remote sensing

MDPI

Article

Developing a Generalizable Spectral Classifier for Rhodamine Detection in Aquatic Environments

Ámbar Pérez-García [1,*], Alba Martín Lorenzo [1], Emma Hernández [1], Adrián Rodríguez-Molina [1], Tim H. M. van Emmerik [2] and José F. López [1]

[1] Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, 35001 Las Palmas, Spain; alba.martin112@alu.ulpgc.es (A.M.L.); ehernandez@iuma.ulpgc.es (E.H.); armolina@iuma.ulpgc.es (A.R.-M.); lopez@iuma.ulpgc.es (J.F.L.)

[2] Hydrology and Environmental Hydraulics Group, Wageningen University, 6708 BP Wageningen, The Netherlands; tim.vanemmerik@wur.nl

* Correspondence: ambar.perez@ulpgc.es

Abstract: In environmental studies, rhodamine dyes are commonly used to trace water movements and pollutant dispersion. Remote sensing techniques offer a promising approach to detecting rhodamine and estimating its concentration, enhancing our understanding of water dynamics. However, research is needed to address more complex environments, particularly optically shallow waters, where bottom reflectance can significantly influence the spectral response of the rhodamine. Therefore, this study proposes a novel approach: transferring pre-trained classifiers to develop a generalizable method across different environmental conditions without the need for in situ calibration. Various samples incorporating distilled and seawater on light and dark backgrounds were analyzed. Spectral analysis identified critical detection regions (400–500 nm and 550–650 nm) for estimating rhodamine concentration. Significant spectral variations were observed between light and dark backgrounds, highlighting the necessity for precise background characterization in shallow waters. Enhanced by the Sequential Feature Selector, classification models achieved robust accuracy (>90%) in distinguishing rhodamine concentrations, particularly effective under controlled laboratory conditions. While band transfer was successful (>80%), the transfer of pre-trained models posed a challenge. Strategies such as combining diverse sample sets and applying the first derivative prevent overfitting and improved model generalizability, surpassing 85% accuracy across three of the four scenarios. Therefore, the methodology provides us with a generalizable classifier that can be used across various scenarios without requiring recalibration. Future research aims to expand dataset variability and enhance model applicability across diverse environmental conditions, thereby advancing remote sensing capabilities in water dynamics, environmental monitoring and pollution control.

Keywords: dye tracking; rhodamine; artificial intelligence; band selection

check for updates

Citation: Pérez-García, Á.; Martín Lorenzo, A.; Hernández, E.; Rodríguez-Molina, A.; van Emmerik, T.H.M.; López, J.F. Developing a Generalizable Spectral Classifier for Rhodamine Detection in Aquatic Environments. *Remote Sens.* **2024**, *16*, 3090. https://doi.org/10.3390/rs16163090

Academic Editor: Akira Iwasaki

Received: 30 June 2024
Revised: 9 August 2024
Accepted: 15 August 2024
Published: 22 August 2024

1. Introduction

Understanding the water movement and particle transport in the ocean and rivers is crucial for giving a fast response to environmental disasters. Although numerous mathematical models predict how these agents are transported, they often come with high computational costs and may not be adaptable to the highly variable conditions encountered in real-world scenarios, limiting their effectiveness for quick responses [1]. Thus, to understand and track hazardous agents in oceans and waterways, new techniques and tools are necessary [2].

Dye tracers offer a simpler and effective solution. They are widely used to study the transport and dispersion of particles in aqueous environments, such as in the ocean [3], coastal areas [4,5], rivers [6], and lakes [7]. Rhodamine is a synthetic dye commonly used as a tracer in environmental studies due to its distinctive colour and high visibility. Its applications include monitoring water flow, tracking pollutant dispersion, and studying hydrodynamic processes in aquatic environments. Rhodamine is a key tool for understanding complex water movement and contamination patterns. Current methods of measuring rhodamine concentration involve collecting samples for later laboratory analysis [8] or using fluorimeters for in situ measurements [9]. However, these techniques are costly and spatially limited, capturing information only at the local levels [10]. In this context, aerial and near-field remote sensing emerges as a promising alternative, offering greater spatial coverage than traditional in situ sampling methods [11].

Multispectral and hyperspectral remote sensing have proven to be effective in detecting and mapping the concentrations of dye tracers in aquatic environments [12]. Several studies successfully utilized multispectral and hyperspectral sensors onboard piloted aerial platforms to identify rhodamine, also in diverse aquatic environments, including the ocean [13–15], coasts [16], rivers [10,17], and lakes [1]. Some studies focused on comparing the intensity of RGB images [18,19], while others utilized hyperspectral sensors to capture surface reflectance [20]. Combining spectral technology with the emerging use of uncrewed aerial vehicles (UAVs) allows for conducting studies in much detail, with greater flexibility and detecting patterns at sub-meter resolution [12]. In addition, the development of lighter and more specialized optical sensors represents an opportunity to have a sensor with specialized spectral bands for rhodamine detection and concentration estimation at a low price [21].

Despite these advancements, several knowledge gaps remain. The widely used Optimum Band Ratio Analysis (OBRA) [10,22] method for selecting bands in hyperspectral and multispectral sensors to detect rhodamine requires in situ concentration measurements for calibration, limiting its transferability to different locations. Furthermore, most rhodamine studies are conducted in controlled environments, such as experimental channels, water tanks, or deep lakes, where the impact of background reflections is minimized (optically deep waters) [12]. Only a few studies have analyzed real-world scenarios, such as rivers with varying degrees of turbidity or optically shallow waters where the bottom reflects light [12,22]. The optical characterization of rhodamine in aquatic environments presents several challenges, especially in coastal areas, shallow rivers or turbid environments. As a semi-transparent solution, its spectral signature captured by sensors is influenced by the optical properties of suspended components and background reflections. However, more in-depth work is needed to understand the effect of water properties and bottom reflections of optically shallow waters on rhodamine's spectral signature and the transferability of detection models across varied conditions.

In this study, we conducted experiments to measure rhodamine concentrations with the goal of improving detection methods. We first spectrally analyzed the rhodamine in the laboratory with different backgrounds and water types. Using artificial intelligence algorithms, specifically Sequential Feature Selection (SFS) and Random Forest (RF) models, we identified the key spectral bands for rhodamine detection and concentration estimation. We then evaluated the transferability of these bands and the trained classifiers across different water types and backgrounds. This approach aimed to minimize the need for in situ calibration and enhance the robustness of detection models.

In our paper, we present the identification of key spectral bands for rhodamine estimation, the influence of background types on spectral signatures, and the transferability of detection models. We found that combining all samples to train the classification model, and also applying the first derivative [23] to favor the distinction of spectral signatures, improved the transfer of the models to all samples. These findings provide valuable insights for developing effective and affordable remote sensing tools to monitor rhodamine and, by extension, other pollutants in aquatic environments, especially in optically shallow waters. This research is particularly significant, as it supports the development of low-cost, multispectral cameras for environmental monitoring, contributing to more efficient and scalable pollution reduction strategies.

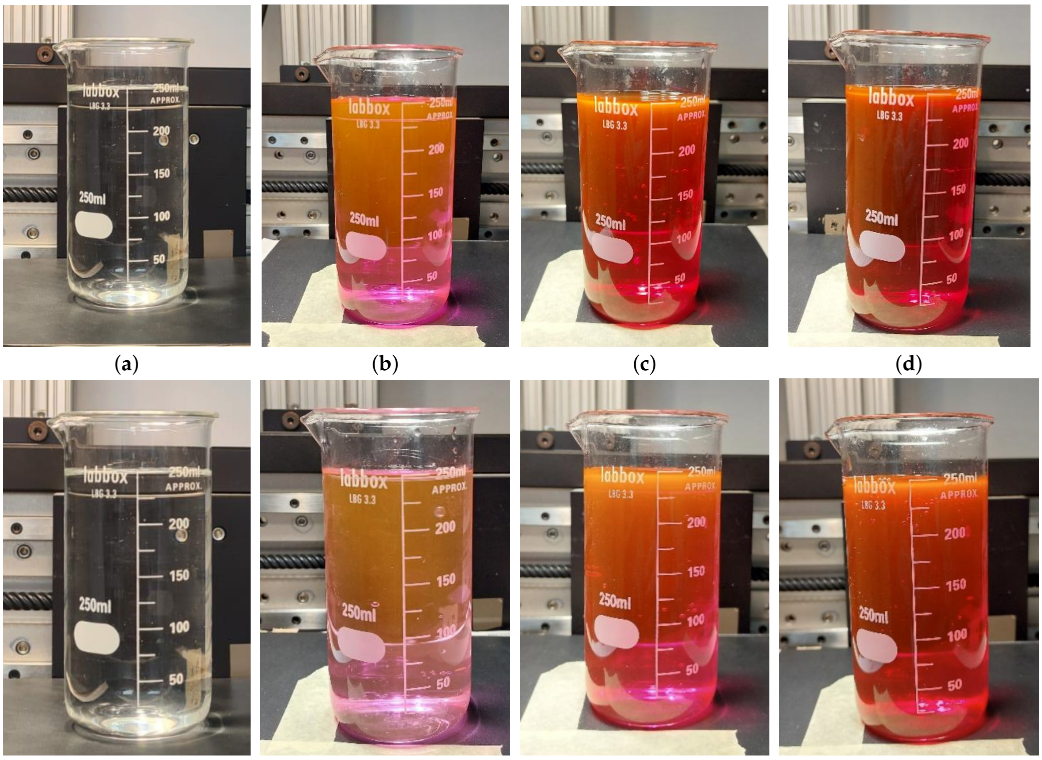
2. Materials and Methods

The methodology followed in this experiment can be divided into preparing the rhodamine samples, performing their spectral analysis, and employing the band selection methodology for classifying rhodamine concentrations. Rhodamine samples were prepared in beakers with distilled and seawater at concentrations of 1 mg/L, 15 mg/L, and 30 mg/L. The spectral signatures of samples on two different backgrounds were obtained with hyperspectral cameras. The spectral signatures of each concentration were provided to the band selection algorithm to obtain the most influential bands for classifying the rhodamine concentration.

