Knowledge-based Engineering and Sciences (2023), **4**, (3), 65–103
doi:10.51526/kbes.2023.4.3.65-103

REVIEW PAPER

A New Benchmark on Machine Learning Methodologies for Hydrological Processes Modelling: A Comprehensive Review for Limitations and Future Research Directions

Zaher Mundher Yaseen*

Civil and Environmental Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
*Corresponding author. Email: z.yaseen@kfupm.edu.sa

(Received 17 November 2023; revised 24 December 2023; accepted 30 December 2023; first published online 31 December 2023)

Abstract

The best practice of watershed management is through the understanding of the hydrological processes. As a matter of fact, hydrological processes are highly associated with stochastic, non-linear, and non-stationary phenomena. Hydrological processes simulation and modeling are challenging issues in the domains of hydrology, climate and environment. Hence, the development of machine learning (ML) models for solving those complex hydrological problems took essential place over the past couple decades. It can be observed, hydrological data availability has increased remarkably, and thus computational resources has led to a resurgence in ML models' development. It has been witnessed huge efforts on the hydrological processes modeling using the facility of ML models and several review researches have been conducted. Literature studies approved the capacity of ML models in the field of hydrology over the classical "traditional models" based on their forecastability, flexibility, precision, generalization, and modeling execution convergence speed. However, although several potential merits were observed in ML model's development, several limitations are allied such as the interpretability of those black-box models, the practicality of the ML models in watershed management, and difficulty to explain the physical hydrological processes. In this survey, an exhibition for all the published review articles on the development of ML models for hydrological processes and recognize all the research gaps and potential research direction. The ultimate aim of the current survey is to establish a new milestone for the interested hydrology, environment and climate researchers on the applications of ML models.

Keywords: Hydrological processes; Machine learning models; Research gaps; Future research directions.

1. Introduction

1.1 Interest for watershed management

Watershed characteristics and management have a complicated and multidimensional interaction. A watershed's features, including its topography, geology, hydrology, and vegetation, are important factors in identifying the management approaches that might be employed to preserve and improve the watershed [1], [2]. For instance, a watershed's hydrology, which includes its flow patterns and water storage capacity, can be influenced by its topography and geology. Determining the best management tactics to safeguard watersheds from erosion, flooding, and other hydrological risks requires knowledge of this information [3]. A watershed's vegetation can be extremely important as well to its management; for instance, trees and other vegetation can remove contaminants from runoff, thereby assisting to stabilise the soil, reduce erosion, and preserve the quality of the water [4]. They also support the watershed's biodiversity by giving wildlife somewhere to live. A watershed's hydrology plays a significant role in the watershed management. Water storage and flow patterns in a watershed can impact the water quality, health of aquatic ecosystems, erosion and sedimentation in the watershed [5], [6]. To reduce adverse effects on water resources, optimum management methods for urban and agricultural development, as well as the placement of water retention structures, can all be informed by an understanding of hydrology [7]. For watershed management plans to be effective, the watershed's specific features must be considered. In a watershed with a high-water table, for instance, a wetland restoration project would be different from one with a low water table.

Making decisions about water resource management involves critical thinking, critical knowledge, and wisdom in addition to interdisciplinary approaches for identifying potential actions and assessing their impacts [8], [9]. The physical science, along with technology, aid in managers' understanding of water resources and environmental issues. Satellite data-based dashboard can serve as a great decision-makers guide throughout the entire process [10]. In other words, a decision-making system can be established in the form of dashboard based on geographic information system, an interactive computer system created to assist decision-makers in using information, documents, data, communication technologies, and models to complete the tasks involved in the decision-making process. In conclusion, a watershed's features—such as its topography, geology, hydrology, and vegetation—are crucial in defining the management approaches that can be applied to preserve and improve the watershed. The creation of efficient management methods that are suited to the peculiar features of a watershed requires an understanding of the distinctive qualities of each watershed.

1.2 The concept of hydrological cycle

Water movement on, above, and below the surface of the earth is called the hydrological cycle or the water cycle [11]. The interdependent and interrelated processes that comprise the hydrological cycle are essential to preserving the planet's water balance. The hydrological cycle's primary processes include transpiration, evaporation, infiltration, and runoff [12]. The process by which water falls from the atmosphere as rain, snow, sleet, or hail is referred to as precipitation. This water can either run off into surface water bodies like rivers and lakes or seep into the ground to replenish groundwater aquifers. The process that releases water vapour into the atmosphere from the Earth's surface is called evaporation [13]. The ocean's surface is the main source of this activity, although it can also come from lakes, rivers, soil, and vegetation. The process by which plants absorb water and release it into the atmosphere through microscopic holes in their leaves is known as transpiration. This mechanism aids in controlling the amount of moisture in the atmosphere and on the surface of the land, making it crucial for the water cycle. Water from surface runoff or precipitation penetrates the ground and integrates into groundwater aquifers through a process called infiltration; this is an essential process for replenishing the groundwater supplies, which are vital to both natural and human systems. Again, water moves from the land surface into surface water bodies, such as rivers and lakes through a process called runoff. This process is crucial for moving water and dissolved nutrients from the land to the oceans, as well as for preserving the water balance in these systems. All these processes are interrelated and dependent on one another; for instance, transpiration and evaporation contribute to the atmosphere's water balance; infiltration and groundwater recharge are necessary for the land surface's water balance; precipitation and runoff are crucial for the surface water bodies' water balance. Understanding the interdependencies among these mechanisms is crucial for the appropriate administration of water resources, as well as for appreciating the effects of climate change and human endeavours on the hydrological cycle [14].

1.3 The motivation of machine learning implication for hydrological process

The hydrology cycle is a complicated nonlinear system typified by processes and events whose dynamics, in addition to the fundamental human interactions, depend on numerous direct causes (such as meteorological and environmental conditions) [15]. As a result, it is commonly known that conceptual, systematic, and results uncertainty are common in hydrological models [16]. The aforementioned facts have led to significant research efforts to build various model structures, such as mathematical, statistical, physical, stochastic, and numerical models. Amidst many technologies and approaches, Machine Learning (ML) models—highly sophisticated nonlinear computer-assisted models with the ability to extract features, patterns, or rules from datasets—have been exponentially growing in use for data-intensive hydrological modelling tasks [17], [18]. Many nations all around the world have experienced drought and water stress in recent years [19]. Water resources are critically declining due to the extreme lack of precipitation and the excessive demand, putting the food security of many nations in jeopardy. Many scientists are considering sustainable ways to enhance the management of water resources and sustainability as a result of this global issue [20]. Some of the more creative technological solutions that have come up include the intriguing ideas for increased automation and flexibility in computer-assisted models.

The application of systems that can assist decision-making in matters like watershed management and operation has opened up new possibilities in the realm of soft computing thanks to the emergence of ML models, Big Data, and high-performance computing approaches [21], [22]. "Learning refers to any process by which a model improves performance from experience," as per Herbert A. Simon. The information could contain secret knowledge that clarifies the principles, laws, and logic of a complicated occurrence. Therefore, ML models are informatively defined as technologies that can automatically extract relevant information from data through advanced analysis. ML models can be used to replicate various hydrological processes considering their inherent capabilities [23]. These algorithms fall into four categories based on the type of training data they use; these are supervised learning, unsupervised learning, reinforcement learning, and semi-supervised learning [24].

With supervised learning, an unknown (input and output) is predicted from known instances (input and output), where the output is labelled. The computer will pick up information more quickly from known samples. These tasks can be classified into two categories: regression, which predicts a certain point on a numerical axis, and classification, which predicts an object's category [25]. In essence, regression is classification in which a number is predicted rather than a category. The Naive Bayes, Decision Tree, Logistic Regression, K-Nearest Neighbours, and Support Vector Machine algorithms are the most often used ML models for supervised learning tasks. Unlike supervised learning, the task of unsupervised learning involves providing the learning system with merely input samples. Considering that the data is unlabelled, the modelling task is just aimed at presenting the information in a way that will enhance understanding. This is frequently called generalisation, dimensionality reduction, or clustering. The commonly used techniques for unsupervised learning are K-means clustering, mean-shift, Singular Value Decomposition (SVD), Principal Component Analysis (PCA), and Latent Dirichlet Allocation (LDA). The basis of reinforcement learning is feedback; only in a dynamic setting is training data sent to a system as feedback. The feedback between the learning system and the interaction experience aids in improving the performance of the task to be learned.

1.4 The survey inspiration and objectives

Hydrologists have been always pursuing research for reliable solutions for future prediction/forecasting of hydrological processes e.g., river flow, rainfall, evaporation, evapotranspiration, infiltration, groundwater level, etc. In addition, other associated variables such as soil temperature/moisture, landslide, and water quality [26], [27]. Physical based models are usually associated with evident limitations such as the serious need to field observations and studying the uniqueness of specific watershed characteristics [28]. On the other hand, mathematical models are incorporated with uncertainties and parametrizing tuning [29]. Over the past couple decades, a massive attention was devoted to promote the feasibility of computer aid models "i.e., machine learning" in the field of the hydrology, climate, environment and earth science [14], [30]–[32]. ML models have been widely introduced to solve hydrological processes and various versions explored such as artificial neural network, support vector machine, genetic programming, fuzzy logic. Nevertheless, the recently explored version of deep learning (DL) models which are vantage to be the mainstream in hydrological processes prediction and leverage to be predominant in hydrology as a tool [33]–[35].

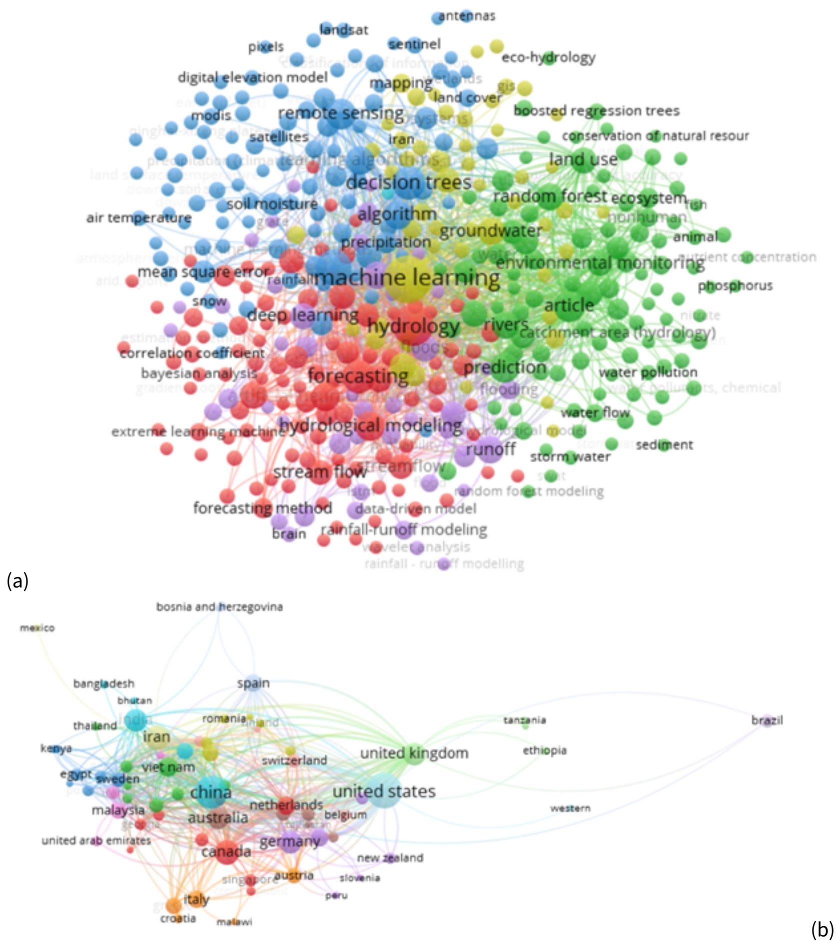

Figure 1: *(a) The main keywords occurrence for the scopus search of machine learning and hydrology; (b) Interested countries in the domain of the hydrological applications using the potential of machine learning models.*

