Knowledge-based Engineering and Sciences (2024), **5**, (2), 75–139
doi:10.51526/kbes.2024.5.2.75-139

REVIEW PAPER

Reservoir Operation based Machine Learning Models: Comprehensive Review for Limitations, Research Gap, and Possible Future Research Direction

Ahmad Fares Al-Nouti,[*,1] Minglei Fu,[2] and Neeraj Dhanraj Bokde[3]

[1]Civil and Environmental Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[2]College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023, China
[3]Renewable and Sustainable Energy Research Center, Technology Innovation Institute, Abu Dhabi, 9639, United Arab Emirates
[*]Corresponding author. Email: Afalnouti@gmail.com

(Received 05 June 2024; revised 30 July 2024; accepted 26 August 2024; first published online 31 August 2024)

Abstract

The operation of dams and reservoirs is critical for water resource management, including flood control, irrigation, hydropower generation, and environmental conservation. Traditional optimization techniques like Dynamic Programming (DP), Linear Programming (LP), and Nonlinear Programming (NLP) have been foundational in managing these operations. However, they often fall short in addressing the complexities of modern water management challenges posed by climate variability and increasing water demands. Machine learning (ML) techniques have emerged as powerful tools to enhance the efficiency and accuracy of dam and reservoir operations. This paper provides a comprehensive review of various ML models, including Neural Networks, Genetic Algorithms, Decision Trees, and Ensemble Methods, highlighting their applications in predicting reservoir inflows, optimizing water release schedules, and improving flood risk management. Notably, ML models like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) have shown significant improvements in forecasting accuracy and operational decision-making. Despite these advancements, several limitations and research gaps persist, including the need for real-time data integration, adaptive learning mechanisms, and models that consider socio-economic and climatic factors. This review underscores the importance of addressing these gaps to develop more robust and generalizable ML models. Future research directions are suggested to focus on hybrid models combining ML with traditional optimization techniques, comprehensive validation across diverse conditions, and the integration of ecological and economic considerations. By systematically identifying and addressing these limitations, this research aims to pave the way for more effective and sustainable dam and reservoir management practices. Leading towards suggesting that enhancing real-time data integration and developing adaptive learning mechanisms are in order to improve model responsiveness.

Keywords: Reservoir operation; Machine learning; Water resources harvesting; Optimization.

1. Introduction

A reservoir is a man-made lake where water is stored, often created by constructing a dam across a river or waterway [1]. Dams are substantial barriers built to control water flow, maintain water supply, regulate flooding, and generate hydroelectric power [2]. The operation of dams and reservoirs involves managing the inflow and outflow of water, maintaining water levels within certain thresholds, and optimizing the usage of stored water for multiple purposes including irrigation, drinking, and electricity generation. Effective operation can mitigate the effects of droughts and floods, enhance water quality, and support sustainable water management [2]. The importance of optimizing the operations of dams and reservoirs cannot be overstressed. As climate variability intensifies and water scarcity issues become more prominent, the efficient management of water resources becomes critical [3]. Optimization of dam operations ensures that water distribution is managed to meet the competing demands of agriculture, urban areas, and ecological systems. It helps in maximizing hydroelectric power generation while minimizing the adverse impacts of such infrastructure on river ecosystems and downstream communities [4].

The optimal operation of dams and reservoirs is a critical aspect of water resource management, ensuring that water supply, flood control, irrigation, and hydropower generation are effectively balanced. Reservoirs play a vital role in regulating water flow and storing water for various uses, while dams are essential for controlling water levels and generating electricity [4]. However, finding the optimal operational strategy for these infrastructures presents several challenges. These challenges include the variability in water inflows, changing weather patterns, conflicting water use demands, and the complex physical and environmental interactions within the reservoir systems [5]. Traditional methods often fall short in addressing these complexities, resulting in suboptimal performance, inefficiencies, and sometimes even failures in water management and distribution.

Machine learning (ML) offers promising solutions to overcome the limitations inherent in traditional optimization methods. ML techniques can analyze large datasets to identify patterns and make predictions that can significantly enhance the decision-making process for reservoir and dam operations [6], [7]. By leveraging historical data, real-time inputs, and predictive analytics, ML models can optimize operational strategies, improve accuracy in forecasting water inflows and demands, and adapt to changing environmental conditions. The flexibility and learning capabilities of ML models make them suitable for handling the non-linear and dynamic nature of water systems, thereby offering a more robust framework for optimizing dam and reservoir operations [8], [9].

A comprehensive review of the literature reveals significant advancements in applying ML to dam and reservoir management. Studies have employed various ML techniques such as neural networks, genetic algorithms, and decision trees to enhance operational strategies. For instance, neural network models have been used to predict reservoir inflows with high accuracy, while genetic algorithms have optimized release schedules to balance water supply and demand effectively [10]. Decision tree models have been applied to forecast flood risks and improve emergency response strategies [11]. Despite these advancements, there are still notable limitations and gaps in the current research. Many studies rely on specific datasets and conditions, limiting the generalizability of their findings. Additionally, the integration of ML models with real-time data and adaptive management strategies remains underexplored [12].

Identifying the limitations and research gaps in the current literature is crucial for guiding future research towards more effective solutions for dam and reservoir operations. Understanding where existing methods fall short can help concentrate efforts on developing new models and approaches that address these weaknesses. This focused research direction is essential for achieving truly optimal operations that can adapt to the increasing variability and uncertainty in water resource management. By systematically reviewing the existing literature and highlighting these gaps, we can pave the way for more innovative and effective ML applications in this field.

The objective of this review is to compile and analyze studies employing various ML techniques in the context of dam and reservoir operations. Through this comprehensive review, we aim to identify the specific limitations and research gaps that currently exist. Furthermore, we will suggest potential future research directions that could address these gaps, thereby enhancing the effectiveness and reliability of ML models in optimizing dam and reservoir operations. This structured approach will provide valuable insights for researchers and practitioners, ultimately contributing to more sustainable and efficient water resource management practices.

2. Reservoir Operation

Reservoir operation encompasses a wide range of activities aimed at managing and optimizing the use of water stored in a reservoir. These activities are crucial for ensuring the safety, efficiency, and effectiveness of the reservoir in meeting various demands such as water supply, flood control, irrigation, and hydropower generation [13]. Effective reservoir operation requires careful monitoring and control of water levels to balance competing needs and mitigate risks. This involves the use of advanced technologies and predictive models to forecast inflow, outflow, and water quality parameters. Additionally, reservoir operation must account for environmental considerations, ensuring that downstream ecosystems receive adequate water flows to maintain their health [14]. The integration of ML and optimization algorithms has significantly enhanced the ability to manage reservoirs dynamically, responding to changing conditions with greater precision. Ultimately, the goal of reservoir operation is to ensure a reliable and sustainable water resource that supports human activities while protecting the environment [15].

2.1 Reservoir Flood Prediction

Reservoir flood prediction involves anticipating potential flood events to manage water levels effectively and mitigate risks. This process uses historical data, real-time monitoring, and predictive models to forecast inflow volumes and possible flood scenarios [16]. Advanced techniques, such as neural networks and ensemble models, enhance the accuracy of these predictions by analyzing complex data patterns and environmental variables [17]. Accurate flood prediction allows for preemptive water releases, maintaining safe reservoir levels and preventing overtopping. Effective flood management ensures the safety of downstream communities, protects infrastructure, and maintains ecological balance by preventing sudden, excessive water releases. This is critical for safeguarding lives and property in flood-prone areas [18].

Reservoir flood prediction is a critical component of water resource management aimed at anticipating and mitigating the impacts of potential flood events. This operation utilizes a combination of historical data, real-time monitoring, and sophisticated predictive models to forecast inflow volumes and possible flood scenario. Flood prediction models often employ advanced ML techniques, such as neural networks and ensemble methods, which are capable of analyzing complex data patterns and environmental variables to deliver accurate forecasts. The primary goal is to maintain reservoir levels within safe limits, thereby preventing the risk of overtopping and ensuring the safety of downstream communities and infrastructure [19], [20].

Accurate flood prediction enables reservoir managers to implement preemptive water releases, thereby reducing the likelihood of sudden, excessive water discharge that can lead to downstream flooding [21]. These proactive measures are crucial for protecting lives and property in flood-prone areas. Additionally, effective flood management helps maintain ecological balance by ensuring that water releases do not disrupt natural habitats or water quality. Advanced flood prediction systems can integrate real-time data from various sources, including weather forecasts, river flow sensors, and satellite imagery, to provide a comprehensive view of the current and future state of the reservoir [22].

Moreover, predictive models can simulate various flood scenarios, allowing managers to assess the potential impacts of different management strategies and make informed decisions [22]. This capability is particularly valuable in the context of climate change, which is expected to increase the frequency and severity of extreme weather events [23]. By leveraging advanced technologies and data analytics, reservoir flood prediction enhances the ability of water managers to respond to emerging threats and ensure the resilience and sustainability of water resource systems. Ultimately, the integration of flood prediction into reservoir operations represents a significant advancement in mitigating flood risks and optimizing the management of water resources [24].

2.2 Monitoring of Reservoir Inflow

Monitoring reservoir inflow involves tracking the amount of water entering the reservoir from rivers, rainfall, and other sources. This operation is essential for managing water levels within the reservoir to ensure a balance between water supply and flood control. Accurate inflow data helps in planning water releases, optimizing storage capacity, and preparing for drought or flood conditions. Advanced monitoring systems and predictive models enhance the accuracy and timeliness of inflow data [25], [26].

Monitoring reservoir inflow is a vital task in reservoir management, essential for maintaining optimal water levels and ensuring that the reservoir meets the demands of various uses, such as irrigation, domestic water supply, and hydropower generation. This operation involves continuously tracking the volume of water entering the reservoir from sources like rivers, rainfall, and snowmelt. Advanced monitoring systems, equipped with sensors, gauges, and remote sensing technologies, provide real-time data on inflow rates and volumes, enabling accurate and timely decision-making [27].

The integration of ML models significantly enhances the accuracy and reliability of inflow predictions. These models analyze historical data and current conditions to forecast future inflows, taking into account factors such as weather patterns, upstream water use, and environmental changes. By providing precise inflow predictions, these models help reservoir managers optimize water storage and release strategies, ensuring that there is sufficient water available to meet demand while also maintaining adequate storage for flood control and drought mitigation [25]–[27].

Real-time inflow monitoring is crucial for dynamic reservoir operations, allowing managers to respond promptly to changes in water availability. For example, during periods of heavy rainfall, accurate inflow data can help prevent reservoir overflow by enabling timely water releases [28]. Conversely, during dry spells, monitoring inflow can assist in conserving water and prioritizing its allocation to critical uses. Additionally, continuous inflow data supports long-term planning and resource management, aiding in the development of strategies to cope with varying hydrological conditions [29]. Advanced inflow monitoring systems can integrate data from multiple sources, including satellite observations, weather forecasts, and hydrological models, to provide a comprehensive view of the reservoir's inflow dynamics. This holistic approach ensures that all relevant factors are considered in inflow predictions, leading to more effective and sustainable water management practices. In summary, monitoring reservoir inflow is a cornerstone of efficient reservoir management, providing the data needed to balance competing water demands and safeguard against extreme hydrological events.

2.3 Water Release Management

Water release management is the operation of controlling the outflow of water from the reservoir through gates, spillways, or turbines. This operation ensures that the reservoir maintains optimal water levels for various uses, including irrigation, domestic water supply, hydropower generation, and environmental flow requirements. Proper water release management balances the needs of different stakeholders while maintaining safety and operational efficiency [30], [31].

Water release management is a crucial operation in reservoir management, involving the strategic control of water outflows to balance various objectives such as water supply, flood control, irrigation, and hydropower generation [32]. This task requires precise decision-making to determine when and how much water to release, based on factors like current reservoir levels, inflow rates, downstream water needs, and environmental considerations. Effective water release management ensures that water resources are utilized efficiently while minimizing risks and meeting the diverse requirements of stakeholders [33], [34].

Advanced algorithms and real-time monitoring systems play a pivotal role in optimizing water release schedules [35]. These technologies enable reservoir managers to analyze vast amounts of data and make informed decisions. Predictive models, for instance, can forecast future inflows and water demands, allowing managers to plan releases that maintain optimal reservoir levels and prevent both shortages and flooding.[36], [37] By leveraging data analytics, water release strategies can be adjusted dynamically in response to changing conditions, enhancing the flexibility and responsiveness of reservoir operations [38].

In addition to managing routine water releases, this operation is critical during extreme weather events. For example, during heavy rainfall, timely water releases can prevent reservoir overflow and downstream flooding [39]–[42]. Conversely, during drought conditions, careful management of water releases ensures that essential water needs are met without depleting reservoir reserves. This balancing act is vital for maintaining the long-term sustainability of water resources and ensuring the resilience of the reservoir system [43], [44].

Water release management also involves coordinating with various stakeholders, including agricultural users, municipal water suppliers, environmental agencies, and energy producers. Effective communication and collaboration are essential to align water release strategies with the needs and priorities of these groups [45], [46]. Also, regulatory requirements and environmental regulations must be considered to ensure that water releases comply with legal standards and support ecological health. Water release management is a complex and dynamic operation that requires a combination of advanced technology, predictive analytics, and stakeholder coordination. By optimizing water release strategies, reservoir managers can enhance the efficiency and reliability of water resource management, ensuring that water is available when and where it is needed most [47].

2.4 Hydropower Generation Optimization

Dams with hydropower facilities are managed to maximize electricity generation while meeting other operational requirements, as detailed in Figure 1. This involves scheduling water releases through turbines to generate electricity during peak demand periods. Hydropower generation optimization takes into account factors such as water availability, energy demand, and grid requirements [48], [49]. Advanced models and control systems help in making real-time decisions to optimize power production [50].

Hydropower generation optimization focuses on maximizing the efficiency and output of electricity production from reservoirs while balancing other water resource management goals [51]. This process involves strategically scheduling water releases through turbines to generate power during periods of peak electricity demand and storing water when demand is lower [52]. The objective is to ensure that hydropower plants operate at optimal capacity, providing a reliable and sustainable source of energy [53].

ML algorithms, such as reinforcement learning and neural networks, play a significant role in optimizing hydropower generation [54]. These models can predict electricity demand patterns, reservoir inflows, and other relevant factors, enabling managers to make data-driven decisions about water release schedules. By analyzing historical data and real-time inputs, ML models can identify the most efficient strategies for energy production, taking into account variables like weather conditions, inflow rates, and electricity prices [55].

Optimizing hydropower generation involves balancing multiple objectives. In addition to maximizing energy output, reservoir managers must consider environmental flow requirements, flood control, irrigation needs, and recreational uses. Advanced optimization techniques help integrate these competing demands, ensuring that water releases support diverse needs while maintaining the overall efficiency of the hydropower system [56]. Hydropower optimization contributes to grid stability and energy security. By generating electricity during peak demand periods, hydropower plants can reduce reliance on fossil fuels and support the integration of renewable energy sources like wind and solar power [57]. This flexibility is particularly valuable in modern energy systems that require reliable backup power to accommodate fluctuations in renewable energy generation.

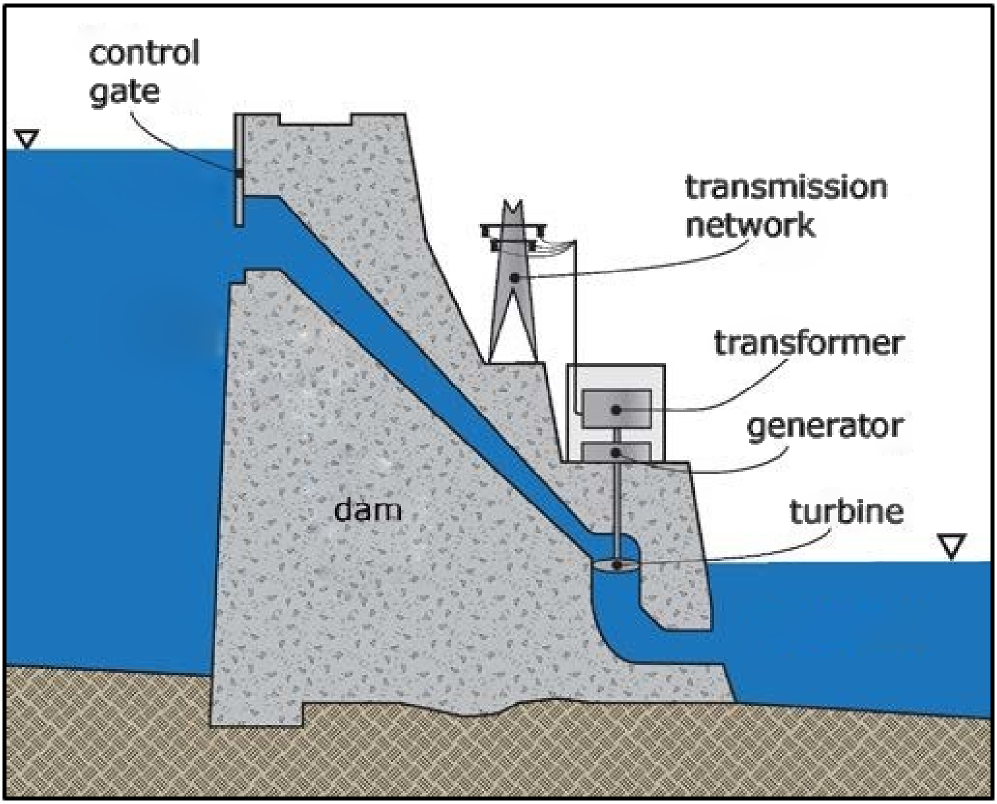

Figure 1: *Hydropower generation components schematic diagram [59].*

Hydropower generation optimization also involves continuous monitoring and adjustment. Real-time data on reservoir levels, inflows, and energy demand allows managers to adapt strategies quickly, responding to changing conditions and unforeseen events. Advanced control systems can automate many aspects of this process, improving responsiveness and reducing the potential for human error. By leveraging advanced ML techniques and real-time data, reservoir managers can achieve a balanced and integrated approach to water resource management, supporting both energy and environmental goals [58].

2.5 Sediment Management

Sediment management involves monitoring and controlling the accumulation of sediments in the reservoir, which can affect water storage capacity and dam operation. Figure 2 shows the areas affected by sediment. Various techniques are employed for sediment management, including dredging, flushing, and sediment bypass systems. Dredging involves physically removing accumulated sediments from the reservoir, which can be resource-intensive but highly effective. Flushing uses controlled water releases to transport sediments downstream, while sediment bypass systems divert sediments around the reservoir, preventing accumulation [60]–[62].

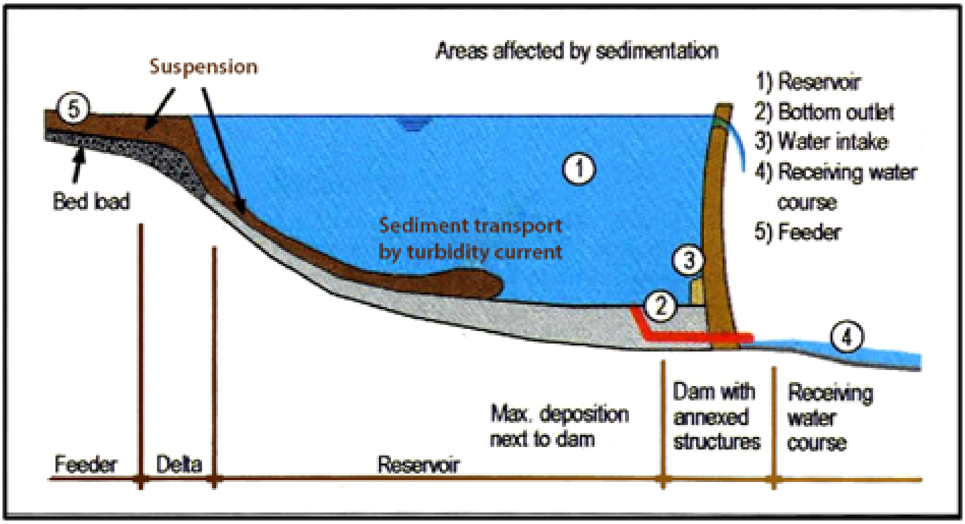

Figure 2: *Areas of the dam affected by sedimentation [68].*

Advanced predictive models and remote sensing technologies enhance sediment management by providing accurate data on sedimentation rates and identifying areas at risk. ML algorithms can analyze historical data and real-time inputs to forecast sediment deposition patterns, enabling proactive management strategies. By predicting where and when sediments will accumulate, managers can schedule maintenance activities more efficiently and reduce the risk of sudden capacity loss [63], [64].

Sediment management is also essential for protecting water quality and maintaining the ecological health of the reservoir and downstream environments. Accumulated sediments can affect water clarity, temperature, and nutrient levels, impacting aquatic habitats and species. By managing sediments effectively, reservoir managers can help preserve biodiversity and support the overall health of the ecosystem [65], [66]. Additionally, sediment management has economic implications. Maintaining reservoir capacity ensures that water storage and hydropower generation can continue at optimal levels, preventing costly disruptions. Effective sediment management strategies can also extend the lifespan of reservoir infrastructure, reducing the need for expensive repairs or replacements. Sediment management is vital for sustaining the operational efficiency and environmental health of reservoirs. By utilizing advanced technologies and predictive models, reservoir managers can implement effective strategies to manage sediments, ensuring the long-term sustainability and functionality of these critical water resources [67].

2.6 Environmental Flow Management

Environmental flow management is the practice of regulating water releases from reservoirs to maintain healthy ecosystems downstream. This approach ensures that the timing, quantity, and quality of water flows mimic natural patterns, supporting the needs of both aquatic and terrestrial habitats [69]. Effective environmental flow management is essential for preserving biodiversity, maintaining water quality, and ensuring the resilience of ecosystems to environmental changes. This involves adjusting water releases to mimic natural flow patterns, supporting fish migration, maintaining water quality, and preserving habitat [70].

One key aspect of environmental flow management is the determination of flow requirements for different species and ecological processes. This involves comprehensive ecological studies to understand the life cycles and habitat needs of various species, particularly fish and invertebrates that are highly dependent on specific flow conditions for spawning, feeding, and migration [71]. These studies help in establishing flow regimes that support the critical ecological functions of rivers and wetlands. Another important component is the integration of real-time monitoring systems. These systems use sensors and remote sensing technologies to collect data on water levels, flow rates, and environmental conditions. Advanced data analytics and ML models analyze this data to predict future conditions and inform management decisions. For example, real-time data on rainfall, temperature, and upstream water use can help predict inflow volumes and optimize water release schedules to meet ecological flow requirements [72]–[74].

Adaptive management strategies are also crucial in environmental flow management. These strategies involve continuously monitoring ecological responses to flow regimes and adjusting management practices based on observed outcomes. This iterative process allows managers to fine-tune water releases to achieve desired ecological objectives while also responding to changing environmental conditions, such as droughts or floods [75]. In addition to ecological considerations, environmental flow management must balance human demands for water, including agriculture, industry, and municipal uses. This requires a holistic approach that integrates ecological and socio-economic factors into water management plans. Collaborative governance involving stakeholders from various sectors ensures that environmental flows are maintained without compromising the water needs of human communities [76], [77].

The benefits of effective environmental flow management extend beyond ecological health. It enhances water quality by diluting pollutants, supports recreational activities such as fishing and boating, and preserves cultural values associated with rivers and wetlands [78]. By maintaining the natural dynamics of water systems, environmental flow management contributes to the sustainability and resilience of ecosystems and human communities alike. Environmental flow management is a complex but vital practice for sustaining riverine and wetland ecosystems. It requires a multidisciplinary approach that combines ecological research, advanced monitoring technologies, adaptive management, and stakeholder collaboration. By ensuring that water flows support both ecological and human needs, this practice plays a crucial role in sustainable water resource management [79], [80].

3. Machine Learning Techniques

ML algorithms are diverse and can be classified into various categories based on their underlying principles, methodologies, and applications. Understanding these categories is essential for selecting the appropriate algorithm for specific tasks, optimizing performance, and addressing the complexities of different data types. This overview provides a brief explanation of key ML algorithm categories, including their characteristics and typical use cases.

Mathematical Programming: It involves formulating and solving optimization problems using techniques like Linear Programming (LP), Nonlinear Programming (NLP), and Integer Programming (IP). These methods are used to find optimal solutions for resource allocation, supply chain management, and decision-making under constraints [81].

Evolutionary Algorithms: Inspired by natural selection, they iteratively improve candidate solutions based on fitness criteria. Common methods include Genetic Algorithms (GA) and Differential Evolution (DE), which are effective for complex optimization problems like scheduling and ML model tuning [82].

Deep Learning and Neural Networks: Deep Learning and Neural Networks consist of layers of interconnected nodes that process data to extract features and patterns. They excel in tasks like image recognition and natural language processing, using models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [83].

Ensemble: Ensemble Learning combines multiple models to improve predictive performance and robustness. Techniques like Bagging, Boosting, and Random Forests reduce overfitting and increase stability, making them suitable for classification, regression, and anomaly detection [84]–[86].

Bayesian Methods: Methods that use Bayes' Theorem to update probabilities based on new data. Algorithms like Naive Bayes and Bayesian Networks are used for probabilistic modeling and decision-making under uncertainty, particularly when data is sparse or incomplete [87], [88].

Decision Tree Algorithms: They create a tree-like model of decisions based on data features. Methods such as CART and C4.5 are used for classification and regression, offering easy interpretability and visualization for exploratory data analysis [89].

