



UWL REPOSITORY

repository.uwl.ac.uk

A Novel AI-based Approach for Modelling the Fate, Transportation and Prediction of Chromium in Rivers and Agricultural Crops: A Case Study in Iran

Montazeri, Ali, Chahkandi, Benyamin, Gheibi, Mohammad, Eftekhari, Mohammad, Waclawek, Stanislaw, Behzadian, Kourosh ORCID logo ORCID: <https://orcid.org/0000-0002-1459-8408> and Campos, Luiza C. (2023) A Novel AI-based Approach for Modelling the Fate, Transportation and Prediction of Chromium in Rivers and Agricultural Crops: A Case Study in Iran. *Ecotoxicology and Environmental Safety*, 263. ISSN 0147-6513

<http://dx.doi.org/10.1016/j.ecoenv.2023.115269>

This is the Published Version of the final output.

UWL repository link: <https://repository.uwl.ac.uk/id/eprint/10241/>

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A novel AI-based approach for modelling the fate, transportation and prediction of chromium in rivers and agricultural crops: A case study in Iran

Ali Montazeri^{a,1}, Benyamin Chahkandi^{b,2}, Mohammad Gheibi^{c,e,3}, Mohammad Eftekhari^{d,4}, Stanisław Waclawek^{e,5}, Kourosh Behzadian^{f,g,6}, Luiza C. Campos^{g,*,7}

^a Department of Civil Engineering, Shahrood University of Technology, Shahrood, Iran

^b School of Civil Engineering, University of Tehran, Tehran, Iran

^c Association of Talent under Liberty in Technology (TULTECH), 10615 Tallinn, Estonia

^d Department of Chemistry, Faculty of Sciences, University of Neyshabur, Neyshabur, Iran

^e Institute for Nanomaterials, Advanced Technologies and Innovation, Technical University of Liberec, Studentská 1402/2, 461 17 Liberec 1, Czech Republic

^f School of Computing and Engineering, University of West London, London W5 5RF, UK

^g Department of Civil, Environmental and Geomatic Engineering, University College London, Gower St, London WC1E 6BT, UK

ARTICLE INFO

Edited by Dr. Hao Zhu

Keywords:

Artificial Intelligence

Bio-magnification

Chromium

Fate and transport modelling

Heavy metal prediction

Water-food-pollution nexus

ABSTRACT

Chromium (Cr) pollution caused by the discharge of industrial wastewater into rivers poses a significant threat to the environment, aquatic and human life, as well as agricultural crops irrigated by these rivers. This paper employs artificial intelligence (AI) to introduce a new framework for modeling the fate, transport, and estimation of Cr from its point of discharge into the river until it is absorbed by agricultural products. The framework is demonstrated through its application to the case study River, which serves as the primary water resource for tomato production irrigation in Mashhad city, Iran. Measurements of Cr concentration are taken at three different river depths and in tomato leaves from agricultural lands irrigated by the river, allowing for the identification of bioaccumulation effects. By employing boundary conditions and smart algorithms, various aspects of control systems are evaluated. The concentration of Cr in crops exhibits an accumulative trend, reaching up to 1.29 µg/g by the time of harvest. Using data collected from the case study and exploring different scenarios, AI models are developed to estimate the Cr concentration in tomato leaves. The tested AI models include linear regression (LR), neural network (NN) classifier, and NN regressor, yielding goodness-of-fit values (R^2) of 0.931, 0.874, and 0.946, respectively. These results indicate that the NN regressor is the most accurate model, followed by the LR, for estimating Cr levels in tomato leaves.

1. Introduction

Human life and civilisation are based on water and available freshwater resources, increasing well-being and the quality of everyday life. Many efforts have been made so far, to balance the relationship between

water, environment and the human civilisation. All these efforts to manage these relations should be in line with the UN Sustainable Development Goals (SDGs) and especially SDG#14 (i.e. Life Below Water (LBW) (Capello, 2022)). LBW is proposed for both governments and private sectors worldwide and a majority of experts aim to meet its

* Corresponding author.

E-mail addresses: ali.montazeri.civil@gmail.com (A. Montazeri), Beniaminch@gmail.com (B. Chahkandi), mohamadgheibi@gmail.com (M. Gheibi), meftekhari85@yahoo.com (M. Eftekhari), stanislaw.waclawek@tul.cz (S. Waclawek), kourosh.behzadian@uwl.ac.uk (K. Behzadian), l.campos@ucl.ac.uk (L.C. Campos).

¹ ORCID: 0000-0002-7949-2874

² ORCID: 0000-0003-4898-6932

³ ORCID: 0000-0003-1987-5790

⁴ ORCID: 0000-0002-6522-8503

⁵ ORCID: 0000-0002-8430-8269

⁶ ORCID: 0000-0002-1459-8408

⁷ ORCID: 0000-0002-2714-7358

requirements by improving the relevant practices in their own areas (Gulseven, 2020). One significant measure for achieving this goal is to protect water resources from wastes and chemicals, i.e. contamination (Joshi and Temgire, 2022). The LBW is fundamental to water resources with major impacts on human life either directly e.g., drinking water or indirectly such as the aquatic, agricultural and industrial activities as illustrated in Fig. 1.

Discharge of pollution and wastewater directly or indirectly into receiving water resources, especially rivers due to human activities are one of the most dangerous threats among all factors threatening water resources (Horn et al., 2022). Moreover, industrial wastewater discharged into the rivers has resulted in increased concentrations of heavy metals including Co, Cr, Cu, Fe, Sb and Zn (Lučić et al., 2022). All these have driven scientists to investigate river catchments and identify the contribution of heavy metals (including fate and transportation) to the pollution of water resources. Fig. 2 shows the interaction between the pollution derived from human activities discharged into the rivers and the threats to the environment and humans. In addition, there might be major threats against rivers with challenging consequences, which require all stakeholders such as governments, the private sector and society, to take the necessary action to mitigate the negative impacts (Lu, 2022; Hernanda and Giyono, 2022). More attention is now given to water resources conservation e.g., LBW in SDG#14 and it can be more crucial in developing countries (Omuku et al., 2022).

The accumulation of heavy metals in living organisms, which is also known as bioaccumulation, can have severe consequences, which can be resulted in the destruction of water-food nexuses in a water body catchment (Lv et al., 2022). Fig. 3 shows this nexus through a cycle of industrial and agricultural activities around the river and their impact on the nearby environment and SDG especially LBW. These conceptual circular impacts can help decision makers to establish sustainable plans for healthy agriculture and the protection of human health. This nexus with its impacts on or affection by intervention in the relevant sectors has been analysed in research works. For example, re-distribution of manufacturing with relevant challenges and opportunities for food-water-energy nexus has attracted a lot of attention by stakeholders and research communities (Veldhuis et al., 2019). More specifically, it aims to understand a conceptual framework for local nexus network for evaluating the sustainability of future localised food systems and their association with energy and water supply (Cottee et al., 2016).

Furthermore, the scarcity of water resources and the necessity of developing water reuse strategies have introduced a new challenge related to the impact of pollution on receiving water bodies. This challenge affects the water-energy-pollution nexus, leading to the emergence of new intervention strategies emphasising the interconnection of these factors (Landa-Cansigno et al., 2020).

Given the significant health hazards associated with heavy metals in water resources, it is crucial to monitor their presence in receiving water bodies and understand their impact on agricultural crops through irrigation. Hence, multiple researchers investigated the sources, transportation and fate of heavy metals in rivers and their impact on water quality factors and plants irrigated with the polluted water. For example, Mokarram et al. (2022) collected experimental samples from a river reach to quantify the water quality of a river and estimate the spatial distribution of heavy metals released by industrial sectors using Kriging methods. The level of heavy metals in soils can also be measured and different factors like the distance of river or type of soil can be considered to evaluate the role of pollution in the agricultural landscape (Shahradnia et al., 2022). Krauss et al. (2002) investigated the potential of using Isoterms as a non-linear method to predict some heavy metal concentrations (Cd, Cu, Pb, and Zn) in wheat grain and leaves. MacFarlane et al. (2003) analysed the accumulative concentration of heavy metals such as Cu, Pb and Zn in different parts of the grey mangrove plant. They studied the relationship between heavy metals and several transferring factors like pH and sediment concentration. Chojnacka et al. (2005) conducted research on the transfer of heavy metal ions (As, Cd, Cr, Cu, Hg, Mn, Ni, Pb, Zn) from contaminated soil to plants with a focus on the transfer factors to better predict whether a given soil is suitable for cultivation of plants. Adams et al. (2004), Ye et al. (2014) and Tang et al. (2018) studied the transfer and prediction patterns of heavy metals in crops by empirical regression models along with edaphic factors such as total heavy metal content, organic matter content and pH. Verma et al. (2007) modelled the cadmium uptake by some vegetables and solved the non-linear partial differential equations through an implicit finite difference method using Picard's iterative technique in MATLAB. Boshoff et al. (2014) developed a regression-based model to predict the heavy metal (As, Cd, Cu, Pb and Zn) concentration in a soil-plant ecosystem in grass and nettle (*Agrostis sp./Poa sp.* and *Urtica dioica L.*) plants using pseudo total metal concentration. Yu et al. (2016) presented a linear model of prediction and distribution of heavy metals (Zn, Cd,

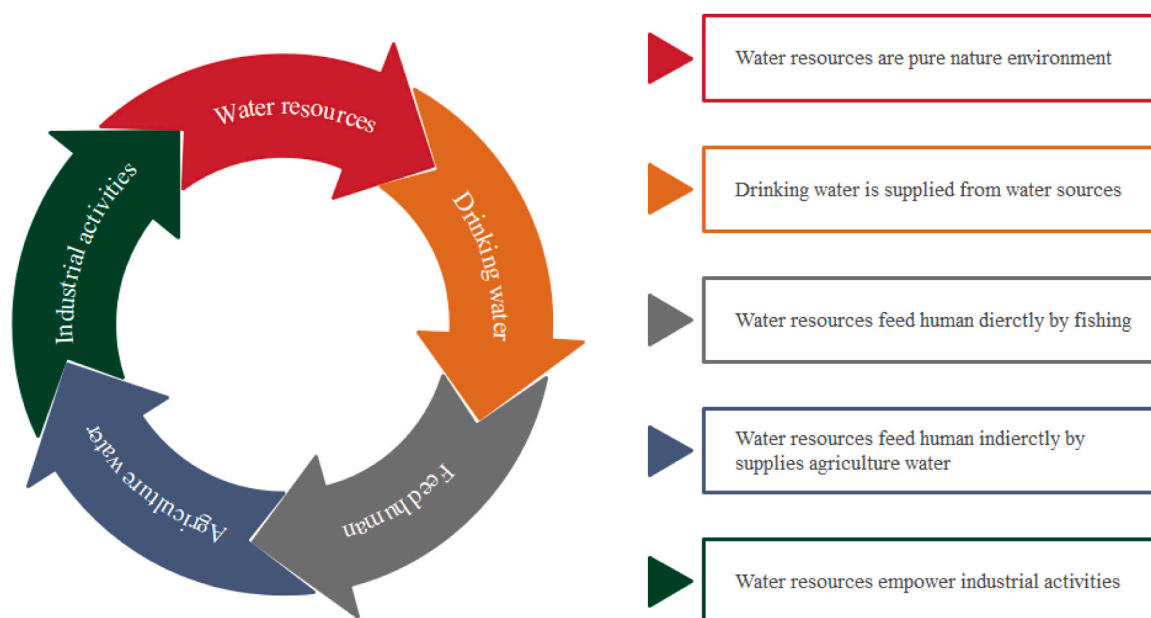


Fig. 1. The impact of the water cycle on human life through direct water use and agricultural and industrial activities.

