environmental and earth sciences proceedings

MDPI

Proceeding Paper

Alleviating Health Risks for Water Safety: A Systematic Review on Artificial Intelligence-Assisted Modelling of Proximity-Dependent Emerging Pollutants in Aquatic Systems †

Marc Deo Jeremiah Victorio Rupin [1,*], Kylle Gabriel Cruz Mendoza [1] and Rugi Vicente Rubi [2,3]

[1] Chemical Engineering Department, College of Engineering, Pamantasan ng Lungsod ng Maynila, General Luna, Corner Muralla St., Intramuros, Manila 1002, Philippines; kgcmendoza.nano@gmail.com
[2] Chemical Engineering Department, College of Engineering, Adamson University, 900 San Marcelino St. Ermita, Manila 1002, Philippines; rugi.vicente.rubi@adamson.edu.ph
[3] Adamson University Laboratory of Biomass, Energy and Nanotechnology (ALBEN), Adamson University, 900 San Marcelino St., Ermita, Manila 1000, Philippines
* Correspondence: rupin.marcdeo14@gmail.com
† Presented at the 8th International Electronic Conference on Water Sciences, 14–16 October 2024; Available online: https://sciforum.net/event/ECWS-8.

Abstract: Emerging pollutants such as pharmaceuticals, industrial chemicals, heavy metals, and microplastics are a growing ecological risk affecting water and soil resources. Another challenge in current wastewater treatments includes tracking and treating these pollutants, which can be costly. As a growing concern, emerging pollutants do not have lower limit levels and can be detrimental to aquatic resources in minuscule amounts. Thus, the assessment of multiple emerging water pollutants in community-based water sources such as surface water and groundwater is a prioritized area of study for water resource management. It provides a basis for the ecological health management of arising diseases such as cancer and dengue caused by unsafe water sources. Accordingly, by utilizing artificial intelligence, wide-range and data-driven insights can be synthesized to assist in water resource management and propose solution pathways without the need for exhaustive experimentation. This systematic review examines the artificial intelligence-assisted modelling of water resource management for emerging water pollutants, notably machine learning and deep learning models, with proximity dependence and correlated synergistic health effects for both humans and aquatic life. This study underscores the increasing accumulation of these emerging pollutants and their toxicological effects on the community and how data-driven modelling can be utilized to assist in addressing research gaps related to water treatment methods for these pollutants.

Keywords: water resource management; artificial intelligence; waterborne diseases; health and safety; emerging pollutants

check for updates

Academic Editor: Helena M. Ramos

Published: 21 February 2025

Citation: Rupin, M.D.J.V.; Mendoza, K.G.C.; Rubi, R.V. Alleviating Health Risks for Water Safety: A Systematic Review on Artificial Intelligence-Assisted Modelling of Proximity-Dependent Emerging Pollutants in Aquatic Systems. *Environ. Earth Sci. Proc.* **2025**, *32*, 7. https://doi.org/10.3390/eesp2025032007

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The continuous industrialization and expansion of production for human consumption have led to the increasing usage of different chemicals that persist in aquatic environments. Emerging pollutants (EPs) such as industrial chemicals, such as fire retardants and surfactants; agrochemicals, such as pesticides and fertilizers [1]; pharmaceuticals and fragrances; and heavy metals are increasingly detected in water systems such as groundwater [1–3], surface water, wastewater, and other streams, such as agricultural runoff [4–6]. As emerging

pollutants, their fate, transport, and accumulation potential are understudied, especially regarding their effects on the environment [7].

Currently, studies in the literature have focused on the detection and evaluation of these emerging pollutants in areas affected by human activity, such as runoffs and wastewater [8]. The growing global concern about these pollutants is mainly focused on their effects on aquatic wildlife and their bioaccumulation, which could lead them to circulate along the food chain. Furthermore, clean water sites, such as riverine and mangrove environments, have been affected by heavy metals [9] and pharmaceuticals [10]. Long-term exposure to these affected water systems is currently understudied and becoming the focus of newer studies. However, the effects of the consumption of these emerging pollutants are well documented, and they can cause toxicological, endocrine-disrupting, and carcinogenic effects when humans are exposed to them through either direct exposure, such as dermal or oral contact or inhalation [11,12], or through food, water, and bioaccumulation in food systems [13–15].

Due to the nature of the fate and transport of these emerging pollutants and their suspected effects, studies are focused on modelling the transport of these pollutants to improve detection, reduce susceptibility to other fauna and flora, and develop action plans for health effects among humans [16]. The most ubiquitous modern modelling tools involve the use of machine learning, supported by geographical data features to bolster the correlation between the theorized health effects and the concentration of pollutants [17]. These models assist in hazard mapping and improving water treatment facilities for better water quality [18,19]. Furthermore, these models can be supported by real-time data collection and the processing of data for sudden increases in emerging pollutants while being connected to geographical positioning systems (GPSs) to improve water resource management. Modelling the fate of these emerging pollutants was first carried out using geographical information system (GIS) models with statistical tools such as the least absolute shrinkage and selection operator (LASSO); now, modelling is supported using machine learning algorithms that use non-linear functions and introduce complexity that further mirrors the reality of emerging pollutant transport.

This paper reports current trends in using both machine learning and geographical data to improve the detection and hazard mapping of emergent pollutants in different water systems such as groundwater, surface water, and wastewater, supported by artificial intelligence models. Furthermore, this study highlights the effects of these emergent pollutants relative to their concentration and location in the local environment. Lastly, current gaps in the literature and possible directions for future research are presented to further research on the proximity-based modelling of emerging pollutants.