2.1. Rhodamine Samples

The Rhodamine Water Tracer, hereafter referred to as rhodamine, is a fluorescent dye primarily used as a tracer in aquatic environments. The company Elittoral [24] from Las Palmas de Gran Canaria, Spain, has procured the rhodamine from ThermoFisher Scientific [25] from Massachusetts, USA, identified by the chemical codes CAS 37299-86-8 and 7732-18-5, with catalogue number 446971000. Initially, the rhodamine is highly concentrated at 20%, or 200 g/L, requiring dilution in water to achieve an appropriate concentration for its discharge into the sea. The beakers are from Labbox [26] in Barcelona, Spain, reference BKT3-250-012. They have a measurable volume of 250 mL, an outer diameter of 60 mm, and a height of 123 mm. When the beaker is filled with 250 mL of liquid, the height of the fluid will be 100 mm.

This study uses 250 mL solutions of seawater and distilled water with varying concentrations of rhodamine provided by Elittoral. Solutions with different concentrations of rhodamine are produced: 1 mg/L, 15 mg/L, and 30 mg/L (Figure 1). In addition to the rhodamine dilutions, pure seawater and distilled water samples are also included in the study. These pure samples serve as reference spectra and are essential for the comparative analysis of rhodamine-contaminated samples.



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

Figure 1. Rhodamine samples for distilled and seawater at different concentrations. (**a**) Distilled 0 mg/L; (**b**) Distilled 1 mg/L; (**c**) Distilled 15 mg/L; (**d**) Distilled 30 mg/L; (**e**) Seawater 0 mg/L; (**f**) Seawater 1 mg/L; (**g**) Seawater 15 mg/L; (**h**) Seawater 30 mg/L.

Hyperspectral signatures of all samples were collected over two different backgrounds. Since the rhodamine solutions are semi-transparent, the background signature likely influences the signature captured by the sensor. We placed two backgrounds underneath the beaker to observe these differences: a white sheet of paper as the light background and a low-reflective black foam material [27] as the dark background.

2.2. Hyperspectral Setup

We took images of the rhodamine samples in the Hyperspectral Laboratory at IUMA [27] in Las Palmas de Gran Canaria, Spain. The system shown in Figure 2 aims to acquire images with pushbroom hyperspectral cameras. It includes a motorized linear stage for linear motion and a light source emitting uniformly. Illumination comes from a 150 W Quartz Tungsten-Halogen (QTH) lamp with broadband emission between 400 nm and 2500 nm (VIS and NIR spectral range). Images were captured using a Specim FX10 camera (Konica Minolta Company, Oulu, Finland) [28]. The FX10 is a hyperspectral camera covering the visible and near-infrared range (VNIR) from 400 nm to 1000 nm, with 224 bands, a spectral resolution of 5.5 nm, a spatial sampling of 1024 pixels, and a field of view (FoV) of 38 degrees. This spectral range was chosen because rhodamine has the greatest response in the VNIR.

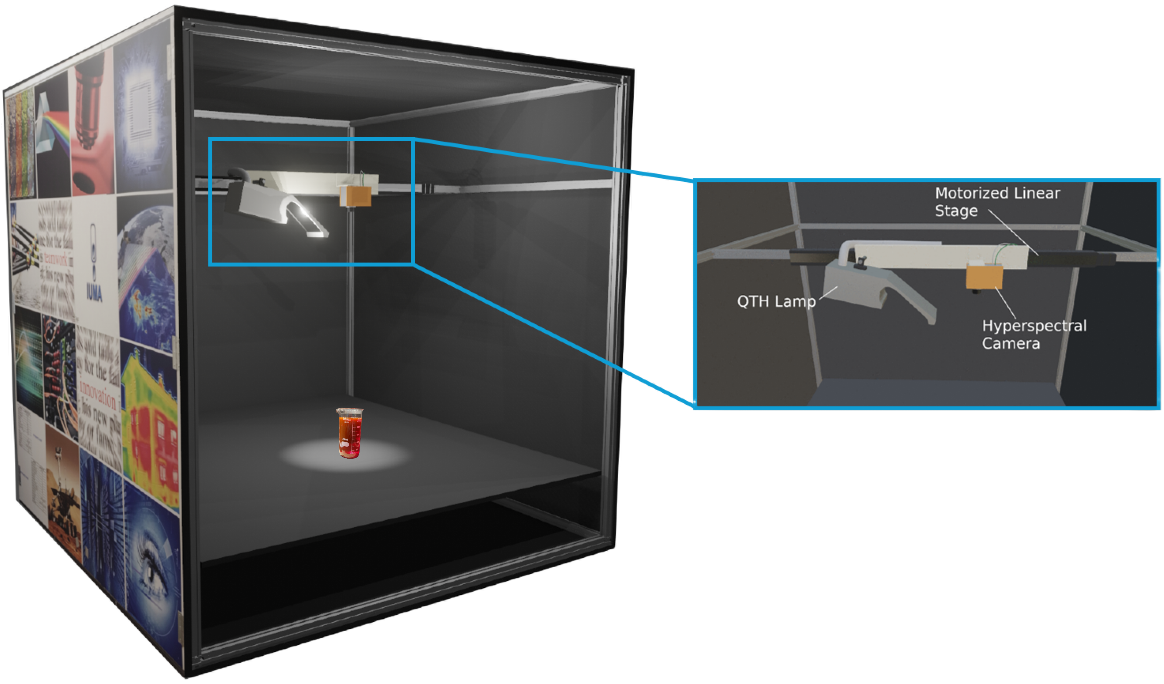


Figure 2. A 3D model of the acquisition system (adapted with permission from [27], under a Creative Commons Attribution (CC BY) 4.0 license. Copyright 2022).

We performed single-point reflectance calibration (Equation (1)) before starting measurements to avoid sensor saturation. This pre-processing involves white and dark (0-photon) references to calculate the reflectance of each pixel from its radiance. The white reference is a high-reflectance Zenith Polymer [29], and the dark reference is obtained by covering the camera lens:

$$reflectance = \frac{radiance - dark}{white - dark} \tag{1}$$

The measurements are taken by scanning the FX10 camera over the different beakers. Since the light source is positioned ahead of the camera as shown in Figure 2, this setup produces glints on the beaker, shadows in certain areas, and increased intensity in others due to internal reflections and external refractions of the glass.

The datasets used for the artificial intelligence model are generated by extracting pixels from the bottom areas of the beaker where no glints are present, while avoiding shadows. However, due to the varying brightness within the beaker, the standard deviation of the classes will be high, indicating variation in intensity, though not in the shape of the spectral signature. The number of pixels per class will not be uniform, as shown in Table 1.

Table 1. Number of pixels per class.

	Dark Background		Light Background		
	Distilled	**Sea**	**Distilled**	**Sea**	**Total**
0 mg/L	346	624	233	299	1502
1 mg/L	857	668	821	545	2891
15 mg/L	1302	406	965	650	3323
30 mg/L	924	346	866	750	2886

Efforts have been made to balance the glints and shadowed areas within the beaker. However, the primary goal of this study was to develop a generalizable classifier, and having data with variability is advantageous, as it better reflects the reality of experimental conditions.

2.3. Methodology

The methodology is divided into training, result analysis, and transfer (Figure 3). During training, data are provided to the band selection model. This model provides evaluation metrics, which are analyzed to determine the optimal number of bands. The band selection model [30] also identifies the best bands and provides pre-trained classifiers for the optimal number of bands. The final stage involves transferring the best bands and pre-trained models to other scenarios to assess their performance and determine if the classifier is generalizable.

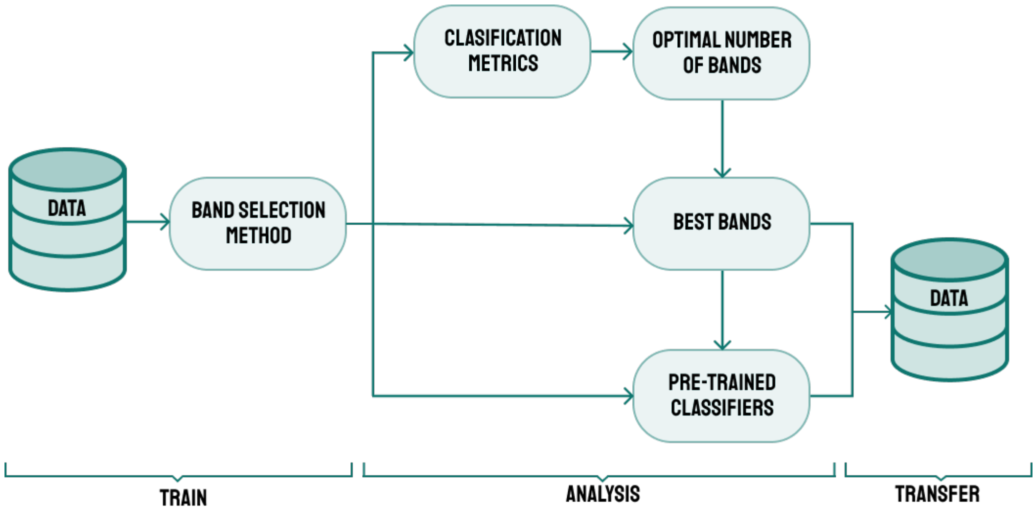


Figure 3. Methodology for transferring results and obtaining a generalizable classifier.

