Chemosphere 352 (2024) 141329

Contents lists available at ScienceDirect

Chemosphere

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

Heavy metals prediction in coastal marine sediments using hybridized machine learning models with metaheuristic optimization algorithm

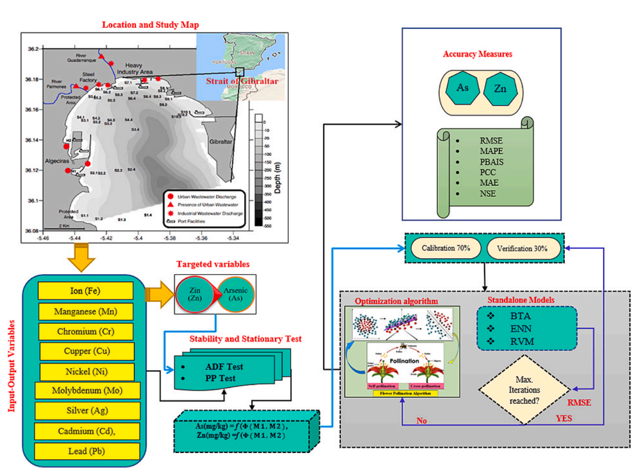
Zaher Mundher Yaseen [a,b,*], Wan Hanna Melini Wan Mohtar [c,d], Raad Z. Homod [e], Omer A. Alawi [f], Sani I. Abba [b], Atheer Y. Oudah [g,h], Hussein Togun [i], Leonardo Goliatt [j], Syed Shabi Ul Hassan Kazmi [k], Hai Tao [l,m,n]

[a] *Civil and Environmental Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia*
[b] *Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia*
[c] *Department of Civil Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600, UKM, Bangi, Selangor, Malaysia*
[d] *Environmental Management Centre, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*
[e] *Department of Oil and Gas Engineering, Basrah University for Oil and Gas, Basra, Iraq*
[f] *Department of Thermofluids, School of Mechanical Engineering, Universiti Teknologi Malaysia, 81310, UTM Skudai, Johor Bahru, Malaysia*
[g] *Department of Computer Sciences, College of Education for Pure Science, University of Thi-Qar, Nasiriyah, 64001, Iraq*
[h] *Information and Communication Technology Research Group, Scientific Research Center, Al-Ayen University, Nasiriyah, 64001, Iraq*
[i] *Department of Mechanical Engineering, College of Engineering, University of Baghdad, Baghdad, Iraq*
[j] *Computational and Applied Mechanics Department, Federal University of Juiz de Fora, 36036-900, Brazil*
[k] *Guangdong Provincial Key Laboratory of Marine Disaster Prediction and Prevention, and Guangdong Provincial Key Laboratory of Marine Biotechnology, Shantou University, Shantou, 515063, China*
[l] *School of Computer and Information, Qiannan Normal University for Nationalities, Duyun, 558000, Guizhou, China*
[m] *Institute of Big Data Application and Artificial Intelligence, Qiannan Normal University for Nationalities, Duyun, 558000, Guizhou, China*
[n] *Faculty of Data Science and Information Technology, INTI International University, 71800, Malaysia*

HIGHLIGHTS

- New hybrid machine learning (ML) models were developed for heavy metals prediction.
- As (mg/kg) and zinc Zn (mg/kg) in marine sediments Algeciras Bay were predicted.
- Akaike and Schwarz information criteria were used for data stationarity appraisal.
- The proposed hybrid ML model exhibited the superior prediction performance.
- The study provides insights for environmental decision-makers for proper management.

GRAPHICAL ABSTRACT



* Corresponding author. Civil and Environmental Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia.

E-mail addresses: z.yaseen@kfupm.edu.sa (Z.M. Yaseen), hanna@ukm.edu.my (W.H. Melini Wan Mohtar), raadahmood@yahoo.com (R.Z. Homod), omer.adnan85@gmail.com (O.A. Alawi), sani.abba@kfupm.edu.sa (S.I. Abba), atheer@alayen.edu.iq (A.Y. Oudah), hussein-tokan@coeng.uobaghdad.edu.iq (H. Togun), leonardo.goliatt@ufjf.br (L. Goliatt), shabiulhassansyed5@gmail.com (S.S. Ul Hassan Kazmi), haitao@sgmtu.edu.cn (H. Tao).

https://doi.org/10.1016/j.chemosphere.2024.141329
Received 26 October 2023; Received in revised form 9 January 2024; Accepted 28 January 2024
Available online 29 January 2024
0045-6535/

ARTICLE INFO

Handling editor: Michael Bank

Keywords:
Artificial intelligence
Heavy metals
Sensitivity analysis
Soil contamination

ABSTRACT

This study proposes different standalone models viz: Elman neural network (ENN), Boosted Tree algorithm (BTA), and f relevance vector machine (RVM) for modeling arsenic (As (mg/kg)) and zinc (Zn (mg/kg)) in marine sediments owing to anthropogenic activities. A heuristic algorithm based on the potential of RVM and a flower pollination algorithm (RVM-FPA) was developed to improve the prediction performance. Several evaluation indicators and graphical methods coupled with visualized cumulative probability function (CDF) were used to evaluate the accuracy of the models. Akaike (AIC) and Schwarz (SCI) information criteria based on Dickey-Fuller (ADF) and Philip Perron (PP) tests were introduced to check the reliability and stationarity of the data. The prediction performance in the verification phase indicated that RVM-M2 (PBAIS = -o.0465, MAE = 0.0335) and ENN-M2 (PBAIS = 0.0043, MAE = 0.0322) emerged as the best model for As (mg/kg) and Zn (mg/kg), respectively. In contrast with the standalone approaches, the simulated hybrid RVM-FPA proved merit and the most reliable, with a 5 % and 18 % predictive increase for As (mg/kg) and Zn (mg/kg), respectively. The study's findings validated the potential for estimating complex HMs through intelligent data-driven models and heuristic optimization. The study also generated valuable insights that can inform the decision-makers and stockholders for environmental management strategies.

1. Introduction

Heavy metals (HMs) are naturally occurring elements in soil, water, air, and living organisms. These metals are considered "heavy" because they have a high atomic weight and density and can be toxic to plants, animals, and humans at high concentrations (Alshehri et al., 2021; Kazemi and Hosseini, 2011; Yaseen, 2021, 2022). The United States Environmental Protection Agency (EPA) highlighted that HMs in the soil could come from various sources, including industrial activities, mining, agricultural practices, and natural weathering of rocks and minerals (EPA, 2007). Some common HMs in soil include Zinc (Zn), lead (Pb), cadmium (Cd), mercury (Hg), Arsenic (As), and chromium (Cr) (Li et al., 2019; Yang et al., 2018). The primary goals of sustainable development goals (SDGs) No. 15 of the 2030 UN Agenda were to safeguard, restore, and advance the sustainable use of terrestrial ecosystems (Sinkakarimi et al., 2020; Song et al., 2009). It is added that; HMs accumulation in the soil can disrupt soil ecosystems and negatively influence plant and animal biodiversity (Sun et al., 2023; Zhang et al., 2020). These, therefore, emphasize adopting sustainable land use practices that can promote soil health and reduce HMs accumulation in achieving the targeted goal (Singh and Kalamdhad, 2011). Knowingly, Zn is not the strongest metal, but its exposure significantly negatively impacts the soil by triggering a decrease in the soil virility and loss of biodiversity.

It can harm plants and reduce crop yields. They can also leach into groundwater, contaminating drinking water sources (Ghadimi, 2014). Also, industrial effluent can contain HMs such as Zn (mg/kg), which can accumulate in soil and adversely affect plant growth and the ecosystem (Alloway, 2008). Similarly, Zn (mg/kg) is a naturally occurring element with the atomic number 30 and the Zn symbol. It is a bluish-white, lustrous metal that is moderately reactive and less toxic than other HMs such as lead, cadmium, and mercury. According to Alloway (2008), Zn can become an HMs pollutant when it accumulates in high environmental concentrations. This can occur through both natural and anthropogenic substances. It is documented that, when present in high concentrations, Zn can cause soil compaction, reducing water infiltration and plant root growth. This can result in decreasing soil porosity and increased soil bulk density, affecting soil structure and overall soil health (Bou Kheir et al., 2010; Tan et al., 2021). However, the Zn is recognized for its significant but dual impact on human health and the environment. While essential in small amounts for various biological functions, Zn is redefined as a potentially toxic element when it accumulates excessively due to natural and human activities. This reclassification is crucial as it underscores the importance of monitoring and managing Zn levels to prevent its detrimental effects, such as soil degradation and health risks, thereby highlighting the complex and pivotal role of Zn in environmental science and public health (Kemper and Sommer, 2002).

In contrast, As is a toxic heavy metalloid that occurs unsurprisingly in the environment, including soil, water, and air (Matschullat, 2000). As exposure can lead to numerous health problems, as well as cancer, skin lesions, cardiovascular disease, and neurological problems (Al-Ghouti et al., 2019). The risk of adverse health effects from As exposure depends on the level and duration of exposure, as well as other factors like age, health status, and genetics (Whitacre, 2008). Even though the element As is mostly found in rocks as a naturally occurring element, soils, and minerals, it also is the agent of man's action on the environment as mining, smelting, and industrial activities can release As into the environment. Pesticides and fertilizers that contain As can also contribute to the contamination of soil and water. As can also be released from wood preservatives, coal-fired power plants, waste incinerators, and the areas where the soil or rock contains high levels of As, the groundwater and surface water can become contaminated (Matschullat, 2000).

HM ions can be removed from water through various techniques, such as chemical precipitation, membrane processes, ion exchange, coagulation, and flotation. However, their high cost often limits these methods (Tao et al., 2023). Some techniques, such as adsorption, specifically using clay minerals, are cost-effective and environmentally friendly methods. Mathematical and conventional approach models are used to supplement laboratory experiments better to understand the relationship between HM properties and removal efficiency (Lu et al., 2022; Sergeev et al., 2019). Yet, conventional statistical models have limitations in computing interactions between elements and analyzing spatial aspects of pollution sources, and predicting the accuracy of HM. Various regressions-based prediction techniques have been suggested to tackle the problems associated with controlling soil and water pollution caused by HMs (Yang et al., 2021). These methods aim to reduce the costs involved in monitoring and provide advance warnings of HMs pollution. To achieve this, developing advanced soft technologies that incorporate computer-aided decision-making is necessary. One such technology is the use of machine learning (ML) models and smart Internet of Things (IoT), which have gained a lot of attention lately for simulating HMs pollution (Azizi et al., 2022). Predictive models can help minimize the cost and resource burden associated with environmental science and engineering, including workforce, finances, time, and space requirements (Meshram et al., 2021; Parsaie et al., 2021; Thomann, 1998). ML models can help to comprehend the unreliable nonlinear patterns of HMs removal using diverse treatment approaches, as compared to traditional models such as mathematical, physical, and empirical models (Azizi et al., 2022; Hafsa et al., 2020; Hu et al., 2020; Palansooriya et al., 2022). However, several factors, including the quantity and type of data available, catchment features, optimization prior to model development, weight minimization, and selection bias in the optimization algorithm used, might affect how well these prediction models for HMs function (Bhagat et al., 2021).

As more data becomes available and computing power increases, ML models will become even more widespread. Numerous urban areas across the globe have experienced varying degrees of soil HMs, as evidenced by studies conducted by Manta et al. (2002) and Luo et al. (2012). These studies have concluded that as industrialization continues to grow, the accumulation of soil HMs in polluted cities also increases, implying a correlation between soil HMs and the process of urban expansion. Zhang et al. (2020) used RF, SVR, and ANN to explore spatial prediction of HMs concentration, including Zn, As, Pb, etc., in China, with the RF model outperforming ANN and SVR models. Similarly, Tan et al. (2021) employed several decision tree algorithms in China to predict HMs concentration (i.e., As, Cr, Pb, and Zn), and the outcomes shown the superiority of MLs models. In 2022, Bazoobandi et al. explored the application of MLs models (ANN, ANFIS, and MLR) to estimate the Cd and Pb in Iran. The outcomes indicated the capability of ANN over other models (Bazoobandi et al., 2019). Recently several technical kinds of literature were established on HMs, showing the growing interest in the field examples (Chen et al., 2023; Sadek et al., 2023; Sari et al., 2023; Shi et al., 2023; Sun et al., 2023; Zhao et al., 2023). Despite several conducted research, the detailed review of HMs regarding methodology, concept, models, experimental and other complications of HMs referred to (Bhagat et al., 2020; Fan et al., 2018; Manta et al., 2002; Yaseen, 2021; Ye et al., 2020; Zhou et al., 2020).

While there have been many studies on using artificial intelligence (AI) based models for developing scenarios pertaining to HM elements, there are still many factors that need to be considered when it comes to the sustainability of these approaches. Presently, there is a deliberation centered around the drawback of conventional regression and traditional MLs methods, and the need for new optimization learning approaches, such as metaheuristic algorithms, is paramount (Yaseen, 2021, 2022; Yaseen et al., 2020). On the other hand, the impact of environmental variables, the types of data used for modeling HMs, the hyperparameters of the models, the types of HMs being simulated, performance metrics, and the architecture of the models themselves. While ML-based optimization has shown promising performance in HM modeling, the lack of thorough discussions about their reliability in terms of function, architecture, and algorithms has raised questions about their applicability compared to standalone AI-based and conventional models and has led to increased interest in a further investigation (Danandeh Mehr et al., 2018; Malik et al., 2021; Tao et al., 2022). In this regard, the primary motivation for using an AI-based heuristic approach in modeling HMs is to better understand and mitigate the risks associated with HMs (As (mg/kg) and Zn (mg/kg)) contamination in the environment, and to develop more effective strategies for monitoring and remediation. It is worth noting that, in this study, the gaps in modelling effectiveness, optimization challenges, reliability validation, strategies for environmental management, and methodological advances in predicting HM in marine sediments through the development and evaluation of hybridized ML models with metaheuristic optimization algorithms. Hence, the primary goal and objectives of this investigation is to explore the feasibility of different standalone ML models viz: Elman neural networks (ENN), decision tree based boosted algorithm (BT), and Bayesian relevance vector machine (RVM) for modeling As (mg/kg) and Zn (mg/kg) in marine sediments. This research suggests that existing standalone schema models are not effective for highly chaotic and nonlinear systems and data uncertainties caused by environmental instabilities. To overcome these weaknesses, the study introduces a new hybrid AI-based model that combines the RVM and RVM-FPA to estimate the As (mg/kg) and Zn (mg/kg) concentration. Additional attention was given to pre-processing techniques based on the nature of data driven-models.

