

## Article

# Machine-Learning-Based Sensor Design for Water Salinity Prediction: A Conceptual Approach

Bachar Mourched <sup>1,\*</sup>, Mariam Abdallah <sup>2</sup>, Mario Hoxha <sup>1</sup> and Sabahudin Vrtagic <sup>1</sup>

<sup>1</sup> College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; sabahudin.vrtagic@aum.edu.kw (S.V.)

<sup>2</sup> Faculty of Science III, Lebanese University, Tripoli 90656, Lebanon

\* Correspondence: bachar.mourched@aum.edu.kw

**Abstract:** This research paper introduces a sensor that utilizes a machine-learning model to predict water salinity. The sensor's concept and design are established through a simulation software which enables accurate modeling and analysis. Operating on the principle of light transmission physics, the sensor employs data collected from the simulation software as input parameters to predict the salinity parameter, serving as the output. The results of the prediction model exhibit excellent performance, showcasing high accuracy with a coefficient of determination value of 0.999 and a mean absolute error of 0.074. These outcomes demonstrate the model's ability, particularly the multi-layer perceptron model, to effectively predict salinity values for previously unseen input data. This performance underscores the model's accuracy and its proficiency in handling unfamiliar input data, emphasizing its significance in practical applications.

**Keywords:** COMSOL; light transmission; machine learning; multi-layer perceptron; water salinity



**Citation:** Mourched, B.; Abdallah, M.; Hoxha, M.; Vrtagic, S. Machine-Learning-Based Sensor Design for Water Salinity Prediction: A Conceptual Approach. *Sustainability* **2023**, *15*, 11468. <https://doi.org/10.3390/su151411468>

Academic Editor: Junwei Ma

Received: 6 June 2023

Revised: 18 July 2023

Accepted: 22 July 2023

Published: 24 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The salinity of seawater is an essential oceanographic parameter, representing the total amount of dissolved salt in seawater. Measuring seawater salinity accurately is essential for understanding ocean circulation, water cycle dynamics, and heat transfer. In this paper, we examine various techniques used for measuring seawater salinity, including their principles, advantages, and limitations. Then, we put forward a new method for detecting water salinity by means of optics and a machine-learning model to achieve precise findings.

The conventional method used to measure seawater salinity is through titration. This technique involves the process of titrating a known volume of seawater with a solution of silver nitrate until a complete reaction occurs with all the chloride ions. When determining the salinity of seawater, the amount of silver nitrate required to carry out titration is proportional to the chloride content [1–3]. Chemo-resistive sensors are used to directly measure seawater salinity [4–6]. An environmentally friendly sensor made of indium-tin oxide (ITO) nanoparticles has been investigated for detecting liquid chemicals in brine [7]. Direct measuring procedures often involve chemical processes to change the solution's chemical properties. For applications that require the chemical properties of the solution to be preserved, these techniques may prove ineffective.

The conductivity method is one of the most common approaches used to determine the salt content of seawater which is based on the correlation between the electrical conductivity and salinity of seawater. A conductivity sensor measures the electrical conductivity of seawater, and salinity is calculated using empirical or theoretical relationships between conductivity and salinity. This approach is known for its straightforward design, ease of implementation, rapid response, and affordability, making it widely used in various industrial settings [8–11]. Conductivity ratio measurements call for a steady stream of water to be pumped through the conductivity cell, like the Sea-Bird CTD profiler. This makes the system more intricate and harder to miniaturize [12]. The conductivity CTD is

also limited in its ability to account for the fraction of dissolved material in seawater that does not conduct electricity [13].

A specific type of sensor that relies on the optical phenomenon known as surface plasmon resonance (SPR) has been proposed in different studies. These sensors have found applications in diverse areas, including medical diagnostics, biochemical sensing, and water pollution monitoring [14–16]. Several studies have employed the SPR technique for developing multi-parameter sensors that can detect seawater salinity and temperature simultaneously [17–20]. Refs. [21,22] have applied Raman spectroscopy to detect seawater salinity.

Since the refractive index (RI) is related to density via the Lorentz–Lorenz relationship, it has been used as an optical technique for salinity sensing [23,24]. Many optical concepts, including the critical angle, the Fraunhofer method, and total internal reflection, have been applied to the development of refractive-index-based optical salinity sensors [25–29].

Operating a densitometer [30,31] or an interferometer [32–34] can also be employed for measuring water salinity. Nevertheless, these approaches may not be suitable for operating under harsh measurement conditions, and they are typically used in laboratory settings [35].

The most used techniques for measuring water salinity are those that do not alter the chemical properties of solutions. Of these methods, optical methods are the simplest, most affordable, and most accurate. Most commercial sensors for measuring salinity rely on the conductivity ratio measurement, such as the widely used Sea-Bird CTD (conductivity, temperature, depth) profiler. These sensors require a reliable pump to maintain a constant flow rate of water through the conductivity cell, which complicates miniaturization and sensitivity to water velocity [36]. Additionally, temperature has a greater impact on conductivity than the refractive index [37]. The TEOS-10 guidelines state that conductivity-based CTD sensors cannot detect seawater's non-conductive dissolved material [13]. As a viable alternative, measuring the refractive index (RI) is an advantageous alternative method. The Lorentz–Lorenz correlation in oceanography links RI to absolute salinity by means of density [23,24]. This has led to an increased focus on developing more precise methods for measuring RI in seawater. The seawater RI equation depends on density and, consequently, on absolute salinity, pressure, temperature, and light wavelength [25,28,29].

Machine-learning models have recently been presented as an alternative method for determining the salinity of water. A sea ice buoy was used to measure chlorophyll levels in the upper ocean, and the nature of the sensor was determined by applying a hybrid machine-learning technique [38]. Yet, a different method was necessary to convert the data into pertinent ocean parameters. Ref. [39] used a neural network trained on a generalized radial basis function to predict the salinity of ocean water. Few studies have examined the measurement of water salinity with the MLP based on optical data. Employing a scattering pattern, an artificial intelligence deep-learning framework is used to identify saline particles with varying salt concentrations, as stated in [40]. In two phases, salt particles were classified. Initial photos of various salt salinities were taken with a Raspberry Pi. The photos were then utilized to train the deep-learning neural network model to categorize saline particles of various ranges. Water salt classification accuracy was 90% using the approach. Due to the insufficient quantity of training data, the deep-learning system categorized incorrectly particles of saline salt. Cubist (version 0.2.2) and corplot (version 0.84) R packages were utilized by [41]. Each rule is coupled with a multiple linear regression model in the rule-based prediction model utilized. Applying the linear regression model, the relationship between soil salinity and MSI-derived surface soil moisture was analyzed. To assess the efficacy of the model, some fundamental metrics were employed, including: (1) the coefficients of determination ( $R^2$ ), (2) root mean square errors (RMSE), (3) normalized root mean square errors (NRMSE) considering the wide range of soil salinity in the study area, (4) ratio of performance to deviation (RPD) and (5) ratio of performance to interquartile distance (RPIQ).

