proceedings

MDPI

Proceedings

Application of Artificial Neural Networks for the Monitoring of Episodes of High Toxicity by DSP in Mussel Production Areas in Galicia †

Andrés Molares *, Enrique Fernandez-Blanco [ID] **and Daniel Rivero** [ID]

CITIC, Faculty of Computer Science, Campus de Elviña, University of A Coruña, 15006 A Coruña, Spain; enrique.fernandez@udc.es (E.F.-B.); daniel.rivero@udc.es (D.R.)

* Correspondence: andres.molares@udc.es

† Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Abstract: This study seeks to support, through the use of Artificial Neural Networks (ANN), the decision to perform closings after days without sampling in the Vigo estuary. The opening and closing of the mussel production areas are based on the toxicity analysis of this bivalve's meat. Sometimes it is not possible to obtain the necessary data for effective closing. If there is evidence of an increase in toxicity levels, "Precautionary Closings" on mussel extraction is done. A small error in the forecast of the state of the areas could mean serious losses for the mussel industry and a huge risk for public health. Unlike in previous studies, this study aims to manage the state of the mussel production areas, whilst the others focused on predicting the harmful algae blooms. Having achieved test sensitivity values of 67.40% and test accuracy of 83.00%, these results may lead to new research that involves obtaining more accurate models that can be integrated into a support system.

Keywords: machine learning; harmful algae blooms; artificial neural networks

1. Introduction

Since 1995, a governmental monitoring program has managed the mussel production areas in Galicia. The creation of this program was necessary because of the high frequency of a phenomenon called Harmful Algal Blooms (HAB), which implies a temporary cessation in the extraction and commercialization of the mussels. The HAB are episodes of a high concentration of algae potentially toxic to humans through mussel consumption. In the Vigo estuary, the most common toxin-producing species are the DSP type, such as the *Dinophysis acuminata* dinoflagellate [1].

A weak point of this process is the absence of sampling during weekends or inclement weather, which sometimes makes it impossible to collect the data to support an effective closing. If there is an indication of an increase in levels of toxicity, the competent authority is legally empowered to proceed to the "Precautionary Closings" on the extraction of bivalve molluscs. Nowadays, the performance of this kind of closing is based on the expertise of government agents. A mathematical model to support the making of these decisions could help experts in complex situations that may cause errors in the decisions made.

Although the previously described situation is focused on the Galician Coast, this scenario is replicated in other major producers around the globe. That is why other works have tried to monitor the HAB episodes using different techniques. To date, those previous works have focused their efforts on predicting biomarkers, such as the concentration of toxic phytoplankton or chlorophyll "a". These studies, although of high scientific interest, do not give concrete support when it comes to monitoring the state of the production areas. The toxicity levels present in mussel meat depend on additional factors, such as retention of toxicity or the relationship between toxic versus non-toxic phytoplankton

present in the medium. These factors, and some others, will be considered in conducting this study to achieve a more practical approach. To do that, a classifier based on Artificial Neural Networks (ANN) [2] is going to be defined to assess the state of the production areas affected by DSP-type toxins, on days with the absence of previous samplings.

2. Results

A summary of the obtained results can be seen in Figure 1. Each row of this table represents a tested model, where the first and second columns define the filters of characteristics used. These filters are represented with the value of the quartile selected for training. The third column shows the architecture of the networks by showing the number of hidden neurons per layer. The fourth column contains the p-value obtained from a Tukey–Kramer paired analysis, after previously performing a ANOVA analysis. Additionally, the remaining columns show the performance measures obtained in the test, that is, the average accuracy, average sensitivity, average kappa coefficient, minimum accuracy, minimum sensitivity, and minimum kappa coefficient.

Table ordered by sensibility									
Correlation filter quartile	Random Forest filter quartile	Neurons per layer	p	Accuracy $\bar{x}$	Sensibility $\bar{x}$	Kappa index $\bar{x}$	Accuracy Min	Sensibility Min	Kappa index Min
-	-	14		83.00%	67.40%	0.591	72.00%	36.00%	0.31
-	-	[10, 20]	1	83.50%	66.40%	0.599	72.00%	32.00%	0.29
-	-	[10, 10]	1	83.30%	66.20%	0.597	71.00%	30.00%	0.27
-	-	8	1	82.50%	63.30%	0.566	71.00%	30.00%	0.26
25	-	14	1	79.30%	60.00%	0.508	63.00%	27.00%	0.18

Figure 1. Statistics of the best models applied in the Vigo. A production area, ordered by descending sensibility. Access to the full table can be obtained here: https://github.com/AndresMolares/XoveTIC2020.git.