The methodology integrates hyperparameterized classifiers with feature selectors to provide optimized bands of interest for classification [30]. The procedure is illustrated in Figure 4. The data are divided into training sets for the feature selectors and classifiers, as well as a validation set. Employing the information provided by the classifiers, the feature selectors determine the bands of interest. Subsequently, the classifiers are retrained exclusively with these bands. Finally, the performance of the updated classifiers is assessed using the validation dataset to obtain classification metrics.

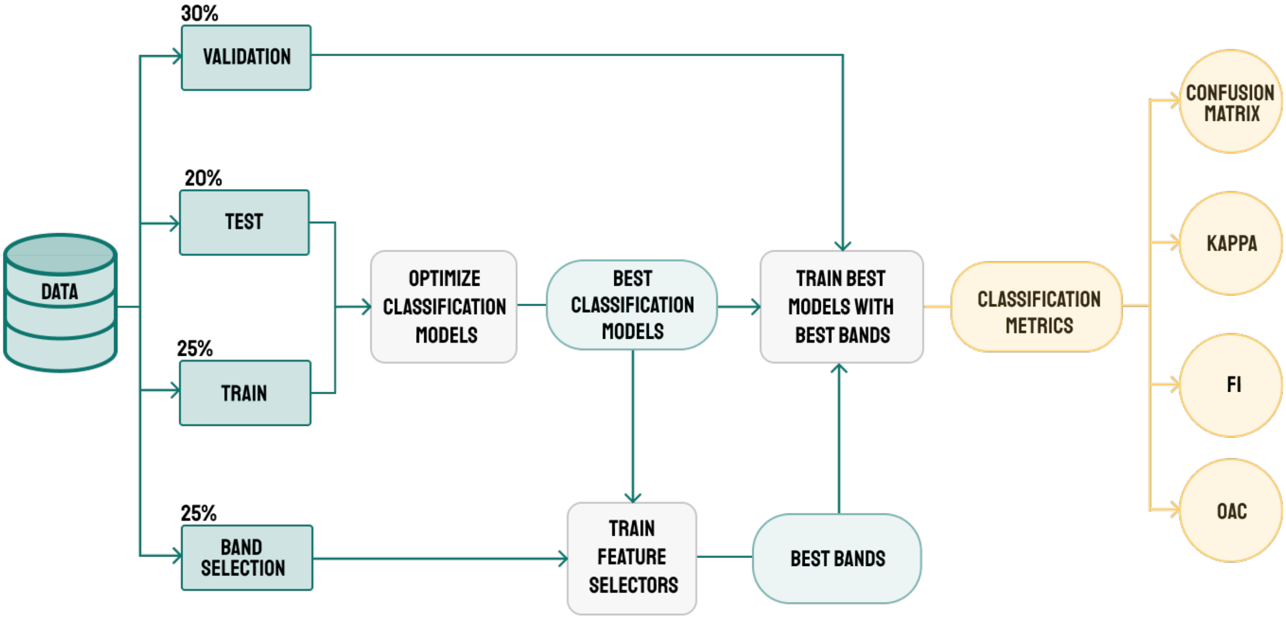


Figure 4. Band selection method (adapted with permission from [30], Copyright 2024, IEEE).

For this study, we combined two feature selectors, Sequential Feature Selector (SFS) and Select From Model (SFM) [31], with three classifiers, Random Forest (RF) [32], Logistic Regression (LR) [33], and Linear Support Vector Machine (SVM) [34]. SFS is a sequential search technique that iteratively adds or removes features to improve the classifier's performance. SFM ranks features based on a model's coefficients or importance, facilitating the selection of efficient feature subsets [31]. RF is a widely used classifier, employing ensemble learning to combine predictions from multiple decision trees [35]. LR is a linear model for binary classification [36]. Linear SVM identifies hyperplanes for optimal class separation [37].

We evaluated the performance of rhodamine concentration classification using the classification metrics on the validation dataset. The overall accuracy (OAC), also called accuracy, represents the proportion of correct predictions out of the total samples [38]. The F1 score is particularly useful with imbalanced class distributions, as it emphasizes the accuracy of the smaller classes [39]. The Kappa statistic (κ) measures inter-rater agreement for categorical items, adjusting for chance agreement, and is particularly relevant for uneven datasets [40]. The confusion matrix provides a detailed decomposition of predicted classification labels versus real labels [41].

The performance of all pairs of feature selectors and classifiers was evaluated for different numbers of bands of the FX10 camera [28]. By plotting the accuracy values for each band on a graph, we were able to use the elbow method [42] to determine the optimal number of bands. This method selects the number of bands at the point where the curve bends, forming an elbow and indicating a slowdown in accuracy improvement [42]. For the optimal number of bands, all classification metrics were calculated.

We also analyzed whether the different feature selector–classifier pairs identify the optimal bands in the same area of the electromagnetic spectrum. For this purpose, we grouped the wavelengths into sections of 25 nm wide for several reasons. First, the spectral signatures of different concentrations of rhodamine are continuous and do not exhibit abrupt changes, making it practical to group the wavelengths for better visualization of large-scale behavior. Second, grouping the bands helps us understand their potential utility in multispectral sensors. The hyperspectral camera used in this study has a spectral resolution of 5.5 nm [28], so it is appropriate to group them, given that multispectral sensors typically have a bandwidth of approximately 20 nm or 30 nm. Third, several studies support that a high level of spectral detail is unnecessary for rhodamine detection, indicating that sensors with broader bands could still provide reliable concentration estimates [20,22].

Finally, we can assess the transferability of the results to different scenarios. We can define different levels of transfer, such as the transfer of the bands of interest or the transfer of pre-trained classification models. Transferring the bands of interest is a straightforward way to transmit part of the knowledge acquired from one sample to another. It involves using the bands identified by the classification model in one scenario to train the classifier (exclusively with those wavelengths) with another scenario, whether it has a different type of water or a different background.

Transferring pre-trained classification models is more complex because these models learn the specific characteristics of each scenario, i.e., the reflectance value of the spectral signature. One strategy that can be employed to improve the transfer of pre-trained models is to provide the classifier with a combination of samples from all types of water and backgrounds for training. Another strategy is calculating the first derivative of the spectral signatures, which helps identify variations and trends in the data, thereby aiding in the differentiation of spectral signatures.

If the spectral behavior is similar enough across different scenarios, it is possible to successfully transfer the models, and the classifiers can be used in multiple scenarios without retraining. This would decrease the required in situ calibration measurements and reduce computational time.

3. Results

3.1. Spectral Analysis—Beaker Influence

An important aspect to consider during the analysis of the results is the influence of the beaker on the spectral signature of its liquid content. The reflections and refractions caused by the glass lead to non-uniform spectral intensity throughout the beaker, resulting in a higher standard deviation in the spectral signatures of each class.

To elaborate on the reflections caused by the beaker, Figure 5 compares the spectral behaviour of the light and dark backgrounds, both with and without the beaker filled with seawater. The results indicate that the primary effect of the background is to modulate the intensity of the reflectance, with greater influence between 400 nm and 500 nm. Therefore, the importance of this spectral range will vary depending on the illumination of the scene and the scattering capacity of the water and the background. It should also be noted that the reflectance value of the dark background increases when the beaker is placed, indicating that reflections and refractions occur in the beaker, which increases the reflectance.

