Supplementary Material for -

Machine Learning Models of the Geospatial Distribution of Groundwater Quality: A Systematic Review

List of Content

List of Abbreviations	Page 02
Text S1: Further explanations regarding GV assessment	Page 04
Text S2: Further explanations regarding the study area of some articles	Page 09
Text S3: An example of improving deep learning transparency	Page 25
Text S4: Further explanations regarding the MOA-based studies	Page 27
Figure S1: Historical (2005-2024) trends of the number of publications per annum of ML models related to groundwater quality. The red curve is an indicative exponential trend	Page 10
Figure S2: Geographical distribution of the study-focus of reviewed published MLGHM studies. Boundaries and names were adopted from generalized shapefiles of the world obtained from the United Nations [1] and Environmental Systems Research Institute (ESRI) [2]	Page 11
Figure S3: The proportion of the target parameters in the surveyed literature. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)	Page 12
Figure S4: Frequency of use of major categories of predictive variables. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)	Page 18
Figure S5: Sankey Diagram showing predictive variables and their frequency of use. See Tables S4 and S5 for details. (Plotted using Sankeymatic [3])	Page 19
Figure S6: The frequency of use of feature selection techniques used in the surveyed literature. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)	Page 23
Figure S7: A schematic example of increasing the transparency of a DL opaque-box model (Feature Selection image reproduced from Podgorski, *et al.* [4], licensed under CC BY 4.0, and the graph of Tansig activation function adopted from Mathworks.com [5])).	Page 26
Table S1: Recent review articles of the ML-based groundwater quality assessment	Page 5
Table S2: The keywords (besides prediction, geospatial, mapping, GIS) used during the search for MLGHM studies	Page 6
Table S3: MLGHM studies with supranational study areas	Page 9
Table S4: Details of the classification of the inputs used in the MLGHM literature	Page 14
Table S5: Components of the included predefined datasets	Page 17
Table S6: Specific predictive variables with the highest frequencies of use	Page 22
Table S7: Classification of the ML models used in the MLGHM literature	Page 24
Table S8: Application of optimization algorithms in the MLGHM literature	Page 27
Table S9: Examples of using MOAs out of the MLGHM literature	Page 27

List of Abbreviations

Full name	Acronym	Full name	Acronym	Full Name	Acronym	Full Name	Acronym
Ada-Boost	AdB	Entropy-weighted water quality index	EWQI	linear discriminant analysis	LDA	random subspace	RS
Adaptive fuzzy regression	AFR	Environmental Systems Research Institute	ESRI	logistic model tree	LMT	Random Tree	RndT
Adaptive neuro-fuzzy inference systems	ANFIS	Escherichia coli	E. coli	logistic regression	LgR	recurrent neural networks	RecNN
Affinity propagation clustering	APC	Evidential belief function	EBF	long short-term memory	LSTM	recursive feature elimination	RFE
Akaike information criterion	AIC	Explainable artificial intelligence	XAI	machine learning	ML	redundancy analysis	RDA
Aluminium	Al	Extreme gradient boosting	XGBoost	machine learning groundwater hazard mapping	MLGHM	Regression Tree	RT
Ammonia	NH_3	Extreme learning machines	ELM	Magnesium	Mg^{2+}	Relative importance	Rel. I
Ammonium	NH_4	Extremely randomized trees / extra trees	ERT	Manganese	Mn	relevance vector machine	RVM
Analytic hierarchy process	AHP	Feedforward backpropagation neural networks	FFBPN	map removal sensitivity analysis	MRSA	Residual sodium carbonate	RSC
Analytic hierarchy process	AHP	Flexible discriminant analysis	FDA	maximum entropy	MaxEnt	ridge regression	RR
Area under curve	AUC	Fluoride	F^-	mean decrease accuracy	MDA	Risk index	RI
Arsenic	As	Forward feature selection	FFS	Mean prediction interval width	MPIW	Root Mean Squared Error	RMSE
Artificial intelligence	AI	Frequency ratio	FR	membership function	MF	Rotation Forest	RotF
Artificial intelligence multiple model	AIMM	Fusion mrf	FMRF	Metaheuristic optimization algorithm	MOA	R-squared	R^2
Artificial neural network	ANN	Fuzzy c-means clustering	FCM	Meta-Model	MM	Safe Drinking Water Act	SDWA
Backward feature selection	BFS	Fuzzy membership framework	FMF	modular neural network	MNN	Seawater intrusion	SWI
Base exchange index	BEX	Fuzzy-catastrophe framework	FCF	multicue MRF	MMRF	Selenium	Se
Bayesian additive regression tree	BART	Gaussian process	GP	multilayer -field	MRF	self-organizing map	SOM
Bayesian generalized linear model	BayesGLM	Gene expression programming	GEP	multi-layer perceptron	MLP	sensitivity analysis	SensA
Bayesian information criterion	BIC	Generalized additive model	GAM	Multiple linear regression	MLR	sequential feature selection	SFS
Bayesian model averaging	BMA	Generalized linear model	GLM	Multivariate discriminant analysis	MulDA	Shannon Entropy	SEn
Bayesian networks	BN	Generalized regression neural network	GRNN	Multivariate regression splines	MARS	Shapley additive explanations	SHAP
Bayesian regularized neural network	BRNN	Genetic algorithm	GA	mutual information	MI	Sobol's main effect index	SMEI
Bayesian ridge regression	BRR	Geographic information system	GIS	naive Bayes	NB	Sodium	Na
Bicarbonate	HCO_3^-	Glm with lasso or elastic net regularization	glmnet	Naive Bayes classifier	NBC	Sodium adsorption ratio	SAR
Biochemical oxygen demand	BOD	Gradient boosting	GBoost	Naive Bayes tree	NBT	sodium bicarbonate	$NaHCO_3$
Boosted regression trees	BRT	Gradient boosting machine	GBM	National Water-Quality Assessment	NAWQA	sodium carbonate	Na_2CO_3
Boosting	Bst	Gradient boosting regression	GBR	neuro-fuzzy	NF	Soluble sodium percentage	SSP
Bootstrap aggregating	Bagging/Bag	Gradient boosting tree	GBDT	Nickel	Ni	stacking ensemble learning	SEL

Boron	B	Grid search	GS	Nitrate	NO_3^-	Stochastic Gradient Boosting	StoGB
Boruta feature selection	BrFS	Gross domestic product	GDP	Nitrite	NO_2^-	Strontium	Sr
Cadmium	Cd	Groundwater hazard	GH	Normalized difference vegetation index	NDVI	Sulphate	SO_4^{2-}
Calcium carbonate	$CaCO_3$	Groundwater hazard mapping	GHM	Olden's connection weights	OCW	supervised committee machine artificial intelligent	SCMAI
Categorical boosting	CatBoost	Group method of data handling	GMDH	Organic Matter	OM	supervised committee machines and fuzzy logic	SCFL
Cation exchange capacity	CEC	Growing neural gas	GNG	Oxidation reduction Potential	ORP	supervised intelligence committee machine	SICM
Chi-square	χ^2	Groundwater quality	GWQ	Partial dependence plot	PDP	support vector machine	SVM
Chloride	Cl^-	Hardness	H	Per- and polyfluoroalkyl substances	PFAS	support vector regression	SVR
Chromium	Cr	Hazard index	HI	Permeability index	PI	technique for order of preference by similarity	TOPSIS
Classification and regression tree	CART	Hierarchal clustering analysis	HCA	Phenol	C_6H_5OH	Terrain Ruggedness Index	TRI
Cluster analysis	CluA	Hybrid neuro-fuzzy inference system	HyFIS	Phosphorus	P	Topographic Position Index	TPI
Coactive neuro-fuzzy inference system	CANFIS	Information gain ratio	IGR	Physics-Informed Machine Learning	PIML	Topographic Wetness Index	TWI
Coefficient of determination	R^2	Informed hazard zonation	IHZ	Pollution index	Pol I	Total dissolved solids	TDS
Committee machine artificial intelligent	CMAI	Instance-based learning	IBL	Pollution Risk	PR	Trial-and-Error	T&E
Conditional mixed mrf	CMMRF	Integrated weight water quality index	IWQI	polynomial regression	PolyR	Tritium	^3H
Convolutional neural network	CNN	Iodine	I	Positive matrix factorization	PMF	Turbidity	Tr
Decision tree	DT	Iron	Fe	Potassium Permanganate	$KMnO_4$	unconventional oil and gas	UOG
Deep belief network	DBN	Jackknife test	JT	Prediction interval coverage probability	PICP	U-Net Enhanced Fourier Neural Operator	U-FNO
Deep learning	DL	Kelly index	KI	Priestley-Taylor Alpha	PTA	Uranium	U
Deep stacked autoencoder	DSA	Kernel extreme learning machine	KELM	Principal Component Analysis	PCA	vulnerability index	VI
Dense non-aqueous phase liquid	DNAPL	K-means clustering	KMC	Python for Long-term Environmental Monitoring	PyLEnM	Ward's hierarchical clustering	WHC
Disjoint aggregating	Dagging/Dag	K-medoid clustering	KMdC	Quadratic discriminant analysis	QDA	Water Quality Index	WQI
Down-sampling with replacement method	DSRM	K-nearest neighbors	KNN	Radial basis function	RBF	Wilcoxon signed-rank	WSR
Dual hybrids	DH	Land use / Land cover	LC / LU	Receiver operating characteristic	ROC	Willmott Agreement Index	WAI
Dynamic evolving neural-fuzzy inference system	DENFIS	Lasso regression	LsR	Radium	Ra	World Health Organization	WHO
Electrical conductivity	EC	Lead	Pb	Radon	Rn	Zinc	Zn
Elman neural network	ELNN	Learning vector quantization	LVQ	random forest	RF		
Entropy model averaging	EMA	Light Gradient Boosting Machine	LGBM	Random search	RndS		

Text S1

Groundwater vulnerability (GV) is divided into "specific GV" and "intrinsic GV" respectively focusing on the contaminant properties and the aquifer's physical characteristics [6,7]. So far, various strategies have been suggested for GV assessment, which can be classified into four groups, namely (i) index-based methods [8], (ii) process-based methods [9], (iii) statistical methods [10], and (iv) hybrid methods [11]. Process-based methods often assess the specific GV, and statistical and index-based methods usually deal with intrinsic GV, whereas hybrid models are used as a combination of these three methods. Further details regarding this classification can be found in Ref.[11].

