Application of Machine learning techniques in environmental governance: A review

Nanyi Peng
School of Life and Environmental Sciences, Minzu University of China, Beijing 100081, China;
18198253477@163.com

Abstract. In recent years, based on the continuous improvement of computer hardware performance, machine learning (ML) technology has made rapid development. Due to the strong ability of ML methods to find complex functions between associated data and the low cost of human and material resources, they are increasingly favored by environmental governance practitioners. Current research has confirmed that the application of ML technology to the field of environmental governance is of great significance in overcoming the difficulties of analysis and practice in traditional environmental work, and can greatly promote the long-term development of such directions as air quality prediction, solid waste classification, pollutant distribution mapping, and intelligent management of water environment. Therefore, by organizing the representative literature in the past five years, this paper describes the current status of the application of ML technology in four different elements of environmental governance, namely, atmosphere, water, soil, and solid waste, and aims to summarize the possible directions of the future development of ML technology and the new development brought to the field of environmental governance by exploring the basic concepts of ML, the advantages and disadvantages of the model and the constraints and other issues.

Keywords: machine learning; environmental governance; smart management; predictive modeling; artificial neural networks.

1. Background to the study

While the wealth of industrial society is growing rapidly, environmental problems are increasingly becoming a shackle on human development. With the rapid development of the country's economy, environmental problems have gradually emerged in China, and environmental management has been put on the agenda and achieved stage-by-stage results. However, air pollution and other problems brought about by it still occur in many areas. To avoid environmental degradation of ecological and human health damage, relying on the traditional human monitoring, regulation, governance methods have been unable to keep pace with the pace of economic development, we urgently need a new technical means to solve this problem.

ML is one of the branches of Artificial Intelligence (AI). With the rapid development of the field of artificial intelligence, ML methods have great potential for application. Unlike mathematical models or statistical methods, ML methods mainly emphasize the mapping relationship between the input data and the desired output response, without the need to identify the complex principle process mechanism. The process can be summarized as first building a standard database, then using a ML program to learn a general pattern of data, and finally building a model and attempting to predict the baseline true value. As the predicted value gets closer to the base truth, the more accurate the model becomes. Through such a continuous learning and optimization process, the ML model can obtain the same learning and problem-solving ability as the human brain [1]. The above process can be described in detail in Figure 1.