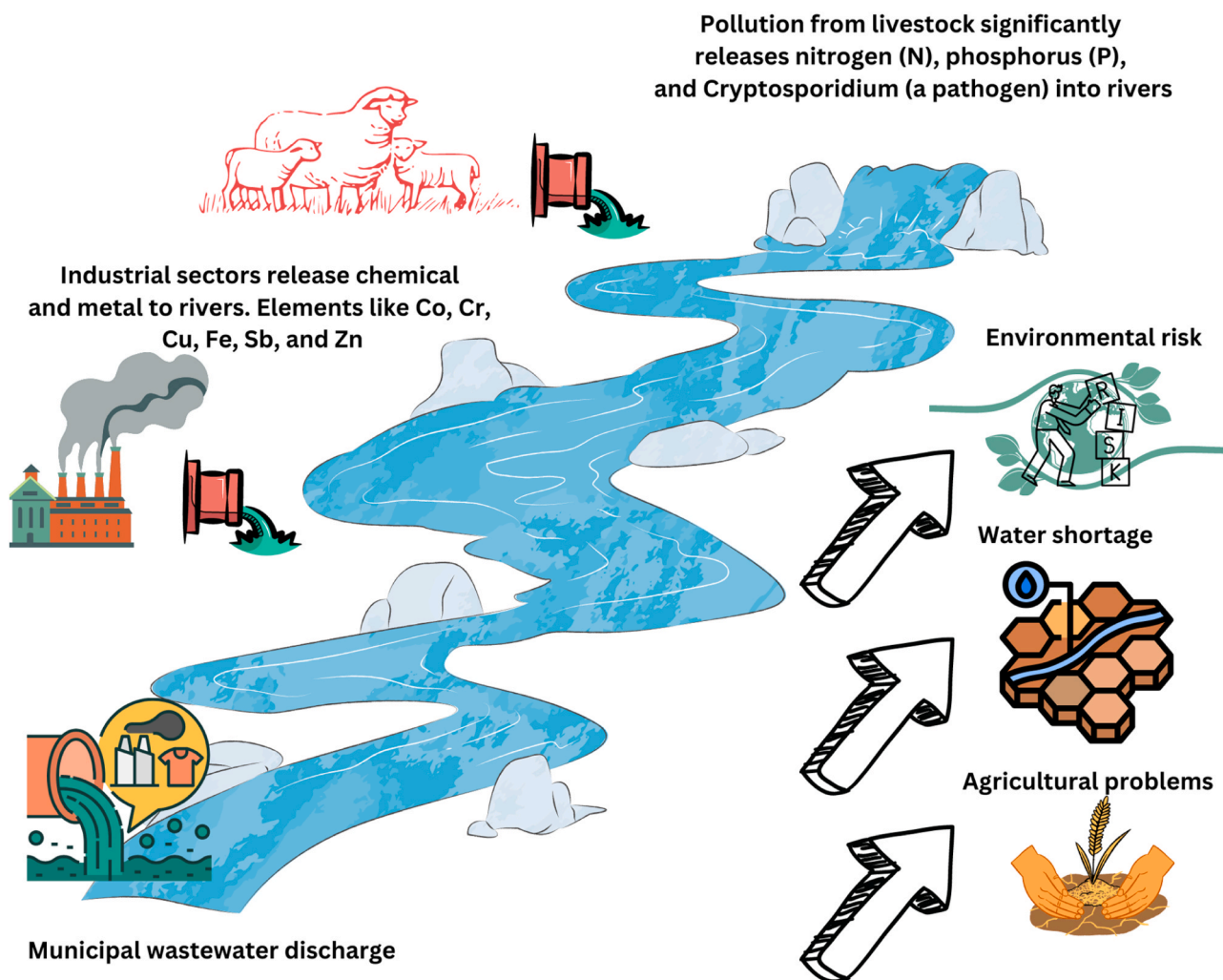


Fig. 2. The interaction of pollution from human activities discharged into a river with the threats to the environment and human demands.

Cu, Pb, Cr, and Ni) in wheat in Tianjin, China. Zhou et al. (2019) also investigated six heavy metal contaminants in soil and the rice-wheat rotation systems in Dingshu an industrial region in China. They further implemented a prediction model of Cd for rice and wheat grains based on the nonedible organs and soil properties. Kumar et al. (2019a) similarly examined a two-factor multiple linear regression model to predict heavy metal uptake by water lettuce (*Pistia stratiotes* L.) from paper mill effluent. In another study by Kumar et al. (2019b), Principal Component Analysis, and Accumulation Nutrient Elements methods were employed as a regression modelling to predict heavy metal uptake by cauliflower (*Brassica oleracea* var. botrytis) grown in soils irrigated by industrial effluent. Zhao et al. (2019) also provided a sigmoid model to predict heavy metal uptake by sunflowers in different stages of growth. Their novelty was to consider the whole biomass-soil system instead of root-soil system employed in previous studies (Thoma et al., 2003; Mathur, 2004; Liang et al., 2009; Wu et al., 2009; Tuovinen et al., 2011). Hu et al. (2020) modelled bioaccumulation of heavy metals (Cu, Cr, Ni, Hg, Cd, As, Pb and Zn) in soil-crop ecosystems and used machine learning algorithms including random forest, gradient boosted machine, and generalised linear to identify the transfer process and controlling factors in the Yangtze River Delta, China. Eid et al. (2020) developed several regression equations to predict the uptake of ten heavy metals (Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb, Zn) by a vegetable (Arugula) in Abha region, Saudi Arabia. Yutao et al. (2022) investigated the concentration of several heavy metals, especially Cr, in the irrigated soil by electroplating factories effluent located in Jiangsu Province, China. They used

back propagation (BP) neural network prediction model and human health risk assessment. Eid et al. (2022) also developed a regression prediction model for evaluating 10 heavy metal uptake into *Hordeum vulgare* L. including roots, foliage and grain. Montazeri et al. (2022) developed a new empirical approach for bio-accumulated Cr in tomato during the growth period and evaluates different 3D mathematical distribution including Polynomial, Interpolant, and Lowest models for estimating Cr concentrations in agricultural crops. Some other experts also traced heavy metals in the food chain by collecting and analysing samples of fish for heavy metals by using atomic absorption spectroscopy (Dehghani et al., 2022). Researchers also investigated the footprint of heavy metals from rivers to human bodies to trace and highlight the importance and biomonitoring of heavy metals including Cr, Co, Cu, As, Hg, and Pb in rivers (Shaabani et al., 2022). In addition to the impact of heavy metals, including Cr on the river ecosystem, they are highly risky to the environment, public and ecological health (Ali et al., 2022). More specifically, heavy metals generally damage all aquatic ecosystems, which requires immediate actions for monitoring and prevention (Ahmed et al., 2022). A review of the most relevant research on this issue is listed in Table1.

As there are a large number of industrial wastewater treatments discharging a high amount of concentration of Cr into rivers, this heavy metal is selected for the analysis in this study. Hence, keywords "chromium", "emission" and "water resources" have been looked out in Scopus databank and then assessed through VOSviewer software. Keywords with more than 200 iterations were extracted and the results associated

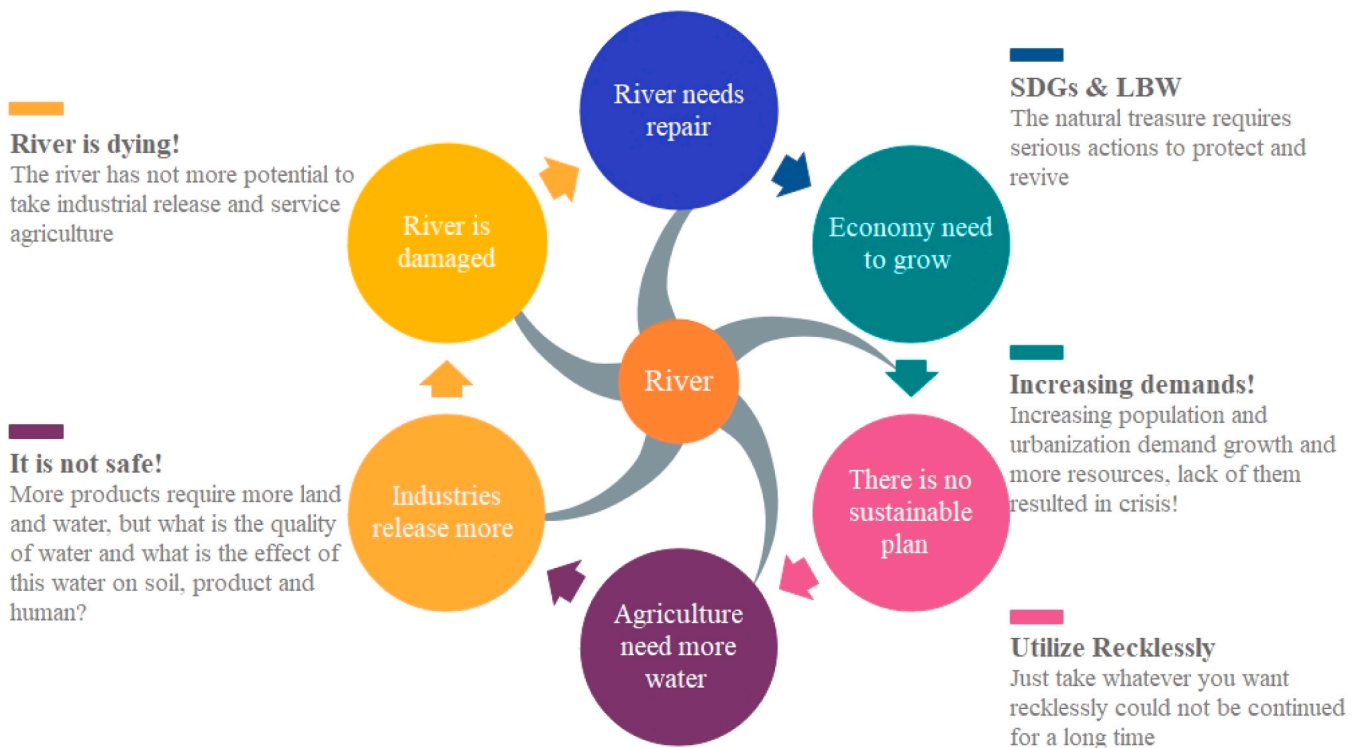


Fig. 3. Defective cycle of the case study river utilisation for gaining more economic growth.

with Cr is depicted in Fig. 4.

Overall, the monitoring and simulation of heavy metals discharged into rivers by industrial activities were analysed thoroughly by various research works. However, the efforts mainly remain in the monitoring phase with a lack of a holistic management approach for taking action against increasing the presence of heavy metals. To overcome this, the present study aimed to analyse the transport and fate of the Cr metal from the beginning (i.e., discharge phase) to the end (i.e., uptake by the crops). This approach starts with a field study and continues with gathering samples for experimental analyses, investigating and analysing the data to create a dataset for developing Artificial Intelligence (AI) and Machine Learning (ML) to forecast pollution spread and finally, supporting managerial systems to make informed decisions.

The next section presents the material and methods including methodology, case study, field research and data collection methods, ML logics, and smart framework creating steps followed by presenting results and discussion by comparing with the latest publications in this area. Finally, the conclusions are drawn by presenting the key findings of the research, followed by making recommendations for future works.

2. Material and methods

2.1. Methodology framework

Fig. 5 shows the methodology based on the framework with all steps suggested in this study. It started with a site investigation to identify industries and the type of potential heavy metals they can potentially discharge into the river. This study selects a heavy metal and traces it from the point of discharge to the bioaccumulation of the plant and agricultural crops based on the conditions of the area (bioaccumulation and its epidemiology-immunology effects for tomato are shown and described in Text A1 and Fig. A6 in Appendix A). The concentration of the heavy metal is also measured across the river to create a comprehensive database of all sampling and potential sources and bioaccumulation in the agricultural crops. Analysing and visualising this database can depict the conditions of the river's health. The ML was then

trained and tested with an existing database to predict the heavy metal for a number of future scenarios. The water quality management of the river can then be delineated based on the list of significant parameters of the bioaccumulation process, along with depicting future scenarios. More details of these steps are described below after presenting the case study in the next section.

2.2. Case study

The methodology in this study is demonstrated by its application to the real-world case study of the river located in the north-western part of Iran (shown in Fig. A1 in Appendix A). The river is the main source of water surface for irrigation of agricultural lands, although it now has no such function due mainly to the lack of proper inflow and over-exploitation of water withdrawal at upstream catchments. Finally joins the Harirud River towards the East at the borderline between Iran and Turkmenistan. The river is 240 km long and passes through mountains, plains and at least four important cities, including Mashhad (the capital of the province) (Davari et al., 2020). A collection of historical and current pictures of the river (Fig. A2 in Appendix A) shows the conditions of the river that have been deteriorated over the time through industrial and urban wastewater discharge and lack of environmental management (Hajinamaki et al., 2016).

