



Application of Artificial Intelligence Models for modeling Water Quality in Groundwater: Comprehensive Review, Evaluation and Future Trends

Marwah Sattar Hanoon · Ali Najah Ahmed  · Chow Ming Fai ·
Ahmed H. Birima · Arif Razzaq · Mohsen Sherif · Ahmed Sefelnasr ·
Ahmed El-Shafie

Received: 27 June 2021 / Accepted: 16 August 2021 / Published online: 2 October 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract This study reported the state of the art of different artificial intelligence (AI) methods for groundwater quality (GWQ) modeling and introduce a brief description of common AI approaches. In addition a bibliographic review of practices over the past two decades, was presented and attained result were compared. More than 80 journal articles from 2001 to 2021 were review in terms of characteristics and capabilities of developing methods, considering data of input-output, etc. From the reviewed studies, it could be concluded that in spite of various weaknesses, if the artificial

intelligence approaches were appropriately built, they can effectively be utilized for predicting the GWQ in various aquifers. Because many steps of applying AI methods are based on trial-and-error or experience procedures, it's helpful to review them regarding the special application for GWQ modeling. Several partial and general findings were attained from the reviewed studies that could deliver relevant guidelines for scholars who intend to carry out related work. Many new ideas in the associated area of research are also introduced in this work to develop innovative approaches and to improve

M. S. Hanoon
College of Technical Engineering, Islamic University, Najaf, Iraq
e-mail: marwa.satar88@gmail.com

A. N. Ahmed (✉)
Institute of Energy Infrastructure (IEI), Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor Darul Ehsan, Malaysia
e-mail: mahfoodh@uniten.edu.my

C. M. Fai
Discipline of Civil Engineering, School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia
e-mail: chow.mingfai@monash.edu

A. H. Birima
Department of Civil Engineering, College of Engineering, Qassim University, Unaizah, Saudi Arabia
e-mail: Ah.mohammed@qu.edu.sa

M. S. Hanoon · A. Razzaq
College of Science, Al Muthanna University, Samawah, AL-Muthanna, Iraq

A. Razzaq
e-mail: arifoobaid@gmail.com

M. Sherif
Civil and Environmental Eng. Dept., College of Engineering, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates
e-mail: MSherif@uaeu.ac.ae

M. Sherif · A. Sefelnasr
National Water and Energy Center, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates

A. Sefelnasr
e-mail: ahmed.sefelnasr@uaeu.ac.ae

A. El-Shafie
Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia
e-mail: elshafie@um.edu.my

the quality of prediction water quality in groundwater for example, it has been found that the combined AI models with metaheuristic optimization are more reliable in capturing the nonlinearity of water quality parameters. However, in this review few papers were found that used these hybrid models in GWQ modeling. Therefore, for future works, it is recommended to use hybrid models to more further investigation and enhance the reliability and accuracy of predicting in GWQ.

Keywords Groundwater quality (GWQ) · Artificial intelligence (AI) · Machine learning (ML) · ANN · ANFIS

Nomenclature

| | | | |
|-----------------|--|---------|--|
| CO ₃ | Carbonate | T | Temperature |
| NO ₂ | Nitrite | RSC | Residual Sodium Carbonate |
| RBF | Radial basis function | RBIAS | Relative Bias |
| MLP | Multi-layer perceptron | SOM | Self-organized map |
| DL | Deep learning | ANEP | Average Normalized Error for Parameter Estimates |
| RF | Random forest | LWPR | Locally weighted projection regression |
| XGBoost | eXtreme gradient boosting | RVM | Relevance vector machines |
| R ² | Determination coefficient | BNN | Bayesian neural network |
| RMSE | Root mean squared error | RE | Reduction of error |
| MARE | Mean absolute relative error | IA | Index of agreement |
| MLP | Multilayer perceptron | KSOFM | Kohonen self-organizing features map |
| GEP | Gene expression programming. | FGQI | Fuzzy-GIS-based groundwater quality index |
| SAR | Sodium adsorption ratio | ASVR | Active Set Support Vector Regression |
| MAPE | Mean absolute percentage error | MAR | Magnesium Adsorption Ratio |
| SI | Scatter index | PMRE | percent mean relative error |
| R | Correlation coefficient | GQI | Groundwater quality index |
| MSE | Mean square error | MARS | Multivariate adaptive regression spline |
| MAE | Mean absolute error | M5 Tree | M5 Tree model |
| FCM | Fuzzy c-means | GA | Genetic Algorithm |
| GP | Grid partition | GEP | Gene expression programming. |
| PSO | Particle swarm optimization. | TOC | Total organic carbon |
| NF | Neuro-fuzzy | NSE | Nash-Sutcliffe efficiency |
| SAR | Sodium absorption ratio | WHO | World health organization |
| BNs | Bayesian networks | LMI | Legates and McCabe index |
| MTEs | Mixtures of Truncated Exponentials | SDR | Standard deviation ratio |
| XGB | Extreme gradient boosting | WI | Willmott index of agreement |
| VAF | Variance account for | NE | Normalized error |
| PAEE | Percent Average Estimation Error | MLR | Multiple linear regression |
| PS | Potential Salinity | SEM | Structural equation modeling |
| DENFIS | Dynamic evolving neural-fuzzy inference system | GIS | Geographic information system |
| ESP | Exchangeable Sodium Percentage | FCT | Fuzzy Clustering Technique |
| SVR | Support vector regression | ACOR | Ant colony optimization for continuous domains |
| GMDH | Group method of data handling | Mmce | Mean misclassification error |
| | | AARE | Average absolute relative error |
| | | GRNN | generalized regression neural network |
| | | ASE | average squared error |
| | | RSC | Residual sodium carbonate |
| | | PSVM | Probabilistic Support Vector Machine |
| | | MAR | Magnesium adsorption ratio |
| | | KR | Kellys ratio |
| | | BPNN | Back-propagation neural network |
| | | DE | Differential evolution. |
| | | GP | Gaussian Process |
| | | RT | Random tree |
| | | PBIAS | Percent of bias. |
| | | PSVMs | Probabilistic support vector machines |

| | |
|---------|--|
| PNNs | Probabilistic neural networks |
| DO | Dissolved oxygen |
| TA | Total alkalinity |
| PBIAS | Percent of bias. |
| BOD | Biological oxygen demand |
| LSSVM | Least square support vector machine |
| COD | Chemical oxygen demand |
| SOM | Self-organizing map |
| FFNN | Feed forward neural network |
| FNN-SVR | Fuzzy neural network-based support vector regression |
| CE | Coefficient of efficiency |
| AIC | Akaike information criterion |
| KNN | K-nearest neighbor |
| WNN | Wavelet neural network |
| MFIS | Mamdani Fuzzy Inference System |
| As | Arsenic |
| ELM | Extreme learning machine |
| MLP | Multi-layer perceptron |
| MABE | Mean absolute bias error |
| PCR | Principal component regression |
| BR | Bayesian regulation |
| RR | Recharge rate |
| A | Abstraction |
| AVR | Abstraction average rate |
| LT | Lifetime |
| GWL | Groundwater level |
| AT | Aquifer thickness |
| DSWS | Depth from the surface to well screen |
| DSSL | Distance from sea shoreline |
| TR | Total rainfall |
| RH | Relative humidity |
| Tmin | Minimum temperature |
| GPR | Boosted regression tree |
| Tmax | Maximum temperature |
| Tavg | Average temperature |
| TPH | Total Petroleum Hydrocarbon |
| W | Average wind speed |
| NSGA-II | Non-dominated sorting genetic algorithm-II |
| Wmin | Minimum wind speed |
| MT3D | Modular three-dimensional transport model |
| Wmax | Maximum wind speed |
| ICC | Initial chloride concentration |
| GPR | Gaussian process regression |
| CGA | Continuous genetic algorithm |
| PSO | Particle swarm optimization. |
| DE | Differential evolution. |
| ROC | Receiver operating characteristics |

| | |
|------|------------------------------------|
| AUC | Area under the ROC curve statistic |
| FWQI | Fuzzy water quality index |
| TPR | True positive rate |
| SC | Specific conductance |
| WQI | Water quality index |
| SDT | Single decision tree |
| DTF | Decision tree forest |
| DTB | Decision treeboost |
| RP | Redox potential |
| SSE | Sum of squared errors |
| SOM | Self-organizing map |

1 Introduction

Water is necessary for life and is critical for irrigation, drinking water, and industry. Therefore, many techniques have been proposed to monitor water quality parameters, which is important for assessing the hydrological status of water and its management activities (Varotsos et al. 2020a; Varotsos et al. 2019; Krapivin et al. 2017).

Ground-water is an essential element of the natural water resources system and people have used it continuously since antiquity. The progress in drilling and pumping knowledge led to easy use of groundwater and enables people to reach a store of extremely deep aquifers. Ground-water is inexpensive compared with treating surface water. Along with these advantages, it includes nutrients that are beneficial for health. The use of groundwater could be very helpful to address a water scarcity in regions where surface water is restricted. Also, it could be utilized to complement surface water resources.

Human activities influence the quality of ground water supplies because of the dangerous chemical substances generated via agricultural and manufacturing practices. Nitrate (NO_3^-) is a common chemical contaminant that exists in various aquifers worldwide (Nadiri et al. 2019). NO_3^- has ability to penetrate into ground water therefore, the high concentrations of it $\text{NO}_3^- > 10$ mg per liter are dangerous and cause and harmful effect to humans (RadFard et al. 2019). Evaluating the groundwater pollution caused by NO_3^- contamination is a difficult matter due to the complexity of associated uncertainties and NO_3^- transport (Vadiati et al. 2016). Moreover, another common pollutant in groundwater is salinity, comprising various ions,

including chloride, magnesium, sodium, nitrate, bicarbonate, calcium, and sulfate. The simulation experiments performed in (Varotsos and Krapivin 2018; Krapivin et al. 2021) showed that heavy metals, oil hydrocarbons, and radionuclides are considered as primary contaminants of water. Therefore, assessment of groundwater pollutants such as heavy metals is essential to human health. Toxic ions, for instance, bromide (Br⁻), boron (B), and iron (Fe) can accumulate at advanced level (Nasr and Zahran 2014). Additionally, some of the heavy metals consider as micronutrients could cause harmful health impacts when they have contents exceeding the allowable limits in drinking water (Prasanna et al. 2011; Prasad et al. 2014; Sobhanardakani 2016; Varotsos et al. 2021).

Pollution could limit and influence the development of groundwater for potable uses. Therefore, ensuring quality of water and its accessibility has a great effect on quality of life. Ground-water is a main water resource in semiarid regions; its quality and accessibility are major concerns for hydrogeologists and environmental managers (Jiang et al. 2009; Bain et al. 2014). Changes in GWQ could be produced via changing ground water flow characteristics that could leads to dissolution and transportation of various minerals within aquifers (Hem 1985; Swarna Latha and Nageswara Rao 2012; Knoll et al. 2017).

This study aims to provide a comprehensive review and discuss the application of AI models for assessing the GWQ. In this context, application of many AI models is assessed. Moreover, a comprehensive literature review and meta analysis are conducted to discuss the past trends and future opportunities of AI models for predicting the GWQ.

1.1 Groundwater Quality (GWQ) Parameters

To determine the GWQ, the chemical listed in Table 1 should be considered:

In addition, the assessment of GWQ must take into consideration the following drinking water parameters: (Dohare et al. 2014; Varotsos et al. 2020b)

1. **pH:** is the inverse logarithm of H⁺ ion concentration. The pH scale ranges from 0 to 14, and pH values 7-14 are alkaline, from 0-7 are acidic, and 7 is neutral. The pH of mostly drinking water is between 4.4–8.5.

Table 1 Substances naturally discovered in some groundwaters that could impact GWQ

| Substances | Type of problem |
|--|--|
| Irons (Fe ⁺² or Fe ⁺³) | Toilet fixtures, encrustations and staining of laundry |
| Manganese (Mn ⁻²) | Toilet fixtures, encrustations and staining of laundry |
| Silica (SiO ₂) | Encrustations |
| Chloride (Cl ⁻) | Portability and Corrosion |
| Fluoride (F ⁻) | Caused fluorosis |
| Nitrate (NO ₃ ⁻) | Methemoglobinemia |
| Sulfate (SO ₄ ⁻²) | Portability |
| Dissolved oxygen (DO) | Corrosion |
| Carbon dioxide (CO ₂) | Corrosion |
| Miner constituents | Health domains and Portability |
| Calcium and Magnesium (Ca ²⁺ , Mg ²⁺) | Encrustations |

2. **Turbidity:** refers to the interference of some suspension of particles in water with the passing of light. It is produced via a broad set of suspended particles. It is measured either via its impact on the transmission of light, known as turbiditymetry, or via it is impact on the scattering of light, called nephelometry. According to IS: 10500–2012, an appropriate and acceptable limit are one and five.
3. **Total dissolved solids (TDS):** A variation of TDS is utilized to determine the ability of filter solids via assistance of a filtrate. Also, it could be estimated from the conductivity measurement in water samples. The standard and permitted limits according to IS: 10500–2012 are 500 and 2000 mg/l respectively.
4. **Electrical Conductivity (EC):** is the capability of water to hold an electrical current and differs both with number and kinds of ions the solution comprises. On the other hand, the conductivity of distilled water is lower than 1 μmhos per cm. Such conductivity relies on the existence of ions, their total concentrations, temperature of liquid flexibility, valency, and comparative concentration. Most inorganic acid solutions are good conductors such as salts and bases.
5. **Total hardness (TH):** According to IS: 10500–2012 the desirable and permissible limits for hardness are from 200 to 600 mg per liter, respectively. The impact of it size can be seen on utensils and hot water systems in boilers, etc. The amount of

TH for drinking water is categorized in terms of corresponding CaCO_3 concentrations as soft -0-60 mg/, moderate - 60-120 mg/, tough-120-180 mg/, extremely tough - >180 mg/.

6. **Sulfate (SO_4^{2-}):** Sulfate ions are present in natural water, and the majority of these ions are also soluble in water. Several sulfate ions are generated during the oxidation phase of their ores, and they are often found in industrial wastes. A UV spectrophotometer is used to calculate the amount of sulfate. According to IS: 10500–2012, the desirable limit for sulfate is 200 mg/ liter and the permissible limit is 400 mg/.
7. **Nitrate (NO_3^-):** is found in raw water and is mostly a form of the N_2 compound (of its oxidizing state). It is provided by chemical and fertilizer plants, animal waste, rotting crops, and domestic and industrial waste. A UV Spectrophotometer is used to determine the amount of nitrate. According to IS: 10500–2012, the desired limit for nitrate is max.45, and there is no relaxation of the allowable limit.
8. **Total alkalinity (Alk):** The number of the components in water that appear to raise the pH to the alkaline side of neutrality is referred to as alkalinity. It is usually expressed as milligrams per liter as calcium carbonate (m as CaCO_3) and is determined by titration with uniform acid to a pH of 4.5. CaCO_3 , PO_4 , and hydroxides are common alkalinity-increasing materials found in water.
9. **Chloride (Cl^-):** can be found in all forms of natural and raw water. It is derived from agricultural operations, industrial activities, and Cl-stones. Because of human activity, its concentration is high. According to IS: 10500–2012, the desirable limit for Cl^- is 250 and the permissible limit is 1000 mg/l.
10. **Fluoride (F^-):** is found in nature as fluorspar, rock phosphate, triphite, and phosphorite crystals, among other things. The temperature of the region, as well as the presence of accessory minerals in the rock minerals assemblage from which the groundwater circulates, are factors that regulate the concentration of F^- . According to IS: 10500–2012, the desirable limit for F^- is 1 mg/l and the permissible limit is 1.5 mg/l.
11. **Boron:** is found in nature as boric acid and boric acid salts. Weathering causes it to be expelled from rocks and soils and end up in water. It also enters

the soil and groundwater through domestic landfills that are not properly sealed. It is a common indicator compound that signals the presence of other potentially dangerous compounds. According to IS: 10500–2012 the desirable limit for boron is 0.5 and 1 mg/l the permissible limit.

12. **Phosphate (PO_4):** is a necessary plant nutrient that often regulates aquatic plant growth in freshwater. Because of the poor solubility of natural phosphate minerals and the ability of soils to absorb phosphate, groundwater usually absorbs only a minimal amount of phosphorus.
13. **Chemical Oxygen Demand (COD):** is a measurement of the amount of oxygen available for the chemical oxidation of organic matter using a powerful chemical oxidant. High COD levels can cause oxygen depletion due to microbe decomposition to levels harmful to aquatic life. COD determination has an advantage over BOD determination in that the outcome can be achieved in approximately 5 h as opposed to the 5 days needed by the BOD test.
14. **Zinc (Zn):** The earth's crust contains approximately 0.05 g/kg of zinc. Its most common mineral is sphalerite (ZnS), which is usually combined with other sulfide elements. Zinc poisoning symptoms in humans include vomiting, dehydration, electrolyte deficiency, and stomach discomfort. Zinc chloride has been linked to acute renal failure.

1.2 Problem Statement

Alterations in ground water chemical composition are affected via a number of geochemical processes and human activities that may be complicated to identify (Almasri and Kaluarachchi 2005; Yesilnacar and Sahinkaya 2012). Therefore, it's essential to comprehend the procedures that affect GWQ for efficient management of water resource and for management and protection of aquatic environments and such considerations could have the most important influence on sustainable development of countries (Batayneh et al. 2016; Fijani et al. 2017; Sheikhy Narany et al. 2014; Srinivas et al. 2015; Srinivasamoorthy et al. 2013; Krapivin et al. 2018). Conceptual or physical based models are conventionally major tools; nevertheless, they have several practical limitations, including the requirement for a

large volume of data and input parameters. GWQ assessment required a large number of samples with chemical, physical, and biological variables (Belkhiri et al. 2018). However, in various zones, large scale sampling become unfeasible because of it is expensive and a lack of facilities; thus, lower cost and quick monitoring approaches are required. Conventional techniques of GWQ estimation are based on mathematical model that have lower accuracy because they typically consider the linear relationship between dependent and independent parameters. Because of limitations in GWQ simulation, new computational methods are necessary to enhance estimate accuracy (Maroufpoor et al. 2020). Additionally, in various instances, data have limitations, and achieving accurate prediction is more essential than comprehending basic mechanisms; in contrast, the artificial intelligence (AI) models could be an appropriate alternate.

1.3 Research Objectives

To build new and better artificial intelligence methods for GWQ modeling, it's vital to examine what has been carried out with artificial intelligence approaches and recent research, and there is a necessity for scholars to understand what other researchers have performed in this respect. Consequently, the contributions of the current review are summarized as following:

1. This survey is summarizing 81 research studies based on artificial intelligence approaches and their advance application, growing over the past two decades (2001-2021), input-output data considerations, study areas, performance metrics, predicting period, and assessment.
2. The survey has emphasized the gaps in application of artificial intelligence approaches for predicting water quality in groundwater research and models that have not been investigated enough with hydro-chemical, and hydrogeological data of groundwater.
3. Lastly, the review shows the benefits and limitations of almost all models, the necessity for further research in different fields, concerns with data gathering, and recommendations for forthcoming improvements. Therefore, this survey study introduces a snapshot of historical evolution of AI methods over time. In depth explanation of AI models' architectures and mathematical theories have not been described in this survey. However, proper

references have been mentioned in background part of each AI method category for the readers who want additional detailed information.

The GWQ modeling employing AI approaches for any study includes the following steps: (1) data collection from remotely sensed/field, (2) dividing data into sets training, validation, and testing of AI methods, (3) selecting essential parameters that are important for training AI models, and (4) evaluating of attained output and comparing it with actual dataset then finding the precision of predictor, as in Fig.1. Two sorts of input parameters involving the hydro-chemical and aquifer characteristics parameters, or their patterns have been applied to predict the GWQ. The hydro-chemical parameters involve sodium (Na^+), temperature (T), calcium (Ca^{2+}), magnesium (Mg^{2+}), bicarbonate (HCO_3^-), SO_4^{2-} , Cl^- , TH, TDS, dissolved solids (DS), EC, pH, F^- , potassium (K), PO_4 , and Fe. Such parameters have been gathered via measures filed. The hydrogeological variables refer to aquifer characteristics such as land use. AI methods used for predicting the water quality of groundwater could be categorized as: (1) artificial neural network (ANN) models, (2) fuzzy logic (FL) based models, (3) support vector machines (SVM) models, (4) hybrid models, and (5) machine learning models.

Several papers (more than 80) have been published to predict the water quality in groundwater by AI models since 2001, as seen in Fig. 2. These studies have been published in international journals from well-known publishers such as Elsevier, Springer, IWA, Wiley, etc., between 2001 and 2021. The studies were obtained from searching the web utilizing the related keywords and were selected because they were published in famous international journals in the hydrology field, water resources field and AI sciences. It could be observed from Fig. 2 that AI models have had obvious growth recently, especially in 2017–2021 which gained the highest percent of published papers. Furthermore, many architecture variations have been noticed in addition to this exponential increase in applications of artificial intelligence approaches. The continued progress indicates the necessity for further exploration because every model has its own limitation and advantages.

