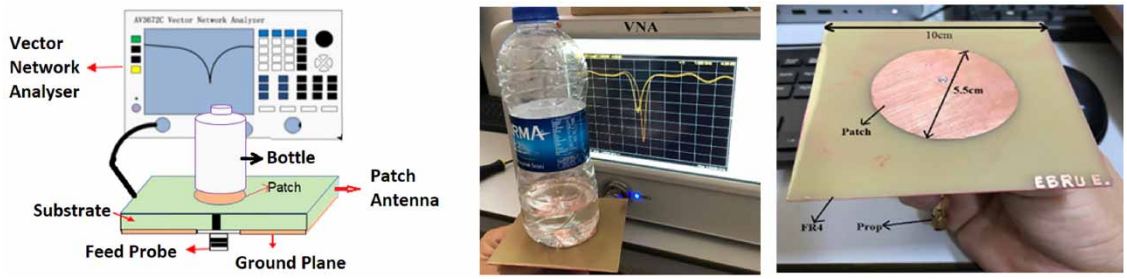

Figure 1 | Experimental setup.

output. It is possible to calculate the voltage values of incoming and outgoing signals by the following equations.

$$V_{r1} = S_{11}V_{i1} + S_{12}V_{i2} \tag{1}$$

$$V_{r2} = S_{21}V_{i1} + S_{22}V_{i2} \tag{2}$$

where V_{i1} and V_{i2} represent the voltage values applied to the circuit, and V_{r1} and V_{r2} represent the voltage values reflected from the circuit. In real applications, voltage measurements are difficult to obtain due to the wavelength of the signal at high frequencies. However, if the activated and reflected voltage is normalised with the square root of characteristic impedance (Z_0), the square root of the power is obtained using the following equations.

$$b_n = \frac{V_m}{\sqrt{Z_0}} \tag{3}$$

$$a_n = \frac{V_{in}}{\sqrt{Z_0}} \tag{4}$$

Using the power expression, the two-port circuit can be expressed as in the following equations.

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{5}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 \tag{6}$$

S parameters of a linear two-port circuit are shown in Figure 2. Impedance connection is made to the ports according to the parameter to be calculated. For example, if there is no reflected sign, that is, when $a_2 = 0$, S_{11} can be calculated. In order for a_2 to be zero, an impedance equal to the transmission impedance must be connected to the second port so that the signal applied on the impedance is used immediately and there is no signal reflection back. In this case (when $a_2 = 0$), Equation (7) becomes:

$$S_{11} = \frac{b_1}{a_1} \tag{7}$$

Antennas send electrical signals from any transmitter into free space by converting them into electromagnetic waves. These types of antennas are called transmitting antennas. Similarly, they receive electromagnetic waves from free space and convert these waves into electrical signals and transmit them to receivers. Such antennas are called receiving antennas. The antenna used in this study is both a transmitter and a receiver. There are some antenna parameters that need to be considered in microstrip antenna design. These are antenna patch, ground plane, feeding method, dielectric constant and thickness of the dielectric layer. The patch is a conductive structure that allows electromagnetic waves to absorb or radiate and is positioned on a dielectric layer. A circular copper patch was preferred in the antenna. The reason why the circular patch was preferred is its symmetrical radiation characteristic, which is not found in other types of patches. The dielectric layer is the material between the antenna radiation patch and the underside where the ground plane is located. FR-4 material with a thickness of 1.6 mm and a dielectric constant of 4.4 was used as a dielectric layer in the antenna. If the dielectric value of the dielectric layer is high in patch antennas, surface waves occur. For this reason, material with a very high dielectric

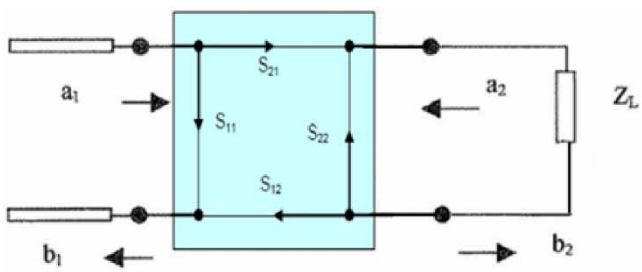

Figure 2 | *S* parameters with incident and reflected waves in a two-port linear circuit.

value was not preferred. The conductive material used in the ground plane has a material similar to the material used in the radiant patch. The coaxial probe feeding method was preferred as the feeding method. The reasons why this method was preferred are that it is easy to manufacture, has low artificial radiation characteristics, low interference radiation and less line losses.

The flowchart of the methodology used in this study is given in Figure 3. First, using four different classification algorithms, various salt water and pure water samples were classified. Then, the performances of these algoritms were evaluated using a set of performance metrics obtained with and without 10-fold cross validation.

3. CLASSIFIERS AND PERFORMANCE METRICS

In the following subsections, decision trees and performance metrics used in this study are explained.

3.1. Decision trees

The Rep tree algorithm (Quinlan 1987), originally proposed by Quinlan, uses regression tree logic. It creates many trees in different iterations and chooses the best tree among these trees. It uses the principle of gaining information by entropy in the creation of the regression tree, and achieves the least error with pruning (Devasena 2014; Srinivasan & Mekala 2014).

Hoeffding tree was proposed by Hulten *et al.* (2001). It uses the statistical value known as the Hoeffding boundary in each node of the decision tree to decide how to split the node. It reads each sample once and processes at an appropriate time interval.