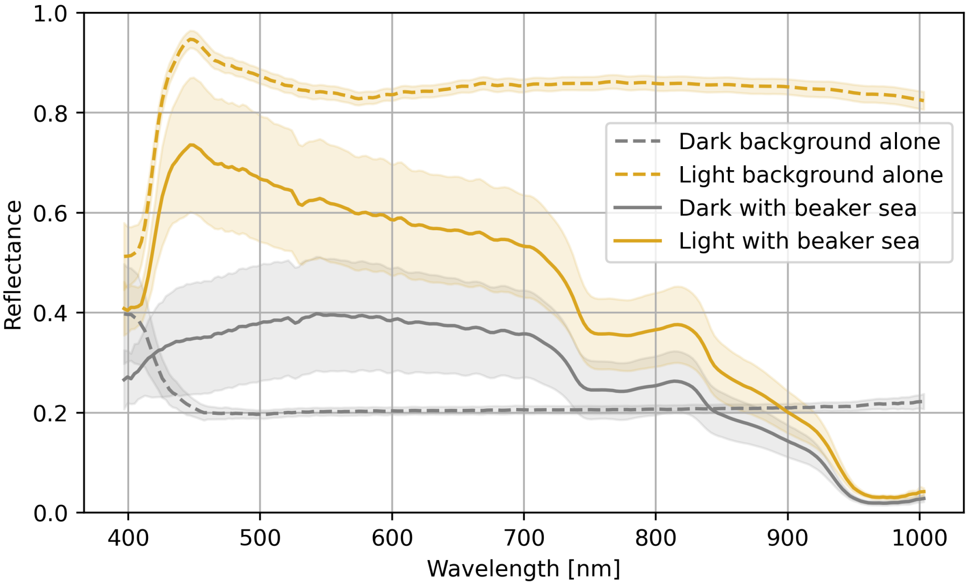
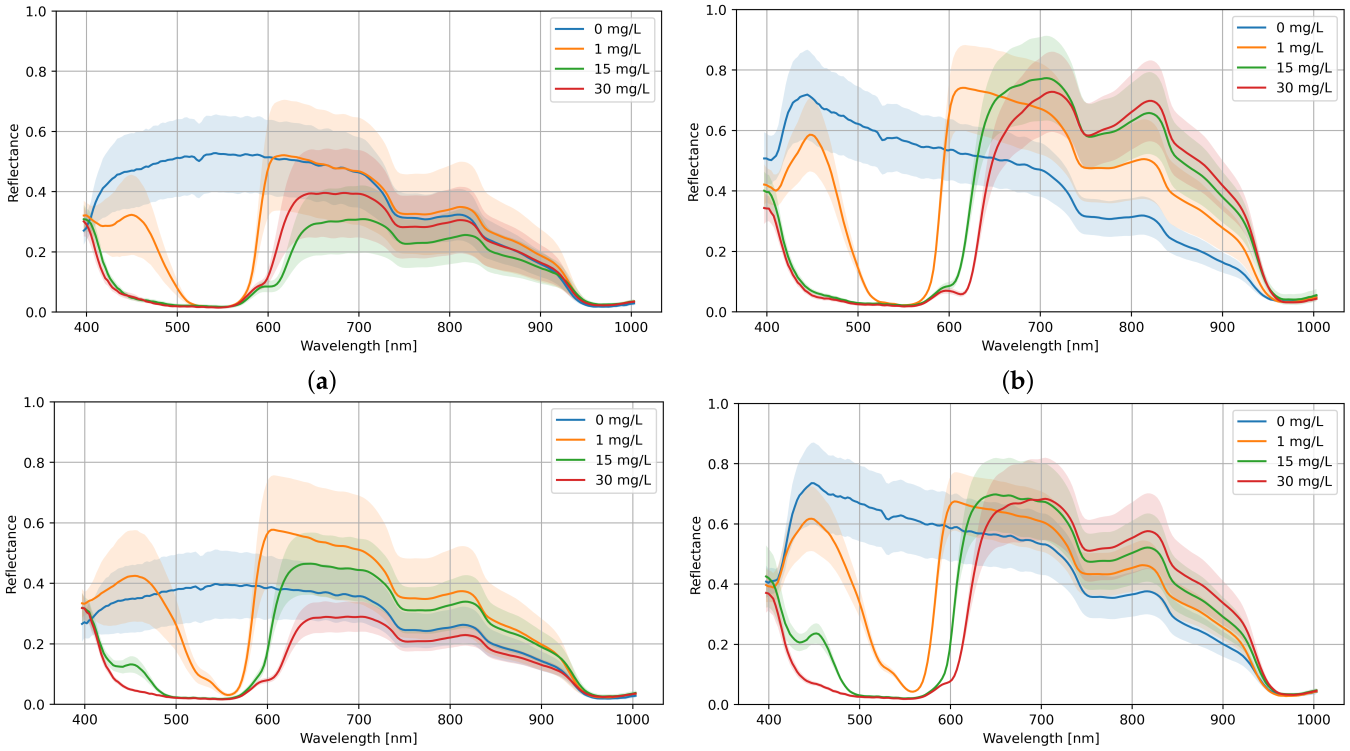


Figure 5. Mean spectral signature and standard deviation (shaded in the corresponding colour) of the backgrounds with and without the beaker.

3.2. Spectral Analysis—Rhodamine

We observed variations in spectral response to rhodamine concentrations at different wavelengths (Figure 6). We had four samples that combined distilled and seawater with light and dark backgrounds. In all samples, an increase in rhodamine concentration resulted in a shift of the reflectance peak around 600 nm towards longer wavelengths, giving the solution an increasingly orange hue. Initially, no differences were observed between distilled water and seawater samples, but noticeable differences emerged between light and dark backgrounds, especially at low rhodamine concentrations or when it was absent.



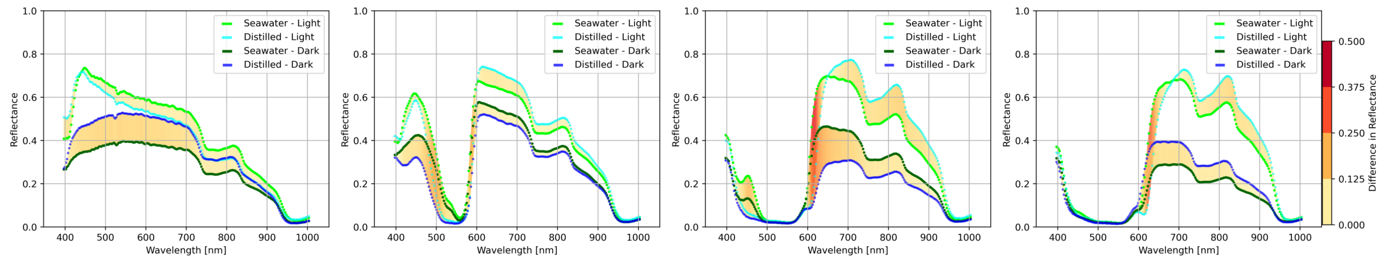
(**a**) (**b**)

(**c**) (**d**)

Figure 6. Mean spectra with standard deviation (shaded in the corresponding colour) for each concentration and sample. (**a**) Distilled water with a dark background; (**b**) distilled water with a light background; (**c**) sea water with a dark background; (**d**) sea water with a light background.

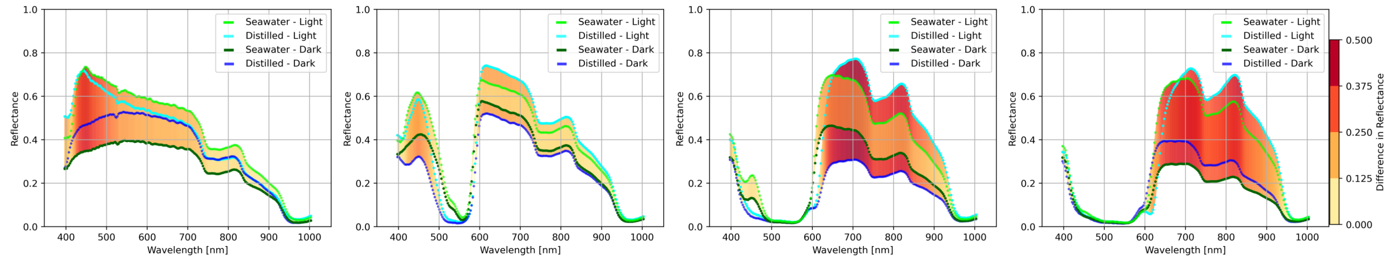
The most significant difference for both backgrounds was found at a concentration of 15 mg/L, specifically in the wavelength range of 600 nm to 650 nm (Figure 7). The samples exhibited a spectral difference of between 0.1 and 0.3 reflectance units; only the 15 mg/L sample with a light background with 0.37 reflectance units exceeded this value. The spectral differences between the two types of water were insufficient to conclusively distinguish between them.

More significant differences were observed when comparing the same water sample on the two different backgrounds, with discrepancies varying between 0.2 and 0.47 reflectance units. In Figure 8, a substantial decrease in reflectance is observed in the samples on a dark background, particularly pronounced in the red and near-infrared range (600 nm to 900 nm). The 15 mg/L and 30 mg/L samples show the most noticeable drops, reaching a difference between the spectral signatures of 0.47 and 0.41 for distilled water. Additionally, in the pure seawater samples, there is a reflectance peak associated with the blue colour, around 450 nm, on the light background that is not observed on the dark background, resulting in a difference in the spectral responses of 0.39.



a) *b)* *c)* *d)*

Figure 7. Spectral difference between distilled and seawater for the two backgrounds: (**a**) 0 mg/L; (**b**) 1 mg/L; (**c**) 15 mg/L; (**d**) 30 mg/L.



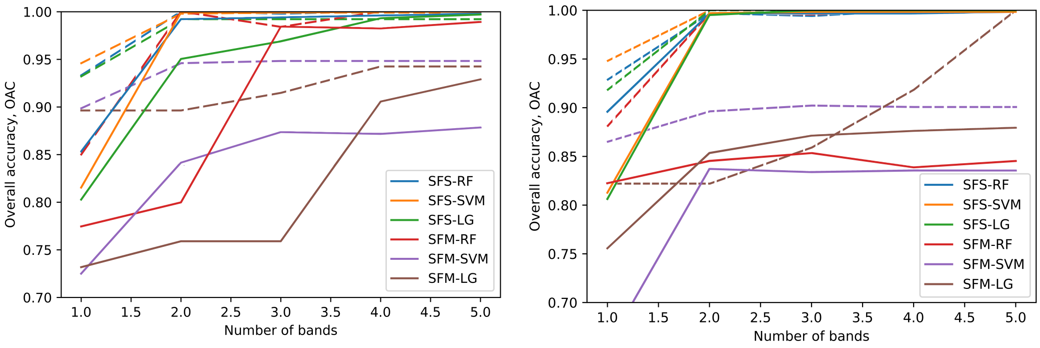
a) *b)* *c)* *d)*

Figure 8. Spectral difference between backgrounds: (**a**) 0 mg/L; (**b**) 1 mg/L; (**c**) 15 mg/L; (**d**) 30 mg/L.