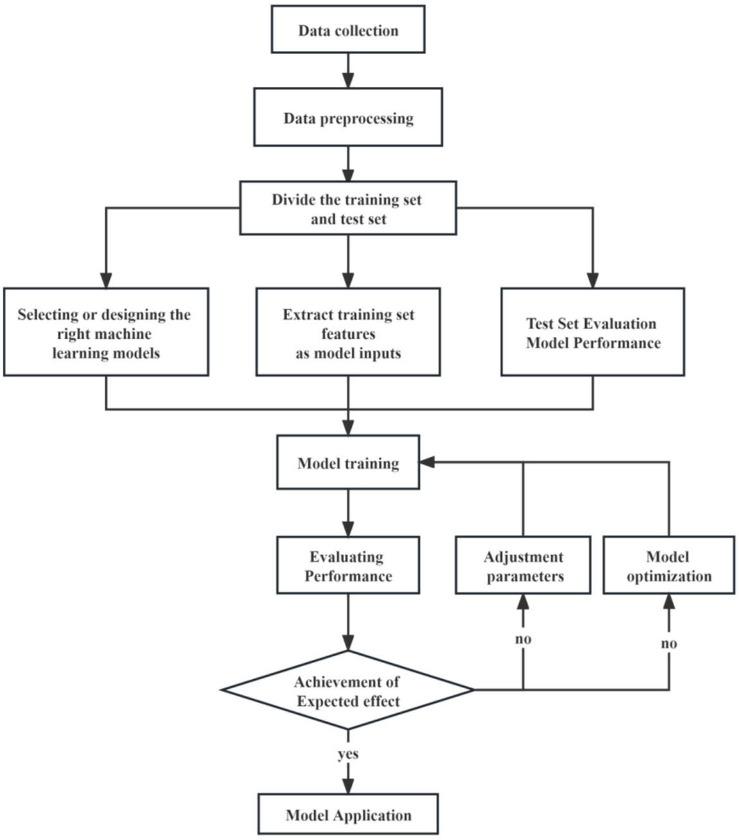

Fig. 1 Training process of ML model

From the perspective of learning methods, ML algorithms can be categorized into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Currently, the commonly used ML algorithms in the field of environmental governance are supervised learning, such as Artificial Neural Networks (ANN), Random Forests (RF), and Support Vector Machines (SVM) [2]. In terms of specific inputs to environmental governance research, a study explored the performance of ANN with an adaptive neuro-fuzzy inference system (ANFIS) in predicting heavy metal concentrations and found that it can have very low error values [3], when dealing with large number of nonlinear and complex datasets, ANN has a lower root mean square error (RMSE) and higher correlation coefficient of concordance (CCC) compared to the Random Forest Regression (RFR) and the Support Vector Machine (SVM) regression. Consistency Correlation Coefficient (CCC) [4].

The advantage of the application of ML models in the field of environmental governance is that it can accurately identify certain nonlinear relationships without a priori knowledge by learning rich historical data [5]. For example, ANN is good at fitting nonlinear relationships between variables, but the model is more complex and contains numerous hyperparameter adjustments [6].Support Vector Machines (SVMs) were initially used to solve binary classification problems, but nowadays kernel tricks are more often used to solve nonlinear regression problems [7]. In addition, tree-based models, including Random Forests (RF), Gradient Boosted Regression Trees (GBRT), and Extreme Gradient Boosting (XGB), are also popular in solving nonlinear problems.

However, despite the obvious application advantages of ML methods, there are many difficulties in its technical implementation. For example, traditional neural network algorithms based on the gradient descent method may face problems such as being trapped in local minima, low learning rates, and overfitting. The traditional SVM algorithm may be inefficient when observing too many samples.From the perspective of feature extraction, the implementation of ML methods often requires domain experts to design and extract features on raw data, and then go through tedious hyperparameter tuning to achieve high-quality prediction, which also poses a problem for accurate model building and wide application [8].

This paper summarizes the research of ML algorithms in environmental governance in four aspects, namely, atmospheric governance, water governance, soil governance, and solid waste governance, and discusses the limitations and shortcomings of ML algorithms and their branch derivatives in environmental governance, and points out the possible development prospects of this direction in the future, to providing a reference basis for the further research and application of ML algorithms in the field of environmental governance.

2. Research Overview

2.1 Applied to Atmospheric Management

The Atmosphere is an important element of the environment. Since the 20th century, with the increase of human industrial activities and the continuous use of fossil fuels, the problem of air pollution has become increasingly serious. In particular, the widespread haze triggered by pollutants such as carbon monoxide, nitric oxide, hydrocarbons, and volatile organic compounds in China around 2016 made air pollution a thorny environmental problem. Since then, with the explosive growth of air monitoring data and the increasing demand for fast and accurate prediction, ML-related models have been rapidly developed and applied to air pollution research, which has attracted the attention of the global scientific community [9].

In order to enhance atmospheric control, regular monitoring and analysis of air pollutant concentrations as well as prediction of future air quality is necessary. The basic process of predicting air quality is to combine historical data on atmospheric composition with up-to-date real-time data from ground-based sensors, and then create a more accurate model supported by suitable ML algorithms to predict where the data will go next. However, since the atmosphere is a complex and variable system containing a range of pollutants and complex chemical processes in which the relationships between pollution components and drivers are not simply linear, traditional linear-based statistical regression models, such as parametric regression models, moving averages, exponential smoothing, and autoregression, are no longer applicable in terms of prediction accuracy, handling of complex data, and keeping the model maintained and updated [10]. To solve this problem, Y.A. Li et al [11] proposed an adaptive neuro-fuzzy weighted limit learning machine (ANFIS-WELM) based on the fusion of a weighted limit learning machine (WELM) and an adaptive neuro-fuzzy inference system (ANFIS). They emphasized the application of nonlinear models in atmospheric concentration prediction.The experimental results show that the ANFIS-WELM model has good prediction accuracy and real-time performance, and multi-step time series prediction is realized on this basis. The study of Rajnish Lahoria et al also discarded the traditional linear statistical regression method, and they utilized the architecture based on the N-BEATS neural network, which fully considered the temporal variations among pollutants, meteorological conditions, and internal influencing factors, and realized the prediction of various air pollutants including NO_2, O_3, and CO. However, their study has some limitations in their results due to the lack of real-time wind data and failure to analyze wind-borne air pollution from other regions.