Random Forest, proposed by Breiman, performs well even for classifying large amounts of data. Instead of branching nodes selected from the best attributes in the dataset, it branches all nodes by selecting the best of randomly received attributes in each node. Each dataset is generated from the original dataset with displacement and no pruning is involved (Breiman 2001). The reason why Random Forest is faster and more accurate than other algorithms is that it increases the classification rate by generating more than one decision tree.

Logistics Model Tree (LMT) combines logistic regression and decision tree (Lavrač 2003; Landwehr *et al.* 2005). While ordinary decision trees form a fragmented fixed model, LMT is a decision tree with a linear regression model whose

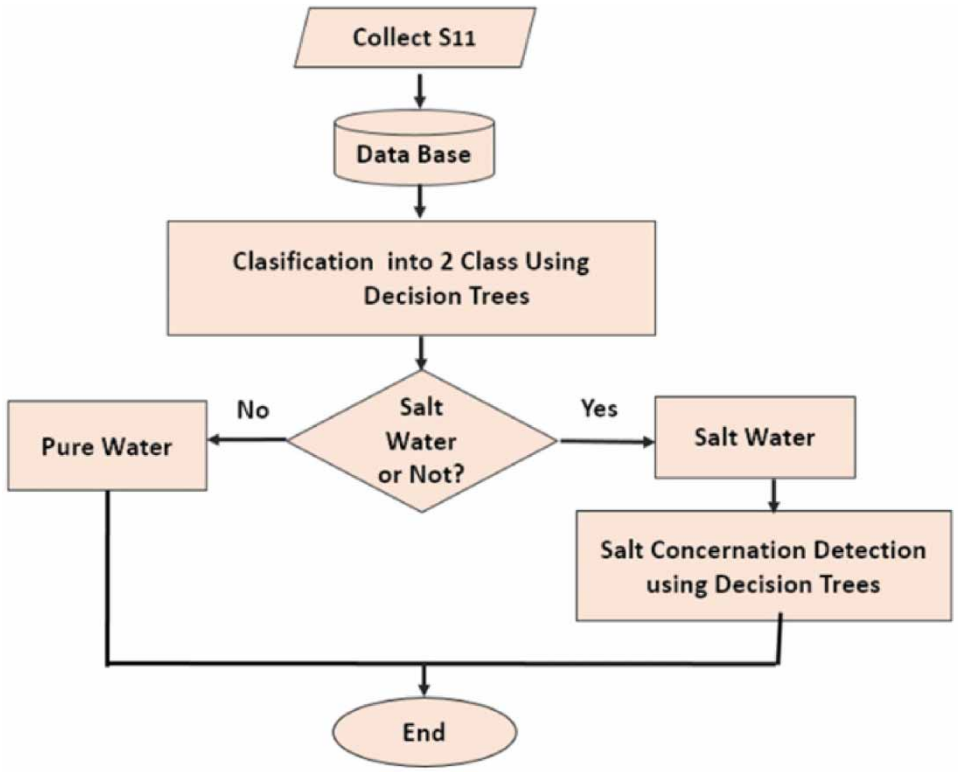

Figure 3 | Flowchart of the methodology.

leaves provide a piecewise linear regression (Lavrač 2003). The LogitBoost algorithm can be used to generate a logistic regression model at each node of the tree and then the node is separated using C4.5 criteria. With each LogitBoost execution, it restarts from its results via the parent node. Finally, the tree is pruned (Sumner *et al.* 2005).

3.2. Performance metrics

Following are the four outcomes in a classification process: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). In this study, TP occurs when the sample predicted as salt water is actually salt water. TN occurs when the sample predicted as not salt water is actually not salt water. FP occurs when the sample predicted as salt water is actually not salt water. Finally, FN occurs when the sample predicted as not salt water is actually salt water. In determining the saline concentration, the correctly predicted concentration amounts represent TP value. Based on TP, TN, FP and FN, various performance metrics can be calculated. Accuracy is the ratio of the number of correctly classified samples (TP + TN) to the total number of samples (TP + TN + FP + FN). Recall is the ratio of the number of correctly classified positive samples (TP) to the total number of positive samples (TP + FN). Precision is the ratio of the number of correctly classified positive samples (TP) to the total number of TP + FP. F-Score is the harmonic mean of the recall and precision metrics and can have a maximum of 1 and a minimum of 0. It is difficult to compare two models with low recall and high precision and vice versa; but, F-Score is used to make them comparable. The Kappa statistic is a numerical value comparing the expected and observed values. The observed accuracy rate is obtained by dividing the number of correctly classified samples by the total number of samples. The expected accuracy means that the classification algorithm is successful. Using these values, Kappa value can be calculated using Equation (8). Kappa statistic provides a measure of how closely a classification algorithm classifies by controlling its expected accuracy.

$$\text{Kappa} = \frac{\text{Observed accuracy value} - \text{expected accuracy value}}{1 - \text{expected accuracy value}} \tag{8}$$

There are some error criteria that are used to determine how accurately the predictions made by the algorithms. In this study, Mean Absolute Error (MAE) and Matthews Correlation Coefficient (MCC) were used. The MAE is one of the most frequently used error criteria and is calculated by taking the average of the absolute value of each difference between the actual value and the predicted value for that sample in the entire samples of the dataset. The MCC is a measure of the quality of classifications, introduced by biochemist Brian W. Matthews in 1975 (Matthews 1975). It is often used when imbalanced datasets are used. It is calculated using the following equation.