By exploring the Scopus database for the keywords search "machine learning AND hydrology", 632 research articles displayed for this search. Although, ML is not essentially dated back term for few decades back; however, the research outcome clearly revealed the popularity of those advanced technologies in the domain of hydrology. The total number of keywords for the research results was 5975, Figure 1 displayed the occurrence of 6 keywords with 532 keywords. What can be understood from this biographical presentation, ML models have been explored nearly to all related hydrological process that are possibility experienced at certain watershed area (Figure 2). Over 80 countries have been noticed to show interest in the applicability of ML models and their remarkable solution for related hydrological processes. This was one of the significant motivations for the current review article where the current direction would be more important to be recognized for hydrological scientists and data science developers.

Hydrological sciences have seen a noticeable shift in the past few decades towards the use of computer aid models "machine learning", mostly for forecasting, prediction, and optimisation. There is a global movement encouraging scientists to consider environmentally friendly solutions in the field of hydrology by utilising cutting-edge technology like ML and the Internet of Things (IoT). Different approaches have been developed and used, and they are suitably designed as a result of the variety of ML models used in hydrological sciences. The goal of this review is to establish new standards for ML applications in the field of hydrology. The primary goal of the study is to identify the critical views that interested hydrological scholars need to know to fill the research gap. The current review's general outcome shows that ML models are superior, can generalise, are unique, and that awareness should be given serious consideration.

Figure 2: *The hydrological processes simulated over the literature review over the past couple decades using machine learning models.*

2. Literature review

In the current section, Table 1 presents the summary of the surveyed review papers and identifies their number of reviewed papers, topics coveted, time spam for the survey, ML models, presented limitations, and recognized possible future research directions.

Table 1: *The surveyed review articles on the applications of machine learning models for simulating hydrological processes.*

Ref.	No. Papers	Topics	Time Span	Models	Research Limitations	Possible Future Research Direction
[36]	96	> Forecasting water level in surface, water formations. > Creating models, mapping, and evaluating the risk of floods. > Simulation of sediment movement within river networks. > Anticipating water demand in urban areas. > Simulation of fluid movement through hydraulic structures. > Modeling fluid and sediment dynamics within sewer systems	2000-2019	> SWIM > ANN > MLPNN > GRNN > RBNN > GMDH > ELM > PSO-RBNN > Hybrid-RBNN > DT > GRU > MLP > LSTM > WNN	> ML models require long historical data and high-quality data for optimal performance. > The black box nature of neurocomputing models hinders the explicit identification of inner physical relations, limiting certain applications. > One of the Limitation in ML models' is the competence in the applications with insufficient data, particularly in estimating flood dynamics and sediment concentrations in poorly gauged or ungauged catchments. > Few studies focus on long-term forecasts in sediment transport modeling, with challenges in realistic forecasting due to the use of 'future' information in historical re-forecasting. > Limited neurocomputing models for supporting long-term predictions in operational hydrodynamics and morpho-dynamics forecasting. > Challenges in generalization and accuracy of neurocomputing models for water demand prediction, with models often trained and validated on specific case studies. > Difficulty in cross-comparing results and limited validity for different settings in water demand prediction models.	> Explore novel ML models (e.g., deep learning) for hydrological and hydraulic sciences. > Ensure robust supervision of ML models, e.g., integrating ML with optimization algorithms. > Conduct a comparative study on the efficacy of soft computing models versus hard computing models in hydrological sciences. > Expand research topics to include groundwater modeling, irrigation systems, water quality simulation, precipitation forecast, evaporation estimation, and rainfall runoff processes. > Integrate neurocomputing models with GIS for a seamless link between pre-processing andpost-processing. > Investigate the hybridization of neurocomputing models with other soft computing concepts for improved performance. > Invest in research for the development of more efficient and accurate hybrid models. > Encourage further research on the application of DL models (e.g., ESN, CNN) in various hydrological and hydraulic scenarios. > Need for future comparative studies to assess the sensitivity, portability, and robustness of state-of-the-art neurocomputing models across different case studies and uncertainties.

[37]	NA	> Cultural Barriers: Explores ML-PBM differences in EES and examines the impact on objectives like prediction. > Hybrid Models: Merges ML-PBM for effectiveness and advocates co-creation over borrowing. > Knowledge-Driven ML: Addresses ML challenges in EES and emphasizes knowledge-driven and data-driven approaches	NA	> DL	> Cultural barriers pose challenges to collaboration between ML and PBM. > Differences in modeling objectives limit the applicability of models in specific contexts. > ML's lack of interpretability hinders effective communication of modeling outcomes. > PBM scalability requires substantial modifications for diverse datasets. > Limited process representation in PBM poses challenges in capturing real-world system complexities. > The absence of model coevolution hampers the potential for transformative innovations.	> Explore coevolutionary modeling to integrate ML and PBM strengths. > Develop models tailored for complex, multi-dimensional spaces to enhance understanding. > Focus on estimating and quantifying uncertainty in both ML and PBM models to improve overall model certainty. > Explore initiatives to break cultural barriers and foster collaboration between ML and PBM communities. > Develop interdisciplinary education programs for ML and Earth and Environmental Sciences to ensure sustainable training.

[38]	160	> Overview of ANN models in hydrological variable forecasting. > Components of ANN modeling: data preprocessing, input determination, model characteristics, and assessment methods. > Hybrid ANN models: introduction, taxonomy, and practical applications. > Current obstacles in using ANN models in hydrology. > Recommendations in hydrological variable forecasting using ANN models.	1998-2015	> ANN modeling for hydrological applications. > Using soft computing approaches for hydrological variable modeling. > Hybrid ANN models in hydrology. > Hybrid ANN models with focus of data intensive. > Hybrid ANN models with focus of model intensive. > Hybrid ANN models with focus of technique intensive.	> The application of new approaches in ANN is usually limited to specific indices, hindering their assessment across diverse architectures and variables. > Restricted the availability of appropriate and long-term data for water quality parameters presents a challenge for developing models and techniques. > Evaluation of hybrid ANN models influenced by the absence of standardized methodologies for combining ANN with alternative models and techniques. > Shortage of a systematic method for identifying the number of hidden layers in ANN models presents a limitation in the current state of research.	> Develop systematic approaches for determining the number of hidden layers in ANN models. > Create new hybrid models by combining ANN with alternative models and techniques from advanced research in various engineering fields. > Apply recent approaches to develop new models and techniques for short-term or missing data, particularly for water quality parameters. > Extend the application of new algorithms and methods to various architectures of ANN models for a comprehensive evaluation. > Conduct new studies to find the optimum number of neurons systematically for each type of hydrological variable, moving beyond the trial-and-error approach.

[39]	NA	> A workflow to address pitfalls and challenges in applying ML models to hydrology.	NA	> ANN > SVM > ELM > RBF	> Increasing the number of observations substantially raises computational demands, posing a limitation in terms of scalability.] > Workflow may present challenges in terms of interpretability due to the inherent complexity of machine learning models. > There is a potential for bias in the estimation of prediction accuracy, as evidenced by the need to empirically estimate bias in the results. > The study relies on standard statistical metrics for model evaluation, which may have limitations in capturing all aspects of model performance.	> Understanding variable selection algorithms' tendencies with changes in training data. > Exploring theory-guided data science to enhance ML model interpretability. > Extending the workflow's application to finer temporal resolutions and encouraging experimentation in forecasting various processes. > Computational complexity of the proposed workflow, especially with regards to variable selection algorithms. > Impact of increasing the number of observations and modeling at finer temporal resolutions on workflow runtime.
[40]	101	> Streamflow forecasting using AI-Models > Streamflow modeling with ANN > SVM approach for streamflow forecasting > FL for streamflow prediction > Evolutionary computing methods in streamflow modeling > Wavelet-complementary modeling for streamflow prediction	2000 - 2015	> ANN > SVM > Fuzzy Logic > EC (GA, PSO) > W-AI	> ANN models face challenges like slow learning, local minima, and overfitting. > RBFNN models have limitations in short database forecasting. > DNN models encounter difficulties accurately modeling hydrological data. > SVM models struggle in short-term scenarios. > Fuzzy Logic models have limitations in handling complexity and interval data.	> Development of a new architecture for streamflow forecasting to enhance prediction accuracy and efficiency. > Preprocessing time series frequency, with a focus on integrating FOS techniques for improved data preparation. > Application of SI as a modern optimization approach for refining forecasting models and enhancing overall system performance
[41]	42	> AI-RS Publications trends. > RS data evolution. > Applications of ML methods in processing remote sensing data for mineral exploration.	2003 - 2021	> ML (SVM, DL, ANN, RF, Clustering, Regression analysis, Dimensional reduction technics)	> Data-driven techniques may risk over-fitting and problem dependence. > The complexity and high dimensionality of ML algorithms pose challenges for geologists in their application to geological data. > ML models still face a challenge in accurately quantifying uncertainty in predictions. > Complexity in ML models, coupled with a lack of clear guidelines for settings and result interpretation, presents challenges	> Investigation of uncertainty in predictions within remote sensing-based model applications. > GANs can be used to tackle class-imbalanced issues in geo-science and RS by generating or reconstructing data for imbalanced classes and missing regions. > Accuracy-computational effort trade-off in data-driven techniques, addressing risks like over-fitting and problem dependence through additional research.