2. Method

This study followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines to review the relevant literature and studies on the proximity-based artificial intelligence modelling of emerging pollutants. Using PRISMA guidelines for literature reviews helps in recognizing the available literature through a deductive approach, further improving the focus of the study from the overall available literature through the initial search and identification. Figure 1 summarizes the process of selecting studies from the available body of knowledge.

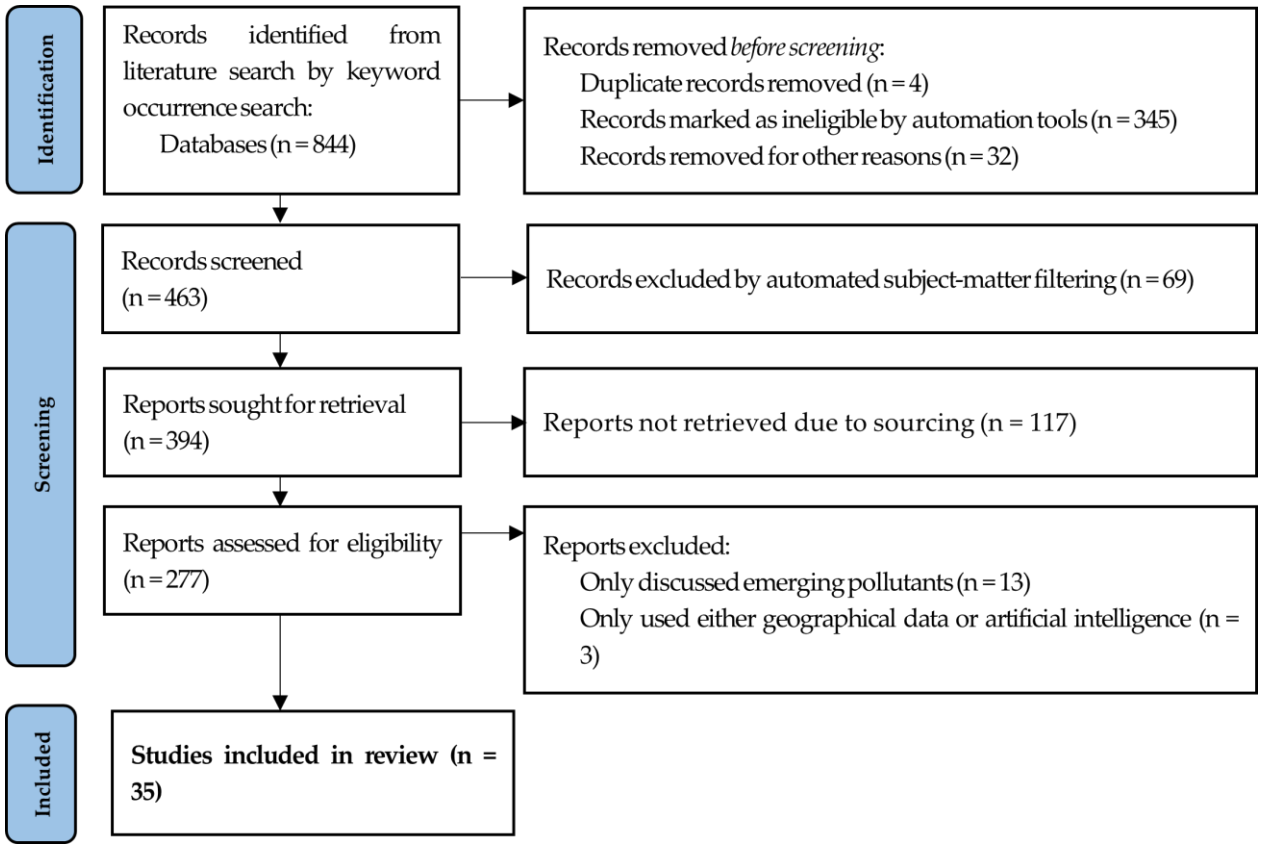


Figure 1. Filtering of the relevant literature using the PRISMA guidelines.

The keyword search included the terms "emerging pollutants", "water resources", "aquatic resources", "pharmaceuticals in water", "microplastics in water", "persistent organic compounds", "machine learning and artificial intelligence", and "geographical information systems". The combination of these terms led to a total of 844 studies, of which 117 were not accessible. The inaccessibility of these 117 studies does not limit our understanding of the proximity-dependent modelling of emerging pollutants due to the repeated usage of common artificial intelligence algorithms such neural networks and random forests. However, these 117 studies do provide better context for regions in East Asia and the Mediterranean as the main locales of the studies.

Out of 277 studies, 13 studies used experimental methods on the identification of emerging pollutants in water; 3 studies only used machine learning without geographical data or geographical data only without artificial intelligence; and 242 studies discussed emerging pollutants in soil or air resources. In this review, a total of 35 studies were collected and analyzed.

3. Proximity-Based Machine Learning for Emerging Pollutants in Water Resource Management

In the current available literature, both groundwater and surface water are reported to be tainted with differing amounts of EPs based on their distance from anthropogenic activities [20]. These pollutants are also found in soil and other resources related to water usage. These activities can be agricultural, industrial, or residential, and EPs can be attributed to the continued use of products such as pesticides, pharmaceuticals, mining equipment, and solvents. Depending on physicochemical properties [16], such as volatility, pH, and turbidity, and macroscopic features, such as topographical and spatiotemporal data, the concentration, persistence, and accumulation potential of emerging pollutants can

differ greatly between source points. These EPs carry various levels of risk that need to be accounted for to assist in their possible removal [21].