Table S1. Recent review articles of the ML-based groundwater quality assessment

Ref.	Keywords
Geetha, *et al.* [12]	water quality parameters; water quality index; surface water; groundwater; monitoring; modeling; artificial intelligence
Torres-Martínez, *et al.* [13]	Machine learning; Groundwater quality; Supervised learning; Predictive modelling; Water pollution indicators
Li, *et al.* [14]	Groundwater pollution; Machine learning; Groundwater quality; Pollution source identification; Pollution risk; Hydrochemical evolution
Luo, *et al.* [15]	Machine learning; Surrogate model; Groundwater contaminant transport modelling; Artificial neural network
Nordin, *et al.* [16]	Groundwater monitoring; Groundwater parameters; Statistical modelling; Artificial intelligence; Artificial neural network
Bhowmik, *et al.* [17]	Machine Learning; Groundwater contamination; Arsenic; Fluoride; Nitrate; Climate Change
Pandya, *et al.* [18]	-
Haghbin, *et al.* [19]	-
Huang, *et al.* [20]	Groundwater quality assessment; Groundwater quality prediction; Machine learning; Groundwater modeling
Hanoon, *et al.* [21]	Groundwater quality (GWQ), Artificial intelligence (AI), Machine learning (ML), ANN, ANFIS
Chowdhury, *et al.* [22]	Contamination; Groundwater quality index; GIS; MODFLOW; Machine learning prediction
Kumari, *et al.* [23]	ANN; Groundwater; FFNN; LM algorithm; Groundwater Modelling
Haggerty, *et al.* [24]	Machine learning; Deep learning; Groundwater quality; AI

Table S2. The keywords (besides prediction, geospatial, mapping, GIS) used during the search for MLGHM studies

Source	Keywords
Google Scholar	**groundwater** (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **artificial intelligence** (/ data mining)
ScienceDirect	**groundwater** (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **artificial intelligence** (/ data mining) + **nitrate** (/ other contaminants)
SpringerLink	**groundwater** (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **artificial intelligence** (/ data mining) + **Iran** (United States, India, Bangladesh / other critical areas)
Science	**groundwater** (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **artificial intelligence** (/ data mining) + **nitrate** (/ other contaminants) + **Iran** (United States, India, Bangladesh / other critical areas)
Wiley Online Library	**groundwater** (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **machine learning** (/ supervised / unsupervised / model's specific name)
	groundwater (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **machine learning** (/ supervised / unsupervised / model's specific name) + **nitrate** (/ other contaminants)
	groundwater (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **machine learning** (/ supervised / unsupervised / model's specific name) + **Iran** (United States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) + **pollution** (/ contamination / vulnerability / hazard / risk) + **machine learning** (/ supervised / unsupervised / model's specific name) + **nitrate** (/ other contaminants) + **Iran** (United States, India, Bangladesh / other critical areas)

		Search 1	Search 2	Search 3	Search 4	Search 5	Search 6	Search 7	Search 8
Web of Science	**Row 1**	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)	**groundwater** (/ aquifer / well)
	Row 2	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)	**pollution** (/ contamination / vulnerability / hazard / risk)
	Row 3	**artificial intelligence** (/ data mining)	**artificial intelligence** (/ data mining)	**artificial intelligence** (/ data mining)	**artificial intelligence** (/ data mining)	**machine learning** (/ supervised / unsupervised /	**machine learning** (/ supervised / unsupervised /	**machine learning** (/ supervised / unsupervised /	**machine learning** (/ supervised / unsupervised /

						model's specific name)	model's specific name)	model's specific name)	model's specific name)
	Row 4		**nitrate** (/ other contaminants)	**Iran** (United States, India, Bangladesh / other critical areas)	**nitrate** (/ other contaminants)		**nitrate** (/ other contaminants)	**Iran** (United States, India, Bangladesh / other critical areas)	**nitrate** (/ other contaminants)
	Row 5				**Iran** (United States, India, Bangladesh / other critical areas)				**Iran** (United States, India, Bangladesh / other critical areas)

Scopus	**groundwater** (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **artificial AND intelligence** (/ data AND mining)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **artificial AND intelligence** (/ data AND mining**)** AND **nitrate** (/ other contaminants)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **artificial AND intelligence** (/ data AND mining**)** AND **Iran** (United AND States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **artificial AND intelligence** (/ data AND mining) AND **nitrate** (/ other contaminants) AND **Iran** (United AND States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **machine AND learning** (/ supervised / unsupervised / model's specific name)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **machine AND learning** (/ supervised / unsupervised / model's specific name) AND **nitrate** (/ other contaminants)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **machine AND learning** (/ supervised / unsupervised / model's specific name) AND **Iran** (United AND States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) AND **pollution** (/ contamination / vulnerability / hazard / risk) AND **machine AND learning** (/ supervised / unsupervised / model's specific name) AND **nitrate** (/ other contaminants) AND **Iran** (United AND States, India, Bangladesh / other critical areas)

Copernicus Publications	**groundwater** (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **artificial intelligence** (/ data mining)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **artificial intelligence** (/ data mining) **nitrate** (/ other contaminants)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **artificial intelligence** (/ data mining) **Iran** (United States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **artificial intelligence** (/ data mining) **nitrate** (/ other contaminants) **Iran** (United States, India, Bangladesh / other critical areas)
Nature Portfolio	**groundwater** (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **machine learning** (/ supervised / unsupervised / model's specific name)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **machine learning** (/ supervised / unsupervised / model's specific name) **nitrate** (/ other contaminants)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **machine learning** (/ supervised / unsupervised / model's specific name) **Iran** (United States, India, Bangladesh / other critical areas)
	groundwater (/ aquifer / well) **pollution** (/ contamination / vulnerability / hazard / risk) **machine learning** (/ supervised / unsupervised / model's specific name) **nitrate** (/ other contaminants) **Iran** (United States, India, Bangladesh / other critical areas)

Text S2

Further examination performed regarding the study area of some articles is as follows:

(i) In two efforts, viz. Ref.[25] that focused on Holocene multi-aquifers of Ganges Delta and Ref.[26] that focused on Indo-Bangla Ramsar site, the study areas span across both Bangladesh and India. Hence, both countries are counted for these two studies.

(ii) There were several studies with supernational study areas (i.e., multi-country, continent-scale, and global-scale, as listed in Table S3). Whilst all these efforts are included in the main dataset of this study, only multi-country studies (Refs.[27-30]) were counted in terms of the study area. In other words, continent-scale and global-scale studies were omitted from geographical trend analysis because these studies essentially focus on highlighting crucial parts of the world/continents. Hence, restricting the geographical trend analysis to local and national study areas may yield a better understanding of their relative frequency of examination.

(iii) Turkiye, being a transcontinental country, is counted for both Europe and Asia continents.

Table S3. MLGHM studies with supranational study areas

Cite	Target	Model	Year	Area
Wood, *et al.* [31]	Ca, Mg, Na, K, HCO3, Cl, SO4, Sr, Fe, NO3 (as N), F, Br, Si, DO, pH, Specific conductance, Temperature	KNN	2020	Global
Podgorski and Berg [32]	As	RF	2020	Global
Amini, *et al.* [33]	As	ANFIS, MLR	2008	Global
Podgorski and Berg [34]	F^-	RF	2022	Global
Amini, *et al.* [35]	F^-	ANFIS, MLR	2008	Global
Usman, *et al.* [36]	F^-	NB-stacking	2024	Africa
Hájek, *et al.* [37]	pH, Ca	RF	2020	Europe
Podgorski, Kracht, Araguas-Araguas, Terzer-Wassmuth, Miller, Straub, Kipfer and Berg [27]	^3H	RF	2024	Africa's Sahel (Mauritania, Mali, Niger, Chad, Senegal, Guinea, Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Central African Republic)
Podgorski, Araya and Berg [28]	Mn, Fe	RF, BRT	2022	Southeast Asia and Bangladesh
Araya, Podgorski and Berg [29]	salinity	RF	2023	Horn of Africa (Djibouti, Eritrea, Ethiopia, Kenya, Somalia)
Winkel, Berg, Amini, Hug and Annette Johnson [30]	As	LgR	2008	Southeast Asia (Bangladesh, Cambodia, Vietnam, Indonesia, Myanmar, Thailand)

S1. Trend Analysis

S1.1 Historical Trend

The number of MLGHM papers published per annum over the period 2005-2024 is shown in Figure S1. There is a consistent rise in the rate of publication of such papers, with sharp incremental increases in 2017, 2022 and 2024. These trends may be due to (i) growing environmental concerns all over the world, (ii) advances in computational intelligence and ML developments, (iii) facilitated interdisciplinary collaboration among experts, and (iv) greater availability of secondary geochemical and geographical data (e.g., through remote sensing).

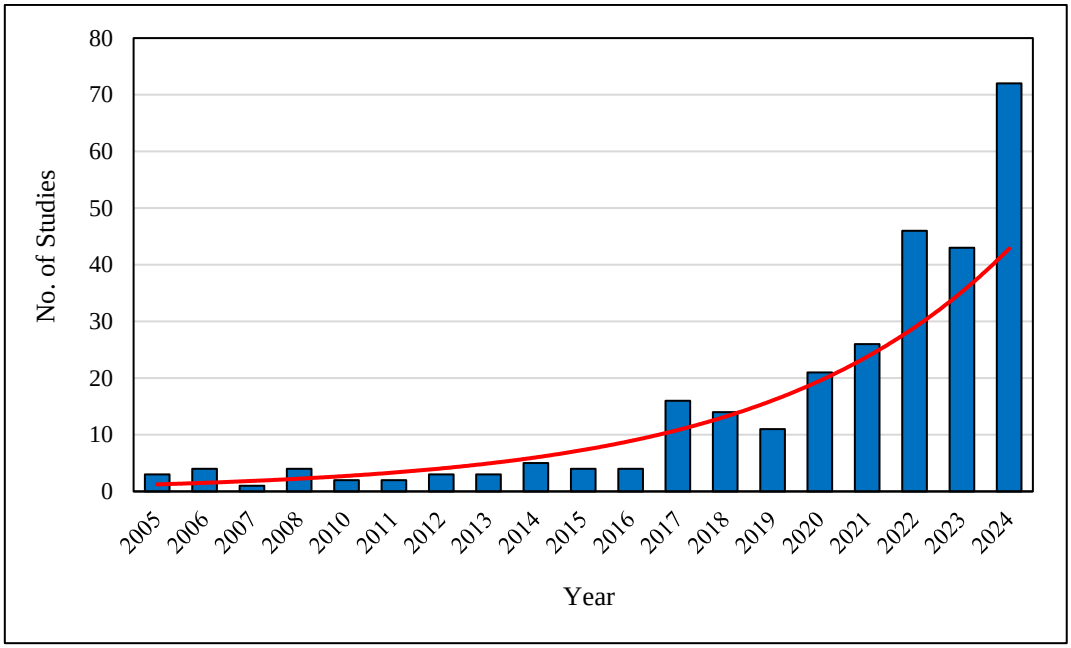


Figure S1. Historical (2005-2024) trends of number of MLGHM publications per annum. The red curve is an indicative exponential trend.

S1.2 Geographic Trend

The distribution by country-of-study of reviewed MLGHM papers is shown in Figure S2. Asia (195 published studies) was the most studied part of the world, having Iran (71 studies), India (46 studies), China (25 studies), and Bangladesh (13 studies) as the most frequently studied countries. Africa (54 published studies) gained the second rank with Egypt (8 studies) and Tunisia (6 studies) leading. North America (43 studies) appeared next, comprising United States (42 studies) and Mexico (1 study). In Europe (16 studies), Spain (6 studies) was the most studied country. Lastly, Uruguay (1 study) was the only representative of South America.

