Computers and Electronics in Agriculture 197 (2022) 106956

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Machine learning in management of precautionary closures caused by lipophilic biotoxins

Andres Molares-Ulloa [a,*], Enrique Fernandez-Blanco [b], Alejandro Pazos [b,c], Daniel Rivero [b]

[a] Universidade da Coruña, Department of Computer Science and Information Technology, Faculty of Computer Science, 15071 A Coruña, Spain
[b] Centro de investigación CITIC, Department of Computer Science and Information Technology, University of A Coruña, 15071 A Coruña, Spain
[c] Biomedical Research Institute of A Coruña (INIBIC), University Hospital Complex of A Coruna (CHUAC), A Coruna 15006, Spain

ARTICLE INFO

Keywords:
Machine Learning
Harmful Algal Blooms
Mussels
Aquaculture
Diarrhoeic Shellfish Poisoning

ABSTRACT

Mussel farming is one of the most important aquaculture industries. The main risk to mussel farming is harmful algal blooms (HABs), which pose a risk to human consumption. In Galicia, the Spanish main producer of cultivated mussels, the opening and closing of the production areas is controlled by a monitoring program. In addition to the closures resulting from the presence of toxicity exceeding the legal threshold, in the absence of a confirmatory sampling and the existence of risk factors, precautionary closures may be applied. These decisions are made by experts without the support or formalisation of the experience on which they are based. Therefore, this work proposes a predictive model capable of supporting the application of precautionary closures. Achieving sensitivity, accuracy and kappa index values of 97.34%, 91.83% and 0.75 respectively, the kNN algorithm has provided the best results. This allows the creation of a system capable of helping in complex situations where forecast errors are more common.

1. Introduction

Global mussel production has steadily increased to 2.2 million tonnes in 2018, more than double the amount produced ten years ago (FAO, 2 February 2022). Nearly 94% of global mussel production comes from aquaculture (Avdelas et al., 2021). Young mussels are harvested from the sea and may be grown on suspended ropes; these ropes, which are covered with mussel seed held in place with nylon nets, are suspended either from rafts, or wooden frames, or from longlines with floating plastic buoys. A substantial portion of EU production is farmed on suspended ropes, a technique that can be extended further offshore and which, although very sensitive to plankton blooms, is the only one that could allow further increases in production.

One of the main risks of mussel farming is Harmful Algal Blooms (HABs). HABs are episodes of high concentrations of algae, including some cyanobacteria and microalgae that are potentially toxic for human consumption. This is because there is a risk of poisoning by consuming filter-feeding bivalve molluscs such as mussels that feed on these algae, accumulating the toxins in their meat. To monitor these episodes, there are programs set up in mussel production areas. For the early detection of high toxicity events, these monitoring programmes have fixed sampling points strategically located in the production areas. These high toxicity events can lead to a temporary suspension of mussel harvesting and marketing. The most common toxin-producing species are those of Diarrhoeic Shellfish Poisoning (DSP) type. The most abundant of which is the dinoflagellate *Dinophysis acuminata*) (Vilas et al., 2008).

The opening and closing of the production areas is based on the analysis of the toxicity of the mollusc meat, as established by European legislation (UE6, 2019). Within the monitoring programme, sampling planning uses expert knowledge based on information on endogenous and exogenous factors influencing the proliferation of potentially toxic phytoplankton species. The most compromising point of this process is the absence of sampling during non-working days or when inclement weather does not allow it to be carried out. This leads to situations where it is impossible to collect the data to support an effective closure. If there are indications of a potential increase in toxicity levels, the competent authority is legally entitled to proceed with 'precautionary closures' of bivalve mollusc production areas.

Precautionary closures may become effective after a subsequent analysis verifying the presence of toxins, otherwise, the closure will be lifted. The application or non-application of these measures creates two possible problem scenarios. In the first scenario the precautionary closure is applied even though toxicity values above the legal threshold are not reached. This scenario could lead to economic losses for

* Corresponding author.
E-mail address: andres.molares@udc.es (A. Molares-Ulloa).

https://doi.org/10.1016/j.compag.2022.106956
Received 20 August 2021; Received in revised form 9 March 2022; Accepted 8 April 2022
Available online 15 April 2022
0168-1699/

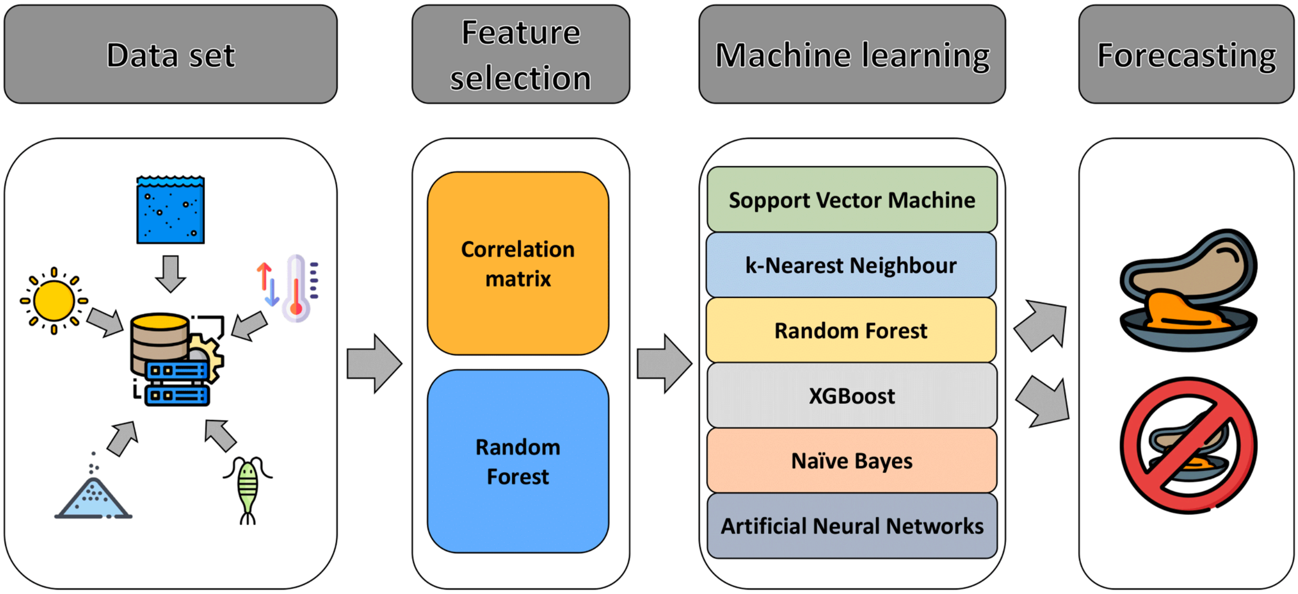

Fig. 1. Schematic representation of the machine learning-based system for predicting harmful algal bloom closures and aiding decision making in mussel farming. This graphic has been designed with resources from Flaticon.com.

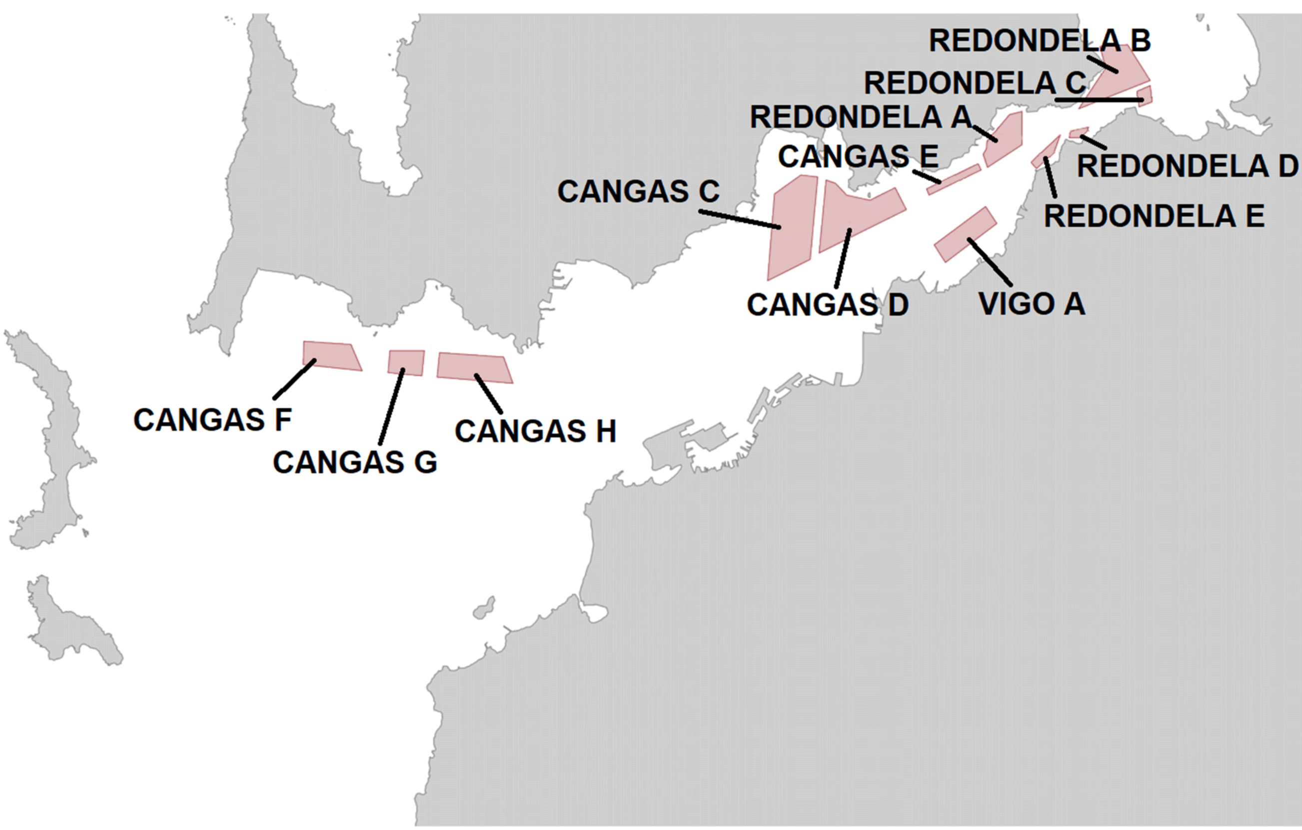

Fig. 2. Map of the production areas of cultivated molluscs in the Vigo estuary. Source: http://193.144.46.136/EstadoZonas/Default.aspx?tmapa = 0.

producers because they are prohibited from working while the area remains closed. In the second scenario no indications of a high toxicity event are detected, but a subsequent analysis shows the presence of toxins. The latter is a much more dangerous situation than the previous one because, during this period of extraction activity, there is a potential risk of introducing contaminated shellfish into the market, with the consequent risk to public health. Today, the implementation of precautionary closures is based on the experience of monitoring experts. The existence of a predictive model could help them make the right decisions in complex situations.

Harmful algal blooms are not only a potential risk to public health, they are also a major problem for the production sector. Work such as that of Di Jin and Porter Hoagland (Jin and Hoagland, 2008) has shown that the development of predictive systems can lead to significant improvements in management strategy and profits for the farming sector. So far, numerous studies have attempted such predictions around the

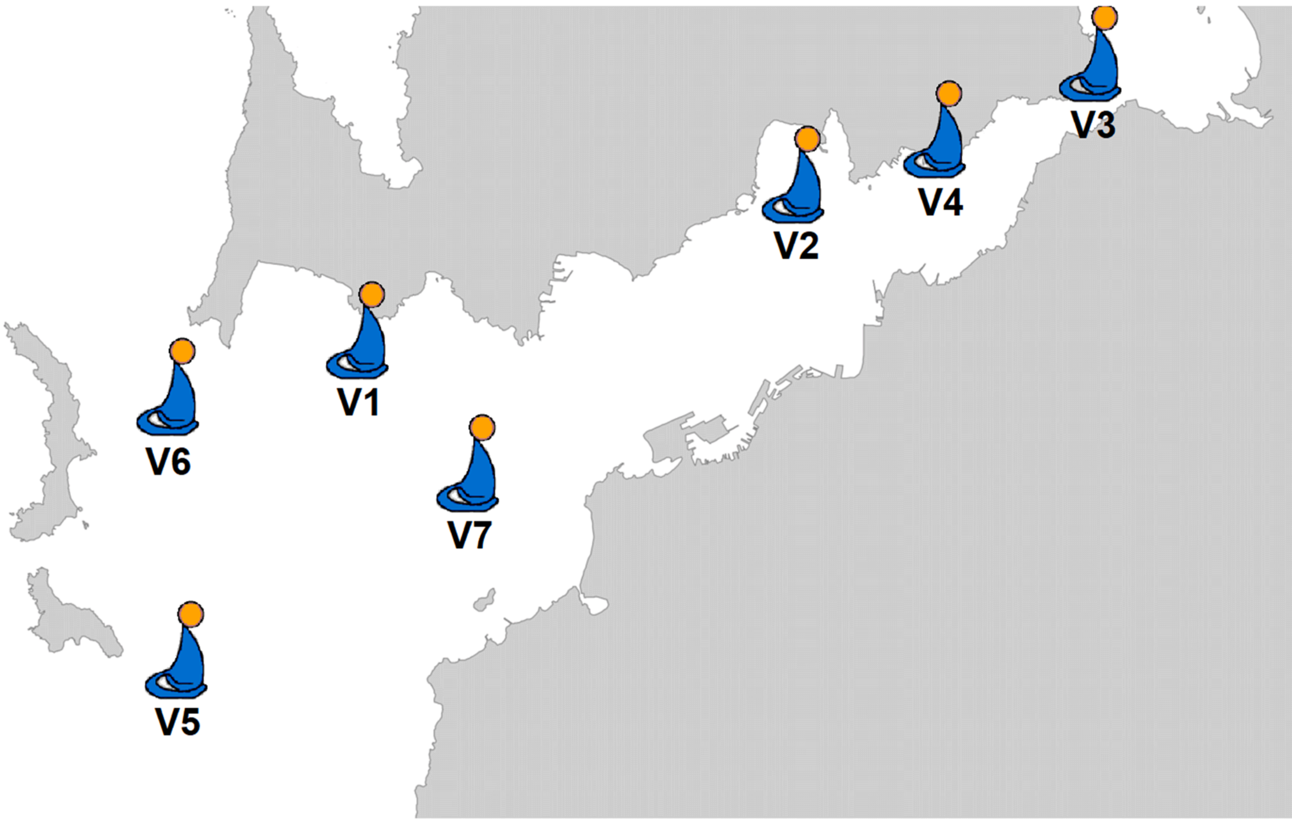

Fig. 3. Map of oceanographic stations located in the Vigo estuary. Source: http://www.intecmar.gal/Ctd/Default.aspx.

Table 1
Table of variables (2004–2018).