These findings highlight the influence of background type on the spectral signature of rhodamine and emphasize the need to consider background effects in remote sensing applications to improve rhodamine detection accuracy.

3.3. Classification

The first step is identifying the optimal number of bands for detecting rhodamine concentrations. As illustrated in Figure 9, OAC increases with the number of bands provided to each feature selector–classifier pair. The OAC achieved is above 80% with few bands across all samples, demonstrating the strong differentiating power of the classifiers. There is no apparent difference between the performance of distilled water and seawater. The models performed better with a light background than with a dark one, achieving greater accuracy faster. Additionally, it is evident that classifiers combined with the SFS feature selector yielded more satisfactory OAC. The optimal number of bands for detecting rhodamine using the elbow method [42] was two for all samples.



(**a**) (**b**)

Figure 9. OAC of all model combinations for dark (solid line) and light (dashed line) backgrounds. (**a**) Distilled water; (**b**) seawater.

The SFS feature selector outperformed SFM, and both RF and SVM classifiers demonstrated excellent performance as Table 2 shows. For the light background, perfect performance was achieved in the three metrics, indicating a correct classification of all pixels in the validation subset. These results were to be expected, as the four classes are well differentiated and the number of pixels is limited.

Table 2. Metrics for all models with two bands of interest. The best result for each sample is marked in bold.

		SFS			**SFM**		
		OAC	**F1**	κ	**OAC**	**F1**	κ
Distilled water Dark background	**RF**	0.992	0.992	0.989	0.800	0.797	0.716
	LR	0.950	0.950	0.930	0.759	0.721	0.652
	SVM	**0.999**	**0.999**	**0.999**	0.842	0.799	0.773
Distilled water Light background	**RF**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	LR	0.992	0.992	0.989	0.896	0.863	0.850
	SVM	0.998	0.998	0.997	0.946	0.941	0.923
Seawater Dark background	**RF**	**0.997**	**0.997**	**0.995**	0.845	0.844	0.785
	LR	0.995	0.995	0.993	0.853	0.854	0.798
	SVM	**0.997**	**0.997**	**0.995**	0.837	0.838	0.775
Seawater Light background	**RF**	0.997	0.997	0.996	0.997	0.997	0.996
	LR	**1.000**	**1.000**	**1.000**	0.822	0.764	0.748
	SVM	**1.000**	**1.000**	**1.000**	0.896	0.880	0.856

We identified the optimal bands for each model (Figure 10). Two spectral areas of interest are highlighted, ranging from 400 nm to 500 nm and from 550 nm to 650 nm. These findings align with the regions of interest identified in the spectral analysis of different rhodamine concentrations. Samples with light backgrounds show more bands of interest in the blue region of the spectrum, whereas those taken on dark backgrounds tend to have significant bands in the yellow-orange region.

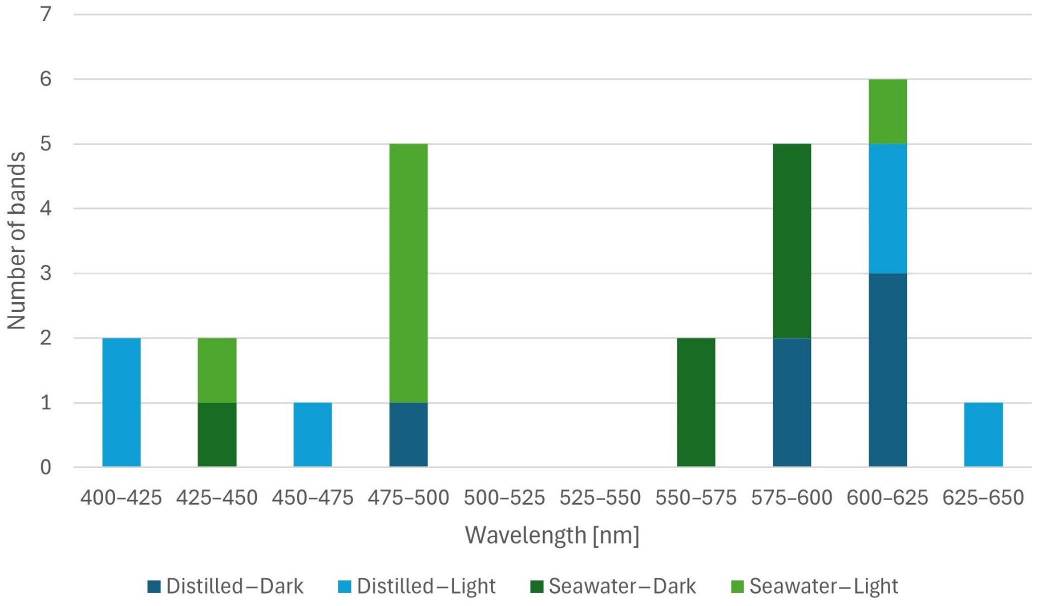


Figure 10. Spectral areas of interest, identified by grouping the two most significant bands for each combination of SFS with RF, LR, and SVM.

The background affects the spectral signature of the rhodamine and the band selection. However, transferring the trained artificial intelligence model from one scenario to another with a different background may be possible.

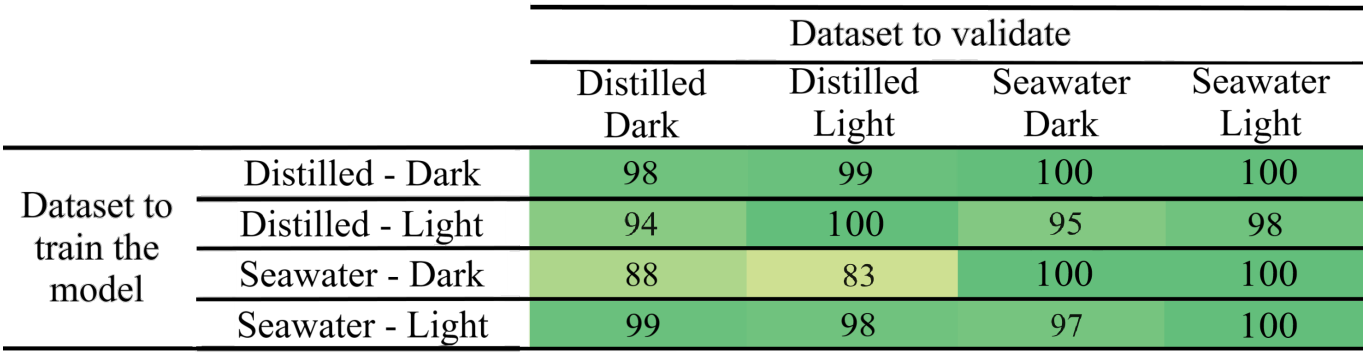
3.4. Transferability

We differentiated a section for the transfer of the bands of interest (Section 3.4.1) and another for the transfer of the pre-trained models (Section 3.4.2). In addition, a third section (Section 3.4.3) was included, where we improved the model transfer by combining all the samples and calculating the first derivative to train the classifier.

3.4.1. Band Transfer

The two most influential bands obtained with each model and sample were transferred to the rest of the samples to analyze the effectiveness of transferring spectral bands between different samples in the dataset. Section 3.3 shows that combining SFS with different classifiers led to the best results. Therefore, for this experiment, we exclusively analyzed the results of the SFS-RF combination. The other classifier combinations yielded similar results without a significant impact on the transferability of the bands and required more computational time.

The band transfer was successful for all samples as shown in Figure 11. The best results were obtained when validating the models with seawater samples, with superior performance observed when transferring from models trained with dark background samples. For distilled water samples, better performance was achieved when transferring bands from models trained also with distilled water samples.



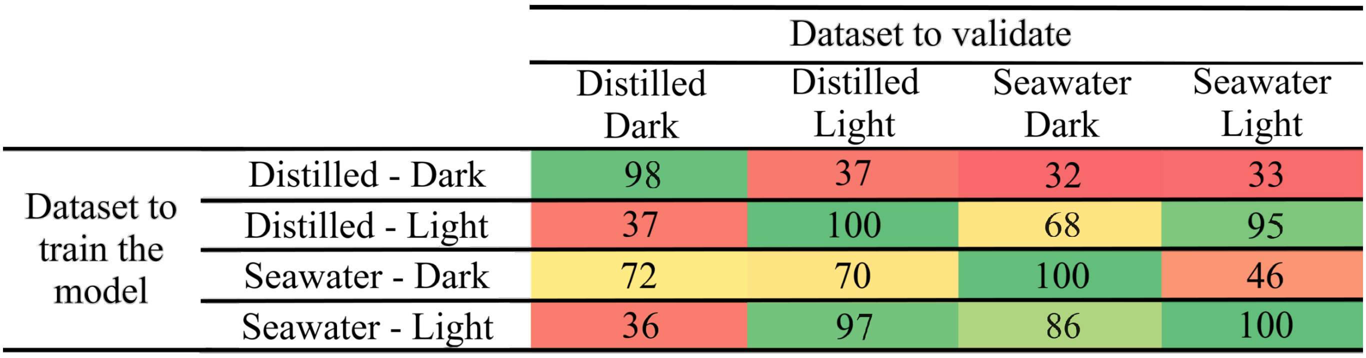
		Dataset to validate			
		Distilled Dark	Distilled Light	Seawater Dark	Seawater Light
Dataset to train the model	Distilled - Dark	98	99	100	100
	Distilled - Light	94	100	95	98
	Seawater - Dark	88	83	100	100
	Seawater - Light	99	98	97	100

Figure 11. Accuracy obtained by transferring bands of interest from one sample to another. The colours indicate performance: green tones for accuracies above 80%, yellowish for 60–80%, orange for 40–60%, and red for accuracy below 40%.