Dimensionality Reduction: Techniques to reduce the number of features in a dataset while preserving information. PCA, LDA, and t-SNE help mitigate the curse of dimensionality, enhancing data visualization and improving model performance [90], [91].

Regression: algorithms that predict continuous numerical values from input features. Linear Regression, Polynomial Regression, and Ridge Regression establish relationships between variables, widely used in finance, economics, and environmental modeling [92]–[94].

Regularization: Methods that prevent overfitting by adding a penalty term to the loss function. Ridge Regression (L2), LASSO (L1), and Elastic Net combine penalties to constrain model complexity, improving generalization for regression and classification tasks [95].

Clustering: Clustering algorithms group similar data points into clusters. Techniques like K-Means, Hierarchical Clustering, and DBSCAN are used for exploratory data analysis, pattern recognition, and market segmentation, identifying natural groupings in data [96], [97].

Instance-Based: Instance-Based learning, such as K-Nearest Neighbors (KNN), makes predictions based on the closest instances [98].

An organizational chart illustrating the techniques of ML, along with the reviewed algorithms, is presented in Figure 3. Additionally, other ML algorithms and mathematical methods, such as TLBO and Dynamic Programming, are incorporated. These methods are included due to their significant role in optimizing and/or predicting reservoir operational processes.

4. Literature Review

In the past, the optimization of dam and reservoir operations heavily relied on traditional optimization techniques such as dynamic programming (DP), linear programming (LP), and nonlinear programming (NLP). These methods were foundational in developing operational strategies that aimed to balance the multifaceted demands on water resources, such as irrigation needs, hydroelectric power generation, and flood management. However, as the complexity of water management challenges has grown due to factors like climate change, population growth, and increased environmental concerns, there has been a significant shift towards more advanced solutions. Currently, the focus has shifted to utilizing ML algorithms, which offer enhanced predictive capabilities and adaptive learning opportunities. This modern approach allows for more dynamic and precise management of reservoir operations, catering to the evolving needs and constraints of water resource management.

4.1 Mathematical Programming

Mathematical programming serves as a cornerstone for optimizing various engineering systems, particularly in the study and management of water resources. Techniques like dynamic and linear programming allow for the formulation of optimal solutions to complex problems involving the allocation and management of water in reservoirs and basins. These methods rely on constructing mathematical models that can process large datasets to simulate and predict outcomes under different operational scenarios. By integrating these models with real-world data, engineers and researchers can devise strategies that enhance efficiency, sustainability, and decision-making accuracy in water resource management.

The details in Table 1 are a review of mathematical programming applications across different case studies related to water management. This review elaborates on the optimizers used, the type of

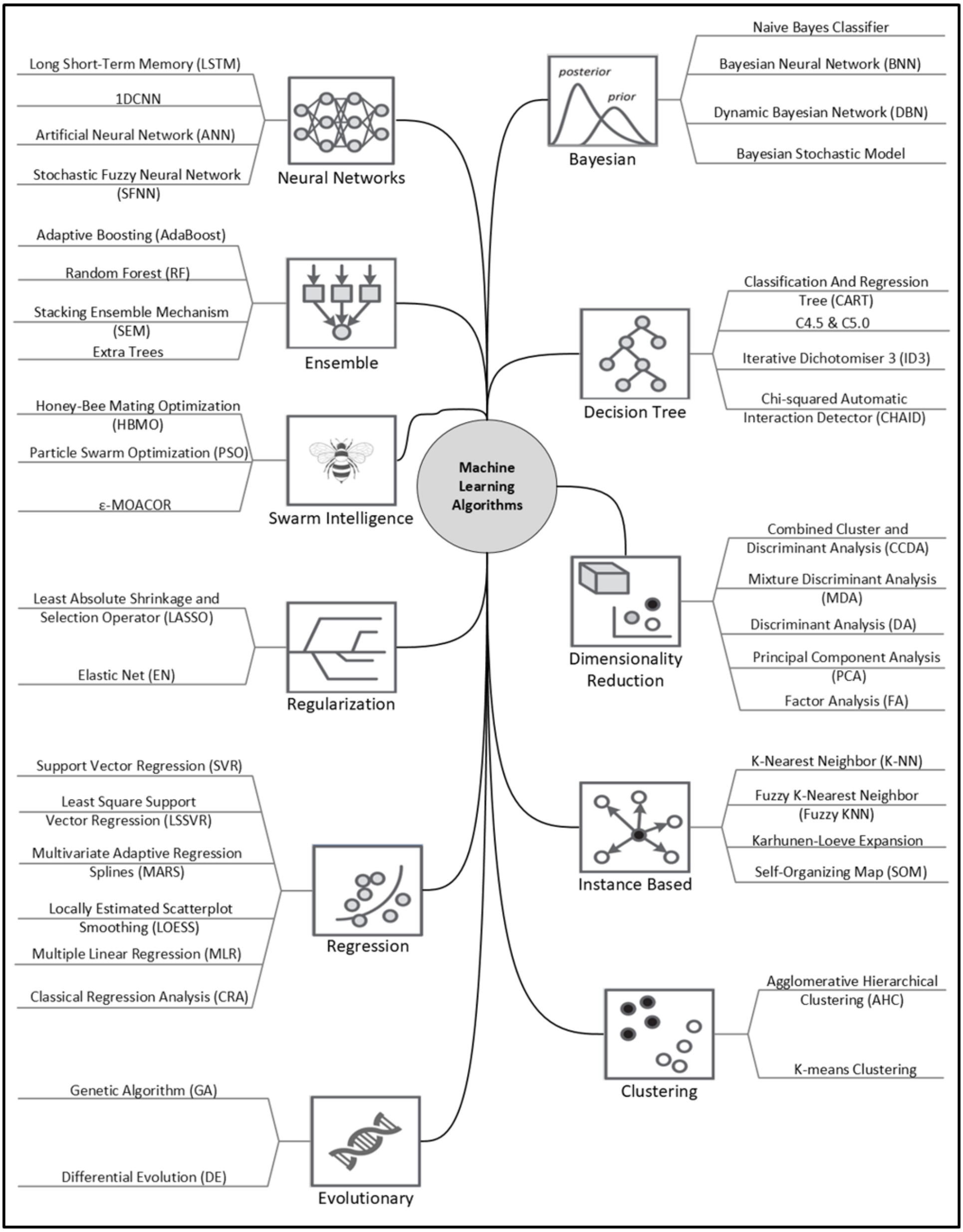

Figure 3: *Reviewed Machine Learning Algorithms classified according to learning style.*

data processed, and the main conclusions derived, thereby showcasing the versatility and impact of mathematical programming in this field.

Table 1: *Overview of Mathematical Programming Applications in Reservoir Operation.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[99]	Finite Dams	Dynamic Programming	Reservoir capacity, discrete inputs, probability distributions	The critical numbers and return functions were characterized, leading to stopping rules for infinite time horizons, improving computational efficiency for large problems.	Study emphasizes the use of successive approximations to determine optimal release rules and critical numbers, focusing on computational savings.
[100]	Multipurpose Reservoir	Dynamic Programming	Seasonal inflows, discharge, and reservoir volume; piecewise linear, concave return functions	Optimal regulation policies were characterized with efficient bounds for special cases and conditions for finite and infinite-time horizon optimality.	The study extends previous research by incorporating seasonal variations and more complex return function structures.
[101]	Finite Dams	Dynamic Programming	Reservoir levels, inflow data, discrete releases	The study presents a dynamic programming approach to derive optimal release policies under varying conditions.	Highlights the use of critical numbers to formulate optimal policies and ensure computational efficiency.
[102]	Columbia River Basin	Nonlinear Programming	Hydroelectric flow data, power generation capacities	The method provided effective optimization for power generation, balancing power deficits across time intervals.	Employed a complex penalty formulation to handle constraints, proving practical for large-scale operations.
[103]	Hypothetical Basin	Linear Programming (LP)	Surface and groundwater interaction data	Integrating detailed simulation of surface and groundwater interactions improves conjunctive water management, providing more efficient use of water resources.	The model effectively uses LP to develop optimal reservoir operation rules and conjunctive use strategies, enhancing water resource management.
[104]	Five-Reservoir System in China (Yangtze River Basin)	Parallel Dynamic Programming (DP)	Reservoir inflow, storage, release, hydropower generation, system efficiency	The parallel DP algorithm optimized the joint operation of the five-reservoir system, resulting in a 4.96x10^8 KW h/year increase in energy production. The system provided more secure and reliable output production, relieving stress at river confluence points and enhancing overall efficiency.	The algorithm proved to be scalable and efficient, with wall clock time reduced from 266.83 h to 1.54 h using 350 peer processes. RMSE values were not explicitly mentioned, but the improvements in operational efficiency and energy production highlight the algorithm's effectiveness.
[105]	Kataj and Latian and Laar Dams	Mixed Integer Linear Programming (MILP)	Reservoir storage and release data, underground water usage	MILP technique achieves optimal reservoir operation, minimizing underground water use.	Results show 15.9% more reservoir storage, 11.6% more outflow, and 21.7% less overflow compared to historical operations. This leads to improved hydropower efficiency and increased stored water in reservoirs.

The studies listed predominantly rely on traditional optimization techniques, which may struggle with computational efficiency when applied to complex or large-scale reservoir systems. Accuracy and availability of data are crucial, and any limitations here can significantly impact model performance. The rigidity of traditional models like DP, LP, and NLP may not accommodate the dynamic and uncertain nature of real-world reservoir systems. Simplifications necessary for these models to function, such as piecewise linear assumptions, may not accurately reflect the intricacies of water management. Furthermore, the specific focus of each study on particular types of dams or basins limits the generalizability of the findings across different contexts or geographic locations.

Traditional literature lacks exploration into the integration of ML algorithms, which could significantly enhance predictive accuracy and adaptability in reservoir management. There is a notable absence of real-time optimization frameworks capable of handling live data streams for on-the-fly decision making. Uncertainty in environmental factors and inflows remains insufficiently addressed, posing a challenge for robust optimization under varying conditions. Comprehensive models that simultaneously consider ecological, quality, and socioeconomic factors are scarce. Lastly, the long-term impacts of climate change on reservoir operations have not been adequately factored into optimization models, representing a significant gap in current research.

To address the limitations of traditional optimization techniques studied from 1960 to 1980, current research focuses on integrating advanced ML algorithms to enhance predictive accuracy and computational efficiency in reservoir management. As we continue to develop real-time optimization frameworks capable of processing live data streams for immediate decision-making, future research should prioritize managing environmental variability and inflow fluctuations through comprehensive uncertainty analysis. Additionally, incorporating ecological, water quality, and socioeconomic considerations will ensure more holistic and sustainable reservoir management practices. Researchers need to focus on the long-term impacts of climate change, developing models that can predict and mitigate these effects over extended periods. Expanding the geographical scope of studies and testing models on various types of dams and basins will enhance the generalizability and applicability of findings, ensuring broader relevance and effectiveness in diverse contexts.

4.2 Evolutionary Optimization Algorithms

Evolutionary optimization algorithms are crucial tools in the field of reservoir operation and management, as they help tackle complex optimization problems that are otherwise intractable using traditional methods. These algorithms, including genetic algorithms (GA), differential evolution (DE), and others, are designed to mimic natural evolutionary processes to iteratively search for more optimal solutions. They have proven particularly effective in managing the multifaceted challenges of reservoir systems, such as water allocation, sediment management, and ecological flow requirements. These techniques can adapt to varying conditions and incorporate multiple objectives, making them ideal for the dynamic and often uncertain environments associated with water resources.

In Table 2, an extensive review of the application of evolutionary optimization algorithms in reservoir management is detailed. The table explores various case studies where these algorithms have been employed, highlighting the data types used, the specific optimization problems addressed, and the outcomes achieved, thus emphasizing the adaptability and efficiency of these algorithms in optimizing reservoir operations.

Table 2: *Applications of Evolutionary Optimization Algorithms in Reservoir Management*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[106]	Mahabad Dam	DE, GA, TLBO	Environmental flows, agricultural demand, and reservoir operations data	DE algorithm performed better than GA and TLBO in terms of efficiency and meeting agricultural and environmental demands.	DE algorithm shows faster convergence and better performance in optimizing reservoir operations under tested scenarios.
[107]	Networking Reservoirs System	CASO and CGA	Water levels, inflow data, reservoir simulation data.	CASO was more efficient in converging to optimum rule curves than CGA, with a 40% faster convergence rate.	CASO and CGA used for optimizing rule curves; CASO proved to be more efficient, reducing overflow durations significantly.
[108]	Changheba Rockfill Dam	GA, NN, OD, FEM	Piezometric head and flow rate measurements	The inverse modeling successfully identified the hydraulic conductivities of foundation rocks and clarified leakage channels.	The approach combined neural networks and genetic algorithms to reduce computational cost and improve result reliability.
[109]	Bocac Dam	GA	Dam behavior data, sensor measurements	The adaptive system effectively manages sensor failures by adjusting the regression model in real-time to maintain prediction accuracy.	The system excels in real-time adjustments during sensor malfunctions, ensuring robust predictions despite variable sensor availability.
[110]	Shih-Men Reservoir	Binary-Coded GA, Real-Coded GA	Water release, reservoir levels, hydrological data	Genetic algorithms effectively improved the rule curves, surpassing the performance of existing M-5 curves in terms of water deficit reduction and hydropower efficiency.	The study showcases the efficiency of real-coded over binary-coded GA, with real-coded providing slightly more accurate and efficient outcomes.
[110]	Tapu Reservoir	Genetic Algorithm (GA)	Reservoir operation data, sedimentation rates, historical water inflow and outflow data	The genetic algorithm optimized the operation rule curves and flushing schedule, improving water management and reducing sedimentation impacts efficiently.	The study demonstrates how GA can effectively handle complex reservoir operations by optimizing both water supply and sediment flushing efficiency.
[112]	Shih-Men Reservoir	Constrained GA (CGA)	Reservoir operation data, hydrological data, ecological base flow requirements	The CGA significantly improved performance in reservoir operations, achieving lower water shortage indices and better ecological flows.	The study demonstrated that CGA could efficiently meet both human and ecological demands by optimizing operation rules.
[113]	Mula Reservoir	Genetic Algorithms (GA), (PSO), Differential Evolution (DE), Artificial Bee Colony (ABC)	Reservoir inflow, Storage, Demand, Release	The integrated meta-heuristic approach effectively optimized the reservoir operations policy. DE and PSO provided the best performance with significant improvements in convergence speed and solution quality. DE achieved the best mean value of 125,825.22 with a standard deviation of 0.07.	The integration of Dynamic Programming with meta-heuristics enhanced the overall performance, reducing computation time and improving the accuracy of the results. RMSE for DE was significantly lower compared to other algorithms.
[114]	Mohamed V Reservoir	Genetic Algorithm (GA) with a new objective function and smoothing constraint	Reservoir inflow, Storage, Demand, Release	The improved GA model effectively optimized the operating curves and hedging rules for the multipurpose reservoir. The new objective function combining the maximum annual deficit and the frequency of shortage provided better performance than the conventional SSD function. The optimization resulted in reduced vulnerability and improved reliability of water supply, especially under drought conditions.	The study highlights the significance of the smoothing constraint in avoiding large fluctuations in operating curves, leading to more practical and stable reservoir management. RMSE values were not explicitly mentioned, but the overall performance metrics showed significant improvements in terms of reduced maximum deficits and frequency of shortages.

The studies exhibit limitations, including dependence on specific optimization algorithms that may not generalize across varying dam operation contexts. Performance metrics are focused on efficiency and convergence, potentially overlooking robustness to environmental unpredictability. Real-time data integration shows promise but is limited by sensor data reliability and computational constraints. While CGA demonstrates ecological consideration, the application to diverse ecosystems with distinct needs remains untested.

The research lacks comprehensive models accounting for extreme environmental variability and long-term climate change impacts. There's an opportunity for algorithms that leverage real-time data to make predictive adjustments. Integrating socio-economic factors into these models remains unexplored, as does the synergy between different ML techniques for more holistic management. Practical field application and validation of these optimized models are not adequately documented.

To overcome these limitations, future research should focus on developing and testing more generalized optimization algorithms that can adapt to various dam operation contexts. Emphasizing robustness to environmental unpredictability alongside efficiency and convergence in performance metrics will provide more comprehensive evaluations. Improving real-time data integration requires enhancing sensor data reliability and addressing computational constraints to fully leverage real-time predictive adjustments. Expanding the application of ecological considerations to diverse ecosystems with distinct needs will ensure broader ecological relevance. Additionally, incorporating socio-economic factors into optimization models will provide a more holistic approach to dam management. Exploring the synergy between different ML techniques can enhance the effectiveness of these models. Finally, conducting practical field applications and thorough validation of optimized models will bridge the gap between theoretical research and real-world implementation, ensuring the developed algorithms are robust, reliable, and effective in diverse scenarios.

4.3 Neural networks and Deep learning

Neural networks and deep learning are pivotal in advancing the automation and optimization of reservoir operations. These technologies leverage large amounts of data to train models that can predict outcomes and optimize processes far more efficiently than traditional methods. Neural networks, including variations like LSTM and convolutional neural networks, are adept at recognizing patterns in sequential data, making them ideal for time-series forecasting such as predicting reservoir levels or inflow rates. Deep learning, with its ability to process and learn from complex and high-dimensional data, enhances decision-making capabilities in real-time operations, accommodating various operational constraints and environmental considerations.

In Table 3, a detailed exploration of how neural networks and deep learning are employed in the management of reservoirs is presented. This includes case studies that demonstrate the integration of these technologies with other optimization techniques, such as genetic algorithms and fuzzy systems, to enhance predictive accuracy and operational efficiency. The table underscores the transformative impact of these advanced computational tools in reservoir management, highlighting their roles in facilitating more informed and strategic decision-making processes.

Table 3: *Integration of Neural Networks and Deep Learning in Reservoir Management*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[115]	Three Gorges Reservoir	LSTM1DCNN, -MOACOR (includes ACOR)	Dam discharge temperature data, reservoir operation data, ecological flow data	The study successfully integrates LSTM1DCNN for accurate temperature predictions and -MOACOR for optimizing reservoir operations.	The integration of deep learning with advanced optimization algorithms enhances the management of ecological and human water needs. ACOR is specifically employed within -MOACOR to improve solution exploration and robustness by providing a guided local search strategy.
[116]	Real-time Reservoir Operation	Genetic Algorithm (GA), Adaptive Network-based Fuzzy Inference System (ANFIS)	Reservoir operation data including water demand and inflow conditions	The integration of GA and ANFIS offers superior performance in predicting and optimizing reservoir operations compared to traditional rule curves.	The study highlights the effectiveness of combining GA for optimization and ANFIS for learning optimal release strategies in real-time operations.
[117]	Shihmen Reservoir Operation	Evolving ANN (Artificial Neural Network), Genetic Algorithm (GA)	Operational data including storage levels, inflow, water demands, and energy production needs	The integration of evolving ANNs with GAs optimized operational strategies, leading to more effective management of water releases and energy production.	The system allows for handling multiple decision variables, significantly improving operational performance compared to traditional rule-based methods.
[118]	Barra Bonita Reservoir	Stochastic Fuzzy Neural Network (SFNN)	Reservoir operation data, stochastic inflow data, operational objectives	SFNN successfully optimized the reservoir operation by integrating fuzzy sets and neural networks, outperforming traditional dynamic programming approaches in handling multiple operational objectives and inflows.	The SFNN method effectively incorporates stochastic elements and fuzzy logic, improving the flexibility and efficiency of reservoir management under uncertain conditions.
[119]	Bhakra Reservoir	Fuzzy Neural Network (FNN)	Reservoir inflow, storage, release, hydropower generation	The integration of fuzzy logic and neural networks effectively optimizes hydrogeneration scheduling under uncertain inflows. The model captures the uncertainty in reservoir inflows and provides robust solutions for maximizing hydropower generation.	The fuzzy neural network showed rapid convergence and high accuracy with a mean square error (MSE) of 0.0026 during the test phase. The hybrid system enhances the reliability and efficiency of reservoir management, handling large numbers of input features effectively.
[120]	Soyanggang Dam	ANFIS (Adaptive Neuro-Fuzzy Inference System)	Monthly dam inflow data, weather forecasting information	The ANFIS model successfully predicts monthly dam inflow with high accuracy. The integration of qualitative weather forecasting information into the ANFIS model significantly improves inflow forecasts compared to using past observed data alone.	The study demonstrates the effectiveness of combining neural networks with fuzzy logic to handle both quantitative and qualitative data. The RMSE values for different model configurations indicate the model's high precision, with significant improvements when using weather forecast information.

The studies show several limitations. The integration of advanced algorithms like *LSTM*1*DCNN* and –*MOACOR* in the Three Gorges Reservoir study may be computationally complex and resource-intensive, limiting real-time application. The Real-time Reservoir Operation study's dependence on GA and ANFIS requires high-quality input data, which can affect performance. The Shihmen Reservoir's evolving ANNs with GAs might face scalability issues with larger datasets. The Barra Bonita Reservoir's SFNN method, though effective, may struggle with interpretability due to complexity. Overall, there's a risk of overfitting, reducing the models' ability to generalize to different conditions or reservoirs.

Several gaps are evident. There's a need for extensive validation of these models across different reservoirs and environmental conditions to ensure robustness. Integration of real-time data and adaptive learning is underexplored but crucial for dynamic responsiveness. Combining optimization techniques with socio-economic factors remains unexplored, which could lead to more holistic management solutions. The studies lack focus on interpretability and usability for reservoir operators, needing user-friendly interfaces and clearer outputs. Finally, incorporating climate change impacts into models is critical for long-term effectiveness in sustainable water management.

To address these limitations, future research should focus on developing lightweight versions of advanced algorithms by simplifying model architectures and using efficient computation techniques like pruning and quantization, thereby reducing computational complexity and enhancing real-time feasibility. Ensuring high-quality input data should be achieved through advanced data preprocessing techniques such as outlier detection, normalization, and data augmentation, along with deploying robust data collection systems with high-precision sensors. Additionally, creating scalable versions of evolving ANNs with GAs is crucial, which can be managed by implementing distributed computing frameworks and parallel processing techniques for larger datasets. To improve interpretability, researchers should develop visualization tools to graphically represent model decisions and outcomes, and incorporate explainable AI techniques like SHAP values or LIME. To reduce overfitting, it is essential to apply cross-validation techniques such as k-fold and stratified sampling, use regularization methods like L1 and L2 penalties, and employ ensemble learning strategies like bagging and boosting. Conducting field validation studies across various reservoirs is also recommended, involving collaboration with local authorities to gather diverse datasets and test models in real-world scenarios. Incorporating adaptive learning mechanisms, such as online learning algorithms and reinforcement learning, will enable continuous model updates based on new real-time data, enhancing responsiveness to changing conditions. Developing user-friendly interfaces with intuitive dashboards will aid reservoir operators in decision-making by presenting clear and actionable model outputs. Integrating climate change scenarios into models involves incorporating climate projections and stress-testing models against various scenarios to assess long-term impacts, ensuring sustainable water management strategies. Finally, fostering cross-disciplinary collaboration by forming interdisciplinary research teams and conducting joint studies will lead to the creation of comprehensive models that incorporate socio-economic factors for holistic management solutions.

4.4 Bayesian Learning

Bayesian learning offers a powerful framework for developing predictive models, particularly effective when the amount of data is limited or when incorporating prior knowledge into the model is crucial. It involves using Bayes' Theorem to update the probability estimate for a hypothesis as more evidence or information becomes available. This approach is highly valued for its ability to handle uncertainty and provide estimations of confidence in the predictions made by the model.

Table 4: *Applications of Bayesian Techniques in Environmental Modeling*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[121]	Nakdong River, South Korea	Naïve Bayes Classifier	Flow data, water quality parameters (BOD, T-P), drought conditions (SPI), monthly data	Naïve Bayes classifier effectively predicted the likelihood of achieving target water quality (TWQ) under varying flow conditions with an accuracy of 72.67% for BOD and 76.46% for T-P during validation. The model demonstrated the importance of considering flow and drought conditions along with water quality parameters for effective watershed management.	The Naïve Bayes classifier provided a quick and efficient method for predicting water quality exceedances, making it a valuable tool for decision-making in watershed management. However, the model's performance is dependent on the quality and comprehensiveness of input data, and further improvement can be achieved through data accumulation and integration of additional variables.
[122]	Serpent River basin and Chute-du-Diable basin, Quebec, Canada	Bayesian Neural Network (BNN)	Daily precipitation, temperature, river flow, and reservoir inflow data	Bayesian Neural Network (BNN) outperformed both standard ANN and HBV-96 models in simulating mean, peak, and low river flows and reservoir inflows. BNN provided reliable predictions with confidence intervals, reducing overfitting and handling parameter uncertainty effectively.	BNN showed superior performance by incorporating uncertainty estimation and avoiding overfitting issues. It provided more accurate and reliable forecasts compared to traditional ANN and conceptual models. However, the model's complexity and computational cost are higher, requiring substantial data quality and quantity.
[123]	Shihmen Reservoir Operation	Bayesian Stochastic Model, SVM	Reservoir operation data, water quality and quantity data	The integration of the Bayesian stochastic model with SVM resulted in effective reservoir operations, enhancing water quality and management. The SVM model successfully refined operation rules by predicting optimal release strategies.	This approach highlighted significant improvements in managing water demands and quality, providing a robust solution for real-time reservoir operations.
[124]	Multi-Reservoir System	Dynamic Bayesian Network (DBN)	Historical flood data, reservoir operation models	The implementation of DBNs for flood control demonstrated effective management of uncertainties, significantly improving the prediction and control of flood risks by allowing for real-time adjustments based on dynamically changing conditions. The method notably enhances the decision-making process under uncertainty, achieving predictive accuracies that effectively support operational adjustments.	The study showcases DBNs as a powerful tool for real-time risk management in complex hydrological systems, leveraging temporal and spatial data variability.
[125]	Heihe River Basin	Dynamic Bayesian Network (DBN)	Reservoir operation data, environmental flow requirements	The decision support model reduced the ecological flow shortage or overflow rate and the economic loss rate by 5% and 6%, respectively.	This model demonstrates how the integration of Dynamic Bayesian Networks can effectively manage reservoir operations, reducing uncertainty and improving ecological and economic outcomes.