To provide assistance in removing these persistent organic pollutants, several studies utilized geographical data with artificial intelligence to delineate concentration hotspots for future treatment efforts. In developing machine learning (ML) models for these emerging pollutants, the most common form of ML utilized comprised supervised models and ensemble models, such as random forests and support vector machines, while deep learning models, such as neural networks, are now being utilized. Meanwhile, the most common equation used for geographical features involves groundwater vulnerability models such as the depth of water, recharge, aquifer media, soil media, topography, impact of the vadose zone, and hydraulic conductivity (DRASTIC) model [22]. Figure 2 summarizes the ubiquitous components of a proximity-dependent artificial intelligence model.

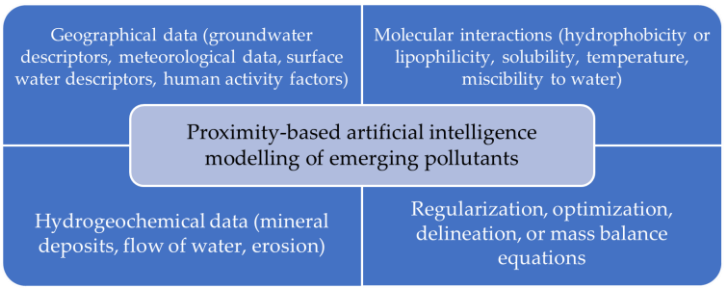


Figure 2. Components of proximity-based artificial intelligence model.

Similarly, for surface water and wastewater, distance is the primary representative feature for machine learning. Furthermore, additional features such as Shapley Additive Explanation (SHAP) values can be applied to machine learning, which involve geographical data to better identify input contributions to the overall prediction [23]. Meanwhile, rainfall amount and quarterly data have been used in studies as features to address heavy rainfall and seasonal changes. With these additional features added to artificial intelligence models, their accuracy and predictive power increased when simulating emerging pollutant fate and transport. As different features are caused by pollutants in the focus areas and the available geographical data, multiple artificial intelligence models have been observed to be tested across various studies, but unsupervised and deep learning models were preferred in this study.

3.1. Pesticides and Persistent Organic Pollutants (POPs)

Agricultural runoff and industrial wastewater contributed to the greatest discharge of pesticides and other POPs in surface water throughout the hydrological cycle [24]. Furthermore, different forms of pesticides, such as organochlorines, pyrethoids, organophosphates, carbamates, and biological pest repellants, are widely used in the market. Consequentially, the continuous use of organochlorines affects the food chain due to their lipophilic nature and the fact that they tend to be stored in animal fats, while organophosphates are water-soluble and percolate through soil towards aquifers and groundwater, affecting their quality [25]. Meanwhile, pesticides such as carbamates tend to slide off in soil and are washed off in surface water. Evidenced in riverine environments in the Tengi River Basin in Malaysia, four pesticides proved nearly impossible to remove in drinking water supplies through conventional water treatment facilities, and these continue to circulate within the ecosystem, leading to further accumulation [26].

Dichlorodiphenyltrichloroethane (DDT) and its metabolite (DDE) were modelled using random forest classification with soil, water, and geographical data within the Mediterranean coastline. The study was able to delineate concentrations of up to 1.93 µg/L of

DDT-DDE at close distances to the Vinalopo River and 1.19 µg/L for drainage canals in the Segura River [4]. On the other hand, 1,2,3-trichloropropane, a nematicide in groundwater, was modelled in California Central Valley, using Random Forests as the ML model [1]. After the analysis, the estimated TCP amount increased by 110,000 kg up to 4.3 million kg in relation to its position relative to collection sites and industrial areas [1].

In Quanzhou Bay, the persistence of polycyclic aromatic hydrocarbons (PAHs) and organochlorine pesticides was investigated using geographical data and the least absolute shrinkage and selection operator, or LASSO, depicting Quanzhou Bay as having a low-to-medium risk for the dispersion of the two POPs considered in the study [27]. Random forests were applied to groundwater in Colorado to track perfluorooctane sulfonates and perfluorooctanates to detect risks within the area, including the major use of private wells for water, reaching different levels of precision from 55% to 90% depending on the risk category of the area, with moderate risk at 55% precision, high risk at 71% precision, and low risk at 90% precision [28]. Next, in the Yeongsan River Basin, six simulated pesticides—acetamiprid, hexaconazole, aldicarb, sulfone, azoxystrobin, metalaxyl, and bentazon—were modelled with long short-term memory (LSTM) and convolutional neural networks (CNNs) using a soil and water assessment tool (SWAT) with R^2 values of 0.99 and 0.76 for training and validation, respectively [29].

Currently, artificial intelligence and machine learning models delineate persistent organic pollutants within areas near agricultural runoffs more effectively when compared to purely geographical data or hydrogeochemical data (Soil and Water Assessment Tools [SWATs], DRASTIC) or statistical data analysis with or without simulations (multivariate statistics, Toxic Substances in Surface Waters [TOXSWA]). This can be attributed to the complexity associated with artificial intelligence models [30–32]. These data models can be on par with machine learning models but require larger datasets and fewer complexities. Meanwhile, in producing data for action, stakeholder health can be monitored through human health risk analysis with proximity-based machine learning models [17] or molecular descriptors for translocation and accumulation within root and tissue systems of flora [33].

3.2. Heavy Metals

A study in Abakaliki, Nigeria, showed that high concentrations of heavy metals were found in active mining sites from the mineralization of sulphide-rich ores, with the concentrations of lead as Pb^{2+} (11.42 mg/L), As (4.13 mg/L), Se^{2+} (2.68 mg/L), Cd^{2+} (15.67 mg/L), and Hg^{2+} (2.60 mg/L) being relatively higher in groundwater than the recommended concentrations for drinking water. The associated high values of these heavy metals were attributed to their low mobility while being freed in active mine sites. The presence of heavy metals in drinking and potable water systems is dangerous due to their carcinogenic nature and other diseases-causing capabilities, such as Parkinson's disease and selenosis [34].

