

## NRC Publications Archive Archives des publications du CNRC

### Paralytic shellfish poisoning risk assessment in the west coast of Canada

Bi, Chang; Pan, Youlian; Zhang, Xuekui

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. /  
La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version  
acceptée du manuscrit ou la version de l'éditeur.

For the publisher's version, please access the DOI link below. / Pour consulter la version de l'éditeur, utilisez le lien  
DOI ci-dessous.

#### **Publisher's version / Version de l'éditeur:**

<https://doi.org/10.1016/j.jhazmat.2025.140459>

*Journal of Hazardous Materials*, 500, C, pp. 1-9, 2025-11-12

#### **NRC Publications Archive Record / Notice des Archives des publications du CNRC :**

<https://nrc-publications.canada.ca/eng/view/object/?id=ff4a2ee5-01b6-45d8-aaa0-50b6d976445d>

<https://publications-cnrc.canada.ca/fra/voir/objet/?id=ff4a2ee5-01b6-45d8-aaa0-50b6d976445d>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.



**Questions?** Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the  
first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez  
la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous  
n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



# Paralytic shellfish poisoning risk assessment in the west coast of Canada

Chang Bi<sup>a</sup>, Youlian Pan<sup>a,b,\*</sup> , Xuekui Zhang<sup>a,\*\*</sup> 

<sup>a</sup> University of Victoria, 3800 Finnerty Road, Victoria, BC V8W 2Y2, Canada

<sup>b</sup> Digital Technologies Research Centre, National Research Council Canada, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada

## HIGHLIGHTS

- A novel autoregressive framework forecasts PSP risk from toxin history.
- Stacked ensemble of 11 models achieves high predictive accuracy (AUC  $\approx$  0.95).
- Multivariate toxin profiles significantly outperform total PST for risk prediction.
- The framework is robust when using harmonized legacy (bioassay) monitoring data.
- SHAP analysis identifies key toxins (NEOSTX, C-2) driving regional PSP risk.

## ARTICLE INFO

### Keywords:

Marine biotoxins  
Environmental risk assessment  
Predictive modeling  
*Mytilus edulis*  
Coastal ecosystem health

## ABSTRACT

Paralytic Shellfish Poisonings (PSPs) are a major public health concern requiring robust environmental monitoring. We developed and validated a machine learning framework to assess PSP risk in blue mussels (*Mytilus edulis*) along Canada's west coast. Our study comprised three experiments that tested the ability of 11 models to forecast risk using historical toxin data (2000-2020). Results showed that lower detection thresholds and the use of multivariate toxin profiles significantly improved predictive accuracy. Tree-based algorithms, in particular, excelled with this detailed data. A stacked ensemble model consistently matched the best individual model's performance, achieving an AUC (area under receiver operating characteristic curve) over 0.912 across all experiments and offering a robust solution for operational forecasting. Model interpretation revealed that recent toxin history and specific compounds like Neosaxitoxin (NEOSTX) and N-sulfocarbamoyl gonyautoxin-3 (C-2) were the most important predictors, aligning with regional ecological dynamics. This framework provides a powerful, data-driven tool for enhancing early warning capabilities and supporting proactive risk management.

## 1. Introduction

Harmful algal blooms (HABs) pose a significant and increasing threat to water quality and public health worldwide [1,2]. Among the various toxins produced by these blooms, Paralytic Shellfish Toxins (PSTs) are of particular concern due to their potential to accumulate in shellfish and cause serious illness in humans [3]. Ingestion of contaminated shellfish can lead to Paralytic Shellfish Poisoning (PSP), a potentially fatal illness characterized by neurological symptoms such as tingling, numbness, paralysis, and respiratory failure [4]. In the Pacific Northwest, these blooms are often dominated by *Alexandrium catenella* and cause PSTs to accumulate in filter-feeding shellfish. This poses risks to human consumers and impacts local economies reliant on shellfish harvesting and

aquaculture [5]. In Canada, closures to shellfish harvesting are initiated when PST levels exceed the regulatory limit ( $80 \mu\text{g STX.eq } 100 \text{ g}^{-1}$ ) [6]. Current monitoring approaches primarily involve laboratory analyses (e.g. HPLC and mouse bioassay), which are time-consuming and costly. This can lead to delays to effective public health protection, underscoring the value of developing proactive environmental management strategies.

In response to such challenges across various environmental domains, machine learning has emerged as a powerful tool for predictive analysis and risk assessment of hazardous materials. Its applications are diverse, ranging from identifying spatiotemporal patterns and assessing ecological risks associated with heavy metal contamination in water bodies and soils [7,8] to predicting the bioaccessibility of contaminants

\* Correspondence to: National Research Council Canada, 1200 Montreal Road, Ottawa ON K1A 0R6, Canada.

\*\* Corresponding author.

E-mail addresses: [youlian.pan@nrc-cnrc.gc.ca](mailto:youlian.pan@nrc-cnrc.gc.ca) (Y. Pan), [xuekui@uvic.ca](mailto:xuekui@uvic.ca) (X. Zhang).

<https://doi.org/10.1016/j.jhazmat.2025.140459>

Received 5 July 2025; Received in revised form 7 November 2025; Accepted 10 November 2025

Available online 12 November 2025

0304-3894/Crown Copyright © 2025 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

like polycyclic aromatic hydrocarbons in food [9]. Furthermore, machine learning models are increasingly utilized for quantitative prediction in chemical risk assessment, offering new approaches for assessing toxicological impacts and broader environmental hazards [10]. Such data-driven approaches collectively enhance environmental monitoring and facilitate more accurate health and ecological risk assessments.

For instance, previous studies have investigated deep learning models based on chemical fingerprints [11] and models incorporating environmental variables such as water temperature and salinity [12,13]. Random forest and boosting models have also been applied to PSP and HAB forecasting, including predicting PST exceedances in Southeast Alaska from hydrographic and meteorological data [14] and forecasting *Alexandrium minutum* blooms in the Adriatic Sea [15]. Nevertheless, the environment-driven models often depend on frequent, high-quality environmental observations (e.g. nutrient levels and chlorophyll-*a*), which are frequently sparse and may be missing at key time points—this data sparsity has been shown to limit algal bloom prediction models even in well-monitored lake systems [16]. Moreover, complex ecological dynamics and species-specific variability have consistently challenged the generalizability and transferability of environment-driven modeling frameworks [17]. For example, models that rely on incorporating environmental factors to predict PSP risk face difficulties. The complex interplay between environmental factors, inherent shellfish physiology, and the intricacies of toxicokinetics can vary significantly across different locations and times [18].

Filter-feeding mussels act as natural integrators of environmental exposures, continuously accumulating toxins in their tissues. This property makes their toxin histories a stable predictor of future PSP risk, compared to sporadic environmental measurements. The importance of mussel-based monitoring is recognized across the globe. For example, NOAA's long-running Mussel Watch Program has demonstrated the effectiveness of mussels as sentinel organisms for contaminant surveillance over decades [19]. In Canada, the Canadian Shellfish Sanitation Program (CSSP), jointly administered by Canadian Food Inspection Agency (CFIA), Department of Fisheries and Oceans (DFO), and Environment and Climate Change of Canada (ECCC), relies on routine mussel biotoxin testing to guide closures to shellfish harvest and protect public health [20]. These precedents provide strong justification for using toxin-history data as predictors in autoregressive forecasting models.