This paper introduces an innovative method that employs both the principles of light refraction and a machine-learning system to enable the real-time measurement of saltwater concentration in a free biochemical environment. A machine-learning model is trained on a large simulation study dataset to find patterns and relationships. Therefore, we are eliminating the experimental complexity of real-time data collection and filtering, which is one of the major challenges of machine learning, to optimize the model. Furthermore, the deep-learning system encountered challenges in certain instances due to the scarcity of training data. This is precisely why our paper focuses on generating synthetic data that can be augmented with real data to enhance prediction accuracy.

In this study, we use COMSOL Multiphysics, a powerful tool widely used in various fields such as energy harvesting, medical devices, and engineering [42–44], to design and analyze our sensor function and obtain the required data. Despite numerous research studies using COMSOL to test seawater sensors, the possibility of integrating machine-learning approaches with COMSOL Multiphysics for seawater sensor development has yet to be thoroughly explored [45,46].

Despite advances in seawater sensor manufacturing, approaches, and methods, a simple, cost-effective, and high-performance sensor that can identify water type (saltwater or pure) and temperature is still essential. Our proposed solution to this issue involves a compact and straightforward laser-based measuring system developed using COMSOL Multiphysics. This system operates on the basis of a fundamental law of physics (Snell's law) for computation and can effectively identify the refractive index of water samples that are being tested. The simulation data collected through this system are then utilized as inputs to a machine-learning model that predicts the salinity concentration of a water sample under different temperatures. This method is highly sensitive, low-cost, and ideal for label-free biochemical detection in real time.

This paper aims to contribute to the field by presenting a novel conceptual approach by combining optics principles with a machine-learning technique for water salinity measurement. We would like to emphasize that our project does not involve a development stage process and does not cover any experimentation, procedures, or instruments.

The following sections examine and analyze the specifics of this research in detail.

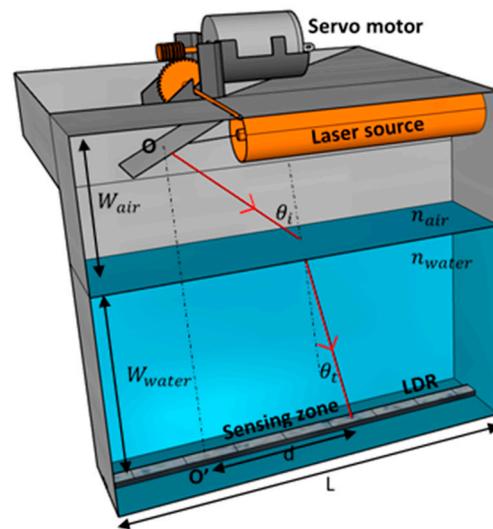
In Section 2, the conceptual framework and method used to make the suggested sensor are explained in detail. Results from COMSOL Multiphysics simulations and analyses of data gathered through machine learning are also presented. Evaluation of the sensor's performance is presented in Section 3, and findings are discussed in Section 4.

## 2. Sensor's Design and Concept

To generate synthetic data for MLP input–output, the sensor principle is simulated using COMSOL Multiphysics. The layout and scale of the sensor are rendered as a 2D rectangular geometry, resembling a water container or tank into which the water sample will be placed for analysis.

The sensor concept revolves around the refraction of laser light as it passes through the water sample until it reaches the bottom section of the tank, known as the sensing zone. Within the zone, sensors are implemented to measure the distance covered by the light. The measured distance depends on several factors, wavelength of light, angle of transmission, refractive index, temperature, and salinity of the water sample.

Parameters and equations crucial for the COMSOL Multiphysics analysis of the sensor's concept and design are presented here. The model has three main parts: a cap that houses the laser, a tank that holds the water sample, and a sensor-equipped sensing zone at the bottom. The sensor model geometry, defined through COMSOL, comprises a 2D design separated into two domains, air (refractive index  $n_{air} = 1$ ) and water (refractive index  $n_w$ ). A 3D illustration of the sensor concept design is shown in Figure 1. However, the study and data obtained using COMSOL are based on a 2D analysis.



**Figure 1.** Illustration depicting the 3D design of the proposed sensor showcasing the positioning of the laser source, servo motor, and the LDR.

The sensor is modeled and designed in COMSOL with a rectangular shape in a 2D configuration, and its dimensions are set to 30 cm for length and 20 cm for width where both air and water widths are equal:  $W_{air} = W_{water} = 10$  cm ( $W_{air} + W_{water} = 20$  cm).

The laser, located at the point of incidence  $O$  within the cap, emits light at an angle of incidence  $\theta_i$ , which can be adjusted using a mirror–servo motor system, as shown in Figure 1. As light travels from air to water, it bends at an angle of refraction  $\theta_t$  with the normal. The relationship between  $n_{air}$ ,  $n_w$ ,  $\theta_i$  and  $\theta_t$  is described by Snell’s law (Equation (1)) implemented in the simulation study.

$$n_{air} \sin\theta_i = n_w \sin\theta_t \quad (1)$$

In order to reach a sensing zone at a distance of  $d$  from the reference point of incidence  $O$ , transmitted light must travel through a medium at an angle of  $\theta_t$ .

The value of  $d$  is influenced by various factors such as the wavelength of the incident light  $\lambda$ , angle of incidence  $\theta_i$ , refractive index of the water sample  $n_w$ , temperature of the water  $T$ , and the depth of air and water in the medium  $W_{air}/W_{water}$ .