Furthermore, heavy metals can also exist in turbid waters and water resources for flora, such as mangroves. Randomized tree models with Local Indicators of Spatial Association (LISA) analysis were proven to predict heavy metals in mangrove settlements at an accuracy of 72.72% and delineate the heavy contamination and percent content of cadmium, lead, arsenic, and chromium [10].

Meanwhile, groundwater analysis in Kaduna State in Nigeria, when coupled with emotional artificial neural networks (EANNs), was able to detect that heightened levels of magnesium, iron, and mercury in groundwater can be attributed to both local mineral content (geochemical) and human activity [3]. Also, groundwater in El Kharga, Egypt, was sufficiently modelled using artificial neural networks at $R^2 = 0.99$ with geographical techniques to model its water quality index and dermal hazard index due to its water

containing iron and manganese, alongside other pollutants [35]. Next, in Saudi Arabia, linear regression models with water quality index and geographical data inputs were able to surmise that heavy metal contamination can be caused by hydrogeochemical effects, and it was deduced that, at most, 30.00% of the observed groundwater system was unfit for drinking [36].

This form of analysis can also be used in surface water for heavy metals such as cadmium, iron, lead, zinc, and copper, as modelled using artificial neural networks (ANNs), random forests (RFs), and support vector machines (SVMs) for three different river systems in Morocco [37]. Also, modelling in highly economic areas for surface water contamination was carried out, such that neural networks and random forest algorithms applied to the water quality indices for the Danube River were sufficiently predictive and used to identify the point of contact for contamination for industrial complexes near surface waters [38]. While chemical interactions are not added as features in the modelling of heavy metals for water resource management, they are usually aggregated to produce a common hazard index or toxicological index to consider their effects after interactions. Notably, these indices can account for both free metal ions and complexed metal ions, as they are analyzed after chelation [39].

Surface water contaminated with heavy metals heightens health risks for communities that use it as a water source. In developing models using human health risks, it can be deduced that several heavy metals such as iron, zinc, chromium, and nickel produce both non-carcinogenic and carcinogenic risks in water samples in Nigeria [40]. Other than the presence of heavy metals in water, correlations with different health effects were also modelled, including geographical data for assessing water quality. Using a geographically weighted random forest (GWRF) model with heavy metals and gastrointestinal system cancers, pollutants such as arsenic were quantified across nine river basins in China with correlations with cancers such as esophageal, gallbladder, and pancreatic cancer [41].

While the analysis of water samples with heavy metals is a growing technology when coupled with the Internet of Things (IoT), such as sensors and testing sites for point detection, extra caution was raised when correlating health data and instances in these models due to ethical considerations regarding sensitive data. By understanding how patient anonymity, data aggregation, and data security can play a role in water resources and ecosystem management, better water resources and health models can be created with artificial intelligence. [42]

3.3. Pharmaceuticals, Hormone-Distrupting Compounds, Fragrances, and Personal Care Products

Pharmaceuticals in wastewater are commonly attributed to anthropogenic use, especially in hospitals and care facilities. The presence of β-estradiol across Potomac and Chesapeake Bay in the United States was modelled with random forests after principal component analysis, amounting to six to seven predictors such as precipitation amount and elevation for estrogenic activity in the context of immediate collection and upstream collection [7]. This development indicated that feature engineering could be an area of interest for modelling the fate and transport of pharmaceuticals such as rainfall data and meteorological and anthropological seasonal changes [43].

In China, thioethers were determined to be heavily reliant on geographical location and dissolved organic carbon when modelled with classification and regression tree analyses. Other regression methods can be used to deduce that fragrance derivatives and odorants such as cresols and benzaldehydes can co-occur in water systems and arise both naturally and from industrial use [44].

3.4. Microplastics

While most emerging pollutants are caused by increased industrial practices using ecological resources, microplastics are purely derived from human activity. The challenge of developing models for microplastic pollution detection was dependent on different polymer types of microplastics. Surface water wetlands were modelled as a secondary treatment after municipal treatment in northern Finland using random forests from the Bayreuth Particle Finder to identify polymer types across the wetland [45]. Three different collection sites were used to identify possible sources of contamination from the initially treated water.

Meanwhile, microplastic trends were analyzed and simulated for municipal wastewater in Phuket using five different convolutional neural networks. Due to the high accuracy derived from simulations, a similar mobile application that uses a phone as a stereomicroscope was created to further improve microplastic analysis [46].

The main challenge in developing models for microplastics transport is its dynamic nature, such that material properties can change, and extensive data collection is required to create more accurate models that can delineate both point detection and seasonal changes. One example of this interaction is the formation of colloidal interactions when dispersed in water, making microplastics transport sensitive to water parameters such as flow, temperature, and amount [47]. These challenges can even affect models developed for detection and analysis, such as the stereomicroscope.

3.5. Summary

In summary, the current form of research using spatiotemporal data with water quality and resource management relies on multiple collection sites, the characterization of economic zones such as heavily industrial or residential areas, water quality indices, hazard indices, and hydrogeochemical factors. The most common models used were random forests (classification models) or neural networks (deep learning models) since they can reliably delineate geographical factors or model non-linear behaviour. Table 1 summarizes the overall findings of this study with the accuracy and precision of artificial intelligence models when supported with spatiotemporal and geographical data.

Table 1. Some findings for modelling with artificial intelligence and geographical data.