Source	Variable	Number of locations	Features generated	Frequency
INTECMAR	Temperature	7	14	Weekly
	Salinity	7	7	Weekly
	Oxygen	7	7	Weekly
	Chlorophyll-a concentration	7	7	Weekly
	Dinophysis acuminata cells abundance	7	7	Weekly
	Dissolved ammonium	7	7	Weekly
	Dissolved phosphate	7	7	Weekly
	Dissolved nitrate	7	7	Weekly
	Dissolved nitrite	7	7	Weekly
	State of production areas	1	1	Daily
METEOGALICIA	Solar irradiation	1	1	Daily
	Sunshine hours	1	1	Daily
	Insolation	1	1	Daily
IEO	Upwelling index	1	1	Daily
-	Seasonality	1	1	Daily

world, notably off the coasts of South Korea (Lee and Lee, 2018), Hong Kong (Yu et al., 2021) and the Persian Gulf (Gholami et al., 2019), in general, these works have focused their efforts on predicting biomarkers such as the concentration of toxic phytoplankton in the water or chlorophyll-a (Deng et al., 2021; Liu et al., 2009).There are studies for the specific case of the Spanish coast (Velo-Suárez and Gutiérrez-Estrada, 2007) and specifically for the Galician coast (Vilas et al., 2014; Aguilar Calderon, 2017; Molares et al., 2020). For the creation of this type of predictive models, the use of different classical techniques has been compared with ML techniques to try to find the co-figuration that best suits this problem (Cruz et al., 2021; Liu et al., 2009). It was determined that ML techniques outperform classical methods.The success of applying machine learning techniques to harmful algal blooms lies in the selection of the relevant data and the pre-processing of the data. The proliferation of harmful algal blooms is influenced by many factors, the most important of which are: temperature, water flow, upwelling, light, nutrients and salinity.

A higher water temperature favours algae proliferation, as well as thermocline stratification favours their concentration (Davis et al., 2009). Excessive water flow and circulation disperses algae concentrations, reducing the occurrence of blooms (Li et al., 2013). The light is necessary for phytoplankton to photosynthesise (Paerl and Paul, 2012). Dissolved nutrients in the water create a favourable environment for algal growth (Paerl and Paul, 2012). The salinity plays an important role in the formation of phytoplankton communities (Gasinaite et al., 2005).

The best results were obtained using the combined CNN and LSTM spatio-temporal classification technique to classify and discriminate between HAB and non-HAB events produced in Florida coastal waters by the algae *Karenia brevis* (Hill et al., 2020). But it is difficult to have such a large volume of data on a regular basis, and even impossible for many

Table 2
Descriptive analysis of input features.

	Maximum chlorophyll 'a' in V1	*D. acuminata* concentration in V1	Ammonium in V1	Phosphate in V1	Nitrate in V1	Nitrite in V1	Average temperature in V1	Thermocline stratification index in V1	Average dissolved oxygen in V1	Halocline stratification index in V1
No. of samples	744	754	752	752	752	751	616	615	614	594
Average value	3.41	276	1.20	0.41	3.73	0.33	14.68	0.63	3.93	1.69
Maximum value	42.76	12040	5.42	1.43	16.62	1.69	20.18	4.42	7.92	6.55
Minimum value	0.07	0	0.05	0.03	0.01	0.01	0	0	0	0
	Maximum chlorophyll 'a' in V2	***D. acuminata* concentration in V2**	**Ammonium in V2**	**Phosphate in V2**	**Nitrate in V2**	**Nitrite in V2**	**Average temperature in V2**	**Thermocline stratification index in V2**	**Average dissolved oxygen in V2**	**Halocline stratification index in V2**
No. of samples	745	756	754	754	754	754	610	610	608	589
Average value	2.61	134	2.07	0.58	4.39	0.42	14.83	0.67	3.75	1.99
Maximum value	22.62	8720	7.55	2.07	19.60	1.82	20.18	4.36	9.35	8.30
Minimum value	0.07	0	0.15	0.08	0.01	0.02	11.22	0	0.00	0
	Maximum chlorophyll 'a' in V3	***D. acuminata* concentration in V3**	**Ammonium in V3**	**Phosphate in V3**	**Nitrate in V3**	**Nitrite in V3**	**Average temperature in V3**	**Thermocline stratification index in V3**	**Average dissolved oxygen in V3**	**Halocline stratification index in V3**
No. of samples	743	755	751	687	752	751	598	598	597	577
Average value	2.33	22	3.19	0.83	4.85	0.51	15.03	0.66	3.62	2.56
Maximum value	22.74	960	10.38	2.33	22.06	2.25	21.62	9.78	7.17	33.53
Minimum value	0.00	0	0.13	0.15	0.22	0.06	8.98	0	-16.34	0
	Maximum chlorophyll 'a' in V4	***D. acuminata* concentration in V4**	**Ammonium in V4**	**Phosphate in V4**	**Nitrate in V4**	**Nitrite in V4**	**Average temperature in V4**	**Thermocline stratification index in V4**	**Average dissolved oxygen in V4**	**Halocline stratification index in V4**
No. of samples	744	755	753	746	753	753	613	613	612	592
Average value	2.32	70	2.62	0.70	4.75	0.47	14.94	0.68	3.72	1.93
Maximum value	26.18	9080	8.66	1.61	19.87	2.27	20.21	5.64	9.00	10.90
Minimum value	0.07	0	0.23	0.08	0.03	0.03	11.17	0	0.01	0
	Maximum chlorophyll 'a' in V5	***D. acuminata* concentration in V5**	**Ammonium in V5**	**Phosphate in V5**	**Nitrate in V5**	**Nitrite in V5**	**Average temperature in V5**	**Thermocline stratification index in V5**	**Average dissolved oxygen in V5**	**Halocline stratification index in V5**
No. of samples	714	743	742	742	742	741	613	613	612	593
Average value	2.94	236	0.83	0.36	3.56	0.30	14.50	0.56	3.90	1.44
Maximum value	25.15	14840	4.69	1.33	17.16	1.50	20.73	3.84	8.78	7.16
Minimum value	0.09	0	0.07	0.03	0.02	0.01	10.95	0	0.01	0
	Maximum chlorophyll 'a' in V6	***D. acuminata* concentration in V6**	**Ammonium in V6**	**Phosphate in V6**	**Nitrate in V6**	**Nitrite in V6**	**Average temperature in V6**	**Thermocline stratification index in V6**	**Average dissolved oxygen in V6**	**Halocline stratification index in V6**
No. of samples	720	749	748	748	747	746	614	614	613	593
Average value	3.49	268	0.92	0.36	3.50	0.31	14.65	0.60	3.96	1.54
Maximum value	43.30	9160	4.48	1.24	18.67	1.73	20.28	5.22	8.11	6.84
Minimum value	0	0	0.05	0.03	0.02	0.01	11.03	0	0.00	0
	Maximum chlorophyll 'a' in V7	***D. acuminata* concentration in V7**	**Ammonium in V7**	**Phosphate in V7**	**Nitrate in V7**	**Nitrite in V7**	**Average temperature in V7**	**Thermocline stratification index in V7**	**Average dissolved oxygen in V7**	**Halocline stratification index in V7**
No. of samples	568	576	574	574	574	574	541	541	541	541
Average value	2.95	311	1.88	0.46	4.43	0.36	14.71	0.63	3.73	1.75
Maximum value	21.90	7280	8.05	1.15	21.51	1.40	20.22	4.21	9.44	7.08
Minimum value	0.07	0	0.07	0.07	0.03	0.03	11.27	0	0.00	0
	Daylight hours	**Insolation**	**Irradiation**	**Upwelling index**	**Week of the year**	**Cangas F state**	**Cangas G state**	**Cangas H state**	**Cangas C state**	**Cangas D state**
No. of samples	247	247	247	763	782	782	782	782	782	782
Average value	6.78	54.05	1517.67	70.40	NA	NA	NA	NA	NA	NA
Maximum value	13.84	91.90	3110	3547.41	53	1	1	1	1	1
Minimum value	0	0	25	-3763.63	1	0	0	0	0	0
	Cangas E state	**Redondela A state**	**Redondela B state**	**Redondela C state**	**Redondela D state**	**Redondela E state**	**Vigo A state**			
No. of samples	782	782	782	782	782	782	782			
Average value	NA	NA	NA	NA	NA	NA	NA			
Maximum value	1	1	1	1	1	1	1			
Minimum value	0	0	0	0	0	0	0			

regions. Therefore, we have studied the effect of sample size (Guallar et al., 2016) and modelling with feature reduction (Rahman and Shahriar, 2013).

As mentioned above, chlorophyll-a concentration is one of the most recurrent biomarkers of potentially toxic phytoplankton proliferation (Rahman and Shahriar, 2013). Chlorophyll-a is related to the concentration of phytoplankton containing this pigment, but not only biotoxin-producing phytoplankton contain it. Therefore, this biomarker may be in error when algal blooms are of non-harmful algae. On the other hand, if the objective is to close mussel production areas as a result of exceeding the legal threshold for the presence of biotoxins (Molares et al., 2020), this could lead to a significant improvement in the accuracy of the prediction.

For this reason, the objective of this study is the creation of a predictive model of high toxicity events in mussel production areas. Consequently, the classification of mussel production areas will focus on whether the presence of lipophilic toxin in mussel flesh exceeds the legal threshold or not. To do this, a comparison of solutions will be carried out, applying different machine learning techniques to predict the state of production areas affected by DSP-type toxins. Taking into account previous studies carried out in the field (Cruz et al., 2021), a total of 6 classification techniques were selected: Artificial Neural Network (ANN), Support Vector Machines (SVMs), k-Nearest Neighbour (kNN), XGBoost, Random Forest and Naïve Bayes. This model can be used by government agencies with responsibilities in the control of shellfish production areas and its use can be of benefit to the mussel industry and the consumer. A workflow of the proposed system can be seen in Fig. 1.

The structure of this paper is defined as follows: It starts with a section on advances in the field of HAB prediction, and in particular in the use of ML techniques for this purpose. In Section 2 the techniques used as well as the configuration of the techniques used are presented. The results of these models can be found in Section 3 and will be interpreted in Section 4. Finally, in Sections 5 and 6 the conclusions obtained and the lines of future work are presented.

Table 3
Distribution of the status of production areas. Non-null sample values refer to samples in which there are no missing values.

	CangasF	CangasG	CangasH	CangasC	CangasD	CangasE	RedondelaA	RedondelaB	RedondelaC	RedondelaD	RedondelaE	VigoA
Samples	783	783	783	783	783	783	783	783	783	783	783	783
Openings	52% (405)	54% (420)	59% (459)	71% (559)	71% (555)	84% (657)	90% (704)	95% (745)	96% (749)	94% (737)	90% (703)	76% (597)
Closures	48% (378)	46% (363)	41% (324)	29% (224)	29% (228)	16% (126)	10% (79)	5% (38)	4% (34)	6% (46)	10% (80)	24% (186)
Non-null samples	175	175	175	175	175	175	175	175	175	175	175	175
Non-null openings	45% (78)	46% (81)	54% (95)	65% (113)	66% (115)	80% (140)	82% (143)	90% (158)	93% (162)	89% (155)	86% (151)	68% (119)
Non-null closures	55% (97)	54% (94)	46% (80)	35% (62)	34% (60)	20% (35)	18% (32)	10% (17)	7% (13)	11% (20)	14% (24)	32% (56)

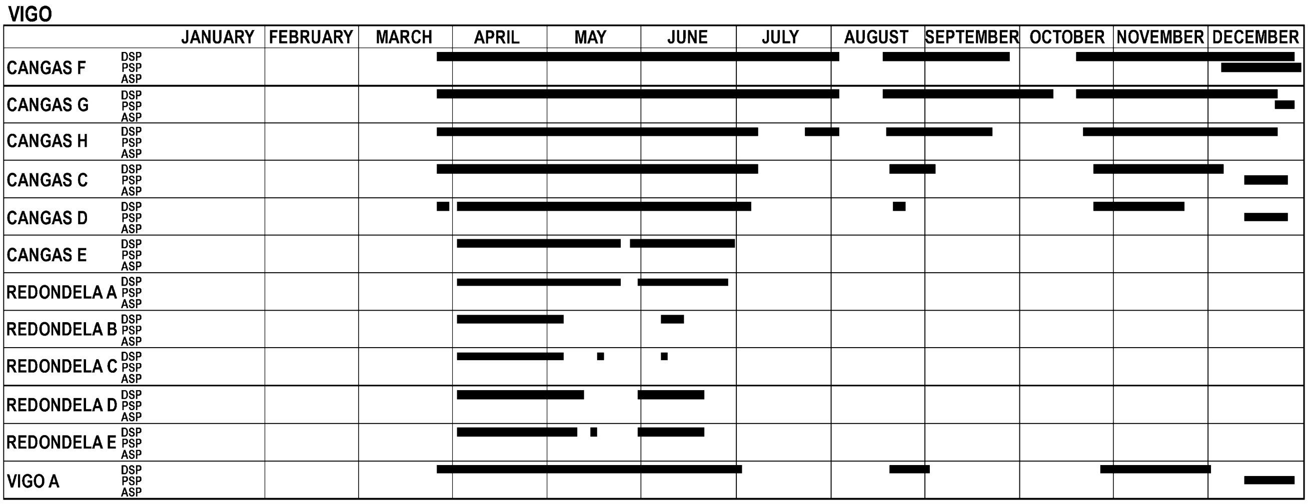

Fig. 4. Distribution of closure episodes caused by HAB in mussel production areas in the Vigo Estuary (2016). Source: http://www.intecmar.gal/Informacion/biotoxinas/Evolucion/DiagramaBateas.aspx.

Table 4
Summary table of the models parameter values used in the grid search.