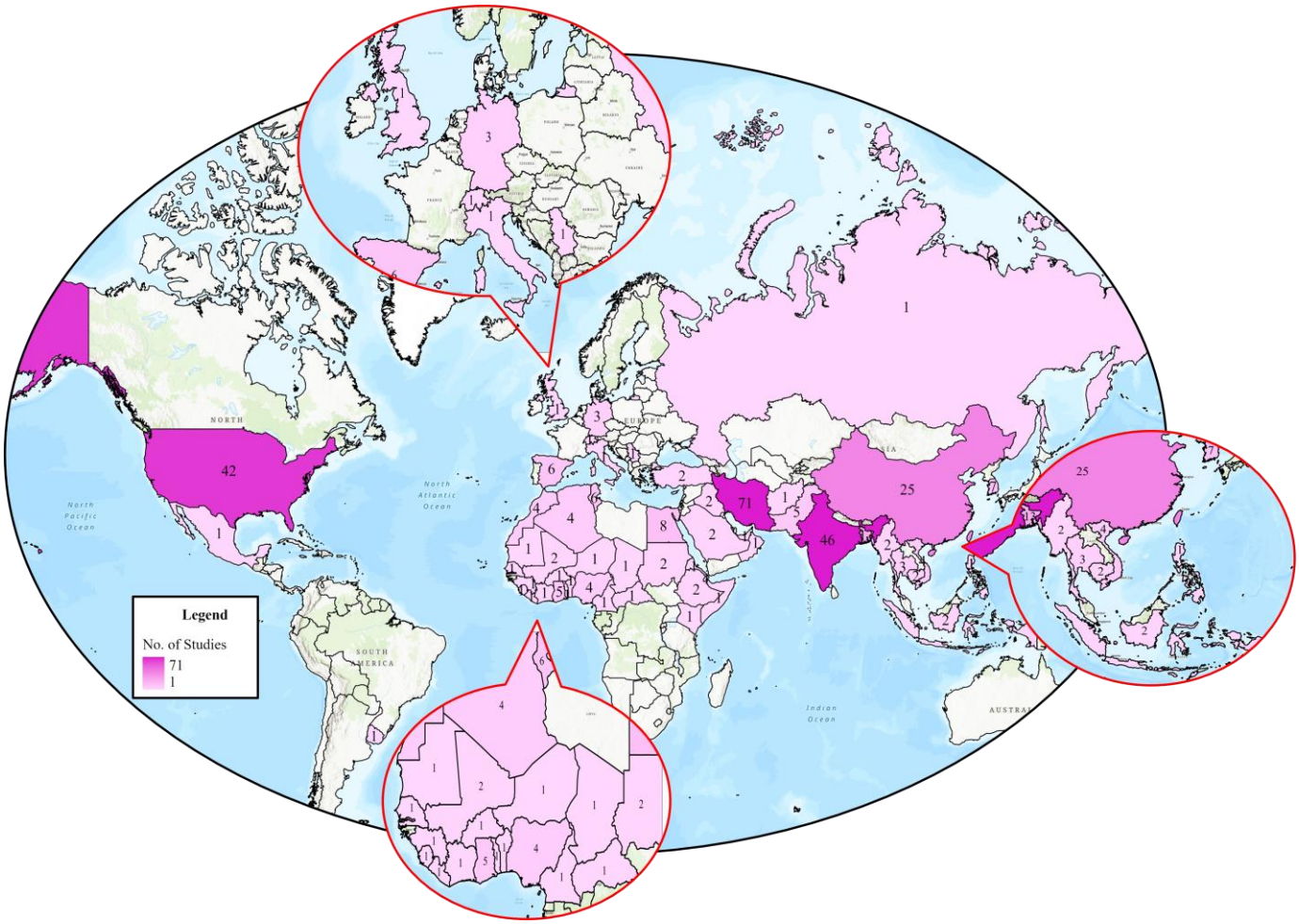


Figure S2. Geographical distribution of the study-focus of reviewed published MLGHM studies. Boundaries and names were adopted from generalized shapefiles of the world obtained from the United Nations [1] and Environmental Systems Research Institute (ESRI) [2]).

Explanations regarding the observed geographical distributions are as follows:

(a) **Asia:** While the quality and quantity of groundwater resources are being negatively affected by excessive exploitation in countries such as India and Bangladesh, the general quality of groundwaters in Iran is good [38]. Hence, the biggest number for the Iranian case studies can be attributed to the significant growth in the publication of science and engineering articles during the past two decades in this country [24,39] rather than being an indication of greater issues with groundwater quality. Moreover, it should be noted that several recent publications with Iranian case studies focused on ML developments by introducing novel ensembles [40,41] and conducting comparisons among the used models [42,43].

(b) **Africa**: Only the northern half of the continent has been explored, reflecting the need for conducting MLGHM studies in the southern half, particularly in critical countries such as Uganda [44].

(c) **Europe:** The tangible difference between Spain and other countries can be explained by the climate of Spain (i.e., facing considerable aridity and drought) that calls for using groundwater for urban water supply and irrigation purposes [45].

(d) **North America:** In the US, groundwater is a significant source of public and domestic self-supply water [46]. Nonetheless, these resources are largely affected by elevated concentrations of geogenic constituents (As, Mg, Sr, Ra, and adjusted gross alpha) [47]. Besides this significance, the availability of adequate water data (e.g.,

National Water-Quality Assessment (NAWQA) Program [48]) can be another reason for the large number of studies in this country.

(e) South America: The quality of groundwater in many Latin American countries is considerably affected by contaminants such as arsenic [49]. It necessitates conducting more MLGHM studies in critical areas such as Uruguay.

S1.3. Target Parameters

The proportion of the target parameters (i.e., predicted variables) in reviewed MLGHM literature is shown in Figure S3. There was a total of 414 predictions (of 55 unique target variables), meaning that many studies had more than one target parameter. NO_3^- with 34% (= 122 studies) is the most frequently considered target, followed by WQI with 16% (= 56 studies) and As with 12% (= 41 studies). The share of other contaminants/parameters from F^- to pH falls within 1-6% (= 5-21 studies). The remaining parameters are collectively labelled 'Others' with 10% (= 35 studies). Each parameter of this label appeared in fewer than 5 studies, representing <1% of the whole. Detailed proportions are illustrated by the pie chart on the right part of Figure S3.

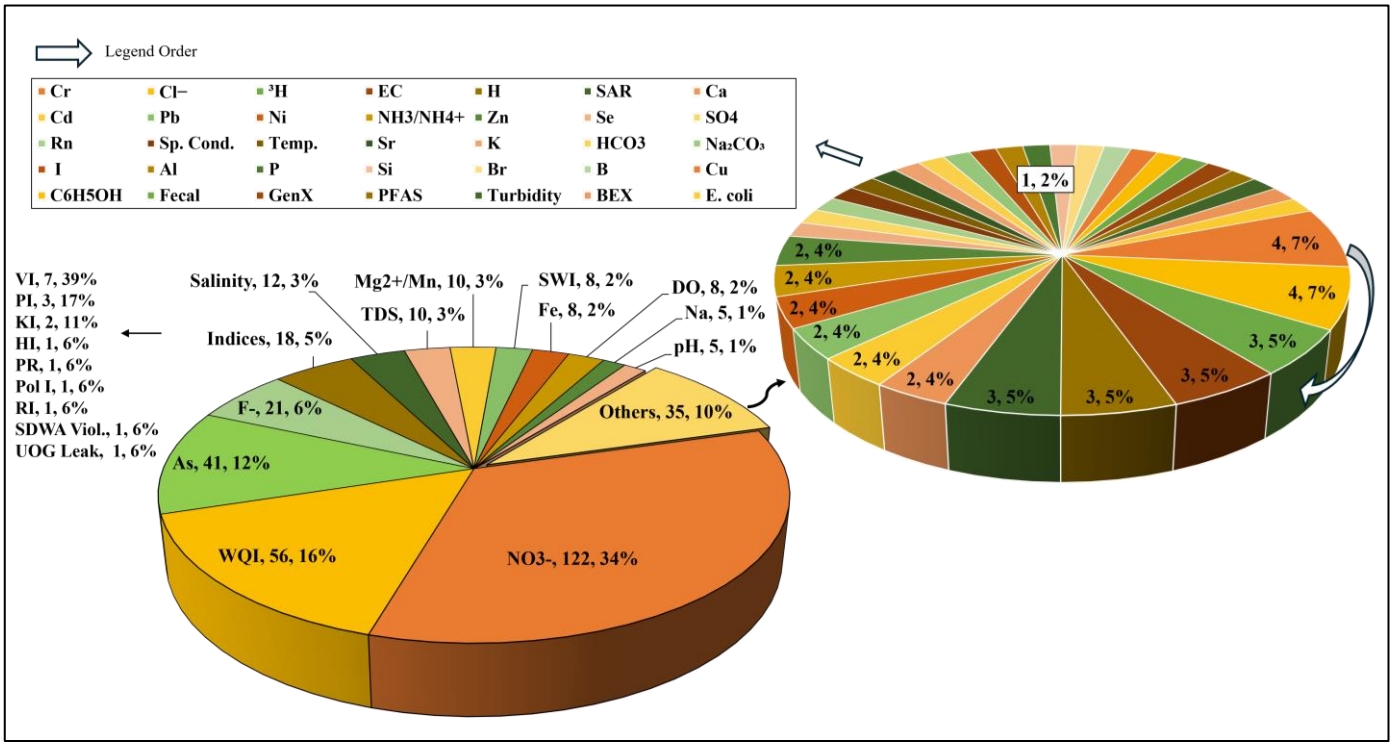


Figure S3. The proportion of the target parameters in the surveyed literature. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)

It is worth noting that while NO_3^- [50], As [51], and F^- [52] are globally known as serious groundwater contaminants, other studied parameters such as H, TDS, and EC are typically (but not always) not considered as contaminants, but are rather treated as key groundwater factors that can affect the level of contamination. Albeit, a combination of them was frequently used for calculating WQI [53], which in this study represents integrated WQI (IWQI) and entropy-weighted WQI (EWQI), too.

S1.4. Predictive Variables

The collected predictive variables were first classified into 9 major categories including (a) Predefined Datasets, (b) Chemical, (c) Topography, (d) Geomorphology, (e) Climate & Weather, (f) Aquifer, (g) LCLU, (h) Anthropogenic, and (i) Time & Coordinate, as detailed in Table S4. While it was aimed for each category to address distinct environmental/geochemical impacts on the groundwater hazard, several of them can be interrelated and share features (e.g., Topography and Geomorphology). Predefined Datasets include eight well-known GHM frameworks, namely DRASTIC, DRASTIC-LCLU, GALDIT, GALDIT-NUTS, GALDIT-MW, GOD, AVI, and SPECTR. The details of these frameworks are presented in Table S5.