2. Study area and sampling locations

The Bay of Algeciras is an ecologically important area that has been heavily altered by human activity. Commercial and industrial growth in the area has led to the establishment of important industrial parks in the northern part of the bay, which results in the release of pollutants, such as heavy metals, into the environment (González-Fernández et al., 2011). The local population generates a significant amount of urban sewage, which is discharged into the bay, and the city of Algeciras lacks wastewater treatment facilities (González-Fernández et al., 2011). The bay is also affected by maritime traffic and commercial trade, which can result in pollution from cargo and fuel spills and loading and unloading operations at the port. Urban and industrial wastewater discharges are received by the rivers Palmones and Guadarranque, which drain into the bay, acting as secondary sources of pollutants to the bay (Fig. 1) (González-Fernández et al., 2011). In December 2006, surface sediments from the top 20 cm layer were composed using a Van Veen grab sampler. The study used 34 sampling stations spread across ten shore-perpendicular transects to cover a large portion of the bay. In port facilities, there were two more sample stations. By combining a minimum of three Van Veen grab duplicates from each sampling station, one representative sediment sample was obtained. The results represent the mean value of the top 20 cm sediment layer. The samples were homogenized with a sterile Teflon spoon in plastic containers and kept refrigerated or frozen until analysis (González-Fernández et al., 2011).

3. The proposed data intelligent approaches

The data-driven models were inspired and developed efficiently by the amount of qualitative and quantitative instances. It is recommended to use measured experimental data or high-resolution reliable data for accurate decision making. In this study, the experimental data was adopted from (González-Fernández et al., 2011). We proposed different standalone models (ENN, BT, RVM) integrated with heuristic optimization (RVM-FPA) to simulate HMs (As, and Zn) subjected to severe anthropogenic activities. The soft computing model has been identified as a potential solution to reduce the significant cost and time associated with experimental-based HMs removal processes. The input-output data used for the development of models consist of ion (Fe), manganese (Mn), chromium (Cr), copper (Cu), nickel (Ni), molybdenum (Mo), silver (Ag), Cd, Pb, As, and Zn (see, Fig. 2). However, to accomplish the best possible achievement of the model, it is crucial that all factors related to the HMs removal process are appropriately incorporated into the input data optimization, to maximize the output and ensure that the model output is as close as possible to the experimental data (Usman et al., 2022).

3.1. Model building, data reliability and validation

The models were developed using a four-step approach. First, the ENN was trained on the dataset to capture complex relationships between the input features and the target variables (As and Zn concentrations). Then, the BT was applied to further refine the predictions, leveraging the strengths of decision trees and boosting techniques. Finally, the RVM was employed to provide probabilistic modeling and account for uncertainty in the predictions. Subsequently, heuristic optimization (RVM-FPA) was used to improve prediction skills. To validate the results, an accurate evaluation process was adopted. A method called stratified *k*-fold cross-validation was used with *k*-subsamples of equal size. This involved randomly dividing the original set of 38 data samples into four equal-sized subsamples, with each subsample being used as validation data once ($k-1$) while the other three were used for training the model. This process was constant four times i.e., k (4) folds, with each subsample being used as the validation data exactly once, and the results from each validation were combined to produce a single estimation result ($k-1$ $(4-1)$). This approach has the benefit that each data sample is used for validation exactly once, while all samples are used for training and validation. The calibration and verification were generated using 70 % and 30 % respectively, other feasible divisions could also employ in the same manner. The development of

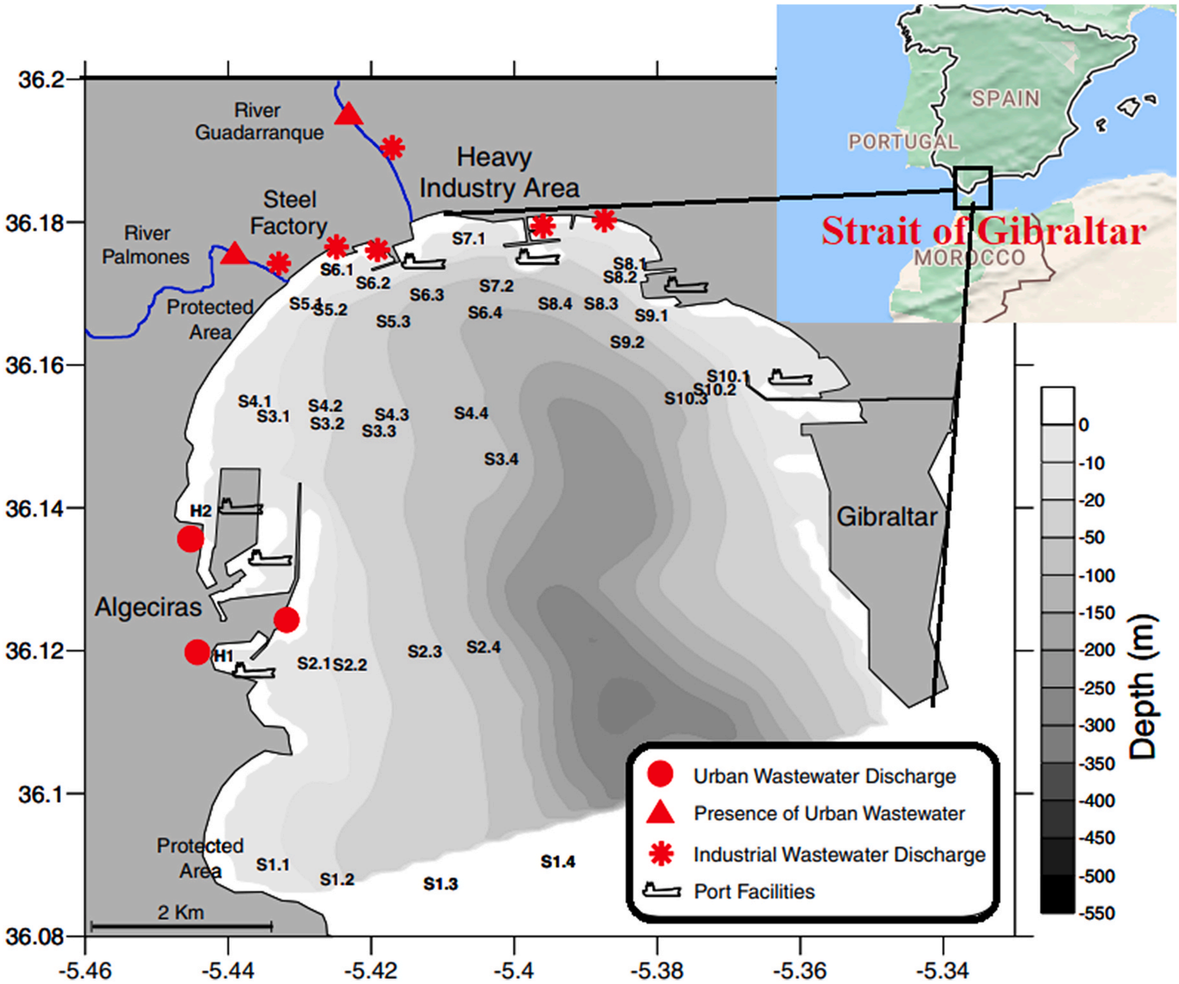


Fig. 1. Study area and sample locations in the Bay of Algeciras (González-Fernández et al., 2011).

models for As (mg/kg) and Zn (mg/kg) were built using feature engineering-based input variables selection. This prevents overfitting and underfitting and ensures that the model is able to make accurate predictions on new, invisible data and generalizations. The use of different evaluation performance metrics is essential to assess the accuracy and precision of predicted models. The results were also validated with the current state-of-the-art literature.

In this study, we employ a comprehensive approach for modeling As (mg/kg) and Zn (mg/kg) levels in marine sediments, ENN, BT, and RVM. For this purpose, Fe, Mn, Cr, Cu, Ni, Mo, Cd and Pb are used as the input variables while the corresponding target variables are As (mg/kg) and Zn (mg/kg). This modeling strategy aims to enhance an understanding and prediction accuracy of As and Zn concentrations in marine sediment samples, contributing to improved environmental monitoring and management efforts. The data was collected as stated above, organized, pre-processed and validated using several techniques. Indeed, it's crucial to discuss the practical implications of the technology developed and understand how these models can be applied in real-world scenarios and their potential benefits. As such, the predictions from models informed environmental management decisions, such as identifying areas with high As and Zn concentrations in marine sediments for targeted remediation efforts or providing early warnings for potential pollution events. Furthermore, it addressed the limitations and challenges of implementing these models in practice and potential future research directions.

Prior to model development, the study employed and recommended pre-analysis data reliability stability tests to ensure the best fit and regularity hypothesis of the data. For stochastic and intricate nonlinear data such as HMs, it is important to conduct a stationarity test of the variables, hence the current study employs both the ADF and PP tests to establish the correct data integration order. These tests should be used to compare stand-alone models, hybrid models, stand-alone approaches, and multivariate approaches. The significance of testing for stationarity in time series data, like HMs measurements in marine sediments, is critical for the reliability and effectiveness of predictive models. Stationarity implies consistent statistical properties over time, which is essential for accurate predictions, as most modeling techniques assume or require this stability. The study's use of Akaike (AIC) and Schwarz (SCI) information criteria alongside ADF and PP tests helps identify if the series is stationary. If data is non-stationary, it may contain trends or seasonal effects that could bias predictions, necessitating methods like differencing or detrending to stabilize the series before modeling (Abdulazeez et al., 2023). Addressing stationarity is significant for ensuring stable predictions, maintaining model validity, and understanding the underlying dynamics in the data. By ensuring the robustness of its predictive models through preliminary stability tests, the study lays a solid statistical foundation for the insights and strategies it provides, crucial for decision-makers and stakeholders in environmental management(Usman et al., 2023). Nevertheless, numerous factors can affect the effectiveness of predictive models, including data optimization, training process iterations, hidden layer quantity, normalization (Eq. (1)), training algorithm type, neuron quantity in hidden layers, initial weights, and transfer function type. To construct an MLs model that performs optimally and fits well with the intended purpose. It is

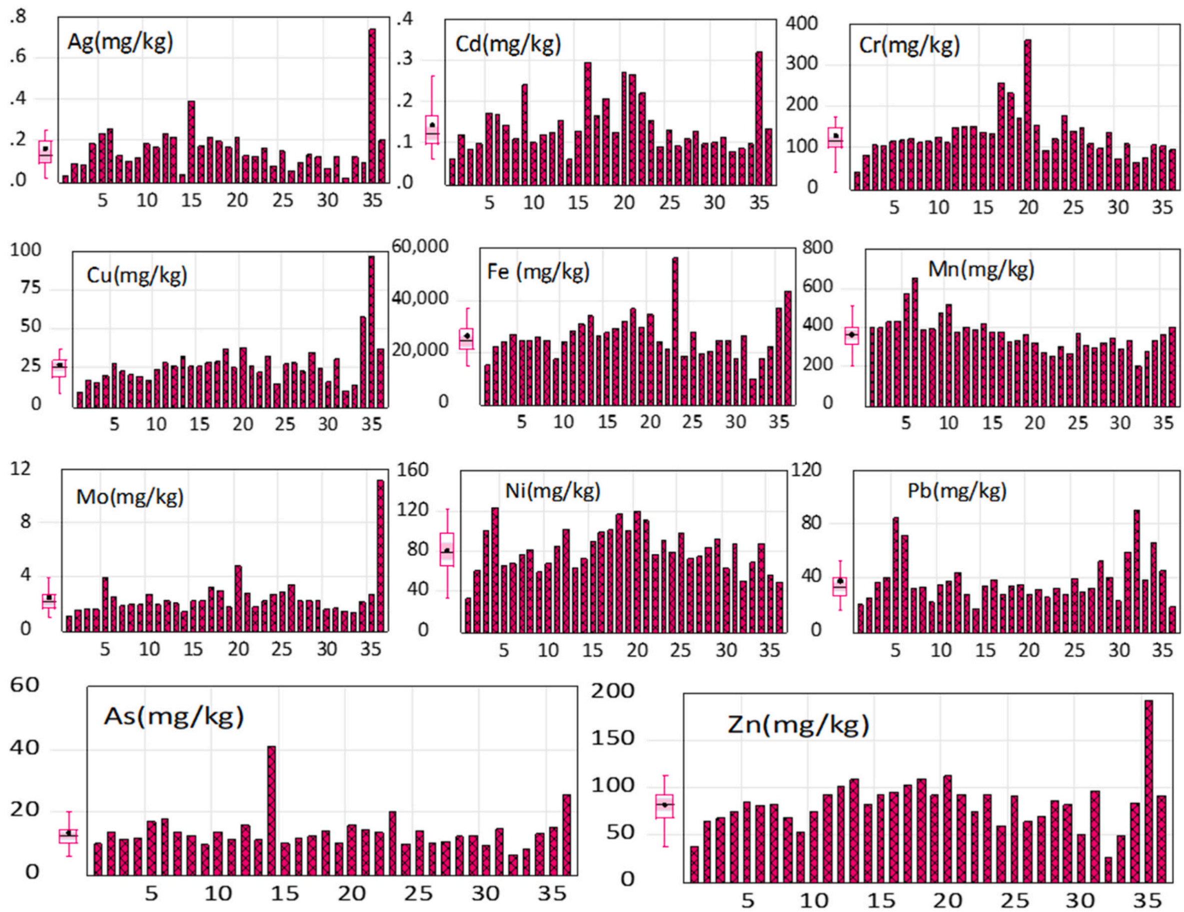


Fig. 2. Embedded bar-box plot for the raw data of input-output variables.

crucial to choose the right algorithm depending on the data properties and model design. The overall flowchart used in this study is presented in Fig. 3.

$$y = 0.05 + \left(0.95\left(\frac{x - \bar{x}}{x_{max} + x_{min}}\right)\right) \tag{1}$$

where y donates normalised data. x is the actual data, $\bar{x}$ is the mean of the measured data, x_{max} denotes the maximum value of the measured data, and x_{min} denotes the minimum value.