[126]	Zayandehrud Dam	Bayesian Networks (BNs)	Annual and monthly dam inflow data, weather forecasting information	BNs efficiently predict annual and monthly dam inflow, considering uncertainties in the data. The model shows a 75% prediction accuracy for annual inflow and 83% for monthly inflow.	Bayesian Networks perform better in predicting inflow ranges rather than numerical values. The RMSE for annual inflow prediction is lower compared to other models, indicating higher accuracy and reliability. The integration of K-means clustering further enhances the model's performance.
[127]	Tankeng Hydropower Station	Dynamic Bayesian Networks (DBNs)	Reservoir inflow, electricity price, hydropower consumption rate	The DBNs model effectively predicts multi-time-scale power output and hydropower generation benefits considering multiple uncertainties. It provides both forward and backward reasoning capabilities, enhancing decision-making for hydropower operations.	The DBNs model demonstrated lower RMSE values compared to traditional models, indicating higher accuracy. The model also provides comprehensive risk assessments for power generation deficiency and power output deficiency.
[128]	Dez and Bakhtiari Reservoirs, Iran	Bayesian Networks (BNs), Varying chromosome Length Genetic Algorithm (VLGA-II)	Monthly inflows, reservoir storages, downstream water demands	BNs are effective for developing monthly operating rules for cascade systems of reservoirs, significantly reducing total damage by 60% compared to fuzzy and classical regression analyses. The average relative error in estimating optimal releases is also reduced by about 30% using BN-based rules.	The integration of VLGA-II with BNs enhances long-term and short-term operation optimization, addressing flood control and agricultural water deficit objectives. The methodology showed significant improvements in operation efficiency and accuracy, with lower RMSE values compared to traditional models.
[129]	Dez Reservoir, Iran	Bayesian Networks (BNs), Dynamic Programming (DP)	Real-time flood forecasting data, reservoir inflows, downstream water demands	BNs combined with DP effectively optimize reservoir operation for real-time flood control. The integration of probabilistic forecasting and optimization reduces peak reservoir releases and downstream flood damages.	The probabilistic approach accounts for uncertainties in inflow forecasts, improving decision-making under uncertain conditions. RMSE values for inflow predictions are significantly reduced, enhancing the accuracy and reliability of flood control operations.

In Table 4, the application of Bayesian techniques in various environmental and hydrological studies is explored. These studies utilize Bayesian methods to enhance prediction accuracy, manage uncertainty, and optimize decision-making processes in complex systems where data may be sparse or highly variable. The table demonstrates how Bayesian networks, stochastic models, and neural networks with Bayesian optimization provide robust solutions for forecasting, resource management, and environmental conservation.

The studies focusing on Bayesian models, such as Naive Bayes Classifier, Bayesian Neural Networks (BNN), and Dynamic Bayesian Networks (DBN), present several limitations. One major limitation is the dependence on high-quality and comprehensive input data, which is crucial for accurate predictions but often challenging to gather. Bayesian models also require substantial computational resources, particularly for complex applications like flood prediction and reservoir inflow forecasting, limiting their scalability for real-time operations. The complexity of these models makes them difficult to implement and interpret, requiring specialized knowledge that may not be readily available in all operational contexts. Additionally, while Bayesian models can effectively handle parameter uncertainty and provide confidence intervals, they may still exhibit sensitivity to initial assumptions and prior distributions, which can influence the outcomes. Integration of additional variables to enhance model performance remains a significant challenge due to data availability and processing constraints.

There are several research gaps identified in the studies focused on Bayesian models. Future research should focus on developing methods to improve data integration and preprocessing to enhance the robustness and accuracy of Bayesian models. Exploring hybrid approaches that combine Bayesian methods with other ML techniques could address the limitations of individual models and improve predictive performance. There is also a need for more efficient algorithms to reduce computational demands and enable real-time application. Comprehensive validation across different hydrological and climatic conditions is essential to ensure the generalizability of these models. Additionally, incorporating socio-economic factors and climate change impacts into Bayesian models could provide more holistic solutions for water management. Enhancing the interpretability and usability of Bayesian models for practical decision-making is crucial, particularly for stakeholders without specialized knowledge. Addressing these gaps will advance the effectiveness and adaptability of Bayesian models in reservoir and dam management, ensuring they can meet the evolving challenges of water resource management.

Recommendations for future research are that future studies should prioritize improving data integration and preprocessing techniques. This can be done by developing advanced methods for outlier detection, data normalization, and augmentation, along with deploying robust data collection systems to ensure high-quality input data. To mitigate the substantial computational resource demands, it is recommended to explore more efficient algorithms and leverage techniques like parallel processing and cloud computing, making these models more scalable for real-time operations. Enhancing the interpretability of Bayesian models is crucial and can be achieved through visualization tools and explainable AI methods, such as SHAP values or LIME, which can make model outputs more understandable for practitioners without specialized knowledge. Exploring hybrid approaches that combine Bayesian methods with other ML techniques, like neural networks or decision trees, can also address the limitations of individual models and improve predictive performance. Comprehensive validation of these models across diverse hydrological and climatic conditions is necessary to ensure their robustness and generalizability. Additionally, integrating socio-economic factors and climate change projections into Bayesian models will provide more holistic and sustainable water management solutions. By incorporating adaptive learning mechanisms, such as online learning and reinforcement learning, Bayesian models can continuously update based on new data, enhancing their responsiveness to changing conditions. These steps will advance the effectiveness and adaptability of Bayesian models in reservoir and dam management, ensuring they can meet the evolving challenges of water resource management.

4.5 Ensemble Algorithms

Ensemble algorithms are a sophisticated set of techniques in ML that improve predictive performance by combining multiple models. These methods typically produce more accurate results than any single model could achieve on its own, by leveraging the strengths and mitigating the weaknesses of various individual models. Ensemble techniques such as bagging, boosting, and stacking are utilized to reduce variance, bias, or improve predictions, making them particularly effective for complex problems where single model predictions may fall short due to overfitting or inherent model limitations.

In Table 5, an array of ensemble learning models and their applications in hydrological forecasting and water management are detailed. The table illustrates how different ensemble techniques, including random subspacing, CatBoost, XGBoost, and Random Forest, are applied to enhance accuracy in predicting water levels, inflow, and other critical environmental factors. These case studies demonstrate the robustness and versatility of ensemble methods in handling diverse data and complex prediction tasks.

The studies on ensemble models, including CatBoost, XGBoost, Random Forest (RF), and Stacking Ensemble Mechanism (SEM), exhibit several limitations. A primary limitation is the dependence on high-quality input data, which is crucial for achieving accurate predictions but can be difficult to obtain. These models also require extensive hyperparameter tuning and computational resources, which can be time-consuming and limit their scalability for real-time applications. Additionally, while ensemble models generally improve predictive performance, they may still exhibit higher errors for specific inflow ranges, as seen with XGBoost and LGBM for moderate inflow values. Another limitation is the complexity of interpreting ensemble model outcomes, particularly when multiple models are combined, making it challenging for non-experts to understand and implement the results effectively. Furthermore, the integration of various data types, such as climatic indices and land-use data, necessitates robust data preprocessing and management, which can add to the complexity and computational burden.

Several research gaps are identified in the studies focused on ensemble models. There is a need for improved methods to handle data sparsity and variability, ensuring robust predictions across diverse hydrological scenarios. Future research should explore the development of more efficient algorithms that reduce computational demands and enhance real-time applicability. Investigating hybrid models that combine ensemble techniques with other ML methods could further improve predictive accuracy and robustness. Comprehensive validation of these models across different geographical locations and climatic conditions is essential to ensure their generalizability. Additionally, incorporating socio-economic factors and climate change impacts into ensemble models could provide more holistic water management solutions. Enhancing the interpretability and usability of ensemble model outcomes for practical decision-making is also crucial. Finally, focusing on integrating additional hydrological and meteorological variables can improve the models' ability to capture complex interactions and predict extreme events more accurately. Addressing these gaps will advance the effectiveness and adaptability of ensemble models in reservoir and dam management.

Table 5: *Applications of Ensemble Learning Techniques in Hydrological Prediction*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[130]	Hydroelectric Reservoir (690 MW)	Ensemble Learning (Bagging, Boosting, etc.)	1-hour ahead water level forecasts, dam metrics	Ensemble models, especially the random subspace model, demonstrated high accuracy with an RMSE of 1.58 and an MAE of 0.24, significantly outperforming LSTM models in reservoir operation forecasting.	The study highlights the superiority of ensemble learning models over LSTM in providing reliable short-term water level forecasts, crucial for effective reservoir management and emergency response in hydroelectric settings.
[131]	Garudeshwar Watershed, India	CatBoost, XGBoost, Light Gradient Boosting Machine (LGBM), Random Forest (RF), MLP	River inflow data, precipitation data, temperature data	CatBoost demonstrated superior performance with the highest accuracy across various metrics including MAE, RMSE, and R^2 values. CatBoost managed categorical and continuous variables effectively, leading to significantly enhanced prediction accuracy. XGBoost and LGBM showed higher prediction errors for moderate inflow values above 10,000 m^3/s.	The study highlighted the effectiveness of ensemble algorithms in hydrological prediction, with CatBoost outperforming other models. However, it also noted that XGBoost and LGBM had higher errors in certain inflow ranges, indicating a need for further refinement in these models.
[132]	Gaborone and Bokaa dams, Botswana	Random Forest Regression (RFR), MLP Neural Network, Vector AutoRegression (VAR)	Monthly dam water levels, rainfall, temperature, climate indices (DSLP, AI, SOI, Niño 3.4), land-use and land-cover data	RFR and MLP-ANN models showed significant correlations between dam water levels and climate factors, with R^2 values of 0.890 to 0.926 for Gaborone and 0.704 to 0.865 for Bokaa. MLP-ANN provided the best prediction results for dam water levels, with the highest R^2 value. The study found that integrating LULC and climate conditions significantly improved prediction accuracy.	While RFR performed better with LULC data, MLP-ANN excelled with climate factors. The study highlighted the effectiveness of hybrid models, such as VAR-ANN, for capturing both linear and nonlinear relationships in time-series data. However, accurate predictions depend on high-quality input data and thorough model tuning.
[133]	Sri Ram Sagar Project (SRSP), Telangana, India	Stacking Ensemble Mechanism (SEM) incorporating XGBoost, LGBM, and RF	Daily reservoir inflow, rainfall, evaporation, climate indices (GBI, NINO 3, NINO 3.4, NINO 4, GH, DMI)	SEM outperformed individual models with KGE values of 0.94 (training) and 0.89 (testing), demonstrating superior prediction accuracy and effective simulation of peak inflow events.	SEM is highly effective for inflow prediction, but relies on high-quality input data and thorough hyperparameter tuning. The study indicates potential for further improvements by incorporating additional hydrological and meteorological variables.
[134]	Sirikit Dam	XGBoost	Reservoir inflow, precipitation, and humidity	The XGBoost models demonstrated excellent predictive performance with a daily RMSE of 8.3666 and a monthly RMSE of 372.6547, significantly refining the accuracy of inflow forecasts. This performance highlights the effectiveness of XGBoost in handling complex hydrological data and enhancing decision-making in reservoir management.	These results highlight the robustness of XGBoost in handling complex hydrological data, enhancing decision-making in reservoir management.

[135]	Baozhusi Hydropower Station	VMD-BiLSTM (Variational Mode Decomposition - Bidirectional Long Short-Term Memory)	Daily reservoir inflow data, weather forecasting information	The VMD-BiLSTM model outperforms traditional and other hybrid models in predicting daily reservoir inflow. It achieves an RMSE of 64.783 and NSE of 95.7%, indicating high prediction accuracy and reliability.	The combination of VMD with BiLSTM enhances the model's ability to handle non-stationary and nonlinear inflow data. The ensemble approach reduces noise and improves forecast stability, making it a preferred tool for reservoir inflow prediction.
[136]	Krasnodar Krai Rivers (Pshish and Mzymta)	Ensemble Algorithms (M5P, XGBoost, MLP)	Water level data from automated hydrological complexes, weather forecasting information	Ensemble algorithms like XGBoost and MLP show high prediction accuracy for water levels with significant improvement over single models. The optimal lead time for predictions varies between 15 to 18 hours for Pshish River, with XGBoost showing the best performance.	The ensemble approach enhances the robustness of the model by effectively capturing nonlinear relationships and reducing errors. RMSE values were significantly lower with ensemble algorithms compared to traditional methods, indicating higher reliability and accuracy.
[137]	Carlyle Lake and Lake Shelbyville	Random Forest (RF)	Hydrological and meteorological data, including past, current, and future inflow, storage, and precipitation data	The RF model outperformed other models in predicting reservoir outflow with the lowest RMSE values. RF provided a robust and reliable prediction for both Carlyle Lake and Lake Shelbyville, achieving high accuracy in simulating daily outflow patterns.	RF model showed superior performance with RMSE values of 107.6 (training) and 108.2 (testing) for Carlyle Lake, and 34.5 (training) and 35.9 (testing) for Lake Shelbyville. The ensemble approach of RF effectively handled large datasets and complex nonlinear relationships, making it suitable for long-term reservoir management.

To address the limitations identified in studies on ensemble models like CatBoost, XGBoost, Random Forest (RF), and Stacking Ensemble Mechanism (SEM), future research should prioritize developing methods to handle data sparsity and variability. This can be achieved by implementing advanced data augmentation techniques and robust preprocessing systems to ensure consistent input quality across diverse hydrological scenarios. Reducing the computational demands of these models is essential, which can be accomplished by optimizing algorithms and leveraging parallel processing and cloud computing for more efficient hyperparameter tuning and scalability in real-time applications. Exploring hybrid models that combine ensemble techniques with other ML methods, such as neural networks or Bayesian models, can further enhance predictive accuracy and robustness. Comprehensive validation across different geographical locations and climatic conditions is crucial to ensure model generalizability. Integrating socio-economic factors and climate change impacts into ensemble models will provide more holistic water management solutions. Enhancing the interpretability of ensemble model outcomes can be done through visualization tools and explainable AI methods like SHAP values or LIME, making results more accessible to non-experts. Additionally, integrating various hydrological and meteorological variables, including climatic indices and land-use data, will improve the models' ability to capture complex interactions and predict extreme events more accurately. By addressing these research gaps, the effectiveness and adaptability of ensemble models in reservoir and dam management can be significantly advanced, ensuring they meet the evolving challenges of water resource management.

4.6 Decision-tree Algorithms

Decision-tree algorithms are a fundamental class of ML techniques known for their simplicity and effectiveness in handling classification and regression tasks. These algorithms partition the data into subsets based on different criteria, using simple decision rules inferred from the features. In the context of reservoir management, decision trees can analyze various operational data points, such as hydrological data, inflow rates, and energy production levels, to make informed decisions about water release schedules, energy generation, and flood control. The advantage of decision trees lies in their ability to provide transparent, understandable models that can be easily validated and adjusted according to new data or changing operational requirements.

In Table 6, the utilization of decision-tree algorithms in reservoir management is examined through several case studies. Each study highlights the adaptability and precision of decision trees in forecasting and real-time operational decision-making. The table reviews different enhancements to traditional decision tree models, such as integration with other ML techniques or improvements in validation methods, illustrating their robustness and reliability in complex environments.

The studies focusing on decision-tree algorithms reveal several limitations. Decision-tree models, while effective in forecasting and control, can be prone to overfitting, especially with complex and varied data. This necessitates robust cross-validation techniques, as highlighted in the CART model study. The reliance on historical and operational data means these models may not always adapt well to sudden or extreme changes in environmental conditions. Additionally, the scalability of these models to larger, more diverse datasets can pose a challenge. The studies also point to varying effectiveness across different scenarios, indicating a need for careful model selection and tuning based on specific use cases.

Table 6: *Applications of Decision-Tree Algorithms in Reservoir Management.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[138]	Tianshengqiao Cascaded Hydropower Plants	Decision Tree Algorithms (C4.5, CHAID, ID3-IV)	Hydrological data, energy production data, operational schedules	The study validated that the decision tree algorithm C4.5 was the most effective, showing minimal error rates in forecasting generation schedules for cascaded hydropower plants.	The C4.5 algorithm outperformed other decision tree algorithms by providing quicker and more accurate predictions for hydro scheduling, enhancing operational efficiency.
[139]	Rizhao Reservoir	Decision Tree (C4.5)	Reservoir operation data, rainfall, inflow data	The DT-based method enabled effective real-time reservoir flood control operation, demonstrating adaptability and the ability to guide operational decisions efficiently.	The study highlights the application of DT-based rules in real-time operations, improving the handling of dynamic inflow scenarios and ensuring operational adaptability.
[140]	Jiroft Dam	ANN with PSO and HS, MLR, Decision Trees (CDT, RDT)	Piezometric water level, reservoir inflow, evaporation rates, sluice gate outflow, intake outflow, total outflow	ANN-PSO showed the best performance for Piezometer 20 and Piezometer 28 with correlation coefficients (R) of 0.990 and 0.945 respectively; MLR provided optimal results for Piezometer 30 with an R of 0.945.	These models demonstrate varying degrees of effectiveness, with ANN-PSO generally outperforming others in terms of prediction accuracy for specific piezometers.
[141]	Shasta Lake	Dynamic Merging (DMerge), Decision Trees (AdaBoost, Extra Trees)	Hydropower release data, hydrological variables, climate indices	The DMerge method outperformed traditional decision tree methods (AdaBoost, RF, Extra Trees), demonstrating superior accuracy in simulating hydropower discharge with a CORR of 0.959, RMSE of 551.169 m^3/s, and KGE of 0.934 during validation.	DMerge efficiently integrates outputs from various decision tree models, dynamically adapting to changes in data inputs and model performance, enhancing prediction accuracy.
[142]	Major California Reservoirs	CART combined with Shuffled Cross-Validation, Random Forest	Reservoir operation data, hydrological data	The enhanced CART model with shuffled cross-validation outperformed both the standard CART and Random Forest models in simulating reservoir outflows, particularly in capturing peak flow events more accurately.	The use of shuffled cross-validation in CART significantly reduces the risk of overfitting and improves the model's ability to generalize across different reservoir conditions.
[143]	Jiroft Dam, Iran	Classification and Regression Tree (CART), Chi-squared Automatic Interaction Detector (CHAID)	Evaporation data, dam input water volume, dam output water volume	Both CART and CHAID models were effective in predicting water sales revenue. The CART model performed better in terms of accuracy, with a lower RMSE of 649.97 during the training phase and 894.25 during the testing phase. CHAID showed slightly less accuracy with RMSE values of 630.005 and 834.60 for training and testing phases, respectively.	Decision Tree algorithms like CART and CHAID are highly suitable for modeling economic features of water revenue, providing detailed decision-making support. The RMSE values indicate that these models can predict revenue with reasonable accuracy.
[144]	Fill Dam (specific name not provided)	Decision Tree Analysis (CART)	Seepage data, water level, daily rainfall, antecedent 5-day rainfall	Decision Tree Analysis using the CART algorithm effectively models the relationship between seepage, water level, and rainfall. The study classified seepage data into rainfall-free and rainfall-occurring groups, identifying key variables impacting seepage rates.	The model demonstrated high accuracy in predicting seepage under varying conditions. RMSE values were not explicitly mentioned, but the results indicated a significant reduction in prediction errors. The approach provides clear insights into the factors affecting seepage, aiding in effective dam management.

[145]	High Aswan Dam Reservoir, Nile River	Policy Tree Optimization Model	Inflow data, outflow data, evaporation data, historical reservoir level data	The policy tree optimization model effectively optimizes reservoir operations by balancing flood control, water supply, and hydropower generation. The model demonstrates significant improvements in operational efficiency, reducing overflow incidents and enhancing water storage management.	The Decision Tree approach used in the policy tree model provides a robust framework for making informed operational decisions. RMSE values indicate high accuracy in predicting reservoir levels and optimizing water release policies, leading to more reliable reservoir management.

Several research gaps are evident in the decision-tree algorithm studies. There is a need for improved methods to prevent overfitting and enhance the generalizability of these models across different reservoir conditions. Integrating real-time data streams and adaptive learning mechanisms can improve model responsiveness to dynamic changes. Further exploration into hybrid models combining decision trees with other ML techniques, such as ensemble or neural network methods, could enhance predictive accuracy and robustness. Comprehensive validation across different geographical and climatic conditions is essential to ensure model reliability. Finally, incorporating socio-economic and climate change factors into decision-tree models would provide more holistic and sustainable water management solutions.

To address the limitations identified in studies focusing on decision-tree algorithms, future research should consider several innovative approaches. Developing decision-tree algorithms that incorporate uncertainty quantification can improve robustness in the face of sudden environmental changes, which can be achieved by integrating probabilistic methods that provide confidence intervals for predictions. Utilizing transfer learning techniques can help these models adapt to new datasets with limited historical data, enhancing applicability across diverse scenarios. Implementing advanced feature selection methods, such as genetic algorithms or mutual information-based techniques, can identify the most relevant variables and reduce model complexity. Exploring multi-objective optimization can balance trade-offs in water management, such as efficiency, sustainability, and economic factors, thereby developing more comprehensive decision-making frameworks. Additionally, creating modular and interoperable software platforms for decision-tree models will facilitate integration with existing water management systems, enhancing usability for practitioners. Encouraging participatory modeling involving stakeholders can ensure models address practical needs and incorporate valuable local knowledge. Lastly, focusing on interpretable ML methods, such as rule-based models derived from decision trees, will make results more accessible to non-experts. By adopting these strategies, future research can enhance the robustness, adaptability, and practical applicability of decision-tree algorithms in reservoir management.

4.7 Regularization

Regularization techniques in ML are crucial for improving model performance by reducing overfitting, enhancing generalization to unseen data, and simplifying models by penalizing complexity. These techniques adjust the learning process to discourage complex models unless the data provides sufficient support for such complexity. Common regularization methods include LASSO (Least Absolute Shrinkage and Selection Operator), Ridge Regression, and Elastic Net, which combine the penalties of LASSO and Ridge. These methods are particularly useful in scenarios with high-dimensional data or when the risk of overfitting is substantial.

In Table 7, the role of regularization techniques in managing and optimizing reservoir operations is detailed. The table showcases various case studies where regularization methods have been integrated with other ML models to enhance predictive accuracy and operational efficiency in reservoir management. Each entry exemplifies how these techniques not only improve model predictability but also help in handling large datasets, managing sparsity, and maintaining robustness under varying operational conditions.

Table 7: *Impact of Regularization Techniques on Reservoir Management Models.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[146]	Flood Embankments Monitoring	Elastic Net (EN), ANN, Gauss-Newton, Level Set	Electrical impedance tomography data, conductivity measurements	The hybrid method combining EN with ANN significantly improved the accuracy of the reconstructed images, achieving high-resolution imaging of flood embankments with an R-squared value of 0.9993 in the testing set.	While EN effectively reduces input data dimensions, the performance is highly dependent on the quality of the initial measurements, and the approach may be computationally intensive.
[147]	Kanayama Dam and Satsunaigawa Dam	Elastic Net (EN), Fully Connected Neural Network (FCNN), Recurrent Neural Network (RNN), Random Forest (RF)	Hydrological data, rainfall data, inflow data	The EN model achieved high prediction accuracy with Nash-Sutcliffe (NS) coefficients of 0.7 or greater, demonstrating robustness in predicting dam inflow during severe floods.	EN effectively handled sparse data and improved the accuracy of inflow predictions without the need for highly accurate predicted rainfall. However, its performance can be affected by the quality and resolution of input data.
[148]	Texas Reservoirs	LASSO, Adaptive LASSO, Group LASSO	Reservoir levels, storage capacities, conservation pool heights	LASSO methods, particularly adaptive LASSO, effectively identified key predictors for reservoir levels, achieving a high predictive accuracy with an Adjusted R^2 of 0.9112 and reducing the model complexity by eliminating irrelevant variables.	The study demonstrates LASSO's capability to improve model interpretability and prediction accuracy, although its performance may vary with the quality of the input data and the specific configuration of reservoir characteristics.
[149]	Colorado River Basin	Least Absolute Shrinkage and Selection Operator (LASSO)	Streamflow data, dam attributes, fish biodiversity metrics	LASSO regression identified that 63% of the variation in flow alteration was explained by network-level attributes and 37% by local dam properties. This indicates that the location and cumulative upstream regulation are more significant in determining flow alteration impacts.	The study emphasizes the need for basin-wide reoperation strategies to mitigate flow alteration and its effects on biodiversity. Limitations include the potential impact of smaller, unaccounted-for dams and the challenges of fully capturing climate-driven changes.
[150]	Sedre River, Turkey	LASSO combined with Multi-Stage Genetic Programming (MSGP)	Streamflow data, historical monthly streamflow records	The MSGP-LASSO model significantly improved streamflow forecasting accuracy, achieving an NSE of 0.64 during the training period and 0.59 during the testing period, demonstrating a marked improvement over traditional GP models.	The hybrid model effectively handles complex hydrological data, but the presence of negative predictions in low-flow periods indicates the need for further refinement.
[151]	Tennessee River Basin, USA	LASSO, FCM (Fuzzy C-means), DBN (Deep Belief Network)	Monthly streamflow, rainfall, SST, climate indices	The LASSO-FCM-DBN model significantly improved streamflow prediction, achieving an R^2 of 0.86 and RMSE of 305 during validation, demonstrating higher accuracy and stability compared to traditional ANN models.	The hybrid model's performance may be affected by the quality and granularity of input data, and further refinement may be required for low-flow predictions.