Table S4. Details of the classification of the inputs used in the MLGHM literature

Predefined Datasets	**Chemicals (Water)**		**Topography**	**Geomorphology**	**Climate / Weather**	**Aquifer**	**Anthropogenic**	**LCLU**	**Time & Location**
DRASTIC	Mg	Fe	Elevation	Geology / Lithology	Precipitation	Aquifer type	Direct Anthropogenic	LU / LC / LCLU	Location
DRASTIC-LCLU	Cl	PO_4^{3-}/PO_4^{2-}	Dis. to Drainage (+ Density)	Dis. to Fault (+ Density)	Evapotranspiration	Aquifer Thickness	Dis. to Contaminant Centres	NDVI	Time
GALDIT	SO_4^{2-}	As	Terrain	Dis. to Waterbodies (Rivers, Lake, Sea + Density)	Elements in Atm. & Precipitation	Aquifer Media	Agriculture & Irrigation		
GALDIT-NUTS	pH	Salinity	Slope (°, Aspect, Length)	Dis. to Lineament (+ Density)	Evaporation	Hydraulic Conductivity	Animal Effect (Manure)		
GALDIT-MW	Na	CO_3^{2-}	Others (Water-Related)	Soil Chemistry	Aridity	Recharge			
GOD	EC	SAR		Soil Type	Climate / Weather	Groundwater Depth			
AVI	HCO_3^-	Zn		Soil Texture	PTA Coefficient	Transmissivity			
SPECTR	K	DO		Others	Water Wilting Point	Others – GW and Wells			
	Ca	Alkalinity			Others	Vadose Zone			
	NO_3^-	Turbidity				Others - Aquifer			
	TDS	Cr							
	Hardness	Water Temp.							
	F	Others							
Details:									

	Others:	**Terrain:**	**Soil Chemistry:**	**Others:**	**Others – GW and Wells:**	**Direct Anthropogenic:**		
	Permeability	Watershed Features	Organic	Land Surface	Well Depth	Anthropogenic		
	Hg	TRI	Soil Geochemistry (Elements)	Temp.	Well Screen Features	Population/Households		
	Ba	Convergence	Soil pH	Solar Radiation	Water Systems	Anthropogenic Nitrogen		
	Al	Profile Curvature	Soil CEC	Water Wilting	Operation Time of Wells	GDP		
	Cu	Plan Curvature	Others	Point	Borehole Yield	Dis. to Roads		
	Cd	Landform		Water Vapor	Groundwater Exploitation			
	Pb	Roughness	**Soil Texture:**	Pressure	Density of Pumping Well	**Agriculture & Irrigation:**		
	Sr	TPI	Texture	Evapotrans./Prec.	Extraction Capacity			
	Co	TWI	Content (Sand)		Extraction Density	Agronomic Practices		
	B	Others	Content (Silt)		Water Yield Capacity	Irrigation Effect		
	Ni		Content (Clay)		Available Water Capacity	Farm Fertilizer		
	Br	**Slope (°, Aspect, Length):**	Coarse Fragments		Magnitude per Unit Area of Streams			
	NH_3		Water Capacity		Location Condition of Underground Storage Tanks			
	NH_4	Slope (°)	Bulk Density					
	Si	Slope Aspect	Others		Groundwater Vulnerability			
	$CaCO_3$	Slope Length						
	$N0_2^-$	Topography	**Others:**		**Vadose Zone:**			
	^3H		Geomorphology		Vadose Zone (Media - Thickness)			
	ORP	**Others (Water-Related):**			Superficial Thickness			
	Residual Na_2CO_3	Flow Direction			Unsaturated Zone Features			
	Soluble Na (%)	Runoff (Overland Flow)			Depth to Saturated Soil			
	KI	Flow Accumulation			Dis. to Hydraulic Window			
	General Bacteria	Baseflow Index			Redox Impact			
	Fecal Coliform	Others			Residual Water Content			
	$KMnO_4$ Consumption				volumetric Water Content			
	BOD				vadose Residence Time			
	Residual $NaHCO_3$				vadose Water Content			

	Cl/Br				Inverse Air Entry Suction Subsurface Flow Contact Time Subsurface Flow Infiltration Rate **Others - Aquifer:** Aquifer Connectivity GW Leakage Other Hydraulic Features Interannual Variation Synthetic Harmful Coefficient Age of Groundwater			

Table S5. Components of the included predefined datasets

DRASTIC	**GALDIT**	**SPECTR**	**GALDIT-NUTS**
Depth to water	Aquifer type	Topography	GALDIT
Net recharge	Hydraulic conductivity	pH	Net recharge
Aquifer media	Water level (asl)	EC	Topography
Soil media	Dis. to the shore	Hydraulic conductivity	Depth to groundwater
Topography	SWI status	Saturated zone thickness	LU
IoVZ	Aquifer thickness	Recharge	
Hydraulic conductivity			

AVI	**GOD**	**DRASTIC-LU**	**GALDIT-MW**
Hydraulic resistance	Aquifer type	DRASTIC	GALDIT
	Lithology of aquifer	LU	Aquifer Media
	Water table depth		Well Density

In this section, first, the frequency of use of the nine major categories is presented. Next, knowing that most factors of the Predefined Datasets overlap with other major categories, the Predefined Datasets are combined with other major categories to achieve a consistent format of predictive variables. Eventually, a detailed frequency assessment of each single predictive variable is performed.

A total of 3921 predictive variables were used in all reviewed studies. As Figure S4 shows, Chemical category accounts for 32% of predictive variables, making it the largest major category. Next, Geomorphology with 27% is the second largest category, followed by Aquifer, Topography, Climate/Weather, Anthropogenic, LCLU, and Predefined Datasets with respective shares of 13%, 11%, 6%, 5%, 4%, and 2%. Lastly, Time & Location with < 1% is the least popular category. Moreover, the pie chart positioned on the right part of Figure S4 indicates that among Predefined Datasets, DRASTIC with 55 (= 50 DRASTIC + 5 DRASTIC-LCLU) and GALDIT with 11 (= 9 GALDIT + 1 GALDIT-NUTS + 1 GALDIT-MW) times of use are the most popular frameworks, whereas GOD, AVI, and SPECTR were used only once.

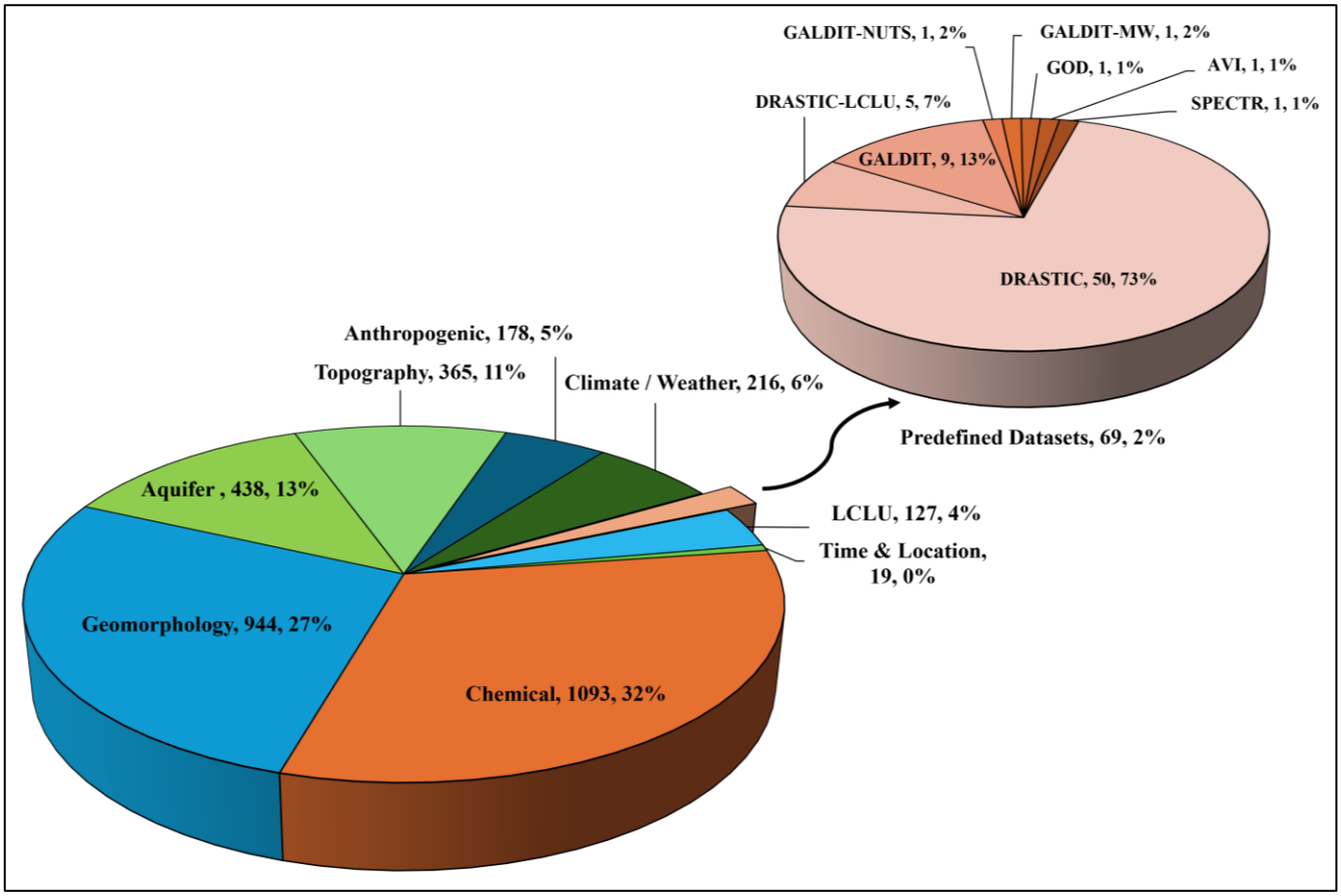


Figure S4. Frequency of use of major categories of predictive variables. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)

8 major categories are located Figure S5 depicts details of the used predictive variables in the form of Sankey Diagram showcasing the inter-category connectivity and frequency of use. Guidelines for interpreting this diagram are

(i) on the left, 8 Predefined Datasets are positioned in the middle-left, and specific predictive variables are shown on the right,

(ii) The thickness of the curves is scaled in proportion to the frequency of the involved predictive variable,

(iii) The number shown beside each major category indicates the summed frequency of the comprised predictive variables. For instance, it can be written for DRASTIC: 350 = 7 predictive variables × 50 time of use,

(iv) The number shown beside each specific predictive variable indicates its summed frequency, which was obtained as the summation of all engaged categories. For instance, it can be written for Slope: 143 = 50 (from DRASTIC) + 5 (from DRASTIC-LU) + 86 (from Topography) + 1 (from SPECTR) + 1 (from GALDIT-NUTS).