The primary pollution sources of the river are shown in the dashboard in Fig. 6. As can be seen, the three wastewater treatment plants (WWTPs), i.e., Oulang and two Parkands, account for 97.3% of the total pollution in the river. These three WWTPs receive urban wastewater and have recently been banned from discharging or bypassing into the river. Therefore, the main threat will be from the rest of the industries shown in Fig. 6b in which the Charmshahr WWTP (CWWTP) is the primary source of pollution in the river. Table A1 in Appendix A provides the flow values of pollution discharge. Note that Charmshahr is an industrial town located in Mashhad with more than 60 branches and 2,200 workers in leather processing. Based on preliminary field investigation and site visits, the discharge of CWWTP into the river has a significant amount of Chromium (Cr) which needs to be monitored and controlled.

Table 1

Main recent research works for modelling and estimating heavy metal concentrations in agricultural crops.

Crop type	Heavy metals	Methodology	Aim	Concentrations	Reference
Wheat	Cd, Cu, Pb, and Zn	Isoterm	Estimation of heavy metal concentration in wheat grain and leaves	< 0.01–1.3 mg Cd/kg, 1.3–6.6 mg Cu/kg, < 0.05–0.30 mg Pb/kg, 8–104 mg Zn/kg of grain 3.2 mg Cd/kg, 111 mg Cu/kg, 4.3 mg Pb/kg, and 177 mg Zn/kg for leaves	Krauss et al. (2002)
Grey Mangrove	Cu, Pb and Zn	Accumulative concentration	Estimation of accumulative concentration relation in different parts of grey mangrove and soil	1.58, 0.79 and 4.51 µg/mol Cu, Pb and Zn for root. 0.14, 0.02, 0.38 µg/mol Cu, Pb and Zn for leaves.	MacFarlane et al. (2003)
Wheat and Barely	Cd	Multiple linear regressions and analysis of variance	Investigation of the relation between soil properties and Cd concentration	0.04–0.2 mg/kg Cd for wheat. 0.008–0.036 mg/kg Cd for barely.	Adams et al. (2004)
Wheat	As, Cd, Cr, Cu, Hg, Mn, Ni, Pb and Zn	Measuring the heavy metal concentration	Finding the correlation of transferring factors and heavy metal concentrations	3.13–9.16 As, 0.07–0.55 Cd, 6.55–26.57 Cr, 6.21–26.2 Cu, 0.57–4.64 Hg, 164.0–1087.0 Mn, 5.24–16.7 Ni, 13.27–47.1 Pb, 0.02–1.80 Se, 21.00–100.70 Zn (all in mg/kg) in soil samples	Chojnacka et al. (2005)
Several vegetables	Cd	Non-linear partial differential equations via an implicit finite difference method	Estimation of cadmium uptake by vegetables	0.980, 0.310, 3.090 and 0.093 mg/kg Cd in carrot, radish, spinach and cabbage respectively	Verma et al. (2007)
Grass & Nettle	As, Cd, Cu, Pb and Zn	Regression-based model using pseudo total metal concentration	Estimation of heavy metal concentrations	Highest amounts (mg/kg) of 11. 5, 60, 130 and 475 for As, Cd, Cu, Pb and Zn in nettle and 10, 217, 633 for Cd, Pb and Zn in grass respectively.	Boshoff et al. (2014)
Rice	Cd	A pot experiment with 19 representative paddy soils	Investigation of the effects of soil properties on bioaccumulation of Cd in rice grains	0.021–0.33 mg/kg in rice grains	Ye et al. (2014)
Wheat	Zn, Cd, Cu, Pb, Cr, and Ni	Linear model	Estimation and distribution of heavy metals	0.025–0.117 Cd, 2.1–4.16 Cu, 0.0–0.24 Pb, 16.0–52.95 Zn, 0.2–0.62 Cr, 0.0–0.52 Ni in grains, 0.027–0.146 Cd, 1.74–4.13 Cu, 0.05–0.69 Pb, 5.28–22.46 Zn, 0.290–0.95 Cr, 0.37–0.41 Ni in leaves, all in mg/kg.	Yu et al. (2016)
Rice & wheat	Cd, Cr, Cu, Ni, Pb and Zn	Empirical model based on the nonedible organs and soil properties	Estimation of Cd in rice and wheat grains	0.04–0.62 Cd, 0.16–0.68 Cr, 0.32–2.47 Ni, 0.26–1.77 Pb, 1.37–7.31 Cu, 26.79–97.08 Zn in wheat grains and, 0.03–0.37 Cd, 0.31–0.93 Cr, 0.25–3.61 Ni, 0.2–1.18 Pb, 2.78–8.87 Cu and 27.97–69.91 Zn in rice grains all in mg/kg	Zhou et al. (2019)
Water lettuce	Cd, Cu, Fe, Pb, Mn and Zn	Two-factor multiple linear regression model	Estimation of heavy metal uptake by water lettuce from paper mill effluent (PME)	0.204 Cd, 0.788 Cu, 14.83 Fe, 0.297 Pb, 10.351 Mn, 4.181 Zn in plant leaves irrigated 100% by PME all in mg/kg.	Kumar et al. (2019a)
Cauliflower	Cd, Cr, Cu, Fe, Mn and Zn	Regression-based model based on principal component analysis, and accumulation nutrient elements methods	Estimation of heavy metal uptake	0.001–1.290 Cd, 0.104–12.580 Cr, 0.313–11.450 Cu, 2.649–22.840 Fe, 2.073–26.750 Mn and 1.640–25.820 Zn, all in mg/kg.	Kumar et al. (2019b)
Sunflowers	Cd, Ni, Pb, and Zn	Sigmoid model	Estimation of heavy metal uptake in different stages of growth	79.54 Cd, 33.89 Ni, 141.63 Pb, 324.65 Zn all in mg/kg	Zhao et al. (2019)
Rice, vegetables, fruit, tubers, bean and tea	Cu, Cr, Ni, Hg, Cd, As, Pb and Zn	Machine learning algorithms including random forest, gradient boosted machine, and generalised linear	Bioaccumulation analysis of heavy metals in soil-crop ecosystems and identifying the transfer process and controlling factors	The average amounts are 14.22 Zn, 1.97 Cu, 0.44 Cr, 0.39 Ni, 0.16 As, 0.05 Pb, 0.05 Cd and, 0.02 Hg all in mg/kg.	Hu et al. (2020)
Arugula	Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb and Zn	Regression-based models	Estimation of the heavy metal uptake	0.2–1.3 Cd, 1–11 Co, 7–88 Cr, 4–27 Cu, 687–17747 Fe, 72–396 Mn, 0.5–3 Mo, 2–27 Ni and 0.1–4 Pb all in mg/kg.	Eid et al. (2020)
Okra	Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, and Zn	Regression-based model	Estimation of heavy metal uptake by the fruits, the leaves, the stems, and the roots of the okra plant	0–2.2 Cd, 0.2–25.1 Co, 0.1–211.5 Cr, 0.8–37.9 Cu, 45.6–27496 Fe, 9.7–937.3 Mn, 0–105.5 Ni, 0–3.9 Pb, 10.1–179.2 Zn in different parts of the plant all in mg/kg.	Eid et al. (2021)
-	Cr and As	Geological accumulation index method, BP neural network	Bioaccumulation analysis for the concentration of several heavy metals in soil and human health risk assessment model	N/A	Yutao et al. (2022)
Barely	Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb and Zn	Regression-based model	Estimation of heavy metal uptake	0.1–1 Cd, 0.1–18 Co, 3–85 Cr, 2–48 Cu, 121–11165 Fe, 0.4–538 Mn, 1–5 Mo, 0.3–42 Ni, 0.1–5 Pb, 6–194 Zn in different parts of the plant all in mg/kg.	Eid et al. (2022)

The technical details of this approach are discussed in the following sections.

2.3. Data collection

This study applies widely used techniques for data collection as spatial monitoring and sampling mainly from the river surface (Islam

et al., 2022). Moreover, the concentration of Cr is measured in the agricultural products irrigated by the river within the downstream lands. Fig. 7 illustrates the sampling network of the Cr measurements as the schematic representation of the catchment, pollution source points, the network/layers for sampling and the location of irrigated land and frequency of sampling from the agricultural crops.

As depicted in Fig. 7a, the sampling scheme is designed by defining

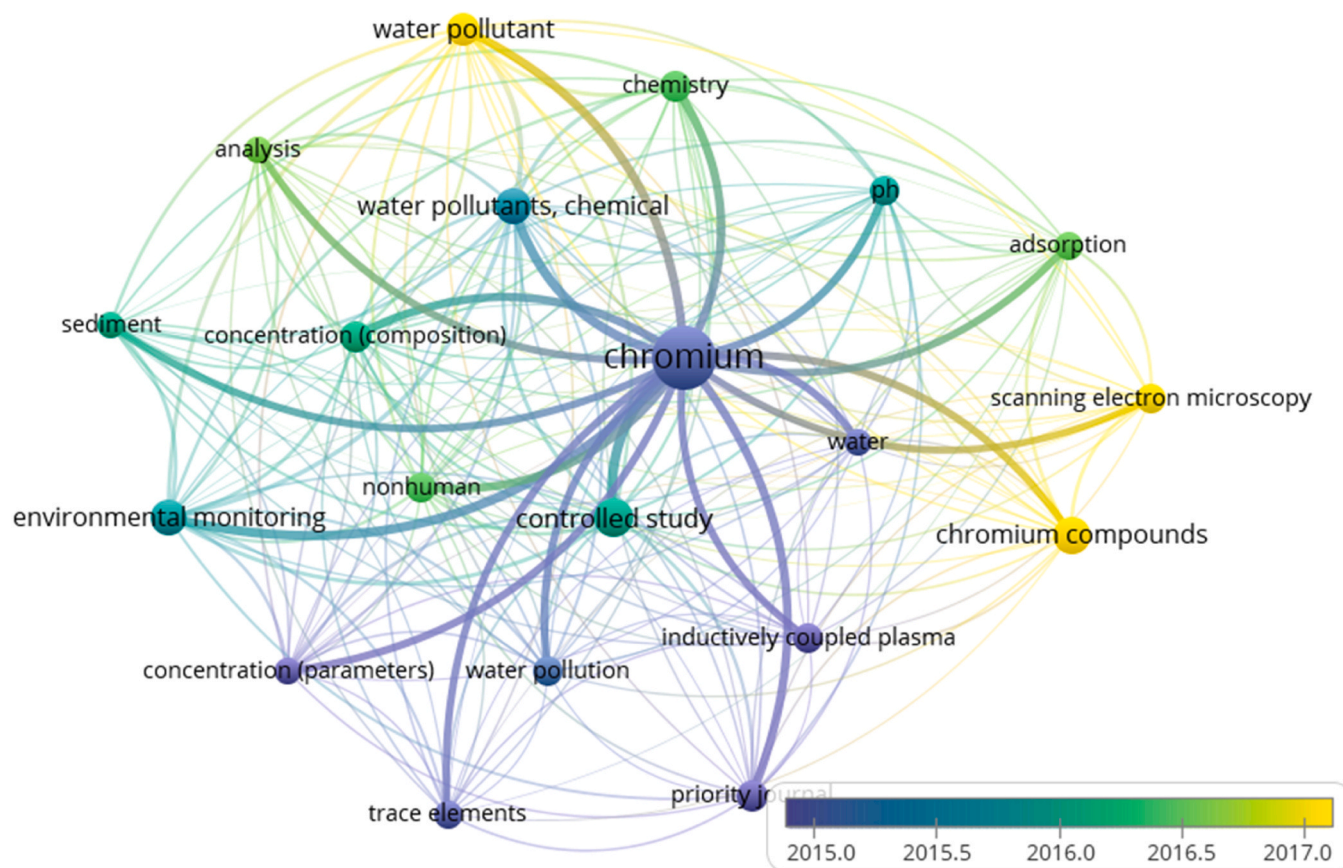


Fig. 4. Keywords associated with chromium with more than 200 iterations in the literature.