[152]	Cahora Bassa Dam, Mozambique	LASSO, XGBoost, ELM, SVR, MARS, GWO	Streamflow data, rainfall, evaporation, humidity	The XGBoost-LASSO hybrid model outperformed other models, achieving an RMSE of 0.124 and an NSE of 0.961 for one-day-ahead predictions, highlighting its efficiency in streamflow forecasting.	While the XGBoost-LASSO model showed high accuracy, the complexity of integrating multiple models and tuning parameters requires significant computational resources. Further research may be needed to refine the model for different hydrological conditions.
[153]	Tennessee River Basin	LASSO, EEMD, DBN	Monthly streamflow data, precipitation, SST, atmospheric circulations	The EEMD-Lasso-DBN model significantly improved the accuracy of monthly streamflow forecasting, achieving an R^2 of 0.922 and an RMSE of 158 during calibration, and an R^2 of 0.885 and an RMSE of 189 during validation.	The combination of EEMD and LASSO for predictor selection and DBN for prediction provides a robust framework for streamflow forecasting, but the model's complexity and computational demands are high.
[154]	Three-River Headwaters Region (TRHR)	Lasso-DBN-Bootstrap (Least Absolute Shrinkage and Selection Operator, Deep Belief Networks, Bootstrap)	Monthly streamflow data, weather forecasting information	The Lasso-DBN-Bootstrap model significantly improves the performance and stability of streamflow forecasting. It effectively screens predictors, models complex relationships, and assesses forecasting uncertainty. The model outperforms traditional forecasting methods, achieving lower RMSE and higher accuracy.	Regularization via Lasso enhances the predictor selection process, reducing overfitting and improving model interpretability. The integration with DBN and Bootstrap methods further enhances the robustness and reliability of forecasts. The RMSE values indicate a substantial improvement over conventional models.
[155]	Haihe River Basin	Lasso Regularization integrated with LSTM, TTS, RF	Streamflow data, precipitation, temperature, humidity, wind speed	The Lasso Regularization method significantly enhances the performance of LSTM, TTS, and RF models by selecting key features and reducing overfitting. The Lasso-LSTM model shows superior stability and generalization capabilities in streamflow prediction.	The Lasso method effectively reduces the dimensionality of input features, leading to improved model accuracy and generalization. The Lasso-LSTM model achieves an NSE of 0.84 on the test set, with lower RMSE values compared to models without Lasso integration.
[156]	Zambezi River Basin, Cahora-Bassa Dam, Mozambique	Elastic Net Regularization (EN)	Streamflow data, precipitation, evaporation, relative humidity	The Elastic Net Regularization (EN) model significantly enhances the predictive capabilities of ML models for streamflow forecasting. It effectively selects relevant features and prevents overfitting, leading to improved accuracy in flow predictions.	The EN model demonstrated superior performance, with lower RMSE values compared to traditional models, indicating high accuracy in predicting streamflow. The integration of EN with other ML models provided robust and reliable forecasts, crucial for reservoir management.

The studies focused on regularization models, including Elastic Net (EN), LASSO, and their combinations with other techniques, exhibit several limitations. The effectiveness of these models is highly dependent on the quality and completeness of the input data, which can be variable and sparse. Regularization methods like EN and LASSO can help mitigate overfitting but may struggle with extremely noisy or incomplete datasets. Hybrid models combining regularization with other techniques, such as ANN or DBN, can be computationally intensive and complex to implement, potentially limiting their practical application. Additionally, these models may require extensive tuning of hyperparameters, and their performance can be sensitive to the specific configuration of reservoir characteristics.

The studies reveal several research gaps in the application of regularization models. There is a need for improved methods to handle data sparsity and noise, ensuring the robustness of the models across diverse scenarios. Further exploration into real-time data integration and adaptive learning mechanisms could enhance the responsiveness and adaptability of these models to dynamic changes in reservoir conditions. Developing more efficient and user-friendly techniques for hyperparameter tuning is essential to simplify the implementation and improve the performance of hybrid models. Comprehensive validation of these models across different geographical locations and environmental conditions is necessary to ensure their generalizability. Finally, incorporating socio-economic and climate change factors into regularization models could lead to more holistic and sustainable water management solutions.

To shrink the limitations of regularization models such as Elastic Net (EN) and LASSO, future research should focus on several key areas. Developing methods to incorporate uncertainty quantification into regularization models can provide more robust predictions under varying data quality and completeness, using techniques like Bayesian regularization. Implementing ensemble methods that combine multiple regularization models can help mitigate the impact of noisy or incomplete datasets by leveraging the strengths of different algorithms. Creating modular frameworks that allow easy integration of regularization models with other ML techniques can reduce computational complexity and facilitate practical application. Exploring the use of meta-learning approaches can streamline the process of hyperparameter tuning, making it more efficient and less dependent on extensive manual intervention. This can be achieved by developing meta-models that learn the optimal hyperparameter configurations based on past experiences. Additionally, developing interactive tools that allow users to visualize and interpret the influence of hyperparameters on model performance can enhance usability. Research should also focus on the development of adaptive regularization techniques that dynamically adjust the regularization strength based on real-time feedback from the model's performance, which can be implemented through reinforcement learning algorithms. Ensuring comprehensive validation of these models across various geographical locations and environmental conditions is crucial, which can be achieved by establishing international collaborations and data-sharing platforms. Lastly, incorporating multi-criteria decision analysis into regularization models can provide a more balanced approach to water management by considering various conflicting objectives such as economic, environmental, and social factors. This can be done by integrating decision-support systems that allow stakeholders to weigh different criteria and assess trade-offs in model predictions. By addressing these areas, future research can enhance the robustness, flexibility, and practical utility of regularization models in reservoir management.

4.8 Dimensionality Reduction

Dimensionality reduction is a critical process in data preprocessing, especially in environments dealing with large datasets, such as those common in reservoir management and water quality monitoring. Techniques like Principal Component Analysis (PCA), Factor Analysis (FA), and Discriminant Analysis (DA) are used to simplify datasets by reducing the number of variables under consideration. This not only helps in enhancing computational efficiency but also improves model performance by focusing on the most relevant features. These methods are particularly valuable in environmental science and hydrology, where high-dimensional data can include a wide range of physical, chemical, and biological parameters.

In Table 8, the application of various dimensionality reduction techniques in the analysis of water quality and reservoir operations is explored. The table presents case studies detailing how these methods have been effectively utilized to manage the complexity of environmental data, allowing for more efficient processing and clearer insights into water management challenges. Each case study illustrates the method used, the type of data processed, and the significant outcomes of applying dimensionality reduction, thereby demonstrating their utility in making environmental data more manageable and interpretable.

The studies on dimensionality reduction models exhibit several limitations. The effectiveness of methods like PCA, FA, and MDA is highly dependent on the quality and completeness of input data, which can vary and may limit the models' performance. These techniques can oversimplify complex datasets, potentially overlooking subtle but important variations. The computational resources required for implementing these models, especially when dealing with large and multi-dimensional data, can be significant. Additionally, while dimensionality reduction can optimize monitoring networks and reduce data collection costs, it might not capture all relevant variations, leading to potential information loss. The applicability of these models can also vary across different hydrological and geographical contexts.

Several research gaps are evident in the studies on dimensionality reduction models. There is a need for improved methods to handle data sparsity and variability, ensuring robustness across diverse scenarios. Further exploration into integrating real-time data streams and adaptive learning mechanisms could enhance the models' responsiveness to dynamic changes in environmental conditions. Developing hybrid models that combine dimensionality reduction techniques with other ML methods could improve predictive accuracy and robustness. Comprehensive validation across different geographical locations and climatic conditions is essential to ensure generalizability. Additionally, incorporating socio-economic factors and climate change impacts into these models could provide more holistic and sustainable water management solutions. Enhancing the interpretability and usability of these models for practical application in monitoring and decision-making is also crucial.

To address the challenges associated with dimensionality reduction models like PCA, FA, and MDA, future research should explore several innovative strategies. First, developing context-aware preprocessing techniques tailored to specific environmental conditions can significantly enhance model performance. Additionally, employing advanced methods such as manifold learning and nonlinear dimensionality reduction can capture complex patterns without oversimplification. Reducing computational overhead through hardware acceleration and efficient algorithmic optimizations will make these techniques more practical for large-scale applications. Furthermore, integrating multi-source data fusion from various sensors and databases can provide a more comprehensive dataset, thereby improving robustness. Exploring the synergy between dimensionality reduction and unsupervised learning methods, such as clustering or anomaly detection, can offer new insights and enhance predictive capabilities. Real-time adaptation through the incorporation of streaming data and online learning algorithms will ensure models remain responsive to new information. Extensive field validation, facilitated through international collaboration, will be crucial for testing these models across diverse hydrological and geographical contexts. Moreover, incorporating socio-economic indicators and climate projections will enable the development of more holistic and future-proof water management solutions. Finally, developing intuitive visualization tools and interactive platforms will make these models more accessible to decision-makers, allowing for better integration into existing monitoring and management frameworks. By focusing on these approaches, future research can overcome current limitations and significantly improve the utility of dimensionality reduction models in environmental management.

Table 8: *Applications of Dimensionality Reduction Techniques in Water Quality and Reservoir Management.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[157]	Jezioro Kowalskie Reservoir	Principal Component Analysis (PCA), Factor Analysis (FA), CA, DA	Water quality parameters (dissolved oxygen, BOD, COD, pH, conductivity, hardness)	PCA and FA effectively reduced the dimensionality of the dataset, retaining key variables that explained over 70% of the variance in water quality parameters.	The study demonstrates that multivariate statistical methods can optimize monitoring networks by reducing the number of measurement control points without significant loss of information. The method's effectiveness is limited by the initial quality and distribution of the input data.
[158]	Firuzkuh County, Iran	Mixture Discriminant Analysis (MDA), RF, SVM, MARS, EN	Sediment control factors, topographic, hydrological, geological, and anthropogenic data	MDA effectively classified and identified significant factors for sediment control, achieving a high classification accuracy of 93.7%. Although MDA performed well, the Elastic Net model, despite being a regularization algorithm, had superior performance with an accuracy of 96.6%, making it the most accurate in predicting suitable sites for check dams.	The study demonstrates the efficacy of MDA in handling complex, multi-dimensional data for environmental modeling, but highlights that Elastic Net's ability to regularize and select features contributed to its superior performance. However, both methods require high-quality input data and computational resources.
[159]	River Tisza, Hungary	Combined Cluster and Discriminant Analysis (CCDA), PCA, CA	Water quality parameters (e.g., dissolved oxygen, BOD, pH, nutrients) from 14 sampling sites over 30 years	CCDA effectively reduced the dimensionality of the dataset by identifying homogeneous groups, leading to the potential reduction of sampling locations from 14 to 11, improving monitoring efficiency while retaining 70% of the variance in water quality parameters.	The study highlights the capability of CCDA in optimizing monitoring networks by reducing redundancy and cost, but its effectiveness is dependent on the quality and consistency of the input data.
[160]	Kralkızı, Dicle, and Batman Dam Reservoirs, Turkey	PCA, Factor Analysis (FA), Cluster Analysis (CA), Discriminant Analysis (DA)	Water quality parameters (e.g., dissolved oxygen, BOD, COD, pH, conductivity, nutrients) collected over a year	PCA and FA reduced the data set's dimensionality, identifying five key factors that explained 80% of the total variance. DA was effective for data reduction, identifying nine parameters for temporal variations and eight for spatial variations, allowing for a reduction in sampling efforts while maintaining data integrity.	While the multivariate techniques effectively simplified the dataset, PCA/FA did not significantly reduce the number of measured parameters, indicating a need for continued comprehensive monitoring to capture all relevant variations in water quality.
[161]	La Grande River, Tono Dam (Japan), Various Reservoirs, Multireservoir Systems	Principal Component Analysis (PCA), Non-negative Principal Component Analysis (NPCA)	Multireservoir power system trajectories, Stochastic long-term multireservoir operation, Reservoir storage trajectories	PCA and NPCA reduced the original problem complexity by focusing on major components, making it solvable by dynamic programming. NPCA-based approach provided a better representation of the Pareto front compared to selecting a subset of objectives. PCA significantly reduced the number of state variables, making dynamic programming feasible for large-scale problems. PCA reduced dimensionality by focusing on components explaining the majority of variance, ensuring efficient optimization.	The reduced model accounted for 96.6% of the sample variance; PCA explained a high percentage of variance while maintaining interpretability. NPCA demonstrated superior solution diversity and consistency with improved RMSE performance. PCA accounted for 85% of variance with first components, greatly reducing computational complexity. PCA effectively reduced the problem size while maintaining a high level of variance explanation, improving model efficiency.

[162]	Tono Dam, Japan	Non-negative Principal Component Analysis (NPCA)	Daily operation data of a multi-purpose water reservoir	The NPCA aggregates the original multiple objectives into a reduced number of principal components. This transformation aids in optimizing the decision-making process with lower dimensional objectives.	The NPCA-based approach provides a better representation of the Pareto front, ensuring consistency and solution diversity. It effectively reduces computational complexity.
[163]	South Yellow Sea Basin	Principal Component Analysis (PCA), Least Squares Support Vector Machine (LSSVM)	Well log data including gamma ray (GR), sonic travel time (DT), spontaneous potential (SP), resistivity (RT), neutron porosity (NPHI)	PCA improved the performance of LSSVM and ANFIS-SCM by reducing model complexity and preventing overfitting. PCA-LSSVM outperformed PCA-ANFIS-SCM in predicting water saturation with higher accuracy and generalization ability.	PCA-LSSVM achieved superior performance metrics with an RMSE of 0.022 during training and 0.059 during testing. The use of PCA reduced dimensionality while maintaining 80.3% of the variance, enhancing the overall efficiency and reliability of the models.
[164]	Reservoir management in Quebec, Canada (La Grande River)	Principal Component Analysis (PCA)	Multireservoir storage trajectories, inflows	PCA effectively reduces the dimensionality of the state variables in the stochastic long-term multireservoir operating problem. The PCA-based approach reduces the original 10-state variables to 4 principal components, accounting for 96.6% of the sample variance.	The PCA approach demonstrated significant improvement in computational efficiency while maintaining high accuracy in reservoir management. The eigenvalues and corresponding variance contributions show that the majority of the variance is captured by a few principal components. RMSE values indicate high predictive accuracy.

4.9 Clustering

Clustering algorithms are a fundamental aspect of unsupervised learning used to group a set of objects in such a way that objects in the same group (or cluster) are more similar to each other than to those in other groups. This technique is invaluable in various fields such as ML, pattern recognition, image analysis, and bioinformatics, where it helps to identify natural groupings among data without prior knowledge of the group definitions. Clustering is particularly useful in exploratory data analysis, allowing researchers and analysts to discover underlying patterns, group subjects with similar behaviors, and segment data into distinct parts for further study or targeted action.

In Table 9, a variety of clustering methods used across different studies is highlighted. These methods include K-means, Hierarchical Clustering, and more sophisticated two-step approaches that enhance data analysis capabilities in different contexts. The table showcases how these clustering techniques are applied to hydrological data, environmental management, and infrastructure monitoring, demonstrating their effectiveness in extracting meaningful information from complex datasets.

The studies focusing on clustering models, such as K-means, hierarchical clustering, and two-step clustering, reveal several limitations. Firstly, these models are highly dependent on the quality and completeness of the input data, which can affect the accuracy of the clustering results. The selection of initial centroids and the number of clusters significantly influences the performance of methods like K-means, making them sensitive to initialization parameters. Furthermore, the computational intensity required for clustering large datasets can be substantial, limiting real-time application and scalability. Another limitation is the complexity of interpreting clustering outcomes, particularly for non-expert users, which may hinder practical implementation. Additionally, clustering techniques may not capture the temporal dynamics of hydrological data effectively, necessitating the integration with time-series models for comprehensive analysis.

Several research gaps are evident in the studies focused on clustering models. There is a need for improved methods to handle data variability and sparsity, ensuring the robustness of clustering results across diverse environmental scenarios. Future research should explore hybrid models that combine clustering with other ML techniques to enhance predictive accuracy and robustness. Developing more efficient algorithms that can handle large datasets and provide real-time clustering is crucial. Comprehensive validation of these models across different geographical locations and climatic conditions is necessary to ensure their generalizability. Incorporating socio-economic factors and climate change impacts into clustering models could provide more holistic and sustainable water management solutions. Enhancing the interpretability and usability of clustering results for practical decision-making is also important. Additionally, focusing on integrating clustering methods with time-series analysis can improve the models' ability to capture temporal dynamics in hydrological data. Addressing these gaps will pave the way for more effective and adaptable clustering models in reservoir and dam management.

Table 9: *Utilization of Clustering Techniques in Data Analysis.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[165]	Brazilian Hydroelectric System	K-means Clustering	Monthly streamflow data, reservoir inflows	The application of K-means clustering in the monthly streamflow generation model effectively reduced computational effort in the mid-term operation planning model. Clustering techniques enabled the selection of representative inflow sequences from a large set, preserving the statistical properties of the original data.	The K-means clustering approach proved efficient in handling large datasets and reducing computational load, making it suitable for complex hydroelectric systems. However, the quality of clustering results depends on the accurate selection of initial centroids and the number of clusters.
[166]	River Brahmaputra Basin, India	K-means Clustering, Agglomerative Hierarchical Clustering (AHC),ARIMA, SARIMA	Rainfall data, runoff data, river discharge data	K-means and AHC clustering techniques effectively identified patterns in the hydrological data, leading to improved runoff forecasting models. The ARIMA model achieved an average R^2 value of 0.92 across eight models, indicating high accuracy in runoff prediction.	Clustering techniques enhanced the understanding of hydrological patterns, improving the predictive capabilities of regression models. Accurate clustering depends on proper selection of initial parameters and sufficient data quality. The combined approach of clustering and regression analysis is highly useful for water resources planning and development.
[167]	Narmada River Basin, India	Hierarchical Clustering (HC), Thiessen Polygon Method	Rainfall data, discharge data, Digital Elevation Model (DEM) data	Hierarchical Clustering identified key rain gauge stations, optimizing the rain gauge network by grouping stations with similar rainfall patterns. This resulted in improved flood forecasting accuracy with an NSE value of 0.963 and an R^2 of 0.9636 for 2012.	The HC approach effectively identified significant rain gauge stations, enhancing the efficiency of flood forecasting models. However, the method requires high-quality input data and computational resources, and its performance may be affected by missing or incomplete data.
[168]	Eyvashan Earth Dam, Iran	K-means Clustering, Two-Step Clustering	Settlement data from precision instruments, water levels, embankment levels	K-means and two-step clustering effectively identified settlement patterns, allowing for improved monitoring and prediction of settlement behavior. The maximum settlement of 809 mm occurred at the mid-level of the dam core, which was 1.2% of the dam height.	Clustering methods enhanced the interpretability of settlement data, helping to identify critical areas. However, the effectiveness of clustering is limited by the quality and resolution of input data, and the method requires long-term monitoring data for accuracy.
[169]	Eyvashan Earth Dam, Iran	K-means Clustering, Two-Step Clustering, RF, MARS, GMDH	Settlement data from precision instruments, embankment levels, water levels	The study showed that spatiotemporal clustering, combined with AI models, significantly improves the prediction of dam settlement behavior. K-means clustering identified critical settlement patterns, with a maximum observed settlement of 809 mm (1.2% of the dam height) at the middle level of the dam core. GMDH demonstrated superior predictive performance with RMSE of 1.6947 and DC of 0.9837.	Clustering methods enhanced the interpretability and accuracy of settlement predictions. The integration with AI models provides robust results for monitoring dam stability. However, the approach requires high-quality input data and computational resources.

[170]	Xinyihe Reservoir, China	K-means Clustering	Rainfall, inflow, water levels, sediment concentration	K-means clustering was used to identify patterns in sediment concentration data. This method effectively grouped similar data points, enhancing the understanding and management of sediment dynamics in the reservoir.	The application of K-means clustering significantly improved the model's ability to predict sediment concentrations. The RMSE values indicate higher accuracy and reliability in sediment concentration forecasts.
[171]	Streamflow Forecasting in Semi-arid Region, Tigris River, Iraq	K-means Clustering, Support Vector Regression (SVR), Generalized Regression Neural Network (GRNN)	Streamflow data, rainfall, temperature	K-means clustering was used to classify the input data into different clusters before applying SVR and GRNN for forecasting. This approach improved the performance of the models by identifying and grouping similar patterns in the data.	The integration of K-means clustering with SVR and GRNN resulted in a significant reduction in RMSE values, demonstrating the effectiveness of clustering in improving model accuracy. The ELM model outperformed the SVR and GRNN models, with RMSE reductions of 21.3% and 44.7%, respectively.

To address the limitations identified in studies on clustering models such as K-means, hierarchical clustering, and two-step clustering, future research should consider several new approaches. Developing algorithms that incorporate adaptive clustering techniques can help models dynamically adjust the number of clusters and initial centroids based on the data, improving accuracy and reducing sensitivity to initialization parameters. Enhancing data quality through the use of sophisticated preprocessing methods like noise filtering and missing data imputation can ensure more reliable clustering results. Leveraging advanced ML techniques, such as ensemble learning, can help mitigate the computational intensity required for large datasets, enabling more efficient and scalable clustering solutions. Implementing real-time clustering frameworks, perhaps through the use of streaming data architectures, will allow for continuous updating and analysis, crucial for dynamic environmental conditions. Additionally, integrating clustering models with spatial analysis tools can provide deeper insights into geographical patterns and trends, making the outcomes more actionable. Developing interactive visualization platforms that present clustering results in an intuitive manner can make these models more accessible to non-experts, facilitating practical implementation. Incorporating temporal data analysis within clustering models will help capture the dynamics of hydrological data more effectively, providing a comprehensive understanding of temporal changes. Finally, including socio-economic and climate change factors within clustering frameworks can lead to more sustainable and holistic water management strategies. By focusing on these innovative strategies, future research can significantly enhance the robustness, applicability, and practical utility of clustering models in reservoir and dam management.

4.10 Regression

Regression techniques form a fundamental part of predictive modeling, focusing on estimating the relationships among variables. These methods are extensively utilized in various scientific, engineering, and economic fields to forecast future events based on historical data. The objective of regression is to model the target variable as a function of one or more independent variables, providing a quantitative assessment of the relationships among those variables. This modeling enables decision-makers to understand how changes in predictor variables affect the outcome variable and to make predictions or forecasts based on statistical analysis.

In Table 10, a diverse array of regression models used across different case studies is detailed. This compilation includes both linear and nonlinear regression methods, highlighting their application in predicting outcomes in natural resource management, hydrological forecasting, and environmental science. The table serves as an extensive resource for understanding the versatility and effectiveness of regression techniques in handling complex and multidimensional data.

The studies on regression models, including Multivariate Adaptive Regression Splines (MARS), Multiple Linear Regression (MLR), and Support Vector Regression (SVR), exhibit several limitations. One key limitation is the dependency on high-quality, comprehensive input data, which is essential for accurate predictions but often challenging to obtain. These models can struggle with nonlinear relationships and complex interactions in the data, necessitating the integration of hybrid approaches for improved performance. Additionally, the computational intensity required for training and validating these models can be significant, limiting their scalability and real-time application. Another limitation is the models' sensitivity to input variability, which can lead to reduced accuracy under varying environmental conditions. Furthermore, while regression models can provide robust long-term forecasts, they may not capture short-term fluctuations effectively.

Table 10: *Applications of Regression Techniques in Predictive Modeling.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[172]	Çoruh River Basin, Turkey	Multivariate Adaptive Regression Splines (MARS), TLBO, ABC, Classical Regression Analysis (CRA)	Streamflow values and suspended sediment load (SSL) data from two gauging stations	MARS was the most accurate model for predicting SSL, achieving RMSE values between 35% and 39% for test datasets, and even lower errors (7% to 15%) for another dataset. TLBO and ABC also performed well but were outperformed by MARS.	The study demonstrates the effectiveness of heuristic and meta-heuristic models for SSL prediction, with MARS showing superior performance. The approach requires comprehensive data for training and validation to ensure accuracy.
[173]	Detention Dams in Iran	Multivariate Adaptive Regression Splines (MARS), Gene Expression Programming (GEP), Group Method of Data Handling (GMDH)	Hydraulic parameters, sediment characteristics, sediment trap efficiency data	MARS model showed superior accuracy in predicting trap efficiency (TE) of detention dams with an R^2 value of 0.95 and RMSE of 5.79, outperforming other models like GEP and GMDH. The most effective parameters identified were the ratio of flood volume to sediment volume (VF/VS), mean diameter of sediment size (D50), and specific gravity (Gs).	The MARS model effectively maps complex relationships between hydraulic parameters and TE, providing a robust tool for dam performance analysis. However, the accuracy of the model depends on the quality of input data and comprehensive parameter tuning.
[174]	Pailugou catchment, China	Multivariate Adaptive Regression Splines (MARS), Support Vector Regression (SVR), M5 Model Tree (M5Tree)	River flow data, precipitation data, temperature data	The MARS model provided accurate river flow forecasts with R values above 0.90 and NS values above 0.80 for 1-, 2-, and 3-day lead times. The M5Tree model outperformed both MARS and SVR, achieving the highest NS value of 0.917 for 1-day ahead forecasts.	The study shows the effectiveness of MARS in handling complex nonlinear relationships in hydrological data. However, the M5Tree model demonstrated superior overall performance, indicating the potential for hybrid modeling approaches.
[175]	Lower Murrumbidgee River, Australia	Generalized Additive Models (GAM), Locally Estimated Scatterplot Smoothing (LOESS)	River discharge data, rainfall data, SPI, SFI indices	River regulation and water diversion significantly impact hydrological drought characteristics. The GAM with LOESS terms revealed that upstream regulation reduced drought severity at Wagga Wagga but increased it downstream at Balranald. The SFI model showed that regulation mitigated droughts in the upstream region while exacerbating them downstream due to increased water diversion.	The study highlights the need for balanced water management to address the conflicting needs of upstream and downstream users. The integration of SPI and SFI indices provided a robust framework for analyzing the impacts of regulation on drought characteristics. The complexity of the models requires high-quality input data and computational resources.