Based on the search, Fig.3 delineates that the most frequently used models involve in ANN methods, with 52 papers followed by machine learning models with 21 papers. Also, it would be observed that the lowest

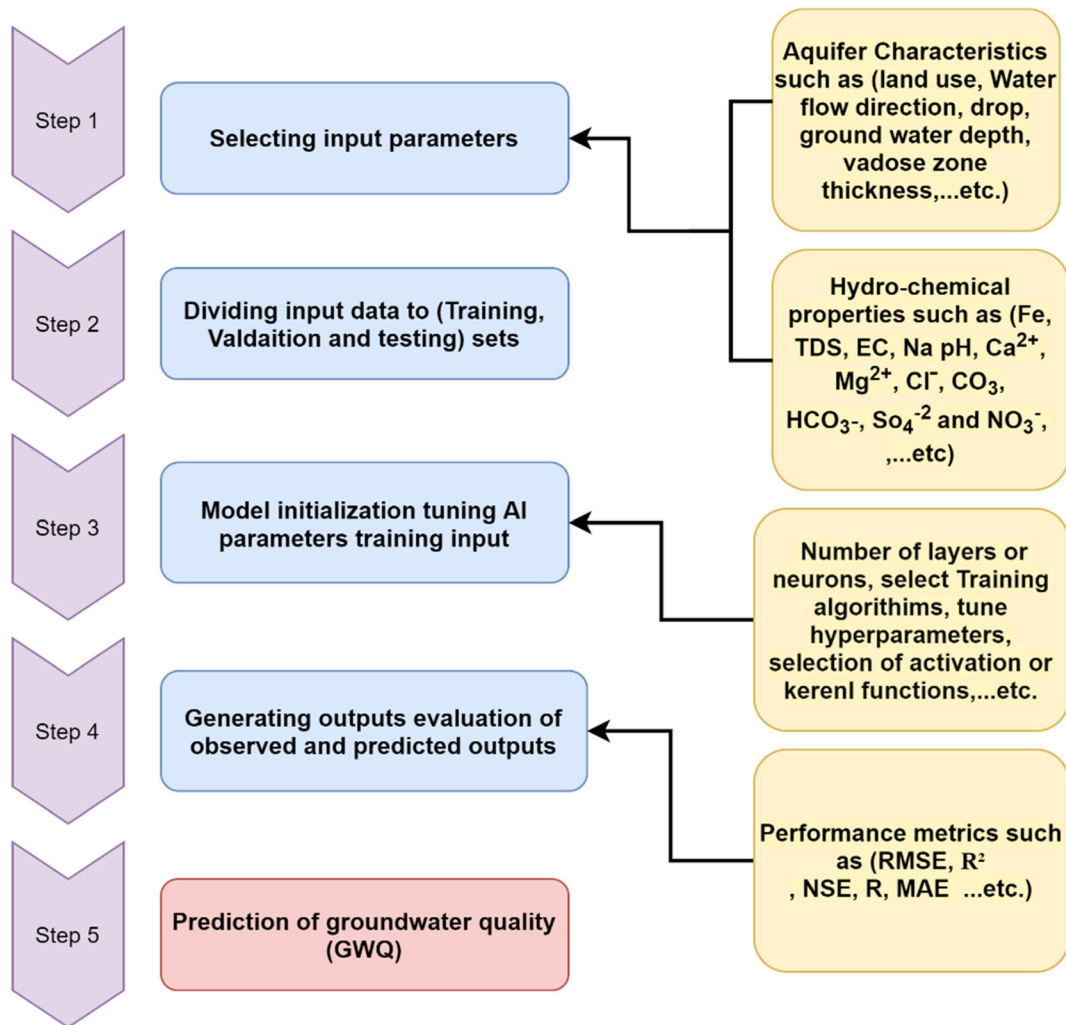


Fig. 1 Procedure of applying AI models for GWQ modeling

frequency of papers included in hybrid models compared with single models with only nine papers. The RBF, MLP and back propagation neural network (BPNN) are the most common ANN methods applied in the groundwater field. Moreover, application of machine learning methods substantially in ceased from 2017 to 2021, whereas SVM model have been infrequently applied recently. Even though a varied rate is observed in all publications, there is a substantial growing trend in applying AI models to predict GWQ from 2017, as shown in Fig.3.

In Table 2 many among the most widely used AI approaches for GWQ modeling are discussed. The methods include artificial neural network (ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machines (SVM), hybrid models, and various

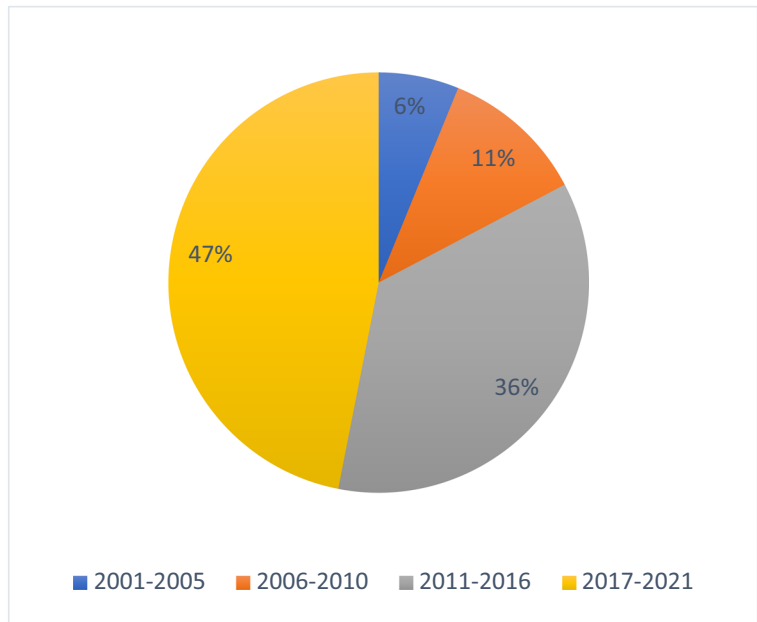
machine learning (ML) techniques. Firstly, a brief overview for every AI method is introduced and then the related conducted studies are cited and reviewed. This is followed by assessment and evaluation, conclusion, and future trends in research.

2 Artificial Intelligence Methods

2.1 Artificial Neural Network (ANN) for GWQ Modeling

Is a resilient method, using ANN inspired by brain processes to solve a complicated problems by making a relationship between input and output datasets. This relationship is produced via neurons. The connection

Fig. 2 Number of published papers (in the current study) that used AI models in field GWQ modeling



between inputs and outputs are created in training during the iteration process that is terminated after the specified criterion are satisfied. Feed forward neural networks (FFNN), that is easiest kind of neural network, is classifying to single layer perceptron (SLP) and MLP. SLP is makes linear connection between input and output, is the improper model for complicated and non-linear problem, whereas MLP provide a robust connection between input variable and output variable to modeling a nonlinear problem. Datasets are pass across the loop in order to get valid relation for input-output. The BPNN is well-known sort for MLP that is use a

back-ward method of adjusting parameter of networks. RBFNN is other model of ANN based model that is used the RBF as an activation function. RBFNN model was primarily presented by (Lowe, Broomhead, and Royal Signals and Radar Establishment 1988). Figure 4 shows a simple structure of ANN model.

2.1.1 Bibliographic Review

Current research in predicting GWQ reported that ANN can present as a superior replacement for

Fig. 3 Number of times AI methods are applied among reviewed papers for predicting GWQ

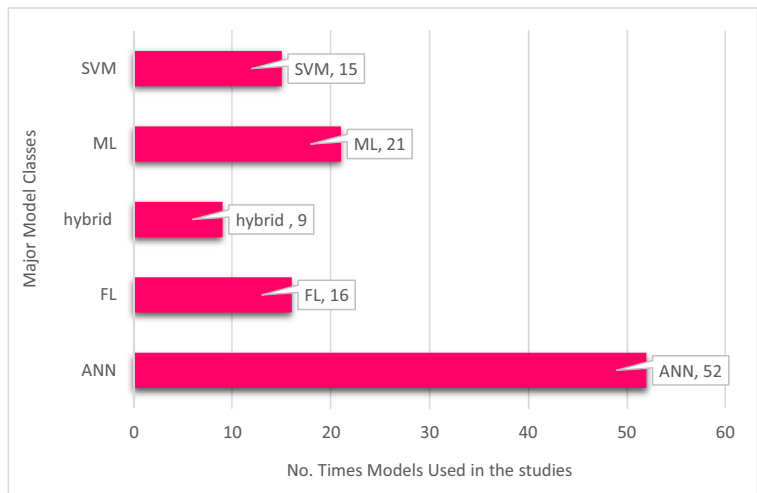


Table 2 Summary of studies that reported the application of AI models and their learning methods for modeling water quality in groundwater

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|---------------------------------------|------------------------------------|------------------------------|-----------------------------|--|--|---|
| 1 | (Yoon-Seok Hong * and Institute 2001) | Urban Water 3 | Mt Eden Aquifer, New Zealand | KSOM | Rain, Zn, Ca ²⁺ , Pb, Mg ²⁺ , Cl ⁻ , NO ₃ -N, K, Na, SiO ₂ , Cu, Cr, Cd/ NO ₃ | NA | 1999 |
| 2 | (Singh and Datta 2004) | Environmental Forensics | India | ANN | Specification of hydrogeologic parameters, boundary conditions for the aquifer, etc./ source of pollution | R, Total error, AARE, NE, ANEP, SD, PAEE | A period of ten years divide to 40 time steps (each of three months' interval) |
| 3 | (Kuo et al. 2004) | Water Research | Taiwan | ANN | EC, TDS, Cl ⁻ , SO ₄ ⁻² , Na +, K +, and Mg ²⁺ and the As pollutant factor including concentrations of Alk, TOC/seawater intrusion and As pollutant | RMSE | 28 wells, seasonally 1992–1998 |
| 4 | (Khalil et al. 2005) | Water Resources Research | Washington | ANN, SVM, LWPR, RVM | Nitrogen loadings from manure and fertilizers/ NO ₃ | CE, MSE, RMSE | The pool of 440 patterns |
| 5 | (Dixon 2005) | Applied Geography | USA | FL-GIS, GPS, remote sensing | Recharge, application rate elevation, potentiometric surface, soils, Clay, confining unit geology, pesticide, well water quality, land use, well location/groundwater vulnerability | NA | 55 well |
| 6 | (Tutmez et al. 2006) | Computers and Geosciences | Mersin, Turkey | ANFIS | TDS/EC | NA | September through November 2002 Samples |
| 7 | (Sahoo et al. 2006) | Science of The Total Environment | USA | ANN | Time (month) of sample collection, categorical indices of depth to aquifer material, aquifer sensitivity to pesticide contamination, well depth, depth to water from land surface, pesticide leaching class, and additional travel distance in the saturated zone (such as distance from land surface to midpoint of well screen)/ pesticide | Detection efficiency, R, mean error, RMSE, VAF | A monitor well network has 191 wells in 57 counties in Illinois. Samples from 159 wells were available to analyses. |
| 8 | (Wang et al. 2006) | Environmental Geology | Huantai County-China | ANN (BP-NN) | Depth to ground water, soil organic matter content, soil sand, field nitrogen surplus/ NO ₃ -N | RMSE, R ² | 2002–2004 |
| 9 | (El Tabach et al. 2007) | Mathematical and Computer Modeling | France | ANN | Density, soil permeability, retention properties, viscosity, surface tension of fluids, silty soil thickness, water table depth, pollutant pressure head on the soil surface, soil–pollutant contact time/ contamination of an unsaturated soil | R ² , ASE | 2-year period (September 2000 til April 2002) |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|-----------------------------|--|--|--------------------------|---|-----------------------------|---|
| 10 | (Dahiya et al. 2007) | Journal of Hazardous Materials | India | FSE | pH, TDS, TA, TH, Cl^- , Ca^{2+} , Mg^{2+} , SO_4^{2-} , NO_3^- and F^- /pH, TDS, TA, TH, Cl^- , Ca^{2+} , Mg^{2+} , SO_4^{2-} , NO_3^- and F^- | NA | A total of 42 samples of groundwater |
| 11 | (Purkait et al. 2008) | ISEIS Journal of Environmental Informatics | India | ANN, MLR, ASVR | Depth of tube well water, pH, DO, TDS, salinity, redox potential, and specific conductivity /As | RMSE, MAE, PMRE | A total of 85 water samples were collected |
| 12 | (Dixon 2009) | Hydrogeology Journal | Polk County, Florida, USA | SVM, ANN | Recharge of aquifer media, depth to ground water, soil media, effect of vadose zone, pH, hydraulic conductivity, land use, pedality, drainage, hydrologic group, organic matter, topography and bulk density/ NO_3^- | ROC, TPR | 6917 wells |
| 13 | (Gemitzi et al. 2010) | HYDROLOGICAL PROCESSES | South Rhodope Aquifer, Greece | ANN (MLP-BP) | Hydraulic conductivity, depth to the aquifer, soil permeability, impact of the unsaturated zone, land uses/ NO_3^- | RMSE | 2006 |
| 14 | (Bashi-Azghadi et al. 2010) | Expert Systems with Applications | Tehran, Iran | PNNs, PSVMs | Aquifer type, number of cells in the X direction, number of cells in the Y direction, average storage coefficient, average hydraulic conductivity, Cell dimensions, average thickness of the aquifer /temporal and spatial variations of the TPH concentration | NA | 37 wells |
| 15 | (Gholami et al. 2011) | Applied Water Science | Sarcheshmeh cooper mine and Shur River, Iran | BPNN, SVM | EC, pH, Ca^{2+} , TDS, Mg^{2+} , SO_4^{2-} , HCO_3^- , /Fe, Ni | R, RMSE | 56 samples 2006 |
| 16 | (Cho et al. 2011) | Water Research | Cambodia, Laos, and Thailand | ANN, MLR, PCR and PC-ANN | pH, EC, TDS, T, Redox potential /As | NSE, MAE | A total sample were 151 (2008–2010) |
| 17 | (Sirat 2012) | Arabian Journal of Geosciences | USA | ANN | Animals in past, well depth, distance to septic systems, flush window (days), distance to crop land, depth to aquifer material, aquifer type, distance to streams, aquifer class, spill or disposal at site, distance to barnyard, well site topography, onsite pesticide storage /pesticides | R | 1302 domestic and rural hydraulic wells |
| 18 | (Dar et al. 2012) | Revue des Sciences de l'Eau | India | ANN | pH, Cl^- , SO_4^{2-} , Ca^{2+} , Mg^{2+} , Na, HCO_3^- , TDS, CO_3 , K, EC, PO_4 , NO_3^- /F | SD, MSE, Error, correlation | Not mentioned |
| 19 | (Maity et al. 2013) | Environmental Monitoring and Assessment | Maharashtra, India | BNN | EC, TDS, pH, HCO_3^- , Cl^- , SO_4^{2-} , NO_3^- , Ca^{2+} , Mg^{2+} , Na, K, ρ /TDS | R^2 , RMSE, R, RE, IA | 36 water samples collected during March 2011. |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|---------------------------------|---|--------------------------------|--------------------------------|---|------------------------|--|
| 20 | (Moasheri and Tabatabaie 2013) | International Journal of Agriculture | Iran | ANN-GA | PH, EC, Q, T/NO ₃ ⁻ | R, MAE, MSE | 142 wells 2008–2010 |
| 21 | (Moasheri et al. 2013) | International Journal of Agriculture and Crop Sciences. | Iran | ANN | Coordinates of points, EC, pH/TDS | R, MSE, MAE | 10 wells and aqueduct that six months duration in 2010 |
| 22 | (Fijani et al. 2013) | Journal of Hydrology | Iran | Hybrid SFL, MLF, ANNs, and NF. | Net recharge, soil media, topography, depth to water, aquifer media, impact of the vadose zone, and hydraulic conductivity to derive groundwater vulnerability /corrected vulnerability index used NO ₃ -N concentration | R ² , RMSE | 191 data points |
| 23 | (Nikoo and Mahjouri 2013) | Water Resources Management | Iran | FIS, PSVM, FCT and SOM | TDS, EC, pH, CO ₃ , Ca ²⁺ , Mg ²⁺ , Cl ⁻ , SO ₄ ⁻² , HCO ₃ ⁻ , Na ⁺ , K ⁺ , heavy metals and total and fecal Coliform bacteria/NO ₃ ⁻ , Cl and TDS | MARE | Water quality measurements at 102 monitoring wells in the Tehran Aquifer |
| 24 | (Aguilera et al. 2013) | Stochastic Environmental Research and Risk Assessment | Campo de Dall’as, Spain | hybrid BNs with MTEs | EC, NO ₃ ⁻ , Cu, Fe, and pesticides/ EC, NO ₃ ⁻ , Cu, Fe, and pesticides | NA | Total of 125 wells (sampling points) |
| 25 | (Nadiri et al. 2013) | Journal of Hydroinformatics | Maku, Iran. | MFL, ANN, SFL, NF | pH, Na, K, Ca ²⁺ , Mg ²⁺ , CO ₃ , HCO ₃ ⁻ , Cl ⁻ , SO ₄ ⁻² , EC, and SiO ₂ /F- | R ² , RMSE | 38 samples in August of year 2006 and the smallest data set eight samples in July of year 2004 |
| 26 | (Kleçasan et al. 2014) | International Journal of Distributed Sensor Networks | Turkey | ANN | pH, EC, TDS, Cl ⁻ , and TH ₁ /SAR and So ₄ ⁻² density | MSE | 108 sets collected from nine wells once per month for a total period of 1 year, whereas only 107 of them were available since in January 2004. |
| 27 | (Singh et al. 2014) | Journal of Hydrology | India | SDT, DTF, SVM, and DTB | pH, RP, EC, salinity (ppt), TS, TDS, T-Alk, CO ₃ , HCO ₃ ⁻ , T-Hard, calcium hardness (Ca-Hard, magnesium hardness (Mg-Hard, ammonical nitrogen (NH ₄ -N, NO ₃ , NO ₂ , Cl, So ₄ ⁻² , PO ₄ , F, SiO ₂ , K, Na, Ca ²⁺ , Mg ²⁺ /COD | R, MAE, RMSE | The total of 861 water sets were gathered through the pre- and post-monsoon months of 2007–2009. |
| 28 | (Choi et al. 2014) | Journal of Geochemical Exploration | South Korea | SOM, PCA | pH, TDS, Na, K, Ca ²⁺ , Mg ²⁺ , Cl ⁻ , SO ₄ ⁻² , HCO ₃ ⁻ , NO ₃ ⁻ /NO ₂ ⁻ | NA | 2008–2009 |
| 29 | (Rodriguez-Galiano et al. 2014) | Science of the Total Environment | Vega de Grande, Aquifer, Spain | LR, RF | Distance from rivers, distance from points of discharge of urban wastewater, distance from irrigation canal, distance from industries, distance from wastewater treatment plant, distance from dairy | MSE, RMSE | 1980–2004 |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|-------------------------|---|--|--------------------------------------|--|-----------------------------|---|
| 30 | (Alagha et al. 2014) | Environmental Monitoring and Assessment The Egyptian Journal of Aquatic Research | Gaza, Palestine | ANN, SVM | farms, SLOPE, aquifer properties, groundwater depth, hydraulic conductivity, hydraulic gradient, Soil kind, vadose zone thickness, NDVI, land cover, pH, T, EC/NO ₃ ⁻ , EC, TDS, NO ₃ , Cl ⁻ , SO ₄ ²⁻ , hardness, Ca ²⁺ , Mg ²⁺ and Na ⁺ /NO ₃ ⁻ | MAPE, NSE, RMSE | 22 Wells (2007–2010) |
| 31 | (Nasr and Zahran 2014) | Clean - Soil, Air, Water | Alexandria, Egypt | ANN | TDS, pH /groundwater salinity | MSE, R | The samples were harvested in summer, autumn, and winter, 2013 |
| 32 | (Khaki et al. 2015) | Earth Science Informatics | Langat Basin, Malaysia Kerman Plain, Iran | ANN, ANFIS ANFIS-GA and ANFIS-PSO | Na ⁺ , Cl ⁻ , K, pH, HCO ₃ ⁻ , Fe, Ca ²⁺ , Mg ²⁺ NO ₃ ⁻ , SiO ₂ / TDS, EC groundwater level, average rainfall, and local coordinates of the location / Cl ⁻ , EC, and pH | R, MSE | 6 years (2008–2013) |
| 33 | (Jalalkamali 2015) | Environmental Earth Sciences | Sivas, Karabuk, Bartin, Turkey | ANN | pH, EC, Na, K, Ca ²⁺ , Mg ²⁺ , CO ₃ , HCO ₃ ⁻ , Cl ⁻ , SO ₄ ²⁻ , Fe, Mn, Al, NO ₃ ⁻ / water pollution in groundwater | NA | 2000 and 2010 |
| 35 | (Nolan et al. 2015) | Journal of Hydrology | the Central Valley, California, USA | ANN, BN, BRT, RF | The 41 predictor variables represented soils, land use, groundwater age surrogates, and aquifer texture and MODFLOW-simulated vertical water fluxes from previous textural and numerical models of the Central Valley/ NO ₃ ⁻ | R ² , MSE | data set comprised 318 shallow domestic wells, and another 119 wells lacking screened interval data |
| 36 | (Sakizadeh 2015) | Modeling Earth Systems and Environment | Iran | ANN | EC, TDS, Turbidity, pH, TH, Ca ²⁺ , Mg ²⁺ , SO ₄ ²⁻ , PO ₄ , NO ₃ ⁻ , NO ₂ ⁻ , F ⁻ , Fe, Mn, Cu, Cr/WQI | R, MSE | 47 wells during 8 years' time period (2006–2013) |
| 37 | (Vadiati et al. 2016) | Journal of Environmental Management | Iran's Sarab Plain | MFIS | Ca ²⁺ , Mg ²⁺ , Na, Cl ⁻ , NO ₃ ⁻ , SO ₄ ²⁻ /TDS | VAF and RMSE | 2013–2014 |
| 38 | (Wagh et al. 2016) | Modeling Earth Systems and Environment | Nanded tehsil, Maharashtra, India | ANN | pH, EC, TDS, Ca ²⁺ , Mg ²⁺ , Na, K, Cl ⁻ , CO ₃ , HCO ₃ ⁻ , SO ₄ ²⁻ and NO ₃ ⁻ / SAR, RSC, MAR, KR and %Na | R ² , RMSE, MARE | 50 groundwater samples, 2012 |
| 39 | (Ehteshami et al. 2016) | Modeling Earth Systems and Environment | Iran | ANN | soil nitrogen content, soil organic matter and soil carbon content/NO ₃ ⁻ | MSE, R | 2011–2014 |
| 40 | (Arabgol et al. 2016) | Environmental Modeling and Assessment | Arak Plain Aquifer, Iran | SVM | T, EC, groundwater depth, TDS, DO, pH, land use, and season of the year/ NO ₃ ⁻ | RMSE, R | 160 water samples representing 40 different wells (2014–2015) |
| 41 | | | India | | | NA | |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|------------------------------------|---|--|-----------------------|--|----------------------------------|--|
| | (Prajapati and Parekh 2016) | International Journal of Science and Research (IJSR) | | Fuzzy Logic, MFIS | EC, TDS, pH, Cl ⁻ and Ca/ EC, TDS, PH, Cl ⁻ , and Ca ²⁺ | | 509 groundwater wells have been collected in pre monsoon season and 495 wells have been collected in post monsoon season |
| 42 | (Barzegar and Moghaddam 2016) | Modeling Earth Systems and Environment | Tabriz, Iran | MLP, RBFNN, and GRNN | Ca ²⁺ , Mg ²⁺ , Na, Cl ⁻ , SO ₄ ⁻² /EC | R ² , RMSE, MARE | 93 data records of groundwater samples |
| 43 | (Alizamir and Sobhanardakani 2016) | J Adv Environ Health Res | Iran | ANN | Previous heavy metals (As, lead (Pb), and Zn) / (As, Pb, and Zn) concentration | R ² , RMSE | Groundwater samples were collected from 30 different locations including open and tube wells in 3 replicates to evaluate the heavy metal contamination in 2012 |
| 44 | (Heidarzadeh 2017) | Water Supply: Research and Technology - AQUA | Amol-Babol in the Mazandaran province, Iran. | ANN | pH, EC, TH / conservative chemical property of Na | R ² , RMSE | (2132 data) 1987 to 2010 |
| 45 | (Mohammadi et al. 2017) | Pollution | Iran | ANN, ANN-POS | EC, SAR, pH, So ₄ , Ca, Mg, and Na, HCO ₃ , TDS /SAR, EC, and TDS | R ² , RMSE, MAE | 2009–2015 |
| 46 | (Wagh et al. 2017) | Modeling Earth Systems and Environment | Kadava river basin, Maharashtra, India | ANN | EC, TDS, TH, Ca ²⁺ , Mg ²⁺ , Na, Cl ⁻ , CO ₃ , HCO ₃ ⁻ , So ₄ ⁻² , F/ NO ₃ | R ² , RMSE, MARE | 2011 |
| 47 | (Kisi et al. 2017) | Hydrology Research | Neyshabur plain, Iran | ANN-DE, ANN-POS | EC, pH, Na, Ca ²⁺ , Mg ²⁺ , HCO ₃ ⁻ , Cl ⁻ , TH, TDS, SO ₄ ⁻² , SAR/ SAR, and EC | R ² , MAE, RMSE | Monthly groundwater records were collected from 60 observational wells during a 16-year period (1997–2013) |
| 48 | (Yang et al. 2017) | Journal of Hydro informatics | Fujian Province, South China | ANN, WNN | TDS, TH, COD, Cl ⁻ , SO ₄ ⁻² , NO ₃ ⁻ , NO, F/ water quality grade | MAE, RMSE, R ² | 2004–2011 |
| 49 | (Barzegar et al. 2017) | Stochastic Environmental Research and Risk Assessment | Iran | ELM, MLP and SVM | Na, K, Ca ²⁺ , and HCO ₃ ⁻ /F ⁻ | RMSE, R ² , MABE, NSE | 143 water samples were collected in a five-year period, 2004–2008 |
| 50 | (Nadiri et al. 2017) | Science of the Total Environment | Northwest of Tabriz, East Azerbaijan, Iran | LFL, SFL and MFL | Depth to water table, net Recharge, Aquifer media, Soil media, Topography, Impact of the vadose zone, and hydraulic Conductivity/ NO ₃ -N | RMSE, R ² | Each parameter mentioned in study |
| 51 | (Isazadeh et al. 2017) | Environmental Earth Sciences | Guilan Province, northern Iran | ANN, SVM | pH, HCO ₃ ⁻ , Na ⁺ , K ⁺ , Ca ²⁺ , EC, Cl ⁻ , SO ₄ ⁻² , distance from the shoreline, Ground water level/Na ⁺ , SO ₄ ⁻² , EC | RMSE, Error, R | 140 wells (2002–2014) |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|-------------------------------------|--|--|------------------------------------|--|---------------------------------|---|
| 52 | (Nakagawa et al. 2017) | Hydrology Research | Shimabara City in Nagasaki Prefecture, Japan | SOM | Cl ⁻ , NO ₃ ⁻ , SO ₄ ⁻² , HCO ₃ ⁻ , Na ⁺ , K ⁺ , Ca ²⁺ , Mg ²⁺ /Ca-(SO ₄ ⁻² +NO ₃ ⁻), Ca-HCO ₃ | NA | 2012–2015 |
| 53 | (Khudair et al. 2018) | International Journal of Plant & Soil Science | Baghdad city, Iraq | ANN | pH, Cl ⁻ , SO ₄ ⁻² /TDS | R ² | 29 wells |
| 54 | (Rodriguez-Galiano et al. 2018) | Science of the total environment | Vega de Grande, Spain | SVM, LR, RF, CART | Water flow direction, drop, ground water depth, vadose zone thickness, distance from population, transmissivity, module of hydraulic gradient, land cover, NDVI, population, distance to irrigation canals/ NO ₃ | rmsee and AUC | 2014, 2015 |
| 55 | (Zaqoot et al. 2018) | Applied Artificial Intelligence | Gaza Strip, Palestine | ANN (MLP-NN, RBF-NN) | pH, EC, TDS, hardness, Ca ²⁺ , Mg ²⁺ , abstraction rate/NO ₃ ⁻ | RMSE, MAE, R | 15 groundwater wells, 2013–2014 |
| 56 | (Jebastina and Prince Arulraj 2018) | Bulletin of Environmental Contamination and Toxicology | Coimbatore, India | ANFIS | pH, SAR, So ₄ ⁻² , Na, K, Cl ⁻ , Hardness, EC, Ca ²⁺ , Mg ²⁺ , HCO ₃ ⁻ /NO ₃ ⁻ | RMSE, R ² | 6 months interval during the years (2011–2012) |
| 57 | (Alizamir and Sobhamardakani 2018) | Jundishapur Journal of Health Sciences | Iran | ANN-POS and ANN-BR | As, Zn, Pb, Cu/ As, Zn, Pb, Cu | RMSE, R ² , R | 20 different wells |
| 58 | (Jafari et al. 2019) | Water Supply: Research and Technology - AQUA | Tabriz plain, Iran | MLP, GEP, SVM | Na, HCO ₃ ⁻ , Ca ²⁺ , Mg ²⁺ , and SO ₄ ⁻² /TDS | R, RMSE and MAE | 10 years (2002–2012) totaled 1742 samples |
| 59 | (Kisi et al. 2019) | Water Resources Management | Isfahan-Borkhar, Iran | ANFIS | EC, TH, SAR/TH | RMSE, MAPE, R ² , SI | 18 years (1998–2015). |
| 60 | (Aryafar et al. 2019) | Environmental Earth Sciences | Khezri plain, Eastern Iran | ANN, ANFIS and GP | K ⁺ , Na ⁺ , Ca ²⁺ , Mg ²⁺ , SO ₄ ⁻² , Cl ⁻ , HCO ₃ ⁻ , CO ₃ , pH/TH, TDS, and EC | R ² , RMSE, MARE | 2007–2017 (240 sample) |
| 61 | (Vijay and Kamaraj 2019) | International Journal of Research and Analytical Reviews | North of TamilNadu in India | C5.0, Naive Bayes and RF, RStudio. | pH, EC, TH, TDS, Aik, Ca ²⁺ , Mg ²⁺ , SO ₄ ⁻² , NO ₃ ⁻ , Cl ⁻ , BOD, COD, Fe/ pH, EC, TH, TDS, Aik, Ca ²⁺ , Mg ²⁺ , SO ₄ ⁻² , NO ₃ ⁻ , Cl ⁻ , BOD, COD, Fe | NA | Groundwater samples were collected from 30 Locations during month of May 2014 |
| 62 | (Kadam et al. 2019) | Modeling Earth Systems and Environment | India | ANN, MLR | pH, EC, TDS, TH, Ca, Mg, Na, K, Cl, SO ₄ ⁻² , HCO ₃ ⁻ , NO ₃ and PO ₄ /WQI | R ² | 34 samples during pre- and post-monsoon seasons in 2015 |
| 63 | (Rahmati et al. 2019) | Science of The Total Environment | | SVM K-NN, RF | Elevation, distance from stream, hydraulic conductivity, | R ² , RMSE | 34 wells (2015–2016) |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|----------------------------|---|---|---|--|---------------------------|---|
| 64 | (Knierim et al. 2020) | Journal of the American Water Resources Association | Andimeshk-dezful, Khuzestan Province, Iran | BRT | lineament density, land use/ NO_3^- | RMSE, R^2 | 1960 to 2018 |
| 65 | (Singha et al. 2020) | Journal of Contaminant Hydrology | Mississippi River Valley alluvial aquifer North America | DL, ANN | SC and Cl^- , TDS/SC, CI As, Cd, Cu, Cr, Fe, Mn^{2+} , Ni, Pb, and Zn/pollution indices | RMSE, MAE, MSE | Groundwater samples were collected from 226 locations, via dug wells and hand pumping, during April–June 2019 |
| 66 | (Selvaraj et al. 2020) | Environmental Science and Pollution Research | Chhattisgarh, India | MFIS | TH, TDS, pH, Ca^{2+} , Mg^{2+} , K, SO_4^{2-} , NO_3^- + NO_2^- , F, HCO_3^- , CO_3 and Cl^- /FWQI | NA | 30 groundwater samples |
| 67 | (Bedi et al. 2020) | Environmental Monitoring and Assessment | Tamilnadu, India | ANN, SVM, XGB | Multiple hydrogeologic, water quality, and land use /concentration of NO_3^- and pesticide | NA | 303 wells |
| 68 | (Jha et al. 2020) | Water Research | USA | FGQI | NO_3^- , N, Na, Cl^- , K, F^- , SO_4^{2-} , TDS, TH/ NO_3^- , N, Na, Cl^- , K, F^- , SO_4^{2-} , TDS, TH | NA | Seasonal, 24 observation wells during pre-monsoon season (July 2013) and 37 observation wells during post-monsoon season (January 2014) |
| 69 | (Esmacilbeiki et al. 2020) | Cleaner Production | Chhattisgarh state of India | DENFIS, GMDH, MARS, M5 Tree, GEP, ANFIS, MLR, SEM | TH, EC, Na, Cl^- , Ca^{2+} , Mg^{2+} , SO_4^{2-} , HCO_3^- /TH, EC | R, RMSE, MAE, WI, NSE LMI | 2017 |
| 70 | (Kumar et al. 2020) | Environmental Monitoring and Assessment | Kozhikode City, Kerala, India | GP, M5P, RF and RT | TH, SO_4^{2-} , Cl^- , Ca^{2+} , NO_3^- , Mg^{2+} , K+, and Na+ /TDS | RMSE | A total of 234 water samples |
| 71 | (Bui et al. 2020) | Science of The Total Environment | South Italy | hybrid intelligent ANN, FCM clustering, NF-GP, | T, pH, EC, HCO_3^- , F^- , Cl^- , SO_4^{2-} , Na+, K+, Ca^{2+} , Mg^{2+} / NO_3^- , stronzium | R^2 , NSE, RMSE, PBIAS | Datasets include 246 of GWQ samples originating from different wells |
| 72 | (Maroufpoor et al. 2020) | Environmental Science and Pollution Research | Bam Normashir Plain, Iran | | Cl^- , EC, SAR/ NO_3^- | R, RMSE, SI, MAE | 14-year (2003–2016) |