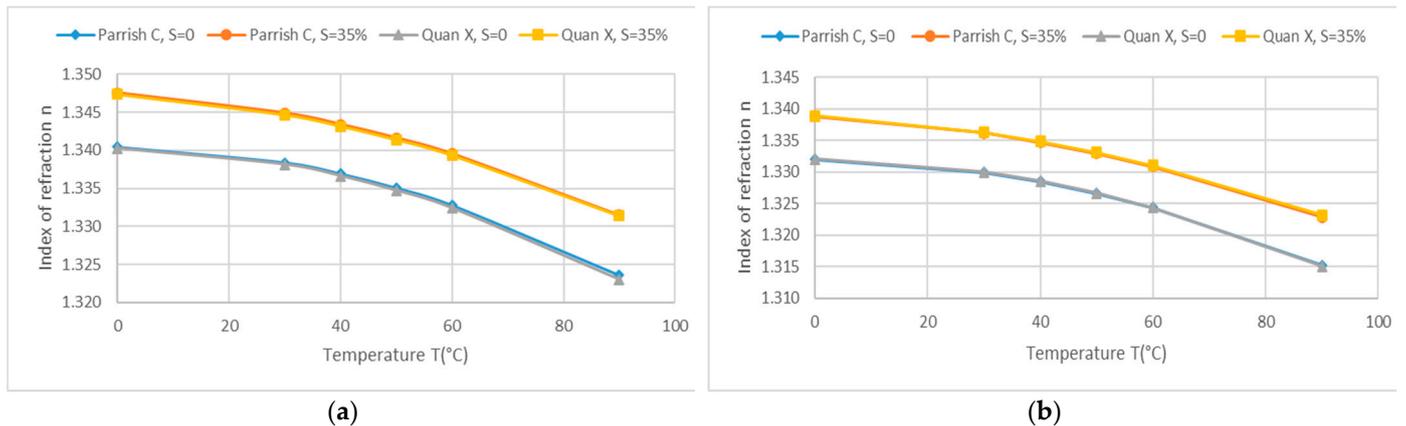
To model the electromagnetic wave propagation, the “Geometrical Optics” physics interface is applied. This study ignores diffraction and reflection at geometry edges and corners by defining wall boundary conditions as “disappear” options for perfect absorption. For precise air–water refraction, the optical path length step is 0.01 cm. The investigation in this study is restricted to the distance  $d$  parameter, dependent on the  $\lambda$ ,  $\theta_i$ ,  $T$ ,  $n_w$  and  $W_{air}/W_{water}$  parameters.

Three distinct wavelengths  $\lambda$  were simulated to determine the distance parameter  $d$  (450 nm for blue, 520 nm for green, and 660 nm for red) and four different incident angles  $\theta_i$  ( $10^\circ$ ,  $20^\circ$ ,  $30^\circ$ , and  $45^\circ$ ). The parameter  $d$  is computed for a temperature  $T$  range of 0 to  $80^\circ\text{C}$  with a  $1^\circ\text{C}$  increment, and for a salinity  $S$  range of 0 to 40% with a 1% increment.

### 3. Simulation Data Validation

The refractive index of the water sample  $n_w$  is a crucial factor in this study and must be defined correctly through COMSOL in order to obtain trustworthy data. Various studies have attempted to determine the refractive index  $n_w$  of salty water. Parrish C [47] proposed an empirical formula, which builds on earlier research [48,49], to estimate  $n_w$  with an accuracy of up to 3–4 decimal places. Reference [26] measured seawater salinity in situ using a total-internal-reflection-based optical refractometer and the empirical equation from the Quan–Fry formula [50]. Both formulae are based on the water’s salinity  $S$ , temperature  $T$ , and the wavelength of light  $\lambda$ .

Refs. [47,50] conducted studies to compare the refractive index of water salinity of  $S = 0$  and  $S = 35\%$  at different temperatures and wavelengths. Figure 2 displays this comparison for  $\lambda = 450$  nm and 660 nm revealing that the calculated refractive index values from both references were identical for both salinity levels and temperatures tested, with a standard deviation of  $34.88 \times 10^{-4}$  for  $\lambda = 450$  nm, and  $33.43 \times 10^{-4}$  for  $\lambda = 660$  nm. Thus, equations in [47,50] are equivalent and can be used interchangeably. For our results and analysis, we have opted to implement the Quan–Fry equation (Equation (2)) outlined in [50] in the simulation study to depict the refractive index of the water sample.



**Figure 2.** Refractive index changes with temperature across different salinity levels for (a)  $\lambda = 450$  nm and (b)  $\lambda = 660$  nm, using two different approaches (Parrish [47] and Quan [50]).

$$n(S, T, \lambda) = n_0 + S(n_1 + n_2T + n_3T^2) + n_4T^2 + \frac{n_5 + n_6S + n_7T}{\lambda} + \frac{n_8}{\lambda^2} + \frac{n_9}{\lambda^3} \quad (2)$$

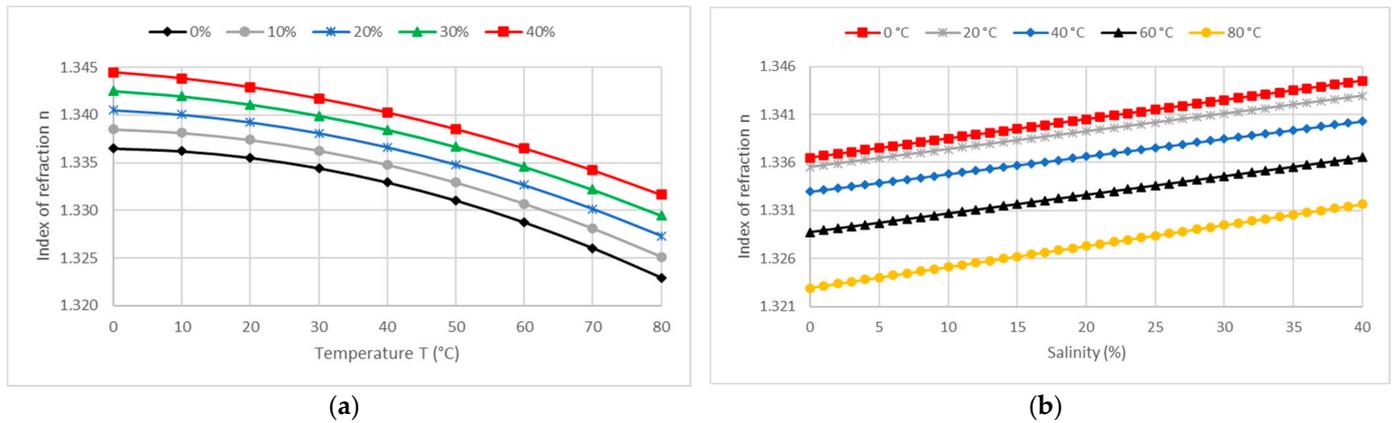
$T, \lambda$ , and  $S$  are the temperature in °C, wavelength in nm, and salinity in %, respectively.  $n_0$  to  $n_9$  are the coefficients determined by  $n_0 = 1.31405, n_1 = 1.779 \times 10^{-4}, n_2 = -1.05 \times 10^{-6}, n_3 = 1.6 \times 10^{-8}, n_4 = -2.02 \times 10^{-6}, n_5 = 15.868, n_6 = 0.01155, n_7 = -0.00423, n_8 = -4382$ , and  $n_9 = 1.145 \times 10^6$ .