[176]	Besiri, Malabadi (Turkey), Hit, Baghdad (Iraq)	Multivariate Adaptive Regression Splines (MARS), Multiple Linear Regression (MLR), Least Square Support Vector Regression (LSSVR), M5 Model Tree (M5Tree)	Monthly streamflow data, periodicity component	The LSSVR model outperformed MARS, M5Tree, and MLR in long-term streamflow forecasting, achieving the highest accuracy with RMSE reduction of 8.95-4.19% for Besiri, 12.8-8.08% for Malabadi, -0.12-4.03% for Hit, and 13.56-10.03% for Baghdad. Including periodicity as an input significantly improved model performance, with R values for LSSVR reaching 0.905 for Besiri, 0.894 for Malabadi, 0.810 for Hit, and 0.879 for Baghdad.	The periodicity component significantly enhances the forecasting accuracy of the models. While MARS and MLR showed good performance, LSSVR consistently provided the best results, demonstrating its robustness in handling complex, nonlinear relationships in hydrological data. The study highlights the importance of incorporating periodic data for improved prediction accuracy.
[177]	480 Catchments in Germany and Austria	Multiple Linear Regression (MLR), Random Forest (RF)	Streamflow data, precipitation data, catchment characteristics	Catchment response variables are the most significant controls of heavy tail behavior in flood distributions. Specifically, a high runoff coefficient and short event time scales are associated with heavier tails. MLR models achieved R^2 values up to 0.607, and RF models provided additional insights into nonlinear relationships.	Understanding catchment-specific responses is crucial for improving flood risk management. The models require comprehensive datasets and high computational resources, highlighting the need for targeted flood mitigation strategies in different catchment areas.
[171]	Tigris River Stream-flow Forecasting	Support Vector Regression (SVR), Generalized Regression Neural Network (GRNN), Extreme Learning Machine (ELM)	Stream-flow data from June 1991 to December 2010	ELM showed a significant improvement over SVR and GRNN in forecasting stream-flow. The RMSE and MAE were reduced by about 17.44–29.78% and 21.3–30.92% respectively when using ELM compared to SVR and GRNN.	The RMSE reduction by ELM indicates a better fit for the observed data, enhancing the prediction accuracy for water resources management decisions.
[178]	Monthly Streamflow Forecasting, Karkheh Dam, Iran	Least-Square Support Vector Regression (LS-SVR), Artificial Neural Network (ANN), Generalized Regression Neural Network (GRNN), K-Nearest Neighbor Regression (KNN)	Monthly inflow, rainfall, snow area extent	LS-SVR performed best in nonlinear conditions, while ANN performed best in linear conditions. The comparative assessment showed that LS-SVR is superior in capturing the nonlinear relationships between predictors and predicted variables.	LS-SVR showed the lowest RMSE values in nonlinear conditions, indicating higher prediction accuracy. ANN performed well in linear conditions but was outperformed by LS-SVR and KNN in nonlinear conditions. The study highlights the importance of selecting appropriate models based on the nature of the data.

Several research gaps are identified in the studies focused on regression models. There is a need for enhanced methods to handle data sparsity and variability, ensuring the robustness of predictions across diverse hydrological scenarios. Exploring the integration of real-time data streams and adaptive learning mechanisms could improve model responsiveness to dynamic changes. Developing more sophisticated hybrid models that combine regression techniques with other ML methods could enhance predictive accuracy and robustness. Comprehensive validation across different geographical locations and climatic conditions is necessary to ensure generalizability. Additionally, incorporating socio-economic factors and climate change impacts into these models could provide more holistic and sustainable water management solutions. Improving the interpretability and usability of regression models for practical application in decision-making is also crucial, as is the need for more efficient computational techniques to facilitate real-time applications. Finally, focusing on capturing both long-term trends and short-term fluctuations in water resource data will be critical for advancing the effectiveness of these models in various water management contexts.

For regression models such as Multivariate Adaptive Regression Splines (MARS), Multiple Linear Regression (MLR), and Support Vector Regression (SVR), future research should explore several new strategies. Implementing transfer learning techniques can help models leverage pre-trained knowledge from related domains, improving performance even with limited high-quality data. Developing models that incorporate multi-fidelity simulations can enhance accuracy by combining high-precision data with lower-quality datasets. Utilizing advanced optimization algorithms, such as genetic algorithms or particle swarm optimization, can improve the efficiency of training processes, making real-time applications more feasible. Additionally, integrating spatial-temporal analysis methods can help models capture both spatial variability and temporal dynamics in water resource data, providing more comprehensive insights. Employing robust statistical techniques to adjust for input variability can enhance model stability and accuracy under diverse environmental conditions. Creating modular frameworks that allow for easy integration of socio-economic and climate change factors can lead to more sustainable water management solutions. Enhancing model transparency through explainable AI techniques will make regression models more interpretable for practitioners, aiding in practical decision-making. Finally, establishing standardized benchmarks and validation protocols across different geographical locations and climatic conditions will ensure the generalizability and reliability of these models. By focusing on these innovative approaches, future research can significantly improve the robustness, scalability, and practical utility of regression models in reservoir and dam management.

4.11 Instance Based Learning

Instance-based learning is a category of learning algorithms that base their prediction on instances or examples from the training dataset rather than attempting to construct a general internal model. These algorithms, such as k-Nearest Neighbors (k-NN), rely on the similarity between new problem instances and instances seen during training, using these similarities to predict the output. This approach is particularly effective for tasks where the decision boundary is irregular and can adapt quickly to changes in input data without the need for retraining.

In Table 11, the use of instance-based learning methods in reservoir management is explored. The table showcases several case studies that demonstrate how these algorithms are applied to predict hydrological outcomes (Out flow and Reservoir's storage levels) based on historical data (currently and previously recorded in-flow data, including rainfall intensities). Each case study details the optimizer used, the type of data analyzed, and the results achieved, emphasizing the practicality and direct applicability of instance-based learning in managing water resources.

Table 11: *Applications of Instance-Based Learning in Reservoir Management.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[179]	Zayandeh-rud Dam, Iran	K-Nearest Neighbour (K-NN), Honey-Bee Mating Optimization (HBMO)	Hydrological data, accumulated inflow data from 1971 to 2001	The optimized K-NN method, enhanced with various techniques including HBMO, showed significant improvement in long-lead hydrological forecasting. The accuracy increased with a correlation coefficient of 96%, a reduction in volume error to 8%, and a decrease in RMSE to 25%.	The use of K-NN combined with HBMO and other optimization techniques effectively improved prediction accuracy. The study highlighted the importance of data preprocessing, optimal selection of neighbors, and distance functions for enhancing K-NN performance.
[180]	Karun-1 Dam, Iran	Fuzzy K-Nearest Neighbor (Fuzzy KNN)	Cavitation damage data, flow characteristics (pressure, velocity)	The Fuzzy KNN model effectively predicted the location and intensity of cavitation damage on the Karun-1 Dam spillway. The model showed high predictive accuracy with Pearson's correlation coefficient of 0.873, MAE of 0.167, NRMSE of 0.110, and CRM of 0.007.	The Fuzzy KNN approach improved prediction accuracy for cavitation damage, providing reliable insights for spillway design and maintenance. The model requires high-quality input data and proper normalization to avoid bias.
[181]	Guadalmellato River Reservoir, Spain	Self-Organizing Map (SOM)	Rainfall, streamflow, reservoir level, water releases	The SOM-based decision support system effectively models and visualizes the complex relationships between key variables involved in reservoir operation. It provides operation profiles that help in identifying and characterizing different conditions and policies, improving decision-making for reservoir management.	The SOM approach allows for an intuitive and helpful interpretation of data, making it a valuable tool for reservoir managers. The method relies on historical data, making it crucial to have accurate and comprehensive records for effective training and analysis.
[98]	10-reservoir system, Columbia River, USA	Karhunen-Loeve (KL) Expansion, Non-dominated Sorting Genetic Algorithm II (NSGA-II)	Reservoir inflows, outflows, storage levels, power generation, fish flow requirements	The spectral optimization model (SOM) significantly reduced the dimensionality of decision variables, achieving better optimization performance with fewer decision variables. For 140 decision variables, optimal performance was obtained with 6 KL terms, and for 3360 decision variables, with 11 KL terms.	The SOM effectively transforms decision variables from the time domain to the frequency domain, reducing the search space and improving computational efficiency. This approach avoids local optima and ensures better convergence and diversity in optimization results. However, it requires prior computation to construct the KL expansion and high-quality input data.
[182]	Bukan Reservoir, Lake Urmia Basin, Iran	K-Nearest Neighbor (KNN), ANFIS	Inflow data, reservoir storage levels, water releases	ANFIS outperforms KNN with a 25% reduction in RMSE, 23% reduction in PWRMSE, 27% improvement in NSCE, and 10% higher correlation coefficient. The ANFIS-based adaptive reservoir operation model achieved an objective function value only 5% better than the KNN-based model, highlighting the efficiency of the adaptation and updating procedure in reducing forecast errors.	The study demonstrates the superiority of ANFIS over KNN for streamflow forecasting in terms of prediction accuracy. However, the marginal improvement in the reservoir operation model's performance suggests that the adaptive updating mechanism effectively mitigates forecast errors, making both methods viable for operational use.

[183]	Monthly Streamflow Forecasting, Three Gorges Region, Yangtze River	SVM with Adaptive Insensitive Factor, Wavelet Denoising, PSO	Monthly streamflow data, rainfall, temperature	The improved SVM model with adaptive insensitive factor and wavelet denoising significantly improves forecasting accuracy. The SVM with PSO optimization demonstrates better generalization ability and higher prediction accuracy compared to conventional SVM and ANN models.	The use of adaptive insensitive factors and PSO in SVM reduces the RMSE by 17.44-29.78%, indicating a substantial improvement in prediction performance. This model is suitable for complex hydrological data series.
[184]	Clair Engle Lake and Bhadra Reservoir	K-Nearest Neighbors (KNN)	Hydrological and climate phenomenon data	KNN showed the least desirable performance for both Clair Engle Lake and Bhadra reservoirs. Despite various parameter settings, KNN was less effective in predicting daily inflows compared to other models such as LSTM and Gradient Boosting Regressor. The model struggled with capturing the complex, non-linear relationships present in the hydrological and climate data.	RMSE values varied significantly, with Clair Engle Lake showing RMSE: 43.56 - 66.27 m^3/s, and Bhadra Reservoir RMSE: 65.18 - 66.27 m^3/s, indicating a higher error margin. KNN's performance was consistent in being the least effective across different seasons and parameter settings, suggesting its limitations in handling the intricacies of reservoir inflow forecasting.
[185]	Shihmen Reservoir, Northern Taiwan	Self-Organizing Map (SOM), Support Vector Machine (SVM), Autoregressive Method (AR)	Outflow sediment concentration, rainfall, inflow and outflow data	The combination of SOM, SVM, and AR (SOSVMAR) significantly improves the accuracy of outflow sediment concentration forecasts. The SOM algorithm effectively clusters the data to identify patterns and salient features, leading to improved forecasting performance by SVM. AR further corrects the forecasted data in real-time.	The SOM algorithm helps in reducing the model's error by 40% at the power plant intake and 35% at the bottom outlet for long lead time forecasting (3 hours). The integration of these algorithms provides robust and reliable forecasts, crucial for reservoir sediment management during desilting operations. RMSE values indicate significant improvements in prediction accuracy.

The analysis of instance-based models reveals several significant limitations. Firstly, the performance of models like K-Nearest Neighbor (KNN) and Self-Organizing Maps (SOM) is highly dependent on the quality and pre-processing of input data, which can vary and influence outcomes. For example, in the Zayandeh-rud Dam study, the K-NN model's accuracy improved significantly with techniques like HBMO and proper data normalization. However, the computational demands, particularly when combined with optimization techniques, can be substantial, posing challenges for real-time application, as seen in the 10-reservoir system study where the KL Expansion required considerable computational effort. Additionally, these models often rely on historical data, which may not always capture sudden or extreme environmental changes. This reliance is evident in studies like the Karun-1 Dam and Shihmen Reservoir, where historical data played a crucial role in model accuracy. The complexity of instance-based models can also present interpretation challenges, making them less accessible to users without specialized knowledge. For instance, the spectral optimization model in the Columbia River study required expertise to interpret the dimensionality reduction results. Furthermore, the need for proper normalization to avoid bias and ensure accurate predictions is critical, as highlighted in multiple studies.

To address these limitations, future research should focus on developing more efficient data pre-processing and normalization techniques that can automatically adjust to varying data quality and types, ensuring robust model performance. Implementing real-time data integration systems, combined with adaptive learning algorithms, will enhance the models' responsiveness to sudden environmental changes and improve their overall accuracy. For example, incorporating IoT sensors for continuous data collection can provide up-to-date information for models like KNN and ANFIS, as demonstrated in the Bukan Reservoir study. Exploring hybrid models that integrate instance-based approaches with advanced ML techniques, such as deep learning or ensemble methods, can yield better predictive performance and robustness, as shown in the studies involving SVR and PSO. Additionally, utilizing cloud-based platforms and parallel processing can alleviate computational demands, making real-time applications more feasible by distributing the computational load. Comprehensive validation across different geographical regions and climatic conditions, such as the diverse datasets used in the Clair Engle Lake and Bhadra Reservoir studies, will ensure these models are robust and generalizable. Incorporating socio-economic factors and climate change projections into instance-based models will result in more holistic and sustainable management solutions. Enhancing the interpretability of these models through intuitive visualization tools and user-friendly interfaces will make them more accessible to non-experts, facilitating broader adoption in practical decision-making scenarios. By addressing these areas, future research can significantly improve the robustness, scalability, and applicability of instance-based models in various environmental contexts, leading to more effective and informed decision-making in reservoir and dam management.

4.12 Hybrid Machine Learning

Hybrid ML techniques represent a significant advancement in the field of reservoir management and optimization. These methods combine the strengths of multiple ML algorithms to improve predictive accuracy, operational efficiency, and robustness in handling complex systems. By leveraging the unique capabilities of different models, hybrid approaches address various challenges such as non-linearity, high dimensionality, and the need for both global and local optimization.

In Table 12, the role of hybrid ML techniques in optimizing reservoir operations is detailed. The table presents multiple case studies where hybrid models have been successfully applied to enhance the management and prediction of reservoir systems. Each entry illustrates the specific hybrid techniques used, the types of data involved, and the main conclusions drawn from these studies. Furthermore, the table describes how hybridization improved the models' performance by combining different optimization strategies, addressing data integration challenges, and providing more reliable predictions under various scenarios

Table 12: *Impact of Hybrid Machine Learning Techniques on Reservoir Management Models.*

Ref.	Case study	Optimizer	Type of data	Main Conclusion	Remarks
[186]	Mangla Watershed, Pakistan	Extreme Learning Machine (ELM), Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO), PSOGWO, PSOGSA	Monthly precipitation and runoff data	The hybrid ELM-PSOGWO model provided the most accurate monthly runoff predictions, reducing RMSE by 38.2% compared to standalone ELM, and significantly outperforming other hybrid models (ELM-PSO, ELM-GWO, and ELM-PSOGSA). The ELM-PSOGWO achieved the lowest RMSE (55.14 m^3/s) and the highest NSE (0.919) in the test phase.	The hybridization combined PSO and GWO to form PSOGWO, enhancing exploration and exploitation abilities. PSO provided local search strength, while GWO contributed global search capabilities, overcoming individual weaknesses. PSOGWO outperformed PSOGSA and standalone models, proving effective in optimizing ELM parameters for more accurate predictions.
[187]	Kenyir Dam, Terengganu, Malaysia	Bayesian Linear Regression (BLR), Boosted Decision Tree Regression (BDTR), Decision Forest Regression (DFR), Neural Network Regression (NNR)	Daily water level, rainfall, hourly sent-out data	BLR outperformed other models with R^2 of 0.998952 for SC1 and BDTR with R^2 of 0.99992 for SC2. BLR and BDTR provided highly accurate water level forecasts.	Hybridization involved using Bayesian Inference combined with linear regression to handle insufficient and incorrectly distributed data. The combination of BDTR with hyperparameter tuning improved prediction accuracy, addressing the limitations of traditional models. For SC2, adding sent-out data further refined predictions.
[188]	Klang Gate Dam	Sim-Heuristic, Harris Hawks Optimization	Precipitation, Temperature, Solar Radiation, Evaporation, Inflow	The sim-heuristic approach effectively optimized the reservoir operations under various climate scenarios (RCP 2.6, RCP 4.5, RCP 8.5). Scenario 3 showed the greatest reliability in satisfying exact demand with 93.54%, as well as the least shortage index and length of water deficit under RCP 4.5. The study demonstrated that using an ensemble of GCMs provided more reliable results in hydrology management compared to a single GCM, emphasizing the importance of considering multiple climate models for accurate forecasting and planning.	The hybridization involved using a combination of Artificial Neural Networks (ANN) and Support Vector Regression (SVR) for downscaling climate data. The ANN and SVR models were used to predict precipitation and temperature variables, achieving high correlation values, with the SVR (Poly Kernel function) showing superior performance. This was followed by the application of the Harris Hawks Optimization (HHO) algorithm to optimize the reservoir operations. The HHO algorithm utilized various strategies to mimic the predatory behavior of Harris hawks, including soft and hard siege tactics, to find the optimal solution for minimizing water deficit while adhering to storage and release constraints.

[189]	Feitsui Reservoir	Artificial Neural Network (ANN) and Genetic Algorithm (GA)	Nutrient loads, precipitation, outflow, water quality data	The combined ANN and GA approach effectively managed water quality in Feitsui Reservoir. The ANN model simulated nutrient load behavior and forecasted total phosphorus (TP) concentration with high accuracy. GA optimized nutrient load control, achieving up to a 60% reduction in TP concentration. The ANN model showed better performance than traditional TP models, capturing seasonal variations effectively. The study concluded that a 10-80% reduction in watershed nutrient loads could maintain the reservoir at oligotrophic or mesotrophic levels.	Hybridization involved integrating ANN and GA for reservoir water quality management. ANN modeled nutrient loads and predicted TP concentrations, leveraging data from watershed loads, precipitation, and outflow. The model was trained with a 6-year dataset (1992-1997) and validated with data from 1998-2000, achieving good correlation with observed values. GA was used as a search strategy to optimize nutrient load reduction rates from the watershed, minimizing the objective function. The hybrid approach allowed the simulation of different scenarios, showing that control schemes could significantly improve water quality.
[190]	Hydropower Multi-Reservoir Systems, Iran	Influential Flower Pollination Algorithm (IFPA), Adaptive Guided Differential Evolution (AGDE), Composite Differential Evolution (CODE), HSLSO, ITLBO, jDE, SATLBO, MS-DEPSO	Water inflow, Reservoir storage, Hydropower generation, System efficiency	The IFPA outperformed other optimization methods in terms of convergence speed and precision in achieving the global solution for multi-reservoir systems. It showed significant improvements in total power production with reduced power deficits. For instance, the IFPA achieved an optimal objective function value of 308.83 for the four-reservoir system, and 1196.92 for the ten-reservoir system, surpassing other advanced methods. Error analysis indicated that IFPA had much lower errors compared to others, ensuring powerful performance in terms of accuracy and stability.	The hybridization involved the integration of the Flower Pollination Algorithm with enhancements to improve both global and local search mechanisms. These included a rank-based mechanism for better transition between global and local searches, and adaptive coefficients to fine-tune the search process. Additionally, hybridization with DE, PSO, and other algorithms allowed leveraging strengths of various methods, ensuring robust optimization for hydropower reservoir operations.
[191]	NA	Hybrid Evolutionary Algorithms	Water inflow, Reservoir storage, Hydropower generation, System efficiency	The review highlights the effectiveness of hybrid evolutionary algorithms in optimizing reservoir operations. It concludes that these algorithms provide better solutions than single algorithms, particularly in handling multi-objective and multi-reservoir systems. The algorithms can solve complex and multidimensional problems with fast convergence rates and improved accuracy.	Hybrid algorithms combine the strengths of different evolutionary algorithms to improve performance. Examples include combining Genetic Algorithms (GA) with Particle Swarm Optimization (PSO) or Differential Evolution (DE) to leverage the advantages of each. These hybrids enhance both global and local search capabilities, leading to more robust and reliable solutions for reservoir optimization.

[192]	Roadford Water Supply System	Genetic Algorithm - Linear Programming (GA-LP)	Historical inflow records of a multi-reservoir system	The GA-LP hybrid approach was effective in optimizing the operation of the Roadford Water Supply System. It provided comparable results to the RELAX algorithm, ensuring non-zero final storages in the larger reservoirs. The method demonstrated potential for generating operating policies in the form of hedging rules without a priori imposition of their form.	The GA-LP hybrid method identifies cost reduction factors (CRFs) via Genetic Algorithms and operational variables via Linear Programming. CRFs are introduced to discourage reservoir depletion in the initial stages of the planning period. The weights for CRFs are determined through a two-step optimization process using GA, which are then used in LP to update operational decisions based on new inflow forecasts, enhancing computational efficiency.
[193]	Monthly Optimization Scheduling of a Hydropower Reservoir	Chaos Genetic Algorithm (CGA)	Historical monthly inflows of the reservoir over 38 years	CGA integrates GA's global search with COA's local search to overcome premature convergence and slow optimization in traditional GAs. It produced higher average annual energy, lower total spill, and faster execution time compared to Dynamic Programming and standard GA, proving its efficiency and accuracy in optimizing hydropower reservoir operations.	CGA hybridizes GA with COA, starting with chaos optimization to ensure a diverse and high-quality initial population. It employs an annealing chaotic mutation to avoid local optima and maintain diversity. This integration balances exploration and exploitation, enhancing global optimization and proving effective for complex reservoir systems like the Chaishitan hydropower reservoir.
[194]	Nagarjunasagar Multi-Purpose Reservoir, India	Genetic Algorithm - Nonlinear Programming (GA-NLP) Hybrid Approach	Historical inflow records and crop water requirements	The GA-NLP hybrid model effectively optimized the reservoir operation for irrigation and hydropower production. It provided superior benefits compared to standard operating policies, especially in drought conditions. The hybrid model improved reliability, maximizing net benefits from crops and hydropower generation.	The GA-NLP hybrid approach uses GA to perform a global search for initial solutions, which are then refined using NLP for faster convergence to the global optimum. This integration leverages GA's global exploration capabilities and NLP's local refinement efficiency, ensuring optimal water allocation and reservoir management.

The PSOGWO approach, despite enhancing exploitation capabilities, might face limitations in global search strength, affecting the optimization of complex reservoir systems. Similarly, while Bayesian Linear Regression and Boosted Decision Tree Regression improved predictions in the Kenyir Dam study, the integration of hybrid models still encounters challenges with hyperparameter tuning and handling different data types for more accurate predictions. The hybrid sim-heuristic and Harris Hawks Optimization approaches for Klang Gate Dam showed promising results but may struggle with the reliability of hydrology management compared to single Global Climate Models (GCMs), impacting accurate forecasting and planning for various climate scenarios.

In the Feitsui Reservoir study, although ANN and GA hybrid models managed nutrient load control effectively, they faced challenges in simulating nutrient load behavior with high accuracy, especially in capturing seasonal variations. The IFPA approach for Hydropower Multi-Reservoir Systems in Iran, despite outperforming other methods, might have limitations in terms of the convergence speed and precision required for optimizing complex multi-reservoir systems, and challenges in integrating multiple local search methods effectively. Lastly, while the general review on hybrid evolutionary algorithms noted their potential in optimizing reservoir operations, they still struggle with handling multi-objective and multi-reservoir systems efficiently, with trade-offs between convergence rates and accuracy needing further exploration.

There is a need for improved integration and validation of hybrid models for diverse data types and varying temporal and spatial scales in reservoir operations. Current studies show potential in individual optimization techniques, but there is a gap in robustly combining global and local search methods for more precise and stable results. Additionally, there is a limited understanding of how hybrid models can be generalized across different climatic and geographical conditions, which impacts the reliability of predictions and operational strategies.

Moreover, there is a necessity to explore more effective hyperparameter tuning methods and address data integration challenges to enhance the predictive capabilities of hybrid models. The existing literature reveals a gap in comprehensive frameworks that efficiently handle multi-objective and multi-reservoir systems. Future research needs to address these gaps to ensure sustainable and efficient water resource management through advanced and validated hybrid optimization models.