Table 2 (continued)

| NO. | Scholar(S) | Journals | Case study | Compared algorithm(s) | Input/output variables | Performance indicators | Range of total data |
|-----|------------------------------|--|---|-------------------------|---|-----------------------------------|--|
| 73 | (Sunayana et al. 2020) | Environment, Development and Sustainability Arabian Journal of Geosciences | Mumbai, India | PSO, NF system ANN | time (month and year), location (latitudes and longitudes), PH, TDS, SO_4^{-2} , Cl^- , NO_3^-/TH | RMSE, R | Data samples for 10 wells, were collected every month for year 2016 and 2017 |
| 74 | (Norouzi and Moghaddam 2020) | Arabian Journal of Geosciences | Iran | RF | TDS, Cl^- , Ca^{2+} , Mg^{2+} , SO_4^{-2} , HCO_3^- , K^+ , Na^+ , and NO_3^-/GQI | MSE | 75 water samples |
| 75 | (Khan et al. 2021) | Groundwater for Sustainable Development | India | ANN | Turbidity, pH, TDS, and EC/ <i>E.coli</i> . | R^2 , MSE | 1301 experimental laboratory |
| 76 | (Bilali et al. 2021) | Agricultural Water Management | Berrechid aquifer in Morocco | RF, ANN, SVR, Adaboost | EC, T, PH/TDS, PS, SAR, ESP, MAR, RSC | R, RMSE, RBIAS | 520 (2009 to 2019) |
| 77 | (Lopez et al. 2021) | Environmental Science & Technology | California | RF | Ca^{2+} , NO_3^- , and SO_4^{-2} concentrations, soil pH, and clay content (weighted average between 0 and 2 m depths)/uranium concentrations | RMSE | A total of 587 ground water sets and an additional 29 ground water sets that contained complete aqueous geochemistry |
| 78 | (Fadipe et al. 2021) | Nigerian Journal of Technology | Nigeria | ANN | turbidity, T, color, DO, TA, TH, Mg , SO_4^{-2} , NO_3^- , NO_2^- , Cl^- , HCO_3^- , CO_3 , Fe, Cr, Mn^{-2} , Pb, Zn and Cu/pH and TDS | R^2 , MSE | A total number of eighteen random chosen dug wells |
| 79 | (Singha et al. 2021) | Chemosphere | Arang of Raipur district, Chhattisgarh, India | ANN, DL, RF and XGBoost | pH, TDS, TH, Ca^{2+} , Mg^{2+} , Na, K, HCO_3^- , Cl^- , SO_4^{-2} , NO_3^- , F ⁻ , and PO_4^{3-} /pH, TDS, TH, Ca^{2+} , Mg^{2+} , Na, K, HCO_3^- , Cl^- , SO_4^{-2} , NO_3^- , F ⁻ , and PO_4 | MSE, R^2 , RMSE, NSE, MAPE, RSR | April, June 2019 total of 226 ground water |
| 80 | (Kassem et al. 2021) | Environmental Earth Sciences | Turkey | ANN | RR, A, AVR, LT, GWL, AT, DSWS, DSSL, TR, RH, Tmin, Tmax, Tavg, W, Wmin, Wmax, ICC/F | RMSE, R^2 | Not mentioned |
| 81 | (Stackelberg et al. 2021) | Groundwater | Northern USA | BRT | CO_3^- content of soils and aquifer materials/pH | %bias, NSE, RMSE, R^2 | Samples from 18,386 wells |

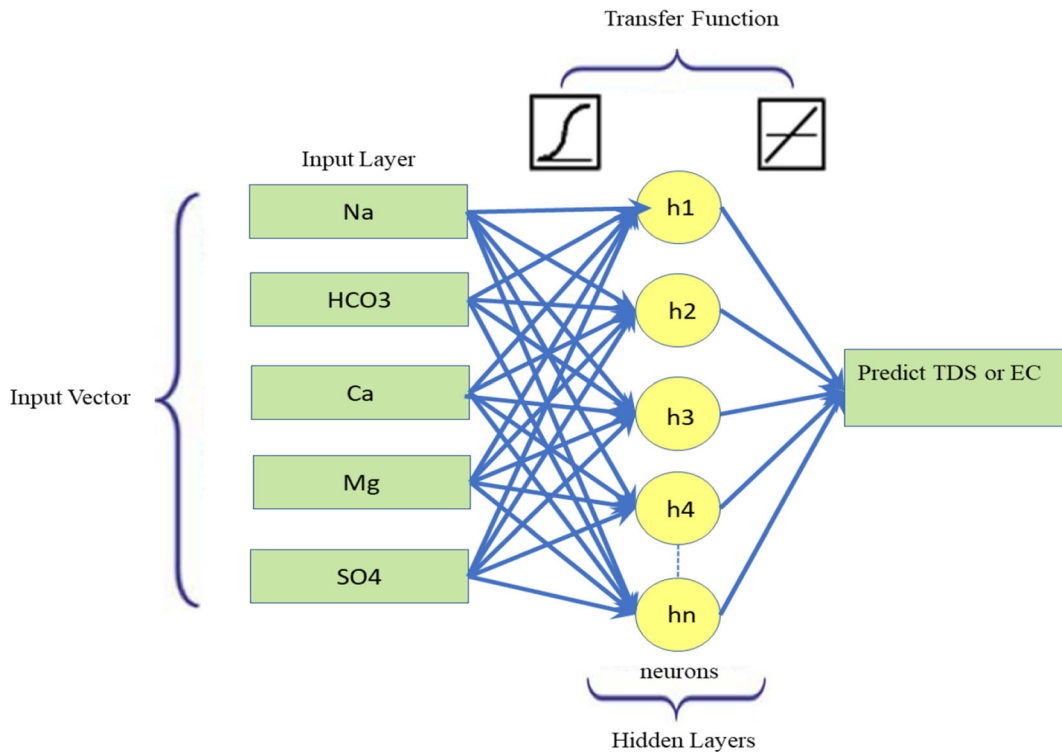


Fig. 4 Schematic view of ANN approach in prediction of GWQ

conventional techniques. Heidarzadeh (2017) used an ANN model to predict the GWQ of the Amol-Babol aquifer using a dataset from 1987 to 2010. Na was studied as an output variable in the ANN approach because of its high concentration in irrigation. Based on pre modeling, the variables pH, EC, and TH were selected as the better inputs. The researchers in (Khudair et al. (2018) investigated applying an ANN model using data for Baghdad, Iraq, for water parameters: pH, Cl⁻, SO₄⁻², and TDS. The findings gained using the ANN model showed high prediction efficiency and the parameters pH and Cl have substantially affected model prediction, and thus are critical factors. Kılçaslan et al. (2014) proposed an ANN method to estimate GWQ and compared it to conventional water quality evaluation techniques; this paper proved the effectiveness and accuracy of the ANN approach. Kuo et al. (2004) was used the BPNN to predict the variation of GWQ in Taiwan. Various kinds of back progression models were created in order to assess their performance. The findings showed that ANN was able to define the complex variation of GWQ and be applied to present accurate predicting. Furthermore, a number of

hidden neurons do not greatly effect performance of the model in training or testing. Sunayana et al. (2020) was evaluated GWQ from 2016 to 2017 at 10 locations in India. The TH was forecasted employing ANN by applying various training algorithms, then the finest one was utilized in the ANN final model to predict GWQ. They quantified the impact for both space and time on TH of ground-water. The ANN used here was able to forecast the TH concentrations of each point. Choi et al. (2014) observed the foremost ions, for instance, Ca²⁺, Mg²⁺, Na⁺, K⁺, HCO₃⁻, Cl⁻, SO₄⁻², and NO₃⁻, in 2008 to evaluate GWQ in South Korea. The hydro-chemical dataset of ground-water samples was analyzed applying the SOM method, which is a type of ANN model. The results revealed that this method could be effectively employed for classifying and characterizing groundwaters in terms of hydro-chemistry and quality on the area size. To predict NO₃⁻ concentrations of ground water wells in the southern area of the Gaza Strip, Zaqoot et al. (2018) used to ANN models MLP-NN and RBF-NN. Both were trained and developed using seven input variables (pH, EC, TDS, Ca²⁺, Mg²⁺, hardness, and abstraction rate), and the results of MLP-NN showed

better performance than RBF-NN. Wagh et al. (2017) applied the ANN approach by BP algorithm to predict GWQ in the Kadava River basin in the Nashik district using parameters EC, HCO_3^- , TDS, TH, Ca^{2+} , Na^+ , Cl^- , CO_3 , Mg^{2+} , F, and SO_4^{2-} . The mimic output tracks the measured and forecast NO_3^- value showed that the ANN is a superior model to manage groundwater resources in the a simpler manner. Wagh et al. (2016) applied ANN in order to predict values of %Na, MAR, SAR, KR, and RSC in the ground water of Nanded Tehsil. The 50 ground-water according to various samples were studied to various physicochemical variables, for example, pH, TDS, EC, Ca^{2+} , Mg^{2+} , HCO_3^- , Na, Cl^- , CO_3 , NO_3 , and SO_4^{2-} during the pre monsoon season in 2012. These parameters were used as input variables in the ANN-based GWQ indices. The results confirm that the ANN model is successfully applied to predict ground-water suitability. Nasr and Zahran (2014) created an ANN with feedforward BP to forecast groundwater salinity in Egypt, which is expressed through TDS, using pH as input parameters. The network output was satisfactory, and simulation can be used for entering new inputs. Khan et al. (2021) forecast the amount of waterborne bacteria in GWQ via forecasting the presence of *E.coli* applying chemical, physical, and microbiological parameters of groundwater. SLA was proposed to examine the pattern of ANN based sensitivity analyses to be determine the significance of all GWQ parameters resulting in prediction of *E.coli* in ground-water. A higher correlation is seen between *E.coli* and a lower pH values, although the lower correlation is found between DO and pH values. The model with turbidity, pH, TDS, and EC as inputs performed better, according to the data. It could be inferred that Grover's algorithm-based superposition models are more effective at predicting each pattern of *E.coli* count in ground-water. Fadipe et al. (2021) assessed the GWQ by predicting the water quality parameters in Boluwaduro community, Ofatedo in Osun State. An ANN model was employed using water samples collected from 18 randomly selected dug wells and subjected to physical, chemical, and microbiological analyses. The ANN design and were trained in some rounds until acceptable outputs was attained with R^2 out of 0.97. The created model for TDS provided a good prediction with TH and Mg^{2+} respectively. Honget al. (2001) applied the KSOFM neural network model to analyze the impact of stormwater infiltration on GWQ. It was concluded that the KSOFM model provided useful analyzing