To validate the data acquired from COMSOL, the refractive index values are cross-referenced with values obtained through Snell's law and trigonometry (Equation (1)), as well as values obtained from [25,26,47]. Figure 3a,b, at  $\lambda = 520$  nm, depict, respectively, the influence of temperature on the refractive index for different water salinities and the influence of water salinity on the refractive index for different temperatures. Notably, as the temperature rises, the refractive index values decrease for all salinity percentages, whereas an increase in salinity results in an increase in refractive indices. The comparison results indicate similarity, thereby confirming the validity of the data acquired through COMSOL.

Data validation in this study involves rigorous measures to ensure the accuracy and reliability of the findings. Firstly, the correctness and reliability of the equation for the refractive index  $n_w$  are verified by comparing it with equations derived from different experimental studies. This comparison serves as a crucial step in confirming the validity of the data used in our analysis. Furthermore, the behavior of  $n_w$  in terms of temperature and salinity is carefully examined and compared with similar studies that exhibit comparable patterns and behaviors. By conducting these thorough validations, we can confidently affirm the integrity and robustness of the simulation results employed in our research, strengthening the credibility of our study's conclusions.

The machine-learning algorithm's parameters are based on the data obtained from the simulation study. The input data in the algorithm can be rearranged to have a specified output depending on the purpose of the study performed. In our study, the main focus of our simulation study is to generate synthetic data related to the salinity measurement. Therefore, the salinity  $S$  is set to be the output parameter. In the following section,

the machine-learning model is detailed, discussing data classification in the model and results analysis.



**Figure 3.** Refractive index changes with (a) temperature across different salinity levels, and (b) salinity across various temperatures.

#### 4. Machine-Learning Model Verification

MLP models are particularly effective in addressing problems involving non-linear relationships between input and output variables. Multiple layers of neurons enable the modeling of non-linear functions by capturing intricate patterns and correlations in data. The backpropagation algorithm adjusts neuron weights and biases based on the difference between predicted and actual outputs to train MLP models. Through this iterative process, the model can learn from the data and enhance its performance over time. Various machine-learning libraries and frameworks, including TensorFlow, PyTorch, and scikit-learn, provide support for MLP models (the latter incorporates a built-in overfitting trigger). These libraries offer efficient implementations of MLPs and a broad range of tools and functionalities for training, evaluating, and deploying the models.

As previously mentioned, the simulation of the laser beam refraction is carried out using COMSOL Multiphysics software, considering modeling different values for various factors such as  $\lambda$ ,  $T$ ,  $\theta_i$ , and  $S$ . The simulation computes the distance  $d$  for every variation in the aforementioned variables. The resulting data from these simulations are combined and restructured to generate an input–output dataset defining the salinity  $S$  as output, presented in Table 1.

**Table 1.** Input–output dataset sample.

	INPUT												OUTPUT	
	$\lambda$ (nm)	450	520	660	450	520	660	450	520	660	450	520	660	$T$ (°C)
$\theta$ (deg)	10	10	10	20	20	20	30	30	30	45	45	45		
$d$ (cm)	3.0687	3.0724	3.0769	6.2764	6.2843	6.2939	9.7901	9.8032	9.819	16.2026	16.2268	16.256	10	8
	3.0716	3.0753	3.0798	6.2826	6.2905	6.3	9.8004	9.8134	9.8293	16.2216	16.2457	16.275	45	15
	3.0743	3.078	3.0825	6.2882	6.2961	6.3057	9.8097	9.8228	9.8387	16.2388	16.2631	16.2925	62	20
	3.0733	3.0771	3.0816	6.2863	6.2942	6.3039	9.8065	9.8196	9.8356	16.2328	16.2572	16.2868	64	27
	3.0634	3.0671	3.0717	6.2651	6.2731	6.2829	9.7714	9.7847	9.8008	16.1683	16.1927	16.2225	8	35
	3.0631	3.0669	3.0715	6.2644	6.2725	6.2823	9.7704	9.7837	9.7998	16.1664	16.1908	16.2206	16	40
	3.0695	3.0732	3.0777	6.2781	6.286	6.2956	9.793	9.8061	9.8219	16.208	16.2321	16.2613	7	3
	3.07	3.0737	3.0782	6.2791	6.287	6.2965	9.7946	9.8077	9.8235	16.211	16.2351	16.2643	3	0

Table 1 displays input–output data parameters from a subset of the entire dataset, with a total row count of 3313. The simulation and training variables used from Table 1’s dataset are as follows:

- Incident angle  $\theta_i$  of 10°, 20°, 30°, and 45°;
- Temperature  $T$  of 30 °C, 40 °C, and 50 °C;
- Wavelength  $\lambda$  of 450 nm, 520 nm, and 660 nm.

Changing these variables will impact the distance  $d$ .

These modified variables will collectively form the input dataset for a multi-layer perceptron (MLP) regression model, which will be used to derive the salinity percentage of water.

Feature scaling involves transforming the values of input elements so they fall within a specified range, typically between 0 and 1 or  $-1$  and 1. While scaling can improve the performance of some machine-learning models, it may not be necessary or beneficial for all datasets, as certain algorithms may be more sensitive. When deciding whether to scale the data, it is important to consider the characteristics of the dataset and the requirements of the machine-learning algorithm. Experimenting with different scaling methods can help determine which approach works better for your specific case. Our variable  $d$  was first rounded to 4 decimal places after which a RobustScaler is applied to all variables as the best option for the feature-scaling method, as it is designed to be robust to outliers in the data. Outliers are extreme values that may be significantly different from other values in the dataset and can potentially affect the performance of machine-learning models. RobustScaler works by centering and scaling the data similarly to other scaling methods, but, instead of using the mean and standard deviation, it uses the median and interquartile range (IQR). The IQR is the range between the first and third quartiles of the data, which is less sensitive to outliers than the full range.

After completing the preprocessing step, the subsequent stage involves initiating the training process using the MLP regressor from the sklearn library. Initially, we specify the input features that correspond to various  $d$  values obtained by simulating  $\lambda$  and  $\theta_i$ . This yields a total of 12 input  $d$  values. Additionally, we incorporate temperature  $T$  as an input parameter, making the total number of input parameters 13. The corresponding output variable represents the salinity percentage  $S$  ranging from 0% to 40%. The MLP Regressor comprises multiple layers of neurons, where each layer contains a group of neurons linked to all neurons in the previous and following layer. Let us denote the input to the first hidden layer as  $h_{11}, h_{12}, \dots, h_{1m}$ , where  $m$  represents the number of neurons in the first hidden layer. The output of the MLP Regressor can be expressed in Equation (3):

$$y = f(z, L) \quad (3)$$

where  $L$  is the number of layers in the network, and  $f$  is the activation function of the output layer.