While ML models can continuously adaptively adjust to historical data in pursuit of higher overall prediction accuracy, the interpretability of these models is often neglected. It as a typical non-mechanistic model (e.g., ANN) does not effectively explain the interactions between various determinants and their effects on air pollution in terms of principles, which will have a significant impact on the credibility and reliability of the results in practical applications. Therefore, in the study of Gu Yuanlin et al [12] they proposed a new interpretable ML model to predict PM2.5. They integrated the feature generation and selection procedure into a hybrid model consisting of a neural network and a nonlinear autoregressive moving average model. The experimental results show that this novel hybrid model can explicitly reveal how PM2.5 prediction is predicted by historical data on weather and seasons, which points to a new direction for the next research.

ML technology has effectively advanced some of this work with its application advantages, and its application in atmospheric management is likely to continue for a long time in the future, as ML models are simpler and easier for new users. With the development of ML modeling, we can expect it to be applied to a wider range of research directions in the future and provide more details to researchers, such as the spatial vertical distribution of a certain pollutant [13].

2.2 ML applied to water management

As an important resource, water is related to the safety of human production and life and the stability of the ecosystem. However, due to the convenience of water bodies that are easy to dispose of waste, they are most likely to be directly and negatively affected by environmental pollution. Especially in recent years, various point and non-point source pollutions caused by frequent industrial, agricultural and human living activities have had malignant consequences on ecosystems and human health. After extensive research confirming that good results can be achieved by putting ML techniques into the study of water pollution and management, one of the hotspots of current research in the field remains the development of more reliable algorithms around ML models into different scenarios.

The assessment and prediction of the quality of water bodies is imperative for the effective management of water resources so that authorities can effectively implement certain interventions to ensure that pollution levels are kept within acceptable limits. The quality of a water body can reflect the pollution status and trends based on pollution indicators. Currently, ML-based methods have been developed for surface water monitoring to assess the water quality index (WQI) with different weights and a combination of physical and chemical factors, with a view to achieving an accurate assessment of polluted water quality and eutrophication, among other issues [14]. While for water quality monitoring of groundwater, El Bilali Ali et al. predicted parameters such as total dissolved solids (TDS), potential salinity (PS), sodium adsorption ratio (SAR), and exchangeable sodium percentage (ESP) of the water body by using physical parameters such as conductivity (EC), temperature (T), and pH as inputs through a ML approach while comparing the results of the Random Forest, Support Vector Regression (SVR), ANN, and Adaptive Boost (Adaboost) four models, and found that the validation performance and prediction performance of the ensemble model Adaboost model and Random Forest model were higher than the other models [15]. In addition, Aslan et al [16] also applied neural network models for water quality assessment in coastal areas and finally compared the assessment results and found that the standard recurrent neural network and long short-term memory (LSTM) models had the best performance. This statement is also confirmed by Lee et al [17], who also chose LSTM model as the best method in conducting experiments for predicting water quality in lakes, improving the feasibility of regional water quality management practices.

The use of ML models to simulate various parameters (e.g., Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Sodium Absorption Rate (SAR), pH, and Nitrate (NO_3) to obtain accurate predictive data is a common application scenario in water pollution research, which significantly reduces the cost of human inputs while achieving very promising results. In addition, the use of ML techniques in other conventional environmental management techniques(e.g.,gas chromatography-mass spectrometry (GC-MS)) for the calculation of complex and time-consuming water quality parameters has also enabled comprehensive and rapid water quality diagnosis [18].