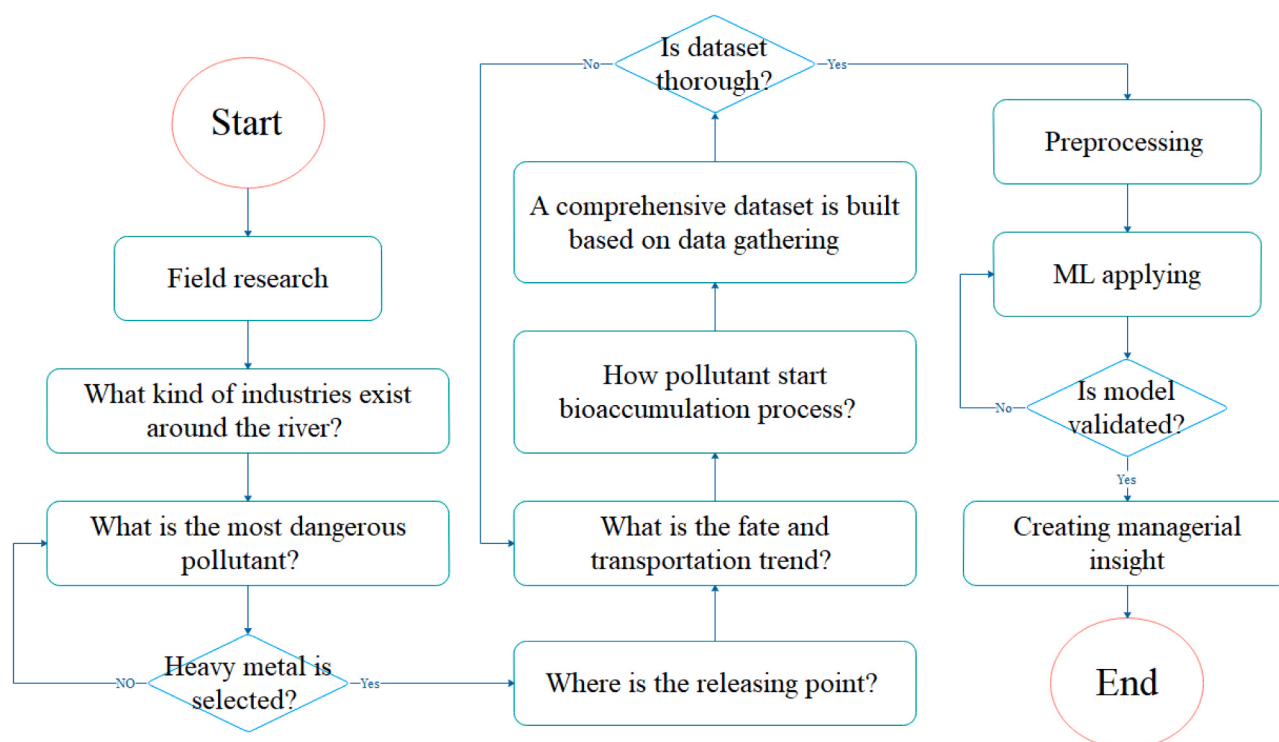


Fig. 5. The flowchart of the methodology.

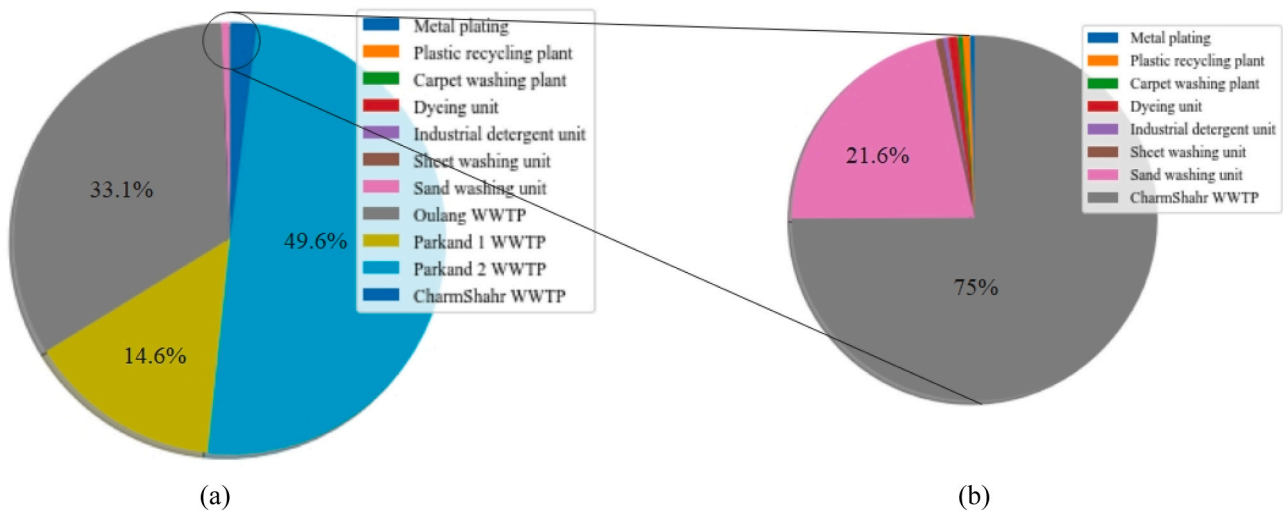


Fig. 6. Share of pollution discharged by industrial sectors into the river of the case study for (a) all industrial units, including three main WWTPs and (b) only small industrial units i.e., except the three main WWTPs.

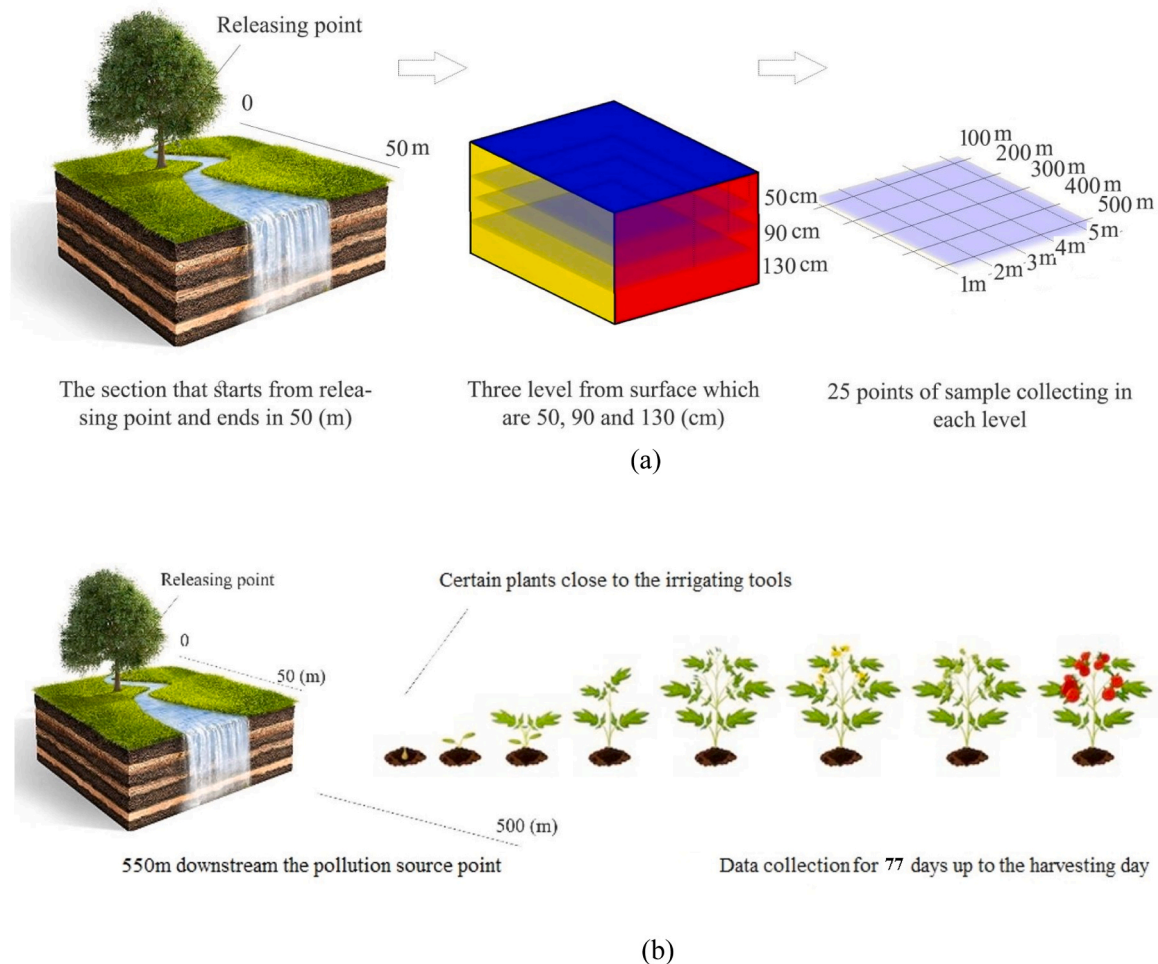


Fig. 7. Sampling scheme of the Cr measurements (a) the pollution source and levels and network of sampling and (b) schematic location of agricultural lands for sampling and frequency of sampling.

25 sampling points in a network of 5 m by 500 m in the river, starting from the pollution source points and extending by 50 m downstream. The sampling is also taken in three levels of depth as 0.5, 0.9 and 1.3 m from the river surface. Therefore, each level is divided into 5 sections in

both directions and 25 sampling points are set, which is a total of 75 samples for measurement. The data were collected in September 2019 – 2020 with even distribution between winter and summer to provide a picture of a general trend of the Cr concentration throughout both

seasons in the river. Therefore, a total of 150 samples are measured in the experimental practices.

The sampling from water is according to the standard of Raven Water Sampling, USA. In each practice, three samples are collected from the river and the mean value of the chromium amount with less than 5% tolerance is reported. Tomato sampling is done manually with experimental instruments (AZ company, Taiwan) as per standard methods (Benedetti et al., 2010; AMBRUS, 1979; Papadopoulos, 2008). All the water and tomato samples are transferred to the lab in icy bottom less than 2 h after the sampling process.

The second part of measurement is for data collection from the agricultural crops (i.e., tomato in this case) irrigated by the river as illustrated in Fig. 7b. Tomato is the main agricultural crop of this area. Hence, the Cr concentration is measured inside the leaves of certain tomato plants every two days for up to 77 days during the summer before harvesting the crop. Note that the location of tomato plants is 500 m downstream of the end of the sampling network of the river, that is 550 m downstream of the pollution source point.

The concentration of total Cr in the lab is measured according to the standard of Perkin Elmer Analyst 700 Atomic Absorption Spectroscopy (AAS) and specific 357.9 nm lamp based on standard methods (Hseu,

2004; Standard methods for the examination of water and wastewater, 2012). The concentration of Cr in tomato is measured based on first acidifying the samples by adding 1 mol L⁻¹ HNO₃ (Merck, Darmstadt, Germany) and then measuring the concentration in the solution by the AAS. In the acidification process of different samples, 2 g from each sample is digested in 20 mL of the declared acid. The hardness of water samples is measured by Portable Water Hardness Testers, HACH, USA (Standard methods for the examination of water and wastewater, 2012). Likewise, the pH of samples is measured by Metrohm 827, Switzerland (Standard methods for the examination of water and wastewater, 2012).