General Settings	
Validation strategy	10-fold cross-validation
Data normalisation	Yes
Artificial Neural Networks	
Number of input neurons	Number of influencing factors
Output neurons	1
Number of hidden layers	1 and 2
Number of neurons in a one hidden layer network	2, 8 and 14
Number of neurons in a two hidden layers network	[10,10] and [10,20]
Activation function output layer	Sigmoid
Hidden layers activation function	Relu
Optimizer	Adam
Learning rate	0,001
Loss function	Binary crossentropy
Batch size	5
Number of epochs	10
Class weighting	Yes
Support Vector Machines	
Kernel type	Lineal, Gaussian and Polynomial
C value	1
Gamma value (gaussian kernel)	0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8
Grade (polynomial kernel)	2
XGBoost	
Gbtree	
Max depth	6
Learning rate	0,3
Dart	
Sample type	uniform
Normalise type	forest
Gblinear	
Updater	coord_descent
k-Nearest Neighbor	
k value	1, 2, 3, 4, 5, 6, 7, 8, 9 and 10
Random Forest	
Number of trees	100, 500, 1000, 1500 and 2000
Naïve Bayes	
Algorithm	Gaussian, Multinomial, Complement and Bernoulli

2. Materials and Methods

2.1. Dataset and its construction

The production status (open/closed) of the crop areas has been used as the target variable.This status of crop areas is assigned according to whether or not the presence of toxin in the mussel tissue exceeds the legal threshold. If the threshold is exceeded, extraction activities in that crop area will cease (closure of the crop area) or, if not, extraction activity will be allowed (opening of the crop area). It was decided to focus the study on predicting the state of the cultivation areas each Monday. This is because no toxin presence analysis is carried out on the previous days (Saturday and Sunday), which is one of the most compromised points of the existing monitoring system. Twelve out of thirteen mussel production areas of the Vigo estuary (Galician coast, Spain) have been selected: Cangas F, Cangas G, Cangas H, Cangas C, Cangas D, Cangas E, Vigo A, Redondela A, Redondela B, Redondela C, Redondela D and Redondela E (see Fig. 2)excluding from this study the production area of Baiona A because it is a polygon that remains unsampled for long periods of time. As the areas are managed independently, and as input variables, we have used a set of environmental and oceanographic data of different nature, recorded by different institutions between 2004 and 2018. The network of sampling points for phytoplankton monitoring coincides, to a large extent, with the stations set up to determine oceanographic conditions, Fig. 2. Weekly, an oceanographic vessel takes samples from points V1, V2, V3, V4, V5, V6 and V7, located in the Vigo estuary. Their distribution can be seen in the Fig. 3). In each sampling point, integrated samples of water between 0 and 15 metres deep to count phytoplankton cells and determine nutrients dissolved in water, were taken. Simultaneously, a multiparametric probe measures the physico-chemical parameters of the water column. The different variables collected in these oceanographic stations, as well as other constant variables for the whole estuary obtained thanks to METEOGALICIA (met, 2021) and the IEO (IEO, April 27, 2021), are shown in the Table 1. All oceanographic stations have been taken into account in order to know which ones offer the data most related to the occurrence and concentration of HAB, as this depends directly on the functioning of factors such as the morphological configuration of the estuary itself or sea currents. By analysing the data collected, it is determined that the sampling frequency of the data collected is mainly weekly, so this metric will be used as a reference for the creation of the models.

The pre-processing of the input data was as follows:

- The weekly information on chlorophyll-a is collected in three samples divided by depth bands: mean chlorophyll-a between 0 and 5 metres, between 5 and 10 metres and between 10 and 15 metres. Since the presence of toxicity in mussel from any part of the culture rope means the total closure of the production area, the maximum value between the three depths was chosen.

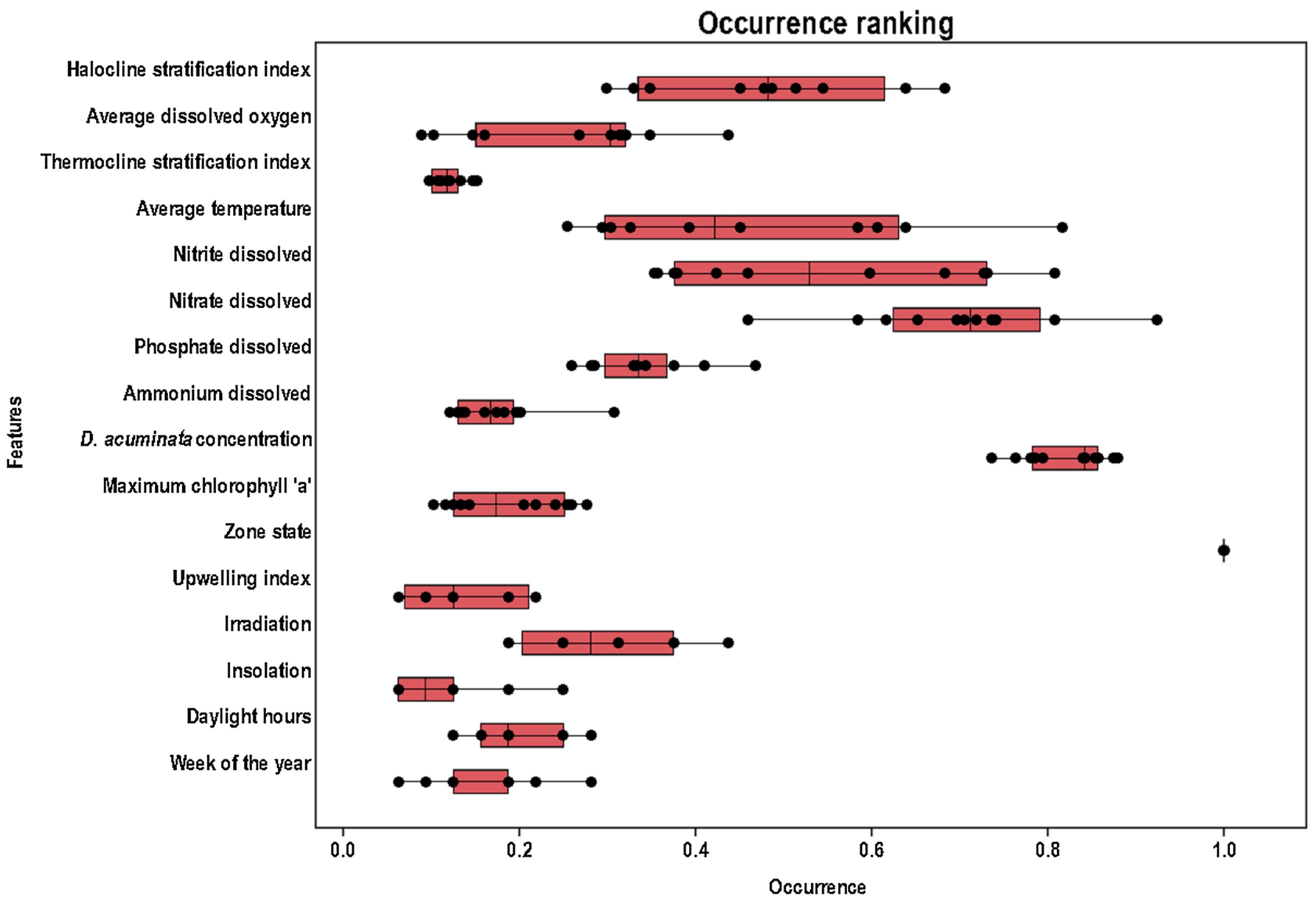

Fig. 5. Summary table with the occurrence of the features after the feature selection processes. Where each point represents the likelihood of a variable being selected as an input feature for a particular production area.

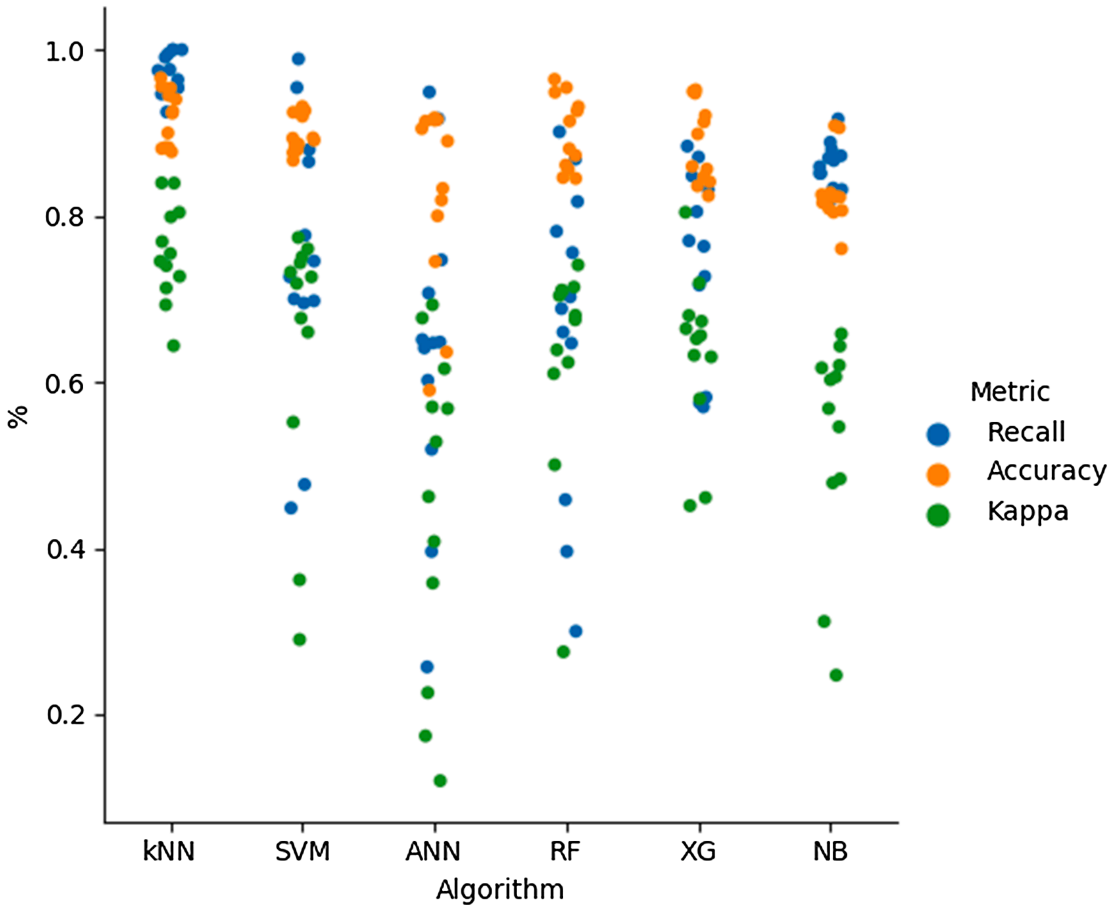

Fig. 6. Combination of recall, accuracy and kappa in the different production zones for each algorithm. The average values for each metric across all folds are shown. In each case, the best performing configurations are represented.

- The count of *Dynophysis acuminata* is a single, weekly value, so information from all available stations was used.
- Nutrient data are collected on a weekly basis and there is only a single piece of data per station, so the count from each oceanographic station was used.
- Environmental values, such as temperature and oxygen, were averaged to unify the information into a single measurement since the data are originally irregular measurements at depths between 0 and 25 metres. Only values up to 12 metres were used for averaging, as this is the length of the mussel ropes. In addition, with the temperature and salinity values, a differential was made between the mean of the first 6 metres and that of the following 6 metres, in order to be able to detect the presence of stratifications, both thermoclines and haloclines.
- The sun data, such as hours of incidence, insolation and irradiation, come from the Meteogalicia weather station, so the data are daily and common for the whole estuary. In order to simplify the input parameters, the weekly average of each of the parameters was calculated.
- The upwelling index data are calculated on a daily basis over four time periods: 00:00 h, 06:00 h, 12:00 h and 18:00 h. In order to simplify the data, the weekly average value was used, thus estimating the predominant value throughout the week.
- To simplify the seasonality into a single value, the date of sampling was transformed, using only the number of the week of the year.
- The specific value of toxins in mussel flesh is a value for which no regular records are covering the whole casuistry in a robust way. Instead, it was concluded that it was possible to classify the status of production areas according to whether the growing area was closed or not. These closures are applied in case the level of toxicity in the mussel flesh exceeds the legal threshold. This information could be obtained by analysing INTECMAR's historical record of closures (INTECMAR, 2 February 2022).

The processing of the 15 years' data resulted in an input dataset of 783 samples. Each of the samples consists of 76 input features. For a more detailed analysis of the input parameters, see Table 2.

This dataset had incomplete samples with missing data for some of the features, so it was necessary to eliminate those rows with such inconsistencies in their data. These samples with missing values were referred to as null samples. After this filtering, a resulting dataset of 175 samples was left. The distribution in the labelling of the samples can be seen in the Table 3. As can be seen in this table, toxicity episodes are more common in the crop areas located in the outer part of the estuary, while their frequency decreases towards the inner parts of the estuary. Fig. 4 shows the behaviour of the HABs that occurred in 2016. An input dataset was created for each of the twelve crop zones; these matrices share 75 of the 76 input features, with the exception of the Friday opening or closing status of the zone to be estimated.

2.2. Machine Learning Models

Based on previous literature, a total of 6 machine learning techniques have been considered: Artificial Neural Networks, Support Vector Machines, XGBoost, k-Nearest Neighbor, Random Forest and Naïve Bayes. These techniques will be tested in order to check which method is the most suitable for the approach proposed in this study. These well-known techniques will be briefly presented below.

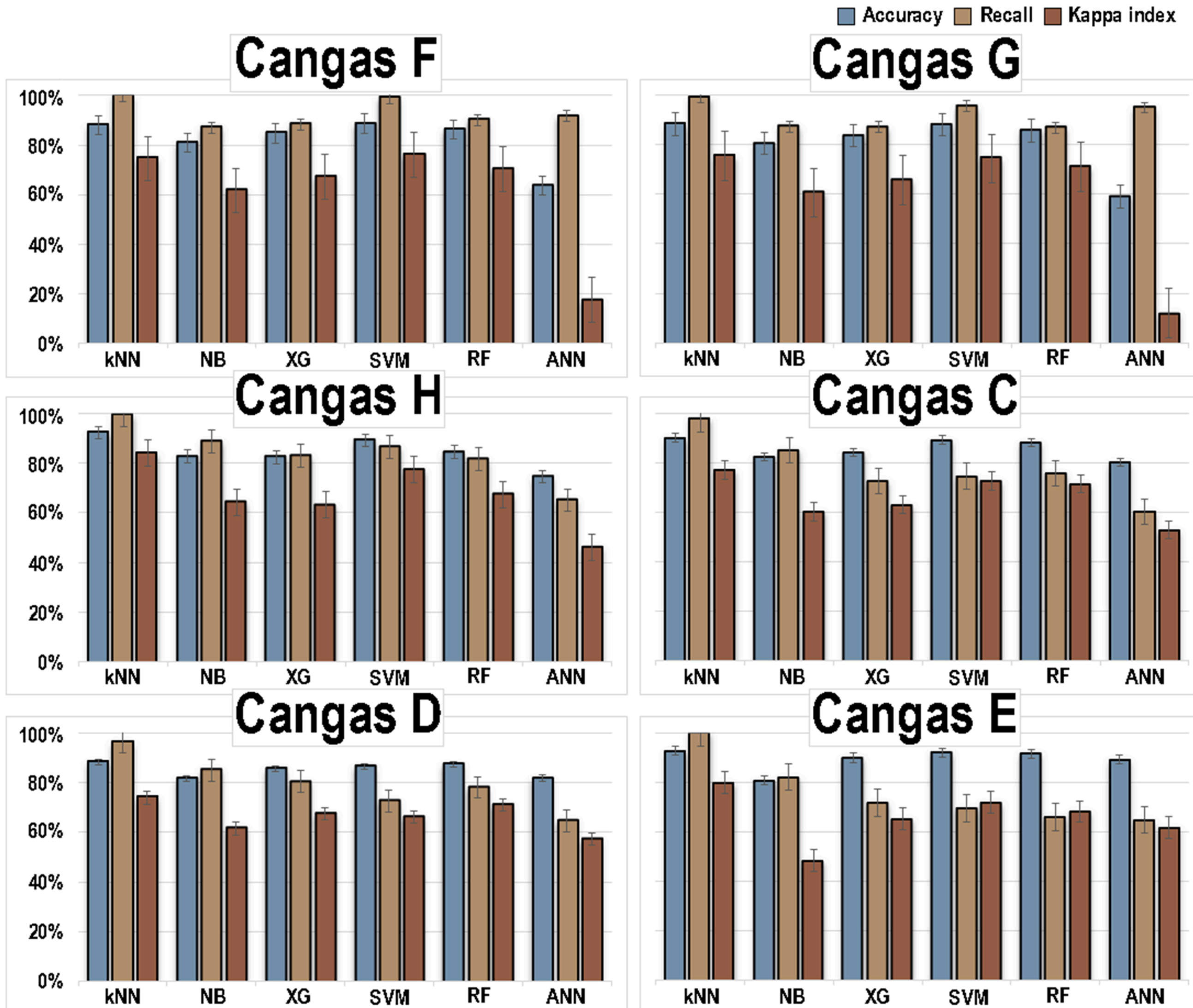

Fig. 7. Combination of recall, accuracy and kappa in the production zones for each algorithm. The average values for each metric across all folds are shown. In each case, the best performing configurations are represented. 1/2.