and diagnosing tools to comprehend a dynamic in GWQ and to extract information included in the multidimensional dataset. It is considered as to have potential not only in GWQ monitoring and diagnosis, but also in further environmental fields. Kassem et al. (2021) used the ANN method with different input combinations of GWQ variables to identify the most important variables impacting of the Cl^- concentration prediction values. Seventeen ANN models were developed through altering inputs variables. Results revealed that ANN 5 model with combination (AVR, ICC, A, LT, GWL, RR, DSWS, W, AT, DSSL) delivered outstanding estimation of prediction values of final Cl^- concentrations. The ANN method illustrated how ANN developing method could be utilized to identifying key variables needed for the most important parameters influencing Cl^- concentrations. Nakagawa et al. (2017) examined the ground water chemistry traits at Shimabara via used of SOM, inputs to this model were the eight main ground water chemical elements (Cl^- , NO_3^- , SO_4^{2-} , HCO_3^- , Na^+ , K^+ , Ca^{2+} and Mg^{2+}). Results indicated that based on chemistry, surface water and groundwater can be classified to five major clusters demonstrating distinctive patterns. Additionally, the five clusters can also be divided to two main water categories, i.e., nitrate-polluted and nonpolluted water. According to Stiff and Piper trilinear diagrams, nitrate polluted water was the $\text{Ca}-(\text{SO}_4 + \text{NO}_3)$ kind, whereas nonpolluted water was classified as $\text{Ca}-\text{HCO}_3$ kind. Wang et al. (2006) developed a BPNN model to mimic spatial distribution of $\text{NO}_3\text{-N}$ concentration in groundwater in the North China Plain by land-use data and site specific hydro geological properties. The GIS tool was used to prepare and process input and output vectors information for BPNN. The outcome indicated that GIS-based BPNN mimic ground water $\text{NO}_3\text{-N}$ concentrations effectively and capture the overall trend of groundwaters NO_3^- pollution patterns. Gemitzi et al. (2010) applied ANN methods combined with GIS for spatial prediction of NO_3^- pollution in groundwater. These involve hydraulic conductivity and depth to aquifer, land uses, fine to coarse grain ratio in the unsaturated area and soil permeability. The method was employed at the South Rhodope Thrace, Greece aquifer. The findings indicated that this method is able to illustrate a spatial pattern of NO_3^- contamination. Alizamir and Sobhanardakani (2016) built the ANN model to predict As, Pb, and Zn concentration in ground water. The Levenberg-Marquardt (LM) algorithm and Bayesian

regularization (BR) algorithm were used as the ANN training algorithms. The performance of these algorithms was assessed, and it was noticed that LM was a great choice for predicting heavy metals concentrations in the chosen monitoring wells. The outcomes showed that the LM algorithm took lesser time for training of network compared with the BR algorithm. Based on results, it was proved that the ANN showed acceptable range in accuracy in predicting heavy metals concentrations in groundwater resources. Sakizadeh (2015) utilized ANN through BR to predict water quality index (WQI), with regard to the concentrations of 16 GWQ variables gathered from 47 wells. The results indicated the effective predicting of WQI via ANNs via used BR algorithm. Also, the sensitivity analyses was applied to illustrate the significance of all parameters in predict WQI through ANN training and it revealed that parameters such as PO_4 and Fe are the most significant parameters in predicting the water quality index. Moasheri and Tabatabaie (2013) measured the value of NO_3^- groundwater qualitative using the dataset of the Birjand plain, which were gathered from 35 wells and aqueducts every six months from 2008 until 2010. ANN was optimized through genetic algorithm (GA) and provided values between the lab result and actual result, which showed the reliability of this approach for predicting the NO_3^- values. Keskin et al. (2015) predicted groundwater pollution sources applying ANN in Turkey. BP and Bee Algorithm (BA) were employed in ANN modeling and 14 hydro chemical datasets were employed. The better ANN classification was achieved with 80% accuracy using the BA algorithm. Singh and Datta (2004) used the universal function approximation property of a multilayer, feedforward ANN to estimate temporal and spatial changing of unknown contamination sources, for instance, hydraulic conductivity, porosity, and dispersivities, and to deliver an accurate estimate for unknown flows and transport parameters. The model was training on patterns of simulated datasets used by the BP algorithm. The set of source fluxes and temporal changing simulating concentrations measurements constitute patterns of trained. A limiting performance evaluation display that introduced method performs well. Sahoo et al. (2006) used BPNN to predict pesticide concentration in ground-water monitored well. The results of the predictive model generated good agreement with actual data in terms of R and pesticide detection efficiency, also good correlation between actual and

predict “classes” groups. The BPNN was applied in order to rank inputs variables with maximum potential to pollute ground water, comprising two original parameters which are depth to aquifer material and pesticide leaching class. Once those parameters were solitary input parameters to the model, they were not capable of predicting pollution. Though, if they were used with further parameters, the prediction performance effectiveness of the model in terms of R, E, ME, RMSE, and pesticide occurrence groups improved. El Tabach et al. (2007) proposed ANN metamodel was applied to evaluate a risk of pollution to attain the ground-water resource underneath the road axis of a high-way project in north of France. Barzegar and Moghaddam (2016) investigated and compared the accuracy of various neural calculating methods, MLP, RBFNN, and GRNN, in predicting of ground-water salinity which is expresses by EC. The performance criteria of the built neural networks model revealed that RBF-NN model has the most excellent performance in predicting EC. Sirat (2012) employed an ANN model in the USA in Illinois, Iowa, and 12 other states for predicting pollution of groundwaters with pesticides. The findings of many trials have shown that this model has predicted the pollution well with-in every minor group within high accuracy. It proved that ANNs models are effective tools for predicting groundwater pollution. (Dar et al. 2012) ANN model was used to comprehend a correlation and sensitivity of every chemical parameter with regard to fluoride. The prediction results indicated that four parameters pH, Cl^- , SO_4^{2-} and Ca^{2+} were able to better influence fluorides compared with the other eight parameters. Also, the Cl^- ions were noted to be the most sensitive parameter and most correlating to fluoride. Kadam et al. (2019) applied ANN and MLR methods for predicting the fitness of GWQ for drinking. The physicochemical parameters were taken in account to compute WQI with the view of producing reliable and accurate modeling to predict WQI-based GWQ. The LM with 3-layer BP algorithm was applied in the ANN structure. Additionally, the MLR model was used to test the effectiveness of the prediction model. The result verified that prediction of ANN is acceptable and confirmed regular satisfactory performance for each season. Moasheri et al. (2013) applied the fusion of ANN and geostatistical to assess more parameters of spatial distribution for GWQ is the TDS plain Birjand more precisely. The analyses of the geostatistical interpolation method and application of

ANN for optimization results of the geostatistical method have been examined. Ehteshami et al. (2016) employed BPNN and RBFNN and the best architecture for the BP model was noted to 4-5-1 and RBF with a spread parameter = 0.5 and MSE = 0.50. Result displayed insignificant variation between architectures. Both ANN models could be consistently predict nitrate contamination in ground-water with satisfactory precision.

2.1.2 Results of Artificial Neural Networks Models

The assessment of many papers on ANN modeling for water quality in groundwater showed the follows points:

- (1) The ANN methods could be expanded simply from univariate to multivariate instances compared to conceptual methods, and model complexity could easily vary through adjusting a transfer functions, learning algorithms, or networks architectures. Similar to any regression model, inputs would depend on experimental proofs or correlational analyses. Also, conclusions of review studies showed that ANN captured the complexity of nonlinear behaviors for GWQ reasonably more accurately than regular regression methods such as LR or MLR.
- (2) The most popular structure of ANN for GWQ modeling is using three or two layers with sigmoid transfer function in the hidden layer and linear transfer function in the output layer. This function is more stable, differentiable, and monotonically growing in its field and is an often used function in predicting GWQ. It should be noted that in most reviewed studies, the ANN architecture and number of hidden neurons were attained through the trial-and-error method.
- (3) The review show that the Levenberg-Marquardt(LM) and Bee Algorithm (BA) algorithms are good training algorithms used to train ANN for GWQ modeling compared with other ANN algorithms. Keskin et al. (2015) compared the performance of BP and BA utilized in ANN training. The best ANN classification was achieved by the BA. The LM algorithm is the alteration of the classical Newton algorithm utilized to find the optimal result to minimizing the problems. The LM algorithm is frequently described as continuous and efficient, and many investigators indicate that it is quicker and less

simply trapping in local minimum than further learning algorithms. Alizamir and Sobhanardakani (2016) compared LM and Bayesian regularization (BR) algorithms in terms of how they perform in GWQ, and it was found that LM was a good choice for forecasting heavy metals concentration in the selected monitoring wells. The simulation results indicated that the LM algorithm took less time for training of the network compared to the BR algorithm.

2.2 Fuzzy Logic- Based Modeling for GWQ Modeling

The integration of the fuzzy inference system (FIS) and the adaptive neural networks (AN) produced model adaptive neuro fuzzy inference systems (ANFIS). These models have the ability for capturing the benefit of both approaches in a single structure. Jang (1993) established the structure and the training process of ANFIS which used the neural network training algorithms to create sets of fuzzy “if and then” rule with proper membership functions (MFs) from specifying input-output. FIS correspond to sets of fuzzy “if and then” rule that have training ability to estimate non-linear function. FIS have two methods called Sugeno and Mamdani. The disparities between such methods occur as the following section explains. Mamdani’s method used fuzzy MFs, while Sugeno’s method utilizes linear/continuous MFs.

ANFIS is the AI approach with adaptable statistical structure that can identify complicated non-linearity and uncertainties because of randomness and inaccuracy among variables, without trying to attain an insight as to the nature of trends. Such method is capable of estimating every actual stable function on the compact sets to every level for precision. Hence, in parameters prediction, wherever provided data in which a system links measurable system variables with the internal system parameters, the function mapped maybe built through ANFIS which approximates the processes for evaluation of an internal system parameters. More details about ANFIS it can be found in supplementary materials and in (Jang 1993). Figure 5 shows ANFIS structure.

2.2.1 Bibliographic Review

Some recent investigations have tried to evaluate GWQ applying FIS. Dahiya et al. (2007) applied FSE to

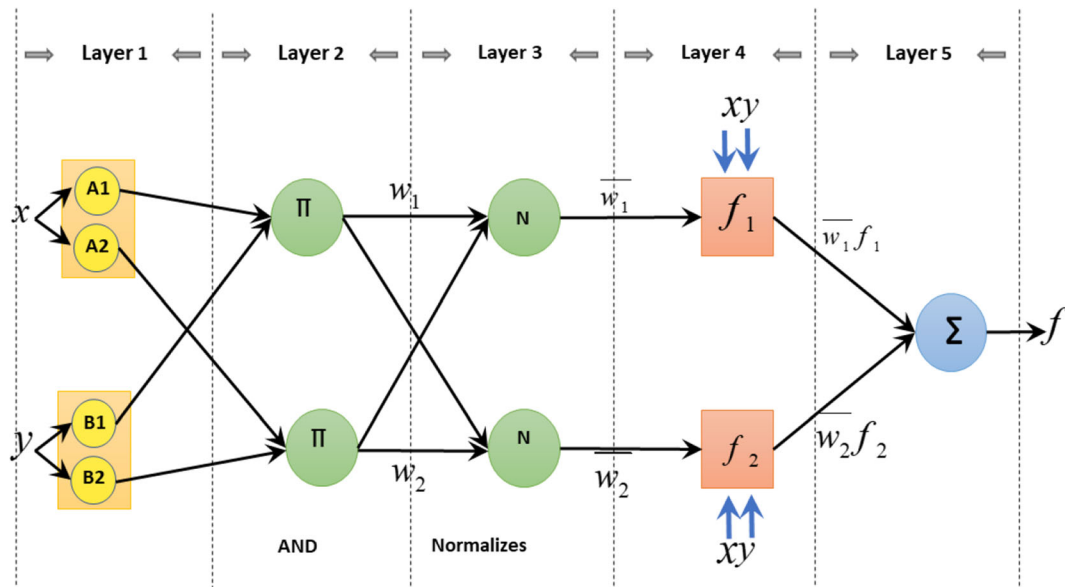


Fig. 5 ANFIS architecture

analyze GWQ. The model was optimized by testing 42 ground water-samples based on ten drinking water quality parameters. Khaki et al. (2015) assessed the ability for using ANFIS and ANN models to assess TDS and the level of EC for GWQ, using the amounts of further present water quality parameters NO_3^- , SiO_2 , Na^+ , pH, Ca^{2+} , Mg^{2+} , K , HCO_3^- , Cl^- , Fe) in Malaysia. The results indicated that both methods have the ability to interpret behavior of water quality parameters in groundwater. (Aryafar et al. 2019) GP was employed to identify relationship between GWQ parameters, using the independent variables (K^+ , Na^+ , pH, Ca^{2+} , Cl^- , CO_3 , HCO_3^- , Mg^{2+} , So_4^{-2} and NO_3^-) in order to predict dependent variables (TH, TDS, EC) for 240 samples of groundwaters gathered from 12 wells in Iran. Also, ANN and ANFIS techniques were developed to further verify the estimation ability of GP through comparing the predicted and actual values for chemical parameters. Findings revealed the superiority of GP acceptable performance was delivered via each model in prediction of intended parameters of water quality. In India (Kumar et al. 2020) applied MLR, SEM, and ANFIS models, through mixing three datasets for various areas and time period in four different methods, various regression approaches designed used the TDS as the output parameter and the input parameters used were (K^+ , SO_4^{-2} , Ca^{2+} , Mg^{2+} , Cl^- , Na^+ , and NO_3^- , TH). Similar regression coefficients were noticed for both SEM and MLR

models, while the ANFIS model demonstrated a better performance than the MLR model. Vadiati et al. (2016) used the MFIS model through commonly accept GWQ indices. The method assessed for it is able to evaluate the drinking water quality for 49 samples gathered in every season from groundwater sources in Iran from 2013 to 2014. Input membership function was described as “desirable”, “acceptable” and “unacceptable” depend on experts’ knowledge and standards and allowable restrictions imposed via WHO. Tutmez et al. (2006) used the ANFIS model of ground-water that EC depend on concentrations for positive charged ions in water. The prediction results indicated that this model outperform over conventional approaches in prediction EC from TDS. Jebastina and Prince Arulraj (2018) investigated NO_3^- pollution in ground water of Coimbatore district of India used 71 observation wells years 2011–2012. The inputs for producing the ANFIS method is average for EC, Na^+ , Ca^{2+} , Cl^- , TH, K and the output was the concentrations of NO_3^- . Outcomes for the best ANFIS model were evaluated via deterministic, geo-statistical and kernel smooth techniques utilizing GIS. Five models developed using the, ANFIS method predicted NO_3^- concentrations. Selvaraj et al. (2020) gathered 30 ground water samples from bore well as dug well source and used chemical parameters to measure FWQI. Five FIS models with various linguistic parameters and five classifications of fuzzy model were

established, for example “excellent”, “good”, “poor”, “very poor” and “not suitable”. Of the findings achieved from such method, six samples were classified to “excellent”, eight samples to “good”, 12 into “poor”, three into “very poor” and one into “not suitable”. Also the result of this method were compared with the output found from the deterministic technique. MFIS was employed for analyses of the GWQ in Nadiad district, India (Prajapati and Parekh 2016). Various groundwater variables such as EC, TDS, PH, Cl and Ca were utilized to analyze GWQ. Different membership functions, “desirable”, “acceptable” and “not acceptable” are used and fuzzy logic rules were defined. The results indicated that MFIS is very useful and effective tool in assessment of GWQ. Dixon (2005) effectively used FL and GIS, GPS, remote sensing to measure ground-water-vulnerability maps. Nadiri et al. (2017) used Larsen fuzzy logic (LFL) in conjunction with Sugeno fuzzy logic (SFL) and Mamdani fuzzy logic (MFL) to assess the ground waters vulnerability in the Varzeqan Plain, in north western Iran. Nikoo and Mahjouri (2013) used PSVM and FIS methods for probabilities water quality zoning which is important to GWQ. The needed data for training PSVM was produced by using FIS in the Monte Carlo analysis. PSVM rained based water quality index give probability of belonging to various water quality classes “very low”, “low”, “medium”, “high” and “very high”. The prediction result was compared with attained results from clustering methods which are FCT and SOM. These outcomes of GWQ zoning were illustrated in maps using using GIS. Nadiri et al. (2013) presented the supervised committee machine with AI method employing the SFL, MFL, ANN, and neuro fuzzy (NF) to predict fluorides concentrations. The findings demonstrated that all of those models have the same fitting to fluorides dataset in the study area, and do not predicting fine for samples in mixing zones.

2.2.2 Results of Fuzzy Logic Models

The assessment for many papers on fuzzy based prediction for water quality in groundwater displays the following issues:

- (1) Of the investigated studies, those using a fuzzy-based model ranked second, while ANFIS ranked highest because it provides the benefits of both ANN and FIS models. ANFIS could managing

non-linearity, uncertainties and fuzziness relate to water quality of groundwater. ANFIS has some benefits in signal identifications, converse coding and noise cancelation. ANFIS has improved result when datasets are normalize (Ahmed et al. 2017).

- (2) ANFIS model has demonstrated superior performance in managing data of water quality in groundwater as it has two categories of parameters (linear and nonlinear) that could be adapted appropriately during training phase (Jang 1993). Beside member function which was illustrates the vital role and right selection lead to finest optimizing for parameters, thus result with highest precision (Jang 1996), an additional vital reason to design the excellent ANFIS model is identifications technique used to split inputs spaces. Common methods employed are subtractive cluster, grid partition and fuzzy c-mean cluster by different functions (Jang et al. 1997). Consequently, this model could do well through nonlinear data alongside with appropriate selection groundwater parameters such as (EC, TDS, PH, Cl, Ca, Mg, NO_3^- ..., etc.) decided via the designer.
- (3) Some models such as MFIS and FWQI were presented for water quality in groundwater modeling. They performed well in managing GWQ. The fuzzy-based model was capable of delivering the highest accuracy of result, and simply incorporating previous knowledge to the model language naturally. This permits parameters to estimate the algorithms in order to give the practical primary values that reduces the opportunity to find needless local minimum. An Additional benefit is the capability to extrapolate that reveals new information revealed from data (Lindskog 1997). The fuzzy rule based model describes input and output relationships more pragmatically and it has physical interpretations that have allowed scholars in GWQ research to select FIS instead of ANN models (Dahiya et al. 2007; Esmailbeiki et al. 2020).

2.3 Support Vector Machine (SVM) for GWQ Modeling

The idea of not being able to separate dataset in support vector networks was established in 1995. This kind of machine training used non-linear inputs mapped to

highest dimension space then ensure higher generalizing. This was model based on the named large margin component and developed according to structure risks minimizing theory. Support vector machine could produce good accuracy compared with KNN, linear and NN classifiers because it does not involve knowledge. The model includes regression model complexity, regularization, and kernels functions (Cortes and Vapnik 1995). Hence, it's called the kernel model. Kernels function and parameters must be carefully chosen to decrease the bound on Vapnik dimensions (Hadian et al. 2021).

The mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem. The extended support vector machine model can solve a function assessment problem via using loss function Vapnik epsilon (ϵ) insensitive and Huber loss function as displayed in Fig. 6 through overall structure for support vector machine (Chen and Pao-Shan 2007; Yaseen et al. 2019).

This model could be design via linear, spline, polynomial, and RBF networks (Suykens and Vandewalle 1999). SVM is commonly used in order to develop the classification models of predicting trends and also solve regression and time series challenges (Raghavendra et al. 2014). Numerous models have been built utilizing kernel architectures as shown in Table 2.