Emerging Pollutants of Interest	Study	Models Used	Performance Matrix
DDT-DDE	[4]	Random Forest (Mutual Information)	Accuracy = 0.815 (0.852)
TCP	[1]	Classification and Regression Tree, Random Forest, Boosted Regression Trees	R^2 = 0.020, 0.44, 0.41
PFOS, PFOA	[28]	Random Forest	Accuracy = 86% (low risk), 80% (medium risk), 90% (high risk)
Acetamiprid, aldicarb sulfone, azoxystrobin, bentazon, hexaconazole, metalaxyl	[29]	Long-term Short Memory (LSTM), Convolutional Neural Networks (CNNs)	For bentazon, LSTM: Training (R^2 = 0.99), Validation (R^2 = 0.93)
Water quality index, mercury, iron, lead	[3]	Emotional Artificial Neural Network	R^2 = 0.89 (Training) and 0.83 (Validation)
Heavy metal pollution index	[37]	Artificial Neural Network, Support Vector Machines	Zn (R^2 = 0.98–0.99), Cd (R^2 = 0.89–0.96), Pb (R^2 = 0.999–0.998)

Table 1. *Cont.*

Emerging Pollutants of Interest	Study	Models Used	Performance Matrix
Estrogenic activity	[7]	Random Forests with Principal Component Analysis	R^2 = 0.00–0.05 (Non-Potomac/Chesapeake Bay Watersheds), 0.02–0.11 (Potomac Temporal Comparison Sites)
Microplastics	[46]	Convolutional Neural Network with Transfer Learning	R^2 = 0.98

4. Current State and Future Outlook for Modelled and Simulated Water Resources

4.1. Ongoing Research on Water Resources and Artificial Intelligence

One aspect of supporting analysis with artificial intelligence was the use of feature importance analysis and feature engineering. Feature importance analysis involved the use of statistical and non-linear tools to reduce resource usage by limiting the number of features or considered variables in modelling. The most ubiquitous models include Shapley Additive Explanations (SHAPs), Taylor decomposition, occlusion analysis, and other models. These techniques employed values that determine relevance and provide variables that greatly affect the accuracy of the model [48]. Through feature importance analysis, the development of models could include locale data on pollutant levels, increased land usage, and meteorological changes.

Machine learning and artificial intelligence are currently growing tools used for scientific inquiry. One area of study on water resources involves the use of neural networks on laboratory data such as surface-enhanced Raman scattering (SERS) [49] or imagery data for plastic fragments and fibres in water [46]. Other laboratory data, such as gas chromatography results from groundwater with organic micropollutants, were able to detect contaminated areas through non-targeted screening techniques with neural networks [50]. These models can provide greater accuracy in the point detection or source detection of pollutants when supported with proximity-based modelling and further contextualize their effects on the area through delineating high levels of pollutants.

Models have been further developed to take a human-focused approach, now considering the effects of emerging pollutant abundance on the current ecosystem, with current research [21] having detecting accumulation in seafood and microplastics in urban environments [51].

4.2. Future Research and Trajectory of Research

Geographically supported models can be further understood as delineating models to identify areas for water treatment and support in order to facilitate clean water access. While these models have reached higher accuracies and reliabilities, there exists a level of complexity in adding more features into current artificial intelligence-based models. The largest challenge that computational studies face is the complexity and resource trade-off. By understanding how computational resources can be further utilized while having more features to quantify water quality, accuracy can be maintained while memory utilization stays low. This area of research will aid in the use of these models in remote areas in which high computational usage is not preferred.

However, current key research areas are now shifting towards human-focused modelling of water resources with assistance from other computational techniques, such as game theory with Shapley value-supported learning and transfer learning. Other research

could also focus on other ecological resources, such as soil and air, with current studies bolstering our understanding of their interdependence and how cycles can contribute to the distribution of emerging pollutants. Modelling these pollutants raises a challenge due to ethical challenges in using human data. These challenges can include data privacy, aggregation, and the anonymization of sensitive data. Meanwhile, current challenges in proximity-dependent modelling include seasonal changes and long-term modelling.

Overall, geographically supported artificial intelligence for water resource management, especially for emerging pollutants, is a growing and promising field. The key areas of study focus on feature engineering for modelling, techno-economic analysis for treatment, the effectiveness of natural water treatment facilities, correlation with health endemics, and socioeconomic correlations with water quality. By addressing the computational trade-off, geographical simulations with artificial intelligence can greatly assist in dealing with large-scale water treatment projects, even in currently unreachable areas, by finding the overall fate and transport of pollutants. Eventually, artificial intelligence will be used to optimize resource utilization by reducing the costs associated with analogue simulations.

