



## OPEN ACCESS

## EDITED BY

Da Tian,  
Anhui Agricultural University, China

## REVIEWED BY

Mengying Zhang,  
Shanghai Institute of Microsystem and  
Information Technology (CAS), China  
Pei Peng,  
Rutgers, The State University of New  
Jersey, United States  
Zhaoxia Duan,  
Hohai University, China

## \*CORRESPONDENCE

Juai Wu,  
✉ wujuai@njupt.edu.cn

## SPECIALTY SECTION

This article was submitted to Bioprocess  
Engineering,  
a section of the journal  
Frontiers in Bioengineering and  
Biotechnology

RECEIVED 18 March 2023

ACCEPTED 27 March 2023

PUBLISHED 31 March 2023

## CITATION

Wu J and Zhao F (2023), Machine  
learning: An effective technical method  
for future use in assessing the  
effectiveness of phosphorus-dissolving  
microbial agoremediation.  
*Front. Bioeng. Biotechnol.* 11:1189166.  
doi: 10.3389/fbioe.2023.1189166

## COPYRIGHT

© 2023 Wu and Zhao. This is an open-  
access article distributed under the terms  
of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original author(s)  
and the copyright owner(s) are credited  
and that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Machine learning: An effective technical method for future use in assessing the effectiveness of phosphorus-dissolving microbial agoremediation

Juai Wu<sup>1\*</sup> and Fangzhou Zhao<sup>2</sup>

<sup>1</sup>College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, China, <sup>2</sup>School of Environmental and Biological Engineering, Nanjing University of Science and Technology, Nanjing, China

The issue of agricultural pollution has become one of the most important environmental concerns worldwide because of its relevance to human survival and health. Microbial remediation is an effective method for treating heavy metal pollution in agriculture, but the evaluation of its effectiveness has been a difficult issue. Machine learning (ML), a widely used data processing technique, can improve the accuracy of assessments and predictions by analyzing and processing large amounts of data. In microbial remediation, ML can help identify the types of microbes, mechanisms of action and adapted environments, predict the effectiveness of microbial remediation and potential problems, and assess the ecological benefits and crop growth after remediation. In addition, ML can help optimize monitoring programs, improve the accuracy and effectiveness of heavy metal pollution monitoring, and provide a scientific basis for the development of treatment measures. Therefore, ML has important application prospects in assessing the effectiveness of microbial remediation of heavy metal pollution in agriculture and is expected to be an effective pollution management technology.

## KEYWORDS

machine learning, microbial remediation, agricultural pollution, assessment and prediction, crop yield

## 1 Introduction

The safety of agricultural land as the basis for human survival and development is of increasing concern (Roy et al., 2022). In recent years, the levels of heavy metals in agricultural land have been increasing due to industrial activities, fertilizer application, coal burning, and other human activities (Xu et al., 2014). Heavy metals are characterized by non-biodegradability, high toxicity, and ease of accumulation, and are the most common type of pollutants in the agricultural environment, seriously affecting ecosystem function, food security, and human health (Sharma et al., 2022). In addition, higher levels of heavy metals may lead to changes in soil structure and nutrient loss, which in turn may affect crop quality and yield (Fei et al., 2022). Therefore, it is necessary to analyze the effects of heavy metal pollution on agricultural land and to assess the level of heavy metal pollution in agricultural land.

## 2 Phosphate solubilizing microorganisms as a green technology for improving crop growth and remediating heavy metal contamination

Microorganisms play an important role in the growth of crops in agricultural fields, and can enhance crop growth by improving the soil environment, promoting nutrient cycling, and improving plant immunity in a variety of ways. Phosphorus solubilizing microorganisms (PSM) are a group of microorganisms that can convert soil organic phosphorus compounds or insoluble inorganic phosphorus compounds into plant-available phosphorus elements. As one of the first beneficial functional microorganisms discovered in agricultural fields, the main functions of PSM include 1) increasing the effective soil phosphorus content through the release of organic acids, biological enzymes, and a series of other secretions to promote plant nutrient uptake and growth and development (Alori et al., 2017). 2) Forming a symbiotic relationship with plant roots, using inter-root substances secreted by plants for subsistence, as well as secreting hormones and enzymes (indoleacetic acid, 1-Aminocyclopropane-1-carboxylate deaminase, etc.) to help plants absorb nutrients (Rawat et al., 2020). 3) Produce plant antibodies (e.g., cell wall lysis enzymes, antibiotics, etc.) to suppress plant diseases, activate the plant's immune system, and enhance plant resistance to diseases (Zaidi et al., 2016). 4) Decomposing organic matter and promoting soil aeration and water retention capacity, thus improving soil structure and nutrient retention (Tian et al., 2021). 5) Phosphate solubilizing microorganisms are able to adsorb, accumulate and complex heavy metal ions as well as toxicity reduction through cellular self and secretions (extracellular polymers, glutathione, etc.) (Chen et al., 2023b). 6) The use of phosphate solubilizing microorganisms can also reduce the use of chemical fertilizers and pesticides, reducing the negative impact on the environment (Ahemad, 2015). In addition, the good affinity of PSM allows it to be used together with other materials to improve soil properties and remediate environmental pollution (Chen et al., 2019; Feng et al., 2022; Lai et al., 2022; Chen et al., 2023a).