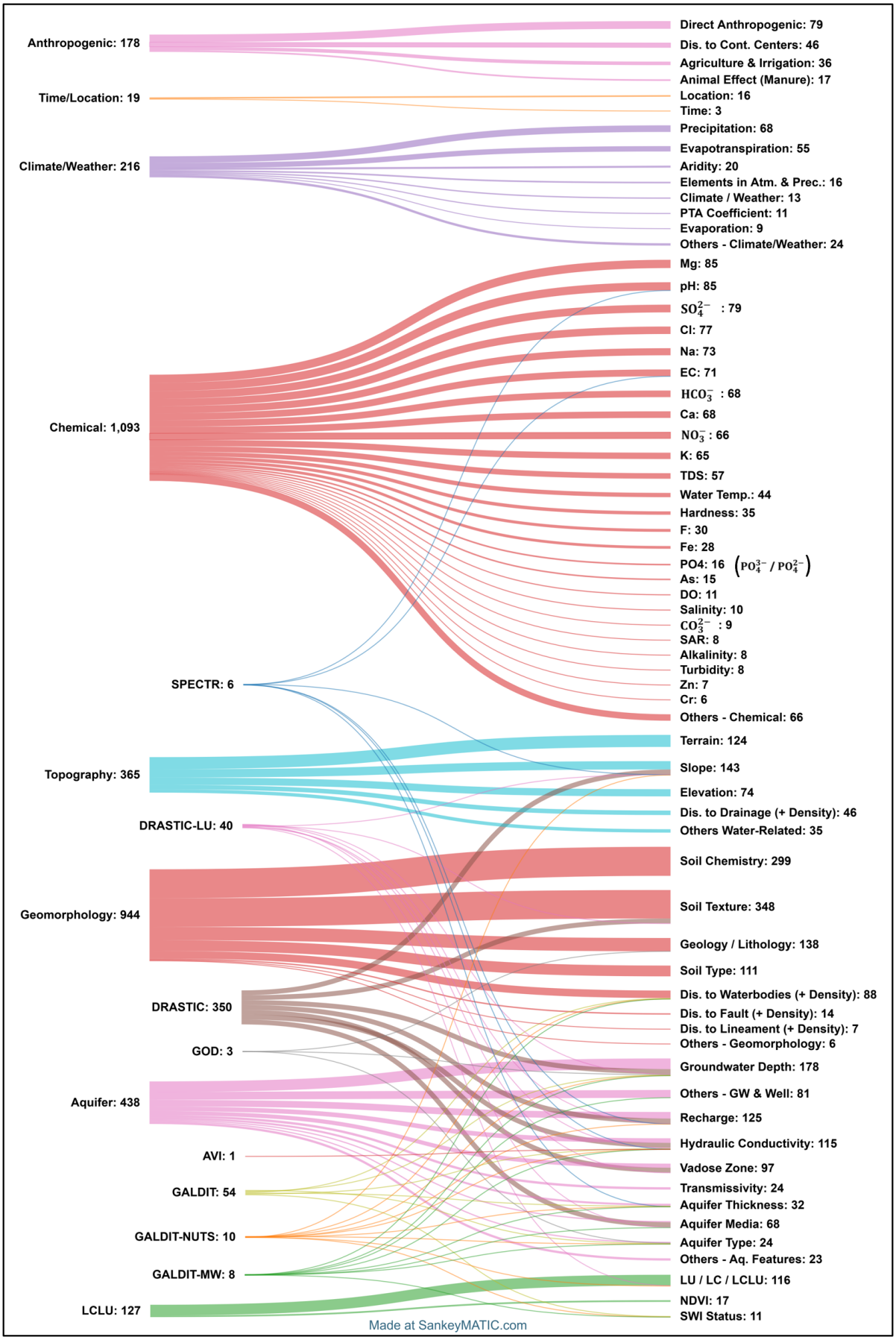


Figure S5. Sankey Diagram showing predictive variables and their frequency of use. See Tables S4 and S5 for details. (Plotted using Sankeymatic [3])

Explanations regarding predictive variables and Figure S5 are as follows:

(a) Anthropogenic: Anthropogenic activities are among the key factors that affect the environment [54], particularly GWQ through different ways, e.g., agricultural and industrial activities [55]. In this category, Direct Anthropogenic – mostly addressing the effect of Population/Households – was the most used factor followed by Distance to Contamination Centres that were expressed as the distance to various pollution sources including wastewater sources, industrial centres, cemeteries, and mines.

(b) Time/Location: Time and location (i.e., spatial coordinates) were considered in a few studies. A possible reason for not including time is the low temporal contamination variability of groundwater resources. As for the location, the essence of geospatial analysis accounts for the distribution of contaminants and influential factors, hence, no urgent need for spatial coordinates to be specifically taken into calculations.

(c) Climate: Climate change, apart from impacting recharge, is related to GWQ through affecting groundwater temperature, which alters chemical reactions within the aquifer [56]. Precipitation, evapotranspiration, and aridity status were the most common distinguished predictive variables from the Climate/Weather category.

(d) Chemical: The importance of water's physicochemical characteristics in GWQ is evident [55,57]. The most frequently used Chemical predictive variables were (major ions of) Mg, Cl, SO_4^{2-}, Na, HCO_3^-, K, Ca, NO_3^-, F, and Fe, as well as the groundwater characteristics of pH, EC, TDS, temperature, and H. It is pertinent to note that, pH and EC were the only chemical attributes that were explicitly linked to a Predefined Dataset (SPECTR).

(e) Topography: Topographical features have received significant attention as they can influence surface water-aquifer interaction and contaminant transport [58]. Slope (represented by Topography in GALDIT-NUTS, DRASTIC, and SPECTR) was the most favoured factor of this category, followed by Terrain Characteristics (notably including TWI), and Elevation.

(f) Geomorphology: Geomorphology accounts for the influence of various factors, each playing a pivotal role in determining GWQ. For instance, geological settings dictate the aquifer capacity and dilution of minerals [59,60]; soil is a natural filter and its characteristics (type, texture, and chemistry) influence the cleanliness of the water that reaches the aquifer [61]; lineaments and faults in the region can contribute to the water flow and the reservoir's capacity [62,63]; and nearby waterbodies such as rivers, regulate water's recharge and chemical composition [64]. Geomorphological features are highly regarded in Predefined Datasets such as GOD, DRASTIC, and GALDIT. In this category, soil characteristics (Soil Texture and Soil Chemistry) appeared most frequently, tailed by Geology/Lithology, Soil Type and Distance to Waterbodies (+ their Density).

(g) Aquifer: This category addresses aquifer properties along with groundwater's (and well's) hydraulic characteristics that contribute to delineating different elements of GWQ [65]. Groundwater Depth, Recharge, and Hydraulic Conductivity were the most frequent aquifer features that have been input to ML models. The reason for this popularity can be sought in different features such as water exploitation and soil saturation status that can be explained by these three factors, particularly Groundwater Depth [66,67]. Figure S5 shows that most factors from the Aquifer category are widely linked to the Predefined Datasets.

(h) LCLU: LCLU affects groundwater resources, notably through changing recharge and water demand [68]. Many studies have addressed the effect of this factor for critical countries such as Bangladesh [69] and India [70]. This category was considered through inputting NDVI and LC/LU/LCLU factors either directly or in extensions of the DRASTIC and GALDIT frameworks.

The most popular specific predictive variables (those with summed frequencies ≥ 70) are expressed in Table S6. According to this table, Chemical and Aquifer categories have the largest number of representatives.

Table S6. Specific predictive variables with highest frequencies of use

Freq.	Chemical	Geomorphology	Aquifer	Topography	Anthropogenic	Climate/ Weather	LCLU	Time & Location
70-100	Mg pH SO_4^{2-} Cl Na EC	Dis. to Waterbodies (+ Density)	Others - GW & Well Vadose Zone	Elevation	Direct Anthropogenic	-	-	-
100-150	-	Geology/Lithology Soil Type	Recharge Hydraulic Conductivity	Terrain Slope	-	-	LU/LC/LCLU	-
150-200	-	-	Groundwater Depth	-	-	-	-	-
> 200	-	Soil Chemistry Soil Texture	-	-	-	-	-	-

S1.5. Feature Selection

When it comes to ML-based modelling, feature selection can play an important role in optimizing the models. Through dimensionality reduction, feature selection enables the model to retain only those predictive variables that have a significant contribution to the task. Proportions of feature selection models used in the reviewed MLGHM literature are shown in Figure S6, acknowledging a total of 348 feature selection applications (with 41 unique methods). 31% (= 100 studies) of articles employed tree-based models - mostly RF. Correlation was the second popular technique with 13% (= 43 studies), followed by PCA, Regression, PDP, and ANN with 11%, 9%, 8%, and 5% of the articles, respectively. Boosting Ensemble and SHAP secured the next position with 4% and 3%, and other methods were spotted in < 3% of the surveyed articles.

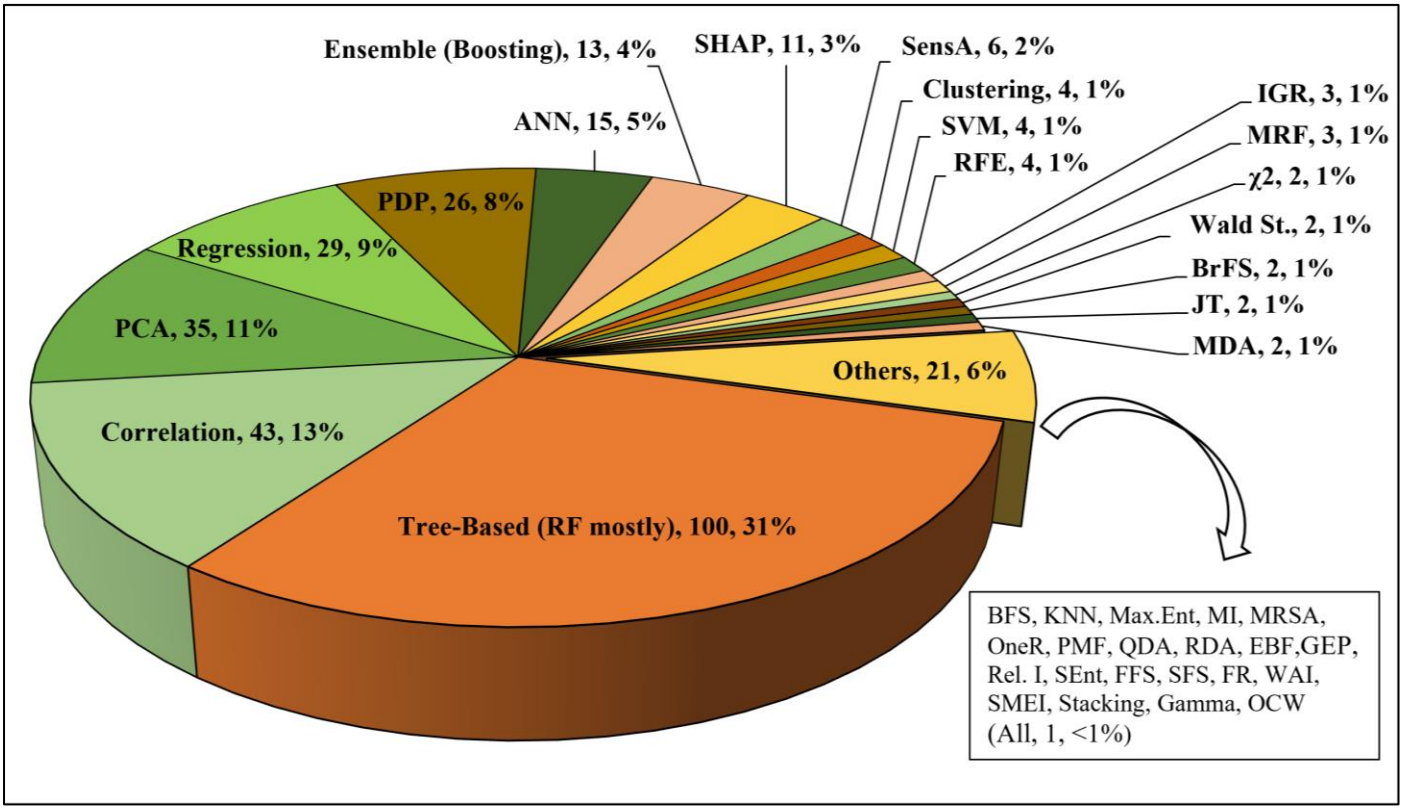


Figure S6. The frequency of use of feature selection techniques used in the surveyed literature. Acronyms can be found in the Nomenclature. (Format of the included texts: Name, Frequency, Percentage of frequency)

Table S7. Classification of the ML models used in the MLGHM literature

Supervised									**Unsupervised**		
Tree-Based	**ANN**	**Ensemble 1**	**SV-Based**	**Regression**	**Ensemble 2**	**Probabilistic**	**Others**	**NF**	**Clustering**	**SOM**	**DL**
RF	ANN	Bst	SVM	LgR	MOA	Gaussian	KNN	NF/ANFIS	FCM	SOM	DBN
DT	DL	Bag	SVR	MLR	MM	Discriminant	GEP	Others	KMC (& K-Me-doid)	GNG	DSA
CART	Others	Dag	RVM	GLMs	DH	Bayesian	PIML		HCA	SOM-FCM	
ERT		RS		MARS		Markovian			APC	SOM-KMC	
M5				Others		Others					
Others											
Details:											
Others:	Others:			Others:		Others:		Others:			
BART				glmnet							
RotF	ELM			BRR							
J48	GMDH			RR							
LMT	ELNN			PolyR				HyFIS			
RT	MNN			AFR				CANFIS			
RndT	BRNN			LsR		MaxEnt		DENFIS			

Text S3

Take supervised DL – a deep MLP neural network – as an evident example of opaque-box ML. This model belongs to the ANN's family, that are innately complicated models due to their configuration and the used non-linear activation functions. However, as shown in Figure S7, the following ideas can be considered to increase the transparency of the model (and the end-user favoured trustworthiness):

- Providing a clear illustration of the model's structure,
- Data visualization and utilizing a feature selection technique that elucidates the role of the input factors used,
- Addressing the hyperparameters that can influence the model's performance,
- Detailing the main internal processes through a shallow simplified (surrogate) version to delineate the calculations/decision-making rules,
- If possible, transforming the solution into explicit mathematical formats.