Future research should focus on developing more sophisticated hybrid models that can seamlessly integrate various optimization techniques, addressing both global and local search capabilities. Emphasis should be placed on enhancing hyperparameter tuning methods and validating these models across diverse climatic and geographical conditions to ensure robustness and accuracy. Furthermore, exploring the use of advanced ML methods for real-time data assimilation and prediction can provide more reliable solutions for reservoir operation optimization. Collaborative efforts between hydrologists and data scientists could lead to the development of comprehensive frameworks that handle multi-objective and multi-reservoir systems efficiently, ensuring sustainable and efficient water resource management.

5. Literature Assessment and Evaluation

In studies where ML algorithms are evaluated and compared, different error criterions are adopted, that includes Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE). Among these, RMSE is often the most preferred error criterion in many ML studies due to its sensitivity to larger errors and its ability to provide a clear measure of the accuracy of predictions [195].

Evaluating and comparing the performance of ML algorithms is crucial in ensuring their reliability and accuracy. Different error metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE), provide valuable insights into how well a model performs. Among these, RMSE is often considered the most preferred error criterion in many ML studies due to its distinctive characteristics [196], [197].

Mean Absolute Error (MAE) calculates the average magnitude of errors in a set of predictions, treating all errors equally regardless of their direction. This metric is straightforward and easy to understand, making it a popular initial choice for evaluating models. However, MAE does not give extra weight to larger errors, which can be a limitation when significant deviations are particularly problematic [198], [199].

Root Mean Squared Error (RMSE) stands out due to its sensitivity to larger errors. RMSE is calculated by taking the square root of the average squared differences between predicted and observed values. This squaring process penalizes larger errors more heavily, making RMSE particularly useful in applications where large prediction errors are more consequential. For instance, in fields like environmental modeling and energy consumption forecasting, larger errors can have significant implications, thus making RMSE a preferred metric [196], [198].

Mean Squared Error (MSE), which is the square of RMSE, is another commonly used metric that emphasizes larger errors. MSE's primary advantage is its ease of use in mathematical formulations, but its interpretation is less intuitive because of the squaring effect, which can make RMSE more favorable for reporting results. The penalty for larger errors in both RMSE and MSE can highlight models that may appear adequate under less sensitive metrics but fail to perform well in real-world applications where accuracy is critical [196], [200].

Coefficient of Determination (R^2) measures the proportion of variance in the dependent variable that is predictable from the independent variables. It is widely used in regression analysis to evaluate the goodness of fit. However, R^2 alone can be misleading, especially in non-linear models or when comparing models with different numbers of predictors. It does not provide a direct measure of prediction error, thus necessitating the use of additional metrics like RMSE for a comprehensive evaluation [201], [202].

Mean Absolute Percentage Error (MAPE) expresses prediction accuracy as a percentage, which can be advantageous for interpretability in business and economic applications. However, MAPE can be problematic when dealing with data that include zero values or very small values, as it can lead to infinite or undefined values [203], [204].

The preference for RMSE in ML literature is well-justified. RMSE provides a clear measure of prediction accuracy by heavily penalizing larger errors, which is essential in many real-world applications. In environmental modeling, for instance, accurate predictions of phenomena like air quality and water levels are crucial, and RMSE's sensitivity to larger errors ensures that models providing the most accurate predictions are identified [205]–[207]. This sensitivity helps in refining models that might otherwise be deemed adequate under less sensitive metrics such as MAE or MAPE. In energy consumption forecasting, RMSE is favored because it accurately captures the importance of peak demand predictions. Accurate peak demand predictions are necessary to prevent outages and optimize resource allocation. The penalty on larger errors inherent in RMSE helps highlight models that perform better in predicting these critical values [208].

The theoretical justification for using RMSE lies in its mathematical properties. Derived from the Euclidean distance, RMSE is a natural measure of average error magnitude. Its sensitivity to larger errors aligns well with the goals of many predictive modeling applications, where avoiding significant deviations from actual values is paramount. Additionally, RMSE is dimensionally consistent with the data being measured, which aids in maintaining interpretability and comparability across different studies and datasets [196], [198].

Overall, RMSE's ability to penalize larger errors more heavily than MAE and its natural alignment with the Euclidean distance make it a preferred choice in many ML studies. Its usage across various fields underscores its versatility and importance in ensuring accurate and reliable model performance evaluation. As ML continues to advance, RMSE remains a cornerstone metric for assessing the quality of predictive models, providing a robust measure of accuracy that is essential for effective decision-making in numerous applications.

In the realm of dam operations, several advanced computational models have demonstrated exemplary accuracy and reliability, crucial for effective water management and operational planning. Among these, the Multivariate Adaptive Regression Splines (MARS) stands out for its versatility in hydrological modeling, achieving remarkable Nash-Sutcliffe efficiency scores as high as 0.917 for 1-day forecasts, thereby proving its efficacy in adapting to complex, non-linear relationships in river flow and precipitation data [174]. Similarly, Support Vector Regression (SVR), enhanced with M5 Model Trees (M5Tree), has showcased robust performance, particularly noted for its precision in streamflow predictions, which are essential for reservoir operation and strategic water release decisions [174], [176]. Further, the Stacking Ensemble Mechanism (SEM) integrates multiple predictive models, including Bi-LSTM, CNN, and RF, achieving KGE values of 0.94 during training and 0.89 during testing, thus underscoring its high predictive accuracy and effectiveness in daily reservoir inflow predictions [133]. Lastly, the Fuzzy Inference System (ANFIS) outperforms traditional models like KNN with a 25% reduction in RMSE, offering refined forecasting capabilities that enhance the adaptability and precision in managing reservoir inflow and water storage levels [179], [182]. These models collectively enhance the decision-making process in dam operations, ensuring not only the optimal utilization of water resources but also the safety and sustainability of reservoir systems [98], [103], [133], [174], [176], [181].

The optimization models reviewed in this article were versatile. Some of them were repeated in many different sections and at different uses. It might be a little difficult to recognize which model is meant to be bolded and studied. Due so, a pie chart representing the demonstration of models of interest per section is plotted. The percentage of the used ML models indicates the significance of this review in covering different aspects of the literature and which model currently the authors are interested in. The Pie-charts are shown in Figure 4.

The suite of models employed in dam operations, as evidenced by recent studies, demonstrates a diverse range of capabilities that cater to the multifaceted challenges of hydrological forecasting and water resource management [1], [99], [116], [124]. The Multivariate Adaptive Regression Splines (MARS) is particularly commendable for its adaptability to nonlinear data, making it ideal for complex hydrological contexts where precision is paramount. On the other hand, Support Vector Regression (SVR) supplemented with M5Tree techniques brings a mathematical rigor to handling voluminous data, ensuring that models do not just fit but also generalize well beyond the training datasets. This is crucial in scenarios where long-term forecasting impacts operational decisions [174], [176].

However, the Fuzzy Inference System (ANFIS) stands out for its ability to handle uncertainty and imprecision—common features in environmental data—thereby providing outputs that are not only precise but also practical for real-world applications. Each of these models has its merits and potential drawbacks; the choice of model often depends on the specific requirements of the study, data availability, and the desired precision in outcomes. For instance, while MARS and SVR are superb for their predictive accuracy, RF and SEM offer resilience against data overfitting, and ANFIS offers unmatched handling of fuzzy data, making them indispensable tools in the arsenal of hydrological modeling and dam operation management [116], [182].

In the intricate domain of dam operations, a few computational models have repeatedly proven their worth by surfacing prominently in multiple research studies, highlighting their widespread applicability and trusted performance. The Genetic Algorithm (GA) emerges as a particularly versatile tool, appearing in 20 distinct studies, where it adeptly handles the optimization of complex, nonlinear problems prevalent in water resource management [106], [107], [109]–[112]. Similarly, the Bayesian Networks (BNs), noted in 8, is lauded for its ability to capture and model the intricate relationships within voluminous hydrological datasets. Dynamic Programming also features in 5 studies, prized for its methodical precision in formulating and executing optimal water release strategies under

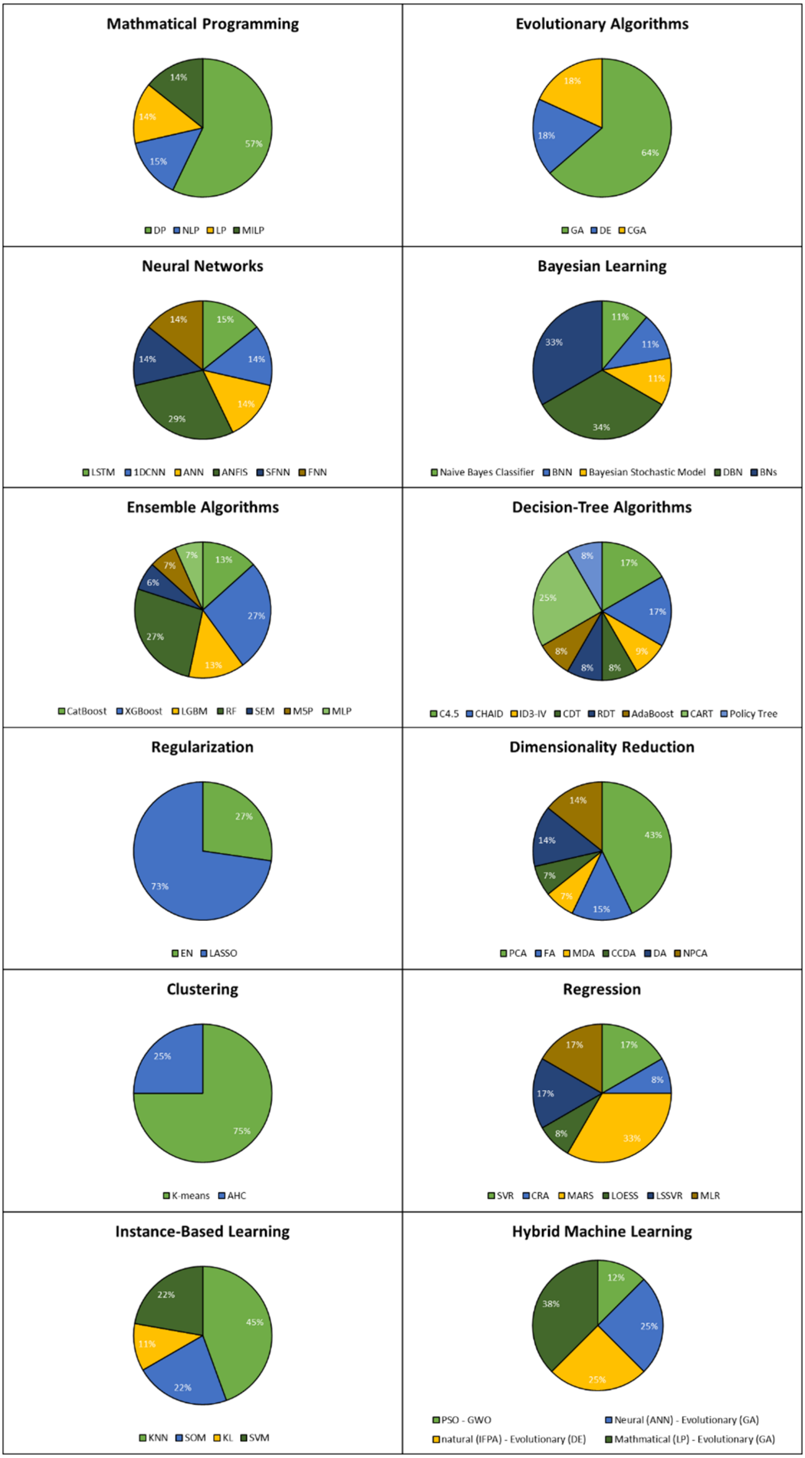

Figure 4: *Optimization models' demonstration percentages according to the reviewed results.*

varied operational scenarios [99]–[101]. These models, by virtue of their repeated application and noted success in addressing diverse challenges, underscore their fundamental role in advancing the efficiency and reliability of dam operations across the globe.

The frequently used optimization techniques and/or ML algorithms as per this review are shown in Figure 5. The number of articles shows how many times these optimizers were used as standalone and/or hybridized techniques. Fuzzy enhanced algorithms are the algorithms that used fuzzy logic to enhance their accuracy whether the learning style was neural network, Regularization or Instance Based learning. Same applies to GA that includes Real-Coded GA or Binary-Coded GA while Boost Algorithms include XGBoost, CatBoost and LGBM. The same applies to Bayesian enhanced algorithms which include DBN, BN, BNN and others.

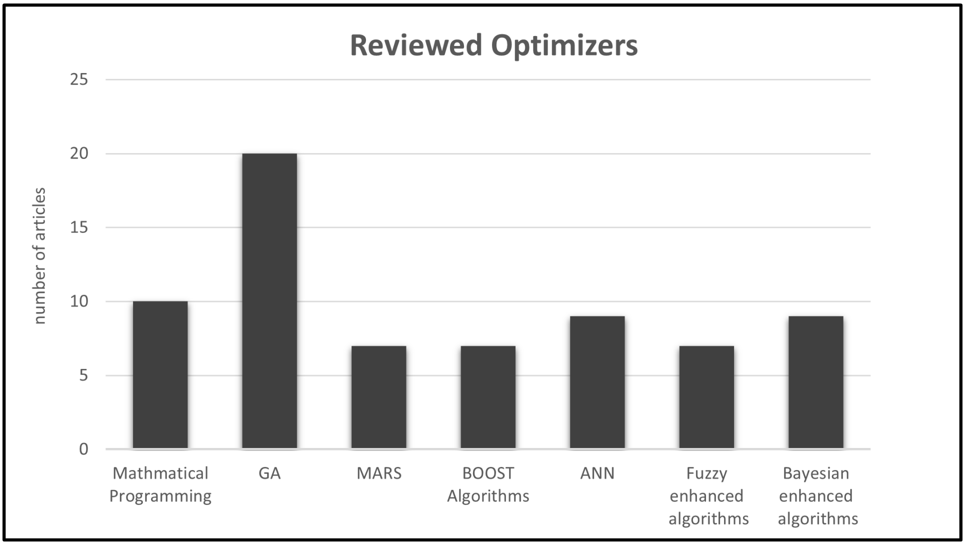

Figure 5: *Mostly used optimization techniques according to the current review.*

Likewise, Figure 6, titled "Reviewed Case Studies global distribution and intensity" plays a critical role in illustrating the geographical diversity and focus of research within the domain of reservoir optimization using ML techniques. This visual representation captures data from various countries, highlighting how research is not only widespread but also varied in intensity across different regions. Notably, Iran emerges as the predominant country in this field of study, with 15 articles dedicated to the topic, which underscores its leadership in implementing advanced technologies for water reservoir management. This is followed by China, which is represented in 8 articles, indicating a significant concentration of research activity.

The map (Figure 6) further serves as a tool for evaluating the current landscape of global research, suggesting areas with high research activities as well as regions that might benefit from increased focus in the future. This distribution insight is vital for scholars and policymakers aiming to understand where knowledge gaps exist and where future studies could be directed to leverage ML for enhancing reservoir management practices. It also suggests a trend where ML techniques are being adopted in diverse contexts, reflecting broader applications in the field.

Despite these successes, significant limitations remain. Ensemble and neural network models, while powerful, often face computational intensity and scalability issues, limiting real-time application. Regularization models, although effective in feature selection, may struggle with data sparsity and noise. Decision-tree algorithms can be prone to overfitting and may require extensive cross-validation

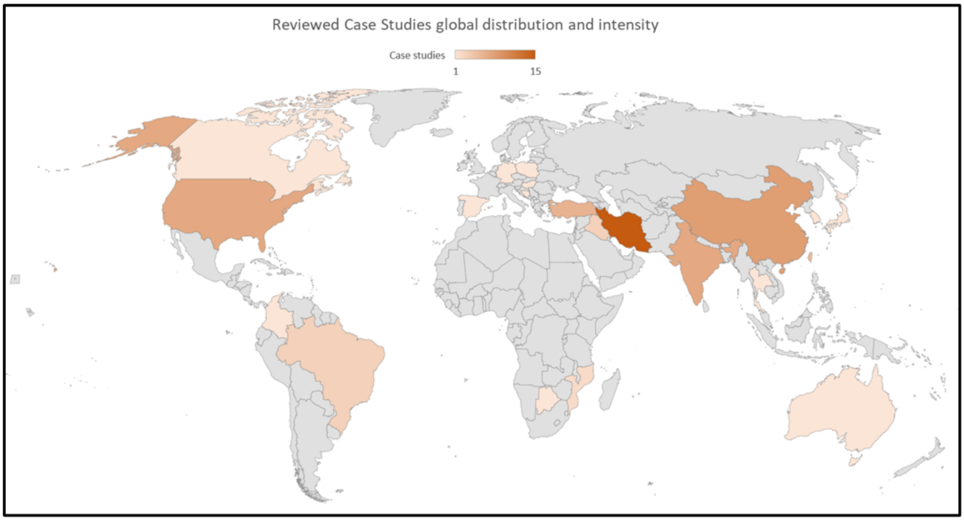

Figure 6: *Map of the regions hosting the reviewed reservoirs/dams and their intensity.*

and tuning to maintain generalizability across different scenarios. Future research should address these limitations by enhancing real-time data integration and developing adaptive learning mechanisms to improve model responsiveness. There is also a need for hybrid models that combine the strengths of various ML techniques to provide more holistic and accurate predictions. Incorporating socio-economic factors and climate change impacts into these models will be critical for sustainable water management. Finally, improving the interpretability and usability of ML models will ensure that they can be effectively utilized by reservoir operators and policymakers.

6. Conclusion

The present study reviewed previous research papers dealing with the application of machine learning techniques in dam and reservoir operations. Applications of ML models in predicting reservoir inflows, optimizing water release schedules, and managing flood risks have been studied extensively over the past years. Previous studies that used ML-based models, such as Neural Networks, Genetic Algorithms, Decision Trees, and Ensemble Methods, were explored as key components in enhancing the efficiency and accuracy of dam operations. Significant attention has been given by researchers to address the complexities posed by climate variability and increasing water demands.

The ML models demonstrated acceptable solutions for optimizing dam and reservoir operations. Based on the results obtained from several literature studies, advanced ML models like Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) showed substantial improvements in forecasting accuracy and decision-making processes. However, there are some drawbacks and shortcomings for these techniques that can be observed during their application, such as computational intensity, scalability issues, and the need for real-time data integration and adaptive learning mechanisms.

This research highlighted the efforts made to develop predictive models based on ML methods. It was illustrated that ML-based models have the ability to predict inflow parameters with a good level of accuracy. Modern models, such as integrated predictive models, are more effective and reliable compared to the classic versions. More recommendations to develop hybrid models combining ML with traditional optimization techniques have been listed. To operate dam and reservoir systems under realistic conditions, a new simulation procedure has been proposed. This procedure involves combining predictive models with optimization algorithms while searching for optimal operating rules and evaluating their performance. This approach aims to pave the way for more effective and sustainable dam and reservoir management practices.

Reference

[1] S. K. Jain and N. I. H. Roorkee, "Introduction to reservoir operation," Natl. Inst. Hydrol. Roorkee, 2019.
[2] C. J. A. Binnie, "The benefit of dams to society," in Long-term benefits and performance of dams, 2004, pp. 3–14. doi: 10.1680/lbapod.32682.0001.
[3] S. Maxwell, "Historical Water Price Trends," J. AWWA, vol. 102, no. 4, pp. 24–28, Apr. 2010, doi: 10.1002/j.1551-8833.2010.tb10086.x.
[4] S. Ko, D. G. Fontane, and J. W. Labadie, "MULTIOBJECTWE OPTIMIZATION OF RESERVOIR SYSTEMS OPERATION 1," JAWRA J. Am. Water Resour. Assoc., vol. 28, no. 1, pp. 111–127, Feb. 1992, doi: 10.1111/j.1752-1688.1992.tb03158.x.
[5]A. Badr, Z. Li, and W. El-Dakhakhni, "Dam System and Reservoir Operational Safety: A Meta-Research," Water, vol. 15, no. 19, p. 3427, Sep. 2023, doi: 10.3390/w15193427.
[6] B. Fazli et al., "Improvement of Dam Management in Terms of WAM Using Machine Learning," 2020, pp. 226–236. doi: 10.1007/978-981-15-1971-0.23.
[7] M. A. Ghorbani, R. Khatibi, V. Karimi, Z. M. Yaseen, and M. Zounemat-Kermani, "Learning from Multiple Models Using Artificial Intelligence to Improve Model Prediction Accuracies: Application to River Flows," Water Resour. Manag., vol. 32, no. 13, pp. 4201–4215, Oct. 2018, doi: 10.1007/s11269-018-2038-x.
[8] K. L. Chong et al., "Review on Dam and Reservoir Optimal Operation for Irrigation and Hydropower Energy Generation Utilizing Meta-Heuristic Algorithms," IEEE Access, 2021, doi: 10.1109/ACCESS.2021.3054424.
[9] H. A. Afan et al., "Input attributes optimization using the feasibility of genetic nature inspired algorithm: Application of river flow forecasting," Sci. Rep., vol. 10, no. 1, pp. 1–15, 2020.
[10] D. Zhang et al., "Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm," J. Hydrol., 2018, doi: 10.1016/j.jhydrol.2018.08.050.
[11] J. Xie, J. Zhang, X. Xie, Z. Bi, and Z. Li, "Ensemble of bagged regression trees for concrete dam deformation predicting," IOP Conf. Ser. Earth Environ. Sci., vol. 376, no. 1, p. 012040, Nov. 2019, doi: 10.1088/1755-1315/376/1/012040.
[12] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," J. Hydrol., vol. 598, 2021.
[3] G. Zhao, H. Gao, B. S. Naz, S.-C. Kao, and N. Voisin, "Integrating a reservoir regulation scheme into a spatially distributed hydrological model," Adv. Water Resour., vol. 98, pp. 16–31, Dec. 2016, doi: 10.1016/j.advwatres.2016.10.014.
[14] T. Zhou, B. Nijssen, H. Gao, and D. P. Lettenmaier, "The Contribution of Reservoirs to Global Land Surface Water Storage Variations*," J. Hydrometeorol., vol. 17, no. 1, pp. 309–325, Jan. 2016, doi: 10.1175/JHM-D-15-0002.1.
[15] A. F. Teixeira and A. R. Secchi, "Machine learning models to support reservoir production optimization," IFAC-PapersOnLine, vol. 52, no. 1, pp. 498–501, 2019, doi: 10.1016/j.ifacol.2019.06.111.
[16] N. S. Hsu and C. C. Wei, "A multipurpose reservoir real-time operation model for flood control during typhoon invasion," J. Hydrol., vol. 336, no. 3–4, pp. 282–293, 2007, doi: 10.1016/j.jhydrol.2007.01.001.
[17] M. Campolo, P. Andreussi, and A. Soldati, "River flood forecasting with a neural network model," Water Resour. Res., vol. 35, no. 4, pp. 1191–1197, 1999, doi: 10.1029/1998WR900086.
[18] X. Zhan et al., "Variational Bayesian Neural Network for Ensemble Flood Forecasting," Water, vol. 12, no. 10, p. 2740, Sep. 2020, doi: 10.3390/w12102740.