## 3 Artificial intelligence—machine learning has been initially applied in agricultural environments at this stage

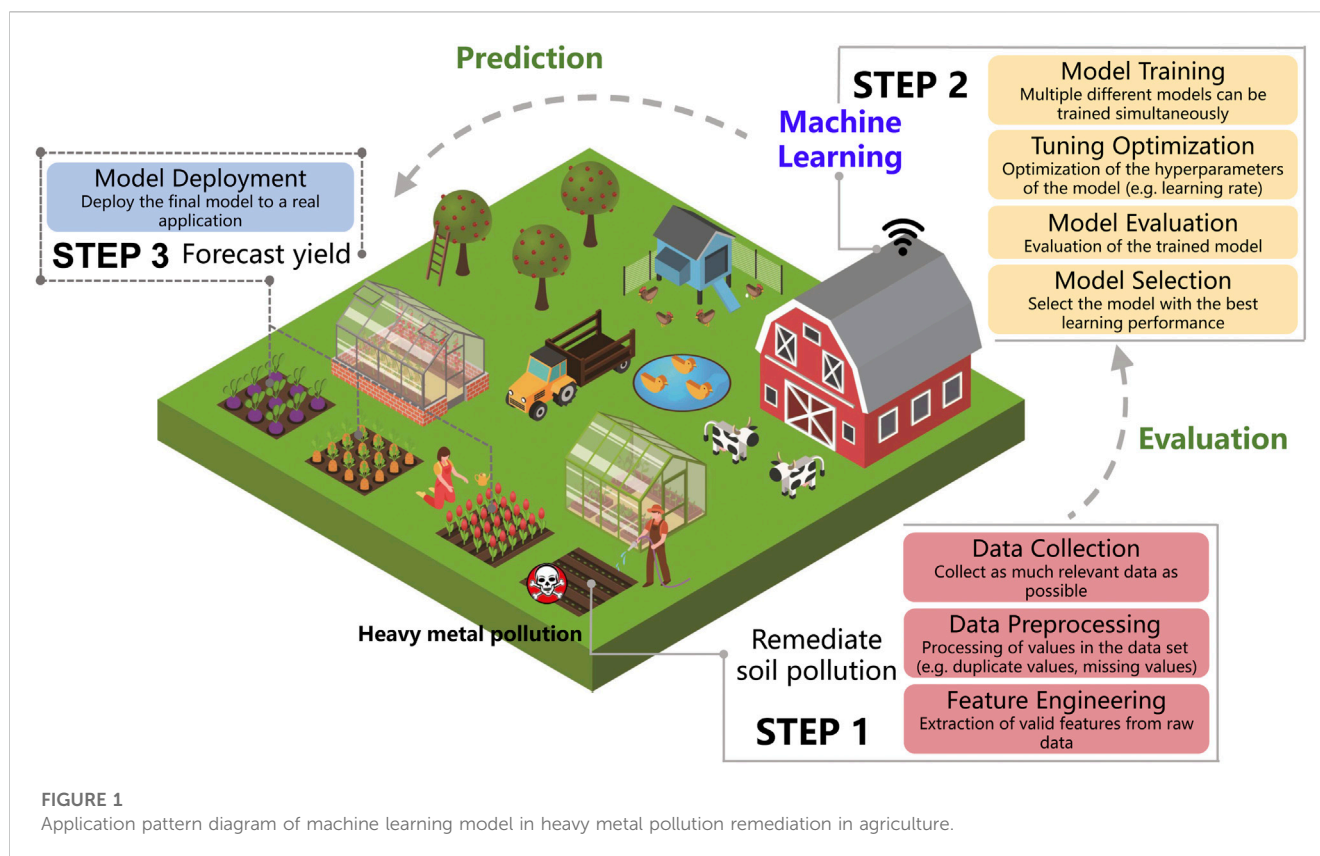
Soil contamination assessment in agriculture is a prerequisite to ensure that crops can be grown properly and safely consumed, and the classical research paradigms include three types: experiment, theory, and computational science. Traditional methods for land contamination assessment include the single factor index, single factor pollution index, potential ecological risk index, and contamination load index (El Azhari et al., 2017; Lu et al., 2021). These methods show some limitations in practical applications due to the lack of adaptability and accuracy of the built mathematical models. The era of big data has given rise to the data-driven paradigm based on big data, which refers to data sets with complex structures, including those that do not have or have not yet mastered their causal relationships. The core of the data-driven