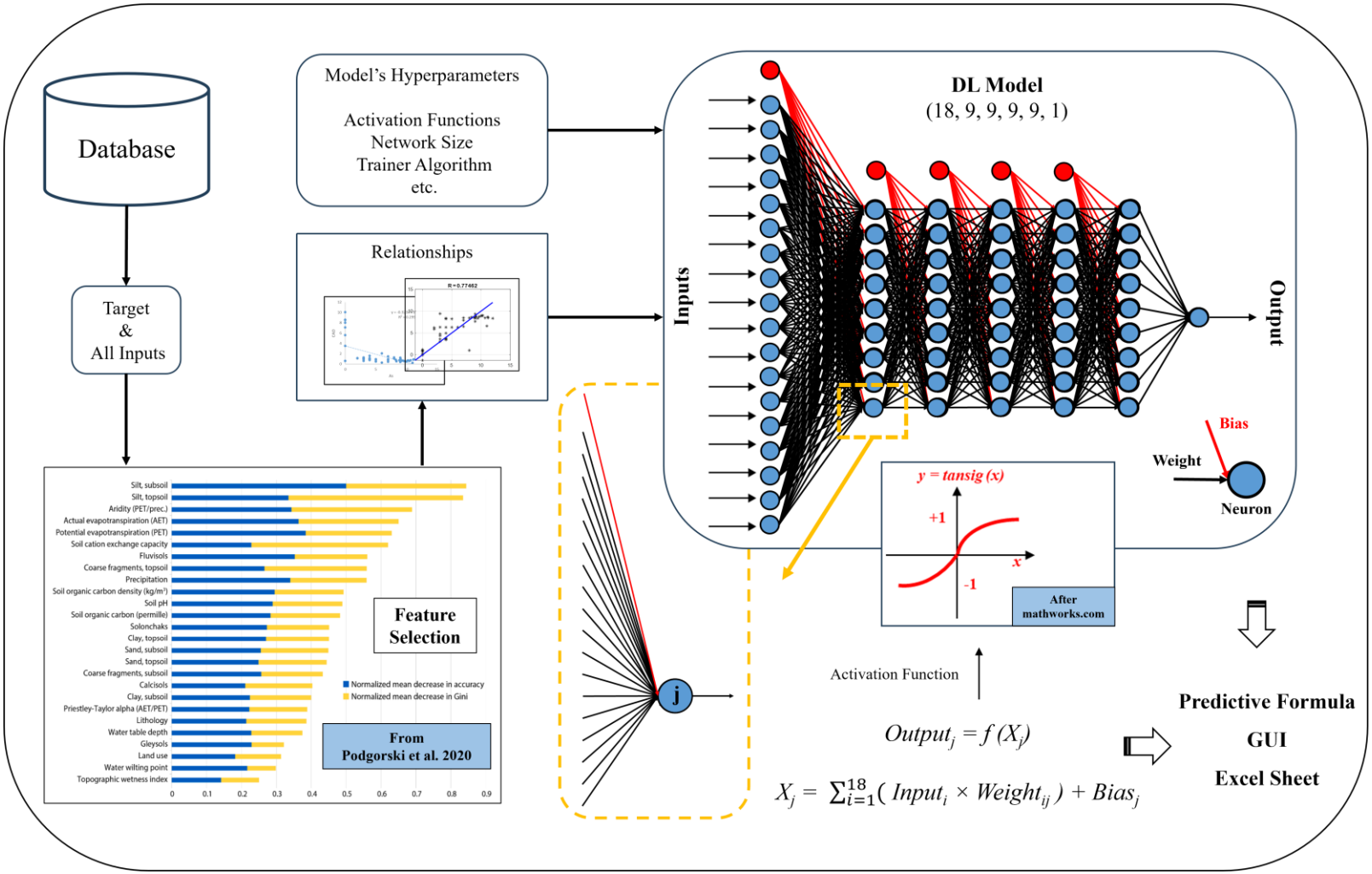


Figure S7. A schematic example of increasing the transparency of a DL opaque-box model (Feature Selection image reproduced from Podgorski, Wu, Chakravorty and Polya [4], licensed under CC BY 4.0, and the graph of Tansig activation function adopted from Mathworks.com [5])).

Text S4

The metaheuristic family has a reputation in optimizing ML models. Taking ANN and ANFIS as examples, when an MOA is coupled with them, it finds the optimized values for weights and biases (in ANN) and membership functions parameters (in ANFIS). This optimization can be particularly helpful when the ML is dealing with a complex and high-dimensional dataset. GS is another famous optimizer – used in many studies e.g., Refs.[71,72] – that, due to its deterministic basis, is different from the MOAs.

While Table S8 gives applications of MOAs in MLGHM studies, the literature contains many other applications of MOAs for optimizing groundwater quality assessment through using (i) non-ML frameworks such as DRASTIC and (ii) numerical predictions of quality indices (without mapping). Some examples of these studies are presented in Table S9. Among these algorithms, GA is a very popular and long-standing technique whose earlier applications can be found in studies by Rao and Jamieson [73] and Inoue, *et al.* [74] for groundwater-related optimizations.

Table S8. Application of optimization algorithms in the MLGHM literature

Ref.	Baseline model(s)	Optimizer algorithm(s)
Fijani, *et al.* [75]	CMAI	GA
Asadi, *et al.* [76]	MLR	GA
Mohammadrezapour, *et al.* [77]	FCM, KMC	GA
Elzain, *et al.* [78]	ANFIS	PSO, DE, GA
Agrawal, *et al.* [79]	NBC, SVM	PSO
De Jesus, *et al.* [80]	ANN	PSO
Pham, *et al.* [81]	RF	Ant, FA, MOE, PSO, GWO
Saha, *et al.* [82]	RF	PSO, GWO, GOA
Jamei, *et al.* [83]	ANFIS	AO, ACO, SMA
Wang, *et al.* [84]	KELM, SVM	PSO
Gholami and Sahour [85]	SOM	GA
Sahour, Khanbeyki, Gholami, Sahour, Karimi and Mohammadi [41]	ANN	PSO, GWO
Gholami, *et al.* [86]	ANN	GA
Bakhtiarizadeh, *et al.* [87]	PolyR, MARS	GA, PSO

firefly algorithm (FA), genetic algorithm (GA), differential evolution (DE) algorithm, particle swarm optimization (PSO), ant optimization (Ant), multiple objective evolutionary (MOE), grey wolf optimization (GWO), grasshopper optimization algorithm (GOA), aquila optimization (AO), slime mould algorithm (SMA), ant colony optimization (ACO)

Table S9. Examples of using MOAs out of (but relevant to) the MLGHM literature

Ref.	Baseline model	Optimizer algorithm
Lakshminarayanan, *et al.* [88]	Basic: DRASTIC	BBO
Nadiri, *et al.* [89]	Basic: DRASTIC	GA
Nadiri, *et al.* [90]	Basic: DRASTIC	GA
Balaji, *et al.* [91]	Basic: DRASTIC	TLBO
Jahromi, *et al.* [92]	Basic: SINTACS	ABC
Lakshminarayanan, *et al.* [93]	Basic: DRASTIC	FA
Siarkos, *et al.* [94]	Basic: DRASTIC-L	GA
Bordbar, *et al.* [95]	Basic: GALDIT	GWO
Alizamir, *et al.* [96]	ML: ANN	GWO
Mazraeh, *et al.* [97]	ML: SVM	HHO

Kisi, *et al.* [98]	ML: ANN	PSO
Harris-Hawkes optimization (HHO), artificial bee colony (ABC), teaching learning-based optimization (TLBO), biogeography-based optimization (BBO)		

References

1. UN. Map of the World. Available online: https://www.un.org/geospatial/content/map-world-1 (accessed on
2. ESRI. World Countries Generalized. Available online: https://hub.arcgis.com/datasets/esri::world-countries-generalized/ (accessed on 10-02-2025).
3. Sankeymatic. Available online: sankeymatic.com (accessed on
4. Podgorski, J.; Wu, R.; Chakravorty, B.; Polya, D.A. Groundwater arsenic distribution in India by machine learning geospatial modeling. *International journal of environmental research and public health* **2020**, *17*, 7119.
5. Mathworks.com. Hyperbolic tangent sigmoid transfer function. Available online: https://uk.mathworks.com/help/deeplearning/ref/tansig.html (accessed on 1 February 2025).
6. Gogu, R.C.; Dassargues, A. Current trends and future challenges in groundwater vulnerability assessment using overlay and index methods. *Environmental geology* **2000**, *39*, 549-559.
7. Mádl-Szőnyi, J.; Füle, L. Groundwater vulnerability assessment of the SW Trans-Danubian central range, Hungary. *Environmental Geology* **1998**, *35*, 9-18.
8. Goyal, D.; Haritash, A.; Singh, S. A comprehensive review of groundwater vulnerability assessment using index-based, modelling, and coupling methods. *Journal of Environmental Management* **2021**, *296*, 113161.
9. Geng, C.; Lu, D.; Qian, J.; Xu, C.; Li, D.; Ou, J.; Ye, Z. A review on process-based groundwater vulnerability assessment methods. *Processes* **2023**, *11*, 1610.
10. Machiwal, D.; Cloutier, V.; Güler, C.; Kazakis, N. A review of GIS-integrated statistical techniques for groundwater quality evaluation and protection. *Environmental Earth Sciences* **2018**, *77*, 681.
11. Taghavi, N.; Niven, R.K.; Paull, D.J.; Kramer, M. Groundwater vulnerability assessment: A review including new statistical and hybrid methods. *Science of the Total Environment* **2022**, *822*, 153486.
12. Geetha, M.; Bonthula, S.; Al-Maadeed, S.; Al-Lohedan, H.; Rajabathar, J.R.; Arokiyaraj, S.; Sadasivuni, K.K. Research trends in smart cost-effective water quality monitoring and modeling: special focus on artificial intelligence. *Water* **2023**, *15*, 3293.
13. Torres-Martínez, J.A.; Mahlknecht, J.; Kumar, M.; Loge, F.J.; Kaown, D. Advancing groundwater quality predictions: Machine learning challenges and solutions. *Science of The Total Environment* **2024**, 174973.
14. Li, X.; Liang, G.; He, B.; Ning, Y.; Yang, Y.; Wang, L.; Wang, G. Recent advances in groundwater pollution research using machine learning from 2000 to 2023: a bibliometric analysis. *Environmental Research* **2024**, 120683.
15. Luo, J.; Ma, X.; Ji, Y.; Li, X.; Song, Z.; Lu, W. Review of machine learning-based surrogate models of groundwater contaminant modeling. *Environmental Research* **2023**, 117268.
16. Nordin, N.F.C.; Mohd, N.S.; Koting, S.; Ismail, Z.; Sherif, M.; El-Shafie, A. Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundwater for Sustainable Development* **2021**, *14*, 100643.
17. Bhowmik, T.; Sarkar, S.; Sen, S.; Mukherjee, A. Application of machine learning in delineating groundwater contamination in present and climate change scenarios. *Current Opinion in Environmental Science & Health* **2024**, 100554.
18. Pandya, H.; Jaiswal, K.; Shah, M. A comprehensive review of machine learning algorithms and its application in groundwater quality prediction. *Archives of Computational Methods in Engineering* **2024**, 1-22.
19. Haghbin, M.; Sharafati, A.; Dixon, B.; Kumar, V. Application of soft computing models for simulating nitrate contamination in groundwater: comprehensive review, assessment and future opportunities. *Archives of computational methods in engineering* **2021**, *28*, 3569-3591.