The output of each hidden layer can be calculated using Equation (4):

$$h_{j,L} = f(Z_j, L) \quad (4)$$

where  $j$  is the neuron index in layer  $L$ ,  $(Z_j, L)$  are the weighted sum of the inputs to neuron  $j$  in layer  $L$ , and  $f$  is the activation function of the hidden layer.

The weighted sum  $(Z_j, L)$  can be calculated using Equation (5):

$$(Z_j, L) = \sum_i (W_{i,j,L} * (h_{i-1}, L)) + (b_j, L) \quad (5)$$

where  $W_{i,j,L}$  is the weight connecting neuron  $i$  in layer  $L - 1$  to neuron  $j$  in layer  $L$ ,  $(h_{i-1}, L)$  is the output of neuron  $i$  in layer  $L - 1$ , and  $(b_j, L)$  is the bias term of neuron  $j$  in layer  $L$ .

Finally, the output of the MLP regressor can be trained by minimizing the mean squared error (MSE) loss function, defined in Equation (6):

$$MSE = \left( \frac{1}{N} \right) \sum (y_i - f(z, L_i))^2 \quad (6)$$

where  $N$  is the number of training samples,  $y_i$  is the true output value for sample  $i$ , and  $(z, L_i)$  is the predicted output value for sample  $i$ .

The weights and biases of the network can be updated using backpropagation and a gradient descent optimization algorithm to minimize the *MSE* loss function. These equations describe the basic structure and training process of the *MLP* regressor. By adjusting the number of layers, the number of neurons in each layer, and the activation functions used, the *MLP* regressor can be tailored to specific regression tasks and provide accurate predictions for a wide range of applications.

In our training, we used the logistic activation function, commonly used in the *MLP* regressor. It is a sigmoidal function that maps any input value to a value between 0 and 1 as defined in Equation (7):

$$f(z) = \frac{1}{1 + e^{-z}} \quad (7)$$

where  $z$  is the weighted sum of the inputs to a neuron.

The logistic function is continuous and differentiable, which makes it suitable for use in the *MLP* regressor, where gradients are required for backpropagation. It can be interpreted as probability, where the output value represents the probability of the input belonging to a particular class. In the case of the *MLP* regressor, the output of the logistic function represents the predicted output value for a given set of input features. The logistic function has the following properties:

- It is a monotonic increasing function;
- Its output is bounded between 0 and 1;
- It is differentiable, and its derivative can be expressed in terms of the function itself as given in Equation (8):

$$f'(z) = f(z) * (1 - f(z)) \quad (8)$$

where  $f'(z)$  is the derivative of the logistic function.

Our *MLP* regressor structure is given below:

- Learning rate = 'adaptive';
- Random state = 0;
- Hidden layer sizes = 3;
- Activation = 'logistic';
- Solver = 'lbfgs';
- Alpha = 0.000001;
- Max iter = 10,000,000.

The 'lbfgs' solver is a popular choice for *MLP* regressors because it can handle large amounts of data and can converge quickly to a local minimum of the cost function. It is particularly advantageous for small- to medium-sized neural networks with many more datasets for training than the inputs of the model.

The alpha parameter is the L2 regularization term added to the cost function, which helps prevent overfitting by adding a penalty to the weights. A smaller alpha, such as 0.000001, suggests a smaller regularization term, which can lead to a more complex model that has greater variance. This can be beneficial if the training data is noisy or complex but can also lead to overfitting if the model is too complex for the data. The "Max iter" parameter specifies the maximum number of iterations for the solver to converge. A larger value, such as 10,000,000, allows the solver to continue iterating until convergence or until the maximum number of iterations is reached. However, a very large number of iterations may lead to overfitting or slow training, especially for large datasets.

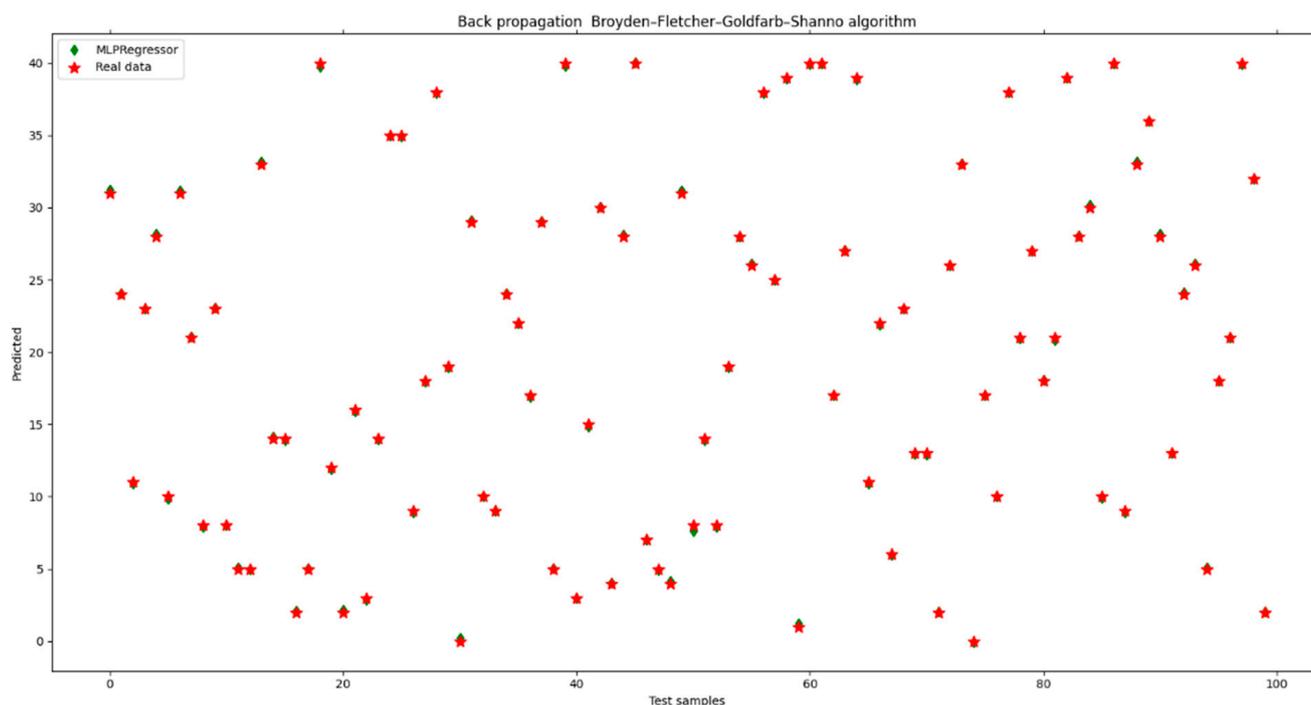
## 5. Results and Discussion

In Table 2, the error statistics show that the *MLP* model is very trustworthy, with  $R^2 = 0.999$  and *MAE* is 0.074. This substantiates the feasibility of the put-forward *MLP* model for input–output mapping.