Note that Sodium Absorption Ratio (SAR) is a parameter used to evaluate the suitability of water for irrigation purposes. It measures the concentration of sodium relative to calcium and magnesium in the soil and is an indicator of the potential for soil structural problems caused by excessive sodium. The SAR is calculated as:

$$SAR = \frac{Na^+}{\sqrt{(Ca^{2+} + Mg^{2+})/2}} \times 100 \quad (1)$$

where Na⁺ = the concentration of sodium ions (in milliequivalents per liter, meq/L); Ca²⁺ = the concentration of calcium ions (in

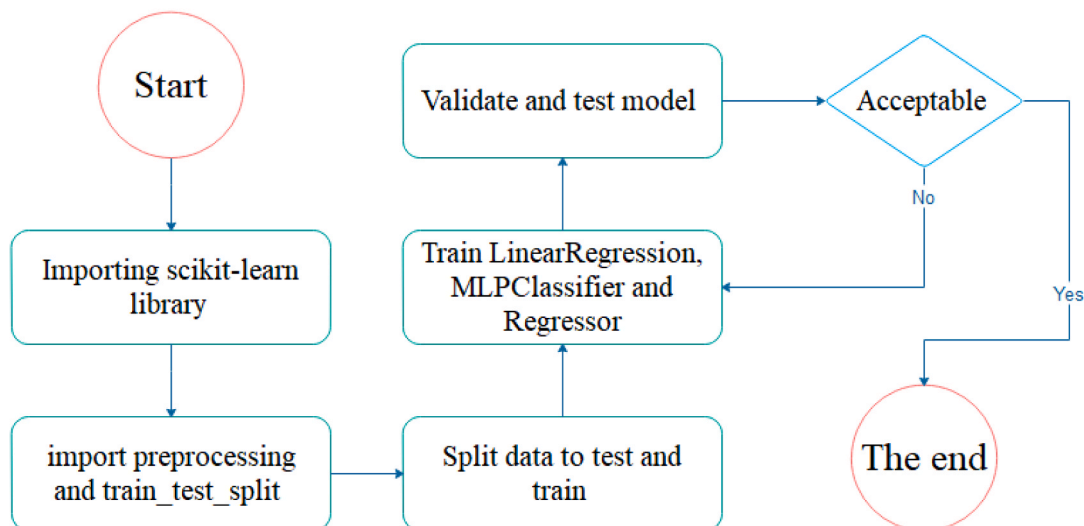
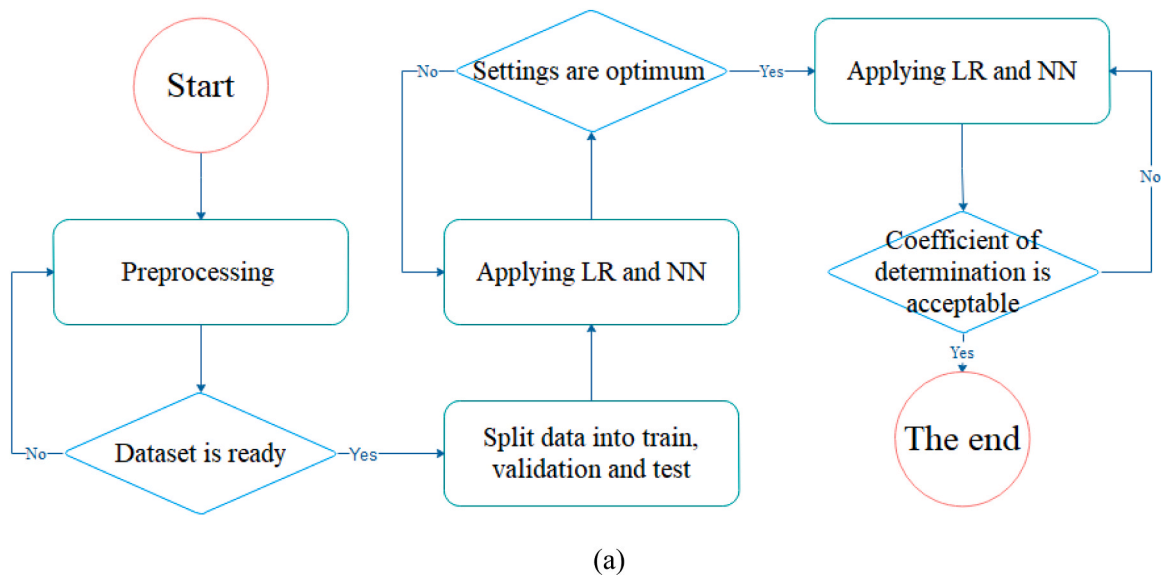


Fig. 8. Steps of the ML used in this study (a) general method and (b) customised by the scikit-learn library.

milliequivalents per liter, meq/L); Mg^{2+} = the concentration of magnesium ions (in milliequivalents per liter, meq/L). In this study, the experimental practices of the SAR measurements were done based on the standard methods and techniques (Sposito and Mattigod, 1977).

2.4. AI application

This study applies AI methods written in Python to estimate the Cr concentration ($\mu\text{g/g}$) in tomatoes at different stages in the river and agricultural plant. Hence, various ML methods are analysed to identify the most suitable one for estimating the concentration of Cr and find their optimum settings. The methods include the ordinary least squares Linear Regression (LR) and Neural Network (NN) multi-layer perceptron classifier and regressor. The combination of these methods is also analysed here to (1) evaluate the potential of regression and classification methods for accumulation prediction; (2) ensure the best ML method is selected that can be used for similar research and finally (3) examine their compatibility with the nature of the dataset.

Fig. 8 shows the steps required for building the two types of the ML used in this study. Fig. 8a is the general steps of the ML starting with pre-processing, dividing data into train and test groups, training model, validating and finally testing the model performance with Coefficient of Determination (R^2). The general steps of the ML method are further customised in Fig. 8b in Python under Jupyter Notebook. The steps from pre-processing to R^2 calculations are utilised by using scikit-learn 1.0.2 library for the LR through fitting a linear model with coefficients to gain minimum residual sum squares between predicted targets and observed records. Moreover, the NN in the library would optimise the log-loss function with Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) or stochastic gradient descent.

Table 2 shows the main features of these two models. The Linear Regression (LR) algorithm in the scikit-learn library provides a powerful tool for modeling linear relationships between variables with several settings and parameters outlined here. The LinearRegression class allows for customisation and fine-tuning of the algorithm's behaviour. The fit_intercept parameter is a boolean value that determines whether to calculate the intercept of the linear regression model. More specifically, when setting to True, an intercept term will be included in the model. Conversely, when setting it to False, it will force the model to pass through the origin (0, 0). The normalise parameter was previously used to normalise the input features although it was marked as deprecated. However, it is now recommended to use the StandardScaler or other appropriate preprocessing methods separately on the input data. The copy_X parameter is another boolean value that determines whether a copy of the input data should be made before fitting the model. Setting it to True ensures that the original input data remains unchanged during the fitting process. The n_jobs parameter controls the parallelism of the

algorithm. By specifying an integer value greater than 1, the computation can be distributed across multiple processors. If it sets to - 1, it will use all available processors for parallel execution. If it sets to None, the algorithm will use the default value i.e., 1. The positive parameter allows the model to enforce non-negativity constraints on the coefficients. When it sets to True, it ensures that the coefficients remain positive during the fitting process. By adjusting these settings in the LinearRegression class, users can tailor the algorithm's behaviour to their specific needs and achieve optimal results when fitting linear regression models.

The MLPClassifier algorithm in scikit-learn provides a range of configurable settings for constructing a Neural Network classifier. Some key selected values for the model include specifying a single hidden layer with 130 neurons using hidden_layer_sizes= (130), employing the rectified linear unit (ReLU) activation function with activation= 'relu' to introduce non-linearity and capture complex patterns, utilising the 'adam' solver that combines stochastic gradient descent (SGD) and adaptive learning rate methods with solver= 'adam', controlling regularisation strength through alpha= 0.0001, setting the initial learning rate to 0.001 with learning_rate_init= 0.001, limiting the number of iterations to 200 using max_iter= 200 to prevent overfitting, and adjusting the batch size automatically based on the training data size with batch_size= 'auto'. These values can be fine-tuned based on the specific dataset and task requirements to optimise the classifier's performance and achieve accurate classification results.

The MLPRegressor algorithm in scikit-learn provides a range of settings to configure a Neural Network regressor. The selected values for the model include a hidden layer with 130 neurons, the 'relu' activation function to introduce non-linearity, and the 'adam' solver as the optimisation algorithm. The alpha parameter controls the regularisation strength, while the learning_rate_init determines the initial learning rate. The max_iter parameter limits the maximum number of iterations during training to 200, and the batch_size is set to 'auto' for automatic adjustment based on the data size. The random_state parameter is set to 1 for reproducibility. Other parameters such as tol, verbose, momentum, nesterovs_momentum, early_stopping, validation_fraction, beta_1, beta_2, epsilon, n_iter_no_change, and max_fun provide additional options for convergence tolerance, verbosity, momentum, early stopping criteria, validation set splitting, and optimization. In total, the MLPRegressor algorithm has 23 parameters that can be adjusted to suit specific needs and optimize performance for regression tasks. The general settings are also optimised to achieve the best results that will be discussed in next sections.

3. Results and discussion

Data collection is the essential and primary step of this study to portray a realistic perspective of the river with the concentration of Cr and how it is taken up to bioaccumulate in the crops, i.e., tomato in the downstream agricultural lands irrigated by the river. Figs. 9–11 show the heatmap of the average concentration of Cr at three depths of the river (also see Figs. A3–A5 and Tables A2–A4 in Appendix A). They are based on the data collection at these depths and five sections in both directions of X (along with the river flow) and Y (perpendicular to the river flow) in both summer and winter seasons. As depicted in Fig. 9, the Cr concentration is significantly high in the areas close to the pollution source, which is around 30 m distance. Records in the same X coordinations are not similar in different Y coordinations, which emphasizes that the Cr concentration is gradually decreasing through both directions, e.g., along with the river flow and perpendicular to the river flow. By comparing the average concentrations of Cr between the two seasons, it is evident that the Cr level is higher in the winter than in the summer. In addition, the rate of reduction in the average Cr concentration is lower in the winter than in the summer, as can be spotted when comparing any two cells between the results of the two seasons.

Fig. 10 shows the records of the average Cr concentration in the second level of depth i.e., 90 cm depth from the surface of the river.

Table 2
Key features of LR and NN methods in the scikit-learn library.

Model	features
LR	<ol style="list-style-type: none">1. It is accepted for predictive modelling and making inference2. There is a high level of collective experience and expertise, including teaching materials on linear regression models and software implementations3. Linear equations have an easy-to-understand interpretation on a modular level (Hastie et al., 2009).
NN (Classifier)	<ol style="list-style-type: none">1. Neural networks require less formal statistical training to develop
NN (Regressor)	<ol style="list-style-type: none">2. Neural networks can implicitly detect complex nonlinear relationships between independent and dependent variables3. Neural networks have the ability to detect all possible interactions between predictor variables4. Neural networks can be developed using multiple different training algorithms (Tu, 1996)

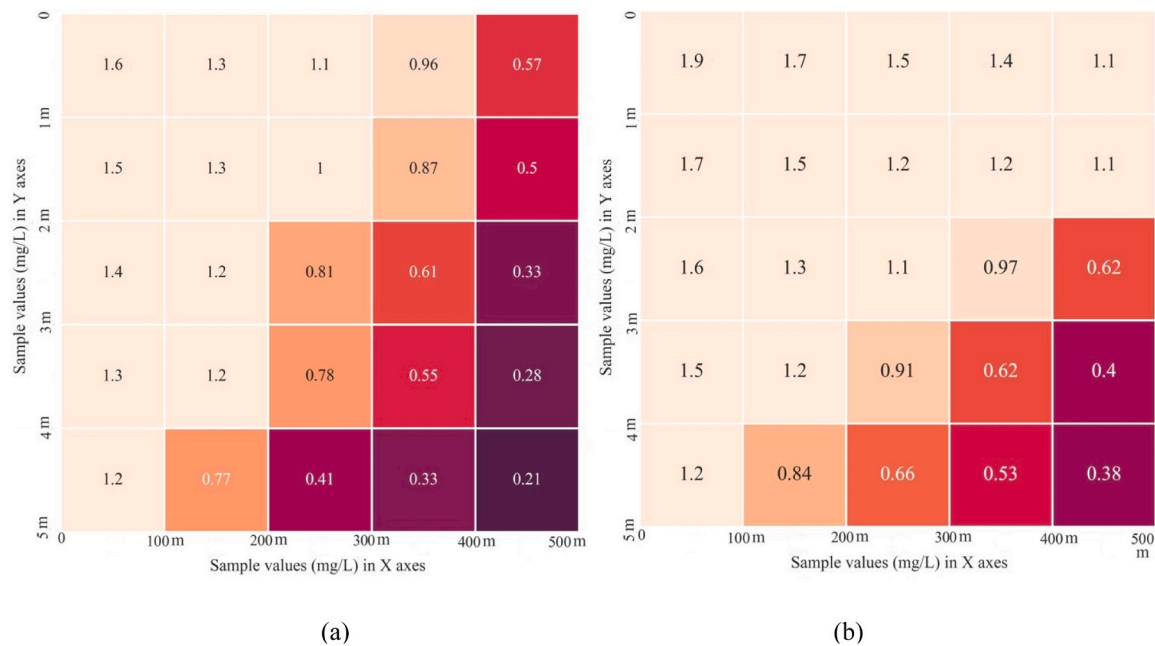


Fig. 9. Heatmap of the average concentration of Cr (mg/L) at the first level (50 cm depth) of the river in (a) the summer and (b) the winter.