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})}} \tag{9}$$

4. RESULTS

When an electromagnetic wave is sent to liquids, interaction between molecules and electromagnetic waves occurs. As a result of this interaction, polarisation and depolarisation occur in molecules with different dielectric permittivity values. These cause a decrease in the magnitude of the wave in the wave velocity, the loss of energy caused by the friction of the directing molecules. In this study, the real part of permittivity ($\varepsilon r'$) is a measure of how much energy from an external electric field is stored in a liquid sample and the imaginary part of permittivity ($\varepsilon r''$) is a measure of how dissipative or lossy a liquid sample is to an external electric field.

In curve fitting algorithms, a mathematical model is created and correlation coefficient (R^2) in this model is determined. Correlation coefficient is between 0 and 1, the closer it is to 1, the more perfect the harmony has been achieved. In this study, curve fitting algorithms were used to determine the relation between salt concentration, permittivity and S parameter. For this purpose, first, the permittivity measurements of liquid samples in the range of 1.5–3 GHz were made with a coaxial probe method. The measurement results are presented in Figure 4. It was observed that when salt content in a liquid sample increased, the permittivity value changed depending on the salt concentration, and there was also a difference between the pure water and salt water graphs. The relation between salt concentration and permittivity is presented in Figure 5. The correlation coefficients between salt concentration and permittivity values, real part and imaginary part,

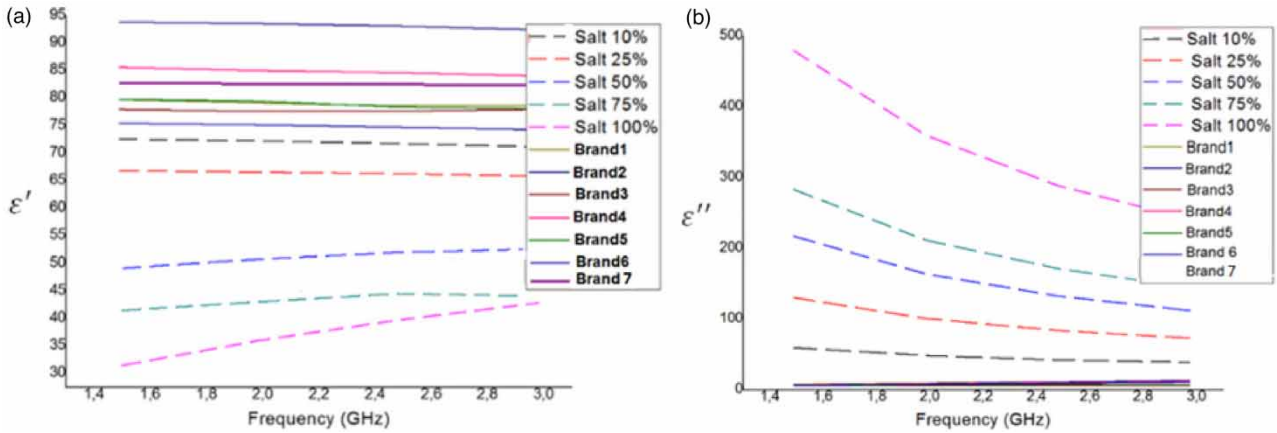

Figure 4 | Permittivity measurement of liquid samples. (a) The real part of permittivity and (b) the imaginary part of permittivity.

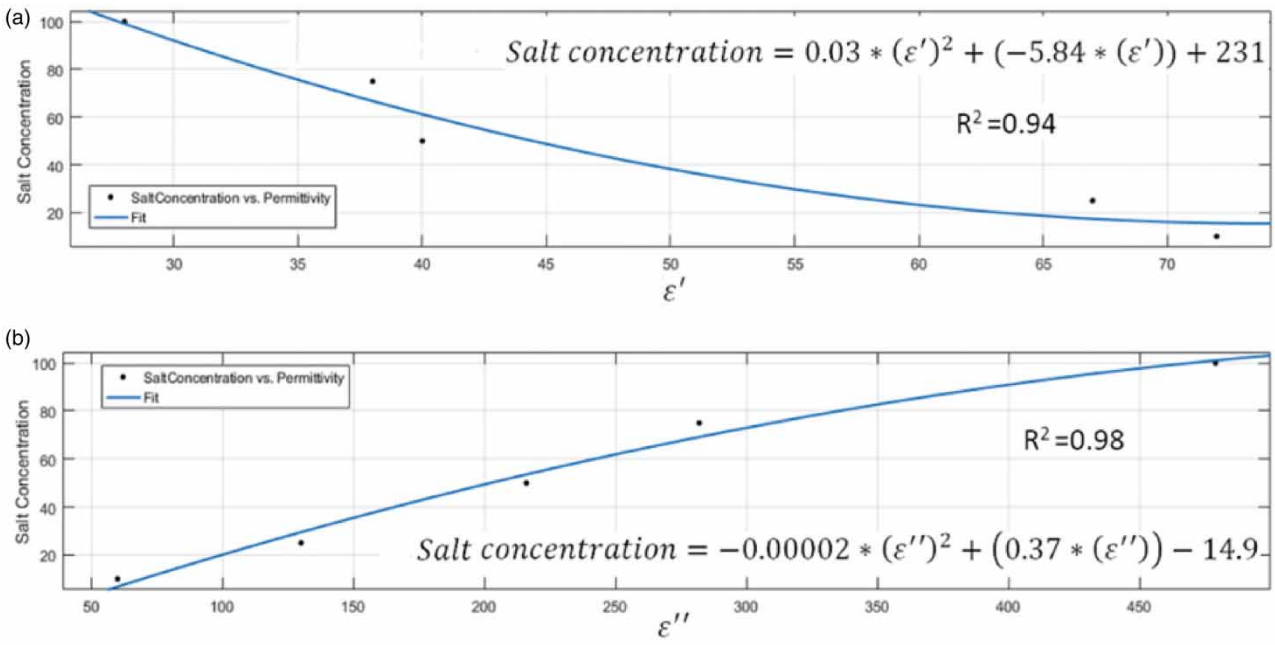

Figure 5 | Relation between salt concentration and permittivity. (a) The real part of permittivity and (b) the imaginary part of permittivity.

were $R^2 = 0.94$ and $R^2 = 0.98$, respectively. These values show that there is a very consistent relation between salt concentration and permittivity value.