approach lies in acquiring knowledge by analyzing large amounts of data. Common data analysis methods include classification, clustering, association (correlation, regression, etc.), discrimination, principal component analysis, statistical inference, etc. (Wu et al., 2020).

In recent years, data mining methods using artificial intelligence algorithms such as machine learning have become a hot research topic. Machine learning method is an advanced data analysis method, which is often applied to analyze the hidden information between input data and output results (Dobbelaere et al., 2021). An integrated model composed of multiple learning algorithms has the advantages of good predictive performance and high interpretability (Zhang et al., 2022). For example, random forest can directly process high-dimensional data without feature selection (Ali et al., 2021). Gradient boosting decision tree use a loss function with robustness to outliers (Yang et al., 2023). These models have been successfully applied to problems such as forecasting, engineering design, and material optimization, saving significant labor and time costs (Tian et al., 2022; Veloso et al., 2022). Currently, machine learning is widely used in agriculture. For example, Hamrani et al. successfully predicted greenhouse gas emissions in agricultural land using nine models separately, such as support vector machine, random forest, etc. Among them, long short-term memory network has the most accurate prediction for both CO<sub>2</sub> ( $R^2 = 0.87$ , RMSE = 30.3) and N<sub>2</sub>O ( $R^2 = 0.86$ , RMSE = 0.19) (Hamrani et al., 2020). Saha et al. combined a machine learning model and multi-criteria decision-making models to assess agricultural land fertility and site suitability, it was found that agricultural land with higher organic carbon content and cation exchange capacity, and low bulk weight was more advantageous (Saha and Mondal, 2022). Some studies have also shown that clay content in the soil, soluble phosphorus, and soluble organic carbon are the main factors affecting phosphorus concentration in groundwater of agricultural land using support vector machine, random forest, and neural network (Yang et al., 2023). These successful examples show that it is feasible to use machine learning methods for prediction, assessment, and analysis of pollution problems in agricultural environments.

## 4 Building machine learning models to predict crop yield and safety after remediation of heavy metal contamination is one of the hot spots for future research

In agricultural land, crop yield is influenced by several factors, including heavy metal contamination concentrations, nutrient elements in the soil, and microbial communities. These factors also have some mechanisms of action among each other. Currently, it is difficult to quantify the effects of these mechanisms of action on crop yield. Therefore, it has become a trend to use machine learning models to simulate and assess soil contamination levels (Wang et al., 2021; Liu et al., 2023). Based on the powerful data analysis methods of machine learning, and integrating the data-driven method (statistical analysis) with the model-driven method (causal analysis), it is possible to link heavy metal contamination with crop yield. It would be an effective



technical tool to explore the effect mechanism of remediated heavy metal contamination on nutrient elements, microorganisms and crop yield. The analysis diagram is shown in Figure 1.

Based on the excellent predictive performance of machine learning models, it is feasible to build a model to predict crop yield. Typical machine learning model structures include supervised learning represented by regression analysis and statistical classification, unsupervised learning represented by generating adversarial network (GAN) and clustering, and reinforcement learning (LeCun et al., 2015; Castaldo et al., 2016; Wang et al., 2020). Currently, deep learning represented by neural network is rapidly developing, which would be more efficient to integrate with supervised learning, unsupervised learning, and reinforcement learning to classify data and identify features. Due to the differences in the datasets, the prediction performance of each model behaves differently. During the model construction process, multiple models are usually selected for simultaneous parameter optimization and prediction, and the model with the best prediction performance is finally selected. Besides, a sufficient amount of data is also one of the important factors to support the success of model prediction. For example, Jhajharia et al. integrated 3,664 sets of yield data of various crops with soil type and rainfall of agricultural land in the Rajasthan region, and the results confirmed that crop yield could be successfully predicted using random forest model ( $R^2 = 0.963$ ,  $RMSE = 0.035$ ) (Jhajharia et al., 2023). Iniyar et al. counted data on precipitation, humidity, temperature, area, soil type, crop type, season, and yield for the last 18 years in the Maharashtra region. After training eight machine learning models such as linear regression, ridge regression, and gradient