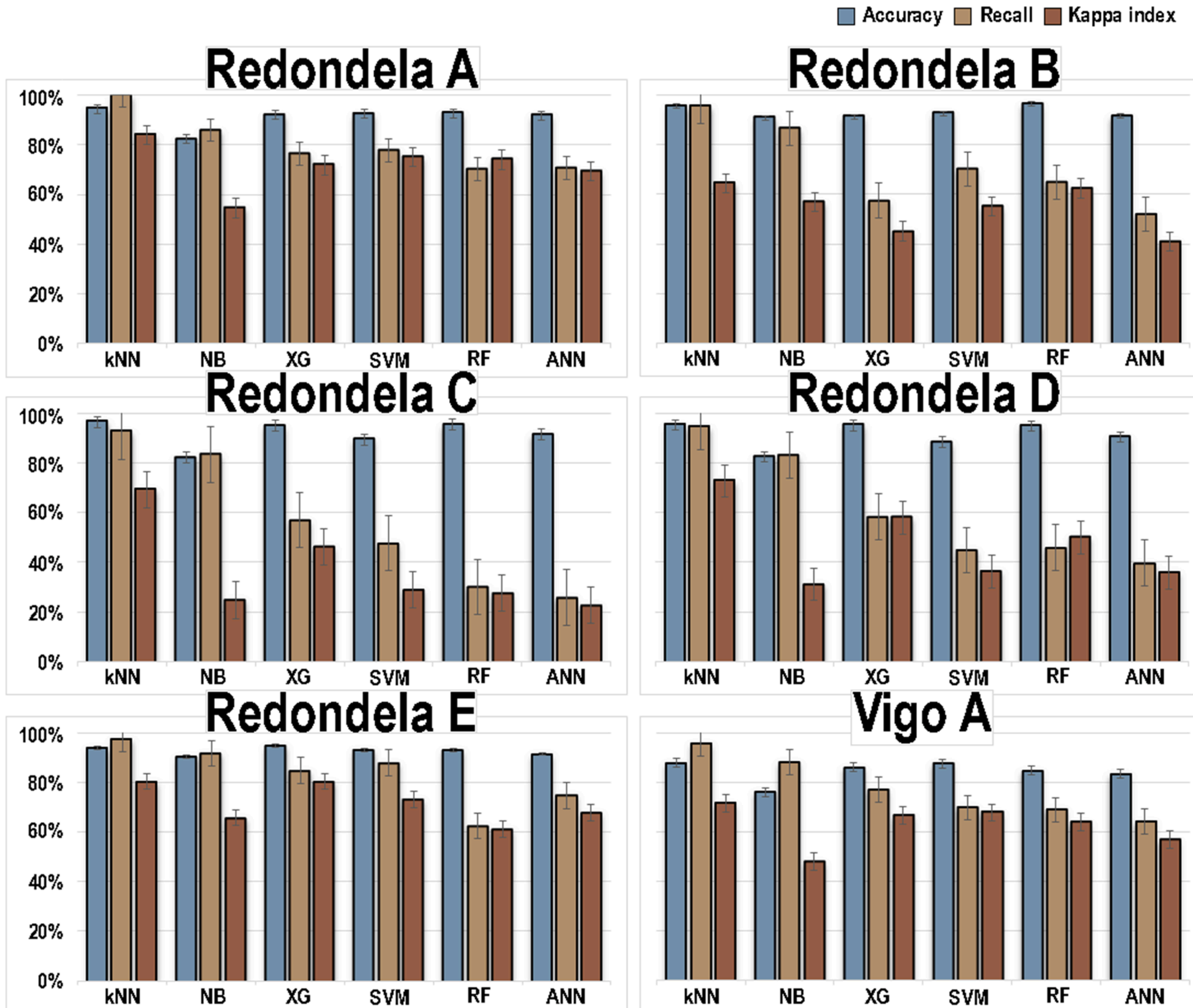

Fig. 8. Combination of recall, accuracy and kappa in the production zones for each algorithm. The average values for each metric across all folds are shown. In each case, the best performing configurations are represented. 2/2.

Table 5
Summary table of the first approach with the models defined as the best in each of the production zones.

Approach 1										
					Recall		Accuracy		Kappa	
Production zone	Corelation filter cuartile	Random Forest filter cuartile	Algorithm	Number of neighbors	μ	σ	μ	σ	μ	σ
Cangas F	-	50	kNN	2	100,00%	0,00%	91,38%	6,37%	0,79	0,14
Cangas G	25	25	kNN	4	99,17%	2,50%	88,50%	7,23%	0,75	0,13
Cangas H	75	75	kNN	2	99,50%	1,50%	91,98%	3,56%	0,83	0,08
Cangas C	50	75	kNN	2	97,61%	2,97%	89,23%	3,64%	0,76	0,09
Cangas D	-	-	kNN	2	96,39%	7,86%	89,23%	6,79%	0,76	0,14
Cangas E	-	-	kNN	2	100,00%	0,00%	92,61%	5,02%	0,80	0,12
Vigo A	50	75	kNN	2	96,32%	3,83%	88,70%	4,01%	0,73	0,09
Redondela A	-	-	kNN	2	100,00%	0,00%	93,93%	3,76%	0,83	0,11
Redondela B	-	-	kNN	2	90,83%	20,56%	90,83%	6,90%	0,64	0,27
Redondela C	50	75	kNN	2	92,50%	16,01%	96,42%	1,92%	0,69	0,13
Redondela D	50	75	kNN	2	93,67%	12,69%	93,69%	3,62%	0,65	0,19
Redondela E	50	75	kNN	2	98,33%	5,00%	93,63%	5,15%	0,82	0,15
Average					**97,03%**	**6,08%**	**91,68%**	**4,83%**	**0,76**	**0,14**

Table 6
Summary table of the second approach with the models defined as the best in each of the production zones.

Approach 2										
					Recall		Accuracy		Kappa	
Production zone	Corelation filter cuartile	Random Forest filter cuartile	Algorithm	Number of neighbors	μ	σ	μ	σ	μ	σ
Cangas F	-	-	kNN	2	100,00%	0,00%	88,10%	7,52%	0,75	0,15
Cangas G	25	-	kNN	2	99,09%	2,73%	87,73%	11,46%	0,74	0,21
Cangas H	50	75	kNN	2	99,50%	1,50%	92,35%	3,54%	0,84	0,08
Cangas C	50	75	kNN	2	97,61%	2,97%	89,98%	3,92%	0,77	0,09
Cangas D	75	75	kNN	2	96,39%	2,58%	87,35%	3,91%	0,73	0,08
Cangas E	-	-	kNN	2	100,00%	0,00%	92,58%	5,09%	0,80	0,14
Vigo A	-	25	kNN	2	95,42%	10,28%	87,71%	6,93%	0,71	0,15
Redondela A	-	-	kNN	2	100,00%	0,00%	94,48%	3,36%	0,84	0,11
Redondela B	-	50	kNN	2	95,42%	10,28%	95,59%	3,20%	0,64	0,24
Redondela C	50	75	kNN	2	92,50%	16,01%	96,60%	1,64%	0,69	0,12
Redondela D	50	25	kNN	2	94,67%	11,08%	95,39%	2,57%	0,73	0,14
Redondela E	-	-	kNN	2	97,50%	7,50%	94,03%	4,43%	0,80	0,16
Average					**97,34%**	**5,41%**	**91,83%**	**4,80%**	**0,75**	**0,14**

2.2.1. *Artificial Neural Networks*

Artificial neural networks (ANNs) are massively parallel interconnected networks of simple (usually adaptive) elements and hierarchical organisation. Artificial neural networks are part of a data analysis technique that, compared to their more rigid and complicated alternatives, offers greater flexibility in processing large volumes of multivariate, non-linear data (White et al., 1992).

2.2.2. *Vector Support Machines*

The classification-regression method Support Vector Machines (SVM) was first proposed by Cortes and Vapnik in 1995 (Cortes and Vapnik, 1995), within the field of computer science. The machine conceptually implements the idea that input vectors are mapped non-linearly into a very high-dimensional feature space. A linear decision surface is constructed in this feature space. The special properties of the decision surface guarantee a high generalisation ability of the learning machine.

2.2.3. *XGBoost*

XGBoost or Extreme Gradient Boosting is an extensible, state-of-the-art application of gradient boosting machines and has been shown to overcome the limits of the computational power of Boosted tree algorithms. Boosting is an ensemble technique in which new models are added to correct errors in existing models. Models are added recursively until no noticeable improvement is found. Gradient boosting is an algorithm in which new models are created to predict the residuals of previous models and then added together to produce a final prediction. It uses a gradient descent algorithm to minimise losses when adding new models (Friedman, 2001).

2.2.4. *k-Nearest Neighbour*

The *k-Nearest Neighbor* (kNN) classifier is an unsupervised machine learning technique for classifying unlabeled observations by assigning them to the class of the most similar labelled examples. The features of the observations are collected for both training and test dataset. The most commonly used metric in the calculations is the Euclidean distance. Another concept is the parameter *k*, which decides how many neighbours will be chosen for the kNN algorithm. The appropriate choice of *k* has a significant impact on the diagnostic performance of the kNN algorithm (Lantz, 2015).

2.2.5. *Random Forest*

Random Forest is an ensemble method, which builds many decision trees that will be used to rank a new instance based on the majority vote. Each node of the decision tree uses a subset of features randomly selected from the original set of features. In addition, each tree uses a different bootstrap data sample, in the same way as bagging. Bagging methods are almost always more accurate than single classifiers. On the other hand, boosting methods can be more accurate than bagging methods but are very sensitive to noise. Random Forest is more robust to noise than boosting methods; performs as well as boosting and sometimes better; and does not overfit (Segal, 2004).

2.2.6. *Naïve Bayes*

Today, the *Naïve Bayes* classifier is used in many applications due to its simple but powerful principle of (Lewis, 1998) accuracy. Bayes' theorem finds the probability of an event occurring given the probability that another event has already occurred. However, this classifier does not take into account the number of occurrences, which is a potentially useful source of additional information. They are called "naïve" because the algorithm assumes that all terms occur independently of each other.

2.3. Performance mesuares

For the analysis of the trained models and their subsequent comparison, 6 statistics were taken into account that were considered relevant when assessing the results (average accuracy, average sensitivity, average kappa coefficient, minimum accuracy, minimum sensitivity and minimum kappa coefficient). In the confusion matrix used to calculate the statistics, closures were defined as positive and openings as negative. Thus, True Positives (*TP*) correspond to those closures correctly classified as closures, True Negatives (*TN*) identify openings classified as such, False Positives (*FP*) represent those openings wrongly classified as closures and, finally, False Negatives (*FN*) are those closures that have been classified as openings.

Calculated according to Eq. (1) accuracy estimates how correctly a binary classification test identifies or excludes a condition. As this is a binary classification paper, this parameter is considered relevant.

$$\frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Not performing a closure when the toxin is present in the mussel poses a higher risk, prioritising the human factor over the economic one. Sensitivity (Eq. (2)) prioritises avoiding misclassifying closures as openings. Sensitivity was therefore the benchmark statistic in this study.

$$\frac{TP}{TP + FN} \tag{2}$$

Cohen's kappa coefficient, calculated according to Eq. (3), is a statistical measure that adjusts for the effect of chance on the proportion of observed agreement between two experts. In this equation, $Pr(a)$ represents the relative observed agreement between the observers, while $Pr(e)$ is the hypothetical probability of agreement by chance. In this study, the model outputs were compared with the labelling performed by the experts to analyse the effect of chance on the models.

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{3}$$

The criteria taken into account when selecting the best models were the values explained above (accuracy, sensitivity and kappa coefficient), as well as the number of features used to make the prediction. A smaller number of input variables would make it easier to make predictions, even on days when certain data are missing. Sensitivity is the most important factor to be taken into account due to the absolute priority of minimising false negatives (as they pose a risk to public health).

2.4. Experimentation setup

By using the strategy of *K-folds* strategy, specifically *10-fold*, yields 10 values of each statistic. The *K-fold cross-validation* procedure randomly divides a dataset into *k* disjoint blocks of approximately equal size, and each block is in turn used to test the model induced from the other $k - 1$ blocks by a classification algorithm. The performance of the classification algorithm is evaluated by the average of the *k*-precisions resulting from the cross-validation of *k*-blocks. This method avoids choosing models with good averages but which perform poorly on certain training blocks, thus ensuring the robustness of the models. The minimum values of the statistics explained above are also taken into account.

Significance analysis was deemed necessary to ensure the robustness of the classification. First, a normality analysis was performed, to ensure that a parametric test can be performed (Sheskin, 2003). When the sample size is at most 50, normality can be tested with the Shapiro–Wilk test test. The Anderson–Darling statistic measures how well the data follow a specific distribution. For a particular data set and distribution, the better the distribution fits the data, the lower this statistic will be. Both the Shapiro–Wilk test and the Anderson–Darling test showed that the sensitivity data for all areas are normal. ANOVA analysis allows multiple means to be compared by studying variances. This was followed by pairwise comparison, in the specific case of this project, with the Tukey–Kramer test. The significance was estimated according to Copenhaver-Holland (Copenhaver and Holland, 1988).

For this study two sets of features were used: one with all 76 input features and another one where the most redundant features were filtered out. In order to do this, a correlation analysis was carried out between the features, and those with a correlation of more than 90% between them were eliminated. This was an empirical approach in which preliminary tests were carried out to eliminate only those variables that really had a very close relationship and leave it to a more purely objective process such as a ranking system to use or assign importance to each. Through this process, influential factors have been sought in less common variables. In this second approach, the 76 input features were reduced to 50.