3. Discussion

After carrying out the study, it can be seen how ANN works better with a large number of characteristics to solve this problem. Although the works carried out to date obtain good results when making HAB predictions on the Galician coast (an overall accuracy between 78.53–82.18% using vector support machines to predict HAB of *Pseudo-nitzschia spp.* [3]), the control of the state of the production areas is conditioned by other external factors, so the definition of the problem changes. As this is the first study that seeks to provide support when estimating the state of the mussel production areas affected by DSP-type toxins, the results are promising (accuracy of 83%). However, to develop models that are precise enough to be integrated into support tools, it would be necessary to develop models with better sensitivity and accuracy values. For this, new machine learning algorithms could be studied, as well as a more exhaustive exploration of the hyperparameter space of the ANN.

4. Materials and Methods

Data from different sources were combined to create the dataset used in this experiment. Those sources contained different values sampled weekly between 2004 and 2018. The result is a dataset with the following variables: seasonality, concentration of chlorophyll "a", *Dinophysis acuminata*, ammonium, phosphate, nitrite, nitrate, water temperature, oxygen in water, salinity, solar irradiation, upwelling index, and the previous state of the production area. These data have been provided by the INTECMAR [4], METEOGALICIA [5], and IEO [6].

From raw data, two types of filtering were applied to choose the most significant features: (a) Applying a correlation matrix of the input variables with the state of the zone (variable objective); and (b) using a Random Forest algorithm as a discriminator. The Random Forest algorithm calculates the importance of a variable, taking into account how much the prediction error increases when the data for that variable is permuted, while all others remain unchanged.

With these methods, blocks of characteristics of quartiles 25, 50, and 75 were obtained. Different experiments have been defined based on the application of each one, another, both, or none of the previously mentioned filtering methods and the architecture of the ANN model. To ensure reliable results, the tests were performed with a 10-fold cross-validation strategy, which was repeated 50 times for each combination of filter methods and classification techniques. The reason for repeating this process is due to the non-deterministic nature of the backpropagation [7] used to train the ANN of each fold. To perform the training, each ANN model was set to use Dense Hidden layers, with the ADAM algorithm as an optimizer and binary cross-entropy as the loss function. Finally, the transfer function of the output layer is a sigmoid function in all cases exposed, while on the contrary, the activation function of the hidden layers is a Relu function.

With that configuration and data in common and having the same input and desired output, five models were tested by changing the number of hidden layers and the number of elements in these. More specifically, the trained and tested models were: One hidden layer with 2, 8, or 14 neurons; and two hidden layers with 10 neurons each and 10 and 20 neurons, respectively.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANN   Artificial Neural Network
DSP   Diarrheal Shellfish Poisoning

References

1. Vilas, F.; Rey, D.; Rubio Armesto, B.; Bernabéu, A.; Méndez, G.; Durán, R.; Mohamed, K.; Rosón, G.; Cabanas, J.; Pérez, F.F.; et al. *La Ría de Vigo: una aproximación integral al ecosistema marino de la Ría de Vigo*; Instituto de Estudios Vigueses: Vigo, Spian, 2008.
2. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133.
3. Vilas, L.G.; Spyrakos, E.; Palenzuela, J.M.T.; Pazos, Y. Support Vector Machine-based method for predicting Pseudo-nitzschia spp. blooms in coastal waters (Galician rias, NW Spain). *Prog. Oceanogr.* **2014**, *124*, 66–77.
4. Instituto Tecnolóxico Para o Control do Medio Mariño de Galicia web page. Available online: http://www.intecmar.gal/intecmar/default.aspx (accessed on 17 January 2020).
5. METEOGALICIA Web Page. Available online: https://www.meteogalicia.gal/observacion/estacionshistorico/historico.action?idEst=14001 (accessed on 17 February 2020).
6. Marnaraia. Available online: http://www.indicedeafloramiento.ieo.es/afloramiento.html (accessed on 17 January 2020).

7. Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Neural Networks for Perception*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 65–93.

8. CESGA Web Page. Available online: https://www.cesga.es/ (accessed on 20 February 2020).