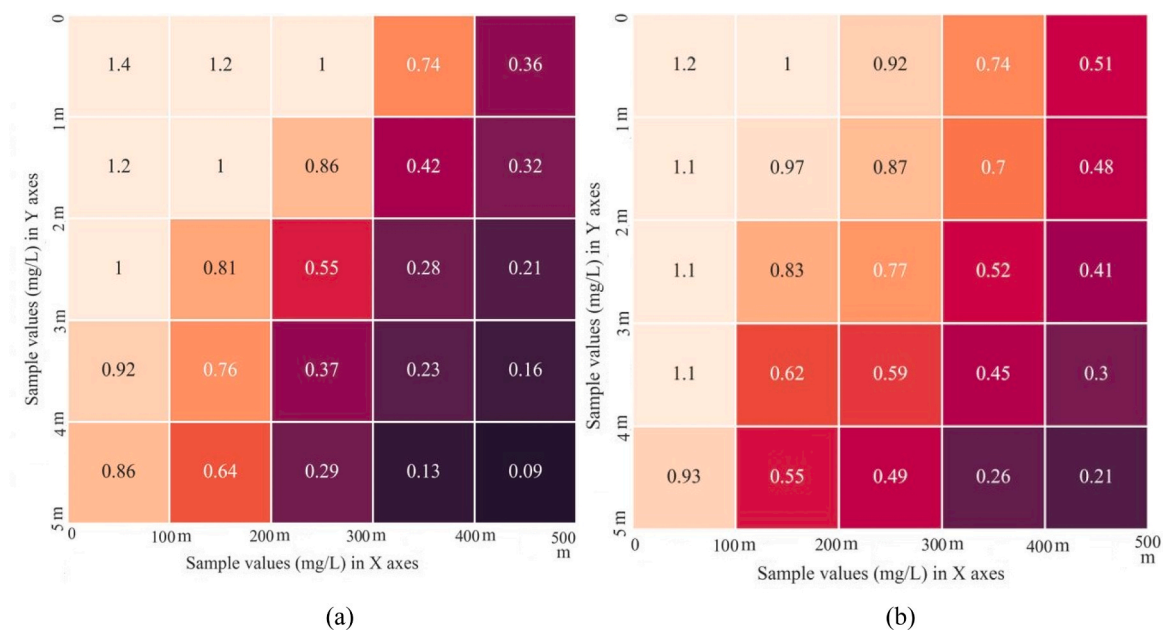


Fig. 10. Heat map of the average concentration of Cr (mg/L) at the second level (90 cm depth) of the river in (a) the summer and (b) the winter.

Fig. 10a reveals that the Cr concentration slightly decreases as depth increases along with the length of the river (i.e., X direction, and perpendicular flow direction i.e., Y direction). The trend of the Cr concentration in Fig. 10b seems to be relatively similar to the results in Fig. 11. This similarity between the records in the summer and winter is observed at the 50 cm level. Note that the average Cr concentration (mg/L) is lower in the second level compared to the first level. This may be because the dilution happens when the pollution is discharged into the water, and hence the pollution concentration reduces through the depth of the river. As depicted in Fig. 11, the average Cr concentration is significantly lower at the third level of depth than at the first and second levels which is because of diffusion phenomena in depth. According to Fick's second law, in all three dimensions with increasing the distance from the main source of the pollution, the flux of contamination is

reduced and the effects of both advection and diffusion are increased. Therefore, with increasing depth, the intensity of mass transfer is increased, and the pollution is much more diluted. This fact also appears in the experimental practices and is depicted in Figs A3–5 in Appendix A. In addition, the concentration ratio of the lowest level to the middle level is equal to the same ratio of the middle level to the highest level. It is clear that the Cr concentration is diluted from the water sampling point before reaching to plant and the pollution level from mg/L is reduced to $\mu\text{g/L}$.

Fig. 12 shows the variation of some parameters obtained by tracing the Cr concentration in the tomato plants for up to 77 days, i.e. the harvesting time. These parameters include temperature (Fig. 12a), TDS (Fig. 12b), pH (Fig. 12c), total hardness (Fig. 12d), Sodium Adsorption Ratio (SAR) (Fig. 12e), mean concentration of Cr in water (Fig. 12f) and

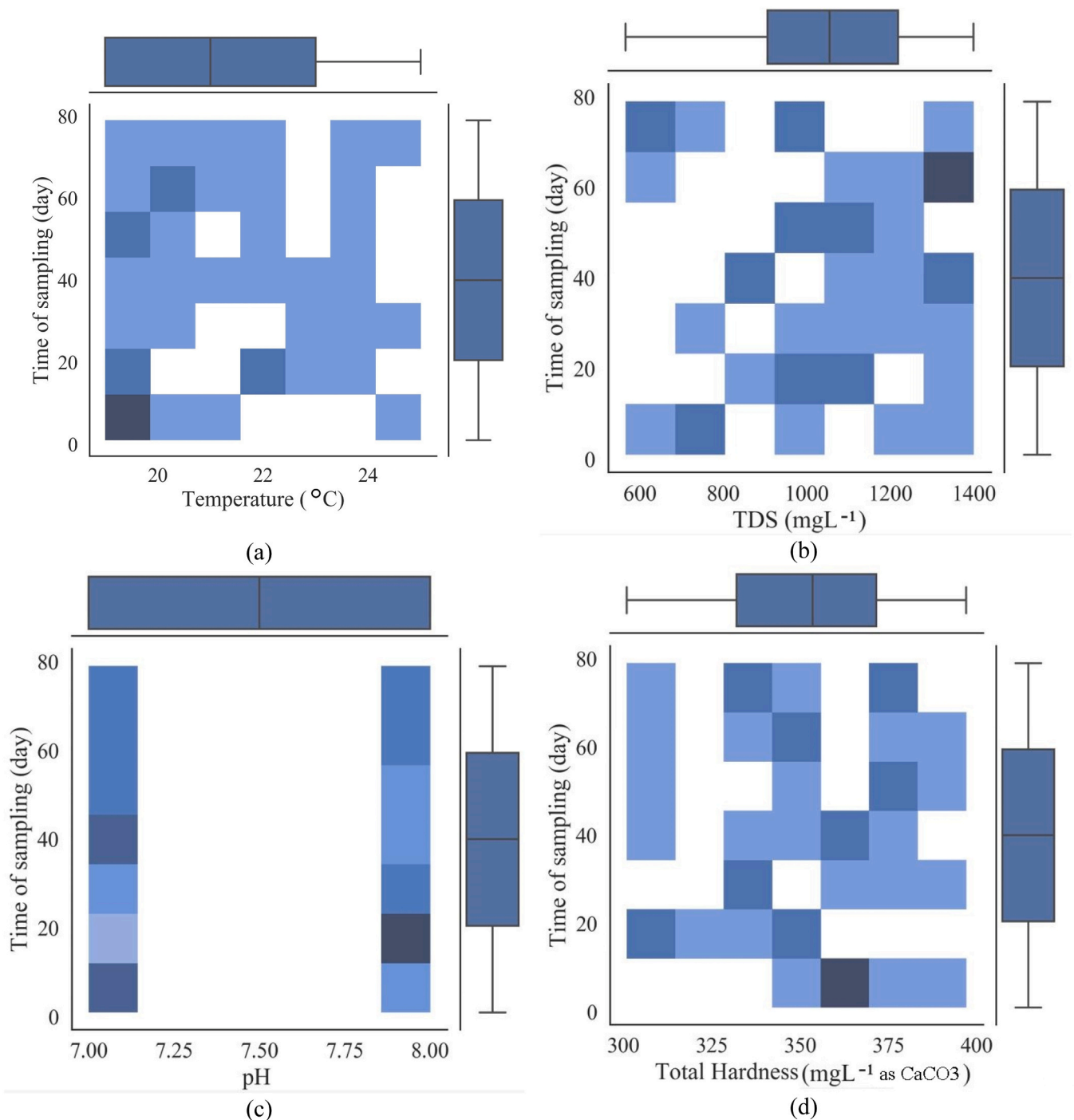


Fig. 11. Heat map of Cr average injection concentration (mg/L) at the third level (130 cm depth) of the river in (a) the summer and (b) the winter.

the Cr concentration in tomatoes (Fig. 12g). Fig. 12a displays that the temperature was mainly around 22 °C although it was up to 24 °C for some limited days. Fig. 12b illustrates Total Dissolved Solids (TDS) which are fairly spread between 600 and 1400 mg/L within the 77 days of the experiment. Fig. 12c shows pH ranging between two integer values of 7 and 8 in the leaves of tomato plants. Variations of total hardness (Fig. 12d) and the mean concentration of Cr (Fig. 12f) in irrigation water samples are similar to TDS in Fig. 12b. Likewise, variation of SAR in Fig. 12e is similar to pH and takes only two values (either 2 or 3). Interestingly, the variation of the Cr concentration in tomatoes (Fig. 12g) is mainly around 0.56 µg/g at the beginning of the experiment and gradually releases during the experiment up to over 1.29 µg/g that

might be related to the the irrigation by the river.

The parameters measured during the 77-day experiment are then used as a database to create ML modelling in this study. Hence, the correlation of these parameters with the concentration of Cr in tomato samples is first analysed using the linear regression (LR) model. The best correlation is obtained for the LR model with the coefficient of determination (R^2) reaching 0.931 for the Cr concentration in the tomato samples. The coefficient of determination for other LR models is 0.655 for the time of sampling, 0.760 for temperature, - 0.011 for TDS, 1.667 for pH, - 0.009 for Total Hardness, 1.479 for SAR, and 1.175 for the mean Cr concentration (µg/L) in the irrigation water sample. This low rate of correlation indicates some parameters such as pH and SAR have

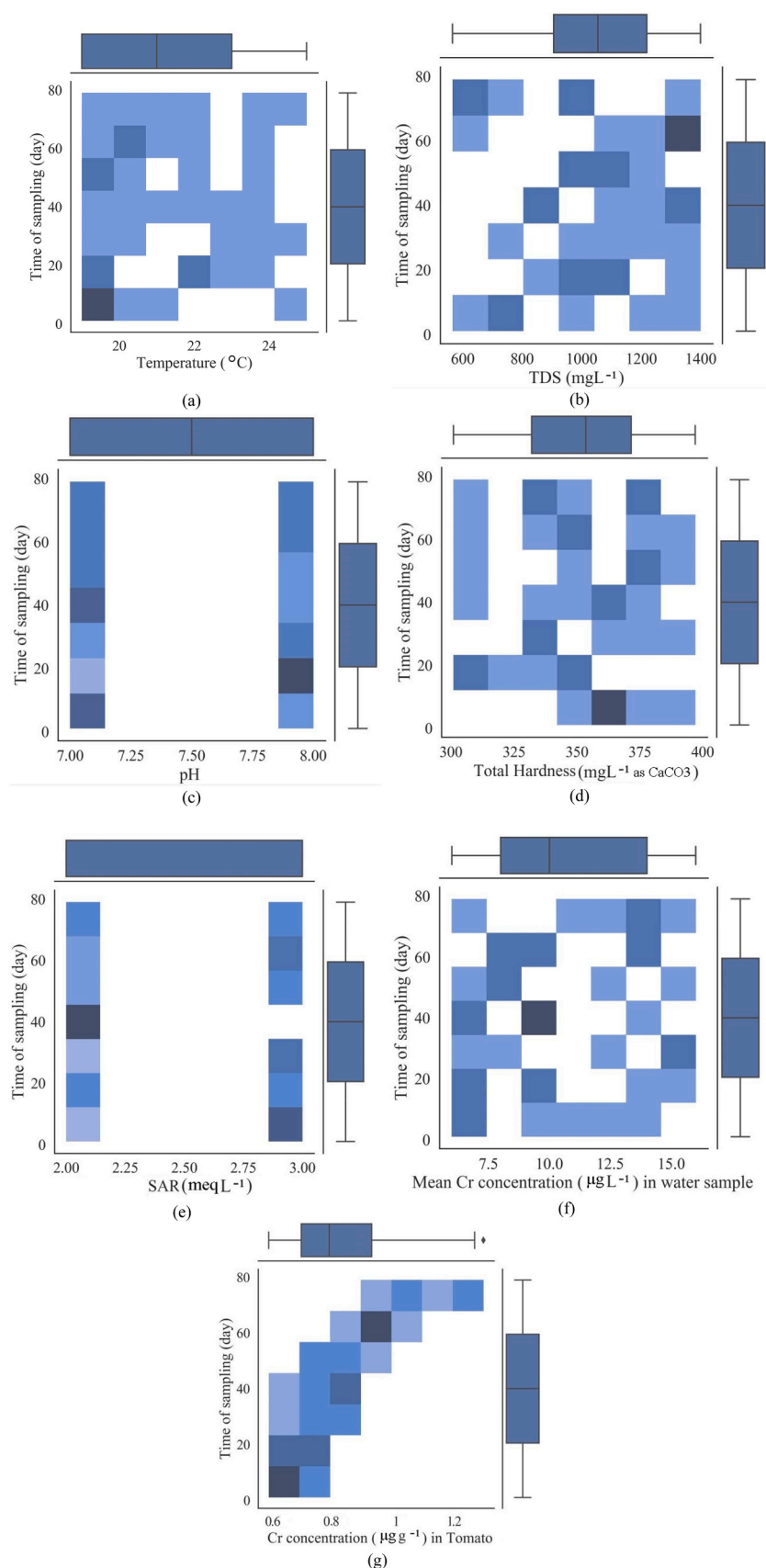


Fig. 12. Parameters collected every two days including (a) Temperature ($^{\circ}\text{C}$), (b) TDS (mg/L), (c) pH, (d) Total Hardness (e) Sodium Adsorption Ratio (SAR), (f) Mean Cr concentration ($\mu\text{g/L}$) in water sample and (g) the concentration of Cr ($\mu\text{g/g}$) in tomato.

no major impacts on the concentration of Cr. The ML model is also trained and tested for the database, as 70% of recorded data is used for training and the remaining 30% is used for testing.