20. Huang, Y.; Wang, C.; Wang, Y.; Lyu, G.; Lin, S.; Liu, W.; Niu, H.; Hu, Q. Application of machine learning models in groundwater quality assessment and prediction: progress and challenges. *Frontiers of Environmental Science & Engineering* **2024**, *18*, 29.
21. Hanoon, M.S.; Ahmed, A.N.; Fai, C.M.; Birima, A.H.; Razzaq, A.; Sherif, M.; Sefelnasr, A.; El-Shafie, A. Application of artificial intelligence models for modeling water quality in groundwater: comprehensive review, evaluation and future trends. *Water, Air, & Soil Pollution* **2021**, *232*, 1-41.
22. Chowdhury, T.N.N.; Battamo, A.Y.; Nag, R.; Zekker, I.; Salauddin, M. Impacts of climate change on groundwater quality: a systematic literature review of analytical models and machine learning techniques. *Environmental Research Letters* **2025**.
23. Kumari, N.; Pathak, G.; Prakash, O. A review of application of artificial neural network in ground water modeling. *Recent Advances in Mathematics, Statistics and Computer Science* **2016**, 393-402.
24. Haggerty, R.; Sun, J.; Yu, H.; Li, Y. Application of machine learning in groundwater quality modeling-A comprehensive review. *Water Research* **2023**, *233*, 119745.
25. Saha, A.; Pal, S.C.; Islam, A.R.M.T.; Islam, A.; Alam, E.; Islam, M.K. Hydro-chemical based assessment of groundwater vulnerability in the Holocene multi-aquifers of Ganges delta. *Scientific Reports* **2024**, *14*, 1265.
26. Biswas, T.; Pal, S.C.; Chowdhuri, I.; Ruidas, D.; Saha, A.; Islam, A.R.M.T.; Shit, M. Effects of elevated arsenic and nitrate concentrations on groundwater resources in deltaic region of Sundarban Ramsar site, Indo-Bangladesh region. *Marine Pollution Bulletin* **2023**, *188*, 114618.
27. Podgorski, J.; Kracht, O.; Araguas-Araguas, L.; Terzer-Wassmuth, S.; Miller, J.; Straub, R.; Kipfer, R.; Berg, M. Groundwater vulnerability to pollution in Africa's Sahel region. *Nature Sustainability* **2024**, *7*, 558-567.
28. Podgorski, J.; Araya, D.; Berg, M. Geogenic manganese and iron in groundwater of Southeast Asia and Bangladesh–machine learning spatial prediction modeling and comparison with arsenic. *Science of the Total Environment* **2022**, *833*, 155131.
29. Araya, D.; Podgorski, J.; Berg, M. Groundwater salinity in the Horn of Africa: spatial prediction modeling and estimated people at risk. *Environment International* **2023**, *176*, 107925.
30. Winkel, L.; Berg, M.; Amini, M.; Hug, S.J.; Annette Johnson, C. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. *Nature Geoscience* **2008**, *1*, 536-542.
31. Wood, W.W.; Smedley, P.L.; Lindsey, B.D.; Wood, W.T.; Kirchheim, R.E.; Cherry, J.A. Global groundwater solute composition and concentrations. *Groundwater* **2022**, *60*, 714-720.
32. Podgorski, J.; Berg, M. Global threat of arsenic in groundwater. *Science* **2020**, *368*, 845-850.
33. Amini, M.; Abbaspour, K.C.; Berg, M.; Winkel, L.; Hug, S.J.; Hoehn, E.; Yang, H.; Johnson, C.A. Statistical modeling of global geogenic arsenic contamination in groundwater. *Environmental science & technology* **2008**, *42*, 3669-3675.
34. Podgorski, J.; Berg, M. Global analysis and prediction of fluoride in groundwater. *Nature Communications* **2022**, *13*, 4232.
35. Amini, M.; Mueller, K.; Abbaspour, K.C.; Rosenberg, T.; Afyuni, M.; Møller, K.N.; Sarr, M.; Johnson, C.A. Statistical modeling of global geogenic fluoride contamination in groundwaters. *Environmental science & technology* **2008**, *42*, 3662-3668.
36. Usman, U.S.; Salh, Y.H.M.; Yan, B.; Namahoro, J.P.; Zeng, Q.; Sallah, I. Fluoride contamination in African groundwater: Predictive modeling using stacking ensemble techniques. *Science of The Total Environment* **2024**, *957*, 177693.
37. Hájek, M.; Jiménez-Alfaro, B.; Hájek, O.; Brancaleoni, L.; Cantonati, M.; Carbognani, M.; Dedić, A.; Dítě, D.; Gerdol, R.; Hájková, P. A European map of groundwater pH and calcium. *Earth System Science Data Discussions* **2020**, *2020*, 1-41.
38. Mohebbi, M.R.; Saeedi, R.; Montazeri, A.; Vaghefi, K.A.; Labbafi, S.; Oktaie, S.; Abtahi, M.; Mohagheghian, A. Assessment of water quality in groundwater resources of Iran using a modified drinking water quality index (DWQI). *Ecological indicators* **2013**, *30*, 28-34.
39. Naghizadeh, M.; Naghizdeh, R. Growth of scientific publications in Iran: Reasons, impacts, and trends. *The Development of Science and Technology in Iran: Policies and Learning Frameworks* **2017**, 75-86.

40. Barzegar, R.; Moghaddam, A.A.; Baghban, H. A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. *Stochastic environmental research and risk assessment* **2016**, *30*, 883-899.
41. Sahour, S.; Khanbeyki, M.; Gholami, V.; Sahour, H.; Karimi, H.; Mohammadi, M. Particle swarm and grey wolf optimization: enhancing groundwater quality models through artificial neural networks. *Stochastic Environmental Research and Risk Assessment* **2024**, *38*, 993-1007.
42. Sahour, S.; Khanbeyki, M.; Gholami, V.; Sahour, H.; Kahvazade, I.; Karimi, H. Evaluation of machine learning algorithms for groundwater quality modeling. *Environmental Science and Pollution Research* **2023**, *30*, 46004-46021.
43. Barzegar, R.; Razzagh, S.; Quilty, J.; Adamowski, J.; Pour, H.K.; Booij, M.J. Improving GALDIT-based groundwater vulnerability predictive mapping using coupled resampling algorithms and machine learning models. *Journal of Hydrology* **2021**, *598*, 126370.
44. Wilson, G.J.; Muloogi, D.; Hamisi, R.; Denwood, T.; Bhattacharya, P.; Nuwategeka, E.; Gooddy, D.C.; Polya, D.A.; Huck, J.J.; Richards, L.A. Surface-derived groundwater contamination in Gulu District, Uganda: Chemical and microbial tracers. *Science of the Total Environment* **2024**, *955*, 177118.
45. Llamas, M.R.; Custodio, E.; De la Hera, A.; Fornés, J. Groundwater in Spain: increasing role, evolution, present and future. *Environmental Earth Sciences* **2015**, *73*, 2567-2578.
46. Lindsey, B.D.; Fleming, B.J.; Goodling, P.J.; Dondero, A.M. Thirty years of regional groundwater-quality trend studies in the United States: Major findings and lessons learned. *Journal of Hydrology* **2023**, *627*, 130427.
47. Erickson, M.L.; Miller, O.L.; Cashman, M.J.; Degnan, J.R.; Reddy, J.E.; Martinez, A.J.; Azadpour, E. Status of water-quality conditions in the United States, 2010–20. *USGS Professional Paper* **2025**, 1894.
48. USGS. National Water-Quality Assessment (NAWQA) Program. Available online: https://www.usgs.gov/mission-areas/water-resources/science/national-water-quality-assessment-nawqa (accessed on
49. Bundschuh, J.; Litter, M.I.; Parvez, F.; Román-Ross, G.; Nicolli, H.B.; Jean, J.-S.; Liu, C.-W.; López, D.; Armienta, M.A.; Guilherme, L.R. One century of arsenic exposure in Latin America: A review of history and occurrence from 14 countries. *Science of the total Environment* **2012**, *429*, 2-35.
50. Wick, K.; Heumesser, C.; Schmid, E. Groundwater nitrate contamination: factors and indicators. *Journal of environmental management* **2012**, *111*, 178-186.
51. Shaji, E.; Santosh, M.; Sarath, K.; Prakash, P.; Deepchand, V.; Divya, B. Arsenic contamination of groundwater: A global synopsis with focus on the Indian Peninsula. *Geoscience frontiers* **2021**, *12*, 101079.
52. Kumar, R.; Ali, S.; Sandanayake, S.; Islam, M.A.; Ijumulana, J.; Maity, J.P.; Vithanage, M.; Armienta, M.A.; Sharma, P.; Hamisi, R. Fluoride as a global groundwater contaminant. In *Inorganic Contaminants and Radionuclides*; Elsevier: 2024; pp. 319-350.
53. Uddin, M.G.; Nash, S.; Rahman, A.; Olbert, A.I. A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. *Water Research* **2022**, *219*, 118532.
54. Geng, Y., Liu, Y., Li, P., Sun, J., Jiang, Y., Pan, Z., Li, Y.-Z., & Zhang, Z. Anthropogenic activity and climate change exacerbate the spread of pathogenic bacteria in the environment. *Science* **2025**, *11*, doi:10.1126/sciadv.ads4355.
55. Liu, F.; Zou, J.; Liu, J.; Zhang, J.; Zhen, P. Factors controlling groundwater chemical evolution with the impact of reduced exploitation. *Catena* **2022**, *214*, 106261.
56. Benz, S.A.; Irvine, D.J.; Rau, G.C.; Bayer, P.; Menberg, K.; Blum, P.; Jamieson, R.C.; Griebler, C.; Kurylyk, B.L. Global groundwater warming due to climate change. *Nature Geoscience* **2024**, 1-7.
57. Abanyie, S.K.; Apea, O.B.; Abagale, S.A.; Amuah, E.E.Y.; Sunkari, E.D. Sources and factors influencing groundwater quality and associated health implications: A review. *Emerging Contaminants* **2023**, *9*, 100207.