As for the prediction of other data reflecting the deterioration of water bodies, Mesut Yilmaz et al [19] tested the ability of ML techniques to predict disease outbreaks in farms by modeling three popular ML techniques, namely logistic regression (LR), SVM, and plain Bayesian (NB), using the physicochemical properties of water and heavy metal content as the study object. The experimental results showed that all models produced successful results, with the most successful model being SVM, possessing 93.3% accuracy [20].A similar study was done by Yilmaz Mesut et al. who found that multinomial logistic regression analytical model predicted the likelihood of disease outbreaks with a higher accuracy of 95.65% [21]. In estimating water salinity metrics, ML regression algorithms can improve the performance of monitoring water salinity models by utilizing nonlinear transformations to describe the relationship between near-infrared spectral data and the target variable.Patthranit Wongpromrat et al. used a linear multivariate approach (Partial Least Squares Regression PLS) and nine ML methods (e.g., SVR, DT, RF AB, GB, BME, ERT, etc.) to develop a model for monitoring salt water pollution in rivers. The results showed that the accuracy of monitoring results of all ML models ranged between 0.87 and 0.97, and the relative percentage difference (RPD) ranged from 2.80 to 6.00. So far, all these models demonstrated high precision and accuracy [22].

However, one of the main drawbacks of ML techniques is their "black-box" structure: due to the internal non-interpretability of the model, there are no boundary conditions for model training, and prediction in this case may have serious negative impacts, and may even produce physically inconsistent predictions. In the study of pollutant transport in simulated water systems, Daan Bertels et al. [23] took into account the "black box problem" of ML techniques and proposed an interpretable ML model architecture based on LSTM units, which successfully implemented the principle of mass conservation in the model architecture. Compared with the established reference neural network model, this model shows higher accuracy.

In summary, the current research work on ML in water governance mainly focuses on water quality prediction, water salinity estimation, and parameter simulation based on ML models, and more challenges may be encountered in the future to widely carry out the pervasive application, which will also be the next step of the ML models in the water environment governance needs to be overcome.

2.3 ML applications for soil management

The application of ML to soil governance is less common compared to water and atmosphere. For environmental governance of soil, traditional laboratory methods for measuring ionic composition are the most accurate and reliable for assessing soil physicochemical property parameters (e.g., Table 1) and pollutant content, but their application is often limited by high cost and long time-consumption, and it is difficult to do in-situ measurements for long periods in the presence of a large number of soil samples. Considering that the prediction of soil ionic composition is a complex process, ML methods, in turn, can bypass the internal principle variables and provide simple solutions for nonlinear and multivariate functions, and have started to be used by more and more scholars in recent years [24].

Table 1 Common soil physical and chemical properties

Typology	Causality
Soil physical properties	thickness, subsoil gravel content, mechanical composition, texture, bulk density, density, sand content, mud content, clay content, total carbon content, total organic carbon content.
Soil chemical properties	organic matter content, carbon to nitrogen ratio, pH (H_2O), total exchangeable acids, soluble salt ion composition, CEC soil, Al_2O_3 content, Fe_2O_3 content, Cu content, Zn content, Fe content, Mn content, and B content.

Utilizing ML methods to manage soil and study soil contamination in different regions and scales is a current research hotspot in the field. On the one hand, studies have confirmed the superiority of ML methods in studying the spatial distribution characteristics of pollutants. Xiyue Jia et al [25] combined ML methods with satellite hyperspectral images to build ML models of random forest (RF), adaptive boosting, extreme gradient tree (xgboost), and gradient descent boosting (GDB) to predict soil arsenic content. The final Random Forest RF method presented the best performance (r = 0.78), with a 30% reduction in prediction error compared to the Geostatistical Interpolation Method (GIM) used for conventional measurements. This proves that the ML-based method is a high-precision and low-cost method for mapping arsenic contamination in brownfield soils. Similarly, Yifei Qiu et al [26], in comparing the accuracy and applicability of different ML models for predicting the spatial distribution of soil microplastics (MPs), found that the RF model was also the best choice

(R=9007.48), which better explained the magnitude of influence of impact factors on the occurrence of soil microplastics, and was able to combine the bivariate localized Moran's I index and the Normalized Difference in Vegetation Index (NDVI) trends to map the spatial distribution of soil microplastic pollution in the study area.