S parameter measurements of the liquid samples with the proposed measurement system are given in Figure 6. The relation between the peak points (S_{max}) expressing the maximum amplitude values from the S parameter measurements and the permittivity values is presented in Figure 7. It can be seen that increase in salt concentration in water is inversely proportional to ε' and directly proportional to ε''. As the salt concentration increases, ε' decreases and ε'' increases. The relation between S_{max} value and salt concentration is shown in Figure 8. It is seen that S_{max} and salt concentration change inversely. A correlation coefficient of 0.93 indicated that the fit was high and the correlation allowed to determine the relation between salt concentrations or permittivity using S parameter measurements.

Since there is a strict relation between salt concentration and S parameter, both deciding whether a liquid sample was salt water or pure water and predicting salt concentration using S parameter measurements is possible. To prove this

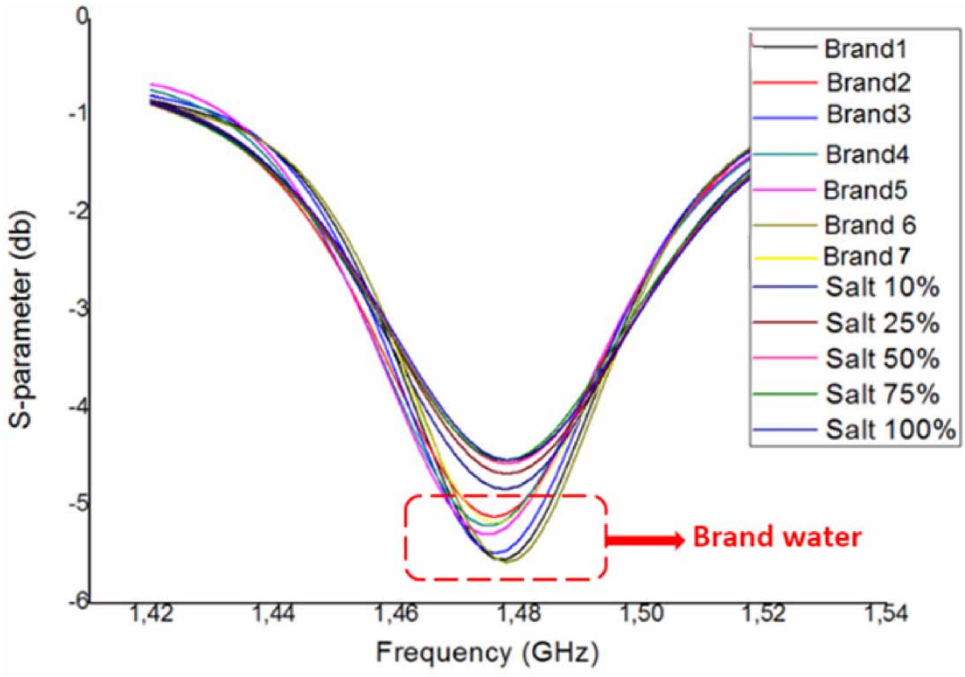

Figure 6 | *S* parameter measurements.