Author Contributions: M.D.J.V.R.—conceptualization, draft preparation, writing, and review; K.G.C.M.—draft preparation, reviewing; R.V.R.—conceptualization, writing, and review. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data are contained in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hauptman, B.H.; Naughton, C.C.; Harmon, T.C. Using Machine Learning to Predict 1,2,3-Trichloropropane Contamination from Legacy Non-Point Source Pollution of Groundwater in California's Central Valley. *Groundw. Sustain. Dev.* **2023**, *22*, 100955. [CrossRef]
2. Karimzadeh Motlagh, Z.; Derakhshani, R.; Sayadi, M.H. Groundwater Vulnerability Assessment in Central Iran: Integration of GIS-Based DRASTIC Model and a Machine Learning Approach. *Groundw. Sustain. Dev.* **2023**, *23*, 101037. [CrossRef]
3. Vivan, E.L.; Bashir, F.M.; Eziashi, A.C.; Gammoudi, T.; Dodo, Y.A. Ground Water Quality Evaluation Using Hydrogeochemical Characterization and Novel Machine Learning in the Chikun Local Government Area of Kaduna State, Nigeria. *Water Sci. Technol.* **2023**, *88*, 1875–1892. [CrossRef]
4. Melendez-Pastor, I.; Lopez-Granado, O.M.; Navarro-Pedreño, J.; Hernández, E.I.; Jordán Vidal, M.M.; Gómez Lucas, I. Environmental Factors Influencing DDT–DDE Spatial Distribution in an Agricultural Drainage System Determined by Using Machine Learning Techniques. *Environ. Geochem. Health* **2023**, *45*, 9067–9085. [CrossRef]
5. Omeka, M.E. Evaluation and Prediction of Irrigation Water Quality of an Agricultural District, SE Nigeria: An Integrated Heuristic GIS-Based and Machine Learning Approach. *Environ. Sci. Pollut. Res.* **2024**, *31*, 54178–54203. [CrossRef] [PubMed]
6. Salele, B.; Dodo, Y.A.; Sani, D.A.; Abuhussain, M.A.; Sayfutdinovna Abdullaeva, B.; Brysiewicz, A. Run-off Modelling of Pervious and Impervious Areas Using Couple SWAT and a Novel Machine Learning Model in Cross-Rivers State Nigeria. *Water Sci. Technol.* **2023**, *88*, 1893–1909. [CrossRef] [PubMed]
7. Gordon, S.; Jones, D.K.; Blazer, V.S.; Iwanowicz, L.; Williams, B.; Smalling, K. Modeling Estrogenic Activity in Streams throughout the Potomac and Chesapeake Bay Watersheds. *Environ. Monit. Assess.* **2021**, *193*, 105. [CrossRef] [PubMed]
8. Mojtahedi, A.; Dadashzadeh, M.; Azizkhani, M.; Mohammadian, A.; Almasi, R. Assessing Climate and Human Activity Effects on Lake Characteristics Using Spatio-Temporal Satellite Data and an Emotional Neural Network. *Environ. Earth Sci.* **2022**, *81*, 61. [CrossRef]
9. Omeka, M.E.; Igwe, O.; Onwuka, O.S.; Nwodo, O.M.; Ugar, S.I.; Undiandeye, P.A.; Anyanwu, I.E. Efficacy of GIS-Based AHP and Data-Driven Intelligent Machine Learning Algorithms for Irrigation Water Quality Prediction in an Agricultural-Mine District within the Lower Benue Trough, Nigeria. *Environ. Sci. Pollut. Res.* **2024**, *31*, 54204–54233. [CrossRef]