In this study, both standalone and hybrid models were compared using NSE (Nash–Sutcliffe efficiency), PCC (Pearson correlation coefficient), MAPE (mean absolute percentage error), MAE (mean absolute error), RMSE (root mean square error), and PBAIS to understand the strength of each model combination. Model evaluation criteria is an essential tool for assessing the performance of AI-based models in prediction tasks (Alamrouni et al., 2022; Alhaji et al., 2022; Legates and McCabe Jr, 1999; Yassin et al., 2022). Hence, these metrics can make better decisions about which models to use, how to optimize them, and how to advance the quality of the data used to train them.

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(HM_{(p)} - HM_{(o)}\right)^2}{\sum_{i=1}^{N}\left(HM_{(p)} - HM_{(o)}\right)^2} \tag{2}$$

$$PCC = \frac{\sum_{i=1}^{N}\left[HM_{(o),i} - \overline{HM_{(p)}}\right]\left[\widehat{HM}_{(o),i} - \widetilde{HM}_{(p)}\right]}{\sqrt{\sum_{i=1}^{N}\left[QHM_{(o),i} - HM_{(p)}\right]^2\left[\widehat{HM}_{(o),i} - \widetilde{HM}_{(p)}\right]^2}} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(HM_{(p)} - HM_{(o)}\right)^2} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{N}\left|HM_{(p)} - HM_{(o)}\right|}{N} \tag{5}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{N}\left|\frac{HM_{(o)} - HM_{(p)}}{HM_{(o)}}\right| \tag{6}$$

$$PBIAS = \frac{\sum_{i=1}^{N}\left(HM_{(o)} - HM_{(p)}\right)}{\sum_{i=1}^{N}HM_{(p)}} \tag{7}$$

whereby; $HM_{(p)}, HM_{(o)}, \overline{HM_o}$ is considered as the predicted, observed and averaged values, respectively for the HM element, where in this study referring to As (mg/kg) and Zn (mg/kg).

3.2. Data reliability and stationarity

It is also important to apply stability, normality, and stationarity tests to direct raw data, even though it increases computational costs and time (Musa et al., 2021). In the context of HM prediction in environmental science, data reliability is paramount, as the consequences of incorrect predictions can be detrimental to ecosystems and human health. It is essential to ensure the highest level of data reliability when developing AI models for HMs prediction. For this purpose, the

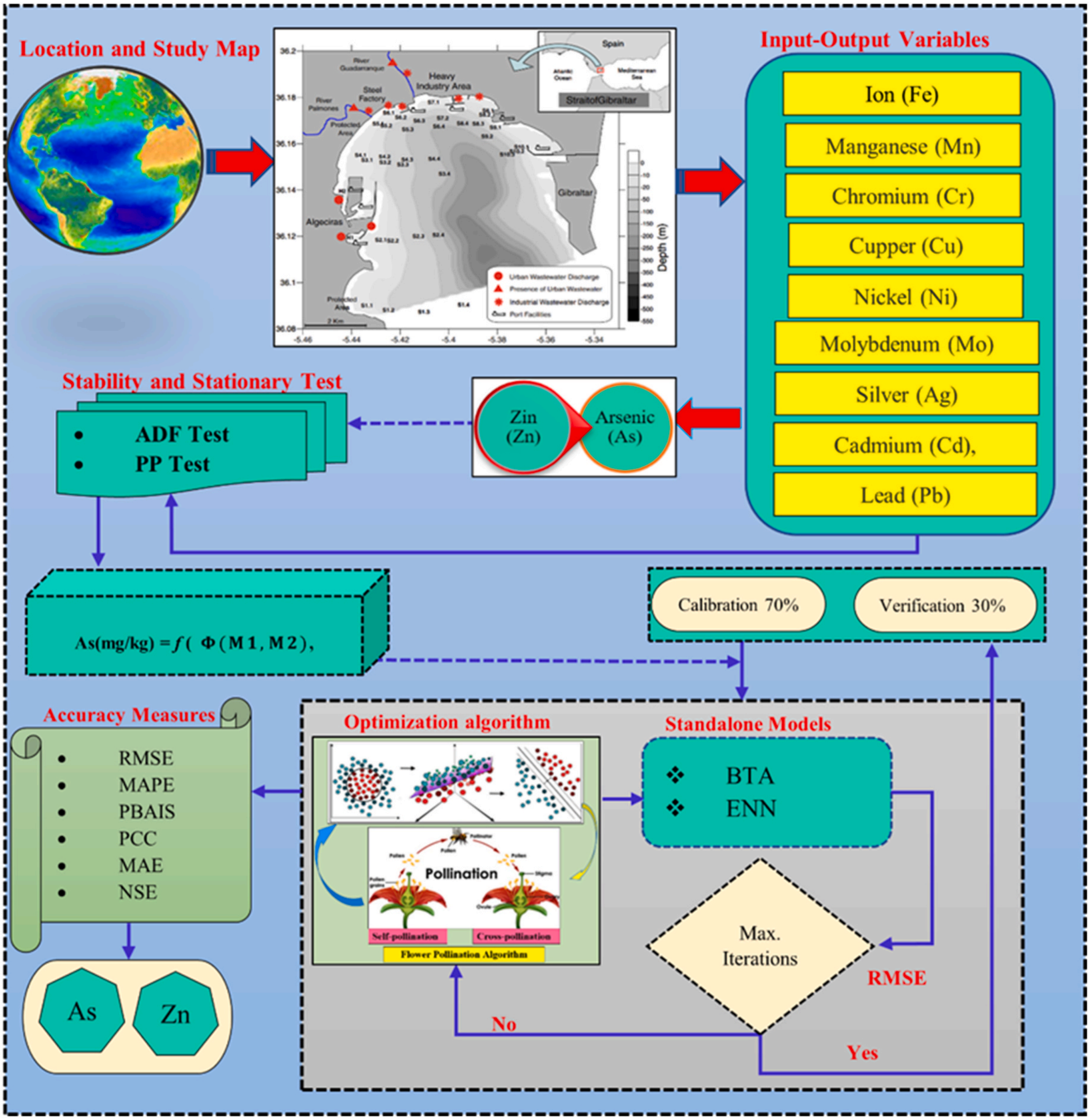


Fig. 3. Flow chart of the study.

Jarque-Bera test was used as a statistical test to check whether sample data have the skewness and kurtosis matching a normal distribution. In essence, it's a goodness-of-fit test specifically fitted for normal distributions. The Jarque-Bera test statistic combines these two measures and follows a chi-squared distribution with two degrees of freedom under the null hypothesis that the data are normal. The test showed that the parameters are not uniformly distributed except for Ni with a probability of 0.8733. The Jarque-Bera test suggests that the data do not significantly deviate from normality since the null hypothesis of normality is not rejected ($p > 0.05$) (Fig. 4). For stationarity test, which is most crucial for our sample data, the time series exhibiting statistical properties such as mean, variance, and autocorrelation that are constant over time. In Fig. 4, both the input and outputs (As and Zn) are stationary in nature with the P-value of 0.000 except with Ni as mentioned above. Both tests are concerned with the time-dependent structure of the data, not the distribution of the data values. So, a time series can be non-normal or not uniformly distributed but still be stationary if its statistical properties do not change over time. On the contrary, a time series could be normally distributed and still be non-stationary if its statistical properties do change over time (Alamrouni et al., 2022; Mati, 2021; Mati et al., 2019).

3.3. Relevance vector machine (RVM)

Relevance Vector Machine (RVM) is an ML algorithm used for regression and classification tasks. It is a probabilistic approach to MLs that uses Bayesian inference to evaluate the posterior distribution of the model variables (Bai et al., 2014). RVM is a sparse method that automatically selects the most relevant features and produces models with fewer parameters than traditional methods like SVM. The basic idea behind RVM is to find a function that maps the input data to the output data by minimizing a regularized error function (Lu and Zhou, 2023). The function is represented as a linear combination of basic functions that depend on the input variables. The parameters of the function and the relevance of the basic functions are learned from the data using a Bayesian approach. However an improved learning method is used to develop the RVM model to overcome the challenge of large matrix inversion in the training stage, making it easier to apply to large datasets (Biswas et al., 2019). Hyperparameters are used in the RVM regression framework to regulate the weights related with the input set, and their

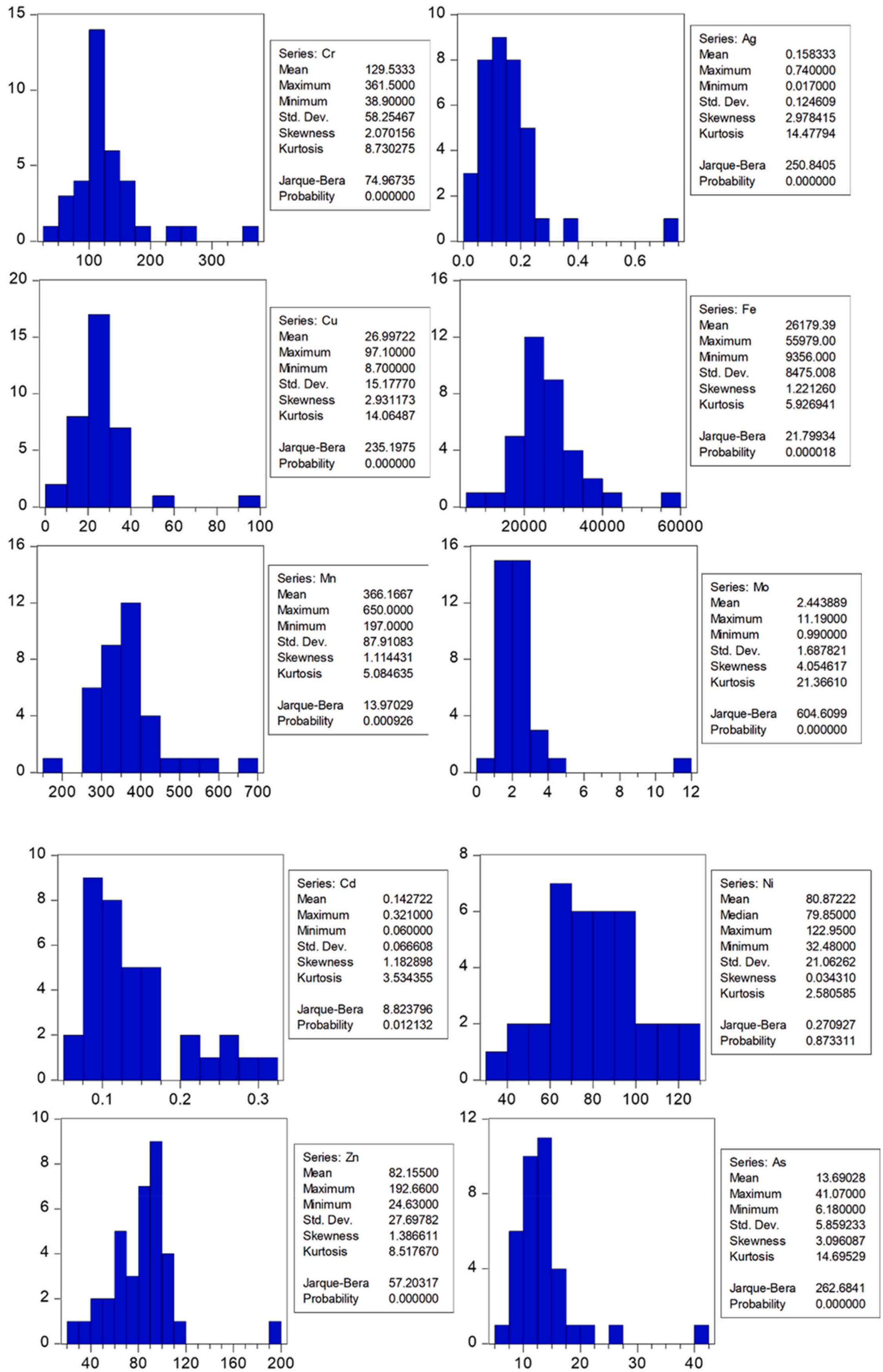


Fig. 4. Parametric analysis based on reliability and stationarity approach.

optimization is critical in building vigorous and precise models that avoid overfitting or underfitting. RVM has been successfully employed in numerous applications such as classification, communication, and prediction (Bonakdari et al., 2019; Chengyun et al., 2018).