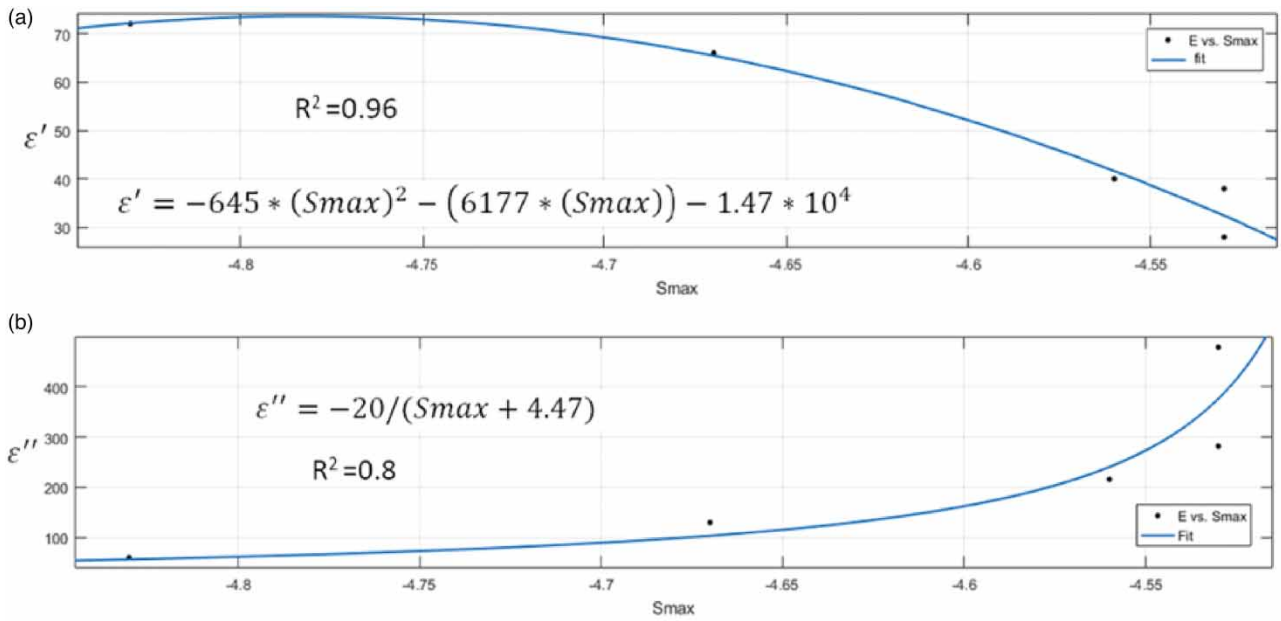

Figure 7 | Relation between S_{max} and permittivity. (a) The real part of permittivity and (b) the imaginary part of permittivity.

experimentally, *S* parameter measurements of seven different brands of bottled water and different concentrations of salt water were made. All measurements were made two times and a dataset was created. Using the dataset, first it was predicted whether a liquid sample was salt water or pure water. Then, the sample's salt concentration was predicted. The results were also validated using 10-fold cross validation. Tables 1 and 2 list the results in terms of the performance metrics employed in this study and the total number of correct and incorrect classifications for each of the classification algorithms.

In predicting whether a liquid sample was salt water or not, Precision, Recall, F-Score and MCC values were 1 in all algorithms except Rep tree. This shows that the algorithms made perfect classifications. On the other hand, Rep tree achieved a

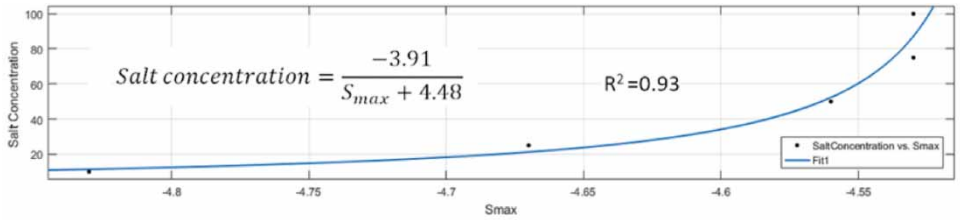

Figure 8 | Relation between salt concentration and S_{max}.

Table 1 | Performance metrics obtained when the aim was to predict whether liquid samples were salt water or pure water

Hoeffding Tree

Class	Without Cross validation				With Cross validation			
	Precision	Recall	F-Score	MCC	Precision	Recall	F-Score	MCC
Salt water	1	1	1	1	1	1	1	1
Pure water	1	1	1	1	1	1	1	1
Average	1	1	1	1	1	1	1	1

LMT

Class	Without Cross validation				With Cross validation			
	Precision	Recall	F-Score	MCC	Precision	Recall	F-Score	MCC
Salt water	1	1	1	1	1	1	1	1
Pure water	1	1	1	1	1	1	1	1
Average	1	1	1	1	1	1	1	1

Random Forest

Class	Without Cross validation				With Cross validation			
	Precision	Recall	F-Score	MCC	Precision	Recall	F-Score	MCC
Salt water	1	1	1	1	1	1	1	1
Pure water	1	1	1	1	1	1	1	1
Average	1	1	1	1	1	1	1	1

Rep Tree

Class	Without Cross validation				With Cross validation			
	Precision	Recall	F-Score	MCC	Precision	Recall	F-Score	MCC
Salt water	0.83	1	0.90	0,84	1	0,80	0.88	0.83
Pure water	1	0.85	0.92	0.84	0.87	1	0.93	0.83
Average	0.93	0.91	0.91	0.84	0.92	0.91	0.91	0.83

precision of 0.83, a recall of 1, an F-Score of 0.9 and an MCC of 0.84. The recall value of 1 in predicting whether salt water or not indicates that Rep tree correctly classified all the salt water samples. However, there is a decrease in its recall value in classifying pure water samples. This shows that Rep tree made wrong classifications in predicting the classes of the salt water samples. The value of F-Score also supports this. The fact that Rep tree's precision value was 1 indicates that it correctly classified all the pure water samples. When cross validation was applied, a decrease in MCC was observed as the success in predicting salt water samples decreased. Figure 9 presents the total number of correct and incorrect classifications for each of the classification algorithms when the aim was to predict whether the liquid samples were salt water or pure water.