[30]	NA	> ML Applications in Hydrology	NA	> ML (KNN, Regularized Linear Models, ANN, CNN, SVM, DTL, RF, GBM, M5 and M5-cubist, SG)	> Hydrological datasets often have correlations, impacting model tuning, and cross-validation. > In hydrology, comparisons usually rely on basic metrics, possibly missing crucial aspects. > ML growth in hydrology is tied to data availability challenges. > Emphasizing model generalization, a perfect fit on training data may not extend well. > Hydrological processes' complexity increases different method comparisons.	> ML and DL gain importance in hydrology, with a focus on reproducibility through code sharing. > Explore synergies between ML and PB models, prompting new questions in mechanistic modeling. > The low maturity of DL prompts extensive research, especially in the context of learning from large unsupervised data for "Big Data Hydrology." > Anticipated rapid expansion of ML and DL in hydrology includes ongoing investigations into SVMs, CNNs, and RF, with expectations for new algorithms. > Successful big data analysis in hydrology relies on continuous monitoring and extension of hydrological data, particularly in revealing patterns over decades.
[14]	NA	ML as a powerful tool to help society to adapt the climate changes in various ways such as: > Electricity systems > transportation > buildings and cities > industry, farms and forests > carbon dioxide removal > climate prediction, societal impacts > solar geoengineering > individual action > collective discussion > education > finance	NA	> ML	> Transportation lies in its insufficient progress in reducing CO2 emissions, with much of the sector deemed challenging to decarbonize. > ML requires the need for data access and cleaning, and managerial caution due to potential costly consequences, favoring risk-averse strategies. > Uncertainty and lack of clear guidance for ML practitioners wish to apply ML to tackle climate change despite their recognized power for technological progress. > The highlighted impactful applications do not constitute a single fix for climate change. > There are areas where ML is inapplicable to climate change solutions. > ML can be applied in ways that exacerbate the issue of climate change. It is widely used to expedite activities like fossil fuel exploration, and some ML models are energy-intensive to train and run.	> Enhance ML for flexible energy systems, accelerating the transition to carbon-free sources. > Leveraging ML to optimize strategies for reducing transport activity, improving vehicle efficiency, exploring alternative fuels and electrification, and facilitating modal shifts towards lower-carbon options. > Leveraging ML to enhance energy efficiency strategies by modeling data on energy consumption and optimizing energy use in smart buildings. > ML for industrial emissions reduction shows promise when there is accessible high-quality data, firms share information, processes are adjustable, and incentives align with emission reduction goals. > ML can be used to monitor forest and peatland health, predicting fire risks, and promoting sustainable forestry, emphasizing alignment with decarbonization goals.

[42]	180	> Review of employing ML Models for Predicting Flooding Events	2008 - 2017	> ANN > MLP > ANFIS > WNN > SVM > DT > EPS	> ML modeling for flood prediction is in its early stages, limiting the maturity and robustness of existing models. > The success of novel ML models is contingent on the effective use of soft computing techniques, potentially limiting advancements if not properly implemented. > ANNs in flood modeling has drawbacks like low accuracy, parameter tuning challenges, slow learning response, and difficulties in physical interpretation, especially in precipitation and peak-value prediction.	> Explore advanced hybridization and ensemble techniques for improved performance and robustness in flood prediction models. > Investigate the application of data decomposition techniques to enhance the accuracy of flood prediction datasets. > Explore the use of ensemble methods and optimizer algorithms to improve model generalization and reduce uncertainty in flood prediction. > Incorporation of soft computing techniques for designing novel learning algorithms in flood prediction models. > Research on spatial flood prediction using ML models to address the unique challenges in predicting flood locations.
[43]	NA	> Data-driven Modeling and Computational Intelligence Methods in Hydrology	NA	> M5 MT > ANN > SVM > FRBS	> Demand for computer memory in FRBS increases exponentially with a growing number of input variables. > The effectiveness of DDM in a scientific domain area depends on factors such as the availability of a large sample of data.	> Combine diverse models, integrating physical models for synergy in hybrid approaches > Use computational intelligence for optimal, adaptive structures in hybrid models.
[44]	129	The study incorporates DL methods in the water sector, focusing on: > monitoring > management > governance > communication of water resources.	2018 - 2020	> DL	> ANN Tasks may require impractically large hidden layers. > The water domain lacks quality benchmark datasets, hindering collaboration and model improvement. > Water data from authorities are dispersed with temporal and spatial mismatches, making acquisition challenging. > Studies claiming DL use often rely on traditional ANN approaches, raising doubts about DL application. > DL and water studies lack method details, slowing progress and reproducibility.	> Collaborate for DL-based hydrological forecasting using real-time data. > Create open datasets to overcome the lack of DL-ready water field data. > Utilize centralized AI frameworks for custom solutions in hydrological applications. > Implement edge computing for sensor data processing, encouraging innovation. > Use intelligent assistants to extract knowledge from vast hydrological datasets. > Integrate DL with virtual and augmented reality for immersive experiences in hydrological analysis.

[45]	NA	> How well do ML models perform without hydrologists? > Applying rational feature selection for enhanced hydrological forecasting.	NA	> MLR > M5P-MT > MLP > ANN > LSTM	> ML models, including ANNs, lack clear explanations for predictions and hydrological processes. > Performance deteriorates with increasing model complexity. > ML models exhibit poor performance, falling below the predictability threshold on a 5-day lead time. > Attempts to build a universal, data-driven hydrological model with automatic structure selection have been unsuccessful.	> Explore methods to enhance interpretability of nonlinear hydrological models. > Develop visualization techniques for complex models, particularly for black-box models like artificial neural networks. > Conduct comparative analyses of different model structures, especially in contrasting catchment conditions. > Test hypotheses on the efficiency of rational feature selection in catchments with specific characteristics. > Explore ways to incorporate hydrologist expertise in the model training process. > Assess ML model generalization across diverse natural conditions.
[46]	NA	> Applications of SVM in Hydrology	NA	> ML (SVM)	> SVM linearizes data through kernel transformation, making results accuracy independent of expert judgment for non-linear input data. > SVM heuristic selection of kernel function and hyperparameters, relying on a time-consuming trial-and-error process. > Nonlinear SVR model complexity hampers easy understanding and interpretation, resulting in a slower training process compared to linear models. > Poor model extrapolation occurs with past data inconsistency, as the model heavily relies on past records as support vectors. > SVM produces only point predictions and lacks design for probabilistic forecasts.	> Extend SVMs to tackle hydrologic inverse problems incorporating a physical understanding of geological processes, such as density estimation. > Explore SI techniques, like artificial bee colony and ant colony optimization, for hybridization with SVM to achieve global optimal results in parameter selection. > Investigate the Cloud SVM training mechanism in a cloud computing environment with MapReduce for large datasets, aiming for efficient and cost-effective watershed model calibration. > Focus on future hydrology modeling sophistication, encompassing a wide range of natural phenomena to understand watershed.

[47]	113	> ML in hydrology	2002 - 2021	> AdaBoost > XGB > ANN > MLR > BMA > SVM > W-ANN > SGB	> Limited data on soil and aquifer properties challenges hydrogeological modeling. > Ensemble methods are underused in hydrological drought modeling, with sparse literature on machine learning as base learners. > Ensemble techniques, particularly bagging and boosting, may show inefficiency in rare cases. > Advanced ensemble methods like Adaboost, XGB, and Dagging are still limited in recent studies. > Ensemble modeling mainly relies on tree structures, neglecting models like ANFIS, GMDH, GEP, deep echo state, and ELM.	> Explore diverse ML models in hydrological ensemble modeling. > Assess bagging and boosting methods' performance in hydrological modeling. > Integrate ensemble strategies for improved learning algorithm performance. > Extend ensemble learning to diverse hydrological sciences. > Appraise ensemble learning importance in various hydrological aspects. > Comprehensive comparisons between ensemble models, individual models, and common hydrological methods are lacking.
[48]	NA	> The article explored optimizing information flow, rules of use, and efficient data utilization, emphasizing ML methods like DL and active learning. > Specifically applied ML techniques to enhance water resources management.	NA	> ANN > GPR > SVM > Lasso > Clustering > DL	> Small datasets and climate-induced changes hinder adaptability. > Complex model relationships limit physical understanding and reliability. > Limited training data affects model performance across different regions. > Hydrology suffers from inadequate data representation due to scarcity. > Varying data types and accuracies pose challenges to ML models. > Many models lack methods to manage uncertainty in hydrology's limited data scenarios.	> Improve ML adaptability to changing conditions and limited hydrological data through enhanced generalization and spatial adaptability. > Develop inherently interpretable ML models specific to hydrology and geosciences for better decision-making. > Integrate physical knowledge into ML by blending process-based modeling with data-driven approaches. > Address data limitations and biases by refining learning strategies for scarce and imbalanced datasets.

[49]	36	> Understanding scales for better watershed management. > ML role in predicting river water quality. > Advanced ML applications in water quality modeling. > Enhancing ML model selection, explainability, uncertainty quantification. > Factors in using ML such as scale, data, resources, stakeholder needs.	2008 - 2021	> ML > ANN > LSTM > XGBoost > KGML	> ML models prioritize statistical relationships over physical consistency. > Limitations in extrapolating beyond available data. > Challenges integrating scientific knowledge into ML. > Struggles predicting extreme events. > Complex water quality data requires advanced ML representations.	> Advancements in model selection and hyperparameter optimization for improved water quality models. > Integration of domain knowledge into ML methodologies for more guided and accurate predictions. > Exploration of transfer learning techniques to leverage knowledge from related domains or datasets. > Development of new data representations suitable for handling complex water quality data. > Explicit treatment of extreme events within models for enhanced predictive capabilities. > Focus on uncertainty quantification techniques for more reliable predictions and assessments.
[50]	25	> This research emphasizes advancing AI's role in sediment transport applications	2001 - 2014	> ANN > ANFIS > SVM > Fuzzy Logic > Wavelet-AI Integrated Model	> ANN faces challenges like overfitting, slow learning, susceptibility to local minima, and struggles with complex, non-stationary, dynamic, and nonlinear time series. > SVM drawback is the selecting suitable kernel parameters, especially for Gaussian kernels and the insensitive loss function. > SVM training and testing sessions are time-consuming, unsuitable for real-time applications. > GA have long training times, hindering quick outcomes. > GA can be sluggish for real-time applications due to complex solutions. > Random convergence of solutions is a drawback in GA, impacting their effectiveness regardless of the fitness function. > GA consume considerable time due to the involvement of numerous parameters in the optimization process.	> Utilize OPLS and bidirectional O2PLS for data preprocessing to maintain crucial information while eliminating unwanted variations. > Develop two sub models based on discharge values to accurately capture sediment behavior variations. > Implement hybrid models like ANN-wavelet to handle non-stationary and complex time series data effectively, employing tools such as FOS for efficient noise elimination. > Enhance the backpropagation in ANN by employing PSO, ACO, and other algorithms to overcome local minima-maxima issues and improve sediment transport estimation models.