3.4. Elman neural network (ENN)

The ENN, a subset of artificial neural networks (ANNs), consists of three layers: input, hidden, and output (Jia et al., 2019). Unlike other ANN techniques, Elman NN has the added advantage of incorporating feedback into its architecture, making it a popular choice for feedback NN. This feedback capability enhances the network's training ability and reduces the likelihood of failure due to local minima (Benaafi et al., 2022). The Elman NN is particularly useful for highly chaotic datasets and has excellent generalization capabilities. The feedback mechanism connects the performance of the hidden layers with their inputs, enabling the network to remember complex time-varying parameters based on historical data (Ren et al., 2018; Toha and Tokhi, 2008). One of the possible applications of ENN in agriculture and management practices is the prediction of Zn accumulation in soil (Huang et al., 2018). The author found that ENN could accurately predict the accumulation of Zn in soil under different management scenarios. Fig. 5 illustrates the ENN structure.

3.5. Boosted Tree algorithms (BTA)

Boosted Tree algorithms (BTA) are a type of ensemble learning method that combines multiple decision trees to improve predictive accuracy (Bokde et al., 2021). The basic idea is to iteratively train weak decision trees and then combine them into a single strong model that is capable of making more accurate predictions (Alnahit et al., 2022). The boosting process involves sequentially adding decision trees to the ensemble while adjusting the weights of the training examples based on their classification errors in previous iterations. This process helps the model to focus on the most challenging examples and gradually improve its performance. There are several variations of the BTA, but the most popular ones are Gradient Boosting and AdaBoost (Bokde et al., 2021) (see Fig. 6). Gradient Boosting uses gradient descent optimization to minimize the loss function while building each tree, whereas AdaBoost adjusts the weights of the training examples based on their classification errors and then trains the next tree to focus on the misclassified examples (Taffese et al., 2015; Umar et al., 2022). The BTA has several advantages over other MLs, including its ability to handle a wide range of data types and to handle missing values effectively. They also have high predictive accuracy and are relatively robust to overfitting. However, they can be computationally expensive and may require careful tuning of the hyperparameters to accomplish optimal performance (Lee et al., 2017).

3.6. Flower pollination algorithms (FPA)

The FPA is a nature-inspired optimization algorithm that simulates the pollination behavior of flowering plants. It was first introduced by Yang in 2012. The algorithm starts by generating an initial population of candidate solutions, which are represented as a set of n-dimensional vectors (Yang, 2012). These solutions are then evaluated based on a fitness function that measures their quality. One of the key advantages of the FPA algorithm is its ability to handle optimization problems with complex search spaces and non-linear, non-convex functions. It has been successfully applied in a diversity of domains, comprising engineering, finance, and machine learning. In nature, plants have evolved to optimize their pollination process to ensure the survival of the fittest (Chiroma et al., 2015). The FPA algorithm mimics this process by representing candidate solutions as pollens and using cross-pollination and biotic mechanisms to search the solution space. The algorithm has as its starting point a single flower that only yields one pollen, which is a potential solution. Pollen movement is modeled using a Levy flight. The FPA has been successful in solving complex optimization problems in various fields, comprising engineering and finance (Abdel-Basset and Shawky, 2019; Gupta and Dalal, 2022). Fig. 7 shows the diagram of the modeling process for hybrid RVM.

The objective of this study was to optimize the parameters of the RVM model in order to increase its accuracy. The classical RVM model used a trial-and-error method to select parameters, but this method could not always find the optimal parameters for every dataset. To address this issue, the Flower Pollination Algorithm (FPA) was coupled with the classical RVM model to create a hybrid bio-inspired model that could find optimal parameters more effectively. The FPA algorithm was used to find optimized parameters and fix the membership functions, resulting in high accuracy predictions of health measures. In the hybrid RVM-FPA model, the population and location of each answer were modified by the algorithm and returned to the best location. The weight

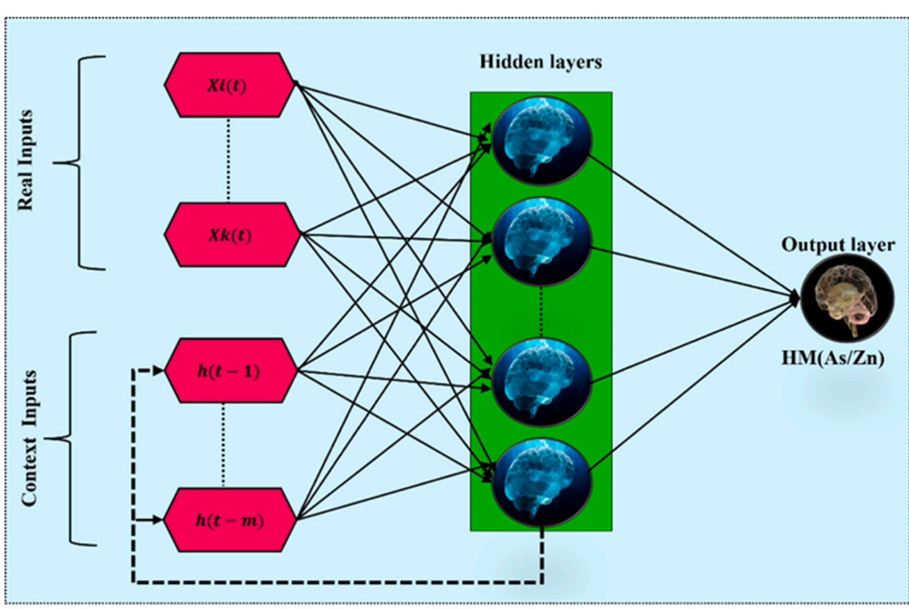


Fig. 5. Structure of Elman neural network.

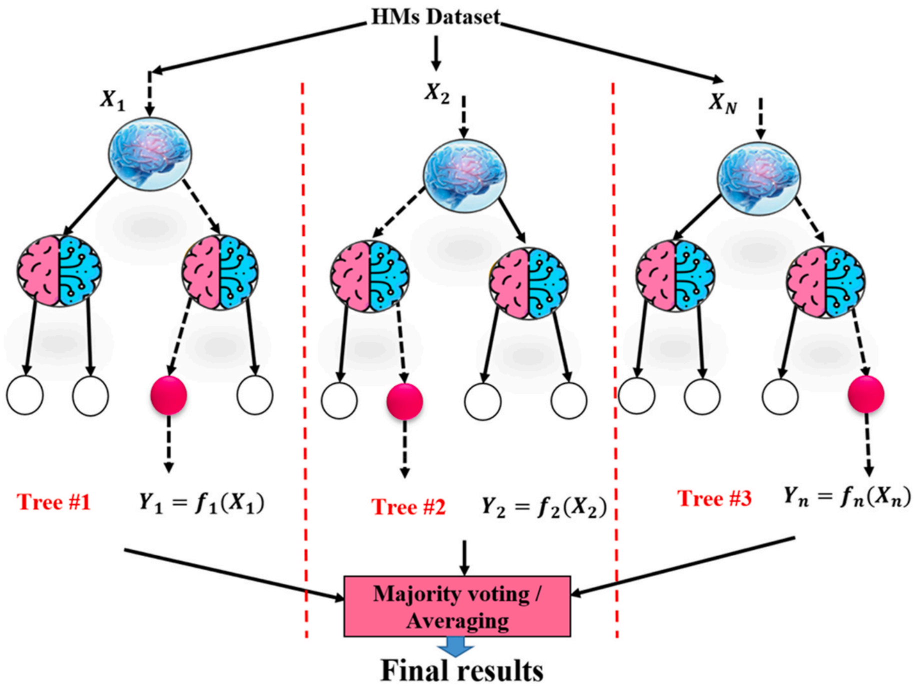


Fig. 6. Propose architecture of Boosted Tree algorithms.

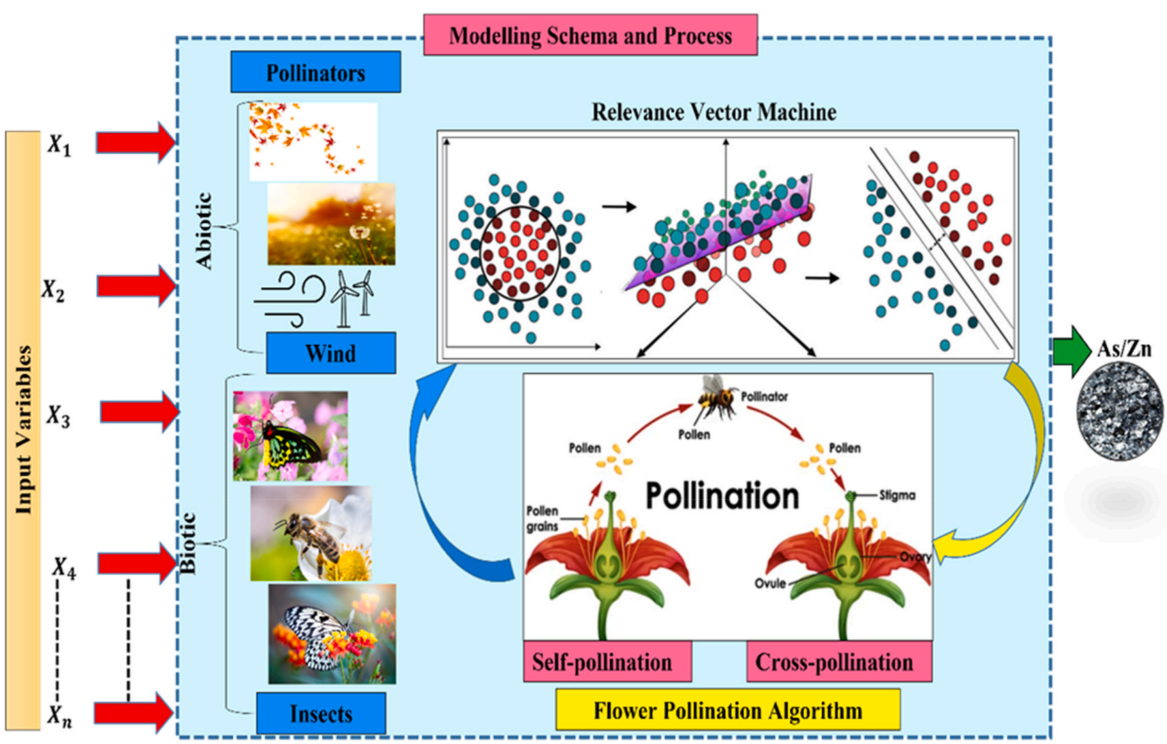


Fig. 7. Schematic diagram of the modeling process for hybrid RVM.

of the RVM was then efficient based on the optimal values achieved from the FPA algorithm. This approach showed promise in improving the accuracy of the RVM model. In each iteration of the algorithm, the candidate solutions undergo a pollination process that mimics the transfer of pollen between flowers. This process involves three main steps: randomization, pollen dissemination, local search. These steps are repeated until a stopping criterion is met, such as a maximum number of iterations or reaching a desired level of convergence.

4. Modeling results and discussion

The pollution of HMs, including As and Zn can have severe environmental and health impacts, including contamination of soil and water, harm to aquatic life, and adverse effects on human health, such as neurological damage and cancer. In the present day, advanced technologies such as AI and IoT are being used in the industrial sector to improve prediction and modelling of HMs pathways to minimize their impact to the ecosystem. Investigations have shown that MLs can be a useful tool for optimizing and controlling these processes. However, further work is needed to assess the economic viability of implementing such technology. This section discusses the outcomes of analyzing data, preparing data, evaluating sensitivity, and developing independent data-driven models.

4.1. Preliminary results

To accurately conduct the stochastic process and time series analysis when developing a model, it is important to ensure the stability and consistency of the dataset. The Augmented Dickey-Fuller (ADF) and Philip Perron (PP) analysis was used to test unit roots and ensure stationarity of all the HMs variables (see Table 1). This was done to achieve more reliable and valid results. Zhou et al. (2017) stability analysis is crucial to both computational intelligence and numerical analysis. Recently, stability analysis has been used in feature selection approaches, and linear feature selection approaches were employed in the present research.

The outcomes of the unit root tests are shown in Table 1. ADF and PP, reveal that all the variables are in a state of non-stationarity at level I (1) and I (2), except for Ag and Zn, which is stationary at level I (0). The method develops Schwarz info criterion with least square exogenous approach. The results of Mo attained the stationarity level at 2nd difference which surprising owing to it is correlation with As of 29 %. The results are supported by the fact that the probability and t-statistics values exceed the critical values for all levels of implication. According to Hair et al. (2010), a stationary test is considered valid when Cronbach's alpha values exceed 0.7. From Tables 1 and it is apparent that the mean and variance of all variables, except for Ag and As, are not constant, and therefore, we need to generate the first difference to achieve stationarity. It is important to note that ADF and PP enable analysts to choose appropriate modeling techniques and make more accurate predictions about future values of the series. Besides the ADF and PP, test input variables combination were determined using conventional corro-analysis. The results of this combination were presented in Eqs. (8) and (9). Based on auto-correlation function, 5 % and 10 % confident internal were set for As and Zn, respectively, creating two main combinations of M1 and M2 as presented in Fig. 8.

$$As\left(\frac{mg}{kg}\right)=\begin{cases} M_1=\Phi(Fe+Mo+Mn+Zn+Cu)\\ M_2=\Phi(Fe+Mo+Mn+Zn++Cr+Cu)\\ \qquad M=\text{stand for}\\ \text{model}\end{cases} \tag{8}$$

Table 1
ADF tests for input-outputs variables.