The critical knowledge gap, therefore, is that while environment-driven models have been extensively tested, few studies have systematically evaluated whether toxin histories themselves, which are already routinely collected, can provide a robust basis for forecasting. This study addresses this gap by developing and validating an assessment framework that uses historical toxin concentrations and multivariate toxin profiles in sentinel mussels (*Mytilus edulis*) to predict future risk, thereby systematically evaluating an autoregressive toxin-history approach. The primary objective of our approach is to provide an early warning capability, which could significantly improve shellfish safety management by allowing for more timely and targeted interventions. This study benchmarks 11 diverse models, including tree ensembles, deep learning, and statistical baselines, and demonstrates that autoregressive toxin-history forecasting is both feasible and operationally robust.

Given the temporal dynamics often observed in toxin accumulation and depuration in shellfish, recent historical data—including toxin concentrations and compound profiles from preceding days—are inherently strong predictors of the near-future risk [21,22]. This study develops and evaluates a framework that leverages this temporal dynamic pattern to forecast PSP events with a defined lead time. The aim is not to replace essential laboratory assays but rather to develop a system that serves as a valuable early warning tool for proactive risk assessment and management. Such a system could potentially optimize the frequency and spatial targeting of resource-intensive sampling and analyses, thereby improving the overall efficiency and responsiveness of shellfish safety monitoring programs and assisting management decisions.

The problem formulation is described in 2.2. A collection of machine learning and statistical models is compared, and an ensemble model is developed and introduced in 2.3. The details about the three experiments are shown in 2.4. The results are presented in 3 and discussed in 4.

## 2. Materials and methods

### 2.1. Data

We acquired 20 years (2000–2020) of PST data on blue mussels from the Canadian Food Inspection Agency (CFIA), including 372 sites on the west coast of Canada. The data distributions were highly uneven and sparse since the data were collected based on the monitoring requirements [21]. The data covered three time periods. First, mouse bioassay was used to measure the PST level during 2000–2010. The measurement was recorded as the total amount of PST with a detection limit of  $40 \mu\text{g } 100 \text{ g}^{-1} \text{ STXdHCl eq}$ . The second period (2011–2015) was featured with a transition; HPLC (High Performance Liquid Chromatography) was gradually introduced to replace mouse bioassay. After 2015, more detailed measurements of individual PST compounds became available with a detection limit of  $1 \mu\text{g } 100 \text{ g}^{-1} \text{ STXdHCl eq}$ . This included N-sulfocarbamoyl-11-hydrosulfate toxins (C-1, C-2), gonyautoxin-1 through -5 (GTX-1, GTX-2, GTX-3, GTX-4, GTX-5), Neosaxitoxin (NEOSTX), Saxitoxin (STX), decarbamoylgonyautoxin-2 and -3 (dcGTX-2, dcGTX-3), decarbamoylsaxitoxin (dcSTX), and a summation of the 12 compounds as the total amount of PST.

### 2.2. Problem formulation

This study aimed to forecast whether the risk of PST level for a site exceed the threshold at the target time, based on its historical data. We formulate this problem as a binary classification based on a time series of PST levels. As indicated in Fig. 1, the history,  $X_{T,s}$ , is defined as a time series with a fixed duration of  $c$  days and occurs  $d$  days (*history distance*) before the target event at time step  $T$ , which is

$$X_{T,s} = \{x_{t,s} : t \in [T-d-c, T-d-1], s \in S\}$$

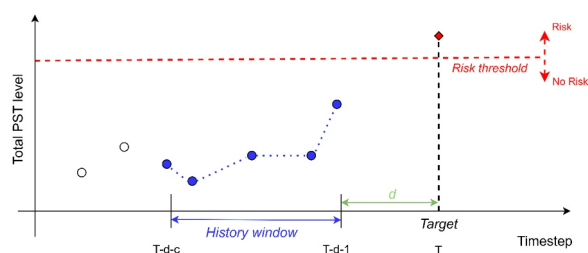
where  $t$  represents the time step in days and  $S$  represents the data collection for a site. The time span between  $T-d-c$ , and  $T-d-1$  is referred to as the *history window*. The target is a binary risk indicator represented by  $Y_{T,s} \in \{0, 1\}$ , where 0 represents no risk and 1 at risk. The objective is to learn a classifier such that

$$Y_{T,s} = f(X_{T,s}) = f(x_{t=T-d-c,s}, \dots, x_{t=T-d-1,s})$$

### 2.3. Method

#### 2.3.1. Data selection and quality control

The raw monitoring data was collected in a highly irregular manner, with time gaps between adjacent data points at a single site ranging from days to months. To ensure the data was suitable for time-series modeling, we established a set of quality control criteria to select



**Fig. 1.** Scheme of problem formulation. The task is to predict whether the total PST value at timestep  $T$  would exceed a predefined risk threshold based on the history time series between timesteps  $T-d-c$  and  $T-d-1$ .

valid data segments. For any given target prediction time  $T$ , the corresponding historical data was included only if it met the following conditions:

- **Maximum gap constraint:** The time between any two consecutive measurements within the history window must be less than 21 days. This threshold, hereafter referred to as *max gap*, was selected based on a sensitivity analysis detailed in (Supplementary File A.1) to ensure an optimal balance between data quality and representation coverage.
- **Signal presence:** At least one measurement within the history window must have a value greater than the minimum value present in the dataset, ensuring that the model learns from meaningful signals rather than null data.
- **Boundary points requirement:** To prevent extrapolation, there must be at least one real measurement within *max gap* days prior to the start of the history window and at least one between the end of the history window and the target date.

This rigorous selection process ensures that every time-series segment used for training and testing is underpinned by sufficiently dense and well-bounded empirical data.

### 2.3.2. Data preprocessing

After selecting high-quality data segments, the data was pre-processed to create uniformly sampled time series suitable for machine learning models. The data within each selected history window was linearly interpolated to a daily frequency using the *interp1d* function from the *SciPy* library [23].

Linear interpolation was chosen because it is a conservative and robust method for this application. Unlike more complex algorithms, a linear approach does not introduce artificial oscillations between data points, which could create misleading patterns. This method assumes a constant rate of change between measurements. This is a reasonable first-order approximation for the biological processes of toxin accumulation and depuration over the short time frames enforced by our *max gap* criterion.

### 2.3.3. Dataset assembly and splitting

To construct the final dataset from the preprocessed time series, we parsed the data for each site over the entire study period. For every potential timestep  $T$ , the corresponding history window was evaluated against the quality control criteria defined in 2.3.1. If the criteria were met, the interpolated history,  $X_{T,s}$ , was generated. This feature set, along with its corresponding binary risk indicator,  $Y_{T,s}$ , was then added to the final dataset.

The key parameters for this process were defined as follows:

- **History window size:** 63 days
- **History distance:** 7 days
- **Risk threshold:**  $80 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq, consistent with the Canadian regulatory limit.

Finally, to ensure that the models were evaluated on unseen future data, we chronologically divided the entire dataset into training and testing sets. This approach prevented data leakage and provides a realistic assessment of the models' forecasting performance. The specific time periods used for training and testing in each of the three experiments are detailed in 2.4.