3.4.2. Train Model Transfer

Transferring pre-trained classification models is complex due to the limited variability in the training samples of the dataset, which can easily lead to overfitting and lack of generalization. The accuracy results obtained differ significantly from those seen in band transfer. As shown in Figure 12, the model trained on distilled water with a dark background is not transferable to other samples, achieving an accuracy between 32% and 37%. However, the model trained on distilled water with a light background transfers perfectly to seawater with a light background, with an accuracy of 95%. This model also achieved 68% accuracy when transferred to seawater with a dark background. The model trained on seawater with a light background shows 97% accuracy when validated on distilled water with a light background and 86% accuracy when validated on seawater with a dark background. Meanwhile, the model trained on seawater with a dark background achieved 70% accuracy on both distilled water samples but fell below 50% on seawater with a light background.

The average accuracy, considering all transferred models, was 70%. Models show higher accuracy when transferring within the same background type, achieving 76% average accuracy for dark background samples and 98% accuracy for light ones. In contrast, models trained on dark backgrounds achieved only 47% accuracy when applied to light background samples, and those trained on light backgrounds achieved 57% accuracy when transferred to dark background samples. This indicates that model transfer is performed effectively with the same background but different water types. Additionally, better results are obtained with light backgrounds compared to dark backgrounds, whereas performance decreases with different backgrounds.



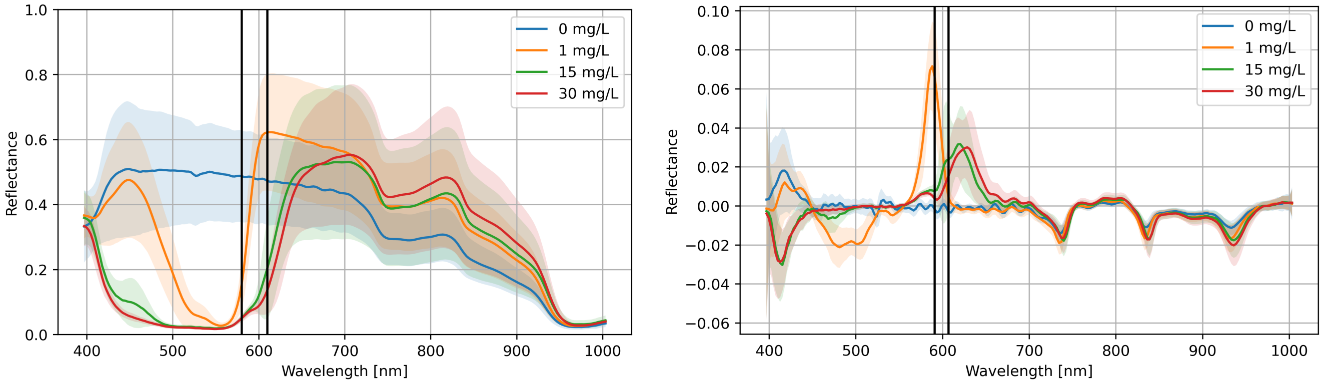
		Dataset to validate			
		Distilled Dark	Distilled Light	Seawater Dark	Seawater Light
Dataset to train the model	Distilled - Dark	98	37	32	33
	Distilled - Light	37	100	68	95
	Seawater - Dark	72	70	100	46
	Seawater - Light	36	97	86	100

Figure 12. Accuracy obtained by transferring trained classifiers from one sample to another. The colours indicate performance: green tones for accuracies above 80%, yellowish for 60–80%, orange for 40–60%, and red for accuracy below 40%.

3.4.3. Improving Model Transfer Combining the Samples and Applying the First Derivative

Two approaches improve the model transfer. The first involves creating a new dataset that combines all types of samples with different water types and backgrounds, thereby avoiding overfitting the classifier. Secondly, the first derivative highlights variations in spectral signatures, enhancing the identification of concentrations. These new datasets will be referred to as combined sample (CS) and combined derivative (CD), respectively.

Analyzing the average spectral signature of CS (Figure 13), we can already anticipate that distinguishing between the 15 mg/L and 30 mg/L samples will be challenging because their spectral signatures are very similar. This similarity occurs because darker backgrounds attenuate light reflection, causing the reflectance of the 30 mg/L sample on a dark background to resemble the reflectance of the 15 mg/L sample on a light background. In the average spectrum of the first derivative, 15 mg/L and 30 mg/L are the closest concentrations, whereas the signature of the 1 mg/L concentration is well differentiated. The bands of interest are 580 nm and 610 nm for the combined samples and 591 nm and 607 nm for their first derivative.

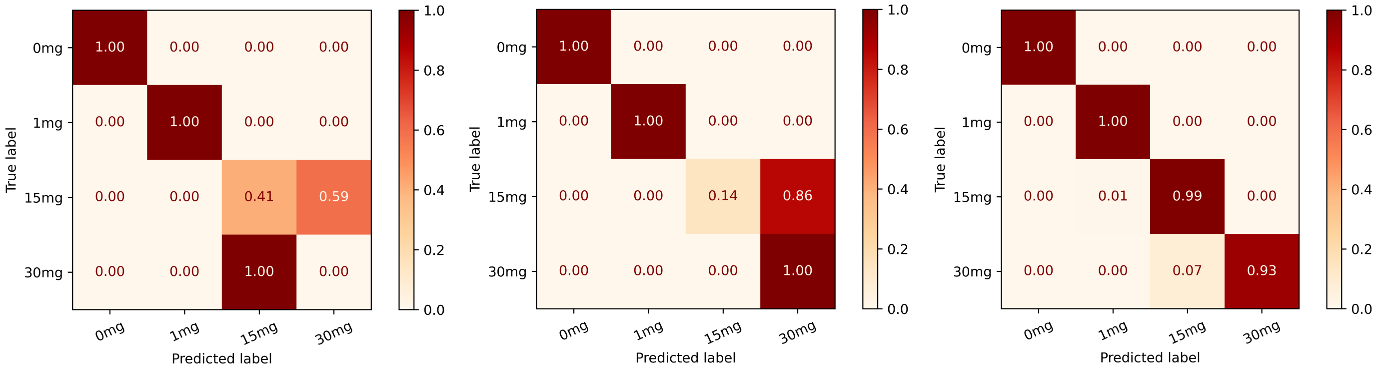


(**a**) (**b**)

Figure 13. Mean spectra and standard deviation (shaded in the corresponding colour). The two best bands are indicated with black vertical lines. (**a**) Combined samples (580 and 610 nm); (**b**) the first derivative of the combined samples (591 and 607 nm).

The transfer of models trained with CS shows higher accuracy than the results presented in earlier sections. When applying the first derivative, the results are excellent, surpassing 85% accuracy in all cases except for the distilled samples with a light background, which only reaches 50%.

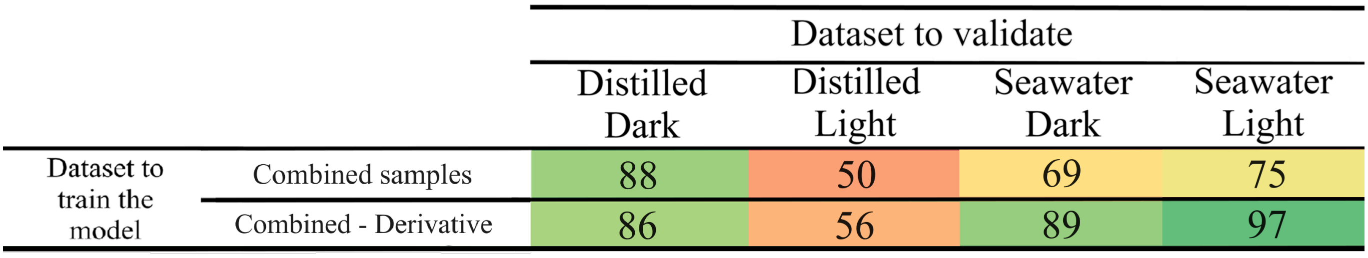
The confusion matrices for the best and worst cases provide additional insight into how the CS and the derivative influence classification (Figure 14). When training with CS and validating with light background samples, there is significant misclassification between the 15 mg/L and 30 mg/L, especially for distilled water. The separation of these two concentrations improves notably when the derivative (CD) is applied.



(**a**) (**b**) (**c**)

Figure 14. Confusion matrix training with CS and CD for the best and worst scenarios in Figure 15. (**a**) CS validating on distilled light; (**b**) CS validating on seawater light; (**c**) CD validating on seawater light.

This experiment again demonstrates that the background affects the reflectance of translucent solution spectral signatures. However, training a model with varied samples enhances classification performance, and using the first derivative aids in better identifying rhodamine concentrations.