Variables	t-Statistic	5 % Critical Value	Prob	Decision
Fe (mg/kg)	−4.8	−2.9	0.0003	I (0)
	−10.1	−2.9	0.000	I (1)
Mn (mg/kg)	−2.6	−2.9	0.0855	I (0)
	−7.1	−2.9	0.000	I (1)
Cr (mg/kg)	−3.7	−2.9	0.0073	I (0)
	−6.9	−2.9	0.000	I (1)
Cu (mg/kg)	−4.1	−2.9	0.0031	I (0)
	−6.9	−2.9	0.000	I (1)
Ni (mg/kg)	−4.2	−2.9	0.002	I (0)
	−7.3	−2.9	0.000	I (1)
Mo (mg/kg)	−1.6	−2.9	0.460	I (0)
	−3.9	−2.9	0.0047	I (1)
	−7.7	−2.9	0.000	I (2)
Ag (mg/kg)	−5.6	−2.9	0.000	I (0)
Cd (mg/kg)	−4.8	−2.9	0.0004	I (0)
	−9.1	−2.9	0.000	I (1)
Pb (mg/kg)	−4.1	−2.9	0.0034	I (0)
	−6.0	−2.9	0.000	I (1)
As (mg/kg)	−6.0	−2.9	0.000	I (0)
Zn (mg/kg)	−4.8	−2.9	0.0004	I (0)
	−8.6	−2.9	0.000	I (1)

$$Zn\left(\frac{mg}{kg}\right)=\begin{cases} M_1=\Phi(Cu+Ag+Fe+Cd+Cr)\\ M_2=\Phi(Cu+Ag+Fe+Cd+Cr+Ni)\\ \qquad M=\text{stand for}\\ \text{model}\end{cases} \tag{9}$$

Fig. 9a displays the relationship between different HMs in marine sediments, with As as the target variable. In the context of this study, it can be seen that As offered a positive dependency with all the input variables, for instance, Fe (43 %), Mn (29 %), Cr (13 %), Cu (21 %), Mo (33 %), and Zn (27 %) to improve the predictive models for As concentrations in marine sediments. The values indicate varying degrees of linear relationships with As, with Fe showing the strongest correlation (0.4393) and Cr the weakest (0.1310). In predictive modelling, features (in this case, other HMs) with higher correlations with the target variable As may be more useful predictors. The positive dependency implies that as the concentration of one metal increases, the concentration of As also tends to increase. However, the strength of this tendency varies among the metals. For example, a higher presence of Fe is a stronger indicator of the presence of As than that of Cr.

A similar profile was obtained when Zn as the target variable. Cu has a strong positive correlation (0.6843), suggesting a likely increase in Zn concentrations with Fe. Meanwhile, Cu shows a strong positive correlation (0.8616), making it a highly predictive factor for Zn levels. The Ag also has a very strong positive correlation (0.837614), indicating its predictive value and Cr and Ni exhibit moderate (0.4132) and weaker (0.3096) positive correlations, respectively, indicating less predictive power. The Cd, with a correlation of 0.6304, suggests a strong relationship and could be a good predictor levels (Fig. 9b). These correlations are influential for the ML models discussed in the current study, which highlight the use of hybrid models combining RVM with meta-heuristic algorithms for enhanced prediction accuracy of HM concentrations, emphasizing the importance of such models in environmental management and monitoring strategies.

4.2. Results of data-intelligent algorithms

In this section, computational findings from the data source, preprocessing, and model development are described. To assess As and Zn performance in relation to other HMs, the study used BTA, popular neural network (ENN), and RVM models. The MLs learning models (BTA, ENN, and RVM) were created using the MATLAB 2022a toolbox, and the data pre- and post-processing was done using E-Views 11.0 software, respectively. MATLAB code was created to train and validate the model for ENN modelling. Choosing the best model structure is essential for attaining effective generalization. Hence, for the ENN, hypersensitivity strategies such as the maximum number of iterations (1000), learning rate (0.01), and MSE were utilized (0.0001). The most important factor in creating an ENN is establishing an adequate number of hidden nodes, due to the fact that these layers were discovered by utilizing the formulas (2n1/2 + m) to (2n+1), where n is the number of input neurons and m is the number of output nodes (Fletcher and Goss, 1993). While some investigations have indicated that the formula (n+1) to (n+ 2) is workable to prevent overusing trial-and-error uncertainty, others have suggested that concealed nodes should be in the shape of an oval-shaped structure. As a result, in this study, the optimal structure was found using the range of 2–10 hidden nodes, 20–80 calibration epochs, 1–20 emotional hormones, and activation functions. The BTA was employed with 10 *k*-folds cross validation folds, a prediction speed of 130 obs/sec, and training duration of 20.15 s to guard against overfitting by dividing the data set into fold and assessing accuracy on each fold.

The connections between the HMs parameters may be nonlinear in a complex process because of the unpredictable conditions caused by human activity and the characteristics of the source water that needs to be treated. Table 2 displays the BTA, ENN, and RVM's predicted calibration findings. The pragmatic AI-based and linear models responded

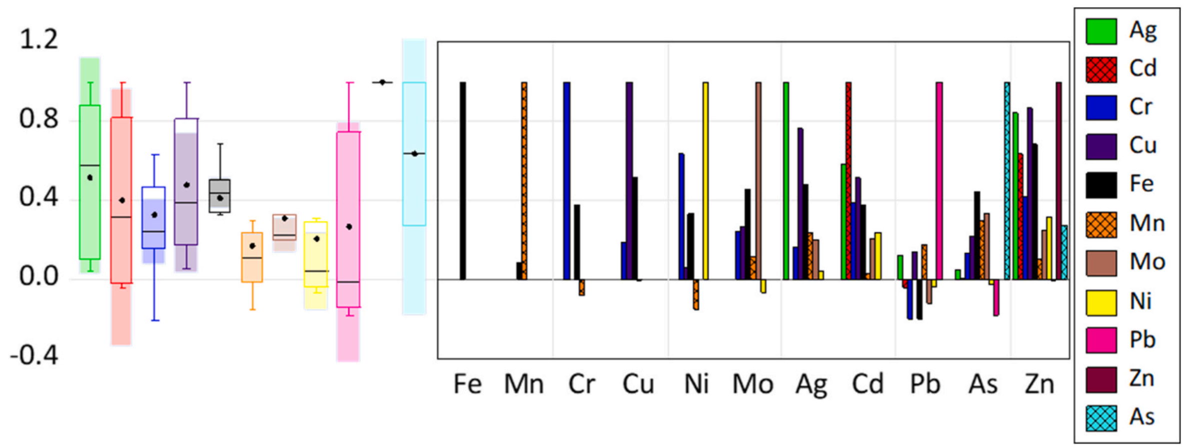


Fig. 8. ACF of input-output variables.

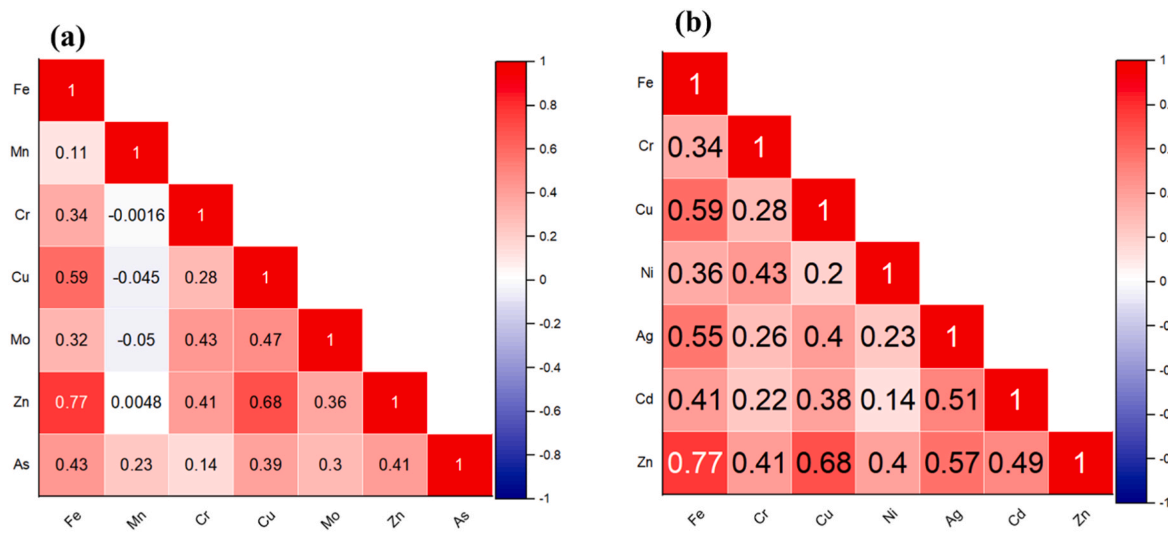


Fig. 9. Decency analysis based on inputs outputs parameters for (a) Ag (b) Zn.

to the various Combo Models As and Zn as shown in Table 3 in their remarkable fashion. Predictive power for As during the calibration phase was 50–60 %, 60–70 %, 70–80 %, 80–90 %, and 0.674–0.12 correspondingly for PCC and MAE. For PCC, NSE, and MAE, the Zn was calculated to be between 80 and 99 %, 90–99 %, and 0.0125 to 0.0322 correspondingly.

As seen in Fig. 10, the probability distribution frequency plots display how frequently values occur in a dataset, are used to comprehend the distribution of data. Looking at the plot makes it simple to spot any outliers, clusters, or peaks that may help to better understand the distribution of the HMs (As, and Zn) data. The numerical comparison indicated that all the models attained desirable accuracy in terms of error (RMSE) and bias (PBAIS). The best outperforming standalone models were RVM-M2 (RMSE = 0.0334, PBAIS = 0.0423) and ENN-M2 (RMSE = 0.0318, PBAIS = −0.04143) for As and Zn, respectively. When comparing the distributions of various variables, it is often important to compare distributions. The distributional differences and similarities can be clearly seen by plotting each variable on the same graph. The plot's ability to spot trends over time, which are utilized to analyze data over time, is another feature. To assess the performance of proposed model, the accuracy obtained is compared with the other outcomes from previous As and Zn modelling studies. Boudaghpour & Malekmohammadi (2020) developed different types of ANN algorithm for prediction of (As and Zn) with the accuracy range between 76 and 86 %. Our outcomes (97 %) were superior to their accuracy results despite reaching the same agreement regarding robustness of MLs. Furthermore, Zhang et al. (2021) obtained the accuracy of 76 % and 92 % for As and Zn, respectively using random forest (RF) which is inconsistent with our finding using BTA algorithm. This is due to the fact that RF as the family of BAT and decision tree are reported promising in the previous literature.

Fig. 11 provides visualization of the prediction accuracy in relation to the MAPE plots. The findings of both calibration and verification showed that estimating the As and Zn variables might be used to create computer-aided instruments for observing the effectiveness of the HMs. For apprehensive users, stakeholders, and decision-makers, the results demonstrated a straightforward modelling technique to accomplish many parameters to establish approximations of the As and Zn outputs. The ENN, RVM, and BTA techniques, like other single model algorithms, show a promising application in the field of engineering sciences as the variables involved in soil engineering hydro-environmental processes. To assess the accuracy of these models, the widely used MAPE was employed due to its independence from scale and ease of interpretation. A MAPE score between 0 and 10 % is generally considered reliable, and all models tested in this study fell within this range. The quantitative analysis of MAPE revealed that the EANN-M2 model improved upon the error values of another model by between 2 % and 7 %, respectively.

These results indicate that the ENN, RVM, and BTA technique has

Table 2
Results of the standalone models for the calibration and verification phases.

		Calibration Phase					
		NSE	PCC	MAPE	MAE	RMSE	PBIAS
As	BTA-M1	0.7257	0.8519	28.2928	0.0674	0.1263	0.0621
	BTA-M2	0.7823	0.8845	29.1294	0.0683	0.1280	0.0472
	ENN-M1	0.5994	0.7742	26.7429	0.0745	0.1772	0.1776
	ENN-M2	0.7715	0.8784	104.4710	0.2012	0.2498	−0.3308
	RVM-M1	0.8726	0.9341	27.1628	0.0735	0.1711	0.1598
	RVM-M2	0.8710	0.9333	26.5440	0.0737	0.1724	0.1777
Zn	BTA-M1	0.9393	0.9692	8.5730	0.0210	0.0274	0.0048
	BTA-M2	0.9421	0.9706	8.7077	0.0225	0.0271	0.0081
	ENN-M1	0.9556	0.9776	6.6274	0.0208	0.0255	−0.0160
	ENN-M2	0.9603	0.9800	4.4708	0.0125	0.0221	−0.0020
	RVM-M1	0.9453	0.9723	6.2100	0.0177	0.0234	0.0015
	RVM-M2	0.8369	0.9148	9.7067	0.0322	0.0439	0.0062
Verification Phase							
As	BTA-M1	0.7106	0.8430	17.4391	0.0418	0.0565	−0.0044
	BTA-M2	0.6967	0.8347	19.4722	0.0467	0.0620	0.0359
	ENN-M1	0.5378	0.7333	29.9853	0.0835	0.1720	0.2016
	ENN-M2	0.7274	0.8529	73.8471	0.1010	0.1177	−0.3419
	RVM-M1	0.9352	0.9671	14.6182	0.0229	0.0341	0.0423
	RVM-M2	0.9353	0.9671	14.7307	0.0246	0.0335	0.0465
Zn	BTA-M1	0.9001	0.9488	11.8434	0.0704	0.1422	0.1523
	BTA-M2	0.9095	0.9537	9.5342	0.0643	0.1410	0.1421
	ENN-M1	0.9823	0.9911	11.6929	0.0390	0.0458	0.0100
	ENN-M2	0.9717	0.9858	5.5008	0.0201	0.0318	−0.0143
	RVM-M1	0.9513	0.9754	7.8605	0.0244	0.0322	0.0043
	RVM-M2	0.8401	0.9165	12.7046	0.0427	0.0492	−0.0297

Table 3
Results of hybrid models for predicting As and Zn.