### 2.3.4. Models and ensemble learning

We applied 11 machine learning models to predict  $Y_{T,s}$  based on  $X_{T,s}$ . These models include:

- **Tree-based models (7):** AdaBoost (ADA), LightGBM (LGB), XGBoost (XGB), Random Forest (RF), Extra Trees (ET), Gradient Boosting Classifier (GBC) and CatBoost. [24–30]
- **Deep learning models (2):** Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCN). [31,32]
- **Statistical models (2):** Logistic Regression (LR) and Linear Discriminant Analysis (LDA). [33,34]

To enhance predictive performance and create a more robust forecasting tool, we developed a stacked ensemble model that integrates the outputs of all 11 individual models. This technique leverages the diverse strengths of different algorithms by training a final "meta-model" to learn the optimal way to combine their individual predictions.

The process involves two stages. First, the 11 "base models" are trained on the primary training set. Second, these trained models generate predictions on a separate validation set. These predictions are then used as input features to train a final ElasticNet regression model [35], which serves as the "meta-model." The ElasticNet model learns the best weighting of the base model outputs to produce a single, more reliable forecast. This entire stacked ensemble framework was implemented using the *StackingClassifier* function from the *scikit-learn* (*sklearn*) library [36]. This approach creates a final model that is often more accurate and generalizable than any single model alone.

### 2.3.5. Data imbalance and evaluation metrics

PSP events are infrequent, with only approximately 5 % of PST levels in the dataset exceeding regulatory thresholds. This significant class imbalance often impairs the performance of classifiers, leading to a trivial result that classifies all samples into the majority class.

Due to this imbalance in the data, the *accuracy* metric can lead to a misleading result. For example, a model can achieve high accuracy by simply classifying all samples as negative (the majority class). Other metrics like *recall*, *precision*, and their harmonic mean, *F1-score*, are subject to the selection of a classification cutoff. Each of the 11 models has different sensitivity, and their performance measured by these metrics in a single cutoff value (typically 0.5), could introduce bias. *Receiver Operating Characteristic* (ROC) curve is a visual representation of model performance across all cutoff selections. It is a plot of true positive rate vs the false positive rate at different cutoffs. We selected the *area under the ROC curve* (AUC) as an evaluation metric, because it is invariant to the cutoff selection and robust to class imbalance [37]. A perfect model has an AUC score of 1, meaning that it always correctly assigns positive and negative labels regardless of the cutoff selection and data skewness. The worst AUC score is 0.5, meaning that the model is equivalent to randomly assigning positive or negative labels with equal probability.

## 2.4. Experiments

Three experiments were conducted to measure the performance of our models. The first two experiments use the history of the total PST level as a predictor. The last experiment used data on the history of 12 individual toxin compounds, which sum to the total PST level.

**Experiment I** tested the models' performance using a history of total PST from 2013 to 2020. Both training and testing datasets were obtained using HPLC method. The models were trained with data from 2013 to 2016, and tested on data from 2017 to 2020. Note that although the detection limit of individual toxins was changed to  $1 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq since 2012, the total PST level was reported with a higher bound of  $25 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq. Any total PST value below  $25 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq would be reported as  $25 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq.

**Experiment II** tested the compatibility of the model framework with data collected using different technologies, simulating a common scenario in long-term monitoring programs where legacy data must be integrated. The models were trained on bioassay data (2000–2010) and tested on HPLC data (2013–2020). The detection limit of 2000–2010

data ( $40 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq) was applied to the newer data (collected from 2013 to 2020). To align the datasets, we clipped the HPLC data at the bioassay's detection limit, setting any value below  $40 \mu\text{g } 100 \text{ g}^{-1}$  to be  $40 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq. Invalid samples with  $0 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq were excluded.

**Experiment III** compared the performance of a univariate model, which used the total PST concentration, with a multivariate model, which uses concentrations of 12 available individual PST compounds. In this experiment, models were trained on the data from 2015 to 2016 and tested on the data from 2017 to 2020. The total PST value was calculated by summing up the individual toxin compounds. As a result, the minimum total PST value is  $1 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq, different from that in Experiment I.

## 2.5. Model interpretation

To ensure the interpretability of the complex ensemble model, we used SHAP (SHapley Additive exPlanations) [38]. SHAP is a method from cooperative game theory that explains the output of a machine learning model by assigning each feature an importance value for a particular prediction.

For each prediction, every feature is assigned a SHAP value that represents its contribution. The interpretation of these values is twofold. The magnitude of the SHAP value indicates the strength of a feature's influence. A larger absolute SHAP value means the feature has a greater impact on pushing the model's output away from the baseline. The sign of the SHAP value indicates the direction of influence. For binary classification task in this study, a positive SHAP value means the feature pushes the prediction towards the positive class (at-risk), while a negative SHAP value means it pushes the prediction towards the negative class (no-risk).

To summarize the overall importance of each feature across the entire dataset, we calculated the mean of the absolute SHAP values for each feature. This global measure reflects a feature's average impact on the model's predictions. We used the *KernelExplainer* from the *shap* Python library to compute these values for the ensemble model.

## 3. Results

### 3.1. Performance of individual model

Fig. 2 displays the performance of all models across the four experimental setups. The median AUC for the 11 models in each experiment was 0.897 (Experiment I), 0.901 (Experiment II), 0.909 (Experiment III univariate), and 0.939 (Experiment III multivariate), respectively. The highest AUC score and the corresponding best-performing individual model in each respective experiment was 0.915 (GBC), 0.910 (TCN), 0.930 (LSTM), and 0.949 (GBC), respectively.

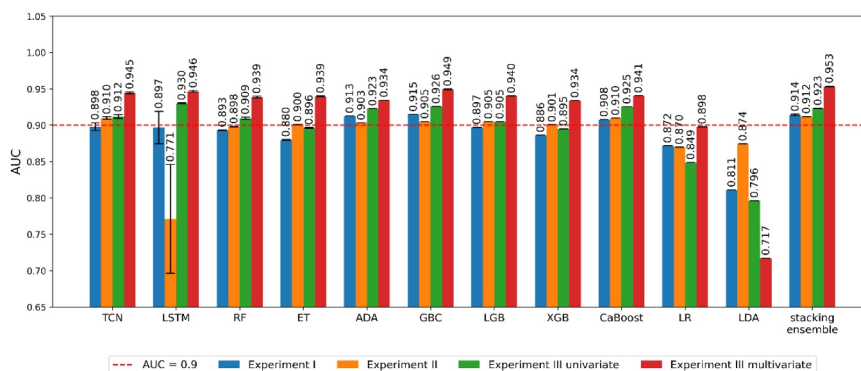


Fig. 2. Model performance for each experiment. The dashed red line shows the median performance of all experiments. TCN, LSTM, RF, ET, ADA, GBC, LGB, XGB, LDA, and LR stands for temporal convolutional neural network, long short-term memory, random forest, extra trees, AdaBoost, gradient boosting classifier, lightGBM, XGBoost, linear discriminant analysis, and logistic regression, respectively.