		Dataset to validate			
		Distilled Dark	Distilled Light	Seawater Dark	Seawater Light
Dataset to train the model	Combined samples	88	50	69	75
	Combined - Derivative	86	56	89	97

Figure 15. Accuracy obtained by transferring trained classifiers from the combined sample. The colours indicate performance: green tones for accuracies above 80%, yellowish for 60–80%, orange for 40–60%, and red for accuracy below 40%.

4. Discussion

Significant differences were observed in the spectral response of samples with light and dark backgrounds in Section 3.2, with discrepancies of nearly 0.5 reflectance units. Small differences were observed based on the water type, but they are not sufficiently representative (between 0.1 and 0.3) to determine their influence. A curve around 810 nm appeared in all samples, consistent with studies [12,22], likely due to a local minimum in liquid water absorption common in shallow waters. The spectral regions most affected by rhodamine concentration were between 550 nm and 650 nm and from 400 nm to 500 nm, independently of the sample. These results aligned with the findings of Clark et al. [11], who observed that upwelling spectral radiance from a water body containing dye decreases in the green portion of the visible spectrum (530–570 nm) due to dye absorption and increases in the red and near-infrared wavelengths (570–750 nm) due to dye reflectance.

It should be noted that the influence of the 400 to 500 nm range is affected by the scene's illumination and the scattering properties of the water and the background. In our case study, this range is particularly impacted by the high reflectance caused by the reflections from the beaker.

The use of classifiers demonstrated high accuracy in differentiating rhodamine concentrations and determining that the optimal number of bands for the classification is two (Section 3.3). Classifiers combined with the SFS feature selector yielded more satisfactory results than SFM, with all classification metrics exceeding 90%. Several samples even achieved 100% accuracy, indicating that all pixels in the validation subset were correctly classified. These results were expected, as the dataset consists of controlled samples measured in a laboratory setting, with a limited number of pixels. This dataset will be extended to include more variability in future work, leading to more realistic classification metrics.

The study identified that for all the samples, the two most influential spectral bands were consistently within the 400–500 nm and 550–650 nm regions, corresponding to the regions identified in the spectral analysis of Section 3.2. We combined the bands identified by the best models (SFS + RF, LR, and SVM) to obtain enough data to determine if any patterns emerged in the bands of interest. A notable difference was observed between samples with light and dark backgrounds regardless of water type. Specifically, the 400–500 nm wavelengths are more relevant on light backgrounds, while the 550–650 nm range is more influential on dark backgrounds. This is a direct consequence of the background's influence and the rhodamine's semi-transparency, which is more critical on dark backgrounds due to lower reflection.

Transferring the results is a crucial phase of the study, as one of the primary objectives is to estimate the rhodamine concentration without specific scene calibration (Section 3.4). The successful transfer of bands of interest determined from one sample to another demonstrated that band transfer is feasible. The transfer of pre-trained models has been highly effective among light background samples, achieving an accuracy exceeding 95%. Nevertheless, the transfer of models between samples with different background types was not as effective, which was expected, given the difference in spectral areas of interest between light and dark backgrounds analyzed in Section 3.3. Also, the spectral range between 600 nm and 650 nm was both an area of interest for the classifiers and a region of significant discrepancy between the spectral signatures of distilled and seawater as shown in Section 3.2. This discrepancy further complicated the transfer between water types.

To overcome the impediments in model transfer caused by background reflectance differences and spectral signature variations due to water type, two essential measures were implemented: training the classifier with samples from different backgrounds and water types, and calculating the first derivative. When applying this approach, accuracy surpassed 85% in all cases except for the distilled light sample (50%). The optimal bands were between 580 and 610 nm. The primary difficulty improved by the first derivative was the differentiation between the 15 mg/L and 30 mg/L concentrations, as the reflectance of the 30 mg/L samples on dark backgrounds decreased, making it resemble the 15 mg/L samples on light backgrounds. This approach in Section 3.4.3 prevented overfitting in the classifier and highlighted variations in the spectral trends, improving the generalizability of the models.

This new proposed methodology can achieve a generalizable model that overcomes the limitations of current models. Clark et al. [11] show that band ratio linearity with rhodamine concentration becomes nonlinear above 0.03 mg/L, a problem that our method overcomes. Additionally, the Optimum Band Ratio Analysis (OBRA) [20], the most widely used method for determining rhodamine concentration, still requires on-site calibration for each specific scenario. Furthermore, the identified spectral areas of interest (400–500 nm and 550–650 nm) and the transfer of models had direct applications in the design and optimization of multispectral cameras [21]. These findings can be integrated into existing camera systems to enhance the detection and quantification of dye concentrations, which is crucial for applications such as tracking pollutant dispersion in water bodies.

Another approach to consider is radiative transfer models (RTMs), such as Hydrolight [43], which simulate light propagation through water and provide insights into environmental factors like depth, bottom types, and constituent concentrations. These models can be useful for predicting the behavior of substances like rhodamine in various aquatic environments, but they rely on precise input parameters that may not fully capture real-world complexity [44]. While complementary to laboratory measurements, which offer controlled and empirical data, RTMs may struggle with the complexity of optically shallow environments, where factors like bottom reflectance and water column effects introduce significant variability [2]. Given these challenges, we chose to rely on laboratory measurements combined with artificial intelligence models in our study. This approach allowed us to more effectively manage the inherent variability and complexity of optically shallow environments, providing a more robust and generalizable method for detecting and analyzing rhodamine in such settings.

Our study presents an exhaustive spectral analysis of different water types and backgrounds to improve the identification of various rhodamine concentrations. The key findings reveal significant spectral discrepancies based on the background type, highlighting the importance of considering background reflectivity in optically shallow waters. The transfer of bands of interest was successful, ensuring that multispectral cameras with a few bands can effectively determine rhodamine concentration. Additionally, training the classification model with combined samples and applying the first derivative enabled the successful transfer of pre-trained classification models. This advancement aimed to develop a classifier that works in several scenarios without the need for on-site calibration, bringing us one step closer to improving the remote detection of dye in aquatic environments.

Future research should focus on further refining the classification models by incorporating more diverse sample types and environmental conditions. Additionally, there is an interest in testing the implementation of a regressor to estimate rhodamine concentration and conducting experiments with lower concentrations in real-world scenarios. Exploring advanced machine learning techniques and integrating them with hyperspectral imaging can provide deeper insights and better transference. Expanding the spectral analysis to include other fluorescent dyes and pollutants can broaden the applicability of this research. Moreover, the development of real-time monitoring systems using these enhanced models could significantly benefit environmental monitoring, pollution control, and water quality assessment.

5. Conclusions

This study provided new insights into the generalization of spectral semi-transparent solutions detection across different water types and backgrounds. The spectral analysis identified relevant regions for rhodamine classification between 400 nm and 500 nm, and 550 nm and 600 nm, related to solution concentration. These wavelengths are commonly found in sensors and satellites such as Landsat 8 [45], Sentinel-2 [46], and WorldView-2 [47]. The significant differences observed between samples on dark and light backgrounds, especially with maximum reflectance captured, underscore the necessity of the spectral characterization of backgrounds when using rhodamine in optically shallow waters.

The classification results are promising, indicating the feasibility of transferring classification results with high accuracy, especially when integrated with the SFS feature selector. Our study identified the two most influential spectral bands consistently within the 400–500 nm and 550–650 nm regions for all samples, correlating with regions identified in the spectral analysis. While the transfer of spectral bands was successful (>80%), the transfer of trained models was only successful among light background samples. To improve these results, a model was trained by combining all samples to avoid overfitting, and applying the first derivative to enhance the identification of variations in the spectral signature. As a result, accuracy surpassed 85% in all cases except for the distilled light sample, where the concentrations of 15mg/L and 30mg/L were misclassified.

Therefore, the transfer of pre-trained classifiers between different backgrounds is feasible. This would potentially eliminate the need for on-site calibration each time rhodamine dispersion is studied in aquatic environments, saving materials and providing a generalizable classifier. Future research will focus on extending the dataset to include more variability and backgrounds, exploring the practical applicability of the models in outdoor scenarios. This expansion will ensure that the model can accurately generalize across diverse aquatic environments. By doing so, we can advance towards establishing standardized protocols for remote sensing and monitoring solutions dispersion, facilitating broader applications in environmental monitoring and water dynamics.

Author Contributions: Conceptualization, Á.P.-G., T.H.M.v.E. and J.F.L.; methodology, Á.P.-G. and A.M.L.; software, Á.P.-G. and A.M.L.; validation, A.R.-M. and E.H.; formal analysis, Á.P.-G. and A.M.L.; investigation, Á.P.-G. and A.M.L.; resources, E.H. and J.F.L.; data curation, A.R.-M. and A.M.L.; writing—original draft preparation, Á.P.-G. and A.M.L.; writing—review and editing, A.R.-M. and E.H.; visualization, Á.P.-G. and A.M.L.; supervision, T.H.M.v.E. and J.F.L.; project administration, J.F.L.; funding acquisition, J.F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was completed while Ámbar Pérez-García was beneficiary of a predoctoral grant (ID: TESIS2021010059) given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Informacion (ACIISI)" of the "Consejería de Universidades, Ciencia e Innovación y Cultura", part-financed by the European Social Fund Plus (FSE+) "Programa Operativo Integrado de Canarias 2021–2027, Eje 3 Tema Prioritario 74 (85%)". Adrián Rodríguez-Molina and Emma Hernández have benefited from a predoctoral grant given by the "Vicerrectorado de Investigación y Transferencia de la Universidad de las Palmas de Gran Canaria" during the research. We would also like to thank the OASIS-HARMONIE project, under contract PID2023-148285OB-C43 from "Proyectos de Generación de Conocimiento" 2023, and the PERSEO project, under contract CPP2021-008527, from "Programa de Colaboración Público-Privada" of "Gobierno de España", 2023–2025.