**Table 2.** MLP model regressor errors.

Coefficient of Determination $R^2$	0.999
Mean Squared Error MSE	0.009
Root Mean Squared Error RMSE	0.094
Mean Absolute Error MAE	0.074

Figure 4 shows the MLP model’s error parameters and precision, where simulated and predicted test data align across all salinity states with only minimal oscillation. It is essential to note that the evaluation test data were different from the training data.



**Figure 4.** MLP prediction precision on the 100 samples forming the test dataset.

Table 3 provides a tabular representation of the discrepancy between the error in the salinity prediction using a novel input parameter and the MLP predictions. It is noteworthy that the data used for this analysis was not employed during the training or testing phase. Furthermore, the results displayed in Table 3 demonstrate the model’s ability to accurately predict the salinity values for untrained input data, indicating high performance levels. This highlights the relevance of the model’s ability to accurately process unfamiliar input data.

**Table 3.** Salinity percentage  $S$  (%).

Real	8	15	20	27	35	40	3	0
Predicted	7.97926	14.9341	19.9869	26.9965	34.946	39.8294	3.07405	0.016663
Error difference	0.02074	0.0659	0.0131	0.0035	0.054	0.1706	0.07405	0.016663

Few studies have examined the measurement of water salinity using MLP data generated from the optical principle. Reference [40] used convolutional neural networks and transfer learning to classify salt particles in seawater. The study achieves an accuracy of 90% and an f-score of 87%. However, due to the limited amount of training data, the deep-learning system suffered from the misclassification of saline salt particles. For this reason, our paper generates synthetic data that can be enriched with real data for improved

prediction. As evidenced by Figure 4 and Table 3, the model demonstrates a strong correlation between variable  $d$  and salinity, given that our research makes use of synthetic data devoid of anomalies. The  $d$  values are derived using COMSOL and the equations presented in [47,50].

Figure 4 depicts the error parameters and MLP precision, demonstrating a close alignment between the actual and predicted data across all salinity states, with minimal oscillation observed at certain data points. A more comprehensive statistical analysis of the model is presented in Table 3, where the model is tested with entirely new and unseen input data. The results displayed in Table 3 indicate that the model's predictions are satisfactory across the entire salinity range (0 to 40%), but, due to the use of synthetic data, model robustness for the real on-site measurements will decrease the accuracy [51,52].

## 6. Conclusions

In conclusion, this research study presents a machine-learning laser-based sensor design for accurate water salinity detection and COMSOL 2D-simulation design for synthetic data collection.

- Unlike other approaches that rely on current flow or chemicals, which can be harmful to underwater life and are complex for real on-site evaluations, our proposed method utilizes the light refraction optics concept to generate synthetic data for ML predictions.
- Synthetic data speed up the machine-learning training process, which is a challenging step in all AI applications. COMSOL 2D simulation reduces the time–cost process for obtaining input–output data used in the training process.
- The data obtained from COMSOL is validated by comparing the refractive indices calculated in COMSOL with those determined using Snell's law, trigonometry, and experimental research.
- The obtained results demonstrate the multi-layer perceptron (MLP) model's ability to accurately predict salinity values for previously unseen input data, indicating a high level of precision.
- Water salinity prediction is possible under diverse temperature settings.

By considering these significant findings, this research contributes to the advancement of sensor technology for precise and accessible water salinity detection. The cost-effective design of the sensor could make it easier to make a flexible multi-parameter sensor and can be used as a solution for a variety of purposes, such as life monitoring on contaminants in water, as affordable low-cost IoT technology ensuring its safety even in household water supply systems. Moreover, such a sensor could be implemented in water facilities next to petroleum lines for 24/7 leak detection as part of an IoT technology solution.