The results across the experiments revealed several key insights into the framework's performance and robustness. A particularly interesting finding emerged when comparing Experiments I and II. Despite being trained on data derived from different analytical methodologies (modern HPLC vs. legacy bioassay), the overall predictive potential remained remarkably consistent, with the median AUC across all 11 models being nearly identical (0.897 vs. 0.901).

However, as seen in Fig. 2, this stable median masked significant performance volatility among the individual models. Several algorithms performed differently between the two experiments; for instance, LSTM's performance dropped sharply in Experiment II. This highlighted a critical challenge for operational forecasting: the "best" individual model was often data-dependent and could change as analytical methods evolve. This finding provided the strongest justification for our stacked ensemble approach. The ensemble was designed to be robust against this individual model instability, learning to weigh diverse predictions to produce a single, reliable forecast that consistently performed at or near the top.

Experiment III further demonstrated the significant benefit of using more sensitive and detailed data. The models in the univariate portion, which used total PST data with a lower detection limit, achieved a higher median AUC of 0.909 compared to the first two experiments. The most substantial performance gain was observed in the multivariate part of Experiment III, where the median AUC rose to 0.939. This confirmed that leveraging the profiles of individual toxin compounds provided significantly more predictive power than using only the aggregated total PST value. The performance of tree-based algorithms, in particular, showed a substantial increase in this multivariate setting, underscoring their strength in capturing the intricate patterns from this type of environmental data.

### 3.2. Performance of ensemble learning

Fig. 3 shows that the performance of the ensemble model was close to the best-performing individual model in each experiment, despite the best-performing model being different in each experiment.

### 3.3. Spatial and temporal performance analysis

To evaluate the model's real-world applicability, we analyzed the performance of the ensemble model from Experiment III across different geographical locations and times of the year. The monitoring data was not evenly distributed, with most high-frequency sites concentrated around Vancouver Island and the lower mainland (Figure S2). There was also a slight imbalance in the occurrence of high-risk events, with the positive ratio being slightly higher on the west coast of Vancouver Island. Despite this spatial bias, the ensemble model demonstrated high overall performance across the test sites (Fig. 4). The vast majority of

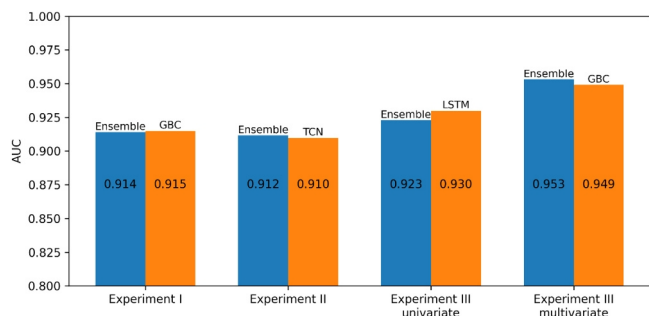


Fig. 3. comparing the performance of the ensemble model and the best individual model in each experiment. GBC, TCN and LSTM stands for gradient boosting classifier, temporal convolutional neural network, and long short-term memory, respectively.

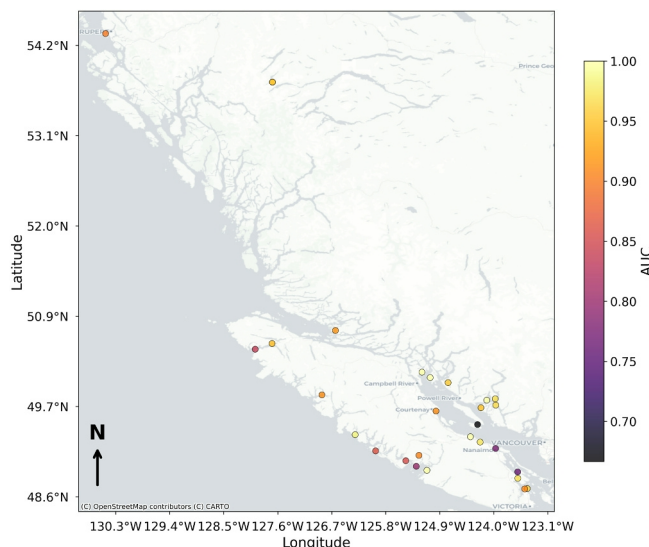


Fig. 4. Spatial distribution of model performance across coastal British Columbia. Points represent monitoring sites colored by AUC values of ensemble model in experiment III according to the scale on the right.

locations showed excellent performance, with AUC values often exceeding 0.90. While a few sites exhibited marginally lower performance, the results confirmed that the model generalized effectively across the primary monitoring region where data was available.

From a temporal perspective, the model’s performance remained strong and stable throughout the typical algal bloom season (Fig. 5). The AUC was consistently high from spring (week 18) through late fall (week 49), covering the periods when PSP events were most common. This

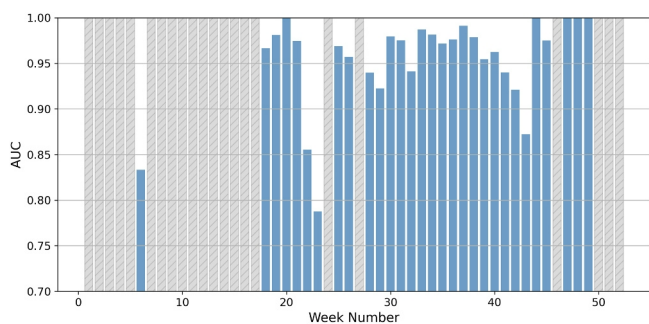


Fig. 5. Weekly AUC performance across the annual cycle in testing set. Blue bars show the average AUC values of each week in the year in the testing set. Diagonal-hatched bars indicate weeks without data.

consistent performance indicated that the model’s predictive accuracy was reliable throughout the seasons of high risk.

### 3.4. SHAP value of the ensemble

We analyzed its decision-making process through SHAP values across different toxins and different weeks of the year. Fig. 6 shows SHAP value and mean concentration of each toxin in the samples of training set. The feature importance showed a significant correlation ( $r = 0.79$ ) with the concentrations of the respective toxins. However, C-2, GTX-3, and GTX-4 had higher feature importance despite having slightly lower concentrations than GTX-1.

Fig. 7 revealed that positive and negative events had different toxin profiles that could affect the feature importance. Despite GTX-1 having a slightly higher concentration than GTX-3 and C2, positive and negative events contained similar percentages of GTX-1, different by 0.6 %. On the other hand, C-2 had a much larger differences, leading to higher feature importance.

Fig. 8 shows the feature importance of each toxin in each time step. The model primarily focused on the most recent information. In particular, the model focused heavily on the toxicity of C-2, GTX-3, GTX-4, STX and NEOSTX within 8 days prior to the target day.

## 4. Discussion

### 4.1. Predictive performance and robustness of machine learning models and ensemble

This study investigates machine learning models for prediction of the PSP outbreaks based on toxins accumulated in blue mussels. The comprehensive exploration across three experiments aims to understand the influences of different data characteristics on the model performance. The overall exceptional performance demonstrates the robustness of the prediction based on historic shellfish toxin data.

The consistently strong performance of tree-based algorithms aligns with prior ecological applications, where random forests and boosting methods have been used to predict PSP exceedances in Southeast Alaska [14] and *Alexandrium* blooms in the Adriatic Sea [15]. These studies confirm the utility of tree ensembles with environmental drivers. Our work extends this evidence by showing that tree-based models also excel when applies in a novel autoregressive context using toxin-history data, a setting that has not been systematically evaluated in previous PSP forecasting research.