[19] O. García-Feal, J. González-Cao, D. Fernández-Nóvoa, G. Astray Dopazo, and M. Gómez-Gesteira, "Comparison of machine learning techniques for reservoir outflow forecasting," Nat. Hazards Earth Syst. Sci., vol. 22, no. 12, pp. 3859–3874, Dec. 2022, doi: 10.5194/nhess-22-3859-2022.
[20] S. S. Sammen, M. Ehteram, Z. Sheikh Khozani, and L. M. Sidek, "Binary Coati Optimization Algorithm- Multi- Kernel Least Square Support Vector Machine-Extreme Learning Machine Model (BCOA-MKLSSVM-ELM): A New Hybrid Machine Learning Model for Predicting Reservoir Water Level," Water, vol. 15, no. 8, p. 1593, Apr. 2023, doi: 10.3390/w15081593.
[21] D. Schwanenberg, F. M. Fan, S. Naumann, J. I. Kuwajima, R. A. Montero, and A. Assis dos Reis, "Short-Term Reservoir Optimization for Flood Mitigation under Meteorological and Hydrological Forecast Uncertainty," Water Resour. Manag., vol. 29, no. 5, pp. 1635–1651, Mar. 2015, doi: 10.1007/s11269-014-0899-1.
[22] V. Ramaswamy and F. Saleh, "Ensemble Based Forecasting and Optimization Framework to Optimize Releases from Water Supply Reservoirs for Flood Control," Water Resour. Manag., vol. 34, no. 3, pp. 989–1004, Feb. 2020, doi: 10.1007/s11269-019-02481-8.
[23] F. Wang, L. Wang, H. Zhou, O. C. Saavedra Valeriano, T. Koike, and W. Li, "Ensemble hydrological prediction-based real-time optimization of a multiobjective reservoir during flood season in a semiarid basin with global numerical weather predictions," Water Resour. Res., vol. 48, no. 7, Jul. 2012, doi: 10.1029/2011WR011366.
[24] D. Nohara, Y. Nishioka, T. Hori, and Y. Sato, "Real-Time Reservoir Operation for Flood Management Considering Ensemble Streamflow Prediction and Its Uncertainty," 2016, pp. 333–347. doi: 10.1007/978-981-287-615-7.23.
[25] P. P. Hadiyan, R. Moeini, and E. Ehsanzadeh, "Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Sefidroud Dam reservoir," Sustain. Comput. Informatics Syst., 2020, doi: 10.1016/j.suscom.2020.100401.
[26] U. Dampage, Y. Gunaratne, O. Bandara, S. De Silva, and V. Waraketiya, "Artificial Neural Network for Forecasting of Daily Reservoir Inflow: Case Study of the Kotmale Reservoir in Sri Lanka," in 2020 5th International Conference on Computational Intelligence and Applications (ICCIA), IEEE, Jun. 2020, pp. 8–12. doi: 10.1109/ICCIA49625.2020.00009.
[27] S. K. Ahmad and F. Hossain, "A generic data-driven technique for forecasting of reservoir inflow: Application for hydropower maximization," Environ. Model. Softw., vol. 119, pp. 147–165, Sep. 2019, doi: 10.1016/j.envsoft.2019.06.008.
[28] W. D. Xu, T. D. Fletcher, M. J. Burns, and F. Cherqui, "Real Time Control of Rainwater Harvesting Systems: The Benefits of Increasing Rainfall Forecast Window," Water Resour. Res., vol. 56, no. 9, Sep. 2020, doi: 10.1029/2020WR027856.
[29] F. H. S. Chiew, S. L. Zhou, and T. A. McMahon, "Use of seasonal streamflow forecasts in water resources management," J. Hydrol., vol. 270, no. 1–2, pp. 135–144, Jan. 2003, doi: 10.1016/S0022-1694(02)00292-5.
[30] B. George, H. Malano, B. Davidson, P. Hellegers, L. Bharati, and S. Massuel, "An integrated hydro-economic modelling framework to evaluate water allocation strategies II: Scenario assessment," Agric. Water Manag., vol. 98, no. 5, pp. 747–758, Mar. 2011, doi: 10.1016/j.agwat.2010.12.005.
[31] T. V. Thang, N. T. Thu Nga, and N. Le Long, "Optimum contributions of hydropower reservoirs to the minimum flow of Vu Gia – Thu Bon river basin," Water Pract. Technol., vol. 15, no. 4, pp. 1083–1095, Dec. 2020, doi: 10.2166/wpt.2020.082.
[32] H. Zamani Sabzi, S. Abudu, R. Alizadeh, L. Soltanisehat, N. Dilekli, and J. P. King, "Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources," Energy Sources, Part A Recover. Util. Environ. Eff., vol. 40, no. 11, pp. 1398–1416, Jun. 2018, doi: 10.1080/15567036.2018.1476934.

[33] L. Hua et al., "Floodwater Utilization Based on Reservoir Pre-Release Strategy Considering the Worst-Case Scenario," Water, vol. 12, no. 3, p. 892, Mar. 2020, doi: 10.3390/w12030892.
[34] N. Jaworski, W. Weber, and R. A. Deininger, "Optimal Reservoir Releases for Water Quality Control," J. Sanit. Eng. Div., vol. 96, pp. 727–742, 1970.
[35] M. S. Hajilal, N. H. Rao, and P. B. S. Sarma, "Real time operation of reservoir based canal irrigation systems," Agric. Water Manag., vol. 38, no. 2, pp. 103–122, Dec. 1998, doi: 10.1016/S0378-3774(98)00061-4.
[36] H. Albo-Salih and L. Mays, "Testing of an Optimization-Simulation Model for Real-Time Flood Operation of River-Reservoir Systems," Water, vol. 13, no. 9, p. 1207, Apr. 2021, doi: 10.3390/w13091207.
[37] D. Che and L. W. Mays, "Development of an Optimization/Simulation Model for Real-Time Flood-Control Operation of River-Reservoirs Systems," Water Resour. Manag., vol. 29, no. 11, pp. 3987–4005, Sep. 2015, doi: 10.1007/s11269-015-1041-8.
[38] A. Karbowski, K. Malinowski, and E. Niewiadomska-Szynkiewicz, "A hybrid analytic/rule-based approach to reservoir system management during flood," Decis. Support Syst., vol. 38, no. 4, pp. 599–610, Jan. 2005, doi: 10.1016/j.dss.2003.10.001.
[39] E. Bednárová, J. Škvarka, P. Václavík, and J. Poórová, "Water management system Liptovská Mara – Bešeňová in the context of climate change," Acta Hydrol. Slovaca, vol. 22, no. 1, pp. 15–21, Jun. 2021, doi: 10.31577/ahs-2021-0022.01.0002.
[40] H. Nouasse, P. Chiron, and B. Archimède, "Contribution to a flood situation management: a supervisory control scheme to reduce disaster impact," Water Supply, vol. 16, no. 3, pp. 587–598, Jun. 2016, doi: 10.2166/ws.2015.160.
[41] X. Zhu et al., "Adaptive flood control operation of the Xin'an Reservoir in future precipitation extremes under climate change," Arab. J. Geosci., vol. 13, no. 15, p. 720, Aug. 2020, doi: 10.1007/s12517-020-05711-1.
[42] B. S. Naz, S.-C. Kao, M. Ashfaq, H. Gao, D. Rastogi, and S. Gangrade, "Effects of climate change on streamflow extremes and implications for reservoir inflow in the United States," J. Hydrol., vol. 556, pp. 359–370, Jan. 2018, doi: 10.1016/j.jhydrol.2017.11.027.
[43] J. I. Kuwajima, F. M. Fan, D. Schwanenberg, A. Assis Dos Reis, A. Niemann, and F. F. Mauad, "Climate change, water-related disasters, flood control and rainfall forecasting: a case study of the São Francisco River, Brazil," Geol. Soc. London, Spec. Publ., vol. 488, no. 1, pp. 259–276, Jan. 2019, doi: 10.1144/SP488-2018-128.
[44] J. C. Carron and H. Rajaram, "Impact of variable reservoir releases on management of downstream water temperatures," Water Resour. Res., vol. 37, no. 6, pp. 1733–1743, Jun. 2001, doi: 10.1029/2000WR900390.
[45] L. Amblard, "Collective action for water quality management in agriculture: The case of drinking water source protection in France," Glob. Environ. Chang., vol. 58, p. 101970, Sep. 2019, doi: 10.1016/j.gloenvcha.2019.101970.
[46] A. Dan Tarlock, "Water Supply as New Growth Management Tool," L. Use Law Zo. Dig., vol. 50, no. 11, pp. 3–7, Nov. 1998, doi: 10.1080/00947598.1998.10395976.
[47] Z. Ravar, B. Zahraie, A. Sharifinejad, H. Gozini, and S. Jafari, "System dynamics modeling for assessment of water–food–energy resources security and nexus in Gavkhuni basin in Iran," Ecol. Indic., vol. 108, p. 105682, Jan. 2020, doi: 10.1016/j.ecolind.2019.105682.
[48] B. Ming, P. Liu, S. Guo, X. Zhang, M. Feng, and X. Wang, "Optimizing utility-scale photovoltaic power generation for integration into a hydropower reservoir by incorporating long-and short-term operational decisions," Appl. Energy, vol. 204, pp. 432–445, 2017.
[49] D. McLaughlin and H. L. Velasco, "Real-time control of a system of large hydropower reservoirs," Water Resour. Res., vol. 26, no. 4, pp. 623–635, Apr. 1990, doi: 10.1029/WR026i004p00623.
[50] M. Chazarra, J. García-González, J. I. Pérez-Díaz, and M. Arteseros, "Stochastic optimiza-

tion model for the weekly scheduling of a hydropower system in day-ahead and secondary regulation reserve markets," Electr. Power Syst. Res., vol. 130, pp. 67–77, Jan. 2016, doi: 10.1016/j.epsr.2015.08.014.
[51] S. J. Pereira-Cardenal, B. Mo, A. Gjelsvik, N. D. Riegels, K. Arnbjerg-Nielsen, and P. Bauer-Gottwein, "Joint optimization of regional water-power systems," Adv. Water Resour., vol. 92, pp. 200–207, Jun. 2016, doi: 10.1016/j.advwatres.2016.04.004.
[52] A. Tayebiyan, T. A. Mohammed Ali, A. H. Ghazali, and M. A. Malek, "Optimization of Exclusive Release Policies for Hydropower Reservoir Operation by Using Genetic Algorithm," Water Resour. Manag., vol. 30, no. 3, pp. 1203–1216, Feb. 2016, doi: 10.1007/s11269-015-1221-6.
[53] Luo Yunxia, Wang Wanliang, and Zhou Muxun, "Study on optimal scheduling model and technology based on RPSO for small hydropower sustainability," in 2009 International Conference on Sustainable Power Generation and Supply, IEEE, Apr. 2009, pp. 1–5. doi: 10.1109/SUPERGEN.2009.5347872.
[54] J. Bernardes et al., "Hydropower Operation Optimization Using Machine Learning: A Systematic Review," AI, vol. 3, no. 1, pp. 78–99, Feb. 2022, doi: 10.3390/ai3010006.
[55] S. D. Latif, A. N. Ahmed, E. Sathiamurthy, Y. F. Huang, and A. El-Shafie, "Evaluation of deep learning algorithm for inflow forecasting: a case study of Durian Tunggal Reservoir, Peninsular Malaysia," Nat. Hazards, vol. 109, no. 1, pp. 351–369, 2021, doi: 10.1007/s11069-021-04839-x.
[56] H. Rahimi, M. K. Ardakani, M. Ahmadian, and X. Tang, "Multi-Reservoir Utilization Planning to Optimize Hydropower Energy and Flood Control Simultaneously," Environ. Process., vol. 7, no. 1, pp. 41–52, Mar. 2020, doi: 10.1007/s40710-019-00404-8.
[57] S. Sterl, P. Donk, P. Willems, and W. Thiery, "Turbines of the Caribbean: Decarbonising Suriname's electricity mix through hydro-supported integration of wind power," Renew. Sustain. Energy Rev., vol. 134, p. 110352, Dec. 2020, doi: 10.1016/j.rser.2020.110352.
[58] O. Olofintoye, F. Otieno, and J. Adeyemo, "Real-time optimal water allocation for daily hydropower generation from the Vanderkloof dam, South Africa," Appl. Soft Comput., vol. 47, pp. 119–129, Oct. 2016, doi: 10.1016/j.asoc.2016.05.018.
[59] R. Viadero, M. Rehbein, and A. Singh, "Hydropower on the Mississippi River," 2017.
[60] G. M. Kondolf et al., "Sustainable sediment management in reservoirs and regulated rivers: Experiences from five continents," Earth's Futur., vol. 2, no. 5, pp. 256–280, May 2014, doi: 10.1002/2013EF000184.
[61] E. Setiawan, S. Hidayat, I. Bagus Giri Putra, M. Bagus Budianto, and Salehudin, "Evaluation of sediment management for two large reservoirs in Lombok island," MATEC Web Conf., vol. 195, p. 05002, Aug. 2018, doi: 10.1051/matecconf/201819505002.
[62] H. Doi, "140. Reservoir Sedimentation Management at the Dashidaira and Asahi Dams," Tunn. Undergr. Sp. Technol., vol. 15, p. 46, 2005.
[63] B. Bhattacharya and D. P. Solomatine, "Machine learning in sedimentation modelling," Neural Networks, vol. 19, no. 2, pp. 208–214, 2006, doi: 10.1016/j.neunet.2006.01.007.
[64] N. Ahmed, S. Mahmud, M. M. Lutfe Elahi, S. Ahmed, and M. Sujauddin, "Forecasting river sediment deposition through satellite image driven unsupervised machine learning techniques," Remote Sens. Appl. Soc. Environ., vol. 13, pp. 435–444, Jan. 2019, doi: 10.1016/j.rsase.2018.12.011.
[65] P. Espa, M. L. Brignoli, G. Crosa, G. Gentili, and S. Quadroni, "Controlled sediment flushing at the Cancano Reservoir (Italian Alps): Management of the operation and downstream environmental impact," J. Environ. Manage., vol. 182, pp. 1–12, Nov. 2016, doi: 10.1016/j.jenvman.2016.07.021.
[66] F. Cattanéo, J. Guillard, S. Diouf, J. O'Rourke, and D. Grimardias, "Mitigation of ecological impacts on fish of large reservoir sediment management through controlled flushing – The case

of the Verbois dam (Rhône River, Switzerland)," Sci. Total Environ., vol. 756, p. 144053, Feb. 2021, doi: 10.1016/j.scitotenv.2020.144053.
[67] A. Palmieri, F. Shah, and A. Dinar, "Economics of reservoir sedimentation and sustainable management of dams," J. Environ. Manage., vol. 61, no. 2, pp. 149–163, Feb. 2001, doi: 10.1006/jema.2000.0392.
[68] G. De Cesare, A. Schleiss, and F. Hermann, "Impact of Turbidity Currents on Reservoir Sedimentation," J. Hydraul. Eng., vol. 127, no. 1, pp. 6–16, Jan. 2001, doi: 10.1061/(ASCE)0733-9429(2001)127:1(6).
[69] B. D. Richter, "Re-thinking environmental flows: from allocations and reserves to sustainability boundaries," River Res. Appl., vol. 26, no. 8, pp. 1052–1063, Oct. 2010, doi: 10.1002/rra.1320.
[70] M. Acreman et al., "Environmental flows for natural, hybrid, and novel riverine ecosystems in a changing world," Front. Ecol. Environ., vol. 12, no. 8, pp. 466–473, Oct. 2014, doi: 10.1890/130134.
[71] A. J. King, D. C. Gwinn, Z. Tonkin, J. Mahoney, S. Raymond, and L. Beesley, "Using abiotic drivers of fish spawning to inform environmental flow management," J. Appl. Ecol., vol. 53, no. 1, pp. 34–43, Feb. 2016, doi: 10.1111/1365-2664.12542.
[72] R. D. Moreno Santillan, D. Yang, H. Wang, M. S. Islam, and S. Z. Farooq, "Estimation of Water Discharge and Its Anomalies Using Remote Sensing," in 2018 37th Chinese Control Conference (CCC), IEEE, Jul. 2018, pp. 10355–10360. doi: 10.23919/ChiCC.2018.8484158.
[73] H. Tongal and M. J. Booij, "Simulation and forecasting of streamflows using machine learning models coupled with base flow separation," J. Hydrol., 2018, doi: 10.1016/j.jhydrol.2018.07.004.
[74] J. D. Tonkin, J. D. Olden, D. M. Merritt, L. V Reynolds, J. S. Rogosch, and D. A. Lytle, "Designing flow regimes to support entire river ecosystems," Front. Ecol. Environ., vol. 19, no. 6, pp. 326–333, Aug. 2021, doi: 10.1002/fee.2348.
[75] C. Allan and R. J. Watts, "Revealing Adaptive Management of Environmental Flows," Environ. Manage., vol. 61, no. 3, pp. 520–533, Mar. 2018, doi: 10.1007/s00267-017-0931-3.
[76] J. Westerink et al., "Collaborative governance arrangements to deliver spatially coordinated agri-environmental management," Land use policy, vol. 69, pp. 176–192, Dec. 2017, doi: 10.1016/j.landusepol.2017.09.002.
[77] R. D. Fish, A. A. R. Ioris, and N. M. Watson, "Integrating water and agricultural management: Collaborative governance for a complex policy problem," Sci. Total Environ., vol. 408, no. 23, pp. 5623–5630, Nov. 2010, doi: 10.1016/j.scitotenv.2009.10.010.
[78] S. Pouso, S. Ferrini, R. K. Turner, Á. Borja, and M. C. Uyarra, "Monetary valuation of recreational fishing in a restored estuary and implications for future management measures," ICES J. Mar. Sci., vol. 77, no. 6, pp. 2295–2303, Nov. 2020, doi: 10.1093/icesjms/fsz091.
[79] P. Olsson, C. Folke, and T. Hahn, "Social-Ecological Transformation for Ecosystem Management: the Development of Adaptive Co-management of a Wetland Landscape in Southern Sweden," Ecol. Soc., vol. 9, no. 4, p. art2, 2004, doi: 10.5751/ES-00683-090402.
[80] H. L. KEOUGH and D. J. BLAHNA, "Achieving Integrative, Collaborative Ecosystem Management," Conserv. Biol., vol. 20, no. 5, pp. 1373–1382, Oct. 2006, doi: 10.1111/j.1523-1739.2006.00445.x.
[81] L. L. Zhang, "Supply chain planning with integrated decision making in resource allocation," in 2015 International Conference on Industrial Engineering and Operations Management (IEOM), IEEE, Mar. 2015, pp. 1–6. doi: 10.1109/IEOM.2015.7093950.
[82] A. Lambora, K. Gupta, and K. Chopra, "Genetic Algorithm- A Literature Review," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, Feb. 2019, pp. 380–384. doi: 10.1109/COMITCon.2019.8862255.
[83] 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, doi: 10.1016/j.engappai.2023.106813.
[84] M. Liu, Z. Wen, R. Zhou, and H. Su, "Bayesian optimization and ensemble learning algorithm combined method for deformation prediction of concrete dam," Structures, vol. 54, pp. 981–993, Aug. 2023, doi: 10.1016/j.istruc.2023.05.136.
[85] X. Li, Z. Wen, and H. Su, "An approach using random forest intelligent algorithm to construct a monitoring model for dam safety," Eng. Comput., vol. 37, no. 1, pp. 39–56, Jan. 2021, doi: 10.1007/s00366-019-00806-0.
[86] L. Wang et al., "Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method," Acta Geotech., vol. 15, no. 11, pp. 3135–3150, Nov. 2020, doi: 10.1007/s11440-020-00962-4.
[87] X. Tang, A. Chen, and J. He, "A modelling approach based on Bayesian networks for dam risk analysis: Integration of machine learning algorithm and domain knowledge," Int. J. Disaster Risk Reduct., vol. 71, p. 102818, Mar. 2022, doi: 10.1016/j.ijdrr.2022.102818.
[88] A. Kalinina, M. Spada, and P. Burgherr, "Application of a Bayesian hierarchical modeling for risk assessment of accidents at hydropower dams," Saf. Sci., vol. 110, pp. 164–177, Dec. 2018, doi: 10.1016/j.ssci.2018.08.006.
[89] H. K. R. Okada, A. R. N. das Neves, and R. Shitsuka, "Analysis of Decision Tree Induction Algorithms," Res. Soc. Dev., vol. 8, no. 11, p. e298111473, Aug. 2019, doi: 10.33448/rsd-v8i11.1473.
[90] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar joseph, "A Review of Dimensionality Reduction Techniques for Efficient Computation," Procedia Comput. Sci., vol. 165, pp. 104–111, 2019, doi: 10.1016/j.procs.2020.01.079.
[91] S. Nanga et al., "Review of Dimension Reduction Methods," J. Data Anal. Inf. Process., vol. 09, no. 03, pp. 189–231, 2021, doi: 10.4236/jdaip.2021.93013.
[92] V. Kumar, P. Samui, A. Burman, and N. Himanshu, "Multivariate Adaptive Regression Spline Based Reliability Analysis of Stability of Durgawati Earthen Dam," 2021, pp. 81–93. doi: 10.1007/978-981-15-6233-4.6.
[93] L. Wang, C. Wu, X. Gu, H. Liu, G. Mei, and W. Zhang, "Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines," Bull. Eng. Geol. Environ., 2020, doi: 10.1007/s10064-020-01730-0.
[94] M. Li, Y. Shen, Q. Ren, and H. Li, "A new distributed time series evolution prediction model for dam deformation based on constituent elements," Adv. Eng. Informatics, vol. 39, pp. 41–52, Jan. 2019, doi: 10.1016/j.aei.2018.11.006.
[95] M. Haggag, "Adjusting the Penalized Term for the Regularized Regression Models," Afrika Stat., vol. 13, no. 2, pp. 1609–1630, Apr. 2018, doi: 10.16929/as/1609.124.
[96] I. Gacko, Z. Muchová, Ľ. Jurík, K. Šinka, L. Fabian, and F. Petrovič, "Decision Making Methods to Optimize New Dam Site Selections on the Nitra River," Water, vol. 12, no. 7, p. 2042, Jul. 2020, doi: 10.3390/w12072042.
[97] J. Hu and F. Ma, "Comparison of hierarchical clustering based deformation prediction models for high arch dams during the initial operation period," J. Civ. Struct. Heal. Monit., vol. 11, no. 4, pp. 897–914, Sep. 2021, doi: 10.1007/s13349-021-00487-8.
[98] D. Chen, A. S. Leon, N. L. Gibson, and P. Hosseini, "Dimension reduction of decision variables for multireservoir operation: A spectral optimization model," Water Resour. Res., vol. 52, no. 1, pp. 36–51, Jan. 2016, doi: 10.1002/2015WR017756.
[99] S. Arunkumar and K. Chon, "On Optimal Regulation Policies for Certain Multi-Reservoir Systems," Oper. Res., vol. 26, no. 4, pp. 551–562, Aug. 1978, doi: 10.1287/opre.26.4.551.
[100] S. Arunkumar, "Optimal regulation policies for a multipurpose reservoir with seasonal input and return function," J. Optim. Theory Appl., vol. 21, no. 3, pp. 319–328, Mar. 1977, doi:

10.1007/BF00933533.
[101] S. Arunkumar, "Characterization of optimal operating policies for finite dams," J. Math. Anal. Appl., vol. 49, no. 2, pp. 267–274, Feb. 1975, doi: 10.1016/0022-247X(75)90178-X.
[102] C. R. Gagnon, R. H. Hicks, S. L. S. Jacoby, and J. S. Kowalik, "A nonlinear programming approach to a very large hydroelectric system optimization," Math. Program., vol. 6, no. 1, pp. 28–41, Dec. 1974, doi: 10.1007/BF01580220.
[103] G. Belaineh, R. C. Peralta, and T. C. Hughes, "Simulation/Optimization Modeling for Water Resources Management," J. Water Resour. Plan. Manag., vol. 125, no. 3, pp. 154–161, May 1999, doi: 10.1061/(ASCE)0733-9496(1999)125:3(154).
[104] X. Li, J. Wei, T. Li, G. Wang, and W. W.-G. Yeh, "A parallel dynamic programming algorithm for multi-reservoir system optimization," Adv. Water Resour., vol. 67, pp. 1–15, May 2014, doi: 10.1016/j.advwatres.2014.01.002.
[105] M. Heydari, F. Othman, and K. Qaderi, "Developing Optimal Reservoir Operation for Multiple and Multipurpose Reservoirs Using Mathematical Programming," Math. Probl. Eng., vol. 2015, pp. 1–11, 2015, doi: 10.1155/2015/435752.
[106] H. Feizi, M. T. Sattari, and H. Apaydin, "A comparative study of different optimization algorithms for the optimum operation of the Mahabad dam reservoir," Results Eng., vol. 21, p. 101664, Mar. 2024, doi: 10.1016/j.rineng.2023.101664.
[107] S. Kosasaeng and A. Kangrang, "Optimum reservoir operation of a networking reservoirs system using conditional atom search optimization and a conditional genetic algorithm," Heliyon, vol. 9, no. 3, p. e14467, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14467.
[108] C.-B. Zhou, W. Liu, Y.-F. Chen, R. Hu, and K. Wei, "Inverse modeling of leakage through a rockfill dam foundation during its construction stage using transient flow model, neural network and genetic algorithm," Eng. Geol., vol. 187, pp. 183–195, Mar. 2015, doi: 10.1016/j.enggeo.2015.01.008.
[109] B. Stojanovic, M. Milivojevic, M. Ivanovic, N. Milivojevic, and D. Divac, "Adaptive system for dam behavior modeling based on linear regression and genetic algorithms," Adv. Eng. Softw., vol. 65, pp. 182–190, Nov. 2013, doi: 10.1016/j.advengsoft.2013.06.019.
[110] F.-J. Chang, L. Chen, and L.-C. Chang, "Optimizing the reservoir operating rule curves by genetic algorithms," Hydrol. Process., vol. 19, no. 11, pp. 2277–2289, Jul. 2005, doi: 10.1002/hyp.5674.
[111] F. Chang, J. Lai, and L. Kao, "Optimization of operation rule curves and flushing schedule in a reservoir," Hydrol. Process., vol. 17, no. 8, pp. 1623–1640, Jun. 2003, doi: 10.1002/hyp.1204.
[112] L.-C. Chang, F.-J. Chang, K.-W. Wang, and S.-Y. Dai, "Constrained genetic algorithms for optimizing multi-use reservoir operation," J. Hydrol., vol. 390, no. 1–2, pp. 66–74, Aug. 2010, doi: 10.1016/j.jhydrol.2010.06.031.
[113] Bilal, M. Pant, and D. Rani, "Large scale reservoir operation through integrated meta-heuristic approach," Memetic Comput., vol. 13, no. 3, pp. 359–382, Sep. 2021, doi: 10.1007/s12293-021-00327-8.
[114] W. El Harraki, D. Ouazar, A. Bouziane, and D. Hasnaoui, "Optimization of reservoir operating curves and hedging rules using genetic algorithm with a new objective function and smoothing constraint: application to a multipurpose dam in Morocco," Environ. Monit. Assess., vol. 193, no. 4, p. 196, Apr. 2021, doi: 10.1007/s10661-021-08972-9.
[115] R. Qiu, D. Wang, V. P. Singh, Y. Wang, and J. Wu, "Integration of deep learning and improved multi-objective algorithm to optimize reservoir operation for balancing human and downstream ecological needs," Water Res., vol. 253, p. 121314, Apr. 2024, doi: 10.1016/j.watres.2024.121314.
[116] L. Chang and F. Chang, "Intelligent control for modelling of real-time reservoir operation," Hydrol. Process., vol. 15, no. 9, pp. 1621–1634, Jun. 2001, doi: 10.1002/hyp.226.