Table 2 | Performance metrics obtained when the aim was to predict the salt concentration of liquid samples

Hoeffding Tree

	Without cross validation						With cross validation					
Class	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)
10 (%) Salt	1	1	1	1			1	0.50	0.66	0.66		
25 (%) Salt	1	1	1	1			0.66	1	0.80	0.76		
50 (%) Salt	1	1	1	1	10	0	1	1	1	1	9	1
75 (%) Salt	1	1	1	1			1	1	1	1		
100 (%) Salt	1	1	1	1			1	1	1	1		
Average	1	1	1	1			0.93	0.90	0.89	0.88		

LMT

Class	Without cross validation						With cross validation					
	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)
10 (%) Salt	1	1	1	1			1	1	1	1		
25 (%) Salt	1	1	1	1			1	0.50	0.67	0.66		
50 (%) Salt	1	1	1	1	10	0	1	0.50	0.67	0.66	8	2
75 (%) Salt	1	1	1	1			1	1	1	1		
100 (%) Salt	1	1	1	1			0.50	1	0.67	0.61		
Average	1	1	1	1			0.90	0.80	0.80	0.78		

Random Forest

Class	Without cross validation						With cross validation					
	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)
10 (%) Salt	1	1	1	1			1	0.50	0.66	0.66		
25 (%) Salt	1	1	1	1			0.66	1	0.80	0.76		
50 (%) Salt	1	1	1	1	10	0	1	1	1	1	9	1
75 (%) Salt	1	1	1	1			1	1	1	1		
100 (%) Salt	1	1	1	1			1	1	1	1		
Average	1	1	1	1			0.93	0.90	0.89	0.88		

(*Continued.*)

Table 2 | Continued

Rep Tree

Class	Without cross validation						With cross validation					
	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)	Precision	Recall	F-Score	MCC	Correct Instances (Total Number)	Incorrect Instances (Total Number)
10 (%) Salt	0.33	1	0.50	0.40			0	0	0	−0.50		
25 (%) Salt	–	0	–	–			0	0	0	−0.25		
50 (%) Salt	–	0	–	–	4	6	0	0	0	−0.25	0	0
75 (%) Salt	0.50	1	0.66	0.61			–	0	–	–		
100 (%) Salt	–	0	–	–			0	0	0	−0.16		
Average	–	0.4	–	–			–	0	–	–		

Five classes were created for predicting the salt concentration of the liquid samples. Without cross validation, all the performance metrics had a value of 1 for Hoeffding Tree, LMT and Random Forest. On the other hand, Rep Tree showed a very bad performance in this classification. When cross validated, the results were worse compared to the classification without cross validation. This indicates that when a liquid sample with a different concentration level is analysed with the proposed system, the concentration level might be predicted incorrectly; although, the decision regarding whether the liquid is salt water or pure water is correct.

Accuracy, Kappa and MAE values of the algorithms are listed in Figures 10–12, respectively. It is seen that without cross validation Hoeffding Tree, LMT and Random Forest achieved an accuracy of 100% and a Kappa value of 1. However, Rep Tree achieved an accuracy of 91% and a Kappa value of 0.83 in predicting whether a liquid sample was salt water or not. On the other hand, in predicting salt concentration it achieved an accuracy of 40% and a Kappa value of 0.25. After cross validation, accuracy rates and Kappa values of Hoeffding Tree and Random Forest were lower. Hoeffding Tree and Random Forest achieved an accuracy of 90% and a Kappa value of 0.87 in.predicting salt concentration. LMT achieved an accuracy of 80% and a Kappa value of 0.75 in predicting salt concentration. Since Rep Tree could not classify any samples correctly, its accuracy was 0%. Hoeffding Tree had the lowest MAE values.

When all the performance metrics are taken into consideration, it is seen that all the algorithms are suitable for predicting whether a liquid sample is salt water or not. Among the algorithms, Rep Tree had the lowest performance in terms of all the

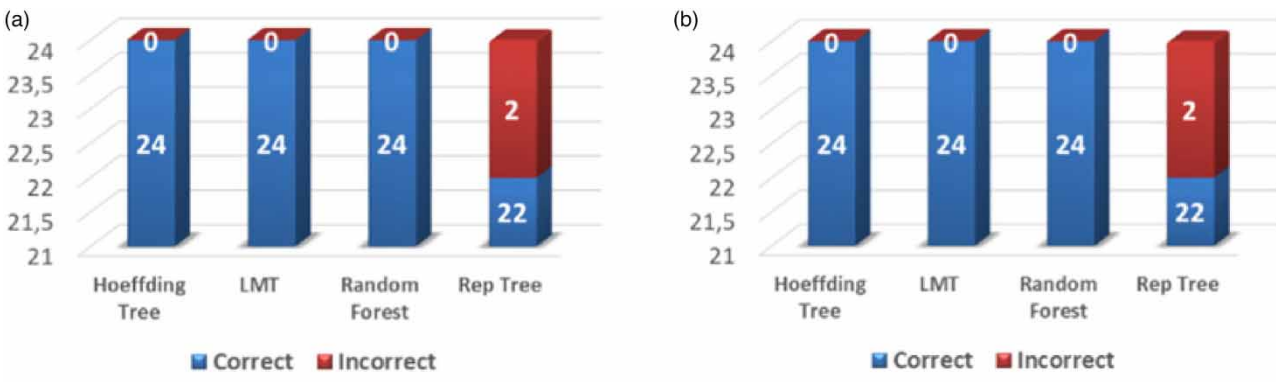

Figure 9 | Predicting whether liquid samples were salt water or pure water (a) without cross validation and (b) with cross validation.