[51]	NA	> Overview of various AI modeling frameworks used in solving river sediment problems	NA	> ANN > SVM > Fuzzy > BN > Wavelet Transform > Nature-inspired hybrid artificial intelligence Models > Ensemble artificial intelligence models > MARS > CART > Regression model > M5MT	> MLR and SRC models try to map hysteresis behavior accurately in sea level relationships, making their comparison with highly non-linear AI models unreasonable. > Previous reliance on past SSL values as model inputs put practical challenges in data collection, particularly during extreme events. > AI models designed for SSL modeling lack applicability to different basins with different morphological and climatic features, limiting their generalization. > The black box nature of AI models confuses the interpretation of their physical foundations, requiring further analysis for a clearer understanding of parameter relationships. > High spatial and temporal variability, as well as skewed distributions in SSL and streamflow data used for modeling, put limitations to the effectiveness of AI models in SSL modeling.	> investigate cost-effective soft computing models for hydrology studies. > integrate nature-inspired optimization algorithms with AI for hybrid predictive models. > Use RFE to enhance AI models for sediment prediction. > Explore nature-inspired optimization algorithms for sediment modeling. > Address AI drawbacks in basin applicability and interpretability for SSL modeling. > investigate ensemble learning to improve AI models in sediment prediction.
[52]	NA	> Applications of DL in hydrology such as hydrologic modeling, flood forecasting, water quality indicators. > Use of CNN in time series modeling. > DL application in data-limited settings, especially in China.	2017 - 2020	> DL	> Limited training data for extreme conditions impacts reliability for rare events and puts the DL in challenges. > DL models may not capture fundamental processes like rainfall-runoff responses, hindering their transferability to other regions. > While applying DL challenges exist to flow and transport modeling in porous media due to instrumentation difficulties and heterogeneity. > Uncertainty quantification creates challenges across different DL model architectures. > Difficulty in reproducing 3D transient solutions in physics-guided ML.	> Necessitate deep integration between DL and physically based models in hydrology research. > Using process-based models to evaluate causal controls and distinguish factors in hydrological processes. > Require in-depth modification of DL algorithms fit to hydrology's specific needs. > Developing techniques for reproducing 3D transient solutions in physics-guided ML.

[53]	209	> AI models Assessing applied in river water quality simulation between 2000 and 2020, covering model structures, input variability, and regional investigations. > AI model effectiveness in addressing complex data characteristics for river water quality monitoring, management, and policymaking.	2000 - 2020	> ANN > Fuzzy Logic > Kernel-Based AI > Complementary Models > Hybrid AI Models	> ANN models need extensive data, prone to overfitting, and struggle with complex environmental data. > Fuzzy Logic model relies on rule-based functions, struggles to prioritize crucial input factors effectively. > SVM model success depends on careful kernel and parameter selection, inefficient with large datasets. > Complementary models depend on wavelet and level selection, leading to unpredictable accuracy. > NI algorithms encounter issues like premature convergence and limited improvement.	> Explore CE systems and DS systems to address complexity, uncertainty, and inconsistency in river water quality management. > Investigate various AI models such as ward NN, hopfield NN, kNN, DT, and DL to improve classification and prediction in river water quality studies. > Develop hybrid models combining AI strengths and optimization techniques for effective handling of complex, nonstationary data in water quality modeling. > Regression models such as GPR, DT, RT, MT, GLM, and ET necessitate further exploration to enhance their classification and prediction capabilities in river water quality research. > Focus on unexplored optimization techniques like bacterial foraging, amoeba-based algorithms to enhance model performance in water quality modeling. > Pay attention to the architecture of model, calibration methods, and data allocation to optimize model performance in river water quality. > Explore WT potential with AI models to extract features and denoise time series data in river water quality. > Incorporate additional variables such as population change, industrial influent to better identify pollutant sources and predict sudden changes in river water quality.

[54]	142	> Introduction to GP and its variants in the context of automatic program generation. > Applications of GP in WRE and its advantages in solving nonlinear WRE problems. > Exploration of advanced GP variants like multigene GP, linear GP, gene expression programming, and grammar-based GP.	1997 - 2018	> GP	> GP has the limitation in accurately modeling spatial velocity fields when facing small training datasets. > GP models often produce complex formulas, making interpretation difficult due to non-linear combinations of variables and constants, thus becoming sensitive to input choices. > Raw data in GP models can lead to dimensional inconsistency, necessitating techniques like DAGP to ensure consistency in models. > GP models lack physical interpretability due to their inherent non-linearity, facing challenges in interpretation using explicit functions. > Overfitting is a common issue in GP, resulting in the evolution of sub-programs that offer minimal or no performance improvement. > Standalone GP and its variants struggle to clearly identify complex relationships in complex systems.	> New GP programs are required to be developed to handle both binary and multi-class classification problems, expanding the software capabilities beyond symbolic regression. > Complexity in GP solutions is crucial, requiring investigation of methods like DAGP and multi-objective optimization to balance accuracy and interpretability while preventing overfitting. > To improve models for complex systems, researchers could focus on hybrid GP models integrating techniques like wavelet transform or moving average filters or coupling GP with other AI approaches or physically based models. > Exploration of potential benefits of other GP variants in WRE, beyond the commonly studied ones like monolithic GP, DAGP, GEP, LGP, MGGP, and GGP is highly encouraged. > Comparative analyses between GP and other AI techniques could offer valuable insights into their performance.
[55]	82	> DL and ML in Hydrological Processes, climate change, and earth	201 - 2018	> ML > DL	> The accuracy drawbacks linked to models. > The limitations of methods used for uncertainty analysis. > Significant computational costs related to employing ML models. > Needs for an extensive amount of data to meet requirements.	> investigate algorithms to enhance ML and DL for modeling hydrological processes. > Employing new fields for effective use of ML and DL methods in studying hydrological processes. > Using hybrid and ensemble techniques to improve ML and DL models in understanding hydrology.
[56]	140	> Flood management technologies. > Image processing in flood management. > ML applications in flood management.	2010 - 2020	> ANN > SVM > MLP > ANFIS > WNN	> Interpretative classification of research articles reveals bias due to subjectivity. > The selected study period is between 2010 to 2020 and may limit the recent developments, potentially impacting the comprehensiveness of the study's results. > Limited use of ML-based methods in post-disaster crisis management.	> Combining image processing and ML, for flood management requires investigation. > Need for AI integration to enhance post-disaster processes. > More research requires to use AI-enabled big data for flood management. > Investigation of models predicting flood recession duration for better recovery and reconstruction planning.

[57]	NA	> ML concepts and methods > Challenges in probabilistic hydrological post-processing. > Predicting in hydrology using ML.	NA	> Quantile > Expectile > Distributional > Regression algorithms.	> Quantile regression algorithms are not ideal for predicting extreme quantiles. > Quantile regression algorithms estimate predictive quantiles separately at different levels, need additional automation and potentially cause quantile crossing.	> Necessitate to explore the applicability of expectile regression algorithms in probabilistic hydrological forecasting. > Explore integrating complete predictive probability distributions in large-scale benchmark tests to investigate. > Meta-learning in hydrological post-processing and forecasting across different time scales and data availability conditions require to investigate.
[58]	NA	>Importance of streamflow gauge data for flood forecasting and risk assessment > Introduction of the Streamflow Hydrology Estimate using ML (SHEM) model > SHEM's reliance on ML and big data processing > Interpolation of estimated discharge and time data for inoperative stream gages.	NA	> ML (SHEM)	> Model faces challenges when historical remote telemetry data for ungauged water catchments are limited to shorter time periods than required for training the ML model. > Limitation to two key data parameters (stream stage and time) to reduce complexity, processing time, and computational requirements, severely impacting the model's analysis scope and accuracy. > Dependence on time and computing resources for building ML correlation indexes and the availability of sufficient and accurate historical streamflow datasets, affecting the effectiveness and utility of the model.	> Extending streamflow estimates research by leveraging SHEM's underlying ML and analytical processes to extrapolate estimated data from ungagged streams and interpolate data estimates from gauged streams with missing data. > Exploring the application of SHEM in remote ungauged catchment areas by incorporating RS technologies like synthetic aperture radar, digital aerial surveillance, and telemetry methodologies to generate stream gage data and historical index datasets worldwide. > Determining optimal locations for physical stream gages, measuring discharge, and conducting remote monitoring in inaccessible areas using SHEM. > Applying a three-phase approach to estimate streamflow for ungauged regions based on the long-term prediction analysis and duration curve prediction research. >Studying the addition of other correlated streamflow parameters, such as topographical attributes and precipitation parameters, to enhance the model's accuracy and efficiency when limited streamflow data histories are available.

[59]	37	> Focus on EC as an advanced ML approach for modeling ET and its progression. > The study aims to establish a new milestone by using the EC algorithm for ET modeling. > Conducting a review to assess the feasibility and potential of EC models in simulating ET across different environments. > Assessment and evaluation of EC models in modeling ET based on the review findings.	2007 - 2019	> ML (EC)	> Estimating ETo reliably with limited meteorological data poses a significant challenge, especially in developing countries. > The limitation of using a two-data-division procedure in modeling ETo is the assessment of methods without independent data sets. > Symbolic regression models using EC are highly complex with numerous mathematical operators. > Existing temperature-based empirical models for estimating ETo are not effective for projecting ETo under climate change. Even with rising temperatures, these models show a decrease in ETo due to the expected decline in diurnal temperature range's influence on ETo.	> Future studies required to develop ETo models that utilize easily available meteorological variables, like maximum and minimum temperature. > Investigate to develop simple temperature based ETo models suitable for reliably projecting ETo under climate change scenarios, considering the projected decline in diurnal temperature range's influence on ETo. > Explore to create a generalized EC-based ETo model to reliably estimate ETo for homogeneous climatic regions by calibrating and validating the model with all available station data in that area. > Utilizing high-resolution meteorological satellite data in ETo modeling, specifically in regions with scarce climatic stations or missing required data, by calibrating satellite data with information obtained from stations to improve spatial modeling.
[60]	NA	> The imperative need to adopt advanced ML methods for solving difficult problems in porous media and geoscience. > Provide a comprehensive review of recently developed methods in ML algorithms and their application in porous media and geoscience.	NA	> ML > AI > DL > ANN > Boosting algorithms > Principal component analysis > Multidimensional scaling > SVM > K-Means Clustering > K-Nearest Neighbor > CNN > RF > Autoencoder > GAN > LSTM	> Main drawback of LSTM is its inability to incorporate spatial information into data. > The drawback regarding input data is that certain areas in porous media and geosciences lack enough data, requiring new solutions to provide the required input for ML.	> Explore the application of recent DL models for understanding complex data and leveraging relevant data. > Development of physics-informed AI models to incorporate complex physics into ML algorithms for porous media and geoscience. > Investigate solutions for input data challenges in areas with insufficient data for porous media and geosciences. > Studying of scientific interpretation of ML results based on experimental and computational methods. > To investigate the customization of current ML models for specific problems in porous media. > Integration of knowledge from big data and ML for better synergy between the fields. > Explore the benchmarking for ML algorithms in Geosciences to enable systematic progress and confident method selection.