10. Proshad, R.; Rahim, M.A.; Rahman, M.; Asif, M.R.; Dey, H.C.; Khurram, D.; Al, M.A.; Islam, M.; Idris, A.M. Utilizing Machine Learning to Evaluate Heavy Metal Pollution in the World's Largest Mangrove Forest. *Sci. Total Environ.* **2024**, *951*, 175746. [CrossRef]
11. Sajjadi, S.A.; Mohammadi, A.; Khosravi, R.; Zarei, A. Distribution, Exposure, and Human Health Risk Analysis of Heavy Metals in Drinking Groundwater of Ghayen County, Iran. *Geocarto Int.* **2022**, *37*, 13127–13144. [CrossRef]
12. Witkowska, D.; Słowik, J.; Chilicka, K. Heavy Metals and Human Health: Possible Exposure Pathways and the Competition for Protein Binding Sites. *Molecules* **2021**, *26*, 6060. [CrossRef]
13. Ali, S.; Ullah, M.I.; Sajjad, A.; Shakeel, Q.; Hussain, A. Environmental and Health Effects of Pesticide Residues. In *Sustainable Agriculture Reviews 48: Pesticide Occurrence, Analysis and Remediation Vol. 2 Analysis*; Inamuddin, Ahamed, M.I., Lichtfouse, E., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 311–336, ISBN 978-3-030-54719-6.
14. Kiran; Bharti, R.; Sharma, R. Effect of Heavy Metals: An Overview. *Mater. Today Proc.* **2022**, *51*, 880–885. [CrossRef]
15. Zaynab, M.; Al-Yahyai, R.; Ameen, A.; Sharif, Y.; Ali, L.; Fatima, M.; Khan, K.A.; Li, S. Health and Environmental Effects of Heavy Metals. *J. King Saud Univ. Sci.* **2022**, *34*, 101653. [CrossRef]
16. Malek, N.H.A.; Yaacob, W.F.W.; Nasir, S.A.M.; Shaadan, N. The Effect of Chemical Parameters on Water Quality Index in Machine Learning Studies: A Meta-Analysis. *J. Phys. Conf. Ser.* **2021**, *2084*, 012007. [CrossRef]
17. Wang, X.; Yu, D.; Ma, L.; Lu, X.; Song, J.; Lei, M. Using Big Data Searching and Machine Learning to Predict Human Health Risk Probability from Pesticide Site Soils in China. *J. Environ. Manag.* **2022**, *320*, 115798. [CrossRef]
18. El-Rawy, M.; Wahba, M.; Fathi, H.; Alshehri, F.; Abdalla, F.; El Attar, R.M. Assessment of Groundwater Quality in Arid Regions Utilizing Principal Component Analysis, GIS, and Machine Learning Techniques. *Mar. Pollut. Bull.* **2024**, *205*, 116645. [CrossRef]
19. Xiangcao, Z.; Su, C.; Xianjun, X.; Ge, W.; Xiao, Z.; Yang, L.; Pan, H. Employing Machine Learning to Predict the Occurrence and Spatial Variability of High Fluoride Groundwater in Intensively Irrigated Areas. *Appl. Geochem.* **2024**, *167*, 106000. [CrossRef]
20. Azizi, K.; Ayoubi, S.; Nabiollahi, K.; Garosi, Y.; Gislum, R. Predicting Heavy Metal Contents by Applying Machine Learning Approaches and Environmental Covariates in West of Iran. *J. Geochem. Explor.* **2022**, *233*, 106921. [CrossRef]
21. Kumar, V.; Kumar, S. ANN-Based Integrated Risk Ranking Approach: A Case Study of Contaminants of Emerging Concern of Fish and Seafood in Europe. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1598. [CrossRef]
22. Ijlil, S.; Essahlaoui, A.; Mohajane, M.; Essahlaoui, N.; Mili, E.M.; Van Rompaey, A. Machine Learning Algorithms for Modeling and Mapping of Groundwater Pollution Risk: A Study to Reach Water Security and Sustainable Development (Sdg) Goals in a Mediterranean Aquifer System. *Remote Sens.* **2022**, *14*, 2379. [CrossRef]
23. Wang, R.; Kim, J.-H.; Li, M.-H. Predicting Stream Water Quality under Different Urban Development Pattern Scenarios with an Interpretable Machine Learning Approach. *Sci. Total Environ.* **2021**, *761*, 144057. [CrossRef] [PubMed]
24. de Souza, R.M.; Seibert, D.; Quesada, H.B.; de Jesus Bassetti, F.; Fagundes-Klen, M.R.; Bergamasco, R. Occurrence, Impacts and General Aspects of Pesticides in Surface Water: A Review. *Process Saf. Environ. Prot.* **2020**, *135*, 22–37. [CrossRef]
25. Syafrudin, M.; Kristanti, R.A.; Yuniarto, A.; Hadibarata, T.; Rhee, J.; Al-onazi, W.A.; Algarni, T.S.; Almarri, A.H.; Al-Mohaimeed, A.M. Pesticides in Drinking Water—A Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 468. [CrossRef] [PubMed]
26. Elfikrie, N.; Bin Ho, Y.; Zaidon, S.Z.; Juahir, H.; Tan, E.S.S. Occurrence of Pesticides in Surface Water, Pesticides Removal Efficiency in Drinking Water Treatment Plant and Potential Health Risk to Consumers in Tengi River Basin, Malaysia. *Sci. Total Environ.* **2020**, *712*, 136540. [CrossRef]
27. Liu, H.; Hu, J.; Tan, Y.; Zheng, Z.; Liu, M.; Lohmann, R.; Vojta, S.; Katz, S.; Liu, Y.; Li, Z.; et al. Identification of Key Anthropogenic and Land Use Factors and Ecological Risk Assessment of Dissolved Polycyclic Aromatic Hydrocarbons (PAHs) and Organochlorine Pesticides (OCPs) in an Urbanized Estuary in China. *Mar. Pollut. Bull.* **2024**, *207*, 116876. [CrossRef]
28. Barton, K.E.; Anthamatten, P.J.; Adgate, J.L.; McKenzie, L.M.; Starling, A.P.; Berg, K.; Murphy, R.C.; Richardson, K. A Data-Driven Approach to Identifying PFAS Water Sampling Priorities in Colorado, United States. *J. Expo. Sci. Environ. Epidemiol.* **2024**. [CrossRef] [PubMed]
29. Yun, D.; Abbas, A.; Jeon, J.; Ligaray, M.; Baek, S.-S.; Cho, K.H. Developing a Deep Learning Model for the Simulation of Micro-Pollutants in a Watershed. *J. Clean. Prod.* **2021**, *300*, 126858. [CrossRef]
30. Taylor, A.C.; Mills, G.A.; Gravell, A.; Kerwick, M.; Fones, G.R. Pesticide Fate during Drinking Water Treatment Determined through Passive Sampling Combined with Suspect Screening and Multivariate Statistical Analysis. *Water Res.* **2022**, *222*, 118865. [CrossRef]
31. Wendell, A.-K.; Guse, B.; Bieger, K.; Wagner, P.D.; Kiesel, J.; Ulrich, U.; Fohrer, N. A Spatio-Temporal Analysis of Environmental Fate and Transport Processes of Pesticides and Their Transformation Products in Agricultural Landscapes Dominated by Subsurface Drainage with SWAT+. *Sci. Total Environ.* **2024**, *945*, 173629. [CrossRef]
32. Zhang, J.; Mahmood, A.; Shao, Y.; Jarosiewicz, P.; Gonsior, G.; Cuellar-Bermudez, S.P.; Chen, Z.; Stibany, F.; Schäffer, A. Combined Simulation on Pesticides Fate, Toxicities and Ecological Risk in Rice Paddies for Sustainable Development Goals Achievements. *Sci. Total Environ.* **2024**, *951*, 175552. [CrossRef] [PubMed]