**Author Contributions:** Conceptualization, B.M. and S.V.; methodology, B.M. and S.V.; software, B.M. and S.V.; validation, B.M. and S.V.; formal analysis, B.M. and S.V.; investigation, B.M. and S.V.; resources, B.M. and S.V.; data curation, B.M. and S.V.; writing—original draft preparation, M.A. and M.H.; writing—review and editing, B.M., S.V., M.A. and M.H.; visualization, B.M., S.V., M.A. and M.H.; supervision, B.M. and S.V.; project administration, M.A. and M.H.; funding acquisition, B.M. and S.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Contact author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Millero, F.J.; Zhang, J.-Z.; Lee, K.; Campbell, D.M. Titration alkalinity of seawater. *Mar. Chem.* **1993**, *44*, 153–165. [CrossRef]
2. Wang, J.; Wang, W.-X. Salinity influences on the uptake of silver nanoparticles and silver nitrate by marine medaka (*Oryzias melastigma*). *Environ. Toxicol. Chem.* **2013**, *33*, 632–640. [CrossRef]
3. Johari, S.A.; Sarkheil, M.; Tayemeh, M.B.; Veisi, S. Influence of salinity on the toxicity of silver nanoparticles (AgNPs) and silver nitrate (AgNO<sub>3</sub>) in halophilic microalgae, *Dunaliella salina*. *Chemosphere* **2018**, *209*, 156–162. [CrossRef]
4. Agha, M.; Ennen, J.R.; Bower, D.S.; Nowakowski, A.J.; Sweat, S.C.; Todd, B.D. Salinity tolerances and use of saline environments by freshwater turtles: Implications of sea level rise. *Biol. Rev. Camb. Philos. Soc.* **2018**, *93*, 1634–1648. [CrossRef]
5. Huber, C.; Klimant, I.; Krause, C.; Werner, T.; Mayr, T.; Wolfbeis, O.S. Optical sensor for seawater salinity. *Anal. Bioanal. Chem.* **2000**, *368*, 196–202. [CrossRef]
6. Huber, C.; Klimant, I.; Krause, C.; Wolfbeis, O.S. Dual Lifetime Referencing as Applied to a Chloride Optical Sensor. *Anal. Chem.* **2001**, *73*, 2097–2103. [CrossRef]
7. An, S.; Ko, D.; Noh, J.; Lee, C.; Seo, D.; Lee, M.; Chang, J. ITO Nanoparticle Chemiresistive Sensor for Detecting Liquid Chemicals Diluted in Brine. *Trans. Electr. Electron. Mater.* **2022**, *23*, 107–112. [CrossRef]
8. Dong, T.; Barbosa, C. Capacitance Variation Induced by Microfluidic Two-Phase Flow across Insulated Interdigital Electrodes in Lab-On-Chip Devices. *Sensors* **2015**, *15*, 2694–2708. [CrossRef]
9. Wu, H.; Tan, C.; Dong, X.; Dong, F. Design of a Conductance and Capacitance Combination Sensor for Water Holdup Measurement in Oil–Water Two-Phase Flow. *Flow. Meas. Instrum.* **2015**, *46*, 218–229. [CrossRef]
10. Zhai, L.; Jin, N.; Gao, Z.; Wang, Z. Liquid Holdup Measurement with Double Helix Capacitance Sensor in Horizontal Oil–Water Two-Phase Flow Pipes. *Chin. J. Chem. Eng.* **2015**, *23*, 268–275. [CrossRef]
11. Huang, X.; Pascal, R.W.; Chamberlain, K.; Banks, C.J.; Mowlem, M.; Morgan, H. A Miniature, High Precision Conductivity and Temperature Sensor System for Ocean Monitoring. *IEEE Sens. J.* **2011**, *11*, 3246–3252. [CrossRef]
12. Ioc, Scor, and Iapso, “The international thermodynamic equation of seawater–2010: Calculation and use of thermodynamic properties,” Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, no. June. 2010. Scientific Committee on Oceanic Research. International Association for the Physical Sciences of the Ocean. Available online: <https://unesdoc.unesco.org/ark:/48223/pf0000188170.locale=en> (accessed on 1 June 2023).
13. McDougall, T.J.; Jackett, D.R.; Millero, F.J.; Pawlowicz, R.; Barker, P.M. A Global Algorithm for Estimating Absolute Salinity. *Ocean. Sci.* **2012**, *8*, 1123–1134. [CrossRef]
14. Xiao, A.; Huang, Y.; Zheng, J.; Chen, P.; Guan, B.-O. An Optical Microfiber Biosensor for CEACAM5 Detection in Serum: Sensitization by a Nanosphere Interface. *ACS Appl. Mater. Interfaces* **2019**, *12*, 1799–1805. [CrossRef]
15. Wang, Q.; Wang, X.-Z.; Song, H.; Zhao, W.-M.; Jing, J.-Y. A Dual Channel Self-Compensation Optical Fiber Biosensor Based on Coupling of Surface Plasmon Polariton. *Opt. Laser Technol.* **2020**, *124*, 106002. [CrossRef]
16. Liu, Q.; Ma, Z.; Wu, Q.; Wang, W. The Biochemical Sensor Based on Liquid-Core Photonic Crystal Fiber Filled with Gold, Silver and Aluminum. *Opt. Laser Technol.* **2020**, *130*, 106363. [CrossRef]
17. Rifat, A.A.; Mahdiraji, G.A.; Sua, Y.M.; Ahmed, R.; Shee, Y.G.; Adikan, F.R.M. Highly Sensitive Multi-Core Flat Fiber Surface Plasmon Resonance Refractive Index Sensor. *Opt. Express* **2016**, *24*, 2485. [CrossRef]
18. Lin, Q.; Hu, Y.; Yan, F.; Hu, S.; Chen, Y.; Liu, G.; Chen, L.; Xiao, Y.; Chen, Y.; Luo, Y.; et al. Half-Side Gold-Coated Hetero-Core Fiber for Highly Sensitive Measurement of a Vector Magnetic Field. *Opt. Lett.* **2020**, *45*, 4746. [CrossRef]
19. Momtaz, M.; Mou, J.R.; Kamrunnahar, Q.M.; Islam, M.d.A. Open-Channel-Based Dual-Core D-Shaped Photonic Crystal Fiber Plasmonic Biosensor. *Appl. Opt.* **2020**, *59*, 8856. [CrossRef]
20. Wang, H.; Dai, W.; Cai, X.; Xiang, Z.; Fu, H.; IEEE, M. Half-Side PDMS-Coated Dual-Parameter PCF Sensor for Simultaneous Measurement of Seawater Salinity and Temperature. *Opt. Fiber Technol.* **2021**, *65*, 102608. [CrossRef]
21. Fan, Y.; Xue, Y.; Wang, Y.; Liu, R.; Zhong, S. Combined LIBS and Raman Spectroscopy: An Approach for Salinity Detection in the Field of Seawater Investigation. *Appl. Opt.* **2022**, *61*, 1718. [CrossRef]
22. Hu, C.; Voss, K.J. In Situ Measurements of Raman Scattering in Clear Ocean Water. *Appl. Opt.* **1997**, *36*, 6962. [CrossRef] [PubMed]
23. Grosso, P. Refractometer Resolution Limits for Measuring Seawater Refractive Index. *Opt. Eng.* **2010**, *49*, 103603. [CrossRef]
24. Malardé, D.; Wu, Z.Y.; Grosso, P.; de Bougrenet de la Tocnaye, J.-L.; Le Menn, M. High-Resolution and Compact Refractometer for Salinity Measurements. *Meas. Sci. Technol.* **2008**, *20*, 015204. [CrossRef]
25. Aly, K.M.; Esmail, E. Refractive Index of Salt Water: Effect of Temperature. *Opt. Mater.* **1993**, *2*, 195–199. [CrossRef]
26. Chen, J.; Guo, W.; Xia, M.; Li, W.; Yang, K. In Situ Measurement of Seawater Salinity with an Optical Refractometer Based on Total Internal Reflection Method. *Opt. Express* **2018**, *26*, 25510. [CrossRef]
27. Qiu, Z. A Simple Optical Model to Estimate Suspended Particulate Matter in Yellow River Estuary. *Opt. Express* **2013**, *21*, 27891. [CrossRef]
28. Röttgers, R.; McKee, D.; Utschig, C. Temperature and Salinity Correction Coefficients for Light Absorption by Water in the Visible to Infrared Spectral Region. *Opt. Express* **2014**, *22*, 25093. [CrossRef]
29. Artlett, C.P.; Pask, H.M. New Approach to Remote Sensing of Temperature and Salinity in Natural Water Samples. *Opt. Express* **2017**, *25*, 2840. [CrossRef]
30. Schmidt, H.; Wolf, H.; Hassel, E. A Method to Measure the Density of Seawater Accurately to the Level of 10<sup>−6</sup>. *Metrologia* **2016**, *53*, 770–786. [CrossRef]