2.3.1 Bibliographic Review

In the field of water quality for groundwater prediction with SVM, (Jafari et al. 2019) TDS of the ground water aquifer at the Tabriz plain was assessed with groundwaters physicochemical parameters such as (Na^+ , Ca^{2+} , HCO_3^- , Mg^{2+} , SO_4^{2-}) using MLP, ANFIS, SVM, and GEP for a period of 10 years from 2002 to 2012. According to the outcomes, GEP, MLP, SVM, and ANFIS methods would be effectively used in predicting TDS alteration and the GEP provided better results compared to the other methods. Alagha et al. (2014) built SVM and ANN models to predict groundwater NO_3^- concentrations applying scant input datasets, the samples used data of 22 municipal wells in Palestine (2000–2010). The conclusions indicated that the developing AI approaches could be effectively used to evaluate impacts of future management scenario

on ground water NO_3^- concentrations, leading to higher acceptable groundwaters sources management and decision making. Gholami et al. (2011) applied the SVM model to forecast concentrations for Ni and Fe, then it was compared with BPNN, and the results indicated that this model could be considered as a suitable algorithm to predict Fe and Ni concentrations. Actually the SVM technique has delivered the best forecast for toxic metals, resulting faster operating time comparing with the ANN model. Dixon (2009) evaluated the ANN and SVM integrated into GIS to identify contaminated wells using 14 GIS derived land use parameter and soil hydrogeologic parameters as preliminary input. Good water quality dataset $\text{NO}_3\text{-N}$ from 6917 wells in USA were applied as an outputs goal class. Findings showed good performance by NN when compared with support vector machine. Arabgol et al. (2016) used the SVM model prediction NO_3 concentrations in groundwaters in Iran. SVM presented as a promising tool to predict NO_3 concentrations utilizing the sets of simply measurable GWQ parameters as input (T, ground water depth, TDS, land use, DO, pH, EC and season in year). The mapping for NO_3 concentrations in ground-water were arranged for all seasons applying training SVM with GIS interpolation scheme. The results indicated that this method can be utilized as a faster, reliable, and cost-effective technique to measure and predict the GWQ. Isazadeh et al. (2017) used SVM and ANN models with various input designs, and the findings confirmed that each model showed good performance when using input well location. Also, the results showed that SVM and ANN generated better accuracy in estimation of three qualitative parameters. Moreover, the significance of auxiliary input variable was determined using Gamma testing. Purkait et al. (2008) used ASVR, ANN and MLR to prediction As contaminations in ground water. Among those models the ANN model that used the BP technique demonstrated more accurate result in predicting As contamination. Also, it is important to note that the architecture of the ANN model used here with 4-layer feedforward BP can be utilized as a satisfactory prediction model for assessing As contamination in groundwater.

2.3.2 Results of Support Vector Machines

- (1) The SVM or SVR are effective ML approaches that have been built and employed for various

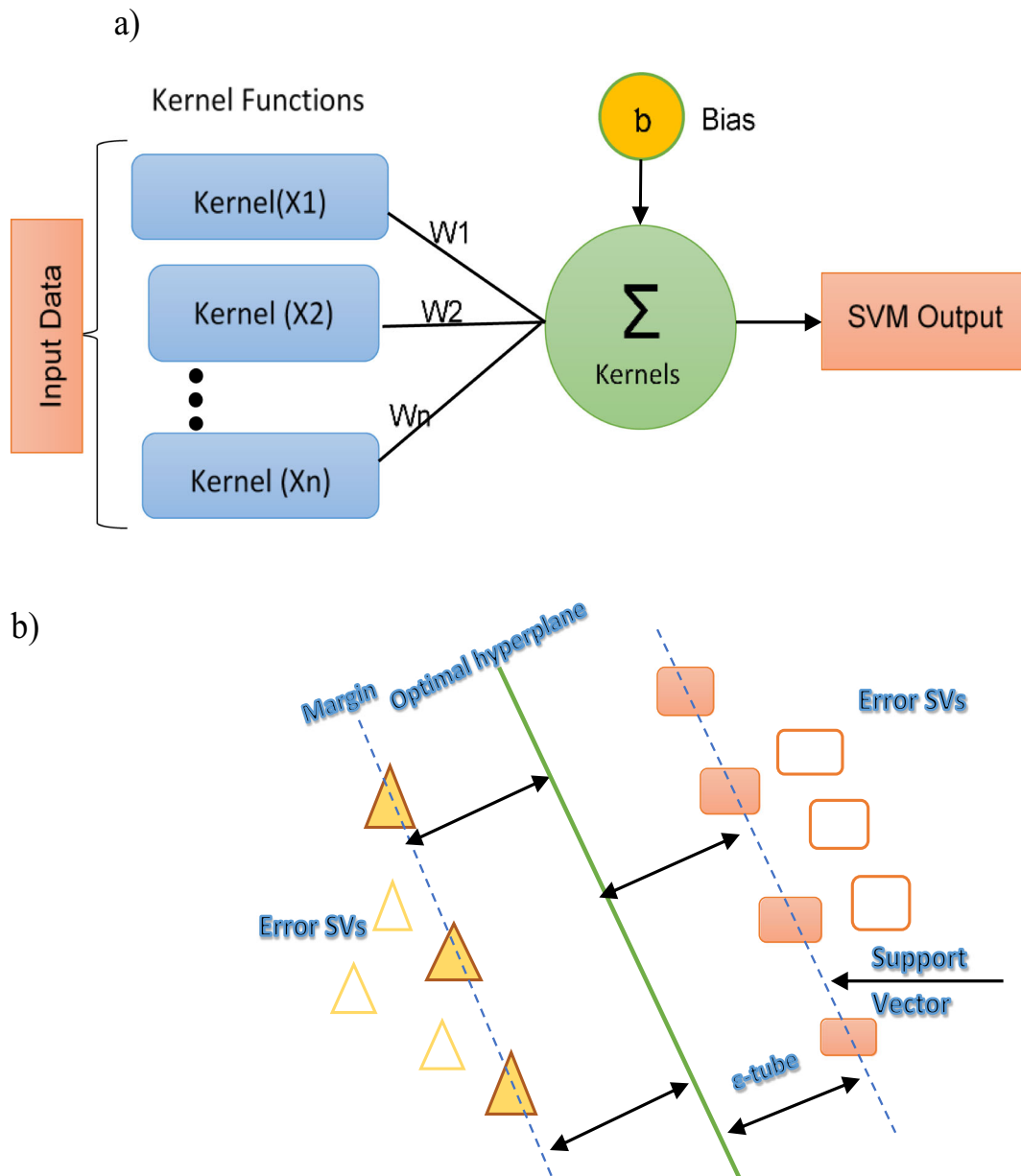


Fig. 6 (a) Typical SVM architecture for groundwater quality modeling, (b) Nonlinear SVM Vapnik ϵ insensitivity loss functions, when predict values are with in tube loss is 0 and outside loss is scale for variation between predicted value and radius of ϵ tube

classification and regression challenges throughout the past two decades. Even though a low number of published studies considered the prediction of GWQ by SVM compared with ANN models, it was applied to predict various time series for the myriad of practical applications in this realm.

- (2) In predicting SVM, properly selecting kernel functions and the values of parameters is very crucial.

In many above-mentioned papers such as (Gholami et al. 2011; Dixon 2009; Arabgol et al. 2016) RBF has been used as the best kernel function. While Sigmoid and RBF were used as the kernel function in (Alagha et al. 2014). However, in some papers, the kernel function used was not revealed. Many researchers have chosen the RBF function as the kernel function for SVR over the

years due to its accuracy and reliability in performance.

- (3) The majority of papers used a trial-and-error method to determine the optimal parameters of the SVM model, with the exception of one analysis (Jafari et al. 2019), which did not mention a tool to determine the optimal parameters of the SVM model.

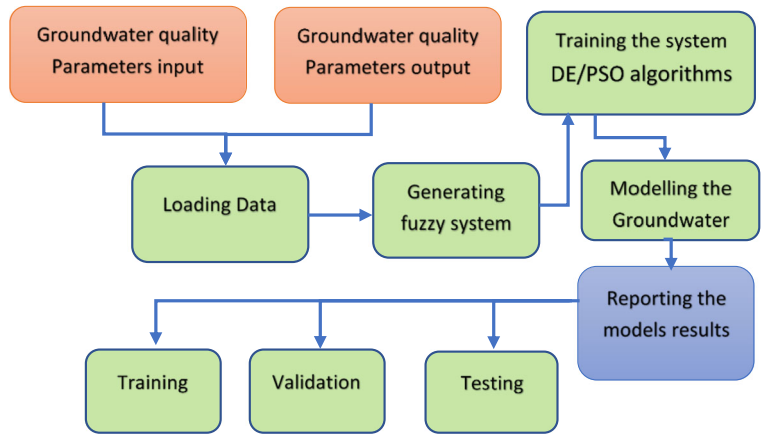
2.4 Hybrid AI Models for GWQ Modeling

Since it was discovered that artificial intelligence methods have some limitations by non-linear process and nonstationary process, many hybrid prediction techniques that comprise data pre-processing/combined various artificial intelligence models to increase abilities for artificial intelligence models. Combined various AI models of various phases of modeling and employing effective approaches for input data preprocessing made the developed of these models more effective. Owing to the ability of geostatistical tools in spatial assessment, hybrid artificial intelligence geostatistical methods were employed in several studies to utilize their capability in mimicking spatiotemporal for GWQ. GA was initially presented during 1962, then significantly enhanced as time progressed. It works on selecting, crosses over, and mutating operators (Goldberg and Holland 1988) of the binary string used for codes; therefore, it used actual variables vector as chromosome, actual variable as gene and actual number as allele. Vectors were decoded and underwent fitting assessment, that help decide if a vector should undergo mutation or crossover. GP models, which are combined GA and GP, it is depend on Darwinian concept for evolution theory introduced by (Koza et al. 1997). It's an independent, organized and domain approach that instinctively solves problems and does require the designer to have in-depth information for solutions structures. The functions of GP are an optimizer and classifiers. Such technique included syntax trees with branch as function with argument and leave of the tree as variables and continual. Many algorithms are dependent on behaviors for the organism, such as particle swarm optimization (PSO) particle swarm optimization uses principle of bird flocks, fish teaching and swarming theory. It depends on the simple idea by least coded. Its benefits over GA consist of improved performance because of groups interactions and memory retention (Eberhat and Kennedy 1995). Figure 7 and 8 show the structure of the hybrid model.

2.4.1 Bibliographic Review

Some researchers attempted to integrated various types of AI models to overcome weaknesses and then enhance their precision. Jha et al. (2020) used a novel hybrid structure integrated with the FL and GIS based GQI to measure GWQ in India using FGQI models (with seven critical variables) for two seasons. The model inputs parameters done by Trapezoidal membership function based on water quality standards for drinking purposes and expert knowledge were classified (desirable, acceptable, and unacceptable). The outcomes were compared with other models and the FGQI model predicted GWQ better than the traditional methods and was more reliable and realistic for GWQ valuation and analyses at large scales. Maroufpoor et al. 2020 applied hybrid models based on neuro fuzzy systems combined with fuzzy c-mean data cluster (FCM) and grid partition (GP) models and ANN combined with PSO algorithm to forecast a spatial distribution for parameters EC, SAR and Cl of ground water. The findings of these methods were compared with the geostatistical method which include kriging, inverse distance weighted (IDW), and RBF. Outcomes indicated that the hybrid model NF-GP gave lower values of both RMSE, and MAE and a higher value of R was found; it was a most appropriate model for predicting the water quality parameters. Aguilera et al. (2013) showed that applying a hybrid model of BNs with MTEs provided a proper result for predicting GWQ. The outcomes attained allow the difference of three classes of sampling, representing three various groups of GWQ and the possibility that the sampling point belongs to all clusters permits the uncertainties in clusters to be measured, and a risk related in term of water quality management. Jalalkamali (2015) used a variety of hybrid approaches, including ANFIS-GA and ANFIS-PSO to model three GWQ parameters in the Kerman plain involving Cl concentration, EC, and PH, assuming different combinations of monthly variables of rainfall and ground water level, as well as three various quality parameters. The findings showed that both models performed well in the spatiotemporal simulation of GWQ. The paper also showed that groundwater variations around the aquifer, and runoff, play important roles in forecasting GWQ. Kisi et al. (2017) employed the ANN-PSO model and the ANN-DE model to predict parameters of ground water qualitative such as SAR and SO_4^{-2} . The prediction results revealed that ANN-DE provided better accuracy

Fig. 7 Steps of hybrid model procedure (Kisi et al. 2019)



comparing with ANN-PSO in predicting these groundwater parameters. Cho et al. (2011) investigated the estimation of prediction performance of various models such as MLR, ANN, PCR and grouping of principal components and ANN (PC-ANN) in predicting of As concentrations and provided evaluation tool for Thailand, Laos, and Cambodia. Prediction precision of PC-ANN shows better performance than other models. Alizamir and Sobhanardakani (2018) used the ANN-PSO approach to predict heavy metals such as (Cu, Zn,

Pb, and As) contamination in ground water sources in Iran. The results showed that this hybrid model could be successfully used for predicting concentrations of heavy metals in ground water resources at the Toyserkan Plain. In (Kisi et al. 2019) the GWQ variables and quality parameters were modeled through simple ANFIS and ANFIS training via evolution algorithm including PSO, DE, CGA and ACOR, and the results of modeling shown that four proposed algorithms improved ANFIS performance in predicting the (TH and EC), and in

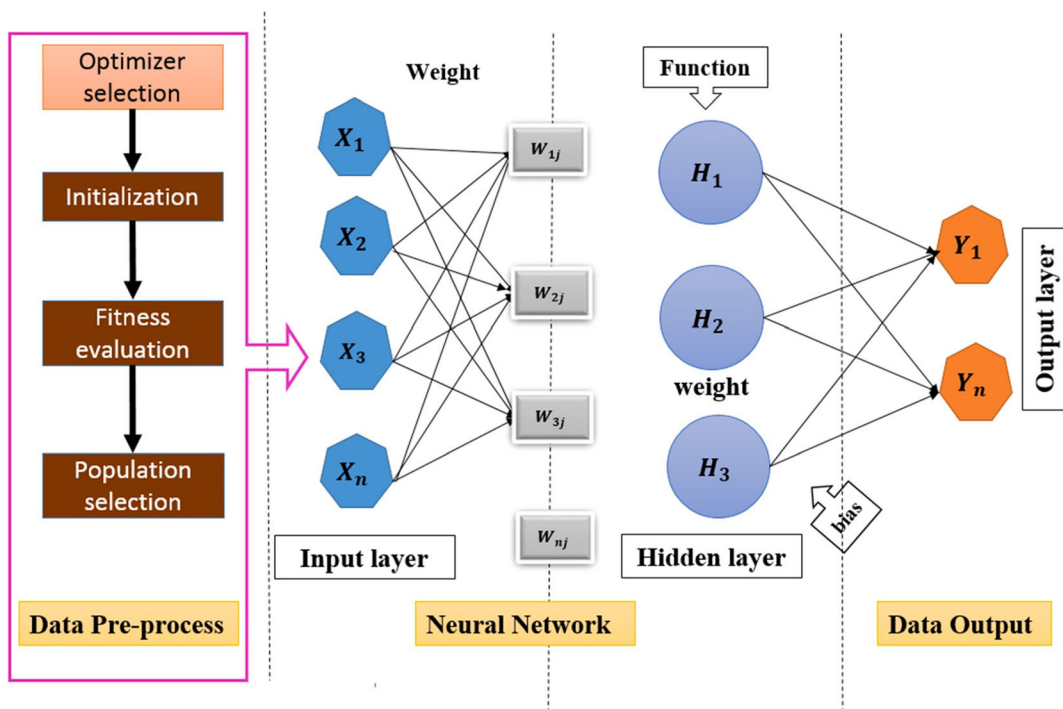


Fig. 8 General structure of optimizers by ANN (hybrid-model) where X = inputs, Y = output, bias is justified via the training process, H = number of hidden layer and w = weight (Tiyasha, Tung, and Yaseen 2020)

algorithms (SAR, PSO, CGA) have the best performance compared with other algorithms. Fijani et al. (2013) coupled SFL, MFL, ANNs, and NF methods to estimate groundwater vulnerability in the Maragheh-Bonab basin of Iran. (Soltani Mohammadi et al. 2017) predicted GWQ parameters such as SAR, EC, and TDS using ANN and ANN- PSO models and compared their outcomes with observed data. The inputs to predict parameter TDS included of EC, SAR, pH, SO_4^{-2} , Na^+ , Ca^{2+} , Mg^{2+} , but for SAR, it included HCO_3^- , pH, Na^+ , and TDS, and as for EC, it involved SO_4^{-2} , pH, Ca^{2+} , SAR, and Mg^{2+} . The results reveal that the highest predicting accuracy for SAR, EC, and TDS was related to model ANN-PSO by using the tangent sigmoid activation function. Bashi-Azghadi et al. (2010) studied a new method to estimate the location and total of leakage from unknown contamination source using data of GWQ. They developed probability prediction models, PSVMs and PNNs for identifying and evaluating core attributes of an unknown ground water contamination source, in term of location, timing and magnitude. In their proposed method the GWQ prediction model derived-out of MODFLOW and MT3D were coupled with a multi objective optimization model NSGA-II. The comparison of accuracy of both models found that the PSVM model was better than PNN. The results revealed in real time monitoring groundwater that the probabilistic mass functions of unknown contamination source location and relative errors in estimation values of leakage depending on the actual concentration of water quality indicators at monitoring wells were better evaluated via training probability in both PSVMs and PNNs models.

2.4.2 Results of Hybrid AI Models

(1) Many papers have reported using PSO in resolving practical optimizing problems in the world such as (Liu et al. 2007; Melin et al. 2013; Selakov et al. 2014; Soltani Mohammadi et al. 2017). PSO could be more efficiently applied to forecast a spatial distribution of parameters of groundwaters due to it has small computational volume, simple operation, is not reliant on problems, high convergence rate, worldwide search ability, able to deal with complex problems and could escape local minimum (Maroufpoor et al. 2020; Jalalkamali 2015). However, in (Kisi et al. 2017) DE algorithm showed better results than PSO which is a

population based stochastic searching method to solve continuing optimizing problems.

- (2) GA is a population based stochastic optimizing tool that could successfully deal with nonlinear and nondifferential water quality of groundwater data problems through attaining optimal solutions via the repetition procedure. The operators could rise to other spaces with look space to discover best new way out. It's a not thorough method till required solution is attained (Jalalkamali 2015). These benefits let them to solving high complex and multiparameter challenges.
- (3) GP delivers stable and effective result (Maroufpoor et al. 2020); although, limited investigation have considered GP model. It was shown good performance once datasets are preprocessed. The explicit and linear representation of this model relies on ease, effectiveness, and genetic operator. Further investigation is required to best utilization of GP model consider they would be simply used, develop through training and solving the complex challenges in unpredicted manners.
- (4) Another combined model was integrated such as BNs with MTEs and Fuzzy Logic with the GIS-based GQI showed good outcomes for prediction GWQ. By merge ANN with PCA in (Cho et al. 2011) a new set of variables and the PC scoring are produced from perpendicular linear transformation of previous data, then these scores utilized in prediction as helpful variables.
- (5) The majority of research papers used one or combinations of multi-dimensional data analyses approach in ANFIS, i.e., GP, PSO, CGA, DE and ACOR cluster. Moreover, CGA and PSO are the superior tools compared with others in handled the classification of GWQ (Kisi et al. 2019).
- (6) Even though many studies have applied used hybrid models in hydrology, most studies conducted for modeling groundwater level (Fahimi et al. 2017; Tiyasha and Yaseen 2020; Rajaei et al. 2019; Tiyasha and Yaseen 2020; Moosavi et al. 2012)

2.5 Machine Learning Models for Modeling Water Quality of Groundwater

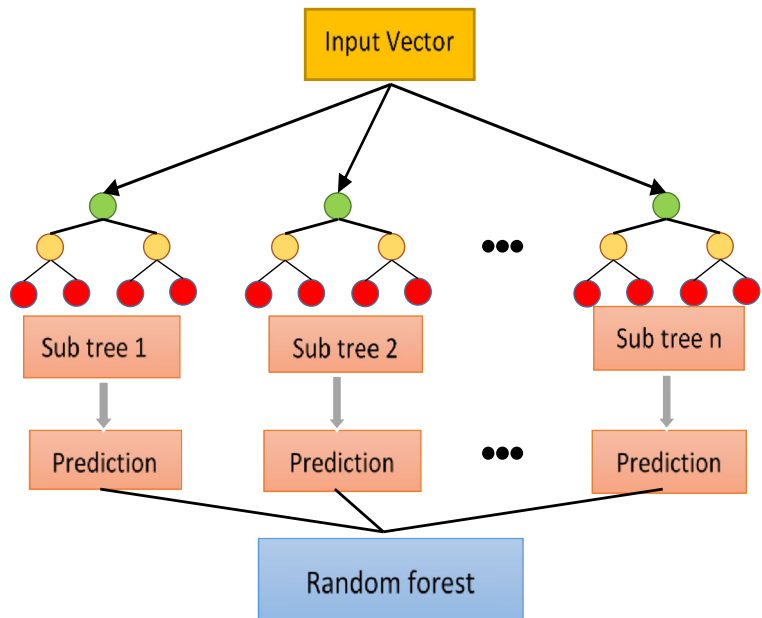
Machine learning (ML) methods for instance, XGBoost, M5P, MARS, GEP and RF are the effective models to

deliver the non-linear relation between input-output parameters. XGBoost approach produces several shallow decision tree, combinations for whole trees generates the highest precision regulation through forecasting (Chen 2016). A decision tree was built via XGBoost algorithms and did not just proceeded to minimize the aim function through consideration of loss function but also protect trees from overfit during regularization procedure. Additionally, flexibility in tuning hyperparameter for XGBoost model make such model very common in various exploration disciplines (Bhagat et al. 2020). The M5P used a binary decision tree (DT) and multi-linear regression to forecast continuous mathematical characteristics. The initial stage involves dividing criteria to produce the decision tree. The second stage involves pruning technique to remove overfitting and building of the linear regression functions (Etemad-Shahidi and Bonakdar 2009). MARS is a regression model built in 1990 (Friedman 1991) which is the combination of regression and recursive partitioning. The method is categorized as a forward and backward subclass selecting process. This model involves product degree, dependent parameters, intercept, basic function and knots. The GEP model is based on the same idea as GP and GA, but it have characteristics of model (Danandeh Mehr et al. 2018). In particular, GEP uses fit testing, genetic operators and individual population (Esau et al. 2019). The RF another model using regression method with combinations of trees predictor. RF is random chosen segment for training data set of all trees (Breiman 2001). The prediction process of RF approach is displayed in Fig. 9. More details about applied other machine learning models can be found in supplementary material. The applications of machine learning models for prediction water quality in groundwater is demonstrated as following sub section.