Data Availability Statement: The dataset used in this study is publicly available on IEEE Dataport. It can be accessed using the following DOI: 10.21227/n788-te91.

Acknowledgments: We would like to thank Elittoral for the preparation and provision of the rhodamine samples in different concentrations used in this study. We would also like to thank Guillermo Valentín Socorro Marrero for his valuable advice.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Powers, C.; Hanlon, R.; Schmale, D.G., III. Tracking of a fluorescent dye in a freshwater lake with an unmanned surface vehicle and an unmanned aircraft system. *Remote Sens.* **2018**, *10*, 81. [CrossRef]
2. Ji, C.; Beegle-Krause, C.J.; Englehardt, J.D. Formation, detection, and modeling of submerged oil: A review. *J. Mar. Sci. Eng.* **2020**, *8*, 642. [CrossRef]
3. Okubo, A. Oceanic diffusion diagrams. In *Deep Sea Research and Oceanographic Abstracts*; Elsevier: Amsterdam, The Netherlands, 1971; Volume 18, pp. 789–802.
4. Pritchard, D.W.; Carpenter, J. Measurements of turbulent diffusion in estuarine and inshore waters. *Hydrol. Sci. J.* **1960**, *5*, 37–50. [CrossRef]
5. Fong, D.A.; Stacey, M.T. Horizontal dispersion of a near-bed coastal plume. *J. Fluid Mech.* **2003**, *489*, 239–267. [CrossRef]
6. González-Pinzón, R.; Haggerty, R.; Dentz, M. Scaling and predicting solute transport processes in streams. *Water Resour. Res.* **2013**, *49*, 4071–4088. [CrossRef]
7. Csanady, G. Turbulent diffusion in lake huron. *J. Fluid Mech.* **1963**, *17*, 360–384. [CrossRef]
8. Atkinson, T.; Davis, P. Longitudinal dispersion in natural channels: I. Experimental results from the River Severn, UK. *Hydrol. Earth Syst. Sci.* **2000**, *4*, 345–353. [CrossRef]
9. Drummond, J.; Davies-Colley, R.; Stott, R.; Sukias, J.; Nagels, J.; Sharp, A.; Packman, A. Retention and remobilization dynamics of fine particles and microorganisms in pastoral streams. *Water Res.* **2014**, *66*, 459–472. [CrossRef]
10. Legleiter, C.J.; McDonald, R.R.; Nelson, J.M.; Kinzel, P.J.; Perroy, R.L.; Baek, D.; Seo, I.W. Remote sensing of tracer dye concentrations to support dispersion studies in river channels. *J. Ecohydraulics* **2019**, *4*, 131–146. [CrossRef]

11. Clark, D.B.; Lenain, L.; Feddersen, F.; Boss, E.; Guza, R. Aerial imaging of fluorescent dye in the near shore. *J. Atmos. Ocean. Technol.* **2014**, *31*, 1410–1421. [CrossRef]
12. Köppl, C.J.; McKnight, U.S.; Lemaire, G.G.; Nørregaard, A.M.; Thiim, T.C.; Bjerg, P.L.; Bauer-Gottwein, P.; García, M. Tracer concentration mapping in a stream with hyperspectral images from unoccupied aerial systems. *Adv. Water Resour.* **2023**, *182*, 104567. [CrossRef]
13. Valerio, C.; Llebaria, A. A quantitative multispectral analysis system for aerial photographs applied to coastal planning. *Int. J. Remote Sens.* **1982**, *3*, 181–197. [CrossRef]
14. Spitzer, D.; Wernand, M. Multispectral remote-sensing of fluorescent tracers-theory and experiments. *Oceanol. Acta* **1983**, *6*, 201–210.
15. Hally-Rosendahl, K.; Feddersen, F.; Clark, D.B.; Guza, R. Surfzone to inner-shelf exchange estimated from dye tracer balances. *J. Geophys. Res. Oceans* **2015**, *120*, 6289–6308. [CrossRef]
16. Filippi, M.; Hanlon, R.; Rypina, I.I.; Hodges, B.A.; Peacock, T.; Schmale, D.G., III. Tracking a surrogate hazardous agent (rhodamine dye) in a coastal ocean environment using in situ measurements and concentration estimates derived from drone images. *Remote Sens.* **2021**, *13*, 4415. [CrossRef]
17. Nekouee, N.; Roberts, P.J.; Schwab, D.J.; McCormick, M.J. Classification of buoyant river plumes from large aspect ratio channels. *J. Hydraul. Eng.* **2013**, *139*, 296–309. [CrossRef]
18. Baek, D.; Seo, I.W.; Kim, J.S.; Nelson, J.M. UAV-based measurements of spatio-temporal concentration distributions of fluorescent tracers in open channel flows. *Adv. Water Resour.* **2019**, *127*, 76–88. [CrossRef]
19. Johansen, K.; Dunne, A.F.; Tu, Y.H.; Almashharawi, S.; Jones, B.H.; McCabe, M.F. Dye tracing and concentration mapping in coastal waters using unmanned aerial vehicles. *Sci. Rep.* **2022**, *12*, 1141. [CrossRef]
20. Legleiter, C.J.; Manley, P.V.; Erwin, S.O.; Bulliner, E.A. An experimental evaluation of the feasibility of inferring concentrations of a visible tracer dye from remotely sensed data in turbid rivers. *Remote Sens.* **2019**, *12*, 57. [CrossRef]
21. Pérez-García, A.; Rodríguez-Molina, A.; Hernández, E.; Vera, L.; López, J.F. Development of Low-Cost Multi-Spectral Cameras for Precision Agriculture. In Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 3466–3469. [CrossRef]
22. Legleiter, C.J.; Sansom, B.J.; Jacobson, R.B. Remote sensing of visible dye concentrations during a tracer experiment on a large, turbid river. *Water Resour. Res.* **2022**, *58*, e2021WR031396. [CrossRef]
23. Developers, N. NumPy Gradient Function. 2024. Available online: https://numpy.org/doc/stable/reference/generated/numpy.gradient.html (accessed on 26 June 2024).
24. Elittoral. 2024. Available online: https://www.elittoral.es/ (accessed on 16 June 2024).
25. Scientific, T. ThermoFisher Scientific Official Website. 2024. Available online: https://www.patheon.com/us/en/home.html (accessed on 28 June 2024).
26. Labbox. Lab Supplies. Available online: https://labbox.eu/ (accessed on 7 August 2024).
27. Morales, A.; Horstrand, P.; Guerra, R.; Leon, R.; Ortega, S.; Díaz, M.; Melián, J.M.; López, S.; López, J.F.; Callico, G.M.; et al. Laboratory hyperspectral image acquisition system setup and validation. *Sensors* **2022**, *22*, 2159. [CrossRef] [PubMed]
28. Specim Spectral Imaging. FX Series Hyperspectral Cameras. Available online: http://www.specim.fi/fx/ (accessed on 6 October 2019).
29. SphereOptics. Zenith Polymer Diffusers. Available online: https://sphereoptics.de/en/product/zenith-polymer-diffusers/?c=79 (accessed on 6 October 2019).
30. Pérez-García, A.; Lorenzo, A.M.; López, J. Spectral Band Selection Methodology for Future Sensor Development. In Proceedings of the 2024 47th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 20–24 May 2024; pp. 152–156. [CrossRef]
31. Kramer, O.; Kramer, O. Scikit-learn. In *Machine Learning for Evolution Strategies*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 45–53.
32. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282. [CrossRef]
33. Wright, R.E. Logistic regression. *Circulation* **1995**, *117*, 18.
34. Hearst, M.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
35. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
36. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.
37. Suthaharan, S.; Suthaharan, S. Support vector machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 207–235.
38. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1.
39. Boughorbel, S.; Jarray, F.; El-Anbari, M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS ONE* **2017**, *12*, e0177678. [CrossRef]
40. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
41. Sammut, C.; Webb, G.I. *Encyclopedia of Machine Learning*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
42. Syakur, M.; Khotimah, B.; Rochman, E.; Satoto, B.D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2018; Volume 336, p. 012017.
43. Mobley, C.D.; Sundman, L.K. *Hydrolight 5 Ecolight 5*; Sequoia Scientific Inc.: Bellevue, WA, USA, 2008; Volume 16.

44. Salesin, K.; Knobelspiesse, K.D.; Chowdhary, J.; Zhai, P.W.; Jarosz, W. Unifying radiative transfer models in computer graphics and remote sensing, Part I: A survey. *J. Quant. Spectrosc. Radiat. Transf.* **2024**, *314*, 108847. [CrossRef]
45. Morfitt, R.; Barsi, J.; Levy, R.; Markham, B.; Micijevic, E.; Ong, L.; Scaramuzza, P.; Vanderwerff, K. Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit. *Remote Sens.* **2015**, *7*, 2208–2237. [CrossRef]
46. Pahlevan, N.; Sarkar, S.; Franz, B.; Balasubramanian, S.; He, J. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sens. Environ.* **2017**, *201*, 47–56. [CrossRef]
47. Corporation, S.I. WorldView-2 Satellite Sensor. 2024. Available online: https://earth.esa.int/eogateway/missions/worldview-2 (accessed on 7 August 2024).