Experiment II aims to address a critical, practical challenge in long-term monitoring: ensuring the valuable legacy data remain useful as analytical techniques evolve. The harmonization of the HPLC and bioassay datasets demonstrates the compatibility between the two. The strong performance (AUC 0.901), similar to that in Experiment I, demonstrates that our predictive framework remains highly effective when

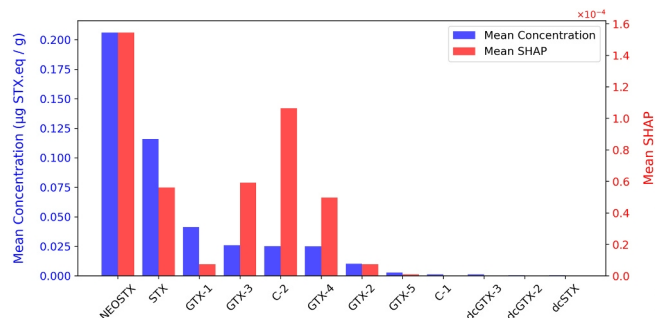
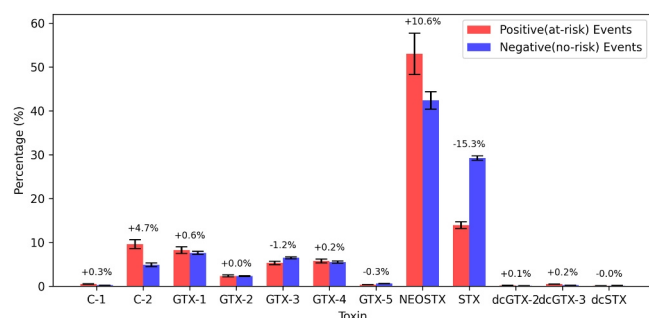
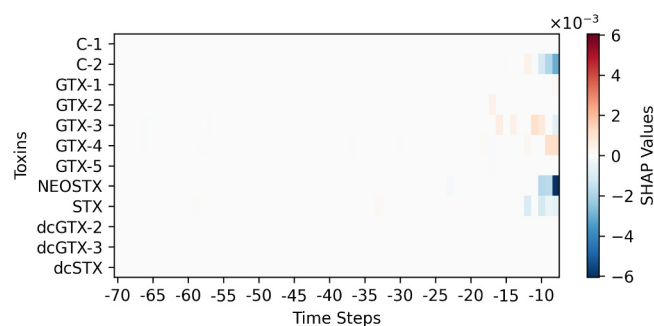


Fig. 6. Comparison of mean toxin concentrations of training set and their feature importance in predicting shellfish toxicity. Blue bars represent mean concentrations, while the red bars indicate feature importance scores. Toxins are ordered by decreasing feature importance from left to right.



**Fig. 7.** Distribution of toxin percentages between positive (at-risk) and negative (no-risk) events in the training set. Bars represent mean percentages with error bars showing standard error. The percentage numbers on top of the bars shows the difference between positive and negative events.



**Fig. 8.** Feature importance of each toxin at each time steps. The darker color represents higher feature importance.

trained on properly harmonized legacy data. This highlights the framework's practical value, allowing historical datasets to remain valuable for forecasting, which is essential for understanding long-term trends and improving risk assessment.

Experiment III explores the impact of evolved analytical methodologies, specifically those providing lower detection limits and more detailed toxin profiles, on model performance. The comparison between Experiment I and Experiment III (univariate) directly addresses the influence of these lower detection limits on predictive performance. Models trained on data derived from these more advanced methods (from post-2015 data where total PST was summed from more sensitively measured individual profiles) consistently outperform those from Experiment I, which has a higher reporting limit of  $25 \mu\text{g } 100 \text{ g}^{-1}$  STXdiHCl eq. This highlights that the ability to detect lower PST concentrations, including minor sulfocarbamoyl analogs indicative of early bloom stages [39], provides crucial, more informative data that significantly enhances model performance. This finding underscores the vital role of analytical sensitivity within monitoring programs for accurate risk assessments. Furthermore, the multivariate aspect of Experiment III confirms the models' capability to leverage these detailed toxin profiles, allowing for a more nuanced understanding of toxin composition effects on predictions.

A key aspect of our framework is the use of a stacked ensemble model, which consistently matches or exceeds the performance of the best individual model in each experiment. Stacking is a powerful and standard technique that improves operational robustness by training a meta-model to optimally combine the predictions from a diverse set of base models. For a monitoring program where reliability is critical, this ensemble approach provides a more resilient and generalizable solution than relying on any individual algorithm.

The novelty of our study lies not in introducing new algorithms, but in reframing PSP prediction as an autoregressive classification problem based on sentinel mussel toxin histories. By systematically

benchmarking 11 models across multiple conditions—including bioassay versus HPLC data, univariate versus multivariate toxin inputs, and varying thresholds—we show that toxin-history forecasting is both feasible and reliable. This formulation complements environment-driven approaches by capitalizing on the mussel's intrinsic role as an environmental integrator.

#### 4.2. The predictive framework in the context of PST monitoring

Timely and effective forecasting of paralytic shellfish poisoning risk is paramount for assessing marine environmental quality and managing risks to public health and aquaculture. While some predictive efforts incorporate environmental drivers (e.g. surface temperature, salinity, nutrient concentrations, chlorophyll *a*) [12,18], our framework utilizes the direct biological response of sentinel species by analyzing their historical PST data.

The blue mussel acts as natural bio-integrators of toxins from microalgae in the sea water. The detailed toxin profiles used in Experiment III serve as rich chemical fingerprints from which our models learn to assess future risk. By forecasting the evolution of toxicity within the sentinel species, this approach provides a practical and direct prediction of shellfish safety. The strength of our models lies in predicting the subsequent accumulation and risk within the shellfish themselves, a critical aspect of effective coastal environmental management. This approach offers several advantages. First, it can be particularly valuable in regions where comprehensive, high-frequency environmental data for all relevant parameters are too sparse or not available. The shellfish effectively "samples" its environment, providing a direct measure of toxicological impact. Second, the rich chemical fingerprints obtained from detailed toxin profiles contain information not only about the total toxicity but also potentially about the stage or type of the harmful algal blooms [11]. Our models learn from them to assess future risk and leverage this direct biological signal from the sentinel species to forecast the evolution of their PST levels. This approach provides a practical and direct prediction of shellfish safety, offering a timely assessment crucial for protecting human consumers and guiding aquaculture operations.

The justification for this approach lies in mussels' role as sentinel organisms: by continuously filtering seawater, they integrate multiple environmental signals into their tissues, creating a cumulative record of exposure. This principle underpins long-standing monitoring programs, such as NOAA's Mussel Watch Program [19] and Canadian Shellfish Sanitation Program (CSSP), which relies on mussel biotoxin testing to guide harvest closures [20]. These precedents support the use of toxin-history forecasting as both scientifically sound and operationally relevant.