Then, in each approach, starting from the raw data, a feature selection process was conducted. This has several advantages. Firstly, we make our model easier to interpret. Secondly, we can reduce the variance of the model and thus the overfitting. Finally, we can reduce the computational cost (and time) of training a model. To carry out the feature selection process, the features were ordered using a ranking process. Two ranking techniques were used for this process:

- Applying a filtering method such as correlation with the variable to be forecast. Using the statistical value to rank order the features, three sets of tests were proposed: one with 25% of the best ranking features, one with 50% and the last one with 75%.
- Use of an embedded method such as the Random Forest algorithm. The tree-based strategies used by Random Forest are naturally ranked according to how they improve node purity. This means a decrease in impurity over all trees (called Gini impurity). Nodes with the highest decrease in impurity occur at the beginning of the trees, while notes with the lowest decrease in impurity occur at the end of the trees. Thus, by pruning the trees below a particular node, we can create a subset of the most important features. After applying this ranking, three sets of tests were proposed: one with 25% of the best ranking features, one with 50% and the last one with 75%.

Different experiments have been defined based on the application of one, both or none of the ranking methods mentioned above. To ensure the reliability of the results, the tests were carried out with a cross-validation strategy of *10-fold*. In order to determine the configuration of the best performing models, a grid search was performed and the parameter values of the models used in the training were adjusted as shown in the Table 4.

3. Results

During the feature selection process, the combined Pearson Correlation and Random Forest techniques were applied. Thanks to this, it was possible to extract the importance that these methods give to the features for the classification process. Fig. 5 shows a summary of the behaviour of these methods throughout the production zones, reflecting the percentage of persistence of each variable after the selection processes. It can be seen that the state of the production zone in the week before the prediction day is the most important characteristic, followed by the concentration of *D. accuminata* and the concentration of dissolved nutrients such as nitrate and nitrite. For each of the production zones, a more detailed overview of the feature selection process can be found in Tables 7–18. These tables show how the data collected at each oceanographic station have a different effect on nearby areas. This is due to how marine currents affect the estuary and how certain stations gain importance over others concerning each production area.

By applying each of the 6 machine learning techniques to the 12 production zones independently, it has been possible to observe the comparative solutions offered by each of these methodologies. In Fig. 6 the values of sensitivity, accuracy and kappa obtained by the best models trained with each algorithm and for each production zone can be seen. Algorithms such as kNN or NB obtain more stable results for all the zones, while algorithms such as SVM, RF and XG, although they show certain stability in the values of accuracy, show great variability in the sensitivity values depending on the production zone. The ANN algorithm is presented as the algorithm with the greatest variability in its results.

For a detailed analysis of how the algorithms behave in each of the production zones, please refer to the Figs. 7 and 8. In these graphs it can be seen that the models perform better in the production areas of Cangas F, Cangas G, Cangas H and Redondela A. While the areas where the models have more difficulties in making predictions are: Redondela B, Redondela C and Redondela D.

The models defined as the best in each of the production zones during the first approach are shown on Table 5 and those of the second approach on Table 6. These tables show the sensitivity, accuracy and kappa values. When applying the ten-folds strategy, it is necessary to show the results as the tuple of mean value and standard deviation of the values obtained in each fold.

4. Discussion

The study of the predictor variables for ML models in the prediction of HAB episodes has been one of the most critical points raised in the literature. To date, there is still no consensus on which are the most influential features, varying considerably depending on the geographical region where it is applied and the ML techniques studied. Chlorophyll-a concentration is one of the most relevant features (Deng et al., 2021; Yu et al., 2021), as it is directly related to phytoplankton abundance, but in this study, it has been clearly surpassed by the concentration of *D. accuminata*. This is due to the fact that this marker is more accurate when estimating the lipophilic toxin, this dinoflagellate being one of its main producers. It is also necessary to highlight the importance of nutrients such as nitrate and nitrite (Yu et al., 2021) and environmental factors such as temperature and salinity (Yñiguez and Ottong, 2020).

The results offered by the kNN machine learning algorithm have been the best for the problem analysed in this work, which is the creation of a predictive model of high toxicity events in mussel production areas (reaching mean sensitivity, mean accuracy and mean kappa index values of 97.34%, 91.83% and 0.75 respectively). Its best values of sensitivity, accuracy and kappa have been higher than those obtained with Random forest, ANN, SVM, Naïve Bayes and XGBoost techniques (see Fig. 6). It should be noted that the average kappa value obtained (0.75) has a substantial degree of agreement according to the scale of values proposed by Landis and Koch (Landis and Koch, 1977).

In the Fig. 5, it can be seen how the SVM, ANN, Random Forest and XGBoost algorithms are more susceptible than kNN and Naïve Bayes to the frequency and duration of mussel harvesting prohibition periods in the production areas. This relationship can be seen in the decrease of the sensitivity values in the areas where these periods are less common (Redondela B, Redondela C and Redondela D), while the values of accuracy remain stable. It is necessary to highlight how the performance of the ANNs tends to offer high values of accuracy and low values of sensitivity for the areas where the state of prohibition of extraction is less common, while in the areas where the number of days of prohibition increases (Cangas F and Cangas G), the model offers an improvement in the values of sensitivity to the detriment of accuracy.

These results reflect the imbalance present in the input data which, in areas such as Redondela C, reach a difference of 7% of positives compared to 93% of negatives. Therefore, areas such as Cangas F, Cangas G and Cangas H, which have a distribution of closures of around 60–40%, always obtain better results than areas where FAN is less frequent and where there are fewer cases for the analysis of this study, such as Redondela B, Redondela C and Redondela D, which have a ratio of closures of around 10%.

5. Conclusions

Although the work carried out to date has obtained good results in predicting biomarkers of FAN, the control of the state of the production areas is conditioned by other external factors, which means that the definition of the problem changes. Some work has used real-time prediction of shellfish and fish mortality events as HAB markers (Yñiguez and Ottong, 2020). But real-time prediction does not provide reaction time to these events. However, in this study we have achieved 3-day predictions while maintaining good results. For this we have used the presence of a toxin level above the risk threshold as a HAB marker. In the Galician coast, some previous works seek to solve this problem (Molares et al., 2020), achieving sensitivity and accuracy values of 67.4% and 83% in the production area of “Vigo A” by applying the ANN technique, while in the present work a significant improvement in the results has been achieved.

The approach of the study has shown that it is possible to estimate the status of production areas affected by marine biotoxin events using machine learning techniques. For this purpose, an extensive historical record of variables related to the occurrence of episodes of high toxicity in mussels has been used. The estimates obtained with the models studied have achieved high values of sensitivity and accuracy, so that the expectations initially set out in this study have been met. It has been found that the machine learning algorithm that offers the best results for the resolution of this specific problem in all the production areas of the estuary is the kNN technique. Its best sensitivity and accuracy values have been superior to those obtained with the techniques of Random forest, ANN, SVM, Naïve Bayes and XGBoost.

The models developed during the study can be used to assess the robustness of the decisions taken by experts when managing the opening or closure of production areas in the absence of recent sampling. This dual assessment mechanism can help experts in complex situations where forecast errors are more likely.

6. Future Works

In this work, 6 different machine learning algorithms were studied to solve the problem. It is proposed to compare the results obtained with other alternative algorithms that can approach the problem from another perspective, such as hybrid machine learning algorithms (Behera et al., 2016).

The study has focused only on the Vigo estuary, as it is one of the most important Galician estuaries for the production of mussels, and because of its geomorphological characteristics that give it a behaviour in the distribution and evolution of algal blooms of great scientific interest. However, the study is continuing with the aim of supporting the rest of the Galician estuaries with mussel production.

In this study, variables identified as relevant in the state of the art have been selected. However, other new variables (e.g. wind, currents, other toxic phytoplankton species, etc.) could be considered as input parameters in the training of machine learning algorithms.

One of the limiting factors in conducting this study has been the amount of missing data from the time series used as data sets. It is therefore considered that the creation of a system capable of obtaining or generating (synthetic data (Chen et al., 2021)) such data could lead to a significant improvement in the results obtained.

Declaration of Competing Interest

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

Acknowledgments

The authors want to acknowledge the support from INTECMAR, who have provide mostly data for this work and CESGA, who allows to conduct the tests on their installations. Funding for open access charge: Universidade da Coruña/CISUG. This work is supported by the "Collaborative Project in Genomic Data Integration (CICLOGEN) PI17/01826 funded by the Carlos III Health Institute in the context of the Spanish National Plan for Scientific and Technical Research and Innovation 2013–2016 and the European Regional Development Funds (FEDER)—"A way to build Europe." This project was also supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (Ref. ED431G/01, ED431D 2017/16), the "Galician Network for Colorectal Cancer Research" (Ref. ED431D 2017/23), Competitive Reference Groups (Ref. ED431C 2018/49) and the Spanish Ministry of Economy and Competitiveness via funding of the unique installation BIOCAI (UNLC08-1E-002, UNLC13-13–3503) and the European Regional Development Funds (FEDER). Enrique Fernandez-Blanco would also like to thank NVidia corp., which granted a GPU used in this work for the preliminary tests.

Appendix A

Tables 7–18

Table 7
Table with the input features associated with each test block in the Cangas F zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓	✓	✓													✓	✓	✓													
Daylight hours	-	✓				✓												✓				✓											
Insolation	-	✓																															
Irradiation	-	✓	✓	✓		✓				✓	✓																						
Upwelling index	-	✓																✓															
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓										
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V3	✓	✓															✓															
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓							✓	✓	✓		✓	✓	✓		✓							
	V5	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓																			
	V6	✓	✓			✓	✓			✓																							
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	
D. acuminata concentration	V1	✓				✓												✓				✓											
	V2	✓	✓																														
	V3	✓	✓	✓	✓	✓	✓																										
	V4	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓			✓	✓	✓					
	V5	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
	V6	✓	✓																														
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓						
Ammonium	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V2	✓	✓			✓	✓			✓								✓	✓	✓		✓	✓			✓	✓	✓					
	V3	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V4	✓				✓												✓				✓											
	V5	✓																✓	✓														
	V6	✓	✓			✓	✓											✓	✓	✓		✓	✓	✓		✓							
	V7	✓	✓			✓												✓	✓			✓	✓										
Phosphate	V1	✓	✓			✓	✓	✓		✓	✓			✓				✓	✓	✓		✓				✓	✓	✓		✓			
	V2	✓	✓															✓															
	V3	✓	✓															✓	✓														
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V5	✓	✓	✓		✓	✓	✓		✓	✓																						
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V7	✓				✓												✓				✓											
Nitrate	V1	✓	✓															✓	✓														
	V2	✓	✓	✓		✓	✓			✓	✓	✓						✓	✓	✓		✓	✓	✓		✓							
	V3	✓	✓			✓	✓											✓	✓			✓				✓	✓	✓					
	V4	✓				✓	✓	✓		✓								✓				✓				✓	✓						
	V5	✓	✓															✓	✓														
	V6	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓										
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																			
Nitrite	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V2	✓	✓	✓		✓	✓	✓		✓				✓	✓																		
	V3	✓																✓															
	V4	✓	✓																														
	V5	✓	✓			✓	✓			✓	✓	✓																					
	V6	✓																✓	✓														
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Average temperature	V1	✓	✓			✓	✓											✓	✓	✓		✓	✓	✓									
	V2	✓	✓	✓		✓	✓			✓								✓	✓			✓				✓							
	V3	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V4	✓	✓			✓				✓								✓				✓				✓							
	V5	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓																✓															
	V7	✓																✓															
Thermocline stratification index	V1	✓	✓	✓		✓	✓																										
	V2	✓	✓			✓	✓											✓	✓			✓	✓	✓									
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V4	✓	✓															✓															
	V5	✓	✓	✓		✓	✓	✓		✓								✓	✓	✓		✓	✓	✓		✓							
	V6	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V7	✓				✓	✓			✓																							
Average dissolved oxygen	V1	✓	✓	✓		✓	✓	✓		✓	✓																						
	V2	✓																✓	✓			✓											
	V3	✓																															
	V4	✓	✓			✓	✓																										
	V5	✓	✓			✓												✓				✓											
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V7	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓									
Halocline stratification index	V1	✓	✓			✓				✓								✓	✓			✓	✓			✓	✓						
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓																			
	V3	✓				✓				✓																							
	V4	✓	✓	✓		✓	✓			✓	✓	✓																					
	V5	✓				✓												✓				✓											
	V6	✓	✓																														
	V7	✓	✓			✓	✓																										

Computers and Electronics in Agriculture 197 (2022) 106956

Table 8
Table with the input features associated with each test block in the Cangas G zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓	✓														✓	✓	✓													
Daylight hours	-	✓	✓			✓												✓	✓			✓	✓			✓							
Insolation	-	✓	✓			✓	✓																										
Irradiation	-	✓	✓			✓	✓			✓	✓																						
Upwelling index	-	✓	✓															✓															
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓			✓												✓	✓	✓		✓	✓										
	V2	✓				✓	✓	✓		✓	✓							✓	✓			✓	✓			✓	✓						
	V3	✓	✓			✓												✓	✓	✓		✓	✓										
	V4	✓	✓			✓	✓											✓	✓			✓	✓			✓	✓						
	V5	✓	✓															✓	✓														
	V6	✓	✓			✓	✓											✓	✓			✓	✓	✓									
	V7	✓	✓			✓	✓											✓	✓	✓		✓	✓			✓	✓						
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓			
	V3	✓	✓			✓	✓			✓	✓			✓				✓	✓			✓				✓				✓			
	V4	✓	✓			✓				✓								✓				✓				✓				✓			
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ammonium	V1	✓	✓	✓														✓	✓														
	V2	✓																															
	V3	✓	✓	✓														✓	✓														
	V4	✓																✓	✓														
	V5	✓	✓			✓	✓											✓	✓	✓		✓	✓	✓									
	V6	✓	✓															✓															
	V7	✓	✓	✓	✓	✓	✓	✓	✓									✓	✓	✓	✓	✓	✓	✓	✓								
Phosphate	V1	✓	✓			✓	✓	✓		✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V2	✓	✓	✓		✓	✓											✓	✓			✓	✓										
	V3	✓	✓															✓	✓														
	V4	✓	✓															✓	✓	✓		✓	✓										
	V5	✓				✓				✓								✓	✓			✓				✓							
	V6	✓				✓	✓			✓								✓	✓	✓		✓	✓			✓							
	V7	✓	✓	✓		✓	✓											✓	✓			✓	✓			✓	✓	✓					
Nitrate	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓																			
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V3	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓																		
	V5	✓	✓			✓	✓	✓		✓				✓				✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓		
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓																		
	V7	✓	✓	✓	✓	✓	✓	✓		✓																							
Nitrite	V1	✓	✓	✓		✓	✓			✓	✓																						
	V2	✓	✓	✓		✓	✓			✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V3	✓	✓	✓		✓	✓	✓		✓																							
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓																						
	V5	✓				✓				✓								✓	✓			✓	✓			✓							
	V6	✓	✓			✓	✓			✓																							
	V7	✓				✓				✓																							
Average temperature	V1	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V2	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
Thermocline stratification index	V1	✓				✓												✓	✓			✓											
	V2	✓				✓												✓	✓			✓	✓										
	V3	✓				✓												✓				✓											
	V4	✓																✓															
	V5	✓																✓															
	V6	✓	✓			✓												✓				✓											
	V7	✓	✓			✓												✓				✓	✓										
Average dissolved oxygen	V1	✓	✓																														
	V2	✓																✓															
	V3	✓																✓															
	V4	✓	✓																														
	V5	✓																✓															
	V6	✓																															
	V7	✓	✓																														
Halocline stratification index	V1	✓	✓			✓	✓																										
	V2	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓						✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V4	✓	✓	✓		✓	✓	✓		✓	✓	✓																					
	V5	✓	✓	✓		✓	✓	✓		✓	✓																						
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓																					
	V7	✓	✓			✓	✓																										