[117] P. Chaves and F.-J. Chang, "Intelligent reservoir operation system based on evolving artificial neural networks," Adv. Water Resour., vol. 31, no. 6, pp. 926–936, Jun. 2008, doi: 10.1016/j.advwatres.2008.03.002.
[118] P. Chaves and T. Kojiri, "Stochastic Fuzzy Neural Network: Case Study of Optimal Reservoir Operation," J. Water Resour. Plan. Manag., vol. 133, no. 6, pp. 509–518, Nov. 2007, doi: 10.1061/(ASCE)0733-9496(2007)133:6(509).
[119] R. Naresh and J. Sharma, "Reservoir operation using a fuzzy and neural system for hydrogeneration scheduling," Int. J. Syst. Sci., vol. 32, no. 4, pp. 479–486, Jan. 2001, doi: 10.1080/00207720119896.
[120] D.-H. BAE, D. M. JEONG, and G. KIM, "Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique," Hydrol. Sci. J., vol. 52, no. 1, pp. 99–113, Feb. 2007, doi: 10.1623/hysj.52.1.99.
[121] W. S. Jung and Y. Do Kim, "Evaluation of Watershed Water Quality Management According to Flow Conditions through Factor Analysis and Naïve Bayes Classifier," Sustainability, vol. 15, no. 13, p. 10038, Jun. 2023, doi: 10.3390/su151310038.
[122] M. S. Khan and P. Coulibaly, "Bayesian neural network for rainfall-runoff modeling," Water Resour. Res., vol. 42, no. 7, Jul. 2006, doi: 10.1029/2005WR003971.
[123] M. Karamouz, A. Ahmadi, and A. Moridi, "Probabilistic reservoir operation using Bayesian stochastic model and support vector machine," Adv. Water Resour., vol. 32, no. 11, pp. 1588–1600, Nov. 2009, doi: 10.1016/j.advwatres.2009.08.003.
[124] J. Chen, P.-A. Zhong, R. An, F. Zhu, and B. Xu, "Risk analysis for real-time flood control operation of a multi-reservoir system using a dynamic Bayesian network," Environ. Model. Softw., vol. 111, pp. 409–420, Jan. 2019, doi: 10.1016/j.envsoft.2018.10.007.
[125] T. Zhou, Z. Dong, X. Chen, and Q. Ran, "Decision Support Model for Ecological Operation of Reservoirs Based on Dynamic Bayesian Network," Water, vol. 13, no. 12, p. 1658, Jun. 2021, doi: 10.3390/w13121658.
[126] P. Noorbeh, A. Roozbahani, and H. Kardan Moghaddam, "Annual and Monthly Dam Inflow Prediction Using Bayesian Networks," Water Resour. Manag., vol. 34, no. 9, pp. 2933–2951, Jul. 2020, doi: 10.1007/s11269-020-02591-8.
[127] J. Chen and P.-A. Zhong, "A multi-time-scale power prediction model of hydropower station considering multiple uncertainties," Sci. Total Environ., vol. 677, pp. 612–625, Aug. 2019, doi: 10.1016/j.scitotenv.2019.04.430.
[128] B. Malekmohammadi, R. Kerachian, and B. Zahraie, "Developing monthly operating rules for a cascade system of reservoirs: Application of Bayesian Networks," Environ. Model. Softw., vol. 24, no. 12, pp. 1420–1432, Dec. 2009, doi: 10.1016/j.envsoft.2009.06.008.
[129] B. Malekmohammadi, R. Kerachian, and B. Zahraie, "Developing monthly operating rules for a cascade system of reservoirs: Application of Bayesian Networks," Environ. Model. Softw., vol. 24, no. 12, pp. 1420–1432, Dec. 2009, doi: 10.1016/j.envsoft.2009.06.008. M. Rezay Nazarzadeh, A. AkhondAli, and A. Daneshkhah, "Optimization of Reservoir Operation for Real-Time Flood Control with Emphasis on Forecast Uncertainty: A case study of Dez Reservoir," J. Hydraul. Struct., vol. 6, no. 3, pp. 92–107, 2020, doi: 10.22055/jhs.2021.34694.1146.
[130] B. Malekmohammadi, R. Kerachian, and B. Zahraie, "Developing monthly operating rules for a cascade system of reservoirs: Application of Bayesian Networks," Environ. Model. Softw., vol. 24, no. 12, pp. 1420–1432, Dec. 2009, doi: 10.1016/j.envsoft.2009.06.008. S. F. Stefenon et al., "Time series forecasting using ensemble learning methods for emergency prevention in hydroelectric power plants with dam," Electr. Power Syst. Res., vol. 202, p. 107584, Jan. 2022, doi: 10.1016/j.epsr.2021.107584.
[131] V. Kumar, N. Kedam, K. V. Sharma, D. J. Mehta, and T. Caloiero, "Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow

Prediction Models," Water, vol. 15, no. 14, p. 2572, Jul. 2023, doi: 10.3390/w15142572.
[132] Y. O. Ouma et al., "Dam Water Level Prediction Using Vector AutoRegression, Random Forest Regression and MLP-ANN Models Based on Land-Use and Climate Factors," Sustainability, vol. 14, no. 22, p. 14934, Nov. 2022, doi: 10.3390/su142214934.
[133] D. Deb, V. Arunachalam, and K. S. Raju, "Daily reservoir inflow prediction using stacking ensemble of machine learning algorithms," J. Hydroinformatics, Apr. 2024, doi: 10.2166/hydro.2024.210.
[134] P. Dornpunya et al., "The reservoir inflow prediction of Sirikit dam using artificial intelligence with machine learning: extreme gradient boosting techique," Artif. Intell., vol. 25, p. 26, 2021.
[135] F. LI, G. MA, S. CHEN, and W. HUANG, "An Ensemble Modeling Approach to Forecast Daily Reservoir Inflow Using Bidirectional Long- and Short-Term Memory (Bi-LSTM), Variational Mode Decomposition (VMD), and Energy Entropy Method," Water Resour. Manag., vol. 35, no. 9, pp. 2941–2963, Jul. 2021, doi: 10.1007/s11269-021-02879-3.
[136] P. A. Belyakova et al., "Forecasting Water Levels in Krasnodar Krai Rivers with the Use of Machine Learning," Water Resour., vol. 49, no. 1, pp. 10–22, Feb. 2022.
[137] G. Qie, Z. Zhang, E. Getahun, and E. Allen Mamer, "Comparison of Machine Learning Models Performance on Simulating Reservoir Outflow: A Case Study of Two Reservoirs in Illinois, U.S.A.," JAWRA J. Am. Water Resour. Assoc., vol. 59, no. 3, pp. 554–570, Jun. 2023, doi: 10.1111/1752-1688.13040.
[138] I. Parvez, J. Shen, I. Hassan, and N. Zhang, "Generation of Hydro Energy by Using Data Mining Algorithm for Cascaded Hydropower Plant," Energies, vol. 14, no. 2, p. 298, Jan. 2021, doi: 10.3390/en14020298.
[139] Y. Diao, C. Wang, H. Wang, and Y. Liu, "Construction and Application of Reservoir Flood Control Operation Rules Using the Decision Tree Algorithm," Water, vol. 13, no. 24, p. 3654, Dec. 2021, doi: 10.3390/w13243654.
[140] R. Salajegheh, A. Mahdavi-Meymand, and M. Zounemat-Kermani, "Evaluating performance of meta-heuristic algorithms and decision tree models in simulating water level variations of dams' piezometers," J. Hydraul. Struct., vol. 4, no. 2, pp. 60–80, 2018, doi: 10.22055/jhs.2018.27833.1092.
[141] T. Yang, X. Liu, L. Wang, P. Bai, and J. Li, "Simulating Hydropower Discharge using Multiple Decision Tree Methods and a Dynamical Model Merging Technique," J. Water Resour. Plan. Manag., vol. 146, no. 2, Feb. 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001146.
[142] T. Yang, X. Gao, S. Sorooshian, and X. Li, "Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme," Water Resour. Res., vol. 52, no. 3, pp. 1626–1651, Mar. 2016, doi: 10.1002/2015WR017394.
[143] M. Zounemat-Kermani, A. Ramezani-Charmahineh, R. Razavi, M. Alizamir, and T. B. M. J. Ouarda, "Machine Learning and Water Economy: a New Approach to Predicting Dams Water Sales Revenue," Water Resour. Manag., vol. 34, no. 6, pp. 1893–1911, Apr. 2020, doi: 10.1007/s11269-020-02529-0.
[144] S.-K. Yun, J. Kim, E.-S. Im, and G. Kang, "Relationships among Seepage, Water Level, and Rainfall of a Fill Dam by Decision Tree Analysis," Geofluids, vol. 2022, pp. 1–12, Jun. 2022, doi: 10.1155/2022/9253324.
[145] E. Goharian, M. Shaltout, M. Erfani, and A. Eladawy, "Developing an Optimized Policy Tree-Based Reservoir Operation Model for High Aswan Dam Reservoir, Nile River," Water, vol. 14, no. 7, p. 1061, Mar. 2022, doi: 10.3390/w14071061.
[146] T. Rymarczyk, G. Klosowski, and P. Tchorzewski, "Analysis of Leaks in Flood Embankments Using Deterministic Methods and Computational Intelligence Algorithms," in 2018 IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, Oct. 2018, pp. 1–6.

doi: 10.1109/IST.2018.8577115.
[147] M. Nakatsugawa, R. Sakamoto, and Y. Kobayashi, "Research on dam inflow prediction during severe flood using machine learning methods," in 22nd Congress of the International Association for Hydro-Environment Engineering and Research-Asia Pacific Division, IAHR-APD 2020: "Creating Resilience to Water-Related Challenges," 2020.
[148] L. Xie, "Estimation of parameters on Texas reservoirs using least absolute shrinkage and selection operator," Int. J. Energy Water Resour., vol. 3, no. 2, pp. 93–104, Jun. 2019, doi: 10.1007/s42108-019-00018-8.
[149] A. Ruhi et al., "How Does Flow Alteration Propagate Across a Large, Highly Regulated Basin? Dam Attributes, Network Context, and Implications for Biodiversity," Earth's Futur., vol. 10, no. 6, Jun. 2022, doi: 10.1029/2021EF002490.
[150] A. D. Mehr and A. H. Gandomi, "MSGP-LASSO: An improved multi-stage genetic programming model for streamflow prediction," Inf. Sci. (Ny)., vol. 561, pp. 181–195, Jun. 2021, doi: 10.1016/j.ins.2021.02.011.
[151] H. Chu, J. Wei, and W. Wu, "Streamflow prediction using LASSO-FCM-DBN approach based on hydro-meteorological condition classification," J. Hydrol., vol. 580, p. 124253, Jan. 2020, doi: 10.1016/j.jhydrol.2019.124253.
[152] A. D. Martinho, C. M. Saporetti, and L. Goliatt, "Approaches for the short-term prediction of natural daily streamflows using hybrid machine learning enhanced with grey wolf optimization," Hydrol. Sci. J., 2023, doi: 10.1080/02626667.2022.2141121.
[153] H. Chu, J. Wei, and J. Qiu, "Monthly streamflow forecasting using EEMD-Lasso-DBN method based on multi-scale predictors selection," Water (Switzerland), 2018.
[154] H. Chu, J. Wei, and Y. Jiang, "Middle- and Long-Term Streamflow Forecasting and Uncertainty Analysis Using Lasso-DBN-Bootstrap Model," Water Resour. Manag., vol. 35, no. 8, pp. 2617–2632, Jun. 2021, doi: 10.1007/s11269-021-02854-y.
[155] Y. Song and J. Zhang, "Enhancing short-term streamflow prediction in the Haihe River Basin through integrated machine learning with Lasso," Water Sci. Technol., vol. 89, no. 9, pp. 2367–2383, May 2024, doi: 10.2166/wst.2024.142.
[156] A. D. Martinho, H. S. Hippert, and L. Goliatt, "Short-term streamflow modeling using data-intelligence evolutionary machine learning models," Sci. Rep., vol. 13, no. 1, p. 13824, Aug. 2023, doi: 10.1038/s41598-023-41113-5.
[157] M. Sojka, J. Kanclerz, T. Dysarz, and J. Wicher-Dysarz, "Application of mulitivariate statistical methods for analysis of physical and chemical factors in reservoir with separated pre-dam zone on the basis of the example of Jezioro Kowalskie," Fresenius Environ. Bull., vol. 24, pp. 1516–1522, 2015.
[158] H. R. Pourghasemi, S. Yousefi, N. Sadhasivam, and S. Eskandari, "Assessing, mapping, and optimizing the locations of sediment control check dams construction," Sci. Total Environ., vol. 739, p. 139954, Oct. 2020, doi: 10.1016/j.scitotenv.2020.139954.
[159] P. Tanos, J. Kovács, S. Kovács, A. Anda, and I. G. Hatvani, "Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account," Environ. Monit. Assess., vol. 187, no. 9, p. 575, Sep. 2015, doi: 10.1007/s10661-015-4777-y.
[160] M. Varol, B. Gökot, A. Bekleyen, and B. Şen, "Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey," CATENA, vol. 92, pp. 11–21, May 2012, doi: 10.1016/j.catena.2011.11.013.
[161] M. Saad and A. Turgeon, "Application of principal component analysis to long-term reservoir management," Water Resour. Res., vol. 24, no. 7, pp. 907–912, Jul. 1988, doi: 10.1029/WR024i007p00907.
[162] M. Giuliani, S. Galelli, and R. Soncini-Sessa, "A dimensionality reduction approach for

many-objective Markov Decision Processes: Application to a water reservoir operation problem," Environ. Model. Softw., vol. 57, pp. 101–114, Jul. 2014, doi: 10.1016/j.envsoft.2014.02.011.
[163] S. Asante-Okyere, C. Shen, Y. Y. Ziggah, M. M. Rulegeya, and X. Zhu, "Principal component analysis (PCA) based hybrid models for the accurate estimation of reservoir water saturation," Comput. Geosci., vol. 145, p. 104555, Dec. 2020, doi: 10.1016/j.cageo.2020.104555.
[164] D. Piróg, J. Fidelus-Orzechowska, Ł. Wiejaczka, and A. Łajczak, "Hierarchy of factors affecting the social perception of dam reservoirs," Environ. Impact Assess. Rev., vol. 79, p. 106301, Nov. 2019, doi: 10.1016/j.eiar.2019.106301.
[165] D. L. D. D. Jardim, M. E. P. Maceira, and D. M. Falcao, "Stochastic streamflow model for hydroelectric systems using clustering techniques," in 2001 IEEE Porto Power Tech Proceedings (Cat. No.01EX502), IEEE, p. 6. doi: 10.1109/PTC.2001.964916.
[166] S. Mishra, C. Saravanan, V. K. Dwivedi, and J. P. Shukla, "Rainfall-Runoff Modeling using Clustering and Regression Analysis for the River Brahmaputra Basin," J. Geol. Soc. India, vol. 92, no. 3, pp. 305–312, Sep. 2018, doi: 10.1007/s12594-018-1012-9.
[167] D. Mehta, J. Dhabuwala, S. M. Yadav, V. Kumar, and H. M. Azamathulla, "Improving flood forecasting in Narmada river basin using hierarchical clustering and hydrological modelling," Results Eng., vol. 20, p. 101571, 2023.
[168] B. Beiranvand, T. Rajaee, and M. Komasi, "Spatiotemporal clustering of dam settlement monitoring using instrumentation data (case study: Eyvashan Earth Dam)," Results Eng., vol. 22, p. 102014, Jun. 2024, doi: 10.1016/j.rineng.2024.102014.
[169] B. Beiranvand, T. Rajaee, and M. Komasi, "Presenting the AI models in predicting the settlement of earth dams using the results of spatiotemporal clustering and k-means algorithm," Sci. Rep., vol. 14, no. 1, p. 10207, May 2024, doi: 10.1038/s41598-024-60944-4.
[170] Z. Feng, W. Niu, R. Zhang, S. Wang, and C. Cheng, "Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization," J. Hydrol., vol. 576, pp. 229–238, Sep. 2019.
[171] Z. M. Yaseen et al., "Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq," J. Hydrol., vol. 542, pp. 603–614, Nov. 2016, doi: 10.1016/j.jhydrol.2016.09.035.
[172] B. Yilmaz, E. Aras, S. Nacar, and M. Kankal, "Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models," Sci. Total Environ., vol. 639, pp. 826–840, 2018, doi: 10.1016/j.scitotenv.2018.05.153.
[173] A. Parsaie, H. M. Azamathulla, and A. H. Haghiabi, "Physical and numerical modeling of performance of detention dams," J. Hydrol., vol. 581, p. 121757, Feb. 2020, doi: 10.1016/j.jhydrol.2017.01.018.
[174] Z. Yin et al., "Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment," Stoch. Environ. Res. Risk Assess., 2018, doi: 10.1007/s00477-018-1585-2.
[175] L. Wen, K. Rogers, J. Ling, and N. Saintilan, "The impacts of river regulation and water diversion on the hydrological drought characteristics in the Lower Murrumbidgee River, Australia," J. Hydrol., vol. 405, no. 3–4, pp. 382–391, Aug. 2011, doi: 10.1016/j.jhydrol.2011.05.037.
[176] Z. M. Yaseen, O. Kisi, and V. Demir, "Enhancing Long-Term Streamflow Forecasting and Predicting using Periodicity Data Component: Application of Artificial Intelligence," Water Resour. Manag., vol. 30, no. 12, pp. 4125–4151, Sep. 2016, doi: 10.1007/s11269-016-1408-5.
[177] E. Macdonald et al., "Event and Catchment Controls of Heavy Tail Behavior of Floods," Water Resour. Res., vol. 58, no. 6, Jun. 2022, doi: 10.1029/2021WR031260.
[178] F. Modaresi, S. Araghinejad, and K. Ebrahimi, "A Comparative Assessment of Artificial Neural Network, Generalized Regression Neural Network, Least-Square Support Vector Regression, and K-Nearest Neighbor Regression for Monthly Streamflow Forecasting in Linear

and Nonlinear Conditions," Water Resour. Manag., vol. 32, no. 1, pp. 243–258, 2018, doi: 10.1007/s11269-017-1807-2.
[179] M. Azmi, F. Sarmadi, and F. Sarmadi, "Improving the Accuracy of K-Nearest Neighbour Method in Long Lead Hydrological Forecasting," Sci. Iran., vol. 23, no. 3, pp. 856–863, 2016, doi: 10.24200/sci.2016.2164.
[180] E. Fadaei Kermani, G. A. Barani, and M. Ghaeini-Hessaroeyeh, "Cavitation Damage Prediction on Dam Spillways using Fuzzy-KNN Modeling," J. Appl. Fluid Mech., vol. 11, no. 2, pp. 323–329, Mar. 2018, doi: 10.29252/jafm.11.02.28356.
[181] R. Rodríguez-Alarcón and S. Lozano, "SOM-Based Decision Support System for Reservoir Operation Management," J. Hydrol. Eng., vol. 22, no. 7, Jul. 2017, doi: 10.1061/(ASCE)HE.1943-5584.0001496.
[182] K. Gavahi, S. J. Mousavi, and K. Ponnambalam, "Comparison of two data-driven streamflow forecast approaches in an adaptive optimal reservoir operation model," Epic Ser. Eng., vol. 3, no. 1, pp. 755–763, 2018.
[183] J. Guo, J. Zhou, H. Qin, Q. Zou, and Q. Li, "Monthly streamflow forecasting based on improved support vector machine model," Expert Syst. Appl., vol. 38, no. 10, pp. 13073–13081, 2011, doi: 10.1016/j.eswa.2011.04.114.
[184] R. Maddu, I. Pradhan, E. Ahmadisharaf, S. K. Singh, and R. Shaik, "Short-range reservoir inflow forecasting using hydrological and large-scale atmospheric circulation information," J. Hydrol., vol. 612, p. 128153, Sep. 2022, doi: 10.1016/j.jhydrol.2022.128153.
[185] M.-J. Chang et al., "Outflow sediment concentration forecasting by integrating machine learning approaches and time series analysis in reservoir desilting operation," Stoch. Environ. Res. Risk Assess., vol. 34, no. 6, pp. 849–866, Jun. 2020, doi: 10.1007/s00477-020-01802-3.
[186] R. M. Adnan, R. Mostafa, O. Kisi, Z. M. Yaseen, S. Shahid, and M. Zounemat-Kermani, "Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization," Knowledge-Based Syst., p. 107379, 2021.
[187] M. Sapitang, W. M. Ridwan, K. Faizal Kushiar, A. Najah Ahmed, and A. El-Shafie, "Machine Learning Application in Reservoir Water Level Forecasting for Sustainable Hydropower Generation Strategy," Sustainability, vol. 12, no. 15, p. 6121, Jul. 2020, doi: 10.3390/su12156121.
[188] V. Lai, Y. F. Huang, C. H. Koo, A. Najah Ahmed, and A. El-Shafie, "Conceptual Sim-Heuristic optimization algorithm to evaluate the climate impact on reservoir operations," J. Hydrol., vol. 614, p. 128530, Nov. 2022, doi: 10.1016/j.jhydrol.2022.128530.
[189] J.-T. Kuo, Y.-Y. Wang, and W.-S. Lung, "A hybrid neural–genetic algorithm for reservoir water quality management," Water Res., vol. 40, no. 7, pp. 1367–1376, Apr. 2006, doi: 10.1016/j.watres.2006.01.046.
[190] I. Ahmadianfar, A. Samadi-Koucheksaraee, and S. Razavi, "Design of optimal operating rule curves for hydropower multi-reservoir systems by an influential optimization method," Renew. Energy, vol. 211, pp. 508–521, Jul. 2023, doi: 10.1016/j.renene.2023.04.113.
[191] J. Adeyemo and D. Stretch, "Review of hybrid evolutionary algorithms for optimizing a reservoir," South African J. Chem. Eng., vol. 25, pp. 22–31, Jun. 2018.
[192] L. F. R. Reis, F. T. Bessler, G. A. Walters, and D. Savic, "Water Supply Reservoir Operation by Combined Genetic Algorithm – Linear Programming (GA-LP) Approach," Water Resour. Manag., vol. 20, no. 2, pp. 227–255, Apr. 2006, doi: 10.1007/s11269-006-8049-z.
[193] C.-T. Cheng, W.-C. Wang, D.-M. Xu, and K. W. Chau, "Optimizing Hydropower Reservoir Operation Using Hybrid Genetic Algorithm and Chaos," Water Resour. Manag., vol. 22, no. 7, pp. 895–909, Jul. 2008, doi: 10.1007/s11269-007-9200-1.
[194] L. K. Karnatapu, S. P. Annavarapu, and U. V. Nanduri, "Multi-Objective Reservoir Operating Strategies by Genetic Algorithm and Nonlinear Programming (GA–NLP) Hybrid Approach," J. Inst. Eng. Ser. A, vol. 101, no. 1, pp. 105–115, Mar. 2020, doi: 10.1007/s40030-

019-00419-2.
[195] W. K. O. Ho, B.-S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," J. Prop. Res., vol. 38, no. 1, pp. 48–70, Jan. 2021.
[196] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," Geosci. Model Dev., vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.
[197] S. S. Abdullah, M. a. Malek, N. S. Abdullah, O. Kisi, and K. S. Yap, "Extreme Learning Machines: A new approach for prediction of reference evapotranspiration," J. Hydrol., vol. 527, pp. 184–195, 2015, doi: 10.1016/j.jhydrol.2015.04.073.
[198] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Clim. Res., vol. 30, no. 1, pp. 79–82, 2005, doi: 10.3354/cr030079.
[199] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," Geosci. Model Dev. Discuss., vol. 2022, pp. 1–10, 2022.
[200] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," PLoS One, vol. 18, no. 2, p. e0279774, Feb. 2023, doi: 10.1371/journal.pone.0279774.
[201] O. Renaud and M.-P. Victoria-Feser, "A robust coefficient of determination for regression," J. Stat. Plan. Inference, vol. 140, no. 7, pp. 1852–1862, Jul. 2010, doi: 10.1016/j.jspi.2010.01.008.
[202] H. Piepho, "An adjusted coefficient of determination (R 2) for generalized linear mixed models in one go," Biometrical J., vol. 65, no. 7, Oct. 2023, doi: 10.1002/bimj.202200290.
[203] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," Int. J. Forecast., vol. 32, no. 3, pp. 669–679, 2016.
[204] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," J. Oper. Res. Soc., vol. 66, no. 8, pp. 1352–1362, Aug. 2015, doi: 10.1057/jors.2014.103.
[205] A. Alqahtani, M. I. Shah, A. Aldrees, and M. F. Javed, "Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality," Sustainability, vol. 14, no. 3, p. 1183, 2022.
[206] H. Liu, Q. Li, D. Yu, and Y. Gu, "Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms," Appl. Sci., vol. 9, no. 19, p. 4069, Sep. 2019, doi: 10.3390/app9194069.
[207] A. Nouraki, M. Alavi, M. Golabi, and M. Albaji, "Prediction of water quality parameters using machine learning models: a case study of the Karun River, Iran," Environ. Sci. Pollut. Res., vol. 28, no. 40, pp. 57060–57072, Oct. 2021, doi: 10.1007/s11356-021-14560-8.
[208] I. Chatunapalak, W. Kongprawechnon, and J. Kudtongngam, "Long-Term Energy Demand Forecasting in Thailand with Ensemble Prediction Model," in 2022 17th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/iSAI-NLP56921.2022.9960242.