boosting, it was found that the long short-term memory network showed the best prediction performance with 86.3% accuracy (Iniyar et al., 2023). Panigrahi et al. successfully used linear regression, decision tree regression, and other models to predict the yield of three different crops of corn, peanuts, and Bengal beans based on the monthly minimum and maximum temperature and annual rainfall in the Telangana (Panigrahi et al., 2023). Therefore, it is feasible to predict the yield of various crops using machine learning models, but the availability of sufficient data support from research institutions or government may be the key to it.

Machine learning models can not only predict crop yields, but also analyze the effect of PSM on remediation of heavy metal contamination or post-remediation soil properties on crop yields through interpretable algorithms such as feature importance, partial dependence, individual conditional expectation, and Shapley additive explanation. For example, by analyzing the partial dependence of a certain heavy metal ion, the concentration threshold for that heavy metal to affect crop growth is determined. Alternatively, using the characteristic importance algorithm, it is possible to obtain the importance weights for the effects of various types of heavy metals on crop yield. Such weighting factors, which can be introduced into the formula for calculating the risk level of heavy metal contamination, make the assessment results more informative. In addition, the joint application of machine learning and high-resolution aerial imaging technology has been shown to be feasible. For example, feature information is obtained based on high-resolution aerial imaging technology, and then machine learning models are used to predict the heavy metal concentrations at unknown points to

finally assess the soil contamination risk levels. The accuracy of this combined application of techniques to assess contamination levels was verified to be significantly better than the traditional kriging interpolation and inverse distance weight interpolation methods (Jia et al., 2021). In conclusion, machine learning models can simultaneously combine soil quality and crop yield so that the two are interlinked, which means that the impact of soil quality on crop growth can be analyzed along with predicting crop yield, correcting adverse factors that affect yield and continuously improving agricultural land quality. Thus, machine learning can be used as an effective technical tool to predict and evaluate the impact of phosphorus dissolving microbially remediated soils on crop yields.

## 5 Discussion

As an efficient and fast data analysis and processing method, machine learning has often been applied to solve various agricultural environmental problems in recent years. Its powerful prediction and analysis capability saves a lot of labor cost and time cost for research, and the application of machine learning in agricultural survey technology will be more promising in the future with the advancement of technology and the increase of data. However, the application direction of the technology is currently more limited. In the future, artificial intelligence, machine learning, and computer vision can be used to identify the growth status of crops and prevent them from being infected with toxic pests to affect their yield. On the other hand, machine learning is essentially a data-driven method, there is currently less training data available to researchers. A large amount of open-source data is also necessary for training models in future research. In this context, data-driven method and model-driven method can complement each other, which includes: 1) introduce casual analysis measures to researches that usually relied entirely on statistical analysis to solve data dependency and improve the analysis applicability and accuracy; 2) introduce statistical analysis measures to researches that usually relied entirely on casual analysis, to improve the analysis efficiency.