31. Woosley, R.J.; Huang, F.; Millero, F.J. Corrigendum to “Estimating Absolute Salinity (SA) in the World’s Oceans Using Density and Composition” [Deep-Sea Res. 193 (2014) 14–20]. *Deep. Sea Res. Part. I Oceanogr. Res. Pap.* **2018**, *142*, 145. [[CrossRef](#)]
32. Zhang, X.; Peng, W. Temperature-Independent Fiber Salinity Sensor Based on Fabry-Perot Interference. *Opt. Express* **2015**, *23*, 10353. [[CrossRef](#)]
33. Wu, C.; Sun, L.; Li, J.; Guan, B.-O. Highly Sensitive Evanescent-Wave Water Salinity Sensor Realized with Rectangular Optical Microfiber Sagnac Interferometer. In Proceedings of the 23rd International Conference on Optical Fibre Sensors, Santander, Spain, 2 June 2014; SPIE Proceedings: Bellingham, WA, USA, 2014. [[CrossRef](#)]
34. Jaddoa, M.F.; Jasim, A.A.; Razak, M.Z.A.; Harun, S.W.; Ahmad, H. Highly Responsive NaCl Detector Based on Inline Microfiber Mach-Zehnder Interferometer. *Sens. Actuators A Phys.* **2016**, *237*, 56–61. [[CrossRef](#)]
35. Li, L.; Xia, L.; Wuang, Y.; Ran, Y.; Yang, C.; Liu, D. Novel NCF-FBG Interferometer for Simultaneous Measurement of Refractive Index and Temperature. *IEEE Photonics Technol. Lett.* **2012**, *24*, 2268–2271. [[CrossRef](#)]
36. Profiling | Sea-Bird Scientific-Overview | Sea-Bird. Available online: <https://www.seabird.com/profiling/family?productCategoryId=54627473767> (accessed on 23 May 2023).
37. Grosso, P.; Menn, M.L.; De Bougrenet De La Tournay, J.-L.; Yan Wu, Z.; Malardé, D. Practical versus Absolute Salinity Measurements: New Advances in High Performance Seawater Salinity Sensors. *Deep. Sea Res. Part. I Oceanogr. Res. Pap.* **2010**, *57*, 151–156. [[CrossRef](#)]
38. Wang, Y.; Guo, J.; Yang, Z.; Dou, Y.; Chang, X.; Sun, R.; Zuo, G.; Yang, W.; Liang, C.; Hao, Y.; et al. Computer Prediction of Seawater Sensor Parameters in the Central Arctic Region Based on Hybrid Machine Learning Algorithms. *IEEE Access* **2020**, *8*, 213783–213798. [[CrossRef](#)]
39. Cipollini, P.; Corsini, G.; Diani, M.; Grasso, R. Retrieval of Sea Water Optically Active Parameters from Hyperspectral Data by Means of Generalized Radial Basis Function Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1508–1524. [[CrossRef](#)]
40. Alshehri, M.; Kumar, M.; Bhardwaj, A.; Mishra, S.; Gyani, J. Deep Learning Based Approach to Classify Saline Particles in SeaWater. *Water* **2021**, *13*, 1251. [[CrossRef](#)]
41. Wang, J.; Ding, J.; Yu, D.; Teng, D.; He, B.; Chen, X.; Ge, X.; Zhang, Z.; Wang, Y.; Yang, X.; et al. Machine Learning-Based Detection of Soil Salinity in an Arid Desert Region, Northwest China: A Comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci. Total Environ.* **2020**, *707*, 136092. [[CrossRef](#)]
42. Mourched, B.; Abdallah, M. Design and Characterization of a New Microscopy Probe Using COMSOL and ANSYS. *IJM* **2022**, *16*, 95–106. [[CrossRef](#)]
43. Mourched, B.; Abboud, N.; Abdallah, M.; Moustafa, M. Electro-Thermal Simulation Study of MOSFET Modeling in Silicon and Silicon Carbide. *IJM* **2022**, *16*, 383–394. [[CrossRef](#)]
44. Mourched, B.; Nativel, E.L.; Kribich, R.; Falgayrettes, P.; Gall-Borruet, P. Study of Light Emission and Collection in a Transparent Dielectric Cantilever-based Near-field Optical Probe. *J. Microsc.* **2015**, *262*, 3–11. [[CrossRef](#)]
45. Mourched, B.; Ferko, N.; Abdallah, M.; Neji, B.; Vrtagic, S. Study and Design of a Machine Learning-Enabled Laser-Based Sensor for Pure and Sea Water Determination Using COMSOL Multiphysics. *Appl. Sci.* **2022**, *12*, 6693. [[CrossRef](#)]
46. Mourched, B.; Hoxha, M.; Abdelgalil, A.; Ferko, N.; Abdallah, M.; Potams, A.; Lushi, A.; Turan, H.I.; Vrtagic, S. Piezoelectric-Based Sensor Concept and Design with Machine Learning-Enabled Using COMSOL Multiphysics. *Appl. Sci.* **2022**, *12*, 9798. [[CrossRef](#)]
47. Index of Refraction of Seawater and Freshwater as a Function of Wavelength and Temperature | Parrish Research Group | Oregon State University. Available online: <https://research.engr.oregonstate.edu/parrish/index-refraction-seawater-and-freshwater-function-wavelength-and-temperature> (accessed on 23 May 2023).
48. Austin, R.W.; Halikas, G. The Index of Refraction of Seawater. Available online: <https://escholarship.org/uc/item/8px2019m> (accessed on 23 May 2023).
49. Bass, M.; DeCusatis, C.; Enoch, J.M.; Lakshminarayanan, V.; Li, G.; MacDonald, C.; Mahajan, V.N.; Stryland, E.V. *Handbook of Optics, Third Edition Volume IV: Optical Properties of Materials, Nonlinear Optics, Quantum Optics*; The Optical Society of America, the McGrawHill Companies: New York, NY, USA, 2009; ISBN 978-0-07-149892-0.
50. Quan, X.; Fry, E.S. Empirical Equation for the Index of Refraction of Seawater. *Appl. Opt.* **1995**, *34*, 3477. [[CrossRef](#)] [[PubMed](#)]
51. Krichen, M.; Mihoub, A.; Alzahrani, M.Y.; Adoni, W.Y.H.; Nahhal, T. Are Formal Methods Applicable to Machine Learning and Artificial Intelligence? In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022. [[CrossRef](#)]
52. Raman, R.; Gupta, N.; Jeppu, Y. Framework for Formal Verification of Machine Learning Based Complex System-of-Systems. *Insight* **2023**, *26*, 91–102. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.