58. Wang, C.; Gomez - Velez, J.D.; Wilson, J.L. The importance of capturing topographic features for modeling groundwater flow and transport in mountainous watersheds. *Water Resources Research* **2018**, *54*, 10,313-310,338.
59. Zarate, E.; Hobley, D.; MacDonald, A.; Swift, R.; Chambers, J.; Kashaigili, J.; Mutayoba, E.; Taylor, R.; Cuthbert, M. The role of superficial geology in controlling groundwater recharge in the weathered crystalline basement of semi-arid Tanzania. *Journal of Hydrology: Regional Studies* **2021**, *36*, 100833.
60. Jude, N.N.; Wotany, E.R.; Agyingi, C.; Nelson, M.A. Geological influence on groundwater quality in volcanic aquifers of Eastern Mount Cameroon, West of the Penda Mboko River. *Discover Applied Sciences* **2024**, *6*, 541.
61. Keesstra, S.D.; Geissen, V.; Mosse, K.; Piiranen, S.; Scudiero, E.; Leistra, M.; van Schaik, L. Soil as a filter for groundwater quality. *Current Opinion in Environmental Sustainability* **2012**, *4*, 507-516.
62. Olomo, O.K.; Danga, O.A.; Aliyu, A.O. Exploration of quality groundwater through lineament delineation in Okene and its surroudings. *Geosystems and Geoenvironment* **2025**, *4*, 100350.
63. Bense, V.; Gleeson, T.; Loveless, S.; Bour, O.; Scibek, J. Fault zone hydrogeology. *Earth-Science Reviews* **2013**, *127*, 171-192.
64. Guo, Z.; Chen, K.; Yi, S.; Zheng, C. Response of groundwater quality to river-aquifer interactions during managed aquifer recharge: A reactive transport modeling analysis. *Journal of Hydrology* **2023**, *616*, 128847.
65. El Osta, M.; Masoud, M.; Alqarawy, A.; Badran, O. Utilizing of aquifer hydraulic parameters to assess the groundwater sustainability in the new reclamation area of Moghra Oasis: Western Desert—Egypt. *Applied Water Science* **2023**, *13*, 238.
66. Sun, L.; Zhang, Y.; Si, H.; Ealotswe, T.K.; Wei, L.; Jin, P.; Song, Z.; Lu, C. Simulation and prediction of shallow groundwater depth in the North China Plain based on regional periodic characteristics. *Environmental Earth Sciences* **2021**, *80*, 1-11.
67. Thapa, R.; Gupta, S.; Guin, S.; Kaur, H. Sensitivity analysis and mapping the potential groundwater vulnerability zones in Birbhum district, India: a comparative approach between vulnerability models. *Water Science* **2018**, *32*, 44-66.
68. Lerner, D.N.; Harris, B. The relationship between land use and groundwater resources and quality. *Land use policy* **2009**, *26*, S265-S273.
69. Siddik, M.S.; Tulip, S.S.; Rahman, A.; Islam, M.N.; Haghighi, A.T.; Mustafa, S.M.T. The impact of land use and land cover change on groundwater recharge in northwestern Bangladesh. *Journal of Environmental Management* **2022**, *315*, 115130.
70. Akshitha, V.; Balakrishna, K.; Udayashankar, H.N. Impact of land use/land cover on groundwater resources in tropical unconfined aquifers of south-western India. *Environmental Research* **2023**, *218*, 114994.
71. Elzain, H.E.; Chung, S.Y.; Venkatramanan, S.; Selvam, S.; Ahemd, H.A.; Seo, Y.K.; Bhuyan, M.S.; Yassin, M.A. Novel machine learning algorithms to predict the groundwater vulnerability index to nitrate pollution at two levels of modeling. *Chemosphere* **2023**, *314*, 137671.
72. Elzain, H.E.; Abdalla, O.; Ahmed, H.A.; Kacimov, A.; Al-Maktoumi, A.; Al-Higgi, K.; Abdallah, M.; Yassin, M.A.; Senapathi, V. An innovative approach for predicting groundwater TDS using optimized ensemble machine learning algorithms at two levels of modeling strategy. *Journal of Environmental Management* **2024**, *351*, 119896.
73. Rao, Z.-f.; Jamieson, D. The use of neural networks and genetic algorithms for design of groundwater remediation schemes. *Hydrology and Earth System Sciences* **1997**, *1*, 345-356.
74. Inoue, K.; Tanaka, T.; Masaki, I. Application of a Genetic Algorithm to the Design of Groundwater Monitoring Well Networks under Conditions of Uncertainty. In Proceedings of the ISOPE International Ocean and Polar Engineering Conference, 2005; pp. ISOPE-I-05-096.
75. Fijani, E.; Nadiri, A.A.; Moghaddam, A.A.; Tsai, F.T.-C.; Dixon, B. Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh–Bonab plain aquifer, Iran. *Journal of hydrology* **2013**, *503*, 89-100.
76. Asadi, P.; Hosseini, S.M.; Ataie-Ashtiani, B.; Simmons, C.T. Fuzzy vulnerability mapping of urban groundwater systems to nitrate contamination. *Environmental modelling & software* **2017**, *96*, 146-157.

77. Mohammadrezapour, O.; Kisi, O.; Pourahmad, F. Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality. *Neural Computing and Applications* **2020**, *32*, 3763-3775.
78. Elzain, H.E.; Chung, S.Y.; Park, K.-H.; Senapathi, V.; Sekar, S.; Sabarathinam, C.; Hassan, M. ANFIS-MOA models for the assessment of groundwater contamination vulnerability in a nitrate contaminated area. *Journal of Environmental Management* **2021**, *286*, 112162.
79. Agrawal, P.; Sinha, A.; Kumar, S.; Agarwal, A.; Banerjee, A.; Villuri, V.G.K.; Annavarapu, C.S.R.; Dwivedi, R.; Dera, V.V.R.; Sinha, J. Exploring artificial intelligence techniques for groundwater quality assessment. *Water* **2021**, *13*, 1172.
80. De Jesus, K.L.M.; Senoro, D.B.; Dela Cruz, J.C.; Chan, E.B. Neuro-particle swarm optimization based in-situ prediction model for heavy metals concentration in groundwater and surface water. *Toxics* **2022**, *10*, 95.
81. Pham, Q.B.; Tran, D.A.; Ha, N.T.; Islam, A.R.M.T.; Salam, R. Random forest and nature-inspired algorithms for mapping groundwater nitrate concentration in a coastal multi-layer aquifer system. *Journal of Cleaner Production* **2022**, *343*, 130900.
82. Saha, A.; Pal, S.C.; Chowdhuri, I.; Roy, P.; Chakrabortty, R. Effect of hydrogeochemical behavior on groundwater resources in Holocene aquifers of moribund Ganges Delta, India: Infusing data-driven algorithms. *Environmental Pollution* **2022**, *314*, 120203.
83. Jamei, M.; Karbasi, M.; Malik, A.; Abualigah, L.; Islam, A.R.M.T.; Yaseen, Z.M. Computational assessment of groundwater salinity distribution within coastal multi-aquifers of Bangladesh. *Scientific Reports* **2022**, *12*, 11165.
84. Wang, Y.; Bian, J.; Sun, X.; Ruan, D.; Gu, Z. Sensitivity-dependent dynamic searching approach coupling multi-intelligent surrogates in homotopy mechanism for groundwater DNAPL-source inversion. *Journal of contaminant hydrology* **2023**, *255*, 104151.
85. Gholami, V.; Sahour, H. A Hybrid Approach of Supervised Self-organizing Maps and Genetic Algorithms for Predictive Mapping of Arsenic Pollution in Groundwater Resources. *Exposure and Health* **2023**, 1-16.
86. Gholami, V.; Sahour, H.; Khaleghi, M.R.; Ghajari, Y.E.; Sahour, S. Combining artificial neural networks and genetic algorithms to model nitrate contamination in groundwater. *Natural Hazards* **2024**, 1-21.
87. Bakhtiarizadeh, A.; Najafzadeh, M.; Mohamadi, S. Enhancement of groundwater resources quality prediction by machine learning models on the basis of an improved DRASTIC method. *Scientific Reports* **2024**, *14*, 1-24.
88. Lakshminarayanan, B.; Ramasamy, S.; Anuthaman, S.N.; Karuppanan, S. New DRASTIC framework for groundwater vulnerability assessment: bivariate and multi-criteria decision-making approach coupled with metaheuristic algorithm. *Environmental Science and Pollution Research* **2022**, *29*, 4474-4496.
89. Nadiri, A.A.; Gharekhani, M.; Khatibi, R. Mapping aquifer vulnerability indices using artificial intelligence-running multiple frameworks (AIMF) with supervised and unsupervised learning. *Water resources management* **2018**, *32*, 3023-3040.
90. Nadiri, A.A.; Norouzi, H.; Khatibi, R.; Gharekhani, M. Groundwater DRASTIC vulnerability mapping by unsupervised and supervised techniques using a modelling strategy in two levels. *Journal of hydrology* **2019**, *574*, 744-759.
91. Balaji, L.; Saravanan, R.; Saravanan, K.; NA, S. Groundwater vulnerability mapping using the modified DRASTIC model: the metaheuristic algorithm approach. *Environmental Monitoring and Assessment* **2021**, *193*, 1-19.
92. Jahromi, M.N.; Gomeh, Z.; Busico, G.; Barzegar, R.; Samany, N.N.; Aalami, M.T.; Tedesco, D.; Mastrocicco, M.; Kazakis, N. Developing a SINTACS-based method to map groundwater multi-pollutant vulnerability using evolutionary algorithms. *Environmental Science and Pollution Research* **2021**, *28*, 7854-7869.
93. Lakshminarayanan, B.; Ramasamy, S.; Yadav, B. Assessing the future groundwater vulnerability of an urban region under variable climatic and land use conditions. *Urban Climate* **2023**, *52*, 101691.
94. Siarkos, I.; Arfaoui, M.; Tzoraki, O.; Zammouri, M.; Hamzaoui-Azaza, F. Implementation and evaluation of different techniques to modify DRASTIC method for groundwater vulnerability assessment: A case study from Bouficha aquifer, Tunisia. *Environmental Science and Pollution Research* **2023**, *30*, 89459-89478.

95. Bordbar, M.; Heggy, E.; Jun, C.; Bateni, S.M.; Kim, D.; Moghaddam, H.K.; Rezaie, F. Comparative study for coastal aquifer vulnerability assessment using deep learning and metaheuristic algorithms. *Environmental Science and Pollution Research* **2024**, 1-15.
96. Alizamir, M.; Kazemi, Z.; Kazemi, Z.; Kermani, M.; Kim, S.; Heddam, S.; Kisi, O.; Chung, I.-M. Investigating landfill leachate and groundwater quality prediction using a robust integrated artificial intelligence model: Grey wolf metaheuristic optimization algorithm and extreme learning machine. *Water* **2023**, *15*, 2453.
97. Mazraeh, A.; Bagherifar, M.; Shabanlou, S.; Ekhlasmand, R. A Hybrid Machine Learning Model for Modeling Nitrate Concentration in Water Sources. *Water, Air, & Soil Pollution* **2023**, *234*, 721.
98. Kisi, O.; Keshavarzi, A.; Shiri, J.; Zounemat-Kermani, M.; Omran, E.-S.E. Groundwater quality modeling using neuro-particle swarm optimization and neuro-differential evolution techniques. *Hydrology research* **2017**, *48*, 1508-1519.