The NN models are developed here for the database collected within the 77-day experiment. Based on the settings in Table 2, R^2 for the NN models is 0.743 for the NN classifier and 0.841 for the NN regressor. These scores are solid statements indicating the dataset is linear and the trend of Cr accumulated in the tomato plants as a simple linear regression is more accurate than a complicated NN model with more than 100 hidden layers. However, by increasing the number of the hidden layers in the NN algorithms to 500, and the maximum number of iterations from 200 to 300, R^2 can be improved to 0.874 and 0.946 which is higher than the LR's R^2 . However, the NN regressor is more sophisticated than the NN classifier. This comparison, to some extent, is done with other research, but here scores are generally higher. For instance, the best score of other studies was reported up to 0.89 for R^2 (Quang et al., 2022). Eid et al. (2022) developed a special linear regression model for estimating the concentration of 10 heavy metals and found the amount of heavy metal in soil, pH, and organic matter content affecting heavy metal concentrations in *Hordeum vulgare* tissues. They achieved the highest R^2 of 0.76 for Cr in the grains harvested after 77 days. Also, the highest and the lowest R^2 in the study were related to the Mn in root and Cu in grains by 0.96 and 0.39 respectively. Hu et al. (2020) estimated and compared the factors controlling the heavy metal (HM) uptake by plants using Random Forest (RF), Gradient Boosted Machine (GBM), and Generalised Linear (GLM) models in soil-crop systems in the Yangtze River Delta, China. Results showed the best prediction for the RF followed by GBM and linear methods. The most important relative variables for estimating the Cr concentration were plant type, elevation, heavy metals in soil, and soil organic materials. The R^2 for estimating the concentration of Cr was 0.59, 0.52, and 0.27 in RF, GBM, and GLM respectively. Novotná et al. (2015) developed regression models for HM uptake into some crops. The influence of measured soil concentrations and soil factors (pH, organic carbon, content of silt and clay) on the Cr concentrations in plants was evaluated using multivariate regressions. The results showed R^2 of 0.55 for Hop and 0.25 for grass mowing. Yu et al. (2016) launched a linear regression model to predict HM concentration in wheat grains. They used pH, organic matter, and salt concentration in the soil as the most affecting factors, but the results did not show a good correlation for Cr concentration. Kumar et al. (2019b) explored the HM uptake by cauliflower through multivariate regression analysis and found that R^2 is equal to 0.8, 0.9 and 0.83 for Cr concentration in root, leaves, and florescence of *B. oleracea*, respectively. Eid et al. (2021) investigated the HM concentration in the okra plant (*Abelmoschus esculentus* (L.) Moench) grown in greenhouse conditions and soil amended with sewage sludge. The Cr concentration was more concentrated in the roots than in any other parts of the plant. The metal bioaccumulation factors were negatively correlated with the pH of the soil and positively correlated with soil organic matter content. They used a regression model and the results showed R^2 equal to 0.79, 0.90, 0.80, and 0.82 for Cr concentration in fruits, leaves, stems, and roots, respectively.

4. Integrated management of the system

Discharging heavy metals by industries into receiving water bodies needs to be monitored and analysed as presented above. However, the whole ecosystem, including the water body, environment, human activities, and wastewater discharge from industries and agriculture, also needs integrated management under a holistic framework to minimise the impact on the environment and human life. Given significant technical efforts and advancements available worldwide to remove the pollution of heavy metals in water bodies, the integrated management should contain a mechanism to prevent the pollution discharge, minimise the negative impacts on aquatic lives and finally minimise the irrigation of agricultural lands by contaminated water. However, the

lack of managerial efforts may be significant despite international conventions. This study suggests that there is an essential need for establishing organisations at both national and local levels run by decision-makers who are aware of SDGs, especially LBW. In this study, a healthy river used to irrigate a major part of one of the greatest provinces in Iran has been affected by unsustainable, short-time industrial and economic purposes. In addition, there is no valid and scientific estimation of the amount of human health damage due to living close to the river or consuming products from those agricultural lands. Following integrated management of the whole system, Fig. 13 shows a conceptual model of what is needed in this context to protect their natural environment and human life. As depicted in the figure, desirable results are achieved by the close collaboration of experts, engagement of stakeholders (including relevant industries and farmers), and decision-makers (including policymakers) who can set out the necessary regulations for water resources conservation and environmental protection. Furthermore, technical analysis and findings should be translated into managerial instruments and instructions utilised by those who have the power to enforce and monitor the integrated system.

5. Conclusions and future prospects

The study examined a new approach for the fate and transportation of Cr in a real-world case study of the river ecosystem by analysing its concentration at the discharge point up to the end of the irrigated plants and its bioaccumulation in agricultural plants irrigated by the contaminated water. In addition to measuring the concentration of Cr, several important parameters (i.e., temperature, pH, SAR, Total Hardness, TDS) of the tomato plant (the most popular crop production in the pilot study) were analysed throughout the 77 days before harvesting tomatoes. Three ML techniques (LR, NN classifier, and NN regressor) were also developed to identify the correlation between the concentration of Cr in the plant leaves and other parameters. Key findings include:

- Pollution levels in the river varied across depths and seasons, with higher concentrations observed in winter due to increased industrial activity and river characteristics.
- Bioaccumulation of Cr in plants initially measured at 0.56 µg/g, gradually increasing to 1.29 µg/g by the harvesting day, likely due to continuous irrigation with contaminated water or nutrients from the accumulated soil.
- Three machine learning techniques demonstrated acceptable coefficient of determination especially NN regressor and LR, indicating their potential for estimating Cr concentration in tomato leaves based on influencing plant parameters.

The study suggests the need for further research to develop an integrated management approach for Cr and other heavy metals in the river, involving stakeholders, creating comprehensive spatial maps of heavy metals, tracing the impact on human consumers, and employing more robust AI methods based on extensive databases for validation and model enhancement.

CRedit authorship contribution statement

Mohammad Gheibi and Kourosh Behzadian devised the project, the main conceptual ideas and proof outline. Ali Montazeri and Mohammad Gheibi worked out almost all of the technical details, mathematical formulation and performed the optimization. Benyamin Chahkandi, and Mohammad Gheibi developed some novel ideas in this research. Mohammad Eftekhari, Stanislaw Wacławek, Mohammad Gheibi, Kourosh Behzadian and Luiza C Campos supervised the designing the methodology, findings of this work and they worked out the validation, editing and reviewing the paper. All authors discussed the results and contributed to the final manuscript.

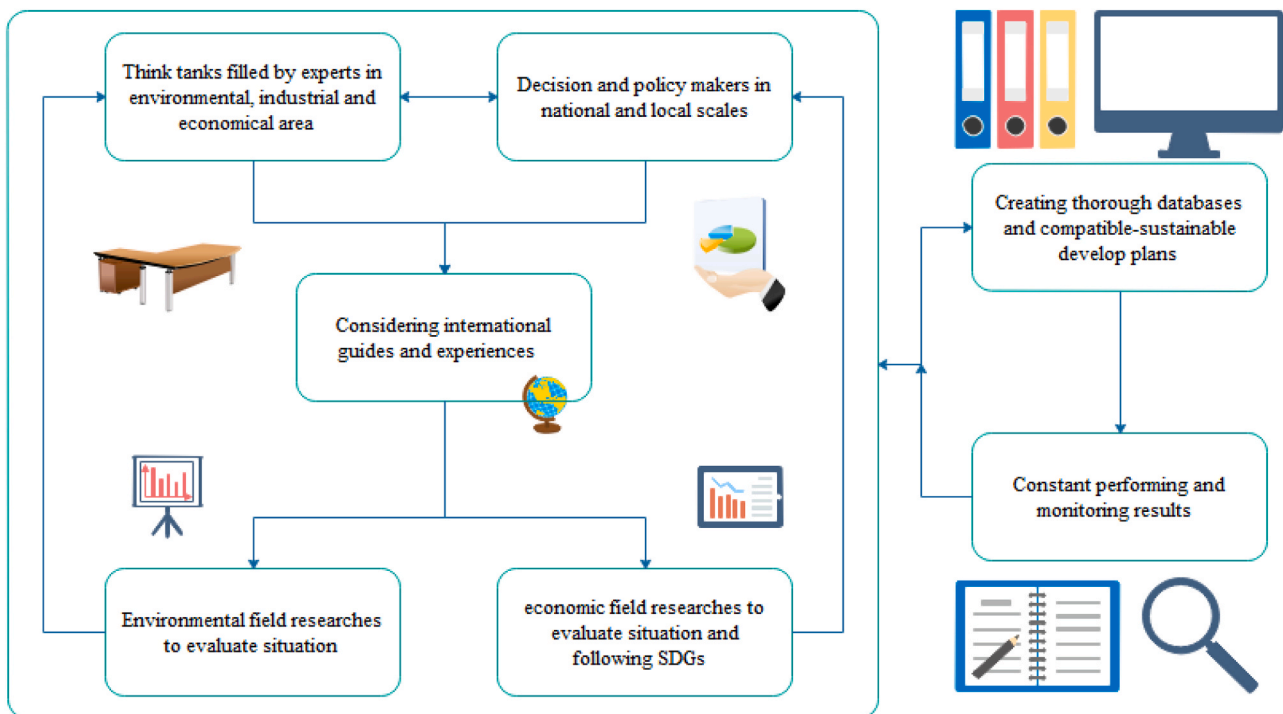


Fig. 13. Conceptual model for establishing an organisation for the protection of the environment and human health.

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

During the revisions of the manuscript, the authors used ChatGPT in order to enhance the language and readability of part of the text only. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

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.

Acknowledgement

The authors also wish to thank the editor and the anonymous reviewers for making constructive comments which substantially improved the quality of the paper.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ecoenv.2023.115269](https://doi.org/10.1016/j.ecoenv.2023.115269).