[61]	550	> Introduction to AI, including its history, major categories, and current development. > Applications of AI in hydropower and dam engineering, emphasizing on predictive modeling, real-time monitoring, optimization, and case studies. > Exploration of current and emerging technologies in dam engineering, covering automated decision-making systems and AI-powered drones for inspection, along with addressing integration challenges.	2002 - 2023	> AI > ML > DL > NN > Ensemble > Regularization > Bayesian > DT > Dimensionality Reduction > Rule System > Instance Based > Regression > Clustering	Digital twins limitations: > The necessity for a thorough and regularly updated digital model poses challenges in its creation and maintenance. > Digital twins heavily depend on accurate and dependable data inputs. > Implementation faces drawbacks due to the scarcity of standardized protocols and frameworks. Internet of things limitations: > Security and privacy issues related to the transmission and storage of data. > Dependence on a powerful and reliable network infrastructure for uninterrupted connectivity. > Issues related to hybridization and compatibility with older, established systems. Drone Technology limitations: >Constraints in flight duration and range, potentially limiting coverage for expansive structures such as large dams. > Impact of regulations and airspace drawbacks on drone operations. > Needs skilled operators and specialized training to ensure safe and efficient drone utilization. Remote Sensing and Satellite Monitoring Technology Limitations: > depends on satellite availability and favorable imaging conditions, which might be restricted in specific geographical areas. > Data processing and analysis demands expertise in remote sensing techniques. > Variations in accuracy and resolution of satellite imagery impacted by sensor types and orbital configurations.	> Understanding and tailored algorithms for AI in dam engineering. > Enhancing data quality and availability in dam engineering, encompassing dam behavior, hydrological conditions, and environmental factors. > Prioritizing algorithmic transparency and ethical considerations in AI applications for dams, emphasizing fairness, lack of bias, and transparency. > Encouraging interdisciplinary collaboration between civil engineers, data scientists, and stakeholders in AI and dam engineering research. > Identifying training programs to equip engineers and stakeholders with AI-related skills for dam engineering. > Exploring methods to incorporate uncertainty quantification into machine learning models used in dam engineering. > Developing explainable AI models for dam engineering to ensure transparency in decision-making processes. > Enhancing data quality and accessibility through standardized collection methods and easier data availability for researchers. > Investigating edge computing for faster processing and improved security in real-time data analysis for dam engineering. > Improving cooperation between AI experts and dam engineers for effective AI model development and application. > Addressing issues of limited data growth in dam engineering through strategies like data enhancement, transfer learning, and collaborative data sharing. > Focus on the role of ML/AI in optimizing existing dam structures, innovative maintenance, and monitoring techniques for enhanced safety, efficiency, and sustainability in dam engineering.

					Cloud virtual and augmented reality technology limitations: > Depends on a stable internet connection for uninterrupted access to cloud resources. > Difficulties in hybridizing virtual and enhanced reality technologies into current workflows. > Needs for proper training and user acceptance to ensure effective utilization. Smart Robotics Technology limitations: > Complicated integration and programming are required for tailored robotic systems. > Challenges in managing intricate or unconventional situations that might require human intervention. > Important initial investment costs associated with the implementation of robotic systems. 3D printing technology limitations: > Limitations in scalability for extensive projects such as large-scale dam construction. > Challenges in guaranteeing material quality and structural integrity. > Considerations about compliance with regulations and codes for extra manufacturing within construction practices. Building Information Modeling Limitations: > Early setup and implementation costs may be high for organizations transitioning to BIM. > Necessitates training and adoption by all project participants to maximize its benefits. > Hybridizing with legacy systems and software interoperability can be problematic. > Dependence on the availability of standardized BIM protocols and workflows.	

[62]	103	> Improvement in hydrologic modeling, especially hybrid wavelet and AI-based models. > Applications and robustness of Wavelet–AI models in predicting hydrologic processes.	2003 - 2014	> Wavelet > AI	> AI-based models face highly non-stationary responses across various frequencies, impacting their effectiveness without proper data pre/post-processing. > Fourier analysis lacks the ability to retain time information during a signal's transformation into the frequency range, making it impossible to identify specific event timing.	> Expand WT applications in hydrology beyond DWT to include CWT for enhanced analysis within all-time scales. > Investigate WT use for spatial data preprocessing in hydrological models, beyond its current focus on temporal preprocessing. > Explore alternative criteria, like energy similarities, to select appropriate mother wavelets and decomposition levels in wavelet-based models for hydrological processes. > More research is required on wavelet–AI models in groundwater and water quality modeling due to restricted existing studies. > Explore benchmark datasets and an archive of wavelet–AI models for certain hydro-climatologic processes to improve transparency and efficiency. > Hybridize wavelet–AI models with physically-based models for spatiotemporal parameter estimation using geomorphologic characteristics. > Develop more review articles to explore applications of hybrid models combining AI with different data preprocessing techniques in hydro-climatologic studies.

[63]	67	> Assessment of AI techniques for modeling groundwater levels. > Review of popular AI techniques for groundwater level modeling and forecasting. > Determination of weaknesses in AI modeling and the significance of reviewing these procedures, especially in groundwater level modeling applications.	2001 - 2018	> ANN > ANFIS > GP > SVM > Hybrid AI techniques	> Traditional conceptual or physical-based models for GWL modeling have drawbacks due to their need for vast data and input parameters, which might hinder accurate predictions when data is limited and prediction accuracy is prioritized over understanding underlying mechanisms. > AI models shown limitations in handling non-linear and non-stationary processes.	> Combination of AI methods with conceptual-numerical models like MODFLOW to address weaknesses of each model, reducing computation time and enhance data availability between models. > Focus on input consideration, emphasizing the significance of GWL time series as a primary input and exploring non-causal wavelets for comprehending hydrological variable interactions. > Exploration of hybrid AI models for GWL simulation, checking different AI techniques at different modeling stages to optimize performance. > Exploration into border effects and causality in wavelet decomposition for GWL modeling, assessing methods to handle decomposed sub-time series effectively. > Investigation of non-causal wavelet algorithms such as trous and maximal overlap DWT for GWL forecasting, addressing potential inaccuracies in wavelet-based hydrological models.

[64]	NA	> Assessment of ML techniques in modeling groundwater quality. > Evaluation of various ML models for predicting groundwater quality parameters. > Assessment of improvements, dominance, and shifts in ML model usage over time in GWQ modeling.	1994 - 2022	> ML > Ensemble > Semi-Supervised > Supervised > Unsupervised > Fuzzy > DT > RF > LR > SVM > ANN > Comparative > DL > SOM > Clustering > Multiple Frameworks > Wavelet > Genetic Algorithms > Bagging Boosting	> Current ML studies' weakness lies in findings limited to certain study areas, causing discrepancies in model performance, and hindering wider applicability. > ANN models face a critical weakness that depends heavily on network structure selection.	> Focusing on best practices in ML application for hydrology, particularly in data handling and model assessment. > Investigating originality in studies, avoiding repetitive descriptions of ANN application without significant contributions. > Studying the potential of DL in GWQ modeling and the integration of lifelong learning and meta-learning for improved performance. > Employing explainable AI (XAI) to interpret and visualize processes in ML models for broader applicability. > Exploring underrepresented models and enhancement techniques for GWQ modeling, including the investigation of genetic algorithms. > Advocating for open-source materials, like code and data, to establish standards and facilitate validation across multiple study areas. > Encouraging the application of ML to new geographical areas for GWQ assessment to test model robustness. > investigating infrequently used parameters, such as anthropological effects or long-term climate conditions, to improve model validity in the face of future changes. > Exploring for practical applications of ML in GWQ management, urging cooperation between data scientists and authorities for effective implementation and assessment.

[65]	39	> DL methods in hydrology and water resources applications. > Integration of physics-based hydrological models with DL (PGDL) > Advancements in DL models for sequential time series data.	2017 - 2023	> DL	> Low availability of labeled datasets for hydrological modeling using DL or ML. > Shortage of interpretability in understanding processes and correlation within DL models. > Challenges in achieving firm predictions by combining domain knowledge with DL. > DL models deal with capturing non-stationary phenomena and uncertainties in hydrological processes. > Struggles in handling high volume, velocity, and accuracy of real-time data for hydrological applications. > Significant computational resources required for DL models with large and complex hydrological datasets. > DL models face challenges in generalizing across diverse regions and conditions. > Difficulties in implementing effective techniques for augmenting hydrological data using DL models.	> Developing models to improve and augment labeled datasets in hydrology. > Generating inherently interpretable DL models for better understanding in hydrological applications. > Enhancing integration approaches for domain knowledge and DL models for accurate predictions. > Investigating novel DL architectures capable of handling non-stationary phenomena. > Exploring scalable frameworks for efficient real-time data processing in hydrology. > Expanding more efficient DL architectures and optimization algorithms for hydrological datasets. > Studying transfer learning methods robust across diverse areas and conditions. > Developing techniques for effective enhancement of hydrological data using synthetic data generation.

[66]	138	> Significance of accurate soft computing methods for groundwater level predicting. > Enhancement in groundwater level prediction using ML models over the past two decades.	2008 - 2020	> ANN > Fuzzy logic > Neuro fuzzy > Kernel > DL > Hybrid ML > DT > Data mining > AI models > Statistical models > NARX	> NARX networks have drawbacks including over-fitting issues and susceptibility to local minima. > The Bayesian algorithm is suggested as a reliable training model to dwindle over-fitting, but its performance might decrease when used in conjunction with early stopping. > Increasing the number of input variables could hinder the reliability and precision of the models. > Exogenous parameters such as sea level rise and GW abstraction significantly affect GWLs.	> Exploration of exogenous parameters like sea level rise and GW abstraction's impact on GWL, particularly in coastal areas. > Improved selection of GWL lags using ACF and PACF approaches for AI modeling approaches. > Employing feature selection techniques to enhance model accuracy by selecting the most important input parameters and eliminating redundant information. > Emphasize predicting yearly GWLs for effective long-term water resource management, especially in dry regions. > Assessment of additional genetic programming variants, like Linear GP or multi-stage GP, for GWL prediction. > Development of DL techniques to impute missing GW values, reducing uncertainty and enhancing data quality for forecasting. > Investigation of hybrid ML models integrating nature-inspired algorithms with standalone ML models to optimize hyperparameters for improved prediction capabilities.

[67]	43	> Reviewing hybrid models integrating AI and optimization techniques specifically for streamflow prediction in hydrology. > Combination of AI and optimization techniques	2009 - 2020	> ANN > SVM > K-NN > ANFIS > GA > PSO > ABC > GWO	ANN limitations: > BPNN model's accuracy reduces during weak inflow reservoir stages. > Shortage of established norms or fixed guidelines for suitable model design. > Over-parameterization and overfitting problems in ANNs without maximal input selection and early interruption strategies. > Limited explanatory capability of ANNs for their test answers. > Uncertainty regarding the number of iterations required for an effective outcome. ANFIS Limitations: > Sensitivity to alterations in cluster counts. > Complexity of coding increases with the inclusion of additional rules. SVM Limitations: > Insufficient efficiency when the number of applications exceeds the number of samples. > Use of K-fold cross-validation for likelihood estimation without a standardized benchmark for the K value illustration. RF Limitations: > Prediction can be slow when tackling a high number of trees. > Even though it performs parametric regression, the model remains a "black box" without uncovering its internal workings. > The model's ability to make inferences beyond the training data is restricted, necessitating the training data to comprehensively represent the forest's variability within the study area. ARMA Limitations: > Long-term forecasting may have inadequate accuracy. > Nonlinear predictions are generally with low accuracy and often not applicable.	> Adoption of more than three different AI models to improve accuracy in study results. > Focus on selecting the best integration of input variables to enhance model performance and result accuracy, with efficient data pre-processing. > Significance of tuning model parameters, like hidden layers, epsilon, and fuzzy rules, utilizing various algorithms rather than relying on trial and error. > Exploring global optimization algorithms, specifically PSO and GA, in conjunction with local algorithms for improved convergence on global or near-global optimum in future studies.