33. Bagheri, M.; Al-jabery, K.; Wunsch, D.; Burken, J.G. Examining Plant Uptake and Translocation of Emerging Contaminants Using Machine Learning: Implications to Food Security. *Sci. Total Environ.* **2020**, *698*, 133999. [CrossRef] [PubMed]
34. Obasi, P.N.; Akudinobi, B.B. Potential Health Risk and Levels of Heavy Metals in Water Resources of Lead–Zinc Mining Communities of Abakaliki, Southeast Nigeria. *Appl. Water Sci.* **2020**, *10*, 184. [CrossRef]
35. Gad, M.; Gaagai, A.; Eid, M.H.; Szűcs, P.; Hussein, H.; Elsherbiny, O.; Elsayed, S.; Khalifa, M.M.; Moghanm, F.S.; Moustapha, M.E.; et al. Groundwater Quality and Health Risk Assessment Using Indexing Approaches, Multivariate Statistical Analysis, Artificial Neural Networks, and GIS Techniques in El Kharga Oasis, Egypt. *Water* **2023**, *15*, 1216. [CrossRef]
36. Alqarawy, A.; El Osta, M.; Masoud, M.; Elsayed, S.; Gad, M. Use of Hyperspectral Reflectance and Water Quality Indices to Assess Groundwater Quality for Drinking in Arid Regions, Saudi Arabia. *Water* **2022**, *14*, 2311. [CrossRef]
37. El Morabet, R.; Barhazi, L.; Bouhafa, S.; Dahim, M.A.; Khan, R.A.; Khan, N.A. Geospatial Distribution and Machine Learning Algorithms for Assessing Water Quality in Surface Water Bodies of Morocco. *Sci. Rep.* **2023**, *13*, 20599. [CrossRef] [PubMed]
38. Georgescu, P.-L.; Moldovanu, S.; Iticescu, C.; Calmuc, M.; Calmuc, V.; Topa, C.; Moraru, L. Assessing and Forecasting Water Quality in the Danube River by Using Neural Network Approaches. *Sci. Total Environ.* **2023**, *879*, 162998. [CrossRef] [PubMed]
39. Hama Aziz, K.H.; Mustafa, F.S.; Omer, K.M.; Hama, S.; Hamarawf, R.F.; Rahman, K.O. Heavy Metal Pollution in the Aquatic Environment: Efficient and Low-Cost Removal Approaches to Eliminate Their Toxicity: A Review. *RSC Adv.* **2023**, *13*, 17595–17610. [CrossRef] [PubMed]
40. Agbasi, J.C.; Egbueri, J.C. Assessment of PTEs in Water Resources by Integrating HHRISK Code, Water Quality Indices, Multivariate Statistics, and ANNs. *Geocarto Int.* **2022**, *37*, 10407–10433. [CrossRef]
41. Gu, W.; Xue, F.; Han, W.; Wang, Z.; Zhao, J.; Zhang, L.; Yang, C.; Jiang, J. Assessment of the Spatial Association between Multiple Pollutants of Surface Water and Digestive Cancer Incidence in China: A Novel Application of Spatial Machine Learning. *Ecol. Indic.* **2023**, *154*, 110897. [CrossRef]
42. Yadav, N.; Maurya, B.M.; Chettri, D.; Pooja; Pulwani, C.; Jajula, M.; kanda, S.S.; babu, H.W.S.; Elangovan, A.; Velusamy, P.; et al. Artificial Intelligence in Heavy Metals Detection: Methodological and Ethical Challenges. *Hyg. Environ. Health Adv.* **2023**, *7*, 100071. [CrossRef]
43. Kavianpour, B.; Piadeh, F.; Gheibi, M.; Ardakanian, A.; Behzadian, K.; Campos, L.C. Applications of Artificial Intelligence for Chemical Analysis and Monitoring of Pharmaceutical and Personal Care Products in Water and Wastewater: A Review. *Chemosphere* **2024**, *368*, 143692. [CrossRef] [PubMed]
44. Wang, C.; Gallagher, D.L.; Dietrich, A.M.; Su, M.; Wang, Q.; Guo, Q.; Zhang, J.; An, W.; Yu, J.; Yang, M. Data Analytics Determines Co-Occurrence of Odorants in Raw Water and Evaluates Drinking Water Treatment Removal Strategies. *Environ. Sci. Technol.* **2021**, *55*, 16770–16782. [CrossRef] [PubMed]
45. Büngener, L.; Postila, H.; Löder, M.G.J.; Laforsch, C.; Ronkanen, A.-K.; Heiderscheidt, E. The Fate of Microplastics from Municipal Wastewater in a Surface Flow Treatment Wetland. *Sci. Total Environ.* **2023**, *903*, 166334. [CrossRef] [PubMed]
46. Akkajit, P.; Sukkuea, A.; Thongnonghin, B. Comparative Analysis of Five Convolutional Neural Networks and Transfer Learning Classification Approach for Microplastics in Wastewater Treatment Plants. *Ecol. Inform.* **2023**, *78*, 102328. [CrossRef]
47. Yang, X.; Tang, D.W.S. Modeling Microplastic Transport through Porous Media: Challenges Arising from Dynamic Transport Behavior. *J. Hazard. Mater.* **2025**, *484*, 136728. [CrossRef]
48. Riyadh, A.; Peleato, N.M. Exploring Spatial and Temporal Importance of Input Features and the Explainability of Machine Learning-Based Modelling of Water Distribution Systems. *Digit. Chem. Eng.* **2025**, *14*, 100202. [CrossRef]
49. Huang, Y.; Yuan, B.; Wang, X.; Dai, Y.; Wang, D.; Gong, Z.; Chen, J.; Shen, L.; Fan, M.; Li, Z. Industrial Wastewater Source Tracing: The Initiative of SERS Spectral Signature Aided by a One-Dimensional Convolutional Neural Network. *Water Res.* **2023**, *232*, 119662. [CrossRef]
50. Ekpe, O.D.; Choo, G.; Kang, J.-K.; Yun, S.-T.; Oh, J.-E. Identification of Organic Chemical Indicators for Tracking Pollution Sources in Groundwater by Machine Learning from GC-HRMS-Based Suspect and Non-Target Screening Data. *Water Res.* **2024**, *252*, 121130. [CrossRef]
51. Chakraborty, T.K.; Rahman, M.S.; Nice, M.S.; Netema, B.N.; Islam, K.R.; Debnath, P.C.; Chowdhury, P.; Halder, M.; Zaman, S.; Ghosh, G.C.; et al. Application of Machine Learning and Multivariate Approaches for Assessing Microplastic Pollution and Its Associated Risks in the Urban Outdoor Environment of Bangladesh. *J. Hazard. Mater.* **2024**, *472*, 134359. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.