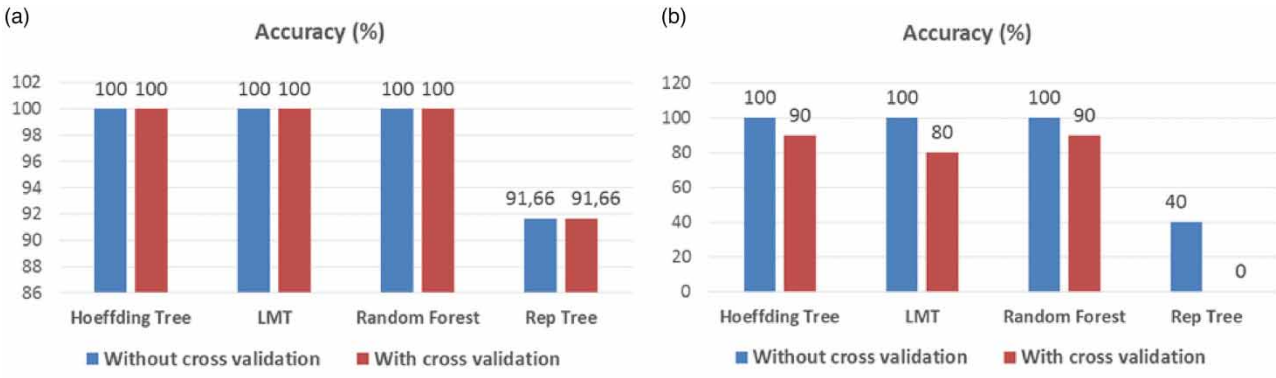

Figure 10 | Accuracy. (a) Predicting whether a liquid sample was salt water or not and (b) predicting the salinity level.

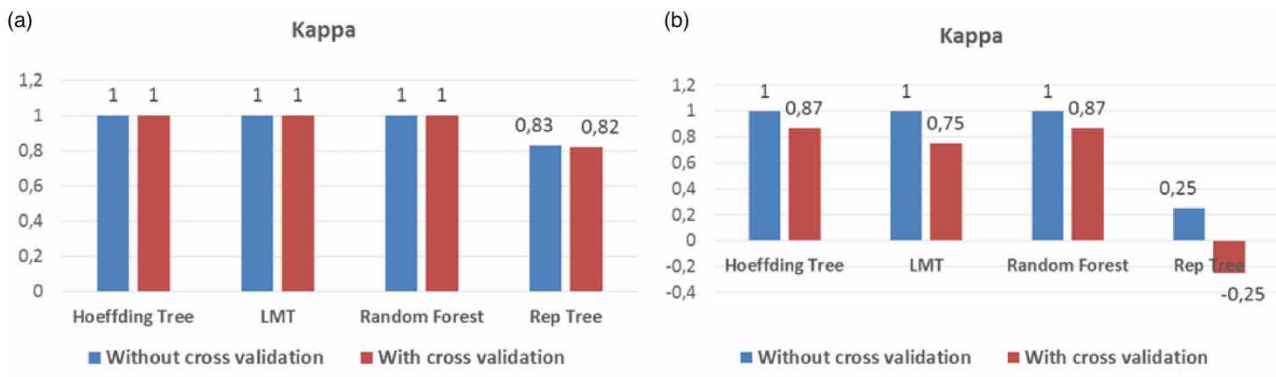

Figure 11 | Kappa. (a) Predicting whether a liquid sample was salt water or not and (b) predicting the salinity level.

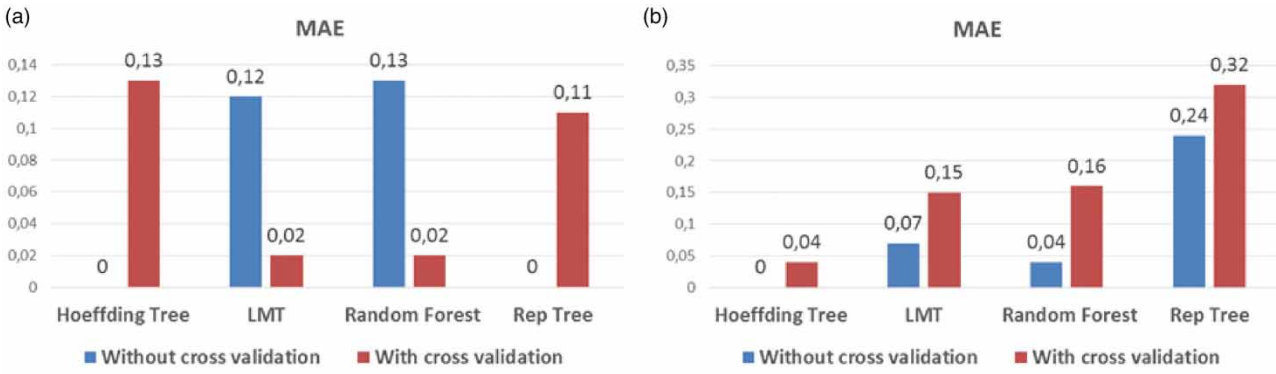

Figure 12 | MAE. (a) Predicting whether a liquid sample was salt water or not and (b) predicting the salinity level.

performance metrics. It had an accuracy rate of 91.66%, while the others achieved an accuracy rate of 100%. While the average values of Precision, Recall, F-Score, MCC and Kappa were 1 in the others, in Rep Tree, the average of these values was 0.92, 0.91, 0.91, 0.83 and 0.82, respectively. In predicting the salt concentration of the liquid samples, Hoeffding Tree, LMT and Random Forest algorithms achieved an accuracy rate of 100%, while Rep Tree achieved an accuracy rate of 40%. As expected, after cross validation, all the algorithms had lower performance. Nevertheless, Hoeffding Tree and Random Forest algorithms achieved an accuracy rate of 90%, a precision of 0.93, a recall of 0.90, an F-Score value of 0.89 and a Kappa value of 0.87. Although they had almost the same metric values, compared to Random Forest, Hoeffding Tree had a lower error rate. Therefore, it can be concluded that Hoeffding Tree made better classifications.