2.5.1 Bibliographic Review

Many investigations have been done in the field of water quality for groundwater using machine learning methods. El Bilali et al. (2021) applied Adaboost, ANN, RF, and SVR methods with 520 samples of dataset relate to 14 ground water quality variables in Berrechid aquifer, Morocco. They were applied EC, T, PH as inputs to prediction PS, SA, MAR, R, RSC, ESP, TDS as outputs. Their calculations showed that the general forecast performance of Adaboost and RF methods were highest than SVR and ANN. Moreover,

generalization capability and sensitive to input analysis showing that ANN and SVR methods have better ability in generalizing and lower sensitivity to input parameters than RF and Adaboost. Lopez et al. (2021) predicted groundwaters uranium concentration via RF regression, using 23 environmental variables from statewide ground water geochemical data base and in publicly accessible maps of soil and aquifer physico-chemical properties in California. They noticed that ground water of concentrations Ca, NO₃, and So₄⁻², soil pH, and clay contents (weighting average from zero- and two-meters depth) were the most significant predictors of groundwaters uranium concentration. Singha et al. (2021) introduced a DL based model for predicting GWQ and compared with other machine learning models, ANN, XGBoost, and RF. A total of 226 ground water samples in India were used. Outcomes demonstrated that the DL method provided better predicting result by give higher precision compared to other ML models, input variable importance calculated through forecast approaches highlighted that DL technique is the most pragmatic and precise model in predicting the GWQ. Bedi et al. (2020) examined performance of leading ML classifiers in order to forecast the NO₃ and pesticide concentrations for ground water wells, used land use, hydrogeologic and water quality data. Each scenario was able to forecast both the NO₃ and pesticide concentration with accuracy equivalent to earlier attempts. The XGB generated the best prediction among other methods. Therefore, this model is able to provide a good alternative to other water quality prediction methods. Rodriguez-Galiano et al. (2018) applied a comprehensive GIS data base of 20 variables involving hydrogeological and hydrological characteristics as input for models to predict NO₃. ML algorithms for example CART, RF and SVM were applied as wrappers considered 4 various subsequent search methods and most features that gained significance from RF and CART models were utilized as the embedding method. RF provided the better performance with mmce = 0.12 and AUC = 0.92. Rodriguez-Galiano et al. (2014) implemented RF to predict the NO₃ contamination. A GIS database including 24 variables associated with essential hydrogeologic proprieties remotely sensed driving forces parameters and physical chemical parameters measured and used as input to design various predictive models of NO₃. Also, RF was compared with LR method used various efficacy measures to guarantee their generalization capability.

Fig. 9 Schematic view of the RF model

The outcomes showed the capability of Random forest to develop accurate model with robust regression abilities. Vijay and Kamaraj (2019) assessed the water quality parameters such as TDS, EC, PH, So_4^{-2} , Cl, NO_3 , CO_3^{-} , HCO_3^{-} , metals ions, trace components. There are two main classifications high and low levels of water pollution noted in Vellore district. This work emphasizes on predicting water quality by applying ML classifiers algorithms C5.0, Naïve Bayes and RF as learner for water quality forecast with good accuracy and efficiency. Rahmati et al. (2019) chose three state of the art machine learning methods including kNN, SVM, and RF were chosen to spatially model ground water NO_3^{-} concentration. The methods were calibrated with NO_3 concentration from 80 wells which equal 70 % of the dataset and later validating with NO_3 concentration from 34 wells which equal 30% of dataset. These methods were compared with uncertainties methods and these results highlighted that kNN model was the better model. Singha et al. (2020) measured concentrations for several heavy metals (HMs) and compute six groundwaters contamination indices. This research work also studied the performance of DL based regression model by comparison research. The findings shown that the DL-PMI scoring lower error than DL based ground water HM contamination model. Comparing of both models DL model provided better performance. Barzegar et al. (2017) applied ELM, MLP, and SVM methods to forecast the level of F^{-} pollution in the

ground-water. The samples were evaluated for EC, pH, main chemical ions, and F^{-} . To develop these approaches, dataset of (Na^+ , K^+ , Ca^{2+} and HCO_3^{-}) concentration was utilized as input and F^{-} concentrations as outputs. The result presented that ELM models performed better than MLP and SVM models. Stackelberg et al. (2021) established a BRT model for predicting pH condition in the US. The prediction result revealed that Ca^{2+} contents of soil and aquifer material robustly control pH once combined with long flow-paths was most alkaline condition. A novel aspect of such a model was the presence of mathematically based estimates of groundwater flow properties (age and flow-path length) as predictor parameters. Knierim et al. (2020) employed BRT models to predict SC and Cl, and TDS were computed from the correlation with SC. The descriptive parameters for BRT model comprised fit location and construction, surficial parameters, and parameters extracted from the ground water flow model, including predicted groundwater ages, and the BRT model was able to capture zones of notarized higher salinity that exceeded the TDS minor extreme pollutant levels for drinking water of 500 mg per liter. Parameters that worked as surrogate for place along groundwaters flow-paths were essential predictors, indicating that much of the controls on dissolving solid is related to rock water collaboration as residence period increases. Khalil et al. (2005) studied various models to predict NO_3 concentrations in ground

water such as ANN, SVM, LWPR, and RVM. The prediction result showed the capability of machine learning to develop an accurate model with robust predictive ability, therefore constituting a useful means for saving efforts in groundwaters pollution modeling and developing model performance. Esmailbeiki et al. (2020) data-intelligent models used data collected from 367 wells in India to chose scenarios to predict EC and TH, which are GMDH, DENFIS, GEP, MARS, and M5 Tree. Outcomes revealed that GEP for input combinations involving TH, Na⁺, Cl⁻, HCO₃⁻ parameters and DENFIS using Ca, Mg, EC parameters provided a better performance in predicting TH and EC, respectively. Maiti et al. (2013) proposed an ANN based method set in the BNN structure and employed it to predict GWQ based on the geo-chemical and geo-physical datasets gathered from the western part of India. This study showed that the trained model could classify validation and test sets with 85% and 80% accuracy, respectively. Based on cross-correlation evaluation and Gibb's diagram of geochemical characteristics, a groundwaters attributes of the study area were classified to sets: "very good", "good", and "unsuitable". BNN model based result indicate that most values of GWQ fall in the range of "good" to "very good" excepting for a few places near the Arabian Sea. Bui et al. (2020) used a novel data mining algorithm that included the GP in the large GWQ data to predict NO₃⁻ pollutant and strontium possible future increase concentrations groundwaters. The prediction result was comparing with RF, M5P, and RT models, it was indicated that GP algorithm outperformed over other algorithms used water quality parameters (EC, T, HCO₃⁻, pH, F⁻, Cl⁻, K⁺, Ca²⁺, Na⁺, Mg²⁺, and SO₄²⁻) as the inputs to the models. Yang et al. (2017) measured the GWQ via MLP-ANN optimizing by BP algorithm and WNN model in the coastal aquifer, South China applying eight parameters, e.g., TDS, TH, COD, Cl, So₄²⁻, NO₃⁻, and F. The findings attained from these approaches demonstrated that WNN has the highest accuracy comparing with ANN. Norouzi and Moghaddam (2020) applied the fuzzy groundwater quality index (GQI) to reduce the uncertainty via the fuzzification of the GQI technique. Moreover, RF algorithm, as the learning approach based on the ensemble of decision tree, was applied to assess the GWQ. For drinking purposes, validating and comparing the methods showed that the method fuzzy groundwater quality index has high accuracy as the more reliable technique in GWQ assessment.

Additionally, the result demonstrated that the RF model could be used as a reliable model for ground water vulnerability, investigation and appropriately manage or monitor of any aquifer. Nolan et al. (2015) employed different machine learning models, BRT, ANN, and BN to predict NO₃⁻ concentrations in California. Prediction result revealed that BRT better than BN and ANN. Whereas BN ranked as second-best predictive model after BRT. Also, comparisons between these models with MLR and RF regression model were done. The results of BRT were compared to RF, MLR had low holdout with R² = 0.07 and described less than half difference in training datasets. Spatial pattern of prediction via BRT model agreed acceptably well with earlier actual patterns of NO₃⁻ occurrence. Singh et al. (2014) built ensembles of classification and regression models and used the ground water hydro-chemistry database at India. SDT, DTF, and DTB methods were compared with the SVM model. The DTF and DTB methods showed more accurate performance than SVM in classification and regression. It should be mentioned that combination of bagging and stochastic gradient boosting algorithms in DTF and DTB methods, respectively resulted in their improved prediction capability. The ensemble models effectively presented the influence for seasonally differences and anthropogenic acts on groundwater hydro-chemistry.

2.5.2 Results of Machine Learning Algorithms

The approaches that fall into the above-mentioned groups are summarized in this section. Machine learning algorithms are common methods after traditional techniques.

- 1) MARS is a flexible model and it has the capability of managing high dimension problems in regression and classification, making it a superior model to testing for predicting water quality of groundwater. It could independently classify additive contribution (Friedman 1991). Also, the M5 Tree model could produce good generalization, best comprehend the parameters relationships, avoid overfitting, and attain good accuracy. Nonetheless they displayed inferior performance in comparison with DENFIS and GEP (Esmailbeiki et al. 2020). GEP encode specific population as linear string for fixed length and later express them as non-linear entities of various shapes and sizes.

- 2) Also, the accurate performance of kNN model in classification was shown, and it performed better than SVM and RF (Rahmati et al. 2019). On the other hand, kNN required more discovery along with research opportunity for hybrid k-NN models.
- 3) There are many variants of DT that are remain to be tested concerning water pollution and GWQ such as naïve bayes trees. Model solve problems (recursive dive and conquer) and generalization of complex issues, leads to increase of applying complex data of GWQ prediction. Additionally DT models are easy to understand (Vijay and Kamaraj 2019).
- 4) BRT model do not require data transformation, are able to be fitted for complex non-linear relationships, and automatically way combines interactions effect between predictors (Elith et al. 2008). The BRT model is, thus, fit suitable for prediction water quality condition complexity, heterogeneous setting for instance the glacial aquifer system (Erickson et al. 2018) and further aquifers setting (Ransom et al. 2017;Knoll et al. 2017). Also, in (Knierim et al. 2020) the BRT model supporting the hypotheses for each surficial or deep sources of salinity. Moreover, BRT proved it has good capability in predicting patterns of NO_3^- (Nolan et al. 2015).

3 General Research Assessment and Evaluation

Analysis of artificial intelligence models shown that these methods generated better result than statistical techniques for example linear and logistic regression as exhibited in Table 2, ANN methods were applied predominantly as a first methods for GWQ modeling. Additionally, use of new classification methods such as KSOM as a type of ANN model for recognizing polluting and nonpolluting wells must increase. The propagation of using the AI approaches began in the late 1990s, and many AI applications involved using various classical AI methods including ANN, ANFIS, SVM and ML. The latest trends include hybrid models, which are an important development of AI models that integrated different basic algorithms for estimation of water quality in groundwater. These combinations make standalone models more reliable and cost effective through decreasing the cost of computation and increasing the accuracy of produced outcomes. A limited number of papers were found to hybrid models in GWQ modeling in this work.

All former review and study papers have regularly emphasized only input variables, which are is considered as crucial for the prediction procedure, and many input optimization methods were applied in order to guarantee the most likely inputs used are qualified to relationships recognized with output variables. Although researchers have studied the input and output relationship for prediction, attention and fewer analyses have been considered in GWQ review papers regarding outputs variables. Because the basic method of these papers is to determine the relationships between the input-output. Also, because of it could be those inputs are utilized for training, optimization and dimensional decrease, and the relationship is determined depending on its sensitivity toward selected outputs; therefore, they are emphasized in this study. Additionally, the outputs are usually chosen based on the issue addressed via the researchers for instance, when the issue is salinity in groundwater, therefore, EC or TDS is carried as output not through using any predefined method. The main output variables, as highlighted as in Fig. 10 represent a number of studies which employed every output variable out of all the considered studies. Consequently, the present review focused on the evaluation of output and which output variable have frequently used to measure water quality in groundwater in most studies. Figure 10 shows that, out of 81 measured papers, most studies reviewed in this study used variable NO_3^- with 20 out of 81 studies as output to measure water quality of groundwater, which indicates how essential this variable is in predicting GWQ. NO_3^- was mostly considered in case for GWQ classification and regression where multiple, or single parameters were evaluated versus government standard as well as classifying in simplified classes. From Fig. 10 the other most popular GWQ variables are TDS, EC, and TH which are an important, a part of many water related reactions. For instance, TDS is crucial water quality and groundwater healthy indicator, and a high concentration of it maybe cause people to suffer from some diseases such as kidney and heart illnesses. Water that has a high-level of solids maybe caused laxative or constipation effects (Sasikaran et al. 2016). Also, the increase in ions concentration improves the EC of water. Usually, EC reveals mineralization and salinity in the water and the value of dissolved solids in water determine EC. The impact of SO_4^{-2} reduction results play a major role in mediating redox conditions and biogeochemical processes for ground water system. Along with Sulfate, Cl which are essential inorganic

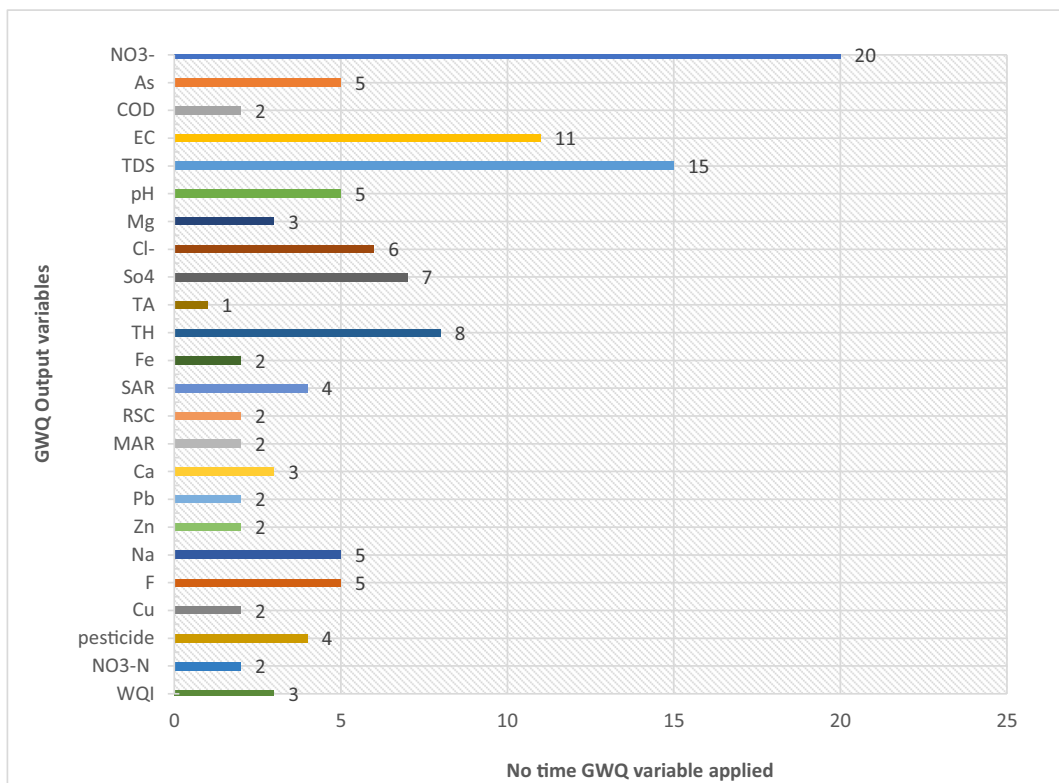


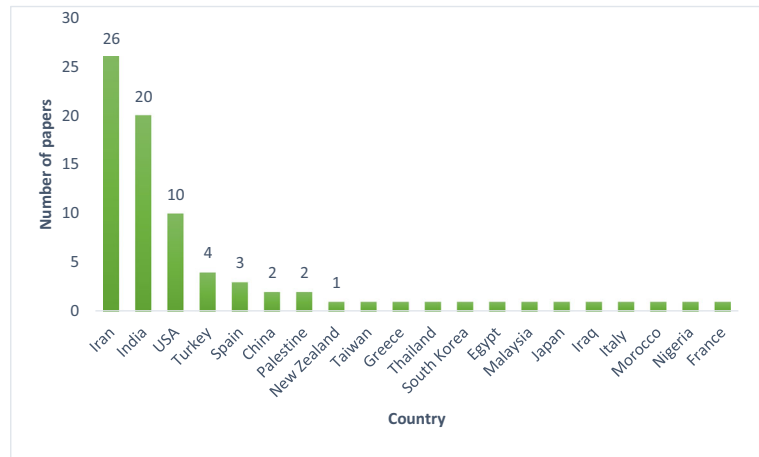
Fig. 10 The output variables used for AI in predicting GWQ

anions, are found in differing concentrations in natural waters and constitute an indicator of contamination. Their presence in groundwater may reveal anthropogenic pollution because they exist in urine and maintenance products. Another output frequently used is WQI which is computed via a complex numerical formula using excessive time and might have incorrect values. This process included using multiple water quality parameters to calculate and compare them with standards values to improve comprehension for the nonscientific community, for example policy makers (Smith 1990). The essential GWQ variable is pH, which is the seventh most used as output variable, hydrogen ion's ability is the most powerful initiators of each chemical and biological processes, including decomposition, toxicity, solubility, and degradation. Only some variables maybe have not directed and cooperative influence. For example, increasing Na + in irrigation water maybe result in secondary alkalization of the soil. Also, the COD variable, is used in the GWQ indicators which have visible microbial and chemical reactions that occur in presence of oxygen in water. COD signifies the need for care or the

scope of rivers contamination. In addition there are many output variables used to measure the GWQ not displayed in Fig. 10 because they were only used once, such as PO₄, Ni, source of pollution, E.coli, uranium,..., etc. According to the reported review, it's obvious that the performing ability of artificial intelligence is reliant on water quality's multidimensional nature. While constructing the regression/classification models, complexities depend on number of various input variables combined such as less is best while excessive input variables increase complexity of the model. Its important to highlight the ability of AI methods in many applications with various model versions as in Table 2, and it could expand dependent upon government policies, geological site, regional population behavior, climate changing and other exclusive scenarios. Therefore, there is no concern in applications of AI models in easy cases, nevertheless, everchanging of GWQ caused by dependent variables could be causing a significant influence on prediction outcome.

Figure 11 demonstrations a number of review studies regarding every country where the study area is situated.

Fig. 11 The sum of studied AI models in GWQ modeling with respect to the countries



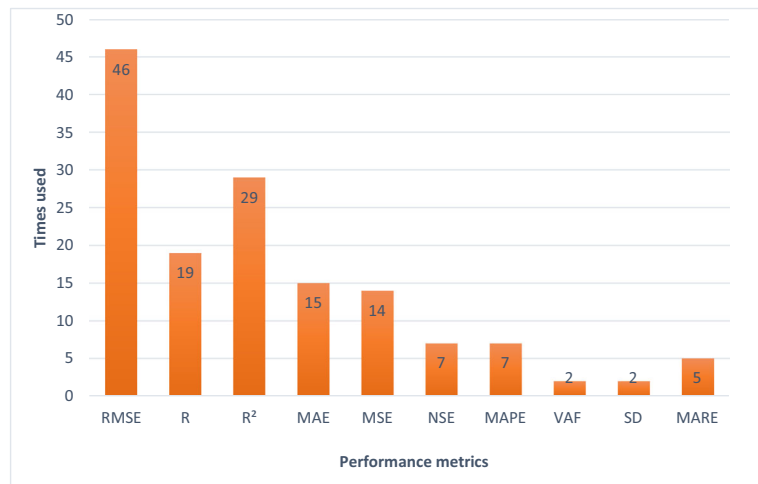
Many study areas are situated in Iran and India, with 26 and 20 out of 81 papers respectively. This point may display the attention of Iranian and Indian academics in water quality in groundwater realm. However, it could be because of dryness or partial dryness of areas as Iran, for example the surface water resources are at a low level, and groundwaters is the greatest obtainable water resources, and consequently GWQ database are more accessible compared to data of surface water. The next category has the USA with 10 papers and Turkey with 4 papers and the remaining countries can be seen in Fig. 11. It's known that the black box AI models are valuable in GWQ modeling, however, they aren't developed utilizing comprehensions on physical process engaged. In this kind of prediction, information regarding fundamental mechanisms is not required and a major aim is attaining predictions accurately.

Performance indicators are statistical measures that help for developers to evaluate and calibrate prediction performance on any platform. In addition, performance precision and effective is translated to comprehensible and expressible forms. Twenty evaluation criteria were used for predicting GWQ as represent in Fig. 12. The quantitative errors measure method is mainly used with AI techniques. The root mean squares errors (RMSE) is the most often applied squared errors measuring system with 46 papers, followed by Determination coefficient (R^2) (29 papers, correlation coefficient (R) with 19 papers, the mean squared errors (MSE) and mean absolute errors (MAE) with 14 and 15 papers respectively, and Nash Sutcliffe efficiency (NSE) and mean absolute percentage errors (MAPE) with 7 papers. Some performance metrics are not illustrated in Fig. 12 because they

not commonly used in many papers such as (%bias, CE, RSR, SI, WI, mmce, AUC, ROC, ..., etc.).