[68]	72	> Investigating a more advanced version of the extreme ML for predicting river flow. > Comparing the EELM model's performance with classical ELM and SVR models using various indicators.	2013 - 2019	> SVR > ELM > ANN	> Random weight initialization in the Single Hidden Layer ELM model affects learning, causing ineffective predictions. > ELM's application lacks memory for crucial hydrological features, delaying advanced DL processes like LSTM-based models. > Limited resources delay experimental research, advocating for CESC models in hydrological science. > ELM's efficiency requires validation beyond numerical accuracy, emphasizing practical implementation for water resource engineering. > Human activities' impacting on catchment behavior is absent in modeling, potentially influencing modeling results. > Additional correlated weather and hydrological variables are necessary for enhanced river flow modeling accuracy.	> Explore DL ELM models with recurrent layers for enhanced weight determination and faster learning. > Investigate LSTM-based ELM models for hydrological forecasting, capturing time-series behavior and spatial features with low complexity. > Focus cost-effective soft computing models for solving real societal obstacles in hydrology. > Study of ELM model justification its efficiency practically for water resource engineering and as an expert system. > Investigate human activities impacting catchment behavior in hydrological modeling. > Investigate additional correlated variables to enhance river flow prediction accuracy. > Develop reliable models to handle missing hydrological data, particularly in developing regions, and reassess model validation techniques for higher performance assessment.

[69]	NA	> Significance of accurate hydrological understanding in managing water resources amidst anthropogenic climate change. > Introduction of PaML as a merging of hydrology and ML paradigms.	NA	> hysical Data-guided ML > Physics-informed ML > Physics-embedded ML > Physics-aware Hybrid learning > HydroPML	Classical Physical Data-guided Neural Networks Limitations: > Limitations in generalizing beyond specific contexts. > Necessitating computational requirements. > Challenges in interpreting physical aspects in the model. > lack of ability to manage sparsity or non-uniformity in temporal or spatial observations. > High computational requirements. > Relies on mesh-based or particle-based models, involving message-passing among small-scale moving and interacting objects. > Limited capabilities and robustness, especially in high-dimensional action spaces, leading to unknown behavior and convergence speeds. Deep Operator Networks Limitations: > Inaccuracy in approximating complex physical dynamics. > Instability and limited generalization. > FFT-based basis functions often lack spatial resolution, specially localized in frequency. > Complexity, delaying interpretability and making training challenging due to computational costs and restricted generalization. Physics-discovery Neural Networks Limitations: > Difficulty arises from complex, non-linear systems and imperfect data that are noisy and incomplete. > Limitations exist in acquiring high-fidelity, noise-free measurements.	> Investigate PeML or hybrid models for better short and long-term rainfall-runoff predictions. > Develop robust spatio-temporal representations for different regions and improve forecast reliability using advanced PaML models. > Address missing data, ungauged basins, and uncertainty in rainfall-runoff predicts by hybridizing ML's transferability with physical models such as PiML, PeML, and PaHL. > Generate a thorough PaML-based hydrodynamic solver spanning domains and incorporating RS and in-situ data. > Employ physics-discovery NNs or data-physics-driven parameter discovery for better calibration of hydrodynamic processes. > Use generative models in PaML for simulating hydrodynamic processes under certain conditions. > Combine physics knowledge into ML to improve uncertainty characterization in hydrodynamic modeling.

[70]	21	> Emphasizing the application of DL in hydrological forecasting parameters. > Investigation of conventional ML models for reservoir inflow and rainfall prediction. > Comparison of AI models used in different hydrology sectors, especially DL and ML techniques.	2018 - 2021	> AI > ML > Supervised > Unsupervised > SVM > RF > BRT > DT > Boosting > DL > LSTM	> ML algorithms face drawbacks in selecting hydrological parameters that strongly correlate with the output for accurate predictions. > DL algorithms need large amounts of data to function effectively.	> Investigate comparative studies employing a wide range of models to identify the most accurate algorithm for predicting reservoir inflow. > Employ diverse hydrological parameters to enhance the precision of forecasting models.
[71]	39	> Comprehending the Challenges of Lake Water-Level Fluctuation Prediction > Progress and Impact of ML in Forecasting Fluctuations > Assessment of different ML Models for Lake Dynamics	2006 - 2020	> ANN > SVM > ANFIS > WA-ANN > WA-ANFIS > WA-SVM > GEP > ELM > DL	> Numerous research mainly used past water-level data without including other significant factors like inflow/outflow, rainfall, and evaporation. > Overlooking influential elements might restrict the accuracy of the models in forecasting lake water levels. > Model Inputs: Emphasize only on past water-level data without considering other key factors like inflow/outflow or rainfall, potentially limiting model precision. > Relying on common indicators such as RMSE, MAE lacks more comprehensive criteria like KGE, limiting a complete model performance assessment.	> Assess how divers ways of splitting data influence model performance. > Utilize evaluation criteria such as KGE, AIC, and SBC for evaluating models. > Employ WDDFF as a benchmark when using hybrid WA and ML models. > Integrate WA, FA, and ML models to gauge their effectiveness together. > Estimate the performance of enhanced ELM models. > Identify the threshold when LSTM-based DL models exceeds other models based on training data length.
[72]	68	> Historical progress of groundwater modeling techniques. > Current application of ML in forecasting groundwater levels. > Prospects and directions for enhancing groundwater modeling with a focus on ML models.	2010 - 2020	> ANN > SVM > DT > Fuzzy > GA > Ensemble > Hybrid	> ML models struggle when handling non-stationary data.	> Expand research on GWL prediction using ML in regions with limited surface water, like Africa, parts of Europe, and South America. > Investigate in more studies the utilization of ensemble learning techniques for enhanced accuracy in GWL prediction, despite their limited usage in the reviewed works.

[73]	18	> Review of rainfall-runoff modeling using ML models. > Assessment of combined and ordinary ML models > Critically reviewing characteristics, advantages, and disadvantages of three commonly used ML approaches for runoff simulation	2010 - 2021	> ANFIS > ANN > SVM	> on-linear intricacy of streamflow needs a model that comprehends these complexities well. > Caution is required with the non-linear nature of ML to avoid overfitting in run-off modeling. > ML approaches can offer higher precision in run-off simulation but require effective input parameter determination for optimal performance.	> Employ newly developed optimization algorithms to enhance ML model abilities. > Explore the non-linear nature of ML to avoid overfitting in run-off modeling problems. > Investigate the potential of ML in computational hydrology, particularly in run-off predicting. > Study hybrid mechanisms combining hydrological knowledge with ML to enhance complex hydrological predictions. > Employ hybrid-based models integrating advantages of both physical-based and ML-based models for run-off simulations in water resource management.
[74]	1451	> Analyzing data's impact on weather, disasters, and smart water systems in environmental management. > Using ML, particularly DL, for autonomous decisions and system improvements in environmental contexts.	2004 - 2018	> DL	> Data cleansing issues due to different data sources and IoT technologies. > Lack of labeled datasets for effective ML. > Discrepancy between data ingestion and generation speeds influencing real-time analysis. > Lack of understanding of DL architectures and best practices. > Necessitate for hybrid ML and physics-based approaches for interpretable solutions. > Hefty cost associated with Big Data platforms. > Absence of data governance and sociotechnical structures for enhanced data quality and accessibility.	> Scaling operations with AI and automated implementation for Smart Data. > Humanitarian benefits through enhanced disaster relief with Big Data and DL. > Different applications showcasing Big Data and AI's potential. > Improving causal inference and reasoning using Big Data and DL. > Enhancing situation awareness for forecasting short-term and gradual Earth system changes.
[75]	NA	> Investigation of TGCO, TGRO, and TGA to address ML drawbacks: mitigating model opacity, improving convergence and generalizability, but lacking evidence on improving transferability.	NA	> ML	> Most ML models employed in hydrogeology are often black boxes, lacking transparency. > The effectiveness of theory-guided models in enhancing the transferability of ML models remains uncertain due to lack of studies. > Limitations of ML models, includes the nature of black-box, limited generalization, hypothetical convergence, and uncertain transferability.	> Additional studies required to fill knowledge gaps on the effectiveness of Theory-guided ML models in prevailing ML drawbacks. > Explore more studies on real configurations for generalization of theory-guided ML models. > Expand transformation functions aligned with theory for full use of TGRO in hydrogeology. > More investigations are required to clarify metrics to quantify black-box nature and transferability prevailing by theory-guided models.

3. Hydrological process-based machine learning insights

Based on the exhibited literature review in this survey, the understanding and prediction of hydrological processes have advanced significantly thanks to computer aid models. Numerous hydrological phenomena, including precipitation, evaporation, snowmelt, soil moisture, streamflow, and groundwater recharge, have been simulated and predicted using these models. The creation of mathematical models known as physically-based models—which replicate the basic physical principles governing hydrological systems—has made some headway. The relationships between precipitation, evaporation, soil moisture, vegetation, and surface runoff for example are simulated by these models, which include the Distributed Hydrological Soil Vegetation Model (DHSVM) and the Soil and Water Assessment Tool (SWAT). Predictions of streamflow, groundwater recharge, and other hydrological variables have been improved by using these models.

The development of ML models (e.g., DL, GP, ANFIS, DT, RF, SVM and ANN), has advanced remarkably for different hydrological problems. These models can be employed to predict future hydrological conditions by learning the correlations between input variables (such as precipitation, temperature, and land use) and output variables (like streamflow and soil moisture) using historical data. Hybrid models have also demonstrated the potential to enhance our knowledge and prediction of hydrological processes because they exploit the best features of physically-based and data-driven models. The use of satellite data for precipitation, evapotranspiration, and snow cover estimation is one example of how remote sensing data has been utilised more frequently in recent years to enhance model performance. In conclusion, there has been a noticeable advancement in understanding and prediction of hydrological processes due to the emergence of mathematical models. Several hydrological processes have been simulated and predicted using physically-based models, data-driven models, and hybrid models and the performance of the models has been enhanced by the use of remote sensing data.

The development and performance of ML models are hampered by several obstacles and limitations, such as bias in variable selection, re-substitution validation, inconsistent validation procedures, resamples for various algorithms, and model selection by the test set [76]. The main issue with ML models is the selection bias of linked parameters, which results in selecting irrelevant predictions. Selection bias is encountered when the same data are used for the selection of the related input predictors and building a model from a training set. This effectively indicates that the process of choosing the variables was not regarded as a component of the model-building procedure [23]. On the other hand, model building without selection bias produces better results for the same data set but the results are sensitive, and the same process of variable selection can produce different results if the training set is slightly different [77]. This is due to the vagueness resulting from the variable selection process can be significantly higher than the uncertainty pertaining to the model. The performance estimates derived from this method, even if the best predictors are chosen, are biased because they fail to account for the uncertainty surrounding the variable selection process and how it affects the outcome [78]. Selection bias is more likely when using ML algorithms for hydrologic modelling and forecasting because of short historical records, a high number of predictors, and sophisticated and potent ML models.