However, there are trade-offs compared with conventional environment-driven approaches. Models that incorporate algal counts, hydrographic, or meteorological data can provide mechanistic insights into bloom initiation and, in some cases, extend forecast lead times. By contrast, toxin-history models primarily reflect recent exposures within the sentinel organism. They therefore excel in operational robustness and scalability but offer less explanatory power for underlying ecological drivers. We therefore view toxin-history and environment-driven approaches as complementary. Importantly, future work should integrate these two paradigms, combining toxin-history predictors with environmental covariates to unite the operational robustness of sentinel-based forecasting with the mechanistic insights of environment-driven models.

#### 4.3. Spatial variation in the model performance

Among the sites with an identifiable location, the model demonstrates overall high performance across British Columbia. The performance shows little variation between most sites in the testing set, despite the spatial data imbalance. However, it should be noted that the data for this spatial analysis are limited to the subset of testing sites with

identifiable geographic coordinates, which are concentrated around Vancouver Island and the lower mainland. The generalization of the results to other areas also needs further study in the future when more data are available.

This robust performance is particularly notable given the inherent complexities and variability often encountered in marine environmental monitoring, which traditionally relies heavily on detailed environmental covariates. Research has shown that environmental predictors of water quality vary significantly by location along BC's coastline [40,41]. These geographical and oceanographic differences create distinct environmental regimes that influence pollutant behavior and model predictability. Furthermore, land-based sources like rainwater runoff and sewage discharges contribute significantly to eutrophication, with their impact varying substantially based on local watershed characteristics [21,42].

However, unlike models that are highly sensitive to these numerous and dynamic external environmental factors, our autoregressive framework, by focusing on the intrinsic temporal patterns of PST accumulation within the sentinel species, inherently accounts for the integration of these complex environmental influences. This approach significantly reduces the dependency on exhaustive external environmental datasets, making our model particularly adaptable and robust in regions where comprehensive environmental monitoring data is sparse or highly heterogeneous. Consequently, our model achieves strong predictive performance while being less susceptible to the variability and measurement challenges associated with diverse environmental covariates.

#### 4.4. Ecological insights and implications for HAB understanding in British Columbia

Beyond its predictive performance, our framework provides valuable ecological insights into regional PST dynamics through the SHAP analysis of the ensemble model. These data-driven observations contribute to a more nuanced understanding of which toxins and timeframes are most critical for predicting risk on Canada's west coast. Regarding specific toxins, the analysis confirmed the importance of several key compounds.

First, while the analysis is correlational, the high feature importance of specific compounds aligns directly with regional ecological dynamics. The primary PSP-causing organism, *A. catenella*, has a toxin signature on the Pacific coast dominated by N-sulfocarbamoyl (C-toxins) and gonyautoxins (GTX-toxins), which can comprise over 90 % of the total toxin content [43–46]. Our model identifies NEOSTX, STX, and especially C-2 as key predictors. This data-driven finding is consistent with literature showing C-2 is often the most dominant compound in *A. catenella* profiles, particularly under the temperature and salinity conditions common to BC's bloom season [44].

Furthermore, the model's temporal focus on the most recent data is consistent with the biological processes of toxin persistence and depuration. Mussels are generally considered rapid detoxifiers of PSTs, meaning their toxin load reflects recent environmental conditions and makes older data less relevant for near-term prediction. Empirical studies demonstrate that *Mytilus* spp. depurate most PSTs within days to weeks once toxicogenic algae are absent, with first-order elimination rates for C1 + C2 of  $\sim 0.08 \text{ d}^{-1}$  (half-life  $\approx 8\text{--}9$  days) observed in both laboratory and field experiments [47]. This rapid clearance implies that mussel toxin loads primarily reflect recent environmental exposure, justifying the model's reliance on shorter historical windows. Mechanistically, fast depuration is facilitated by filtration and gut evacuation, with additional evidence for microbial or enzymatic degradation of absorbed toxins [48]. Comparative data from Atlantic Canada confirm that mussels accumulate and depurate PSTs more quickly than scallops, underscoring their suitability as sentinel species [49]. In contrast, species like geoduck clams have been shown to transform and retain PSTs, leading to longer persistence in tissues [50]. These interspecies

differences support our focus on mussel data for near-term forecasting while cautioning against extrapolating model predictions to slower-clearing bivalves.

Second, the enhanced predictive performance observed in Experiment III (univariate) when using total PST values derived from the summation of individual compounds (effectively reflecting lower detection limits) highlights an important ecological aspect. The ability to detect minor analogs, such as the low-toxicity sulfocarbamoyl toxins, which might be missed with higher reporting thresholds for total PST, appears crucial. These compounds are often associated with the early stages of *Alexandrium* blooms [39]. Their inclusion improves model performance, suggesting that these early-stage toxin signatures are integral to the ecological progression towards more significant, high-toxicity events and serve as important precursors.

In summary, the high model performance during the summer and fall bloom seasons indicates robust temporal performance. The framework serves not only as a forecasting tool but also as an analytical lens, revealing patterns in toxin data that reflect underlying ecological processes. These insights into dominant toxin profiles, early-bloom signatures, and species-specific persistence enhance the regional understanding of PSP events and can directly inform the design of more targeted and efficient environmental monitoring programs.

## 5. Limitations and future work

This study establishes a robust predictive framework for PSP risk using the blue mussel, *Mytilus edulis*, a globally recognized sentinel species for marine biotoxin monitoring. This species-specific focus provides a strong foundation for risk assessment, but it also introduces limitations, since interspecies differences in toxin uptake, biotransformation, and depuration may constrain the generalizability of our findings.

Future research can build upon this framework by incorporating other key monitoring species, such as clams, oysters, or scallops. Such multi-species models would be valuable because the underlying biological processes of toxin accumulation, biotransformation, and depuration are known to vary significantly between bivalves. Extending the framework in this way would account for these interspecies differences, leading to an even more comprehensive and powerful regional risk assessment tool.

Another limitation of our framework is its reliance on toxin-history data alone. While this enhances operational robustness and scalability, it does not capture toxigenic algal bloom initiation or provide mechanistic ecological insights that environment-driven models provide. Future work should therefore focus on hybrid approaches that integrate toxin-history predictors with environmental covariates (e.g., algal counts, hydrographic, or meteorological data). Such integration would combine the practical advantages of sentinel-based forecasting with the extended lead times and ecological interpretability offered by environment-driven models.

A further methodological consideration is that this study uses the default hyperparameters for the base models without extensive tuning. This decision is made intentionally to mitigate the risk of overfitting on a dataset that, while spanning two decades, is sparse and contains a limited number of positive risk events. The framework's core design relies on the diversity and robustness of the ensemble rather than the fine-tuning of a single model. The final meta-model learns to balance the strengths and weaknesses of these diverse models, leading to strong and generalizable performance. However, we acknowledge that with a larger and more comprehensive dataset in the future, a systematic hyperparameter tuning process would be a valuable next step to potentially enhance performance further.

## 6. Conclusion

This study demonstrates the significant potential of machine learning

frameworks as a short-term forecasting tool for Paralytic Shellfish Poisoning (PSP) risk in blue mussels. The central innovation of this work is the validation of an autoregressive, toxin-history approach, which reframes the prediction problem by leveraging the mussel's intrinsic role as a "bio-integrator" of environmental conditions. By systematically benchmarking 11 models, we show that this formulation is a feasible, reliable, and scalable complement to conventional environment-driven forecasting methods.