			Calibration Phase				
		NSE	PCC	MAPE	MAE	RMSE	BIAS
As	RVM-FPA-M1	0.9790	0.9894	0.0042	9.57E-06	1.8E-05	6.2E-06
	RVM-FPA-M2	0.9887	0.9943	0.0059	1.16E-05	2.0E-05	5.1E-06
Zn	RVM-FPA-M1	0.9987	0.9993	0.0094	2.68E-05	3.5E-05	−4.7E-06
	RVM-FPA-M2	0.9968	0.9984	0.0082	2.48E-05	3.2E-05	−1.9E-06
Verification Phase							
As	RVM-FPA-M1	0.9789	0.9893	0.0057	1.06E-05	1.4E-05	−1.8E-05
	RVM-FPA-M2	0.9867	0.9933	0.0057	1.04E-05	1.4E-05	−1.5E-05
Zn	RVM-FPA-M1	0.9899	0.9949	0.0120	4.49E-05	5.3E-05	1.13E-05
	RVM-FPA-M2	0.9779	0.9889	0.0119	4.33E-05	5.2E-05	4.4E-06

potential for accurate HMs predictions. The superior performance of RVM for As and ENN for Zn is likely due to their respective abilities to capture and model the complex, nonlinear patterns of each metal, influenced by the unpredictable conditions of the environment they are measured in. This suggested a need for modified modeling approaches based on the specific characteristics and behaviors of each HM in various environmental contexts. The BTA, ENN, and RVM, offer simplicity and direct insights with faster computation, making them user-friendly and easy to implement. However, they often achieved performance maxima and may not capture complex, nonlinear relationships effectively, limiting their optimization and overall predictive capabilities. In contrast, hybrid models like RVM-FPA provide higher predictive accuracy and robustness by efficiently handling complex data and dynamic conditions. They adapt and tune themselves to diverse datasets, overcoming the limitations of solo models. However, this increased accuracy comes with a trade-off: hybrid models are more complex, require longer computation times, involve a higher risk of overfitting, and demand accurate parameter tuning. The choice between the two depends on the task's specific needs, data nature, desired accuracy, and computational resources. While the hybrid model's superior performance is evident, it requires careful handling to harness its full potential without yielding to its complexities.

4.3. Results of hybrid heuristic algorithms

The second scenario of this work is using hybrid RVM-FPA to optimize the accuracy of single models. The RVM-FPA (M1 and M2) for As and Zn/associate the strengths of RVM model to provide more precise predictions by leveraging the strengths of different models. Table 3 shows that the optimum accuracy was achieved for both As and Zn (98–99 %). This is not unanticipated due to the fact that RVM-FPA could handle a wider range of input data types and formats and more robust to changes in the input data or changes in the environment. In scatter plots and a response plot, Fig. 12 contrasts the performance of the best-developed models with the observed data. As can be observed in Fig. 10, the response plot displays a strong agreement between the actual and expected value (NSE = 0.97–0.99). The MAPE and PBAIS values proved that response plot can display the contemporaneous agreement between two variables (measured and predicted As and Zn. The hybrid FPA are found to be more versatile, allowing for more flexible and adaptable solutions to complex problems such as HMs. The superiority of RVM-FPA over BTA, ENN, and RVM makes it a popular choice in HMs prediction especially for As and Zn.

It is evident, based on the overall discussion, that single models (BAT, ENN, RVM) and hybrid algorithm RVM-FPA are both techniques that can be used for solving optimization problems, despite their predictive differences. Regarding their strengths, standalone models are typically more precise than RVM-FPA. This is because most of them were based on a solid mathematical foundation and are often validated with experimental data. However, standalone models may only sometimes be suitable for complex optimization problems with a large number of variables or constraints, as they can be computationally expensive and time-consuming. In such cases, RVM-FPA may be a better choice, as they can handle more complex problems and are generally faster than standalone models. According to (Tao et al., 2024), hybrid models do not require prior knowledge of the problem or its parameters, making them ideal for problems with unknown or uncertain parameters. The

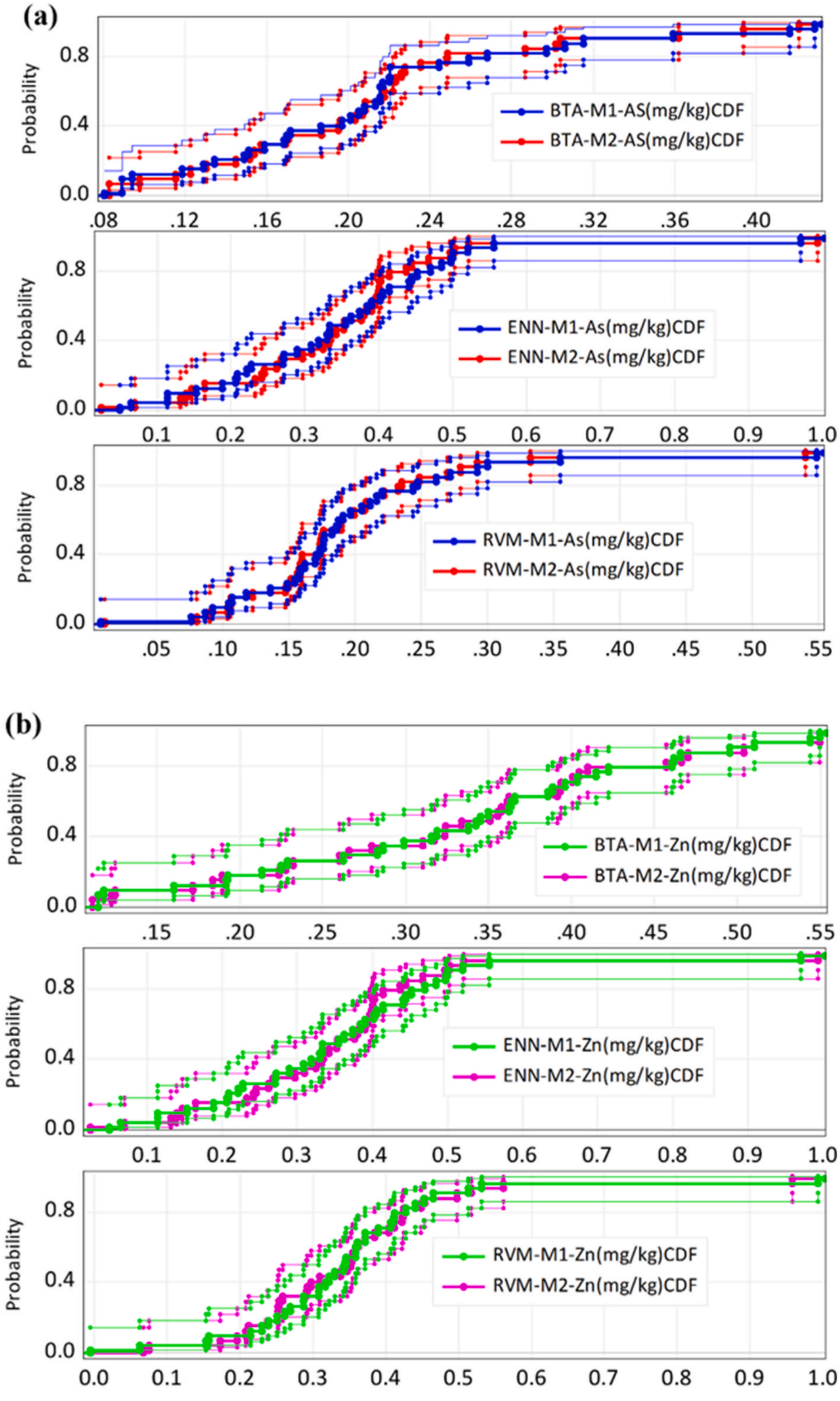


Fig. 10. Probability distribution function for (a) As(mg/kg) and (b) Zn(mg/kg).

overall comparison in terms of goodness-of fit is presented in Fig. 13. From Fig. 13, it is obvious that the accuracy of most of the standalone models was recommended. Generally, both the standalone model and hybrid FPA have their own unique strengths and weaknesses, and the choice of which technique to use depends on the specific problem at hand. Standalone techniques may be preferred for problems that require high accuracy and precision. At the same time, FPA algorithms may be better suited for complex problems with many variables or constraints (Abdel-Basset and Shawky, 2019; Yang et al., 2014).

The hybrid model, particularly the RVM combined with the RVM-FPA, demonstrated significantly improved prediction performance for both As and Zn compared to standalone models. This improvement is quantitatively evident in the predictive accuracy increase of 5 % for As and a substantial 18 % for Zn. The standalone models, while effective, were outperformed by the hybrid approach, which leveraged the strengths of both the RVM's robust modeling capabilities and the

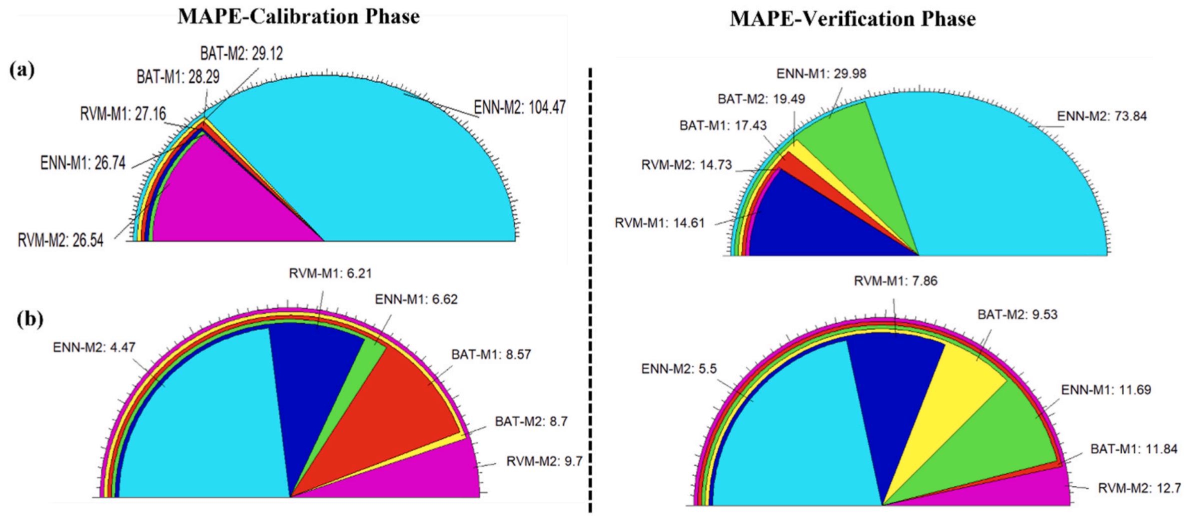


Fig. 11. Error based MAPE values for (a) As(mg/kg) and (b) Zn(mg/kg).

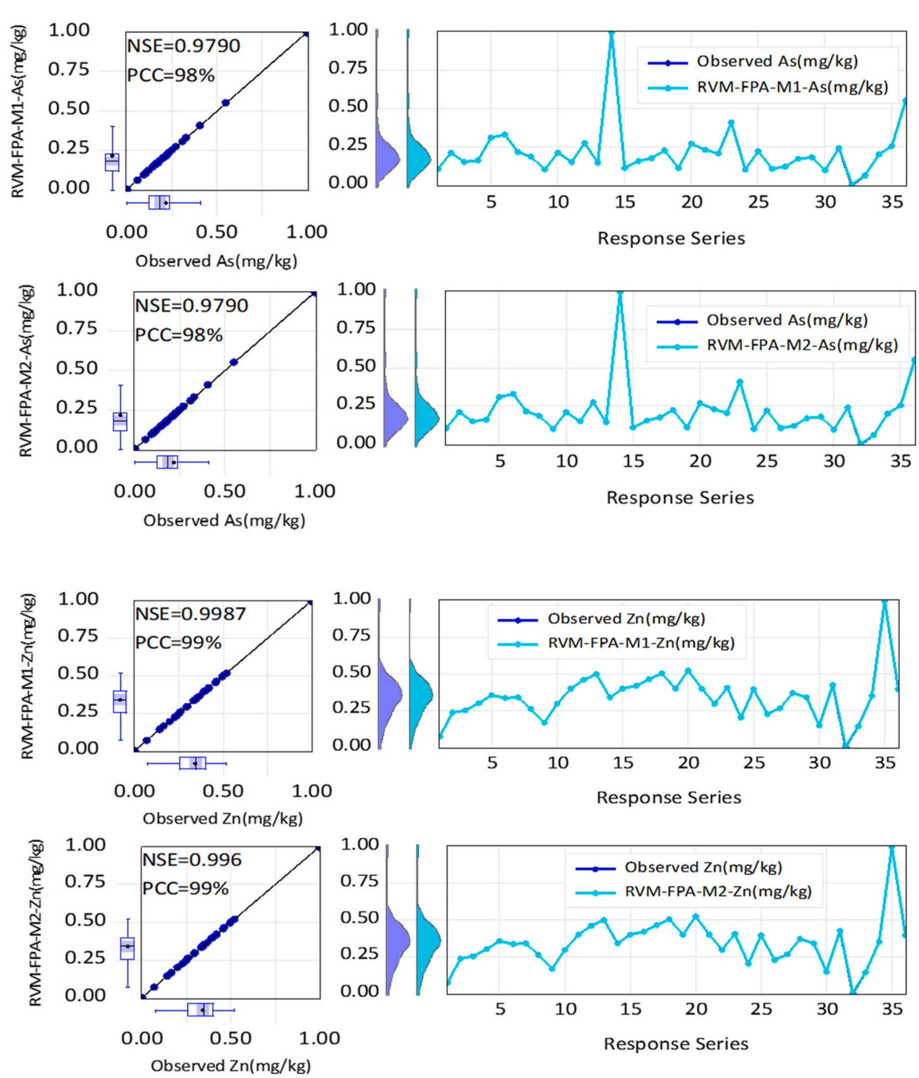


Fig. 12. Scatter plot and response series for hybrid FPA.