On the other hand, in the study of recognizing and determining soil properties, it was found that the ML models developed so far generally suffer from weak generalizability. Soil salinity models developed in one specific region often cannot be applied to another region due to differences in various soil properties (e.g., organic matter content or salt type) [27], which makes it very challenging to develop generalized soil prediction models with high accuracy. To address this issue, Ayan Das et al [28] used an integrated ML approach to extract features from hyperspectral data for soil electrical conductivity (EC) detection, and their innovative incorporation of stacked generalization in the experimental process enabled the model to overcome the generalizability limitation and obtain more robust predictions. Xiao Chao et al [29], on the other hand, compared the optimized input variable combinations of the generalization performance of Random Forest (RF), Support Vector Machine (SVM) and Extreme Gradient Boosting (XGB) models in predicting multiple soil characteristic parameters, such as potential salinity (PS), sodium adsorption rate (SAR), soil temperature (T), soil water content (SWC), and electrical conductivity (EC). The results show that the XGB model performs well when EC, SWC, and T are used as input variables, while the RF and SVM models perform well when EC, T, and pH are used as input variables, but overall the RF and XGB models show better generalization ability than the SVM model.

To summarize, the current application of ML in soil management mainly focuses on ML to identify soil pollution and draw spatial distribution maps. From the current research results, the efficiency of ML models depends not only on the prediction accuracy of the selected model itself, but also on the combination of input parameters and adopted models. In conclusion, ML methods have obvious advantages for improving the efficiency of identifying contamination and reducing the cost of laboratory tests. Even though the prediction of soil ions is still scarce and lacks specific evaluation and management, it is believed that it will be more favored by scholars in the future with the continuous improvement of ML techniques.

2.4 ML applied to solid waste management

The amount of solid waste in China is currently growing at a rate of 10% per year. General solid waste contains types of paper, plastics, textiles, glass, metals, rubber, organics, and sanitary and medical products, which are broadly categorized as recyclables and organics [30]. Globally, the average solid waste generated per person per day in towns and cities alone is 0.74 kilograms. Therefore, the disposal of municipal solid waste (MSW) has become a serious problem in the process of urbanization, and if MSW is not managed properly, it can contaminate the air, water, and soil, leading to serious health and environmental consequences. Currently, the continuous development of ML models has contributed significantly to the advancement of the computer vision field. Earlier studies [31] found that the classification of six wastes could be achieved using Support Vector Machine (SVM) and CNN models, but the achieved accuracy was low, 63% for SVM and only 23% for CNN. However, as model optimization schemes continue to be proposed, ML models show their unique potential for development.

In solid waste identification and classification, ML overcomes the shortcomings of traditional identification and classification techniques, such as long duration, complex operation, and high sample consumption. Based on the excellent characteristics of ML, Huanping Zhang et al [32] developed an optimized hybrid deep learning model for municipal waste classification. They used CNN (AlexNet) feature extraction, Deep Belief Network DBN waste prediction, Optuna hyper-parameter optimization, and finally tested the model with an R2 score of 0.94 and maximum permissible error MPE of 0.02. This optimized hybrid learning model significantly improves the performance of waste recognition and classification compared with the learner model alone. And Junyu Tao et al [33] proposed an approach based on the combination of hyperspectral imaging and

ML models to solve this problem. The results showed that ML helps to utilize and resolve the full range of hyperspectral patterns, especially the artificial neural network showed good classification and regression performance, and the model obtained from the final experiment had an accuracy close to 100% for the classification of organic/inorganic components of solid waste.