RMSE is commonly used because it could reveal numerous deviations, its reliable and provides the highest weights to errors. Outliers must remove R^2 , which is the second most preferred metric, assessing forecasted values generated via such method with respect to observed values, that provide performance of models and range between zero to one. Other set involves the absolute error, which are named: MAE, MSE and MAPE. They're an absolute variance between observed value and predict value. However, they do not model validity, they only provided information about range of errors and they have good sharpness. R and NSE measure a robust relationship among dataset and comparative movements for two data sets by difference units. Various papers have been reported classifying the GWQ to classes named, "excellent", "good", "poor" and "very poor" according to standards of GWQ guidelines of study areas. Among calculated performance indicators, though R^2 and R were commonly applied for model's estimation, they are over sensitive to the highest amounts and insensitively to addition and proportionate variations between models prediction and actual dataset (Moriasi et al. 2007). However, the MAE has been applied as a major measurement because the ultimate values in the forecast dataset are less sensitive than RMSE (Fox 1981). According to studies, NSE is the normalized metric that quantifies the residual variance's relative magnitude in comparison to the calculated data variance (Nash and Sutcliffe 1970). According to Moriasi et al. (2007) the efficiency guidelines of the hydrological model assessment are very

Fig. 12 Performance indicators that have been used for prediction models' evaluation and assessment



successful for the hydrological model, where NSE is more than 0.75. Therefore, this is another extraordinary measure to be examined for such modeling.

4 Conclusion and Future Trends

This study provides a review of the innovations of water resources engineering on applying artificial intelligence models, novelty of water quality in groundwater modeling with various challenging problems in many developing areas and aims to be attained in the forthcoming research, AI modeling approaches have been successfully applied for water quality in groundwater papers, as reviewed via 81 research studies from 2001 till 2021. The AI methods are integrated tools of biological sciences, statistics, and mathematics. These models could overwhelm the particular difficulties via behavior, thinking, and learning. Once integrated with learning they become machine learning. The AI methods could be helpful once it's difficult to develop a sufficient information drive simulating model because of a shortage for ability to acceptably build the mathematic or physic models for the basic processes. AI methods have some important steps involving inputs dataset consideration, inputs dataset partition, control of model characteristics, test set, training set and so on. All these stages will be carefully designed, and the model performance is expected to be reasonable. Even so, it must be noticed that there was no generally accepted method for performing these stages; rather some various research studies performed each stage empirically and/or by trial

and error, taking available data and existing conditions into account. The analysis of the study's findings, which were divided into two distinct sections for example the assessment of each AI group and the overall assessment and evaluation, will serve as a valuable resource for researchers looking to conduct comparable work in a related area, create novel approaches, and increase the quality of modeling. To achieve this aim, the next points are recommended:

- 1.1. AI methods are tools for data and, if proper and competent control stations are developed that can deliver continuous and stable data worldwide, they can achieve the highest precision with consistent, regular and adequate amounts of data.
- 1.2. Several fields of research include data collection challenges and some problems are impossible to solve. Therefore, more AI models should be developed that can address the lack of data and the lack of continuous information. With recent technological advances, such as GIS and remote sensing, the problem of lost data and data, especially unreachable stations, can be solved. These tools may also be used to control resources or to model the environment. For better spatial GWQ modeling, high resolution data from satellites is available.
- 1.3. As to the various GWQ simulation AI models, a specific kind of AI model for a particular problem is practically not to be recommended. It is obvious that a hybrid or coupled model performed better than single AI models. In the various stages of the

GWQ modeling, various types of AI techniques can be evaluated to choose a better AI model in each stage and combine it to achieve an optimum modeling efficiency.

1.4. The AI models can be implemented in the evaluation of groundwater quality in three significant application areas. The fields of application have been divided as:

1. Concentration estimation of different groundwater quality parameters: For reliable concentration estimation in groundwater, AI models such as ANFIS, ANN, SVR, and tree-based models can be effective instruments in this field. According to the literature, the efficiency of these models is comparable and close to each other, but newer models such as ANFIS and tree-based models provide more reliable results. However, because ANFIS employs fuzzy logic, it outperforms it in scenario-based problems.
2. Classification of groundwater quality parameters for instance NO₃ polluting areas: for this aim, the SOM model was used as a powerful model to classify polluting and unpolluting areas, this model has superior visual capabilities, and the SOM model can be an excellent alternative in this field.
3. Optimization of the prediction of groundwater quality parameters: the main focus of this field of application is the determination of hydro economic mechanisms that optimized economic expenditures for environmental purposes and various scenarios. The AI models can be optimized using multi-objective particle swarm optimization (MOPSO), Non-dominated Sorting Genetic algorithm (NSGA-II), and multi-objective shuffled complex evolution metropolis (MOSCEM-UA) algorithms, which are the well-known meta-heuristic optimizing methods to multi objective optimization in water resources management therefore, it can be a great option to find the best results in this area.

1.5. Several AI approaches, such as ELM, SVM, MARS models, have been effectively employed for GWQ estimation but as standalone models, they are not ever combined with bioinspired optimization techniques. They may then be stuck in optimal local solutions and cannot

estimate GWQ accurately. In order to address this problem in future studies, standalone models must be coupled with other algorithm for example evolutionary algorithms such as Ant colony optimization model, genetic algorithm etc., then performance of this novel or hybrid model must be examined in estimate the GWQ. The achieved outcome from hybrid model comparison with standalone model will enable us to determine hybridization effectiveness compared to standalone models and delivers insights into the strengths of hybrid models.

- 1.6. After reviewing all the research papers regarding GWQ, it could be conclude that the contaminations mixing processes of groundwater quality are complex and causing a synergetic reactions produces changes in water quality composition. During the experimental period, other variables such as irregular velocity bed configuration, sediment load, and the dead regions should therefore be addressed.
- 1.7. Because of the low number of studies in GWQ field modeling using hybrid-AI models, its recommended that additional research be conducted on this topic.

Acknowledgements The project was funded from UAE University within the initiatives of Asian Universities Alliance collaboration.

Declaration

Conflict of Interests The authors declare that they have no conflict of interest.

References

- Aguilera, P. A., Fernández, A., Roperó, R. F., & Molina, L. (2013). Groundwater quality assessment using data clustering based on hybrid Bayesian networks. *Stochastic Environmental Research and Risk Assessment*, 27(2), 435–447. <https://doi.org/10.1007/s00477-012-0676-8>
- Ahmed, A., Masrur, A., & Shah, S. M. A. (2017). Application of adaptive Neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River. *Journal of King Saud University - Engineering Sciences*,

- 29(3), 237–243. <https://doi.org/10.1016/j.jksues.2015.02.001>
- Alagha, J. S., Said, M. A. M., & Mogheir, Y. (2014). Modeling of nitrate concentration in groundwater using artificial intelligence approach—a case study of Gaza coastal aquifer. *Environmental Monitoring and Assessment*, 186(1), 35–45. <https://doi.org/10.1007/s10661-013-3353-6>
- Alizamir, M., & Sobhanardakani, S. (2016). Forecasting of heavy metals concentration in groundwater resources of Asadabad plain using artificial neural network approach. *J Adv Environ Health Res*, 4(42), 68–77.
- Alizamir, M., & Sobhanardakani, S. (2018). An artificial neural network - particle swarm optimization (ANN- PSO) approach to predict heavy metals contamination in groundwater resources. *Jundishapur Journal of Health Sciences*, 10(2). <https://doi.org/10.5812/jjhs.67544>
- Almasri, M. N., & Kaluarachchi, J. J. (2005). Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. *Environmental Modelling & Software*, 20(7), 851–871.
- Arabgol, R., Sartaj, M., & Asghari, K. (2016). Predicting nitrate concentration and its spatial distribution in groundwater resources using support vector machines (SVMs) model. *Environmental Modeling and Assessment*, 21(1), 71–82. <https://doi.org/10.1007/s10666-015-9468-0>
- Aryafar, Ahmad, Vahid Khosravi, Hosniyeh Zarepourfard, and Reza Rooki. 2019. “Evolving genetic programming and other AI-based models for estimating groundwater quality parameters of the Khezri plain, eastern Iran.” *Environmental Earth Sciences* 78(3):0. doi: <https://doi.org/10.1007/s12665-019-8092-8>.
- Bain, R. E. S., Wright, J. A., Christenson, E., & Bartram, J. K. (2014). Rural:Urban inequalities in post 2015 targets and indicators for drinking-water. *Science of the Total Environment*, 490, 509–513. <https://doi.org/10.1016/j.scitotenv.2014.05.007>
- Barzegar, R., & Moghaddam, A. A. (2016). Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Modeling Earth Systems and Environment*, 2(1), 1–13. <https://doi.org/10.1007/s40808-015-0072-8>
- Barzegar, R., Moghaddam, A. A., Adamowski, J., & Fijani, E. (2017). Comparison of machine learning models for predicting fluoride contamination in groundwater. *Stochastic Environmental Research and Risk Assessment*, 31(10), 2705–2718. <https://doi.org/10.1007/s00477-016-1338-z>
- Bashi-Azghadi, S. N., Kerachian, R., Bazargan-Lari, M. R., & Solouki, K. (2010). Characterizing an unknown pollution source in groundwater resources systems using PSVM and PNN. *Expert Systems with Applications*, 37(10), 7154–7161. <https://doi.org/10.1016/j.eswa.2010.04.019>
- Batayneh, A., Zaman, H., Zumlot, T., Ghrefat, H., Mogren, S., Nazzal, Y., Elawadi, E., Qaisy, S., Bahkaly, I., Al-taani, A., Batayneh, A., Zaman, H., Zumlot, T., Ghrefat, H., and Mogren, S. (2016). “Hydrochemical Facies and ionic ratios of the coastal groundwater aquifer of Saudi Gulf of Aqaba : Implication for seawater intrusion stable URL : <http://www.Jstor.Org/Stable/43289862> linked references are available on JSTOR for this article : Hydrochemica.”
- Bedi, S., Samal, A., Ray, C., & Snow, D. (2020). Comparative evaluation of machine learning models for groundwater quality assessment. *Environmental Monitoring and Assessment*, 192(12). <https://doi.org/10.1007/s10661-020-08695-3>
- Belkhir, L., Mouni, L., Tiri, A., Narany, T. S., & Nouibet, R. (2018). Spatial analysis of groundwater quality using self-organizing maps. *Groundwater for Sustainable Development*, 7, 121–132.
- Bhagat, S. K., Tiyasha, T., Tung, T. M., Mostafa, R. R., & Yaseen, Z. M. (2020). Manganese (Mn) removal prediction using extreme gradient model. *Ecotoxicology and Environmental Safety*, 204, 111059. <https://doi.org/10.1016/j.ecoenv.2020.111059>
- Bilali, E., Ali, A. T., & Brouziyne, Y. (2021). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, 245, 106625. <https://doi.org/10.1016/j.agwat.2020.106625>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bui, D. T., Khosravi, K., Karimi, M., Busico, G., Khozani, Z. S., Nguyen, H., Mastrocico, M., Tedesco, D., Cuoco, E., & Kazakis, N. (2020). Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm. *Science of the Total Environment*, 715, 136836. <https://doi.org/10.1016/j.scitotenv.2020.136836>
- Chen, T. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. <https://doi.org/10.1145/2939672.2939785>
- Chen, S.-T., & Pao-Shan, Y. (2007). Pruning of support vector networks on flood forecasting. *Journal of Hydrology*, 347(1), 67–78. <https://doi.org/10.1016/j.jhydrol.2007.08.029>
- Cho, K. H., Sthiannopkao, S., Pachepsky, Y. A., Kim, K.-W., & Kim, J. H. (2011). Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Research*, 45(17), 5535–5544. <https://doi.org/10.1016/j.watres.2011.08.010>
- Choi, B.-Y., Yun, S.-T., Kim, K.-H., Kim, J.-W., Kim, H. M., & Koh, Y.-K. (2014). Hydrogeochemical interpretation of south Korean groundwater monitoring data using self-organizing maps. *Journal of Geochemical Exploration*, 137, 73–84. <https://doi.org/10.1016/j.gexplo.2013.12.001>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Dahiya, S., Singh, B., Shalini, G., Garg, V. K., & Kushwaha, H. S. (2007). Analysis of groundwater quality using fuzzy synthetic evaluation. *Journal of Hazardous Materials*, 147(3), 938–946. <https://doi.org/10.1016/j.jhazmat.2007.01.119>
- Dar, I. A., Sankar, K., Dar, M. A., & Majumder, M. (2012). Fluoride contamination – Artificial neural network modeling and inverse distance weighting approach. *Revue Des Sciences de l'Eau*, 25(2), 165–182. <https://doi.org/10.7202/1011606ar>
- Dixon, B. (2005). Groundwater vulnerability mapping: A GIS and fuzzy rule based integrated tool. *Applied Geography*, 25(4), 327–347. <https://doi.org/10.1016/j.apgeog.2005.07.002>
- Dixon, B. (2009). A case study using support vector machines, neural networks and logistic regression in a GIS to identify wells contaminated with nitrate-N. *Hydrogeology Journal*,

- 17(6), 1507–1520. <https://doi.org/10.1007/s10040-009-0451-1>
- Dohare, D., Deshpande, S., Kotiya, A.. (2014). “Analysis of Ground Water Quality Parameters: A Review *Www.Isca.Me.*” *Research Journal of Engineering Sciences* ISSN *Res. J. Engineering Sci* 3(5):2278–9472.
- Eberhat, R., and Kennedy, J. (1995). “A new optimizer using particle swarm theory.” In *Sixth international symposium on micro machine and human science*, (pp. 39–43). Piscataway.
- Ehteshami, M., Dolatabadi Farahani, N., & Tavassoli, S. (2016). Simulation of nitrate contamination in groundwater using artificial neural networks. *Modeling Earth Systems and Environment*, 2(1), 1–10. <https://doi.org/10.1007/s40808-016-0080-3>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Erickson, M. L., Elliott, S. M., Christenson, C. A., & Krall, A. L. (2018). Predicting Geogenic arsenic in drinking water Wells in glacial aquifers, north-Central USA: Accounting for depth-dependent features. *Water Resources Research*, 54(12), 10, 172–10,187. <https://doi.org/10.1029/2018WR023106>
- Esau, I., Miles, V., Varentsov, M., Konstantinov, P., & Melnikov, V. (2019). Spatial structure and temporal variability of a surface urban Heat Island in cold continental climate. *Theoretical and Applied Climatology*, 137(3), 2513–2528.
- Esmailbeiki, F., Nikpour, M. R., Singh, V. K., Kisi, O., Sihag, P., & Sanikhani, H. (2020). Exploring the application of soft computing techniques for spatial evaluation of groundwater quality variables. *Journal of Cleaner Production*, 276, 124206. <https://doi.org/10.1016/j.jclepro.2020.124206>
- Etemad-Shahidi, A., & Bonakdar, L. (2009). Design of rubble-mound breakwaters using M5 ' machine learning method. *Applied Ocean research*, 31(3), 197–201. <https://doi.org/10.1016/j.apor.2009.08.003>
- Fadipe, O. O., L. K. Abidoye, J. O. Adeosun, B. B. Oguntola, and O. Adewusi. (2021) “Simulation of groundwater quality characteristics using artificial neural network.” *Nigerian Journal OfTechnology* 40(1). <https://doi.org/10.4314/njt.v40i1.21>.
- Fahimi, F., Yaseen, Z. M., & El-shafie, A. (2017). Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review. *Theoretical and Applied Climatology*, 128(3–4), 875–903. <https://doi.org/10.1007/s00704-016-1735-8>
- Fijani, E., Nadiri, A. A., Moghaddam, A. A., Tsai, F. T. C., & Dixon, B. (2013). Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh–Bonab plain aquifer, Iran. *Journal of Hydrology*, 503, 89–100. <https://doi.org/10.1016/j.jhydrol.2013.08.038>
- Fijani, E., Moghaddam, A. A., Tsai, F. T. C., & Tayfur, G. (2017). Analysis and assessment of Hydrochemical characteristics of Maragheh-Bonab plain aquifer, northwest of Iran. *Water Resources Management*, 31(3), 765–780. <https://doi.org/10.1007/s11269-016-1390-y>
- Fox, D. G. (1981). Judging air quality model performance. A summary of the AMS workshop on dispersion model performance, woods hole, Mass., 8-11 September 1980. *Bulletin of the American Meteorological Society*, 62(5), 599–609. [https://doi.org/10.1175/1520-0477\(1981\)062<0599:JAQMP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1981)062<0599:JAQMP>2.0.CO;2)
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67. <https://doi.org/10.1214/aos/1176347963>
- Gemitzis, A., Petalas, C., Pisinaras, V., & Tsihrintzis, V. A. (2010). Spatial prediction of nitrate pollution in Groundwaters using neural networks and GIS: An application to south Rhodope aquifer (Thrace, Greece). *Hydrological Processes*, 2274(November 2008). <https://doi.org/10.1002/hyp>
- Gholami, R., Kamkar-Rouhani, A., Doulati Ardejani, F., & Maleki, S. (2011). Prediction of toxic metals concentration using artificial intelligence techniques. *Applied Water Science*, 1(3), 125–134. <https://doi.org/10.1007/s13201-011-0016-z>
- Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3(2), 95–99. <https://doi.org/10.1023/A:1022602019183>
- Hadian, M., Saryazdi, S.M.E., Mohammadzadeh, A., and Babaei, M.. (2021). “Chapter 11 - Application of artificial intelligence in modeling, control, and fault diagnosis.” In *Applications of Artificial Intelligence in Process Systems Engineering*.(pp. 255–323). Elsevier.
- Heidarzadeh, N. (2017). A practical low-cost model for prediction of the groundwater quality using artificial neural networks. *Journal of Water Supply: Research and Technology - AQUA*, 66(2), 86–95. <https://doi.org/10.2166/aqua.2017.035>
- Hem, J.D.. (1985). Study and interpretation of the chemical characteristics of natural water. Vol. 2254. Department of the Interior, US Geological Survey.
- Isazadeh, M., Biazar, S. M., & Ashrafzadeh, A. (2017). Support vector machines and feed-forward neural networks for spatial modeling of groundwater qualitative parameters. *Environmental Earth Sciences*, 76(17). <https://doi.org/10.1007/s12665-017-6938-5>
- Jafari, R., Hassani, A. H., Torabian, A., Ghorbani, M. A., & Mirbagheri, S. A. (2019). Prediction of groundwater quality parameter in the Tabriz plain, Iran using soft computing methods. *Journal of Water Supply: Research and Technology - AQUA*, 68(7), 573–584. <https://doi.org/10.2166/aqua.2019.062>
- Jalalkamali, A. (2015). Using of hybrid fuzzy models to predict spatiotemporal groundwater quality parameters. *Earth Science Informatics*, 8(4), 885–894. <https://doi.org/10.1007/s12145-015-0222-6>
- Jang, J. S. R. 1993. “ANFIS : Adap Tive-ne Twork-based fuzzy inference system.” *IEEE Transactions on systems*, 23(3).
- Jang, J.S.R. (1996). Neuro fuzzy modeling for dynamic system identification. In *Soft Computing in Intelligent Systems and Information Processing*. Proceedings of the 1996 Asian Fuzzy Systems Symposium (pp. 320–25).
- Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [book review]. *IEEE Transactions on Automatic Control*, 42(10), 1482–1484.
- Jebastina, N., & Prince Arulraj, G. (2018). Spatial prediction of nitrate concentration using GIS and ANFIS Modelling in groundwater. *Bulletin of Environmental Contamination and Toxicology*, 101(3), 403–409. <https://doi.org/10.1007/s00128-018-2406-5>