Re-substitution validation, or validating a model using the same data it was trained on, is another frequent problem. This could result in an optimistic performance estimation due to the issue of overfitting that occurs when a model fits the training data too well and applies less well to "unseen data," or data that was not included in the training set [79]. The validation dataset could be used to ensure proper modelling training and prevent overfitting. K-fold cross-validation is one of many approaches that is frequently used to validate models and guarantee generalisation for data that has not yet been observed. Cross-validation, however, has the potential to produce estimates of model performance that are biased [80]. The exposure of the model to the validation dataset (which influences the choice of parameters, model, and hyper-parameter values during the subsequent training and validation phases) provides that the cross-validation estimate is often biased [81]. If a test set proves unusable due to insufficient historical records, the performance estimates derived only from cross-validation will probably be excessively optimistic regarding prediction accuracy. However, as the performance estimates of each model are affected equally, the biased, optimistic estimates from cross-validation can be utilised for comparing among a collection of models and choosing the best model.

The use of distinct validation procedures and resamples for various algorithms presents another challenge in the creation of machine learning models. There must be uniformity in the cross-validation and resampling procedures for the various algorithms under consideration to be equally affected by the performance prediction bias and to have equivalent model behaviours and error ranges. Any data transformation or pre-processing undertaken during the data partitioning between training and validation datasets must be done as part of the cross-validation procedure and not separately beforehand. If the complete training/validation dataset was pre-processed independently beforehand, this might be the result of (i) Knowledge of the mean and variance of the validation set during the training phase, or (ii) Compromising the test data if the entire dataset is used. Leaks of this kind affect not just the choice of the model but also the evaluation of the model's prediction accuracy and generalisation performance on unknown data.

Another concern is the application of the test dataset to algorithm selection or model hyper-parameter optimisation [82]. Model hyper-parameters are ML algorithm settings or tuning parameters that determine a model's flexibility or complexity and help govern algorithm behaviour. Selecting a model based on the test dataset results in an overly optimistic or biased estimate of the model's performance, as well as information leakage from the test set to the model. Test sets should only be utilised to predict the performance of models. While choosing a model and selecting the final model, cross-validation results should be employed since they provide a more reliable viewpoint on how well the model will generalise to new datasets.

Choosing the right input variable is essential to creating reliable ML models. According to recent research, the model's performance can be enhanced by carefully choosing the variables, and this impact varies depending on the problem. The encountered problems in input variable selection are mostly because the variable selection algorithm relies on both the model structure and input data used to select the variables. For example, all potential predictors of streamflow generation processes for a given research region must be considered to create a streamflow forecasting model, especially in mountainous locations with significant variability. When combined with local hydro-meteorological measurements, the use of large-scale climatic indices as predictors enhances the accuracy of streamflow forecasting, according to several studies. Climate indices are usually employed in groups since climate patterns are interrelated and no single index can fully explain all the climatic variability within a river basin. Less interpretable machine learning models can result from constructing data-driven models and evaluating their performance using the above-mentioned traps, since the models may then contain redundant or unnecessary predictors. If the traps are not avoided, the produced model's future prediction accuracy may be overestimated, which could result in unpredictable, uncertain model outputs. Additionally, the model may perform worse when generalising on data that has not yet been observed.

Abbreviations

ACO: Ant Colony Optimization

ABC: Artificial Bee Colony

AIC: Akaike Information Criterion

ANN: Artificial Neural Networks

ANFIS: Adaptive Neuro-Fuzzy Inference System

ARMA: Auto-Regressive Moving Average

BN: Bayesian network
BRT: Boosted Regression Tree
BMA: Bayesian Model Averaging
CART: Classification and regression tree
CE: Coupling Expert
CESC: Cost-Effective Soft Computing
CWT: Continues Wavelet Transform
CNN: Convolutional Neural Networks
DWT: Discrete Wavelet Transform
DS: Decision Support
DTL: Decision Tree Learning
DT: Decision Trees
DDM: Data Driven Model
DAGP: Direct Acyclic Graph Programming
EC: Evolutionary Computing
EES: Earth and Environmental Sciences
EPS: Ensemble Prediction Systems
EC: Evolutionary Computation
EANN: Evolutionary Artificial Neural Network
ELM: Extreme Learning Machine
FRBS: Fuzzy Rule-based Systems
FOS: Fast Orthogonal Search
FA: Firefly Algorithm
GA: Genetic Algorithms
GWQ: Groundwater Quality
GWO: Grey Wolf Optimization
GP: Genetic Programming
GA: Genetic Algorithms
GPR: Gaussian process regression
GBM: Gradient Boosting Machine
GAN: Generative adversarial network
GRNN: Generalized Regression Neural Network
GRU: Gated Recurrent Unit
HydroPML: Hydrology in Physics-aware Machine Learning
Hybrid RBNN: Hybrid Radial Basis Neural Network
KGML: Knowledge-Guided Machine Learning
KNN: k-Nearest Neighbours
KGE: Kling–Gupta efficiency
LM: Multivariate linear
LSTM: Long Short-Term Memory
MLP: Multilayer perceptron
M5P: Modified Decision
MLP: Multilayer Perceptron
M5 MT: M5 Model Trees
MLR: Multivariate Linear Regression
MARS: Multivariate adaptive regression splines
ML: Machine Learning
NCMs: Neurocomputing methods/models
NC: Neurocomputing

NARX: Non-linear auto-regressive network with exogenous input
OPLS: Orthogonal partial least squares
PSO-RBNN: Integrative Particle Swarm Optimization RBNN
PBM: Process-Based Modelling
PSO: Particle Swarm Optimization
PB: Process-Based
PSO: Particle Swarm Optimization
PGDL: Physics-Guided Deep Learning
PAML: Physics-aware Machine Learning
RBNN: Radial Basis Neural Network
RBF: Radial Basis Function
RF: Random Forest
RFE: Recursive Feature Elimination
SDFs: Socio-demographic factors
SWIM: Soil and Water Integrated Model
SG: Stack Generalization
SI: Swarm Intelligence
SVM: Support Vector Machine
SWAT: Soil and Water Assessment Tool
SHEM: Streamflow Hydrology Estimate using Machine Learning
SBC: Schwarz Bayesian Criterion
SGB: Stochastic Gradient Boosting
WT: Wavelet Transformation
WRE: Water Resources Engineering
WA: Wavelet Analysis
WDDFF: Wavelet Data-Driven Forecasting Framework
WT: Wavelet Transform
W-AI: Wavelet-Artificial Intelligence
W-ANN: Wavelet Artificial Neural Networks
WNN: Wavelet Neural Network
XGB: Extreme Gradient Boosting

Conflict of interest: **The author declares no conflict of interest to any party.**

References

[1] P. Gentine, T. J. Troy, B. R. Lintner, and K. L. Findell, "Scaling in Surface Hydrology: Progress and Challenges," J. Contemp. Water Res. Educ., 2012.
[2] L. A. Cuartas et al., "Distributed hydrological modeling of a micro-scale rainforest watershed in Amazonia: Model evaluation and advances in calibration using the new HAND terrain model," J. Hydrol., vol. 462, pp. 15–27, 2012.
[3] G. Wang et al., "Integrated watershed management: evolution, development and emerging trends," J. For. Res., vol. 27, pp. 967–994, 2016.
[4] G. Sun et al., "Forest hydrology modeling tools for watershed management: A review," For. Ecol. Manage., vol. 530, p. 120755, 2023.
[5] A. Beniaich et al., "Spontaneous vegetation as an alternative to cover crops in olive orchards reduces water erosion and improves soil physical properties under tropical conditions," Agric. Water Manag., vol. 279, p. 108186, 2023.
[6] E. P. Anderson et al., "Understanding rivers and their social relations: A critical step to advance environmental water management," Wiley Interdiscip. Rev. Water, vol. 6, no. 6, p. e1381, 2019.

[7] J. R. Karr and I. J. Schlosser, "Water resources and the land-water interface. Water resources in agricultural watersheds can be improved by effective multidisciplinary planning," Science (80-.)., 1978.
[8] H. Xu, A. Berres, Y. Liu, M. R. Allen-Dumas, and J. Sanyal, "An overview of visualization and visual analytics applications in water resources management," Environmental Modelling and Software. 2022.
[9] M. Drogkoula, K. Kokkinos, and N. Samaras, "A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management," Appl. Sci., vol. 13, no. 22, p. 12147, 2023.
[10] J. Sheffield et al., "Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions," Water Resour. Res., vol. 54, no. 12, pp. 9724–9758, 2018.
[11] H. C. Gonçalves, M. A. Mercante, and E. T. Santos, "Hydrological cycle," Brazilian J. Biol., vol. 71, pp. 241–253, 2011.
[12] T. Oki and S. Kanae, "Global hydrological cycles and world water resources," Science (80-.)., vol. 313, no. 5790, pp. 1068–1072, 2006.
[13] Z. M. Yaseen et al., "Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models," Eng. Appl. Comput. Fluid Mech., vol. 14, no. 1, pp. 70–89, 2019.
[14] D. Rolnick et al., "Tackling climate change with machine learning," ACM Comput. Surv., vol. 55, no. 2, pp. 1–96, 2022.
[15] K. J. Beven and M. J. Kirkby, "A physically based, variable contributing area model of basin hydrology," Hydrol. Sci. Bull., 1979.
[16] K. Beven, "Prophecy, reality and uncertainty in distributed hydrological modelling," Adv. Water Resour., vol. 16, no. 1, pp. 41–51, 1993.
[17] H. Tao et al., "Hybridized artificial intelligence models with nature-inspired algorithms for river flow modeling: A comprehensive review, assessment, and possible future research directions," Eng. Appl. Artif. Intell., vol. 129, no. November 2023, p. 107559, Mar. 2024.
[18] Z. Iqbal, S. Shahid, T. Ismail, Z. Sa'adi, A. Farooque, and Z. M. Yaseen, "Distributed Hydrological Model Based on Machine Learning Algorithm: Assessment of Climate Change Impact on Floods," Sustain., 2022.
[19] Z. M. Yaseen and S. Shahid, "Drought Index Prediction Using Data Intelligent Analytic Models: A Review," in Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation, Springer, 2020, pp. 1–27.
[20] G. P. Tsakiris and D. P. Loucks, "Adaptive Water Resources Management Under Climate Change: An Introduction," Water Resources Management. 2023.
[21] R. Chalh, Z. Bakkoury, D. Ouazar, and M. D. Hasnaoui, "Big data open platform for water resources management," in 2015 International Conference on Cloud Technologies and Applications (CloudTech), 2015, pp. 1–8.
[22] Y. Kim, N. Kang, J. Jung, and H. S. Kim, "A Review on the Management of Water Resources Information based on Big Data and Cloud Computing," J. Wetl. Res., 2016.
[23] L. Schmidt, F. Heße, S. Attinger, and R. Kumar, "Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany," Water Resour. Res., 2020.
[24] Ö. Çelik, "A research on machine learning methods and its applications," J. Educ. Technol. Online Learn., vol. 1, no. 3, pp. 25–40, 2018.
[25] C. Castillo-Botón et al., "Machine learning regression and classification methods for fog events prediction," Atmos. Res., vol. 272, p. 106157, 2022.
[26] V. KlemeŠ, "Operational testing of hydrological simulation models," Hydrol. Sci. J., 1986.