In addition, ML models have also made important contributions to the promotion of sustainable solid waste management. The main objective of sustainable solid waste management is to recycle solid waste globally, create jobs in various industries, and promote economic development. For example, MSW can be used as a fuel for energy plants, converting waste into electricity while reducing environmental impacts [34]. With timely access to waste availability information such as waste volumes, waste types, waste generation patterns, and disposal patterns, modern waste treatment systems can be mobilized in time to implement more accurate waste treatment mechanisms. Currently, advanced modeling-based methods are one of the major approaches for estimating waste quantities.ANN methods, support vector machines, and decision tree-based methods have been successfully applied to cross-sectional data, time series data, and panel data with much better performance than multiple linear regression MLR models. The study by Miyuru Kannangara et al [35] used a combination of ML methods and socioeconomic, and demographic parameters, eliminating the process of interpreting the model coefficients to understand the solid waste generation pattern, and the mapping results through the output showed that the error in predicting the generation of solid waste MSW was only 16-23%. Yadong Yang et al [36], based on historical experimental data, hoped to simulate the MSW gasification process with the help of the self-learning function of the ML model, and their experimental procedure, compared with that of Ascher et al's [37], fills the problem of the Ascher et al. model in which the catalyst characteristics are not taken into account. for the lack of consideration of catalyst properties, constructed a more inclusive ML model to predict and understand the gasification process of non-catalyzed MSW, and finally obtained the best model as the gradient boosted regressor GBR model, which was evaluated by the model with the results of $R^2 > 0.926$, RMSE < 6.318, and RRMSE < 0.304, which is a significantly considerable result.

The above results show that ML methods, supported by appropriate variables, can produce highly accurate models for waste classification, generation, and treatment. However, according to the current related research literature, the research of ML on solid waste is still very limited and very uneven, especially focusing on the identification and classification of solid waste, while there is still a lot of room for other aspects to be developed.

3. Summary and outlook

Throughout the above fields, the application of ML models is a key component of many research processes. Due to the generality and universality advantages of ML, its subdivided multiple modeling methods can be widely used for monitoring, predicting, classifying, and processing other environment-related tasks, specifically such as monitoring water pollution, predicting water quality, classifying solid wastes, and utilizing historical data to draw images of the pollution distribution, which have already achieved good results. Therefore, the prospects for the application of ML in the field of environmental governance are very good, and the work that can be continued in depth in the future can be summarized in the following aspects:

Construct a unified and integrated environmental information data platform and open the database sharing interface to solve the data source problems such as large pre-processing cost of model input data, lack of resources, and high difficulty of normalization. Efforts should be made to break the data barriers between different research institutions, universities, and organizations, and to jointly promote the rationalization process in the field of environmental governance with a more open stance.

We should continue to deepen the research of ML algorithms based on the perspective of data science, and update and optimize the models or introduce new algorithmic structures from the underlying logic and technical means to adapt to the more complex new problems in the field of environmental governance in the future. For example, more attempts can be made to innovate ensemble models, develop hybrid algorithms, based on the strengths and weaknesses of existing algorithms as much as possible to avoid the shortcomings, and give play to the accuracy, high efficiency, and generalizability of various algorithms.

Strengthen the integration and development of ML technology and traditional experiments. At present, the use of ML in environmental governance presents an uneven phenomenon, and most of the work is still in the exploration stage. Given the lack of data from some environmental experiments, the realization effect of ML models is significantly limited. If more effective extended information data can be obtained through enhanced computational simulation in the future, it will be an obvious improvement in the extensibility and generalization ability of ML.

References

[1] Liu Y, Yang Q, Li Y, et al. Application of machine learning in organic chemistry[J]. Chinese Journal of Organic Chemistry, 2020, 40(11): 3812.