Table 9
Table with the input features associated with each test block in the Cangas H zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓															✓	✓														
Daylight hours	-	✓				✓												✓	✓			✓											
Insolation	-	✓																															
Irradiation	-	✓	✓			✓	✓																										
Upwelling index	-	✓																✓	✓														
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓			✓	✓											✓	✓	✓		✓											
	V2	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓			✓	✓											✓	✓	✓		✓	✓										
	V4	✓	✓	✓		✓	✓											✓	✓			✓	✓			✓							
	V5	✓	✓															✓	✓														
	V6	✓	✓			✓	✓											✓	✓			✓	✓	✓									
	V7	✓	✓			✓												✓				✓	✓										
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓		✓				✓				✓				✓	✓	✓		✓	✓			✓	✓			✓			
	V4	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓			
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ammonium	V1	✓	✓	✓														✓	✓														
	V2	✓																															
	V3	✓	✓															✓	✓														
	V4	✓																✓															
	V5	✓	✓			✓	✓											✓	✓			✓	✓										
	V6	✓																✓															
	V7	✓	✓	✓														✓	✓	✓													
Phosphate	V1	✓	✓			✓				✓								✓	✓			✓	✓			✓							
	V2	✓	✓			✓	✓	✓		✓	✓							✓	✓	✓		✓	✓			✓							
	V3	✓	✓			✓												✓				✓											
	V4	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓			✓							
	V5	✓	✓			✓	✓			✓								✓	✓			✓	✓			✓							
	V6	✓	✓			✓	✓			✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V7	✓	✓			✓	✓	✓		✓	✓							✓	✓	✓		✓	✓			✓							
Nitrate	V1	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓																		
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V3	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V4	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓																		
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓																		
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓																		
Nitrite	V1	✓	✓			✓	✓			✓																							
	V2	✓	✓	✓	✓	✓	✓			✓	✓							✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V3	✓	✓	✓		✓	✓			✓	✓	✓																					
	V4	✓	✓	✓		✓	✓	✓		✓	✓																						
	V5	✓				✓				✓								✓				✓				✓							
	V6	✓	✓			✓				✓																							
	V7	✓				✓				✓																							
Average temperature	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V2	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓			✓	✓			✓																			
	V5	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓			✓	✓	✓	✓	✓	✓	✓																					
	V7	✓	✓	✓		✓	✓	✓		✓	✓																						
Thermocline stratification index	V1	✓				✓												✓	✓			✓											
	V2	✓				✓	✓											✓	✓			✓	✓										
	V3	✓				✓												✓	✓			✓											
	V4	✓																✓															
	V5	✓	✓															✓	✓														
	V6	✓	✓															✓	✓														
	V7	✓	✓			✓	✓											✓				✓	✓										
Average dissolved oxygen	V1	✓	✓	✓		✓	✓																										
	V2	✓	✓															✓	✓														
	V3	✓	✓															✓	✓			✓	✓										
	V4	✓	✓																														
	V5	✓	✓															✓	✓														
	V6	✓	✓																														
	V7	✓	✓																														
Halocline stratification index	V1	✓	✓	✓		✓	✓	✓																									
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V4	✓	✓	✓	✓	✓	✓	✓																									
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V6	✓	✓	✓	✓	✓	✓			✓	✓																						
	V7	✓	✓	✓		✓	✓	✓		✓	✓	✓																					

Table 10
Table with the input features associated with each test block in the Cangas C zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓																✓															
Daylight hours	-	✓				✓												✓	✓			✓	✓										
Insolation	-	✓	✓																														
Irradiation	-	✓				✓	✓																										
Upwelling index	-	✓	✓	✓														✓	✓			✓	✓										
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓															✓	✓			✓	✓										
	V2	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓									
	V3	✓	✓			✓	✓											✓	✓	✓		✓											
	V4	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V5	✓																✓															
	V6	✓	✓															✓															
	V7	✓	✓			✓	✓											✓	✓			✓	✓										
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓			✓	✓			✓				✓	✓	✓		✓				✓				✓			
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Ammonium	V1	✓																✓	✓			✓											
	V2	✓				✓																											
	V3	✓	✓															✓	✓	✓													
	V4	✓																✓	✓														
	V5	✓	✓	✓														✓	✓			✓	✓										
	V6	✓																✓															
	V7	✓	✓															✓	✓			✓											
Phosphate	V1	✓	✓			✓				✓								✓	✓			✓	✓			✓	✓						
	V2	✓	✓	✓		✓	✓	✓		✓	✓	✓						✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓			✓	✓											✓				✓											
	V4	✓				✓												✓	✓			✓				✓							
	V5	✓	✓			✓				✓								✓	✓			✓	✓			✓							
	V6	✓				✓				✓								✓	✓			✓				✓							
	V7	✓	✓			✓	✓			✓								✓	✓			✓	✓			✓							
Nitrate	V1	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V3	✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓																	
	V4	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓																			
	V5	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓			✓			
	V6	✓	✓			✓	✓	✓		✓	✓																						
	V7	✓	✓	✓		✓	✓	✓		✓				✓	✓																		
Nitrite	V1	✓	✓			✓	✓			✓	✓																						
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V3	✓	✓	✓		✓	✓			✓	✓																						
	V4	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓																		
	V5	✓				✓				✓								✓				✓	✓			✓							
	V6	✓	✓			✓				✓																							
	V7	✓	✓			✓				✓																							
Average temperature	V1	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V2	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V4	✓	✓	✓		✓	✓	✓		✓	✓																						
	V5	✓				✓												✓				✓	✓			✓							
	V6	✓	✓			✓	✓	✓		✓	✓																						
	V7	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
Thermocline stratification index	V1	✓	✓															✓	✓														
	V2	✓																✓	✓														
	V3	✓	✓															✓	✓	✓													
	V4	✓	✓															✓															
	V5	✓	✓															✓	✓														
	V6	✓	✓	✓														✓	✓														
	V7	✓																✓															
Average dissolved oxygen	V1	✓	✓	✓		✓	✓	✓																									
	V2	✓	✓			✓	✓											✓	✓			✓	✓										
	V3	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓			✓	✓	✓					
	V4	✓	✓	✓		✓	✓																										
	V5	✓	✓			✓	✓	✓										✓	✓	✓		✓	✓										
	V6	✓	✓			✓																											
	V7	✓	✓	✓		✓	✓	✓																									
Halocline stratification index	V1	✓	✓			✓	✓	✓																									
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓						✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V4	✓	✓	✓		✓	✓	✓		✓	✓																						
	V5	✓	✓	✓		✓	✓			✓	✓																						
	V6	✓	✓	✓		✓	✓																										
	V7	✓	✓	✓		✓	✓	✓		✓	✓	✓																					

Table 11
Table with the input features associated with each test block in the Cangas D zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓															✓	✓			✓	✓										
Daylight hours	-	✓	✓			✓												✓	✓			✓											
Insolation	-	✓	✓	✓																													
Irradiation	-	✓	✓			✓	✓																										
Upwelling index	-	✓																✓															
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓	✓														
	V2	✓	✓	✓	✓	✓	✓	✓	✓									✓	✓	✓	✓	✓	✓	✓									
	V3	✓	✓			✓	✓											✓	✓			✓											
	V4	✓	✓	✓	✓	✓	✓	✓	✓									✓	✓	✓		✓	✓	✓									
	V5	✓																✓															
	V6	✓																✓															
	V7	✓	✓															✓	✓			✓	✓										
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V3	✓	✓	✓		✓	✓			✓	✓	✓		✓				✓	✓	✓		✓	✓			✓	✓			✓			
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	
Ammonium	V1	✓	✓															✓	✓														
	V2	✓																															
	V3	✓	✓	✓														✓	✓	✓													
	V4	✓																✓															
	V5	✓	✓			✓	✓											✓	✓			✓	✓										
	V6	✓	✓															✓	✓														
	V7	✓	✓			✓	✓	✓										✓	✓	✓		✓	✓										
Phosphate	V1	✓	✓			✓				✓								✓	✓			✓				✓	✓						
	V2	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓				✓	✓	✓		✓	✓	✓		✓	✓			✓			
	V3	✓	✓			✓												✓	✓			✓	✓										
	V4	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓			✓	✓						
	V5	✓	✓	✓		✓	✓			✓	✓							✓	✓			✓				✓							
	V6	✓				✓				✓								✓	✓			✓	✓			✓							
	V7	✓	✓			✓	✓											✓	✓			✓	✓			✓							
Nitrate	V1	✓	✓			✓	✓			✓	✓																						
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓				✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓		
	V3	✓	✓	✓		✓	✓			✓	✓			✓	✓																		
	V4	✓	✓	✓		✓	✓			✓	✓	✓		✓	✓																		
	V5	✓				✓				✓	✓			✓	✓			✓	✓	✓		✓				✓	✓						
	V6	✓	✓			✓	✓			✓																							
	V7	✓	✓			✓				✓	✓																						
Nitrite	V1	✓	✓			✓	✓			✓	✓																						
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓		
	V3	✓	✓	✓	✓	✓	✓			✓	✓			✓																			
	V4	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓																		
	V5	✓	✓			✓	✓			✓								✓				✓				✓							
	V6	✓	✓			✓	✓			✓	✓																						
	V7	✓	✓			✓	✓			✓																							
Average temperature	V1	✓	✓	✓		✓	✓			✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V2	✓	✓			✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V3	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V4	✓	✓			✓	✓			✓																							
	V5	✓				✓												✓	✓			✓	✓			✓							
	V6	✓	✓			✓	✓																										
	V7	✓	✓	✓		✓				✓																							
Thermocline stratification index	V1	✓																✓															
	V2	✓																✓															
	V3	✓	✓															✓															
	V4	✓	✓															✓															
	V5	✓	✓	✓														✓	✓			✓	✓										
	V6	✓	✓	✓														✓	✓			✓											
	V7	✓																✓															
Average dissolved oxygen	V1	✓	✓	✓		✓	✓																										
	V2	✓	✓			✓	✓											✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓			✓	✓						
	V4	✓	✓	✓		✓	✓	✓																									
	V5	✓	✓			✓	✓											✓	✓	✓		✓	✓	✓									
	V6	✓				✓	✓																										
	V7	✓	✓	✓		✓	✓																										
Halocline stratification index	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V5	✓	✓	✓	✓	✓	✓			✓	✓	✓																					
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V7	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓																					

Table 12
Table with the input features associated with each test block in the Cangas E zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓																✓				✓											
Daylight hours	-	✓	✓			✓	✓											✓	✓			✓	✓										
Insolation	-	✓	✓																														
Irradiation	-	✓	✓			✓																											
Upwelling index	-	✓	✓															✓	✓			✓	✓										
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓	✓														
	V2	✓																✓	✓														
	V3	✓	✓															✓	✓	✓													
	V4	✓	✓	✓														✓	✓	✓		✓	✓	✓									
	V5	✓	✓															✓															
	V6	✓	✓															✓	✓														
	V7	✓	✓															✓															
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓				✓	✓	✓		✓	✓	✓		✓	✓	✓		✓			
	V4	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓			
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
Ammonium	V1	✓				✓												✓	✓			✓	✓										
	V2	✓				✓																											
	V3	✓																✓	✓														
	V4	✓	✓															✓	✓														
	V5	✓																✓															
	V6	✓	✓			✓	✓											✓				✓	✓										
	V7	✓																✓	✓														
Phosphate	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓			✓	✓	✓		✓	✓	✓		✓	✓						
	V2	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓	✓			✓							
	V3	✓				✓												✓				✓											
	V4	✓				✓												✓	✓			✓	✓			✓							
	V5	✓	✓	✓		✓				✓								✓	✓			✓				✓							
	V6	✓	✓			✓	✓			✓								✓	✓			✓	✓			✓	✓						
	V7	✓	✓	✓		✓	✓			✓								✓	✓			✓	✓			✓							
Nitrate	V1	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓																
	V5	✓	✓	✓		✓	✓			✓	✓			✓	✓			✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
	V7	✓	✓	✓		✓	✓	✓		✓																							
Nitrite	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V2	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓			✓	✓			✓	✓																						
	V4	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓																		
	V5	✓	✓	✓		✓	✓	✓		✓	✓			✓				✓	✓	✓		✓	✓	✓		✓	✓			✓	✓		
	V6	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓																		
	V7	✓	✓	✓		✓	✓			✓																							
Average temperature	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓							✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V2	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V3	✓	✓	✓		✓	✓			✓	✓																						
	V4	✓	✓	✓		✓	✓			✓																							
	V5	✓	✓			✓	✓											✓	✓	✓		✓	✓										
	V6	✓				✓	✓																										
	V7	✓	✓	✓		✓	✓			✓	✓																						
Thermocline stratification index	V1	✓																✓				✓											
	V2	✓				✓												✓				✓											
	V3	✓				✓												✓				✓											
	V4	✓																✓															
	V5	✓	✓															✓															
	V6	✓																✓	✓														
	V7	✓																✓				✓											
Average dissolved oxygen	V1	✓	✓	✓		✓	✓	✓		✓	✓	✓																					
	V2	✓	✓	✓		✓	✓	✓		✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V3	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V4	✓	✓	✓		✓	✓	✓																									
	V5	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓	✓		✓																											
	V7	✓	✓	✓		✓	✓																										
Halocline stratification index	V1	✓	✓	✓		✓	✓	✓																									
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
	V5	✓	✓	✓	✓	✓	✓	✓																									
	V6	✓	✓			✓	✓	✓																									
	V7	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓																				