Our comprehensive evaluation confirms that tree-based algorithms, particularly when aggregated into a stacked ensemble, consistently achieves excellent predictive performance (AUC  $\approx$  0.95) across multiple experimental conditions. The findings highlight that leveraging multivariate toxin compound profiles significantly enhances predictive performance over univariate totals, and that the framework remains highly effective when trained on properly harmonized legacy data from different analytical methods.

This work validates the hypothesis that the complex environmental drivers of PSP risk are effectively captured within the sentinel species' tissues, making its recent toxin history a direct and powerful input for forecasting. Our SHAP-based interpretation of the ensemble model provides novel ecological insights, revealing that specific toxins, notably NEOSTX, STX, and C-2, and data from the most recent 1–2 weeks are the dominant drivers of prediction. This data-driven finding is supported by established biological mechanisms, aligning with the known toxin signatures of regional *Alexandrium* blooms and the rapid depuration kinetics of mussels.

Ultimately, this framework offers a pathway towards more proactive, data-driven, and resource-efficient PST monitoring strategies along Canada's west coast. By providing timely alerts, such models can enhance public health protection and support sustainable aquaculture industries. Future work should focus on developing hybrid models that integrate these toxin-history predictors with environmental covariates, uniting the operational robustness of our approach with the mechanistic insights of environment-driven models. Expanding the framework to include other key bivalve species would further create a more comprehensive and powerful regional risk assessment tool.

### Environmental implication

This framework provides a robust early warning system for Paralytic Shellfish Poisoning (PSP) events in blue mussels along Canada's west coast, leveraging historical toxin data. By accurately forecasting PSP risk (AUC 0.95), it enhances current monitoring strategies, enabling timely and targeted interventions to protect public health from contaminated shellfish. The model's reliance on direct biological responses within sentinel species reduces dependency on comprehensive external environmental datasets, making it adaptable to data-sparse regions. This proactive, data-driven approach supports sustainable shellfish aquaculture and harvesting industries, ultimately improving coastal ecosystem health management and informing environmental policy.

### CRedit authorship contribution statement

**Chang Bi:** Writing – original draft, Methodology, Investigation, Formal analysis. **Youlian Pan:** Writing – review & editing, Supervision, Funding acquisition, Data curation, Conceptualization. **Xuekui Zhang:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

### Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Google Gemini and ChatGPT in order to improve language and expression. After using these tools, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

### Funding

This work was supported by the Digital Health and Geospatial Analytics program of the National Research Council Canada [grant DHGA-112–1]. XZ was supported by the Canada Research Chair program [950–231363] and Michael Smith Health Research BC Scholar Award [SCH-2022–2553].

### Declaration of Competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xuekui Zhang reports financial support was provided by National Research Council Canada. Youlian Pan reports financial support was provided by National Research Council Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgment

We thank John Hoeve and Sebastien Belanger of CFIA for providing the biotoxins data, Kristen Kirby, Wade Rourke and Edwards Robyn of CFIA for their constructive comments on this study.

### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2025.140459.

### Data Availability

The authors do not have permission to share data. The complete source code used for the analysis in this study has been made publicly available in a GitHub repository to ensure full reproducibility of our methodological pipeline: <https://github.com/changbiHub/PSPRA>. Due to the terms of our data-sharing agreement with the Canadian Food Inspection Agency (CFIA), we are not permitted to share the raw monitoring data used in this research. To facilitate the use of the code, the repository includes a synthetic dataset. It is important to note that this dataset is provided solely as a minimal working example to demonstrate the required input format and ensure the code is fully executable. This data does not mimic the statistical properties of the original dataset and should not be used for performance evaluation or inference.

### References

- [1] Richardson, K., 1997. Harmful or Exceptional Phytoplankton Blooms in the Marine Ecosystem. In: Blaxter, J.H.S., Southward, A.J. (Eds.), in *Advances in Marine Biology*, 31. Academic Press, pp. 301–385.
- [2] Landsberg, J.H., April 2002. The effects of harmful algal blooms on aquatic organisms. *Rev Fish Sci* 10, 113–390.
- [3] Basti, L., Hégaret, H., Shumway, S.E., 2018. *Harmful Algal Blooms and Shellfish. In Harmful Algal Blooms*. John Wiley & Sons, Ltd, pp. 135–190.
- [4] Acres, J., Gray, J., November 1978. Paralytic shellfish poisoning. *Can Med Assoc J* 119, 1195–1197.
- [5] Anderson, D.M., Cembella, A.D., Hallegraeff, G.M., 2012. Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. *Annu Rev Mar Sci* 143–176.
- [6] "Health Canada's maximum levels for chemical contaminants in foods," 2020. [Online]. Available: (<https://www.canada.ca/en/health-canada/services/food-nutrition/food-safety/chemical-contaminants/maximum-levels-chemical-contaminants-foods.html>). [Accessed 2024].
- [7] Wang, S., Li, G., Ji, X., Wang, Y., Xu, B., Tang, J., Guo, C., 2024. Machine learning-driven assessment of heavy metal contamination in the impounded lakes of China's South-to-North Water Diversion Project: Identifying spatiotemporal patterns and ecological risks. *J Hazard Mater* 480, 135983.
- [8] Li, C., Yang, Z., Guan, D.-X., Yu, T., Jiang, Z., Wu, X., Yang, Y., Luan, S., Xu, H., Huang, C., Liangjie, Z., 2025. Spatial-machine learning framework for rapid