optimization prowess of the metaheuristic flower pollination algorithm. This combination allowed for a more understanding and modeling of the complex, nonlinear relationships inherent in the data, leading to superior predictive performance. The hybrid RVM-FPA model's ability to outperform standalone models emphasizes the potential of combining ML with metaheuristic algorithms for enhanced prediction in complex

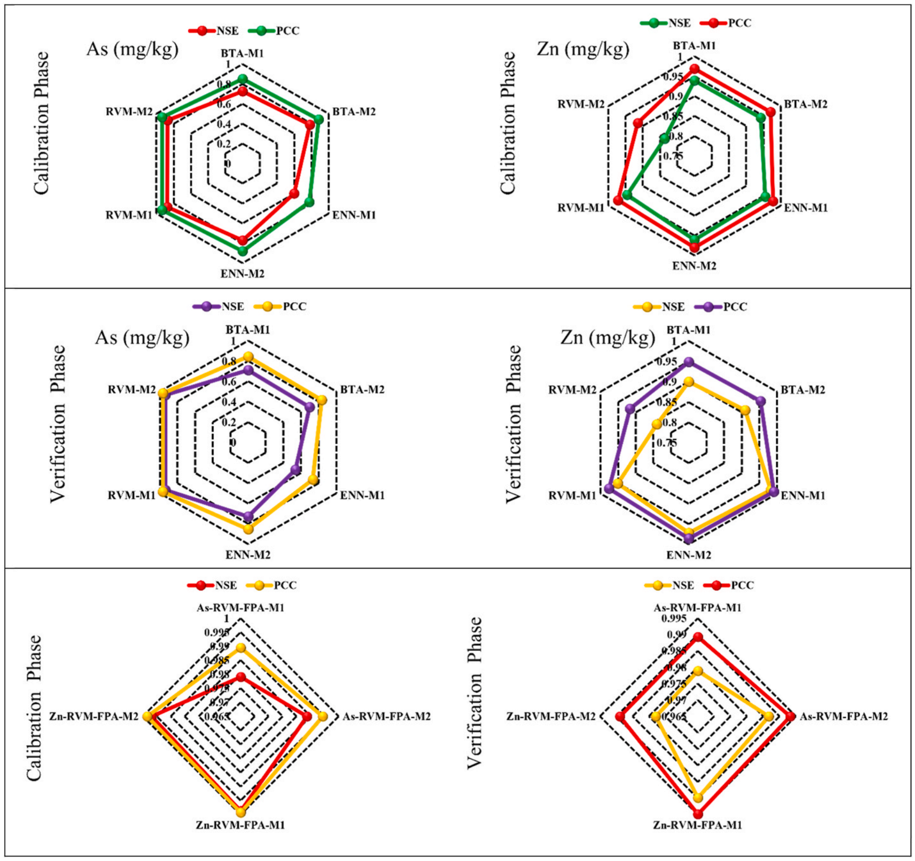


Fig. 13. Radar plot between the observed and predicted approaches for standalone and hybrid learning.

environmental systems.

5. Conclusion

According to Environmental Protection Agency (EPA), HMs monitoring is important to protect human health and the environment from harmful effects of toxic HMs contamination. Due to the progress made by standalone AI-based models applied to environmental engineering, particularly in the simulation of HMs, it is crucial to introduce a new heuristic algorithm (FPA) to measure its predictive feasibility. In this study, different standalone models (BTA, ENN, RVM) integrated with metaheuristic optimization algorithms (FPA) were established to simulate As and Zn in marine sediment due anthropogenic activities. The predictive performance was assessed using several indicators including NSE, PCC, MAPE, RMSE, MAE, and PBAIS to achieve the best simulated interpretation. The accuracy of single model attained within the marginal (53 %) and excellent (97 %) level with predominant laid within the good criteria. It was found that the RVM-FPA-M2 (MAPE = 0.0057, MAE = 1.04E-05) and RVM-FPA-M1 (MAPE = 0.0120, MAE = 4.49E-05) models had the closest agreement for As and Zn, respectively, indicating their strong prediction performance. The RVM-FPA model showed excellent prediction performance for HMs, and the study demonstrated the importance of integrating soft computing and HMs modeling. However, while AI-based models have promising abilities, they also have limitations. For example, linear models cannot capture highly nonlinear relationships, while black-box models like ENN, BAT, and RVM can be computationally costly and entail large amounts of data. The use of modeling systems in practical settings may be improved by integrating graphical presentations, database administration frameworks, and learning modules. Another suggestion was employing other pre-processing such as denoising techniques and nonlinear feature extraction to explore more on the predictability performance. While the current study has taken significant strides in using AI for environmental monitoring and simulation, the full potential of these technologies can only be realized by addressing the existing limitations and continually evolving the models to adapt to the complex and dynamic nature of environmental data.

CRediT authorship contribution statement

Zaher Mundher Yaseen: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Wan Hanna Melini Wan Mohtar:** Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft. **Raad Z. Homod:** Conceptualization, Investigation, Supervision, Visualization, Writing – original draft. **Omer A. Alawi:** Conceptualization, Investigation, Validation, Visualization, Writing – original draft. **Sani I. Abba:** Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Atheer Y. Oudah:** Conceptualization, Investigation, Validation, Visualization, Writing – original draft. **Hussein Togun:** Conceptualization, Investigation, Validation, Visualization, Writing – original draft. **Leonardo Goliatt:** Conceptualization, Investigation, Supervision, Validation, Visualization. **Syed Shabi Ul Hassan Kazmi:** Conceptualization, Investigation, Validation, Visualization, Writing – original draft. **Hai Tao:** Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgements

The authors appreciate the constructive comments given by the respected reviewers and editor. Also, the authors from KFUPM would like to thank the Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, Saudi Arabia for its support.

References

Abdel-Basset, M., Shawky, L.A., 2019. Flower pollination algorithm: a comprehensive review. Artif. Intell. Rev. 52, 2533–2557. https://doi.org/10.1007/s10462-018-9624-4.

Abdulazeez, I., Abba, S.I., Usman, J., Usman, A.G., Aljundi, I.H., 2023. Recovery of brine resources through crown-passivated graphene, silicene, and boron nitride nanosheets based on machine-learning structural predictions. ACS Appl. Nano Mater. https://doi.org/10.1021/acsanm.3c04421.

Al-Ghouti, M.A., Al-Kaabi, M.A., Ashfaq, M.Y., Da'na, D.A., 2019. Produced water characteristics, treatment and reuse: a review. J. Water Process Eng. 28, 222–239. https://doi.org/10.1016/j.jwpe.2019.02.001.

Alamrouni, A., Aslanova, F., Mati, S., Maccido, H.S., Jibril, A.A., Usman, A.G., Abba, S.I., 2022. Multi-regional modeling of cumulative COVID-19 cases integrated with environmental forest knowledge estimation: a deep learning ensemble approach. Int. J. Environ. Res. Public Health 19, 1–22. https://doi.org/10.3390/ijerph19020738.

Alhaji, U., Chinemezu, E., Isah, S., 2022. Bioresource Technology Reports Machine learning models for biomass energy content prediction : a correlation-based optimal feature selection approach. Bioresour. Technol. Rep. 19, 101167 https://doi.org/10.1016/j.biteb.2022.101167.

Alloway, B.J., 2008. Zinc in soils and crop production. Int. Fertil. Ind. Assoc. Paris 139.

Alnahit, A.O., Mishra, A.K., Khan, A.A., 2022. Stream water quality prediction using boosted regression tree and random forest models. Stoch. Environ. Res. Risk Assess. 4 https://doi.org/10.1007/s00477-021-02152-4.

Alshehri, F., Almadani, S., El-Sorogy, A.S., Alwaqdani, E., Alfaifi, H.J., Alharbi, T., 2021. Influence of seawater intrusion and heavy metals contamination on groundwater quality, Red Sea coast, Saudi Arabia. Mar. Pollut. Bull. 165, 112094 https://doi.org/10.1016/j.marpolbul.2021.112094.

Azizi, K., Ayoubi, S., Nabiollahi, K., Garosi, Y., Gislum, R., 2022. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. J. Geochemical Explor. 233, 106921 https://doi.org/10.1016/j.gexplo.2021.106921.

Bai, Y., Wang, P., Li, C., Xie, J., Wang, Y., 2014. A multi-scale relevance vector regression approach for daily urban water demand forecasting. J. Hydrol. 517, 236–245. https://doi.org/10.1016/j.jhydrol.2014.05.033.

Bazoobandi, A., Emamgholizadeh, S., Ghorbani, H., 2019. Estimating the amount of cadmium and lead in the polluted soil using artificial intelligence models. Eur. J. Environ. Civ. Eng. 0, 1–19. https://doi.org/10.1080/19648189.2019.1686429.

Benaafi, M., Yassin, M.A., Usman, A.G., Abba, S.I., 2022. Neurocomputing Modelling of Hydrochemical and Physical Properties of Groundwater Coupled with Spatial Clustering , GIS , and Statistical Techniques.

Bhagat, S.K., Pyrgaki, K., Salih, S.Q., Tiyasha, T., Beyaztas, U., Shahid, S., Yaseen, Z.M., 2021. Prediction of copper ions adsorption by attapulgite adsorbent using tuned-artificial intelligence model. Chemosphere 276, 130162. https://doi.org/10.1016/j.chemosphere.2021.130162.

Bhagat, S.K., Tung, T.M., Yaseen, Z.M., 2020. Development of artificial intelligence for modeling wastewater heavy metal removal: state of the art, application assessment and possible future research. J. Clean. Prod. 250, 119473 https://doi.org/10.1016/j.jclepro.2019.119473.

Biswas, R., Samui, P., Rai, B., 2019. Determination of compressive strength using relevance vector machine and emotional neural network. Asian J. Civ. Eng. 20, 1109–1118. https://doi.org/10.1007/s42107-019-00171-9.

Bokde, N.D., Ali, Z.H., Al-Hadidi, M.T., Farooque, A.A., Jamei, M., Maliki, A.A. Al, Beyaztas, B.H., Faris, H., Yaseen, Z.M., 2021. Total dissolved salt prediction using neurocomputing models: case study of gypsum soil within Iraq region. IEEE Access 9, 53617–53635. https://doi.org/10.1109/ACCESS.2021.3071015.

Bonakdari, H., Ebtehaj, I., Samui, P., Gharabaghi, B., 2019. Lake water-level fluctuations forecasting using minimax probability machine regression, relevance vector machine. Gaussian Process Regression, and Extreme Learning Machine. https://doi.org/10.1007/s11269-019-02346-0. Water Resour. Manag.

Bou Kheir, R., Greve, M.H., Abdallah, C., Dalgaard, T., 2010. Spatial soil zinc content distribution from terrain parameters: a GIS-based decision-tree model in Lebanon. Environ. Pollut. 158, 520–528. https://doi.org/10.1016/j.envpol.2009.08.009.

Boudaghpour, S., Malekmohammadi, S., 2020. Modeling prediction of dispersal of heavy metals in plain using neural network. J. Appl. Water Eng. Res. 8, 28–43. https://doi.org/10.1080/23249676.2020.1719219.

Chen, D., Wang, X., Luo, X., Huang, G., Tian, Z., Li, W., Liu, F., 2023. Delineating and identifying risk zones of soil heavy metal pollution in an industrialized region using machine learning. Environ. Pollut. 318, 120932 https://doi.org/10.1016/j.envpol.2022.120932.

Chengyun, Z., Rong, W., Qiaoying, T., 2018. Prediction model of dissolved oxygen based on SADE-RVM. ACM Int. Conf. Proceeding Ser. 131–134. https://doi.org/10.1145/3265639.3265661.

Chiroma, H., Shuib, N.L.M., Muaz, S.A., Abubakar, A.I., Ila, L.B., Maitama, J.Z., 2015. A review of the applications of bio-inspired Flower Pollination Algorithm. Procedia Comput. Sci. 62, 435–441. https://doi.org/10.1016/j.procs.2015.08.438.

Danandeh Mehr, A., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A.M.A., Yaseen, Z.M., 2018. Genetic programming in water resources engineering: a state-of-the-art review. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2018.09.043.

EPA, E.P.A., 2007. Framework for Metals Risk Assessment.

Fan, M., Hu, J., Cao, R., Ruan, W., Wei, X., 2018. A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence. Chemosphere 200, 330–343. https://doi.org/10.1016/j.chemosphere.2018.02.111.

Fletcher, D., Goss, E., 1993. Forecasting with neural networks. An application using bankruptcy data. Inf. Manag. 24, 159–167. https://doi.org/10.1016/0378-7206(93)90064-Z.

Ghadimi, F., 2014. Prediction of degree of soil contamination based on support vector machine and K-nearest neighbor methods: a case study in arak, Iran. Iran. J. Energy Environ. 5, 345–353. https://doi.org/10.5829/idosi.ijee.2014.05.04.02.

González-Fernández, D., Garrido-Pérez, M.C., Nebot-Sanz, E., Sales-Márquez, D., 2011. Source and fate of heavy metals in marine sediments from a semi-enclosed deep embayment subjected to severe anthropogenic activities. Water. Air. Soil Pollut. 221, 191–202. https://doi.org/10.1007/s11270-011-0782-0.

Gupta, S.K., Dalal, A., 2022. Optimisation of hourly plants water discharges in hydrothermal scheduling using the flower pollination algorithm. Int. J. Ambient Energy. https://doi.org/10.1080/01430750.2022.2142281.