## References

- Ahemad, M. (2015) Phosphate-solubilizing bacteria-assisted phytoremediation of metalliferous soils: A review. *3 Biotech* 5, 111–121. doi:10.1007/s13205-014-0206-0
- Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., and Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Comput. Biol. Med.* 136, 104672. doi:10.1016/j.combiomed.2021.104672
- Alori, E. T., Glick, B. R., and Babalola, O. O. (2017). Microbial phosphorus solubilization and its potential for use in sustainable agriculture. *Front. Microbiol.* 8, 971. doi:10.3389/fmicb.2017.00971
- Castaldo, F., Palmieri, F. A. N., and Regazzoni, C. S. (2016). Bayesian analysis of behaviors and interactions for situation awareness in transportation systems. *IEEE Trans. Intelligent Transp. Syst.* 17, 313–322. doi:10.1109/tits.2015.2466695
- Chen, H. M., Jiang, H. F., Nazhafati, M., Li, L. L., and Jiang, J. Y. (2023a). Biochar: An effective measure to strengthen phosphorus solubilizing microorganisms for remediation of heavy metal pollution in soil. *Front. Bioeng. Biotechnol.* 11, 1127166. doi:10.3389/fbioe.2023.1127166
- Chen, H. M., Min, F. F., Hu, X., Ma, D. H., and Huo, Z. (2023b). Biochar assists phosphate solubilizing bacteria to resist combined Pb and Cd stress by promoting acid secretion and extracellular electron transfer. *J. Hazard. Mater.* 452, 131176. doi:10.1016/j.jhazmat.2023.131176
- Chen, H. M., Wang, Z. J., Tang, L. Y., Su, M., Tian, D., Zhang, L., et al. (2019). Enhanced Pb immobilization via the combination of biochar and phosphate solubilizing bacteria. *Environ. Int.* 127, 395–401. doi:10.1016/j.envint.2019.03.068
- Dobbelaere, M. R., Plehiers, P. P., Van de Vijver, R., Stevens, C. V., and Van Geem, K. M. (2021). Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats. *Engineering* 7, 1201–1211. doi:10.1016/j.eng.2021.03.019
- El Azhari, A., Rhoujjati, A., El Hachimi, M. L., and Ambrosi, J. P. (2017). Pollution and ecological risk assessment of heavy metals in the soil-plant system and the sediment-water column around a former Pb/Zn-mining area in NE Morocco. *Ecotoxicol. Environ. Saf.* 144, 464–474. doi:10.1016/j.ecoenv.2017.06.051
- Fei, X. F., Lou, Z. H., Xiao, R., Ren, Z. Q., and Lv, X. N. (2022). Source analysis and source-oriented risk assessment of heavy metal pollution in agricultural soils of different cultivated land qualities. *J. Clean. Prod.* 341, 130942. doi:10.1016/j.jclepro.2022.130942
- Feng, Y., Zhang, L. L., Li, X., Wang, L. Y., Yusef, K. K., Gao, H. J., et al. (2022). Remediation of lead contamination by *Aspergillus Niger* and phosphate rocks under different nitrogen sources. *Agronomy* 12, 1639. doi:10.3390/agronomy12071639

In conclusion, machine learning can be used to achieve accurate analysis and prediction of agricultural data, improve the efficiency and quality of agricultural production, and promote the development and upgrading of the agricultural industry.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

JW and FZ wrote the manuscript, collected the data, conceived the idea and revised the manuscript. JW led the project.