[27] D. Yang, Y. Yang, and J. Xia, "Hydrological cycle and water resources in a changing world: A review," Geography and Sustainability. 2021.
[28] S. Fatichi et al., "An overview of current applications, challenges, and future trends in distributed process-based models in hydrology," Journal of Hydrology. 2016.
[29] D. Pandi, S. Kothandaraman, and M. Kuppusamy, "Hydrological models: a review," Int. J. Hydrol. Sci. Technol., vol. 12, no. 3, pp. 223–242, 2021.
[30] H. Lange and S. Sippel, "Machine Learning Applications in Hydrology," in Forest-Water Interactions, Springer, 2020, pp. 233–257.
[31] D. Watson-Parris, "Machine learning for weather and climate are worlds apart," Philos. Trans. R. Soc. A, vol. 379, no. 2194, p. 20200098, 2021.
[32] A. E. Thessen, "Adoption of machine learning techniques in ecology and earth science," PeerJ PrePrints, 2016.
[33] M. Fu, T. Fan, Z. Ding, S. Q. Salih, N. Al-Ansari, and Z. M. Yaseen, "Deep Learning Data-Intelligence Model Based on Adjusted Forecasting Window Scale: Application in Daily Streamflow Simulation," IEEE Access, vol. 8, pp. 32632–32651, 2020.
[34] M. Jamei et al., "Multi-step ahead hourly forecasting of air quality indices in Australia: Application of an optimal time varying decomposition-based ensemble deep learning algorithm," Atmos. Pollut. Res., p. 101752, 2023.
[35] T. Tiyasha, T. M. Tung, and Z. M. Yaseen, "Deep Learning for Prediction of Water Quality Index Classification: Tropical Catchment Environmental Assessment," Nat. Resour. Res., pp. 1–20, 2021.
[36] M. Zounemat-Kermani et al., "Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects," Journal of Hydrology. 2020.
[37] S. Razavi et al., "Coevolution of machine learning and process-based modelling to revolutionize Earth and environmental sciences: A perspective," Hydrological Processes. 2022.
[38] F. Fahimi, Z. M. Yaseen, and A. El-shafie, "Application of soft computing based hybrid models in hydrological variables modeling: a comprehensive review," Theor. Appl. Climatol., pp. 1–29, 2016.
[39] A. Gharib and E. G. R. Davies, "A workflow to address pitfalls and challenges in applying machine learning models to hydrology," Adv. Water Resour., vol. 152, p. 103920, 2021.
[40] Z. M. Yaseen, A. El-shafie, O. Jaafar, H. A. Afan, and K. N. Sayl, "Artificial intelligence based models for stream-flow forecasting: 2000–2015," J. Hydrol., vol. 530, pp. 829–844, 2015.
[41] H. Shirmard, E. Farahbakhsh, R. D. Müller, and R. Chandra, "A review of machine learning in processing remote sensing data for mineral exploration," Remote Sens. Environ., vol. 268, p. 112750, 2022.
[42] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," Water (Switzerland). 2018.
[43] D. P. Solomatine, "Data-driven modeling and computational intelligence methods in hydrology," Encycl. Hydrol. Sci., p. 14, 2005.
[44] M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," Water Sci. Technol., 2020.
[45] V. Moreido, B. Gartsman, D. P. Solomatine, and Z. Suchilina, "How Well Can Machine Learning Models Perform without Hydrologists? Application of Rational Feature Selection to Improve Hydrological Forecasting," Water, vol. 13, no. 12, p. 1696, 2021.
[46] S. Raghavendra. N and P. C. Deka, "Support vector machine applications in the field of hydrology: A review," Appl. Soft Comput., vol. 19, pp. 372–386, 2014.
[47] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine

learning paradigms in hydrology: A review," J. Hydrol., vol. 598, 2021.
[48] T. Xu and F. Liang, "Machine learning for hydrologic sciences: An introductory overview," WIREs Water, vol. 8, no. 5, 2021.
[49] C. Varadharajan et al., "Can machine learning accelerate process understanding and decision-relevant predictions of river water quality?," Hydrol. Process., vol. 36, no. 4, p. e14565, 2022.
[50] H. A. Afan, A. El-shafie, W. H. M. W. Mohtar, and Z. M. Yaseen, "Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction," Journal of Hydrology, vol. 541. pp. 902–913, 2016.
[51] H. Tao et al., "Artificial intelligence models for suspended river sediment prediction: state-of-the art, modeling framework appraisal, and proposed future research directions," Eng. Appl. Comput. Fluid Mech., vol. 15, no. 1, pp. 1585–1612, Jan. 2021.
[52] C. Shen and K. Lawson, "Applications of deep learning in hydrology," in Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences, wiley, 2021, pp. 285–297.
[53] Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," J. Hydrol., vol. 585, p. 124670, Jun. 2020.
[54] A. Danandeh Mehr, V. Nourani, E. Kahya, B. Hrnjica, A. M. A. Sattar, and Z. M. Yaseen, "Genetic programming in water resources engineering: A state-of-the-art review," Journal of Hydrology. 2018.
[55] S. Ardabili, A. Mosavi, M. Dehghani, and A. R. Várkonyi-Kóczy, "Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review," in Engineering for Sustainable Future: Selected papers of the 18th International Conference on Global Research and Education Inter-Academia–2019 18, 2020, pp. 52–62.
[56] H. S. Munawar, A. W. A. Hammad, and S. T. Waller, "A review on flood management technologies related to image processing and machine learning," Automation in Construction. 2021.
[57] G. Papacharalampous and H. Tyralis, "A review of machine learning concepts and methods for addressing challenges in probabilistic hydrological post-processing and forecasting," Frontiers in Water. 2022.
[58] T. R. Petty and P. Dhingra, "Streamflow hydrology estimate using machine learning (SHEM)," JAWRA J. Am. Water Resour. Assoc., vol. 54, no. 1, pp. 55–68, 2018.
[59] W. Jing et al., "Implementation of evolutionary computing models for reference evapotranspiration modeling: short review, assessment and possible future research directions," Eng. Appl. Comput. Fluid Mech., vol. 13, no. 1, pp. 811–823, 2019.
[60] P. Tahmasebi, S. Kamrava, T. Bai, and M. Sahimi, "Machine learning in geo-and environmental sciences: From small to large scale," Adv. Water Resour., vol. 142, p. 103619, 2020.
[61] M. A. Hariri-Ardebili, G. Mahdavi, L. K. Nuss, and U. Lall, "The role of artificial intelligence and digital technologies in dam engineering: Narrative review and outlook," Eng. Appl. Artif. Intell., vol. 126, 2023.
[62] V. Nourani, A. Hosseini Baghanam, J. Adamowski, and O. Kisi, "Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review," J. Hydrol., vol. 514, pp. 358–377, 2014.
[63] T. Rajaee, H. Ebrahimi, and V. Nourani, "A review of the artificial intelligence methods in groundwater level modeling," Journal of Hydrology. 2019.
[64] R. Haggerty, J. Sun, H. Yu, and Y. Li, "Application of machine learning in groundwater quality modeling-A comprehensive review," Water Res., p. 119745, 2023.
[65] K. P. Tripathy and A. K. Mishra, "Deep Learning in Hydrology and Water Resources Disciplines: Concepts, Methods, Applications, and Research Directions," J. Hydrol., p. 130458,

2023.
[66] H. Tao et al., "Groundwater Level Prediction using Machine Learning Models: A Comprehensive Review," Neurocomputing, 2022.
[67] K. S. M. H. Ibrahim, Y. F. Huang, A. N. Ahmed, C. H. Koo, and A. El-Shafie, "A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting," Alexandria Eng. J., vol. 61, no. 1, pp. 279–303, 2022.
[68] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "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," J. Hydrol., vol. 569, pp. 387–408, 2019.
[69] Q. Xu, Y. Shi, J. Bamber, Y. Tuo, R. Ludwig, and X. X. Zhu, "Physics-aware Machine Learning Revolutionizes Scientific Paradigm for Machine Learning and Process-based Hydrology," arXiv Prepr. arXiv2310.05227, 2023.
[70] S. D. Latif and A. N. Ahmed, "A review of deep learning and machine learning techniques for hydrological inflow forecasting," Environment, Development and Sustainability. 2023.
[71] S. Zhu, H. Lu, M. Ptak, J. Dai, and Q. Ji, "Lake water-level fluctuation forecasting using machine learning models: a systematic review," Environmental Science and Pollution Research. 2020.
[72] A. I. A. Osman et al., "Past, present and perspective methodology for groundwater modeling-based machine learning approaches," Arch. Comput. Methods Eng., vol. 29, no. 6, pp. 3843–3859, 2022.
[73] B. Mohammadi, "A review on the applications of machine learning for runoff modeling," Sustain. Water Resour. Manag., vol. 7, no. 6, p. 98, 2021.
[74] A. Y. Sun and B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," Environ. Res. Lett., vol. 14, no. 7, p. 73001, 2019.
[75] A. V. D. P. Adombi, R. Chesnaux, and M. A. Boucher, "Review: Theory-guided machine learning applied to hydrogeology—state of the art, opportunities and future challenges," Hydrogeology Journal. 2021.
[76] A. Gharib and E. G. R. Davies, "A workflow to address pitfalls and challenges in applying machine learning models to hydrology," Adv. Water Resour., vol. 152, p. 103920, 2021.
[77] M. Karbasi, M. Ali, S. M. Bateni, C. Jun, M. Jamei, and Z. M. Yaseen, "Boruta extra tree-bidirectional long short-term memory model development for Pan evaporation forecasting: Investigation of arid climate condition," Alexandria Eng. J., vol. 86, pp. 425–442, 2024.
[78] J. Quilty and J. Adamowski, "A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes," Environ. Model. Softw., 2020.
[79] Y. Bi, J. Wu, X. Zhai, G. Wang, S. Shen, and X. Qing, "Discriminant analysis of mine water inrush sources with multi-aquifer based on multivariate statistical analysis," Environ. Earth Sci., vol. 80, pp. 1–17, 2021.
[80] G. Schoups, N. C. Van de Giesen, and H. H. G. Savenije, "Model complexity control for hydrologic prediction," Water Resour. Res., vol. 44, no. 12, 2008.
[81] D. P. Solomatine and A. Ostfeld, "Data-driven modelling: some past experiences and new approaches," J. Hydroinformatics, vol. 10, no. 1, pp. 3–22, 2008.
[82] R. M. A. Ikram, A. A. Ewees, K. S. Parmar, Z. M. Yaseen, S. Shahid, and O. Kisi, "The viability of extended marine predators algorithm-based artificial neural networks for streamflow prediction," Appl. Soft Comput., 2022.