- identification of soil cadmium risk in high geochemical background areas. *J Hazard Mater* 492, 138091.
- [9] Zhang, X., Wang, X., Wu, F., Liang, W., Wang, S., Liang, J., Zhao, X., Wu, F., 2024. Machine learning models to predict the bioaccessibility of parent and substituted polycyclic aromatic hydrocarbons (PAHs) in food: Impact on accurate health risk assessment. *J Hazard Mater* 480, 136102.
- [10] Chandrasekar, V., Mohammad, S., Aboumarzouk, O., Singh, A.V., Dakua, S.P., 2025. Quantitative prediction of toxicological points of departure using two-stage machine learning models: a new approach methodology (NAM) for chemical risk assessment. *J Hazard Mater* 487, 137071.
- [11] Grasso, I., Archer, S.D., Burnell, C., Tupper, B., Rauschenberg, C., Kanwit, K., Record, N.R., 2019. The hunt for red tides: deep learning algorithm forecasts shellfish toxicity at site scales in coastal Maine. *Ecosphere* 10, e02960.
- [12] Shin, J., Kim, S.M., January 2022. Temporal prediction of paralytic shellfish toxins in the mussel *mytilus galloprovincialis* using a LSTM neural network model from environmental data. *Toxins* 14, 51.
- [13] Moore, S.K., Mantua, N.J., Hickey, B.M., Trainer, V.L., February 2009. Recent trends in paralytic shellfish toxins in Puget Sound, relationships to climate, and capacity for prediction of toxic events. *Harmful Algae* 8, 463–477.
- [14] Harley, J.R., Lanphier, K., Kennedy, E., Whitehead, C., Bidlack, A., 2020. Random forest classification to determine environmental drivers and forecast paralytic shellfish toxins in Southeast Alaska with high temporal resolution. *Harmful Algae* 99, 101918.
- [15] Valbi, E., Ricci, F., Capellacci, S., Casabianca, S., Scardi, M., Penna, A., 2019. A model predicting the PSP toxic dinoflagellate *Alexandrium minutum* occurrence in the coastal waters of the NW Adriatic Sea. *Sci Rep* 9, 4166.
- [16] Lin, S., Pierson, D.C., Mesman, J.P., 2023. Prediction of algal blooms via data-driven machine learning models: an. *Geosci Model Dev* 35–46.
- [17] Carey, C.C., Calder, R.S., Figueiredo, R.J., Gramacy, R.B., Lofton, M.E., Schreiber, M.E., Thomas, R.Q., 2025. A framework for developing a real-time lake phytoplankton forecasting system to support water quality management in the face of global change. *Ambio* 54, 475–487.
- [18] Xie, H., Fischer, A.M., Strutton, P.G., July 2021. Generalized linear models to assess environmental drivers of paralytic shellfish toxin blooms (Southeast Tasmania, Australia). *Cont Shelf Res* 223, 104439.
- [19] Kimbrough, K., Johnson, W., Lauenstein, G., Christensen, J., Apeti, D., 2008. An assessment of two decades of contaminant monitoring in the nation's coastal zone. NOAA Tech Memo NOS NCCOS 74.
- [20] "Canadian Shellfish Sanitation Program (CSSP)," Canadian Food Inspection Agency (CFIA), [Online]. Available: (<https://inspection.canada.ca/en/preventive-contr-ols/fish/cssp>). [Accessed 07 09 2025].
- [21] You, S., Xing, L., Lesperance, M., Pan, Y., Zhang, X., 2024. Longitudinal study of paralytic shellfish toxins along Canada's coast. *Environ Res*, 118944.
- [22] Liu, Z., You, S., Xing, L., Durand, G., Moccia, L.P., Mercier, V., Pan, Y., Zhang, X., 2025. OPMS-A web-based ocean pollution monitoring system. *Mar Pollut Bull*, 117510.
- [23] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., 2020. and May, "SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* 17 (3), 261–272.
- [24] Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 119–139.
- [25] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*.
- [26] Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. *Proc 22nd acm sigkdd Int Conf Knowl Discov data Min* 785–794.
- [27] Breiman, L., 2001. Random forests. *Mach Learn* 5–32.
- [28] Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach Learn* 3–42.
- [29] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann Stat* 1189–1232.
- [30] Prokhorenkova, L. a G.G., Vorobev, A., Dorigush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 31.
- [31] Graves, A., Graves, A., 2012. Long short-term memory. *Supervised Seq Label Recurr Neural Netw* 37–45.
- [32] S. Bai, J.Z. Kolter and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018.
- [33] Cox, D.R., 1958. The regression analysis of binary sequences. *J R Stat Soc Ser B Stat Methodol* 215–232.
- [34] Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Ann Eugen* 179–188.
- [35] Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* 301–320.
- [36] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12, 2825–2830.
- [37] Richardson, E., Trevizani, R., Greenbaum, J.A., Carter, H., Nielsen, M., Peters, B., 2024. The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns* 5.
- [38] Lundberg, S.M., Lee, S.-L., 2017. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30.
- [39] Humpage, A., Magalhaes, V., Frosio, S., 2010. Comparison of analytical tools and biological assays for detection of paralytic shellfish poisoning toxins. *Anal Bioanal Chem* 397, 1655–1671.
- [40] Desta, B.N., Tustin, J., Sanchez, J.J., Heasley, C., Schwandt, M., Bishay, F., Chan, B., Knezevic-Stevanovic, A., Ash, R., Jantzen, D., Young, I., February 2024. Environmental predictors of *Escherichia coli* concentration at marine beaches in Vancouver, Canada: a Bayesian mixed-effects modelling analysis. *Epidemiol Infect* 152, e38.
- [41] B.L. Iwabuchi, "A comparison of temperature and salinity conditions on the east and west coasts of Vancouver Island: Implications for intertidal invertebrate population persistence in the face of climate change," 2019. [Online]. Available: (<https://tru.arcabc.ca/islandora/object/tru:5251>). [Accessed 7 April 2025].
- [42] "Phosphorous management in Vancouver Island streams," 2014. [Online]. Available: ([https://www2.gov.bc.ca/assets/gov/environment/air-land-water/water/waterquality/water-quality-reference-documents/phosphorous\\_management\\_vi\\_streams\\_guidance\\_2014.pdf](https://www2.gov.bc.ca/assets/gov/environment/air-land-water/water/waterquality/water-quality-reference-documents/phosphorous_management_vi_streams_guidance_2014.pdf)). [Accessed 2024].
- [43] F.J.R. Taylor, "Harmful algal blooms in the PICES region of the North Pacific," 2002. [Online]. Available: ([https://meetings.pices.int/publications/scientific-reports/Report23/Rep\\_23\\_full.pdf](https://meetings.pices.int/publications/scientific-reports/Report23/Rep_23_full.pdf)).
- [44] Laabir, M., Collos, Y., Masseret, E., Grzebyk, D., Abadie, E., Savar, V., Sibat, M., Amzil, Z., May 2013. Influence of Environmental Factors on the Paralytic Shellfish Toxin Content and Profile of *Alexandrium catenella* (Dinophyceae) Isolated from the Mediterranean Sea. *Mar Drugs* 11, 1583–1601.
- [45] Bui, Q.T.N., Pradhan, B., Kim, H.-S., Ki, J.-S., April 2024. Environmental Factors Modulate Saxitoxins (STXs) Production in Toxic Dinoflagellate *Alexandrium*: An Updated Review of STXs and Synthesis Gene Aspects. *Toxins* 16, 210.
- [46] Krock, B., Seguel, C.G., Cembella, A.D., 2007. Toxin profile of *Alexandrium catenella* from the Chilean coast as determined by liquid chromatography with fluorescence detection and liquid chromatography coupled with tandem mass spectrometry. *Harmful Algae* 6 (5), 734–744.
- [47] Botelho, M.J., Vale, C., Mota, A.M., Gonçaves, M. d L.S.S., 2010. Depuration kinetics of paralytic shellfish toxins in *Mytilus galloprovincialis* exposed to *Gymnodinium catenatum*: laboratory and field experiments. *J Environ Monit* 12, 2269–2275.
- [48] Donovan, C.J., Ku, J.C., Quilliam, M.A., Gill, T.A., 2008. Bacterial degradation of paralytic shellfish toxins. *Toxicon* 52, 91–100.
- [49] Rourke, W.A., Justason, A., Martin, J.L., Murphy, C.J., 2021. Shellfish toxin uptake and depuration in multiple Atlantic Canadian Molluscan species: application to selection of sentinel species in monitoring programs. *Toxins* 13, 168.
- [50] Medina-Elizalde, J., Garcia-Mendoza, E., Turner, A.D., Sanchez-Bravo, Y.A., Murillo-Martinez, R., 2018. Transformation and depuration of paralytic shellfish toxins in the geoduck clam *Panopea globosa* from the Northern Gulf of California. *Front Mar Sci* 5, 335.