5. CONCLUSION

Due to the increasing interest in non-contact detection of salinity, in this study a novel approach for this purpose was proposed. Since the proposed approach relies on software component, i.e. classifiers, as well as hardware components, a set of classification experiments were performed to evaluate the success of different decision tree algorithms in order to determine the most suitable one for salinity detection. The results of the classification experiments show that the Hoeffding tree algorithm achieved the best results and Rep tree had the worst performance. The main limitation of the proposed approach is that it can deliver uncertain results when the concentration of salinity is below 5%. Future work of this study focuses on the development of a VNA-embedded system for the proposed system. Such a prototype system when supported by software application based on the Hoeffding tree algorithm will be able to quickly determine salt concentration in water. It will be a standalone and automated solution and its single requirement to operate will be a power supply.

FUNDING

No funding was received for conducting this study.

CONFLICTS OF INTERESTS

The authors have no conflicts of interest to declare that are relevant to the content of this article.

AUTHORS' CONTRIBUTION

The first author conceived and planned the experiments and prepared the samples. Both authors carried out the experiments and contributed to the interpretation of the results. Both authors discussed the results and contributed to the final manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Bartels, D. & Sunkar, R. 2005 Drought and salt tolerance in plants. *Critical Reviews in Plant Sciences* **24** (1), 23–58.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Chai, B.-H. 2008 Spectroscopic studies of solutes in aqueous solution. *The Journal of Physical Chemistry A* **112** (11), 2242–2247.

Daliri, A. 2010 Circular microstrip patch antenna strain sensor for wireless structural health monitoring. In: *Proceedings of the World Congress on Engineering*.

Deniz, K. & Kadioglu, Y. K. 2021 Geochemistry of salts and the effect of trace elements on human health: Turkey salt resources. *International Journal of Environmental Analytical Chemistry* 1–19 (in press).

Devasena, C. L. 2014 Comparative analysis of random forest, REP tree and J48 classifiers for credit risk prediction. *International Journal of Computer Applications*, 0975–8887. Available from: https://research.ijcaonline.org/icccmit2014/number3/icccmit7033.pdf.

Ekinci, E. I., Clarke, S., Thomas, M. C., Moran, J. L., Cheong, K., MacIsaac, R. J. & Jerums, G. 2011 Dietary salt intake and mortality in patients with type 2 diabetes. *Diabetes Care* **34** (3), 703–709.

Fipps, G. 2003 *Irrigation Water Quality Standards and Salinity Management Strategies*. Texas FARMER Collection.

Hulten, G., Spencer, L. & Domingos, P. 2001 Mining time-changing data streams. In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Jabed, M. A., Paul, A. & Nath, T. K. 2020 Peoples' perception of the water salinity impacts on human health: a case study in South-Eastern Coastal Region of Bangladesh. *Exposure and Health* **12** (1), 41–50.

Lavrač, N. 2003 Machine learning: ECML. In *14th European Conference on Machine Learning*, September 22–26, Proceedings, Cavtat-Dubrovnik, Croatia. Vol. 2837. Springer.

Landwehr, N., Hall, M. & Frank, E. 2005 Logistic model trees. *Machine Learning* **59** (1–2), 161–205.

Sumner, M., Frank, E. & Hall, M. 2005 Speeding up logistic model tree induction. In *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer.

Matthews, B. W. 1975 Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* **405** (2), 442–451.

Meng, Q. 2014 Optical fiber laser salinity sensor based on multimode interference effect. *IEEE Sensors Journal* **14** (6), 1813–1816.

Munns, R. & Tester, M. 2008 Mechanisms of salinity tolerance. *Annual Review of Plant Biology* **59**, 651–681.

Quan, X. & Fry, E. S. 1995 Empirical equation for the index of refraction of seawater. *Applied Optics* **34** (18), 3477–3480.

Quinlan, J. R. 1987 Simplifying decision trees. *International Journal of man-Machine Studies* **27** (3), 221–234.

Srinivasan, D. B. & Mekala, P. 2014 Mining social networking data for classification using reptree. *International Journal of Advance Research in Computer Science and Management Studies* **2** (10), 155–160.

Ustaoğlu, F., Taş, B., Tepe, Y. & Topaldemir, H. 2021 Comprehensive assessment of water quality and associated health risk by using physicochemical quality indices and multivariate analysis in Terme River, Turkey. *Environmental Science and Pollution Research* **28**, 62736–62754.

Villarreal Jiménez, L. R., Rodríguez Rodríguez, A. J., Enríquez Sías, S. U., Elizondo González, C., Barrón González, H. G., Erro Betrán, M. J., Rodríguez Rodríguez, W. E. & Domínguez Cruz, R. F. 2018 An efficient optoelectronic system for remote salinity water sensing. *Applied Mechanics and Materials* **876**, 152–160. https://doi.org/10.4028/www.scientific.net/amm.876.152.

Wong, G. S. & Zhu, S. M. 1995 Speed of sound in seawater as a function of salinity, temperature, and pressure. *The Journal of the Acoustical Society of America* **97** (3), 1732–1736.

Yahaya, N. 2012 Determination of moisture content of hevea rubber latex using a microstrip patch antenna. In: *PIERS Proceedings*.

Zhao, Y. & Liao, Y. 2002 Novel optical fiber sensor for simultaneous measurement of temperature and salinity. *Sensors and Actuators B: Chemical* **86** (1), 63–67.

First received 6 December 2021; accepted in revised form 30 April 2022. Available online 16 May 2022