References

- Adams, M.L., Zhao, F.J., McGrath, S.P., Nicholson, F.A., Chambers, B.J., 2004. Predicting cadmium concentrations in wheat and barley grain using soil properties. *J. Environ. Qual.* 33, 532–541. <https://doi.org/10.2134/jeq2004.0532>.
- Ahmed, A.S.S., Hossain, M.B., Babu, S.M.O.F., Rahman, M., Sun, J., Sarker, M.S.I., 2022. Spatial distribution, source apportionment, and associated risks of trace metals (As, Pb, Cr, Cd, and Hg) from a subtropical river, Gomti, Bangladesh. *Int. J. Sediment Res.* 37 (1), 83–96.
- Ali, M.M., Rahman, S., Islam, M.S., Rakib, M.R.J., Hossen, S., Rahman, M.Z., Kormoker, T., Idris, A.M., Phoungthong, K., 2022. Distribution of heavy metals in water and sediment of an urban river in a developing country: a probabilistic risk assessment. *Int. J. Sediment Res.* 37 (2), 173–187.
- AMBRUS, A., 1979. The influence of sampling methods and other field techniques on the results of residue analysis. In *Pesticide residues* (pp. 6–18). Pergamon.
- Benedetti, R., Piersimoni, F., Bee, M., Espia, G. (Eds.), 2010. *Agricultural survey methods*. John Wiley & Sons.
- Boshoff, M., De Jonge, M., Scheifler, R., Bervoets, L., 2014. Predicting As, Cd, Cu, Pb and Zn levels in grasses (*Agrostis* sp. And *Poa* sp.) and stinging nettle (*Urtica dioica*) applying soil–plant transfer models. *Sci. Total Environ.* 493, 862–871.
- Capello, M.A., 2022. President's Page: advancing sustainability in SEG. *Lead. Edge* 41 (1), 6–7.
- Chojnacka, K., Chojnacki, A., Górecka, H., Górecki, H., 2005. Bioavailability of heavy metals from polluted soils to plants. *Sci. Total Environ.* 337 (1–3), 175–182.
- Cottee, J., López-Avilés, A., Behzadian, K., Bradley, D., Butler, D., Downing, C., Farmani, R., Ingram, J., Leach, M., Pike, A. and De Propriis, L., 2016. The local nexus network: Exploring the future of localised food systems and associated energy and water supply. In *Sustainable design and manufacturing 2016* (pp. 613–624). Springer International Publishing.
- Davari, S., Basafa, H., Toghræi, M., 2020. Kashafrud river (Mashhad plain) from the neolithic to the middle bronze age. *J. Archaeol. Stud.* 12 (3), 41–63.
- Dehghani, A., Roohi Aminjan, A., Dehghani, A., 2022. Trophic transfer, bioaccumulation, and health risk assessment of heavy metals in Aras River: case study—amphipoda-zander–human. *Environ. Sci. Pollut. Res.* 1–10.
- Eid, E.M., Shaltout, K.H., Abdallah, S.M., Galal, T.M., El-Bebany, A.F., Sewelam, N.A., 2020. Uptake prediction of ten heavy metals by *Eruca sativa* Mill. cultivated in soils amended with sewage sludge. *Bull. Environ. Contam. Toxicol.* 104 (1), 134–143.
- Eid, E.M., Shaltout, K.H., Alamri, S.A., Alrumman, S.A., Sewelam, N., Taher, M.A., Hashem, M., Mostafa, Y.S., Ahmed, M.T., 2021. Prediction models founded on soil characteristics for the estimated uptake of nine metals by okra plant, *Abelmoschus esculentus* (L.) Moench, cultivated in agricultural soils modified with varying sewage sludge concentrations. *Sustainability* 13 (22), 12356.
- Eid, E.M., Khedher, K.M., Ayed, H., Arshad, M., Mouldi, A., Shaltout, K.H., Sewelam, N. A., Galal, T.M., El-Bebany, A.F., Alshehri, A.M., 2022. Prediction models based on soil properties for evaluating the heavy metal uptake into *Hordeum vulgare* L. grown in agricultural soils amended with different rates of sewage sludge. *Int. J. Environ. Health Res.* 32 (1), 106–120.
- Gulseven, O., 2020. Measuring achievements towards SDG 14, life below water, in the United Arab Emirates. *Mar. Policy* 117, 103972.
- Hajinamaki, S., Emami, H., Fotovat, A., Haghnia, G.H., 2016. Effects of long-term irrigation with treated wastewater of parkand abad refinery on some soil chemical and biological properties. *Iran. J. Irrig. Drain.* 10 (2), 208–220.
- Hastie, T., Tibshirani, R., Friedman, J.H. and Friedman, J.H., 2009. *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1–758). New York: springer.
- Capello, M.A., 2022. President's Page: Advancing sustainability in SEG. *The Leading Edge*, 41(1), pp.6–7.
- Hernanda, T., Giyono, U., 2022. Environmental legal protection of rivers in the perspective of sustainable development. *J. Jurisprud.* 11 (1), 100–113.

- Horn, S., Vogt, T., Gerber, E., Vogt, B., Bouwman, H., Pieters, R., 2022. HIV-antiretrovirals in river water from Gauteng. In: *Science of The Total Environment*, 806. Mixed messages of wastewater inflows as source, South Africa, 150346.
- Hseu, Z.Y., 2004. Evaluating heavy metal contents in nine composts using four digestion methods. *Bioresour. Technol.* 95 (1), 53–59.
- 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.* 262, 114308.
- Islam, M.S., Khalid, Z.B., Gabar, S.M., Yahaya, F.M., 2022. Heavy metals pollution sources of the surface water of the Tunggak and Balok river in the Gebeng industrial area, Pahang, Malaysia. *Int. J. Energy Water Resour.* 1–8.
- Joshi, S., Temgire, S., 2022. Life below water and the fate of humanity: with special reference to the efforts by adidas towards reducing the effects of plastic waste on the marine life. *Sustain. Agri Food and Environ. Res.* 10, 1.
- Krauss, M., Wilcke, W., Kobza, J., Zech, W., 2002. Predicting heavy metal transfer from soil to plant: potential use of Freundlich-type functions. *J. Plant Nutr. Soil Sci.* 165 (1), 3–8.
- Kumar, V., Singh, J., Kumar, P., 2019a. Heavy metal uptake by water lettuce (*Pistia stratiotes* L.) from paper mill effluent (PME): experimental and prediction modeling studies. *Environ. Sci. Pollut. Res.* 26, 14400–14413.
- Kumar, V., Thakur, R.K., Kumar, P., 2019b. Assessment of heavy metals uptake by cauliflower (*Brassica oleracea* var. botrytis) grown in integrated industrial effluent irrigated soils: A prediction modeling study. *Sci. Hortic.* 257, 108682.
- Landa-Cansigno, O., Behzadian, K., Davila-Cano, D.I., Campos, L.C., 2020. Performance assessment of water reuse strategies using integrated framework of urban water metabolism and water-energy-pollution nexus. *Environ. Sci. Pollut. Res.* 27, 4582–4597.
- Liang, H.M., Lin, T.H., Chiou, J.M., Yeh, K.C., 2009. Model evaluation of the phytoextraction potential of heavy metal hyperaccumulators and non-hyperaccumulators. *Environ. Pollut.* 157 (6), 1945–1952.
- Lu, J., 2022. Can the central environmental protection inspection reduce transboundary pollution? Evidence from river water quality data in China. *J. Clean. Prod.* 332, 130030.
- Lučić, M., Mikac, N., Vdović, N., Bačić, N., Nava, V., Vidmar, J., Milačić, R., 2022. Spatial and temporal variability and sources of dissolved trace elements in the Sava River (Slovenia, Croatia). *Environ. Sci. Pollut. Res.* 1–15.
- Lv, J., Li, Y.P., Huang, G.H., Ding, Y.K., Li, X., Li, Y., 2022. Planning energy economy and eco-environment nexus system under uncertainty: a copula-based stochastic multi-level programming method. *Appl. Energy* 312, 118736.
- MacFarlane, G.R., Pulkownik, A., Burchett, M.D., 2003. Accumulation and distribution of heavy metals in the grey mangrove, *Avicennia marina* (Forsk.) Vierh.: biological indication potential. *Environ. Pollut.* 123 (1), 139–151.
- Mathur, S., 2004. Modeling phytoremediation of soils. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*, 8(4), pp.286–297.
- Mokarram, M., Pourghasemi, H.R., Huang, K., Zhang, H., 2022. Investigation of water quality and its spatial distribution in the Kor River basin, Fars province, Iran. *Environ. Res.* 204, 112294.
- Montazeri, A., Chahkandi, B., Gheibi, M., Behzadian, K., Ardakanian, A. and Ieropoulos, I., 2022. A new empirical approach for modelling fate and transport of Chromium bioaccumulation in irrigated crops: a water-food-pollution nexus.
- Novotná, M., Mikeš, O. and Komprdová, K., 2015. Development and comparison of regression models for the uptake of metals into various field crops. *Environmental Pollution*, 207, pp.357–364.
- Gulseven, O., 2020. Measuring achievements towards SDG 14, life below water, in the United Arab Emirates. *Marine Policy*, 117, p.103972.
- Omuku, P.E., Odidika, C.C., Ozukwe, A.E., Iwuozor, K.O., 2022. A comparative evaluation of rain water obtained from corrugated roofing sheets within Awka Metropolis, Anambra State. *Iranian (Iranica) Journal of Energy & Environment* 13 (2), 134–140.
- Quang, N.H., Quinn, C.H., Carrie, R., Stringer, L.C., Hackney, C.R. and Van Tan, D., 2022. Comparisons of regression and machine learning methods for estimating mangrove above-ground biomass using multiple remote sensing data in the red River Estuaries of Vietnam. *Remote Sensing Applications: Society and Environment*, p.100725.
- Shaabani, Z., Esmaili-Sari, A., Moradi, A.M., Taghavi, L., Farsad, F., 2022. Possible health risk assessment for heavy metal concentrations in water, sediment, and fish species and Turkmen pregnant women's biomonitoring in Miankaleh Peninsula, Iran. *Environ. Sci. Pollut. Res.* 1–17.
- Shahradnia, H., Chamani, A., Zamanpoore, M., 2022. Linking river sediment arsenic to catchment spatial attributes in agricultural landscapes. *Environ. Sci. Pollut. Res.* 29 (2), 2830–2838.
- Sposito, G., Mattigod, S.V., 1977. On the chemical foundation of the sodium adsorption ratio. *Soil Sci. Soc. Am. J.* 41 (2), 323–329.
- Tang, D.D., Yuan, X.Y., Wang, Y.M., Ji, J.F., Wen, Y.B., Zhao, W.F., 2018. Enrichment characteristics and risk prediction of heavy metals for rice grains growing in paddy soils with a high geological background. *J. Agro Environ. Sci.* 37, 18–26. <https://doi.org/10.11654/jaes.2017-0801>.
- Thoma, G.J., Lam, T.B., Wolf, D.C., 2003. A mathematical model of phytoremediation for petroleum-contaminated soil: model development. *Int. J. Phytoremediat.* 5 (1), 41–55.
- Tu, J.V., 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, 49(11), pp.1225–1231.
- Qi, Q. and Song, S., 2022. Development Level Evaluation of Water Ecological Civilization in Yangtze River Economic Belt. *Discrete Dynamics in Nature and Society*, 2022.
- Tuovinen, T.S., Roivainen, P., Makkonen, S., Kolehmainen, M., Holopainen, T., Juutilainen, J., 2011. Soil-to-plant transfer of elements is not linear: results for five elements relevant to radioactive waste in five boreal forest species. *Sci. Total Environ.* 410, 191–197.
- Veldhuis, A.J., Glover, J., Bradley, D., Behzadian, K., López-Avilés, A., Cottee, J., Downing, C., Ingram, J., Leach, M., Farmani, R., Butler, D., 2019. Re-distributed manufacturing and the food-water-energy nexus: opportunities and challenges. *Prod. Plan. Control* 30 (7), 593–609.
- Verma, P., George, K.V., Singh, H.V., Singh, R.N., 2007. Modeling cadmium accumulation in radish, carrot, spinach and cabbage. *Appl. Math. Model.* 31 (8), 1652–1661.
- Wu, S.C., Peng, X.L., Cheung, K.C., Liu, S.L., Wong, M.H., 2009. Adsorption kinetics of Pb and Cd by two plant growth promoting rhizobacteria. *Bioresour. Technol.* 100 (20), 4559–4563.
- Ye, X., Li, H., Ma, Y., Wu, L., Sun, B., 2014. The bioaccumulation of cd in rice grains in paddy soils as affected and predicted by soil properties. *J. Soils Sediment.* 14, 1407–1416. <https://doi.org/10.1007/s11368-014-0901-9>.
- Yu, X., Wang, Z., Lynn, A., Cai, J., Huangfu, Y., Geng, Y., Tang, J., Zeng, X., 2016. Heavy metals in wheat grown in sewage irrigation: a distribution and prediction model. *Pol. J. Environ. Stud.* 25 (1), 413–418.
- Yutao, H., Weilin, S., Juan, J., Shoujun, S., Yang, L., Jinping, W., 2022. Prediction of Soil Heavy Metals Based on BP Neural Network and Assessment of Human Health Risk of an Electroplating Plant. *Asian J. Ecotoxicol.* 2, 278–288.
- Zhao, X., Joo, J.C., Lee, J.K., Kim, J.Y., 2019. Mathematical estimation of heavy metal accumulations in *Helianthus annuus* L. with a sigmoid heavy metal uptake model. *Chemosphere* 220, 965–973.
- Zhou, Y., Jia, Z., Wang, J., Chen, L., Zou, M., Li, Y., Zhou, S., 2019. Heavy metal distribution, relationship and prediction in a wheat-rice rotation system. *Geoderma* 354, 113886.