- Jha, M. K., Shekhar, A., & Annie Jenifer, M. (2020). Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. *Water Research*, 179, 115867. <https://doi.org/10.1016/j.watres.2020.115867>
- Jiang, Y., Wu, Y., Groves, C., Yuan, D., & Kambesis, P. (2009). Natural and anthropogenic factors affecting the groundwater quality in the Nandong karst Underground River system in Yunan, China. *Journal of Contaminant Hydrology*, 109(1), 49–61. <https://doi.org/10.1016/j.jconhyd.2009.08.001>
- Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., & Sankhua, R. N. (2019). Prediction of water quality index using artificial neural network and multiple linear regression Modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*, 5(3), 951–962. <https://doi.org/10.1007/s40808-019-00581-3>
- Kassem, Y., Gökçekuş, H., & Maliha, M. R. M. (2021). Identifying Most influencing input parameters for predicting chloride concentration in groundwater using an ANN approach. *Environmental Earth Sciences*, 80(7), 248. <https://doi.org/10.1007/s12665-021-09541-6>
- Keskin, T. E., Düğenci, M., & Kaçaroğlu, F. (2015). Prediction of water pollution sources using artificial neural networks in the study areas of Sivas, Karabük and Bartın (Turkey). *Environmental Earth Sciences*, 73(9), 5333–5347. <https://doi.org/10.1007/s12665-014-3784-6>
- Khaki, M., Yusoff, I., & Islami, N. (2015). Application of the artificial neural network and Neuro-fuzzy system for assessment of groundwater quality. *Clean - Soil, Air, Water*, 43(4), 551–560. <https://doi.org/10.1002/clen.201400267>
- Khalil, A., Almasri, M. N., McKee, M., & Kaluarachchi, J. J. (2005). Applicability of statistical learning algorithms in groundwater quality modeling. *Water Resources Research*, 41(5), 1–16. <https://doi.org/10.1029/2004WR003608>
- Khan, F.M., Gupta, R., and Sekhri, S. (2021). Superposition learning-based model for prediction of E.Coli in groundwater using physico-chemical water quality parameters. *Groundwater for Sustainable Development* 100580. doi: <https://doi.org/10.1016/j.gsd.2021.100580>.
- Khudair, B. H., Jasim, M. M., & Alsaqqa, A. S. (2018). Artificial neural network model for the prediction of groundwater quality. *International Journal of Plant & Soil Science*, 8(4), 1–13. <https://doi.org/10.9734/ijps/2015/19686>
- Kisi, O., Keshavarzi, A., Shiri, J., Zounemat-Kermani, M., & Omran, E. S. E. (2017). Groundwater quality modeling using Neuro-particle swarm optimization and Neuro-differential evolution techniques. *Hydrology Research*, 48(6), 1508–1519. <https://doi.org/10.2166/nh.2017.206>
- Kisi, O., Azad, A., Kashi, H., Saedian, A., Hashemi, S. A. A., & Ghorbani, S. (2019). Modeling groundwater quality parameters using hybrid Neuro-fuzzy methods. *Water Resources Management*, 33(2), 847–861. <https://doi.org/10.1007/s11269-018-2147-6>
- Klaşlan, Y., Tuna, G., Gezer, G., Gulez, K., Arkoc, O., and Potirakis, S. M.. (2014). ANN based estimation of groundwater quality using a wireless water quality network. *International Journal of Distributed Sensor Networks* 2014. doi: <https://doi.org/10.1155/2014/458329>.
- Knierim, K. J., Kingsbury, J. A., Haugh, C. J., & Ransom, K. M. (2020). Using boosted regression tree models to predict salinity in Mississippi embayment aquifers, Central United States. *Journal of the American Water Resources Association*, 56(6), 1010–1029. <https://doi.org/10.1111/1752-1688.12879>
- Knoll, L., Breuer, L., & Bach, M. (2017). Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Science of the Total Environment*, 668, 1317–1327. <https://doi.org/10.1016/j.scitotenv.2019.03.045>
- Koza, J. R., Bennett, F. H., Andre, D., Keane, M. A., & Dunlap, F. (1997). Automated synthesis of analog electrical circuits by means of genetic programming. *IEEE Transactions on Evolutionary Computation*, 1(2), 109–128.
- Krapivin, V. F., Varotsos, C. A., & Nghia, B. Q. (2017). A modeling system for monitoring water quality in lagoons. *Water, Air, and Soil Pollution*, 228(10). <https://doi.org/10.1007/s11270-017-3581-4>
- Krapivin, V. F., Varotsos, C. A., & Marechek, S. V. (2018). The dependence of the soil microwave attenuation on frequency and water content in different types of vegetation: An empirical model. *Water, Air, & Soil Pollution*, 229(4), 110. <https://doi.org/10.1007/s11270-018-3773-6>
- Krapivin, V. F., Mkrchan, F. A., Varotsos, C. A., & Xue, Y. (2021). Operational diagnosis of arctic waters with instrumental technology and information modeling. *Water, Air, & Soil Pollution*, 232(4), 137. <https://doi.org/10.1007/s11270-021-05068-5>
- Kumar, P. G. D., Viswanath, N. C., Cyrus, S., & Abraham, B. M. (2020). Mixing data for multivariate statistical study of groundwater quality. *Environmental Monitoring and Assessment*, 192(8). <https://doi.org/10.1007/s10661-020-08465-1>
- Kuo, Y. M., Liu, C. W., & Lin, K. H. (2004). Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of Blackfoot disease in Taiwan. *Water Research*, 38(1), 148–158. <https://doi.org/10.1016/j.watres.2003.09.026>
- Lindskog, P. (1997). Fuzzy identification from a grey box modeling point of view. In *Fuzzy model identification* (pp. 3–50). Springer.
- Liu, B., Wang, L., & Jin, Y. (2007). An effective PSO-based Memetic algorithm for flow shop scheduling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(1), 18–27. <https://doi.org/10.1109/TSMCB.2006.883272>
- Lopez, A. M., Wells, A., & Fendorf, S. (2021). Soil and aquifer properties combine as predictors of groundwater uranium concentrations within the Central Valley, California. *Environmental Science & Technology*, 55(1), 352–361. <https://doi.org/10.1021/acs.est.0c05591>
- Lowe, D., Broomhead, D., and Malvern (United Kingdom); Royal Signals and Radar Establishment. 1988. Radial basis functions multivariable functional interpolation and adaptive networks. No. *RSRE-MEMO-4148*. Royal signals and radar establishment Malvern (United Kingdom).
- Maiti, S., Erram, V. C., Gupta, G., Tiwari, R. K., Kulkarni, U. D., & Sangpal, R. R. (2013). Assessment of groundwater quality: A fusion of geochemical and geophysical information via Bayesian neural networks. *Environmental Monitoring and Assessment*, 185(4), 3445–3465. <https://doi.org/10.1007/s10661-012-2802-y>
- Maroufpoor, S., Jalali, M., Nikmehr, S., Shiri, N., Shiri, J., and Maroufpoor, E.. (2020). “Modeling groundwater quality by using hybrid intelligent and Geostatistical methods.”

- Environmental Science and Pollution Research, 27, 28183–97
- Mehr, D., Ali, V. N., Kahya, E., Hrnjica, B., Sattar, A. M. A., & Yaseen, Z. M. (2018). Genetic programming in water resources engineering: A state-of-the-art review. *Journal of Hydrology*, 566, 643–667. <https://doi.org/10.1016/j.jhydrol.2018.09.043>
- Melin, P., Olivas, F., Castillo, O., Valdez, F., Soria, J., & Valdez, M. (2013). Optimal design of fuzzy classification systems using PSO with dynamic parameter adaptation through fuzzy logic. *Expert Systems with Applications*, 40(8), 3196–3206. <https://doi.org/10.1016/j.eswa.2012.12.033>
- Moasheri, S. A., & Tabatabaie, S. M. (2013). Estimate the spatial distribution TDS the fusion method geostatistics and artificial neural networks. *International Journal of Agriculture*, 3(4), 699.
- Moasheri, S.A., Khammar, G., Poornoori, Z., and Beyranvand, Z.. (2013). "Estimate the Spatial Distribution TDS the Fusion Method Geostatistics and Artificial Neural Networks." *International Journal of Agriculture and Crop Sciences*. 410–20.
- Mohammadi, S., Amir, A. S. S., & Naseri, A. A. (2017). Simulation of groundwater quality parameters using ANN and ANN+PSO models (case study: Ramhormoz plain). *Pollution*, 3(2), 191–200. <https://doi.org/10.7508/pj.2017.02.003>
- Moosavi, V., Vafakhah, M., Shirmohammadi, B., & Ranjbar, M. (2012). Optimization of wavelet-ANFIS and wavelet-ANN hybrid models by Taguchi method for groundwater level forecasting. *Arabian Journal for Science and Engineering*, 39(3), 1785–1796. <https://doi.org/10.1007/s13369-013-0762-3>
- Moriassi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Daren Harmel, R., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Nadiri, A. A., Fijani, E., Tsai, F. T. C., & Moghaddam, A. A. (2013). Supervised committee machine with artificial intelligence for prediction of fluoride concentration. *Journal of Hydroinformatics*, 15(4), 1474–1490. <https://doi.org/10.2166/hydro.2013.008>
- Nadiri, A. A., Sedghi, Z., Khatibi, R., & Gharekhani, M. (2017). Mapping vulnerability of multiple aquifers using multiple models and fuzzy logic to objectively derive model structures. *Science of the Total Environment*, 593–594, 75–90. <https://doi.org/10.1016/j.scitotenv.2017.03.109>
- Nadiri, A. A., Norouzi, H., Khatibi, R., & Gharekhani, M. (2019). Groundwater DRASTIC vulnerability mapping by unsupervised and supervised techniques using a Modelling strategy in two levels. *Journal of Hydrology*, 574, 744–759. <https://doi.org/10.1016/j.jhydrol.2019.04.039>
- Nakagawa, K., Amano, H., Kawamura, A., & Berndtsson, R. (2017). Classification of groundwater chemistry in Shimabara, using self-organizing maps. *Hydrology Research*, 48(3), 840–850. <https://doi.org/10.2166/nh.2016.072>
- Narany, S., Tahoora, M. F. R., Aris, A. Z., Sulaiman, W. N. A., Jauhir, H., & Fakharian, K. (2014). Identification of the hydrogeochemical processes in groundwater using classic integrated geochemical methods and geostatistical techniques, in Amol-Babol Plain, Iran. *The Scientific World Journal*. <https://doi.org/10.1155/2014/419058>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nasr, M., & Zahran, H. F. (2014). Using of PH as a tool to predict salinity of groundwater for irrigation purpose using artificial neural network. *The Egyptian Journal of Aquatic Research*, 40(2), 111–115. <https://doi.org/10.1016/j.ejar.2014.06.005>
- Nikoo, M. R., & Mahjouri, N. (2013). Water quality zoning using probabilistic support vector machines and self-organizing maps. *Water Resources Management*, 27(7), 2577–2594. <https://doi.org/10.1007/s11269-013-0304-5>
- Nolan, B. T., Fienen, M. N., & Lorenz, D. L. (2015). A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology*, 531, 902–911. <https://doi.org/10.1016/j.jhydrol.2015.10.025>
- Norouzi, H., & Moghaddam, A. A. (2020). Groundwater quality assessment using random Forest method based on groundwater quality indices (case study: Miandoab plain aquifer, NW of Iran). *Arabian Journal of Geosciences*, 13(18). <https://doi.org/10.1007/s12517-020-05904-8>
- Prajapati, P., and Parekh, F.. 2016. Analysis of groundwater quality parameters using Mamdani fuzzy inference system (MFIS). *International Journal of Science and Research (IJSR)* 5(1), 1335–40. doi: <https://doi.org/10.21275/v5i1.nov153010>.
- Prasad, B., Kumari, P., Bano, S., & Kumari, S. (2014). Ground water quality evaluation near mining area and development of heavy metal pollution index. *Applied Water Science*, 4(1), 11–17. <https://doi.org/10.1007/s13201-013-0126-x>
- Prasanna, M. V., Chidambaram, S., Shahul Hameed, A., & Srnivasamoorthy, K. (2011). Hydrogeochemical analysis and evaluation of groundwater quality in the Gadilam River basin, Tamil Nadu, India. *Journal of Earth System Science*, 120(1), 85–98. <https://doi.org/10.1007/s12040-011-0004-6>
- Purkait, B., Kadam, S. S., & Das, S. K. (2008). Application of artificial neural network model to study arsenic contamination in groundwater of Malda District , Eastern India. *ISEIS Journal of Environmental Informatics*, 12(2), 140–149. <https://doi.org/10.3808/jei.200800132>
- RadFard, M., Seif, M., Hashemi, A. H. G., Zarei, A., Saghi, M. H., Shalyari, N., Morovati, R., Heidarinejad, Z., & Samaei, M. R. (2019). Protocol for the estimation of drinking water quality index (DWQI) in water resources: Artificial neural network (ANFIS) and arc-Gis. *MethodsX*, 6, 1021–1029.
- Raghavendra, N., Sujay, & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19, 372–386. <https://doi.org/10.1016/j.asoc.2014.02.002>
- Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Ahmad, B. B., & Bui, D. T. (2019). Predicting uncertainty of machine learning models for Modelling nitrate pollution of groundwater using Quantile regression and UNEEC methods. *Science of the Total Environment*, 688, 855–866. <https://doi.org/10.1016/j.scitotenv.2019.06.320>
- Rajaei, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level

- modeling. *Journal of Hydrology*, 572, 336–351. <https://doi.org/10.1016/j.jhydrol.2018.12.037>
- Ransom, K. M., Nolan, B. T., Traum, J. A., Faunt, C. C., Bell, A. M., Gronberg, J. A. M., Wheeler, D. C., Rosecrans, C. Z., Jurgens, B., Schwarz, G. E., Belitz, K., Eberts, S. M., Kourakos, G., & Harter, T. (2017). A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. *Science of the Total Environment*, 601–602, 1160–1172. <https://doi.org/10.1016/j.scitotenv.2017.05.192>
- Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (southern Spain). *Science of the Total Environment*, 476–477, 189–206. <https://doi.org/10.1016/j.scitotenv.2014.01.001>
- Rodriguez-Galiano, V. F., Luque-Espinar, J. A., Chica-Olmo, M., & Mendes, M. P. (2018). Feature selection approaches for predictive Modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Science of the Total Environment*, 624, 661–672.
- Sahoo, G. B., Ray, C., Mehnert, E., & Keefer, D. A. (2006). Application of artificial neural networks to assess pesticide contamination in shallow groundwater. *Science of the Total Environment*, 367(1), 234–251. <https://doi.org/10.1016/j.scitotenv.2005.12.011>
- Sakizadeh, M. (2015). Artificial intelligence for the prediction of water quality index in groundwater systems. *Modeling Earth Systems and Environment*, 2(1), 8. <https://doi.org/10.1007/s40808-015-0063-9>
- Sasikaran, S., Sritharan, K., Balakumar, S., & Arasaratnam, V. (2016). Physical, chemical and microbial analysis of bottled drinking water. *Ceylon Medical Journal*. <https://doi.org/10.4038/cmj.v57i3.4149>
- Selakov, A., Cvijetinović, D., Milović, L., Mellon, S., & Bekut, D. (2014). Hybrid PSO–SVM method for short-term load forecasting during periods with significant temperature variations in City of Burbank. *Applied Soft Computing*, 16, 80–88. <https://doi.org/10.1016/j.asoc.2013.12.001>
- Selvaraj, A., Saravanan, S., & Jennifer, J. J. (2020). Mamdani fuzzy based decision support system for prediction of groundwater quality: An application of soft computing in water resources. *Environmental Science and Pollution Research*, 27(20), 25535–25552. <https://doi.org/10.1007/s11356-020-08803-3>
- Singh, R. M., & Datta, B. (2004). Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network. *Environmental Forensics*, 5(3), 143–153. <https://doi.org/10.1080/15275920490495873>
- Singh, K. P., Gupta, S., & Mohan, D. (2014). Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. *Journal of Hydrology*, 511, 254–266. <https://doi.org/10.1016/j.jhydrol.2014.01.004>
- Singha, S., Pasupuleti, S., Singha, S. S., & Kumar, S. (2020). Effectiveness of groundwater heavy metal pollution indices studies by deep-learning. *Journal of Contaminant Hydrology*, 235, 103718. <https://doi.org/10.1016/j.jconhyd.2020.103718>
- Singha, S., Pasupuleti, S., Singha, S. S., Singh, R., & Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, 130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>
- Sirat, M. (2012). Neural network assessment of groundwater contamination of US mid-continent. *Arabian Journal of Geosciences*, 6(8), 3149–3160. <https://doi.org/10.1007/s12517-012-0570-1>
- Smith, D. G. (1990). A better water quality indexing system for Rivers and streams. *Water Research*, 24(10), 1237–1244. [https://doi.org/10.1016/0043-1354\(90\)90047-A](https://doi.org/10.1016/0043-1354(90)90047-A)
- Sobhanardakani, S. (2016). Evaluation of the water quality pollution indices for groundwater resources of Ghahavand Plain, Hamadan Province, Western Iran. *Iranian Journal of Toxicology* 10 (3), 35–40. doi: <https://doi.org/10.29252/arakmu.10.3.35>.
- Srinivas, R., Prashant Bhakar, and Ajit Pratap Singh. 2015. Groundwater quality assessment in some selected area of Rajasthan, India using fuzzy multi-criteria decision making tool. *Aquatic Procedia* 4(Icwrcoe):1023–30. doi: <https://doi.org/10.1016/j.aqpro.2015.02.129>.
- Srinivasamoorthy, K., Vasanthavignar, M., Vijayaraghavan, K., Sarathidasan, R., & Gopinath, S. (2013). Hydrochemistry of groundwater in a coastal region of Cuddalore District, Tamilnadu, India: Implication for quality assessment. *Arabian Journal of Geosciences*, 6(2), 441–454. <https://doi.org/10.1007/s12517-011-0351-2>
- Stackelberg, P. E., Belitz, K., Brown, C. J., Erickson, M. L., Elliott, S. M., Kauffman, L. J., Ransom, K. M., & Reddy, J. E. (2021). Machine learning predictions of PH in the glacial aquifer system, northern USA. *Groundwater*, 1–17. <https://doi.org/10.1111/gwat.13063>
- Sunayana, K. K., Dube, O., & Sharma, R. (2020). Use of neural networks and spatial interpolation to predict groundwater quality. *Environment, Development and Sustainability*, 22(4), 2801–2816. <https://doi.org/10.1007/s10668-019-00319-2>
- Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3), 293–300.
- Swarna Latha, P., & Nageswara Rao, K. (2012). An integrated approach to assess the quality of groundwater in a coastal aquifer of Andhra Pradesh, India. *Environmental Earth Sciences*, 66(8), 2143–2169. <https://doi.org/10.1007/s12665-011-1438-5>
- Tabach, E., Eddy, L. L., Shahrour, I., & Najjar, Y. (2007). Use of artificial neural network simulation Metamodelling to assess groundwater contamination in a road project. *Mathematical and Computer Modelling*, 45(7), 766–776. <https://doi.org/10.1016/j.mcm.2006.07.020>
- Tiyasha, T. M. T., & Yaseen, Z. M. (2020). A survey on river water quality Modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670. <https://doi.org/10.1016/j.jhydrol.2020.124670>
- Tutmez, B., Hatipoglu, Z., & Kaymak, U. (2006). Modelling electrical conductivity of groundwater using an adaptive Neuro-fuzzy inference system. *Computers and Geosciences*, 32(4), 421–433. <https://doi.org/10.1016/j.cageo.2005.07.003>
- Vadiati, M., Asghari-Moghaddam, A., Nakhaei, M., Adamowski, J., & Akbarzadeh, A. H. (2016). A fuzzy-logic based

- decision-making approach for identification of groundwater quality based on groundwater quality indices. *Journal of Environmental Management*, 184, 255–270. <https://doi.org/10.1016/j.jenvman.2016.09.082>
- Varotsos, C. A., & Krapivin, V. F. (2018). Arctic waters has reached a critical point: An innovative approach to this problem. *Water, Air, & Soil Pollution*, 229(11), 343. <https://doi.org/10.1007/s11270-018-4004-x>
- Varotsos, C. A., Krapivin, V. F., & Mkrtchyan, F. A. (2019). New optical tools for water quality diagnostics. *Water, Air, and Soil Pollution*, 230(8). <https://doi.org/10.1007/s11270-019-4228-4>
- Varotsos, C.A., Krapivin, V.F., and Mkrtchyan, F.A.. (2020a). On the recovery of the water balance. *Water, Air, & Soil Pollution* 231(4):170. doi: <https://doi.org/10.1007/s11270-020-04554-6>.
- Varotsos, C. A., Krapivin, V. F., Mkrtchyan, F. A., Gevorkyan, S. A., & Cui, T. (2020b). A novel approach to monitoring the quality of lakes water by optical and modeling tools: Lake Sevan as a case study. *Water, Air, and Soil Pollution*, 231(8). <https://doi.org/10.1007/s11270-020-04792-8>
- Varotsos, C. A., Krapivin, V. F., Mkrtchyan, F. A., & Xue, Y. (2021). Optical spectral tools for diagnosing water media quality: A case study on the Angara/Yenisey River system in the Siberian region. *Land*, 10(4).
- Vijay, S., & Kamaraj, K. (2019). Ground water quality prediction using machine learning algorithms in R. *International Journal of Research and Analytical Reviews*, 6(1), 743–749.
- Wagh, V. M., Panaskar, D. B., Muley, A. A., Mukate, S. V., Lolage, Y. P., & Aamalawar, M. L. (2016). Prediction of groundwater suitability for irrigation using artificial neural network model: A case study of Nanded tehsil, Maharashtra, India. *Modeling Earth Systems and Environment*, 2(4), 1–10. <https://doi.org/10.1007/s40808-016-0250-3>
- Wagh, Vasant Madhav, Dipak Baburao Panaskar, and Aniket Avinash Muley. 2017. "Estimation of nitrate concentration in groundwater of Kadava River basin-Nashik District, Maharashtra, India by using artificial neural network model." *Modeling Earth Systems and Environment* 3(1):0. doi: <https://doi.org/10.1007/s40808-017-0290-3>.
- Wang, M. X., Liu, G. D., Wu, W. L., Bao, Y. H., & Liu, W. N. (2006). Prediction of agriculture derived groundwater nitrate distribution in North China plain with GIS-based BPNN. *Environmental Geology*, 50(5), 637–644. <https://doi.org/10.1007/s00254-006-0237-x>
- Yang, Q., Zhang, J., Hou, Z., Lei, X., Tai, W., Chen, W., & Chen, T. (2017). Shallow groundwater quality assessment: Use of the improved Nemerow pollution index, wavelet transform and neural networks. *Journal of Hydroinformatics*, 19(5), 784–794. <https://doi.org/10.2166/hydro.2017.224>
- Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K.-W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, 569, 387–408. <https://doi.org/10.1016/j.jhydrol.2018.11.069>
- Yesilnacar, M. I., & Sahinkaya, E. (2012). Artificial neural network prediction of sulfate and SAR in an unconfined aquifer in southeastern Turkey. *Environmental Earth Sciences*, 67(4), 1111–1119.
- Yoon-Seok Hong, Michael R. Rosen, and Institute. (2001). Intelligent characterisation and diagnosis of the groundwater quality in an urban fractured-rock aquifer using an artificial neural network. *Urban Water* 3, 53(4), 551–562. <https://doi.org/10.1111/j.1467-8500.1994.tb01504.x>
- Zaqoot, H. A., Hamada, M., & Miqdad, S. (2018). A comparative study of ANN for predicting nitrate concentration in groundwater Wells in the southern area of Gaza strip. *Applied Artificial Intelligence*, 32(7–8), 727–744.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.