Table 13
Table with the input features associated with each test block in the Redondela A zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓				✓												✓				✓											
Daylight hours	-	✓				✓												✓	✓			✓											
Insolation	-	✓																															
Irradiation	-	✓	✓	✓		✓	✓			✓																							
Upwelling index	-	✓																✓															
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓															
	V2	✓	✓															✓															
	V3	✓	✓															✓	✓	✓													
	V4	✓	✓	✓														✓	✓														
	V5	✓																✓															
	V6	✓	✓															✓	✓														
	V7	✓																✓															
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓		✓	✓			✓	✓	✓		✓				✓	✓	✓		✓	✓	✓		✓	✓			✓			
	V4	✓	✓	✓		✓	✓	✓		✓	✓			✓				✓	✓	✓		✓	✓	✓		✓	✓			✓			
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ammonium	V1	✓				✓												✓	✓			✓				✓							
	V2	✓				✓																											
	V3	✓	✓															✓	✓			✓											
	V4	✓	✓															✓	✓			✓	✓										
	V5	✓																✓															
	V6	✓				✓												✓				✓											
	V7	✓	✓															✓	✓														
Phosphate	V1	✓	✓	✓	✓	✓	✓			✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V2	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V3	✓	✓			✓	✓											✓	✓			✓	✓										
	V4	✓	✓	✓		✓	✓			✓								✓	✓	✓		✓	✓			✓							
	V5	✓	✓			✓				✓								✓	✓	✓		✓	✓			✓							
	V6	✓	✓			✓	✓	✓		✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V7	✓	✓			✓				✓								✓	✓			✓	✓			✓							
Nitrate	V1	✓	✓	✓		✓	✓			✓	✓	✓																					
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V5	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																				
	V7	✓	✓	✓		✓	✓	✓		✓																							
Nitrite	V1	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓																		
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V5	✓	✓			✓	✓			✓	✓			✓				✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓																		
	V7	✓	✓	✓		✓	✓			✓																							
Average temperature	V1	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓	✓			✓	✓						
	V2	✓	✓			✓	✓			✓								✓	✓			✓	✓			✓							
	V3	✓	✓	✓		✓				✓																							
	V4	✓	✓			✓				✓																							
	V5	✓				✓												✓	✓			✓	✓										
	V6	✓				✓																											
	V7	✓	✓			✓																											
Thermocline stratification index	V1	✓																✓				✓											
	V2	✓																✓	✓			✓											
	V3	✓				✓	✓											✓	✓			✓											
	V4	✓	✓															✓															
	V5	✓	✓															✓															
	V6	✓	✓															✓	✓														
	V7	✓	✓															✓				✓											
Average dissolved oxygen	V1	✓	✓	✓		✓	✓			✓	✓																						
	V2	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓						
	V3	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓		✓	✓			✓	✓						
	V4	✓	✓			✓	✓																										
	V5	✓	✓			✓	✓											✓	✓			✓	✓										
	V6	✓	✓	✓		✓	✓																										
	V7	✓	✓	✓		✓	✓																										
Halocline stratification index	V1	✓	✓	✓		✓	✓	✓																									
	V2	✓	✓	✓	✓	✓	✓	✓										✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V5	✓	✓	✓		✓	✓																										
	V6	✓	✓	✓		✓	✓	✓																									
	V7	✓	✓	✓		✓	✓	✓		✓	✓																						

Table 14
Table with the input features associated with each test block in the Redondela B zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓				✓												✓				✓											
Daylight hours	-	✓	✓			✓	✓											✓	✓			✓	✓										
Insolation	-	✓																															
Irradiation	-	✓				✓	✓			✓	✓																						
Upwelling index	-	✓	✓															✓				✓											
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓	✓														
	V2	✓	✓															✓				✓	✓										
	V3	✓	✓															✓	✓														
	V4	✓	✓	✓														✓	✓	✓		✓	✓	✓									
	V5	✓	✓															✓	✓														
	V6	✓	✓															✓	✓														
	V7	✓	✓															✓															
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓		✓	✓			✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓			
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓		✓	✓	✓		✓	✓							✓	✓	✓		✓	✓	✓		✓	✓	✓		✓			
	V7	✓	✓	✓		✓	✓	✓		✓	✓			✓				✓	✓			✓	✓	✓		✓	✓			✓			
Ammonium	V1	✓	✓	✓		✓	✓	✓	✓	✓	✓							✓	✓	✓	✓	✓	✓	✓		✓	✓						
	V2	✓	✓			✓	✓																										
	V3	✓	✓			✓	✓											✓	✓			✓	✓										
	V4	✓	✓			✓	✓											✓	✓			✓	✓										
	V5	✓	✓			✓	✓											✓	✓	✓		✓	✓										
	V6	✓	✓			✓	✓	✓		✓	✓							✓	✓	✓		✓	✓			✓	✓						
	V7	✓	✓															✓	✓														
Phosphate	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓		
	V2	✓	✓	✓		✓	✓	✓		✓								✓	✓	✓	✓	✓	✓	✓		✓	✓						
	V3	✓				✓												✓	✓			✓	✓										
	V4	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓	✓			✓							
	V5	✓	✓	✓		✓	✓			✓								✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓				
	V7	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓		✓	✓			✓	✓						
Nitrate	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V7	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓																		
Nitrite	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓																		
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓																				
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V5	✓	✓	✓	✓	✓	✓			✓	✓							✓	✓	✓		✓	✓	✓	✓	✓	✓	✓					
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V7	✓	✓	✓		✓	✓			✓	✓																						
Average temperature	V1	✓	✓			✓	✓			✓								✓	✓	✓		✓				✓							
	V2	✓	✓			✓				✓								✓	✓			✓				✓							
	V3	✓	✓	✓		✓	✓			✓	✓																						
	V4	✓	✓			✓				✓																							
	V5	✓	✓			✓	✓			✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓	✓		✓	✓			✓																							
	V7	✓	✓			✓				✓																							
Thermocline stratification index	V1	✓	✓															✓															
	V2	✓																✓															
	V3	✓				✓												✓				✓											
	V4	✓	✓															✓															
	V5	✓	✓	✓														✓	✓														
	V6	✓																✓															
	V7	✓	✓															✓															
Average dissolved oxygen	V1	✓				✓																											
	V2	✓	✓			✓												✓				✓											
	V3	✓	✓			✓	✓											✓	✓			✓	✓			✓	✓						
	V4	✓																															
	V5	✓																✓															
	V6	✓	✓			✓																											
	V7	✓	✓																														
Halocline stratification index	V1	✓	✓			✓	✓																										
	V2	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓										
	V3	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓	✓		✓	✓						
	V4	✓	✓	✓		✓																											
	V5	✓	✓	✓		✓	✓																										
	V6	✓	✓			✓	✓	✓																									
	V7	✓	✓			✓	✓																										

Table 15
Table with the input features associated with each test block in the Redondela C zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓	✓		✓	✓											✓	✓			✓	✓										
Daylight hours	-	✓				✓												✓	✓			✓											
Insolation	-	✓																															
Irradiation	-	✓	✓			✓	✓																										
Upwelling index	-	✓	✓															✓	✓														
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓															✓															
	V2	✓	✓															✓	✓														
	V3	✓																✓	✓														
	V4	✓	✓	✓														✓	✓	✓		✓	✓										
	V5	✓																✓															
	V6	✓	✓															✓	✓														
	V7	✓																✓															
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V3	✓				✓	✓			✓	✓			✓				✓	✓	✓		✓	✓			✓	✓			✓			
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
	V5	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ammonium	V1	✓	✓			✓	✓											✓	✓	✓		✓	✓			✓							
	V2	✓				✓																											
	V3	✓	✓															✓				✓											
	V4	✓	✓															✓	✓			✓	✓										
	V5	✓	✓			✓	✓											✓	✓			✓	✓										
	V6	✓	✓			✓												✓	✓	✓		✓											
	V7	✓	✓	✓														✓	✓	✓													
Phosphate	V1	✓	✓	✓		✓				✓								✓	✓			✓				✓							
	V2	✓	✓			✓	✓	✓		✓	✓							✓	✓	✓		✓	✓	✓	✓	✓	✓						
	V3	✓																✓	✓														
	V4	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V5	✓	✓			✓				✓								✓	✓			✓	✓			✓							
	V6	✓	✓			✓				✓								✓	✓	✓		✓	✓			✓	✓						
	V7	✓	✓			✓	✓			✓								✓	✓			✓	✓	✓		✓							
Nitrate	V1	✓	✓	✓		✓	✓			✓	✓			✓																			
	V2	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V5	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V7	✓	✓			✓	✓	✓		✓																							
Nitrite	V1	✓	✓			✓	✓	✓		✓	✓			✓																			
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																		
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V5	✓	✓	✓		✓	✓	✓		✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓																			
	V7	✓	✓	✓		✓	✓	✓		✓	✓																						
Average temperature	V1	✓	✓			✓	✓	✓		✓								✓	✓	✓		✓				✓	✓						
	V2	✓	✓			✓												✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓	✓		✓	✓			✓	✓																						
	V4	✓	✓			✓	✓			✓																							
	V5	✓	✓			✓												✓	✓			✓	✓			✓							
	V6	✓	✓			✓	✓																										
	V7	✓	✓			✓																											
Thermocline stratification index	V1	✓																✓															
	V2	✓																✓				✓											
	V3	✓				✓												✓				✓											
	V4	✓																✓															
	V5	✓	✓	✓														✓	✓														
	V6	✓	✓															✓				✓											
	V7	✓																✓															
Average dissolved oxygen	V1	✓	✓			✓	✓																										
	V2	✓				✓	✓											✓	✓			✓	✓										
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V4	✓	✓			✓	✓																										
	V5	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓									
	V6	✓	✓	✓		✓	✓																										
	V7	✓	✓			✓																											
Halocline stratification index	V1	✓	✓	✓		✓	✓	✓		✓	✓	✓																					
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V3	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓		
	V4	✓	✓	✓	✓	✓	✓			✓	✓	✓																					
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V6	✓	✓	✓		✓	✓	✓		✓	✓	✓																					
	V7	✓	✓	✓		✓	✓	✓		✓	✓	✓																					

Table 16
Table with the input features associated with each test block in the Redondela D zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓	✓		✓												✓				✓	✓										
Daylight hours	-	✓				✓	✓											✓				✓											
Insolation	-	✓																															
Irradiation	-	✓	✓	✓		✓				✓	✓																						
Upwelling index	-	✓	✓			✓	✓											✓	✓			✓	✓										
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓	✓															✓															
	V2	✓	✓															✓	✓														
	V3	✓	✓															✓	✓	✓													
	V4	✓	✓	✓	✓	✓	✓	✓										✓	✓	✓		✓	✓										
	V5	✓	✓															✓															
	V6	✓	✓															✓	✓														
	V7	✓	✓															✓	✓														
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V5	✓	✓	✓	✓	✓	✓			✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓						
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V7	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓			✓	✓			✓	✓	✓		✓	✓			✓			
Ammonium	V1	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓				✓	✓						
	V2	✓				✓																											
	V3	✓	✓			✓												✓	✓	✓		✓	✓										
	V4	✓	✓			✓	✓											✓				✓											
	V5	✓	✓															✓				✓											
	V6	✓	✓			✓	✓											✓	✓			✓	✓			✓							
	V7	✓	✓															✓	✓														
Phosphate	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓				✓	✓	✓	✓	✓	✓	✓		✓	✓			✓			
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓		
	V3	✓	✓	✓		✓	✓			✓	✓							✓	✓			✓	✓			✓	✓						
	V4	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V5	✓	✓	✓		✓				✓								✓	✓	✓		✓	✓			✓	✓						
	V6	✓	✓	✓		✓	✓	✓		✓	✓							✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V7	✓	✓	✓		✓	✓	✓		✓	✓							✓	✓	✓		✓	✓			✓	✓						
Nitrate	V1	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓																		
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V5	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V7	✓	✓	✓	✓	✓	✓	✓		✓																							
Nitrite	V1	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V2	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																				
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																
	V5	✓	✓	✓		✓	✓			✓	✓			✓				✓	✓	✓		✓	✓	✓		✓	✓						
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓																			
	V7	✓	✓	✓		✓	✓			✓	✓																						
Average temperature	V1	✓				✓				✓								✓	✓			✓				✓							
	V2	✓	✓			✓	✓			✓								✓	✓			✓	✓			✓							
	V3	✓	✓			✓	✓			✓																							
	V4	✓	✓			✓				✓																							
	V5	✓	✓			✓												✓	✓	✓		✓	✓			✓							
	V6	✓				✓				✓																							
	V7	✓	✓			✓				✓																							
Thermocline stratification index	V1	✓																✓															
	V2	✓																✓															
	V3	✓																✓				✓											
	V4	✓	✓															✓															
	V5	✓	✓	✓														✓	✓	✓													
	V6	✓	✓	✓														✓															
	V7	✓																✓															
Average dissolved oxygen	V1	✓	✓	✓		✓	✓																										
	V2	✓	✓			✓	✓											✓	✓	✓		✓	✓										
	V3	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓		✓	✓	✓					
	V4	✓	✓			✓	✓																										
	V5	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓										
	V6	✓	✓			✓	✓																										
	V7	✓	✓	✓		✓	✓																										
Halocline stratification index	V1	✓																															
	V2	✓	✓			✓	✓											✓	✓			✓	✓										
	V3	✓	✓			✓												✓	✓	✓		✓	✓										
	V4	✓	✓			✓	✓																										
	V5	✓	✓																														
	V6	✓	✓																														
	V7	✓	✓			✓	✓																										

Table 17
Table with the input features associated with each test block in the Redondela E zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓				✓												✓				✓											
Daylight hours	-	✓	✓			✓	✓											✓	✓	✓		✓	✓										
Insolation	-	✓	✓																														
Irradiation	-	✓	✓	✓		✓	✓			✓	✓																						
Upwelling index	-	✓	✓			✓	✓	✓										✓				✓											
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓	✓														
	V2	✓	✓															✓															
	V3	✓	✓	✓														✓	✓	✓													
	V4	✓	✓	✓														✓	✓	✓													
	V5	✓	✓															✓	✓														
	V6	✓	✓															✓															
	V7	✓	✓															✓															
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
Ammonium	V1	✓	✓			✓				✓								✓	✓	✓		✓	✓			✓							
	V2	✓				✓	✓																										
	V3	✓	✓															✓	✓	✓		✓											
	V4	✓	✓															✓	✓			✓	✓										
	V5	✓																✓				✓											
	V6	✓				✓	✓											✓				✓	✓										
	V7	✓	✓															✓	✓														
Phosphate	V1	✓	✓	✓		✓				✓								✓	✓	✓		✓	✓	✓		✓	✓						
	V2	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓		✓	✓	✓		✓	✓						
	V3	✓																✓	✓														
	V4	✓	✓			✓				✓								✓	✓			✓	✓			✓							
	V5	✓	✓			✓				✓	✓							✓	✓			✓	✓			✓							
	V6	✓	✓	✓		✓	✓			✓	✓							✓	✓	✓	✓	✓	✓	✓		✓	✓						
	V7	✓	✓	✓		✓	✓			✓								✓	✓			✓	✓			✓							
Nitrate	V1	✓	✓			✓	✓	✓		✓	✓																						
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓	✓																	
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V5	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓																		
Nitrite	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓																			
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																				
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																	
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																	
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
Average temperature	V1	✓	✓			✓	✓			✓	✓							✓	✓	✓		✓	✓			✓	✓						
	V2	✓	✓			✓	✓			✓								✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓	✓		✓				✓																							
	V4	✓	✓			✓	✓			✓																							
	V5	✓				✓												✓	✓			✓				✓	✓						
	V6	✓				✓				✓																							
	V7	✓	✓			✓				✓																							
Thermocline stratification index	V1	✓																✓				✓											
	V2	✓	✓															✓	✓			✓	✓										
	V3	✓	✓			✓												✓				✓											
	V4	✓	✓															✓															
	V5	✓	✓															✓	✓														
	V6	✓	✓															✓															
	V7	✓	✓															✓															
Average dissolved oxygen	V1	✓	✓			✓	✓			✓	✓																						
	V2	✓	✓			✓	✓											✓	✓			✓	✓										
	V3	✓	✓	✓		✓	✓			✓	✓	✓						✓	✓	✓		✓	✓			✓	✓						
	V4	✓	✓			✓	✓																										
	V5	✓	✓			✓												✓	✓			✓	✓										
	V6	✓	✓			✓	✓																										
	V7	✓	✓	✓		✓	✓																										
Halocline stratification index	V1	✓	✓			✓	✓	✓																									
	V2	✓	✓	✓		✓	✓											✓	✓	✓		✓	✓	✓		✓	✓	✓					
	V3	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓		✓	✓						
	V4	✓	✓	✓		✓	✓																										
	V5	✓	✓	✓		✓	✓																										
	V6	✓	✓			✓	✓																										
	V7	✓	✓			✓	✓																										

Table 18
Table with the input features associated with each test block in the Vigo A zone. The check marks when the feature was used.