## Funding

Scientific Research Project of Nanjing University of Posts and Telecommunications (NO. NY219156).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Hamrani, A., Akbarzadeh, A., and Madramootoo, C. A. (2020). Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci. Total Environ.* 741, 140338. doi:10.1016/j.scitotenv.2020.140338
- Iniyar, S., Akhil Varma, V., and Teja Naidu, C. (2023). Crop yield prediction using machine learning techniques. *Adv. Eng. Softw.* 175, 103326. doi:10.1016/j.advengsoft.2022.103326
- Jhajharia, K., Mathur, P., Jain, S., and Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. *Procedia Comput. Sci.* 218, 406–417. doi:10.1016/j.procs.2023.01.023
- Jia, X. Y., Cao, Y. N., O'Connor, D., Zhu, J., Tsang, D. C. W., Zou, B., et al. (2021). Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field. *Environ. Pollut.* 270, 116281. doi:10.1016/j.envpol.2020.116281
- Lai, W. W., Wu, Y. Y., Zhang, C. N., Dilinuer, Y., Pasang, L., Lu, Y., et al. (2022). Combination of biochar and phosphorus solubilizing bacteria to improve the stable form of toxic metal minerals and microbial abundance in lead/cadmium-contaminated soil. *Agronomy* 12, 1003. doi:10.3390/agronomy12051003
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- Liu, C. T., Fan, H. M., Jiang, Y. Y., Ma, R. M., and Song, S. (2023). Gully erosion susceptibility assessment based on machine learning—A case study of watersheds in Tuquan County in the black soil region of Northeast China. *Catena* 222, 106798. doi:10.1016/j.catena.2022.106798
- Lu, Q. X., Xiao, Q. T., Wang, Y. J., Wen, H. H., Han, B. L., Zheng, X. Y., et al. (2021). Risk assessment and hotspots identification of heavy metals in rice: A case study in longyan of fujian province, China. *Chemosphere* 270, 128626. doi:10.1016/j.chemosphere.2020.128626
- Panigrahi, B., Kathala, K. C. R., and Sujatha, M. (2023). A machine learning-based comparative approach to predict the crop yield using supervised learning with regression models. *Procedia Comput. Sci.* 218, 2684–2693. doi:10.1016/j.procs.2023.01.241
- Rawat, P., Das, S., Shankhdhar, D., and Shankhdhar, S. C. (2020). Phosphate-solubilizing microorganisms: Mechanism and their role in phosphate solubilization and uptake. *J. Soil Sci. Plant Nutr.* 21, 49–68. doi:10.1007/s42729-020-00342-7
- Roy, P., Pal, S. C., Chakraborty, R., Chowdhuri, I., Saha, A., and Shit, M. (2022). Climate change and groundwater overdraft impacts on agricultural drought in India: Vulnerability assessment, food security measures and policy recommendation. *Sci. Total Environ.* 849, 157850. doi:10.1016/j.scitotenv.2022.157850
- Saha, S., and Mondal, P. (2022). Estimation of the effectiveness of multi-criteria decision analysis and machine learning approaches for agricultural land capability in Gangarampur Subdivision, Eastern India. *Artif. Intell. Geosciences* 3, 179–191. doi:10.1016/j.aiig.2022.12.003
- Sharma, P., Dutta, D., Udayan, A., Nadda, A. K., Lam, S. S., and Kumar, S. (2022). Role of microbes in bioaccumulation of heavy metals in municipal solid waste: Impacts on plant and human being. *Environ. Pollut.* 305, 119248. doi:10.1016/j.envpol.2022.119248
- Tian, D., Su, M., Zou, X., Zhang, L. L., Tang, L. Y., Geng, Y. Y., et al. (2021). Influences of phosphate addition on fungal weathering of carbonate in the red soil from karst region. *Sci. Total Environ.* 755, 142570. doi:10.1016/j.scitotenv.2020.142570
- Tian, X. L., Song, S. W., Chen, F., Qi, X. J., Wang, Y., and Zhang, Q. H. (2022). Machine learning-guided property prediction of energetic materials: Recent advances, challenges, and perspectives. *Energ. Mater. Front.* 3, 177–186. doi:10.1016/j.enmf.2022.07.005
- Veloso, M. F., Rodrigues, L. N., and Filho, E. I. F. (2022). Evaluation of machine learning algorithms in the prediction of hydraulic conductivity and soil moisture at the Brazilian Savannah. *Geoderma Reg.* 30, e00569. doi:10.1016/j.geodrs.2022.e00569
- Wang, J. Y., Wu, J., Wang, Z., Gao, F., and Xiong, Z. (2020). Understanding urban dynamics via context-aware tensor factorization with neighboring regularization. *IEEE Trans. Knowl. Data Eng.* 32, 2269–2283. doi:10.1109/tkde.2019.2915231
- Wang, Z. G., Wang, G. C., Ren, T. Y., Wang, H. B., Xu, Q. Y., and Zhang, G. H. (2021). Assessment of soil fertility degradation affected by mining disturbance and land use in a coalfield via machine learning. *Ecol. Indic.* 125, 107608. doi:10.1016/j.ecolind.2021.107608
- Wu, J. J., Liu, G. N., Wang, J. Y., Zuo, Y., Bu, H., and Lin, H. (2020). Data intelligence: Trends and challenges. *Syst. Eng. - Theory and Pract.* 40, 2116–2149.
- Xu, X. H., Zhao, Y. C., Zhao, X. Y., Wang, Y. D., and Deng, W. J. (2014). Sources of heavy metal pollution in agricultural soils of a rapidly industrializing area in the Yangtze Delta of China. *Ecotoxicol. Environ. Saf.* 108, 161–167. doi:10.1016/j.ecoenv.2014.07.001
- Yang, H., Wang, P., Chen, A. Q., Ye, Y. H., Chen, Q. F., Cui, R. Y., et al. (2023). Prediction of phosphorus concentrations in shallow groundwater in intensive agricultural regions based on machine learning. *Chemosphere* 313, 137623. doi:10.1016/j.chemosphere.2022.137623
- Zaidi, A., Khan, M. S., Ahmad, E., Saif, S., Rizvi, A., and Shahid, M. (2016). Growth stimulation and management of diseases of ornamental plants using phosphate solubilizing microorganisms: Current perspective. *Acta Physiol. Plant.* 38, 117. doi:10.1007/s11738-016-2133-7
- Zhang, X., Liu, L., Lan, M. X., Song, G. W., Xiao, L. Z., and Chen, J. G. (2022). Interpretable machine learning models for crime prediction. *Comput. Environ. Urban Syst.* 94, 101789. doi:10.1016/j.compenurbysys.2022.101789