Hafsa, N., Rushd, S., Al-Yaari, M., Rahman, M., 2020. A generalized method for modeling the adsorption of heavy metals with machine learning algorithms. Water 12, 3490. https://doi.org/10.3390/w12123490.

Hair, J., Black, W., Babin, B., Anderson, R., 2010. Multivariate Data Analysis, seventh ed. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Hu, B., Xue, J., Zhou, Y., Shao, S., Fu, Z., Li, Y., Chen, S., Qi, L., Shi, Z., 2020. Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. Environ. Pollut. https://doi.org/10.1016/j.envpol.2020.114308.

Huang, Y., Deng, M., Wu, S., Japenga, J., Li, T., Yang, X., He, Z., 2018. A modified receptor model for source apportionment of heavy metal pollution in soil. J. Hazard. Mater. 354, 161–169.

Jia, W., Zhao, D., Zheng, Y., Hou, S., 2019. A novel optimized GA–Elman neural network algorithm. Neural Comput. Appl. 31, 449–459. https://doi.org/10.1007/s00521-017-3076-7.

Kazemi, S.M., Hosseini, S.M., 2011. Comparison of spatial interpolation methods for estimating heavy metals in sediments of Caspian Sea. Expert Syst. Appl. 38, 1632–1649. https://doi.org/10.1016/j.eswa.2010.07.085.

Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. Environ. Sci. Technol. 36, 2742–2747. https://doi.org/10.1021/es015747j.

Lee, Sunmin, Kim, J.-C., Jung, H.-S., Lee, M.J., Lee, Saro, 2017. Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. Geomatics, Nat. Hazards Risk 8, 1185–1203.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241. https://doi.org/10.1029/1998WR900018.

Li, C., Zhou, K., Qin, W., Tian, C., Qi, M., Yan, X., Han, W., 2019. A Review on Heavy Metals Contamination in Soil: Effects, Sources, and Remediation Techniques. Soil Sediment Contam. https://doi.org/10.1080/15320383.2019.1592108.

Lu, C., Xu, Z., Dong, B., Zhang, Y., Wang, M., Zeng, Y., Zhang, C., 2022. Machine learning for the prediction of heavy metal removal by chitosan-based flocculants. Carbohydr. Polym. 285, 119240 https://doi.org/10.1016/j.carbpol.2022.119240.

Lu, M., Zhou, W., 2023. Quantitative analysis of Cr in soil using laser induced breakdown spectroscopy combined with relevance vector machine method. In: First International Conference on Spatial Atmospheric Marine Environmental Optics (SAME 2023). SPIE, pp. 83–91.

Luo, X.S., Ding, J., Xu, B., Wang, Y.J., Li, H.B., Yu, S., 2012. Incorporating bioaccessibility into human health risk assessments of heavy metals in urban park soils. Sci. Total. Environ. 424, 88–96.

Malik, A., Tikhamarine, Y., Sammen, S.S., Abba, S.I., Shahid, S., 2021. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-021-13445-0.

Manta, D.S., Angelone, M., Bellanca, A., Neri, R., Sprovieri, M., 2002. Heavy metals in urban soils: a case study from the city of Palermo (Sicily), Italy. Sci. Total. Environ. 300 (1–3), 229–243.

Mati, S., 2021. Do as your neighbours do? Assessing the impact of lockdown and reopening on the active COVID-19 cases in Nigeria. Soc. Sci. Med. 270, 113645 https://doi.org/10.1016/j.socscimed.2020.113645.

Mati, S., Civcir, I., Ozdeser, H., 2019. Ecowas common currency. Investig. Económica 78, 89–119.

Matschullat, J., 2000. Arsenic round the world: a review. Sci. Total Environ. 249, 297–312.

Meshram, S.G., Pourghasemi, H.R., Abba, S.I., Alvandi, E., Meshram, C., Khedher, K.M., 2021. A comparative study between dynamic and soft computing models for sediment forecasting. Soft Comput., 0123456789 https://doi.org/10.1007/s00500-021-05834-x.

Musa, B., Yimen, N., Abba, S.I., Adun, H.H., Dagbasi, M., 2021. Multi-state load demand forecasting using hybridized support vector regression integrated with optimal design of off-grid energy Systems—a metaheuristic approach. Processes 9. https://doi.org/10.3390/pr9071166.

Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., Ok, Y.S., 2022. Prediction of soil heavy metal immobilization by biochar using machine learning. Environ. Sci. Technol. 56, 4187–4198. https://doi.org/10.1021/acs.est.1c08302.

Parsaie, A., Haghiabi, A.H., Latif, S.D., Tripathi, R.P., 2021. Predictive modelling of piezometric head and seepage discharge in earth dam using soft computational models. Environ. Sci. Pollut. Res. 28, 60842–60856.

Ren, G., Cao, Y., Wen, S., Huang, T., Zeng, Z., 2018. A modified Elman neural network with a new learning rate scheme. Neurocomputing 286, 11–18. https://doi.org/10.1016/j.neucom.2018.01.046.

Sadek, A.H., Fahmy, O.M., Nasr, M., Mostafa, M.K., 2023. Predicting Cu(II) adsorption from aqueous solutions onto nano zero-valent aluminum (nZVAl) by machine learning and artificial intelligence techniques. Sustainability 15, 2081. https://doi.org/10.3390/su15032081.

Sari, M., Yalcin, I.E., Taner, M., Cosgun, T., Ozyigit, I.I., 2023. An investigation on environmental pollution due to essential heavy metals: a prediction model through multilayer perceptrons. Int. J. Phytoremediation 25, 89–97. https://doi.org/10.1080/15226514.2022.2059056.

Sergeev, A.P., Buevich, A.G., Baglaeva, E.M., Shichkin, A.V., 2019. Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. Catena 174, 425–435. https://doi.org/10.1016/j.catena.2018.11.037.

Shi, L., Li, J., Palansooriya, K.N., Chen, Y., Hou, D., Meers, E., Tsang, D.C.W., Wang, X., Ok, Y.S., 2023. Modeling phytoremediation of heavy metal contaminated soils through machine learning. J. Hazard Mater. 441, 129904 https://doi.org/10.1016/j.jhazmat.2022.129904.

Singh, J., Kalamdhad, A.S., 2011. Effects of heavy metals on soil, plants, human health and aquatic life. Int. J. Res. Chem. Environ. 1, 15–21.

Sinkakarimi, M.H., Rajei, G., Mahdijezhad, M., Hatamimanesh, M., 2020. Health risk assessment of some heavy metals in groundwater resources of birjand flood plain using environmental protection agency (EPA) model. J. Health 11, 183–193. https://doi.org/10.29252/j.health.11.2.183.

Song, Y.C., Sivakumar, S., Nguyen, T.T., Kim, S.H., Kim, B.G., 2009. The immobilization of heavy metals in biosolids using phosphate amendments-Comparison of EPA (6010 and 3051) and selective sequential extraction methods. J. Hazard Mater. 167, 1033–1037. https://doi.org/10.1016/j.jhazmat.2009.01.089.

Sun, Y., Chen, S., Dai, X., Li, D., Jiang, H., Jia, K., 2023. Coupled retrieval of heavy metal nickel concentration in agricultural soil from spaceborne hyperspectral imagery. J. Hazard Mater. 446, 130722 https://doi.org/10.1016/j.jhazmat.2023.130722.

Taffese, W.Z., Sistonen, E., Puttonen, J., 2015. CaPrM: carbonation prediction model for reinforced concrete using machine learning methods. Constr. Build. Mater. 100, 70–82. https://doi.org/10.1016/j.conbuildmat.2015.09.058.

Tan, K., Ma, W., Chen, L., Wang, H., Du, Q., Du, P., Yan, B., Liu, R., Li, H., 2021. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning. J. Hazard Mater. 401, 123288 https://doi.org/10.1016/j.jhazmat.2020.123288.

Tao, H., Abba, S.I., Al-Areeq, A.M., Tangang, F., Samantaray, S., Sahoo, A., Siqueira, H. V., Maroufpoor, S., Demir, V., Dhanraj Bokde, N., Goliatt, L., Jamei, M., Ahmadianfar, I., Bhagat, S.K., Halder, B., Guo, T., Helman, D.S., Ali, M., Sattar, S., Al-Khafaji, Z., Shahid, S., Yaseen, Z.M., 2024. Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: a comprehensive review, assessment, and possible future research directions. Eng. Appl. Artif. Intell. 129, 107559 https://doi.org/10.1016/j.engappai.2023.107559.

Tao, H., Al-Hilali, A.A., Ahmed, A.M., Mussa, Z.H., Falah, M.W., Abed, S.A., Deo, R., Jawad, A.H., Maulud, K.N.A., Latif, M.T., 2023. Statistical and spatial analysis for soil heavy metals over the murray-darling river basin. Chemosphere, 137914. Australia.

Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Salim, H., Sungwon, K., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., 2022. Groundwater Level Prediction Using Machine Learning Models: A Comprehensive Review. Neurocomputing.

Thomann, R.V., 1998. The future "golden age" of predictive models for surface water quality and ecosystem management. J. Environ. Eng. 124, 94–103.

Toha, S.F., Tokhi, M.O., 2008. MLP and Elman recurrent neural network modelling for the TRMS. In: 2008 7th IEEE International Conference on Cybernetic Intelligent Systems. IEEE, pp. 1–6.

Umar, I.K., Nourani, V., Gökçekuş, H., Abba, S.I., 2022. An Intelligent Hybridized Computing Techniques for the Prediction of Roadway Traffic Noise Based on Non-linear Mutual Information 5. https://doi.org/10.1007/s00500-023-07826-5.

Usman, A.G., Işik, S., Abba, S.I., 2022. Qualitative prediction of Thymoquinone in the high-performance liquid chromatography optimization method development using artificial intelligence models coupled with ensemble machine learning. Sep. Sci. Plus 1–9. https://doi.org/10.1002/sscp.202200071.

Usman, J., Salami, B.A., Gbadamosi, A., Adamu, H., GarbaUsman, A., Benaafi, M., Abba, S.I., Othman, M.H.D., Aljundi, I.H., 2023. Intelligent optimization for modelling superhydrophobic ceramic membrane oil flux and oil-water separation efficiency: evidence from wastewater treatment and experimental laboratory. Chemosphere, 138726.

Yang, H., Huang, K., Zhang, K., Weng, Q., Zhang, H., Wang, F., 2021. Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities. Environ. Sci. Technol. 55, 14316–14328. https://doi.org/10.1021/acs.est.1c02479.

Yang, Q., Li, Z., Lu, X., Duan, Q., Huang, L., Bi, J., 2018. A review of soil heavy metal pollution from industrial and agricultural regions in China: pollution and risk assessment. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2018.06.068.

Whitacre, D.M. (Ed.), 2008. Reviews of environmental contamination and toxicology, (Vol. 202).. Springer, New York, NY, USA.

Yang, X.S., 2012. Flower pollination algorithm for global optimization. In: International Conference on Unconventional Computing and Natural Computation. Springer, Berlin, Heidelberg, pp. 240–249. https://doi.org/10.1007/978-3-642-32894-7_27.

Yang, X.S., Karamanoglu, M., He, X., 2014. Flower pollination algorithm: a novel approach for multiobjective optimization. Eng. Optim. 46, 1222–1237. https://doi.org/10.1080/0305215X.2013.832237.

Yaseen, Z.M., 2022. The next generation of soil and water bodies heavy metals prediction and detection: new expert system based Edge Cloud Server and Federated Learning technology. Environ. Pollut. 313, 120081 https://doi.org/10.1016/j.envpol.2022.120081.

Yaseen, Z.M., 2021. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: review, challenges and solutions. Chemosphere 277, 130126. https://doi.org/10.1016/j.chemosphere.2021.130126.

Yaseen, Z.M., Faris, H., Al-Ansari, N., 2020. Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application. Complexity 2020.

Yassin, M.A., Tawabini, B., Al-Shaibani, A., Adetoro, J.A., Benaafi, M., Al-Areeq, A.M., Usman, A.G., Abba, S.I., 2022. Geochemical and Spatial Distribution of Topsoil HMs Coupled with Modeling of Cr Using Chemometrics Intelligent Techniques: Case Study from Dammam Area, Saudi Arabia. Molecules 27 (13), 4220.

Ye, Z., Yang, J., Zhong, N., Tu, X., Jia, J., Wang, J., 2020. Tackling environmental challenges in pollution controls using artificial intelligence: a review. Sci. Total Environ. 699, 134279 https://doi.org/10.1016/j.scitotenv.2019.134279.

Zhang, F., Yang, C., Zhu, H., Li, Y., Gui, W., 2020. An integrated prediction model of heavy metal ion concentration for iron electrocoagulation process. Chem. Eng. J. 391, 123628 https://doi.org/10.1016/j.cej.2019.123628.

Zhang, H., Yin, A., Yang, X., Fan, M., Shao, S., Wu, J., Wu, P., Zhang, M., Gao, C., 2021. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. Ecol. Indic. 122, 107233 https://doi.org/10.1016/j.ecolind.2020.107233.

Zhao, B., Zhu, W., Hao, S., Hua, M., Liao, Q., Jing, Y., Liu, L., Gu, X., 2023. Prediction heavy metals accumulation risk in rice using machine learning and mapping pollution risk. J. Hazard Mater. 448, 130879 https://doi.org/10.1016/j.jhazmat.2023.130879.

Zhou, T., Wang, F., Yang, Z., 2017. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. Water (Switzerland) 9. https://doi.org/10.3390/w9100781.

Zhou, X., Sun, J., Tian, Y., Lu, B., Hang, Y., Chen, Q., 2020. Hyperspectral technique combined with deep learning algorithm for detection of compound heavy metals in lettuce. Food Chem. 321 https://doi.org/10.1016/j.foodchem.2020.126503.