Feature	Oceanographic station	Approach 1																Approach 2															
		Correlation Quartile																															
		0				25				50				75				0				25				50				75			
		Quartile of random forest discriminator																															
		0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75	0	25	50	75
Week of the year	-	✓	✓															✓	✓														
Daylight hours	-	✓				✓												✓	✓			✓											
Insolation	-	✓																															
Irradiation	-	✓				✓	✓																										
Upwelling index	-	✓	✓															✓				✓											
Zone state	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Maximum chlorophyll	V1	✓																✓	✓														
	V2	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓									
	V3	✓	✓															✓	✓	✓		✓	✓										
	V4	✓	✓	✓		✓	✓	✓										✓	✓	✓		✓	✓	✓									
	V5	✓																✓															
	V6	✓																✓															
	V7	✓	✓	✓														✓	✓			✓	✓	✓									
D. acuminata concentration	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
	V3	✓	✓	✓		✓	✓			✓				✓				✓	✓	✓		✓	✓			✓				✓			
	V4	✓	✓	✓		✓	✓	✓		✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓		✓			
	V5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	V6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Ammonium	V1	✓																✓															
	V2	✓				✓																											
	V3	✓	✓															✓	✓														
	V4	✓																✓	✓														
	V5	✓	✓			✓												✓				✓											
	V6	✓																✓															
	V7	✓	✓			✓	✓											✓	✓			✓											
Phosphate	V1	✓	✓	✓		✓				✓								✓				✓	✓			✓							
	V2	✓	✓	✓		✓				✓								✓	✓	✓		✓	✓			✓							
	V3	✓	✓															✓	✓			✓											
	V4	✓	✓			✓	✓											✓				✓				✓							
	V5	✓	✓			✓	✓											✓				✓	✓			✓	✓						
	V6	✓	✓			✓				✓								✓				✓				✓							
	V7	✓	✓	✓		✓	✓	✓										✓	✓			✓				✓							
Nitrate	V1	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
	V2	✓	✓	✓		✓	✓	✓		✓	✓			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		
	V3	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓																		
	V4	✓	✓			✓	✓			✓	✓			✓																			
	V5	✓	✓			✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓		
	V6	✓	✓			✓	✓	✓		✓	✓			✓	✓																		
	V7	✓	✓	✓		✓	✓	✓		✓	✓			✓																			
Nitrite	V1	✓	✓			✓	✓			✓	✓																						
	V2	✓	✓	✓		✓	✓	✓		✓	✓			✓	✓			✓	✓	✓	✓	✓	✓			✓	✓	✓					
	V3	✓	✓	✓		✓	✓			✓	✓																						
	V4	✓	✓	✓		✓	✓			✓	✓																						
	V5	✓	✓			✓				✓								✓	✓			✓				✓							
	V6	✓	✓			✓				✓																							
	V7	✓	✓			✓	✓			✓																							
Average temperature	V1	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	V2	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓																
	V4	✓	✓	✓		✓	✓			✓	✓																						
	V5	✓				✓												✓	✓			✓				✓	✓						
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓																						
	V7	✓	✓	✓		✓	✓	✓		✓	✓																						
Thermocline stratification index	V1	✓																✓	✓														
	V2	✓	✓			✓	✓											✓	✓			✓	✓										
	V3	✓	✓															✓	✓			✓	✓										
	V4	✓	✓															✓															
	V5	✓	✓															✓	✓														
	V6	✓	✓	✓														✓	✓	✓													
	V7	✓	✓															✓															
Average dissolved oxygen	V1	✓	✓			✓																											
	V2	✓	✓			✓	✓											✓	✓	✓		✓	✓			✓	✓						
	V3	✓	✓			✓												✓	✓			✓	✓			✓	✓						
	V4	✓	✓			✓	✓																										
	V5	✓	✓			✓	✓											✓	✓	✓		✓	✓	✓									
	V6	✓				✓																											
	V7	✓	✓	✓		✓	✓	✓																									
Halocline stratification index	V1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
	V2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
	V3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				
	V4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					
	V5	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V6	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓																					
	V7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓																					

References

(2019) Commission implementing regulation (eu) 2019/627 of 15 March 2019 laying down uniform practical arrangements for the performance of official controls on products of animal origin intended for human consumption in accordance with regulation (eu) 2017/625 of the european parliament and of the council and amending commission regulation (ec) no 2074/2005 as regards official controls. URL: http://data.europa.eu/eli/reg_impl/2019/627/2021-01-01.

(2021) Web page of meteogalicia. URL: https://www.meteogalicia.gal/observacion/estacionshistorico/historico.action?idEst=14001.

Aguilar Calderon, V.H., 2017. Predicción de las floraciones algales nocivas (fan) en poblaciones de dinophysis acuminata por redes neuronales artificiales.

Avdelas, L., Avdic-Mravlje, E., Borges Marques, A.C., Cano, S., Capelle, J.J., Carvalho, N., Cozzolino, M., Dennis, J., Ellis, T., Fernandez Polanco, J.M., et al., 2021. The decline of mussel aquaculture in the european union: causes, economic impacts and opportunities. Rev. Aquacult. 13, 91–118.

Behera, R.N., Roy, M., Dash, S., 2016. Ensemble based hybrid machine learning approach for sentiment classification-a review. Int. J. Comput. Appl. 146, 31–36.

Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F., Mahmood, F., 2021. Synthetic data in machine learning for medicine and healthcare. Nat. Biomed. Eng. 5, 493–497.

Copenhaver, M.D., Holland, B., 1988. Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. J. Stat. Comput. Simul. 30, 1–15. https://doi.org/10.1080/00949658808811082.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20. https://doi.org/10.1023/A:1022627411411.

Cruz, R.C., Reis Costa, P., Vinga, S., Krippahl, L., Lopes, M.B., 2021. A review of recent machine learning advances for forecasting harmful algal blooms and shellfish contamination. J. Mar. Sci. Eng. 9. https://doi.org/10.3390/jmse9030283.

Davis, T.W., Berry, D.L., Boyer, G.L., Gobler, C.J., 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of microcystis during cyanobacteria blooms. Harmful Algae 8, 8. https://doi.org/10.1016/j.hal.2009.02.004.

Deng, T., Chau, K.-W., Duan, H.-F., 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. J. Environ. Manage. 284, 112051. https://doi.org/10.1016/j.jenvman.2021.112051. URL: https://www.sciencedirect.com/science/article/pii/S0301479721001134.

FAO (2 February 2022). Food and agriculture organization. URL: https://www.fao.org/in-action/globefish/market-reports/resource-detail/ru/c/1199390/.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Annals Stat. 29, 1189–1232. https://doi.org/10.2307/2699986. URL: http://www.jstor.org/stable/2699986.

Gasinaite, Z.R., Cardoso, A.C., Heiskanen, A.S., Henriksen, P., Kauppila, P., Olenina, I., Pilkaityte, R., Purina, I., Razinkovas, A., Sagert, S., Schubert, H., Wasmund, N., 2005. Seasonality of coastal phytoplankton in the baltic sea: Influence of salinity and eutrophication. Estuar. Coast. Shelf Sci. 65. https://doi.org/10.1016/j.ecss.2005.05.018.

Gholami, Z., Mortazavi, M.S., Karbassi, A., 2019. Environmental risk assessment of harmful algal blooms case study: Persian gulf and oman sea located at hormozgan province, Iran. Human Ecol. Risk Assess.: An Int. J. 25, 271–296. https://doi.org/10.1080/10807039.2018.1501660. URL: https://doi.org/10.1080/10807039.2018.1501660. arXiv:https://doi.org/10.1080/10807039.2018.1501660.

Guallar, C., Delgado, M., Diogène, J., Fernández-Tejedor, M., 2016. Artificial neural network approach to population dynamics of harmful algal blooms in alfacs bay (nw mediterranean): Case studies of karlodinium and pseudo-nitzschia. Ecol. Model. 338, 271–296. https://doi.org/10.1016/j.ecolmodel.2016.07.009.

Hill, P.R., Kumar, A., Temimi, M., Bull, D.R., 2020. Habnet: Machine learning, remote sensing-based detection of harmful algal blooms. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 13, 13. https://doi.org/10.1109/JSTARS.2020.3001445.

IEO (April 27, 2021). Web page of marnaraia proyect. URL: http://www.indicedeafloramiento.ieo.es/afloramiento.html.

INTECMAR (2 February 2022). Historical status of cultivation areas. URL: http://www.intecmar.gal/Informacion/biotoxinas/EstadoZonas/Historico_Batea.aspx.

Jin, D., Hoagland, P., 2008. The value of harmful algal bloom predictions to the nearshore commercial shellfish fishery in the gulf of maine. Harmful Algae 7. https://doi.org/10.1016/j.hal.2008.03.002.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33. https://doi.org/10.2307/2529310.

Lantz, B., 2015. Machine Learning with R: Second Edition.

Lee, S., Lee, D., 2018. Improved prediction of harmful algal blooms in four major south korea's rivers using deep learning models. International Journal of Environmental Research and Public Health, 15. URL: https://www.mdpi.com/1660-4601/15/7/1322. doi:10.3390/ijerph15071322.

Lewis, D.D. (1998). Naive (bayes) at forty: The independence assumption in information retrieval. In: Nédellec, C., Rouveirol, C. (Eds.), Machine Learning: ECML-98 (pp. 4–15). Berlin, Heidelberg: Springer, Berlin Heidelberg. doi:10.1007/BFb0026666.

Li, F., Zhang, H., Zhu, Y., Xiao, Y., Chen, L., 2013. Effect of flow velocity on phytoplankton biomass and composition in a freshwater lake. Sci. Total Environ. 447. https://doi.org/10.1016/j.scitotenv.2012.12.066.

Liu, J., Zhang, Y., Qian, X., 2009. Modeling chlorophyll-a in taihu lake with machine learning models. doi:10.1109/ICBBE.2009.5163072.

Molares, A., Fernandez-Blanco, E., Rivero, D., 2020. Application of artificial neural networks for the monitoring of episodes of high toxicity by dsp in mussel production areas in galicia. Proceedings, 54. doi:10.3390/proceedings2020054012.

Paerl, H.W., Paul, V.J., 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. 46. https://doi.org/10.1016/j.watres.2011.08.002.

Rahman, A., Shahriar, M.S., 2013. Algae growth prediction through identification of influential environmental variables: A machine learning approach. Int. J. Comput. Intell. Appl. 12. https://doi.org/10.1142/S1469026813500089.

Segal, M.R., 2004. Machine learning benchmarks and random forest regression,.

Sheskin, D.J., 2003. Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC. doi:10.1201/9781420036268.

Velo-Suárez, L., Gutiérrez-Estrada, J.C., 2007. Artificial neural network approaches to one-step weekly prediction of dinophysis acuminata blooms in huelva (western andalucía, spain). Harmful Algae 6. https://doi.org/10.1016/j.hal.2006.11.002.

Vilas, F., Rey, D., Armesto, B.R., Bernabéu, A., Méndez, G., Durán, R., Mohamed, K., Rosón, G., Cabanas, J.M., Pérez, F.F., Castro, C.G., Ríos, A.F., Figueiras, F.G., Miranda, A., Riveiro, I., Vergara, A.R., Guisande, C., Reguera, B., Escalera, L., Pazos, Y., Ángeles Moroño, González, J.J., Álvarez, C., Beiras, R., Besada, V., Fumega, J., Ángeles Franco, M., Gómez, M., Quijano, A.G., Nunes, T., Prego, R., Sanz, A.S., Viñas, L., Peleteiro, J.B., Trujillo, V., Bañón, R., Ribó, J., Olmedo, M., Álvarez Blázquez, B., Rodríguez, J.L., Pazó, J., Otero, J.J., Ángel Guerra, Lens, S., Rocha, F., Rodríguez, M.X.V., Blanco, A.P., 2008. La ría de vigo: una aproximación integral al ecosistema marino de la ría de vigo, URL: http://hdl.handle.net/10261/170032.

Vilas, L.G., Spyrakos, E., Palenzuela, J.M.T., Pazos, Y., 2014. Support vector machine-based method for predicting pseudo-nitzschia spp. blooms in coastal waters (galician rias, nw spain). Prog. Oceanogr. 124. https://doi.org/10.1016/j.pocean.2014.03.003.

White, H., et al., 1992. Artificial neural networks. Blackwell Cambridge, Mass.

Yñiguez, A.T., Ottong, Z.J., 2020. Predicting fish kills and toxic blooms in an intensive mariculture site in the Philippines using a machine learning model. Sci. Total Environ. 707, 136173. https://doi.org/10.1016/j.scitotenv.2019.136173. URL: https://www.sciencedirect.com/science/article/pii/S0048969719361698.

Yu, P., Gao, R., Zhang, D., Liu, Z.-P., 2021. Predicting coastal algal blooms with environmental factors by machine learning methods. Ecol. Ind. 123, 107334. https://doi.org/10.1016/j.ecolind.2020.107334. URL: https://www.sciencedirect.com/science/article/pii/S1470160X20312760.