[2] Zhong Shifa,Zhang Kai,Bagheri Majid,Burken Joel G,Gu April,Li Baikun,Ma Xingmao,Marrone Babetta L,Ren Zhiyong Jason,Schrier Joshua,Shi Wei,Tan Haoyue,Wang Tianbao,Wang Xu,Wong Bryan M,Xiao Xusheng,Yu Xiong,Zhu JunJie,Zhang Huichun. ML: New Ideas and Tools in Environmental Science and Engineering.[J]. Environmental science & technology,2021.

[3] Handan O U,Tuba B G,Ercan G, et al. Application of artificial neural networks to predict the heavy metal contamination in the Bartin River.[J]. Environmental science and pollution research international,2020,27(prepublish).

[4] Chen X, Zheng H, Wang H, et al. Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows[J]. Scientific Reports, 2022, 12(1): 12478.

[5] Deng Tianan,Chau Kwok-Wing,Duan Huan-Feng. ML based marine water quality prediction for coastal hydro-environment management[J]. Journal of Environmental Management,2021,284.

[6] Anna Pudełko,Marcin Chodak. Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods[J]. Geoderma,2020,368(C).

[7] Jiang H, Rusuli Y, Amuti T, et al. Quantitative assessment of soil salinity using multi-source remote sensing data based on the support vector machine and artificial neural network[J]. International journal of remote sensing, 2019, 40(1): 284-306.

[8] Zhang X, Shi Q, Wang B, et al. Review of machine learning algorithms in traditional Chinese medicine[J]. Computer Science, 2020, 45(11A): 5.

[9] SUN Yanan, FEI Jinhua. Precise governance of haze pollution based on machine learning[J].Resources Science, 2021, 43(5): 872-885.

[10] Rakholia Rajnish,Le Quan,Quoc Ho Bang,Vu Khue,Simon Carbajo Ricardo. Multi-output ML model for regional air pollution forecasting in Ho Chi Minh City, Vietnam.[J]. Environment international,2023,173.

[11] Yongan Li,Peng Jiang,Qingshan She,Guang Lin. Research on air pollutant concentration prediction method based on self-adaptive neuro-fuzzy weighted extreme learning machine[J]. Environmental Pollution,2018,241.

[12] Gu Yuanlin,Li Baihua,Meng Qinggang. Hybrid interpretable predictive ML model for air pollution prediction[J]. Neurocomputing,2022,468.

[13] Li Yunzhe,Sha Zhipeng,Tang Aohan,Goulding Keith,Liu Xuejun. The application of ML to air pollution research: A bibliometric analysis[J]. Ecotoxicology and Environmental Safety,2023,257.

[14] Ding Yang,Zhao Jinyong,Peng Wenqi,Zhang Jing,Chen Quchang,Fu Yicheng,Duan Maoqing. Stochastic trophic level index model: A new method for evaluating eutrophication state[J]. Journal of Environmental Management,2021,280.

[15] Ali B E,Abdeslam T,Youssef B. Groundwater quality forecasting using ML algorithms for irrigation purposes[J]. Agricultural Water Management,2020(prepublish).

[16] Aslan S, Zennaro F, Furlan E, et al. Extensive study of recurrent neural network architectures with a multivariate approach for water quality assessment in complex coastal lagoon environments: A case study on the Venice Lagoon[J]. 2022.

[17] Lee H W, Kim M, Son H W, et al. Machine-learning-based water quality management of river with serial impoundments in the Republic of Korea[J]. Journal of Hydrology: Regional Studies, 2022, 41: 101069.

[18] Lei Li,Shuming Rong,Rui Wang,Shuili Yu. Recent advances in artificial intelligence and ML for nonlinear relationship analysis and process control in drinking water treatment: A review[J]. Chemical Engineering Journal,2021,405.

[19] Yilmaz Mesut,Çakir Mustafa,Oral Mükerrem Atalay,Kazanci Hüseyin Özgür,Oral Okan. Evaluation of disease outbreak in terms of physico-chemical characteristics and heavy metal load of water in a fish farm with ML techniques[J]. Saudi Journal of Biological Sciences,2023,30(4).

[20] Al-Shari H , Saleh Y A , Odaba A .Comparison of Gradient Boosting Decision Tree Algorithms for CPU Performance[J].Erciyes Tip Dergisi, 2021(37):157-168.

[21] Yilmaz Mesut,Çakir Mustafa,Oral Okan,Oral Mükerrem Atalay,Arslan Tülin. Using ML technique for disease outbreak prediction in rainbow trout (Oncorhynchus mykiss) farms[J]. Aquaculture Research,2022,53(18).

[22] Wongpromrat Patthranit,Phuphanutada Jirawat,Lapcharoensuk Ravipat. Monitoring of salinity of water on the THA CHIN River basin using portable Vis-NIR spectrometer combined with ML algorithms[J]. Journal of Molecular Structure,2023,1287.

[23] Bertels Daan,Willems Patrick. Physics-informed ML method for modelling transport of a conservative pollutant in surface water systems[J]. Journal of Hydrology,2023,619.

[24] Wu Lifeng,Peng Youwen,Fan Junliang,Wang Yicheng,Huang Guomin. A novel kernel extreme learning machine model coupled with K-means clustering and firefly algorithm for estimating monthly reference evapotranspiration in parallel computation[J]. Agricultural Water Management,2020(prepublish).

[25] Jia Xiyue,Hou Deyi. Mapping soil arsenic pollution at a brownfield site using satellite hyperspectral imagery and ML.[J]. The Science of the total environment,2022,857(P2).

[26] Qiu Yifei,Zhou Shenglu,Zhang Chuchu,Qin Wendong,Lv Chengxiang,Zou Mengmeng. Identification of potentially contaminated areas of soil microplastic based on ML: A case study in Taihu Lake region, China[J]. Science of the Total Environment,2023,877.

[27] Kahaer Y, Tashpolat N. Estimating salt concentrations based on optimized spectral indices in soils with regional heterogeneity[J]. Journal of Spectroscopy, 2019, 2019.

[28] Das Ayan,Bhattacharya Bimal Kumar,Setia Raj,Jayasree G.,Sankar Das Bhabani. A novel method for detecting soil salinity using AVIRIS-NG imaging spectroscopy and ensemble ML[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2023,200.

[29] Xiao Chao,Ji Qingyuan,Chen Junqing,Zhang Fucang,Li Yi,Fan Junliang,Hou Xianghao,Yan Fulai,Wang Han. Prediction of soil salinity parameters using ML models in an arid region of northwest China[J]. Computers and Electronics in Agriculture,2023,204.

[30] Kapil Dev Sharma,Siddharth Jain. Overview of Municipal Solid Waste Generation, Composition, and Management in India[J]. Journal of Environmental Engineering,2019,145(3).

[31] Yang M, Thung G. Classification of trash for recyclability status[J]. CS229 project report, 2016, 2016(1): 3.

[32] Zhang H, Cao H, Zhou Y, et al. Hybrid deep learning model for accurate classification of solid waste in the society[J]. Urban Climate, 2023, 49: 101485.

[33] Junyu T,Yude G,Xiaoling H, et al. Combination of hyperspectral imaging and ML models for fast characterization and classification of municipal solid waste[J]. Resources, Conservation & Recycling,2023,188.

[34] Cucchiella F,D'Adamo I,Gastaldi M. Sustainable waste management: Waste to energy plant as an alternative to landfill[J]. Energy Conversion and Management,2017,131.

[35] Kannangara M,Dua R,Ahmadi L, et al. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using ML approaches[J]. Waste Management,2018,74.

[36] Yang Yadong,Shahbeik Hossein,Shafizadeh Alireza,Rafiee Shahin,Hafezi Amir,Du Xinyi,Pan Junting,Tabatabaei Meisam,Aghbashlo Mortaza. Predicting municipal solid waste gasification using ML: A step toward sustainable regional planning[J]. Energy,2023,278(PB).

[37] Ascher Simon,Wang Xiaonan,Watson Ian,Sloan William,You Siming. Interpretable ML to Model Biomass and Waste Gasification.[J]. Bioresource technology,2022,364.