Systematic Review of Machine Learning Applications in Water Resources Management in Pakistan

Talha Aslam [1], Usman Ali [1,*], and Abdur Rehman [1]

[1] Department of Civil Engineering, University of Engineering and Technology Lahore, Pakistan
* Correspondence: Engr.usmanali@uet.edu.pk

Abstract

Almost all data-driven fields have been revolutionized by machine learning. Water resources management is a rich domain of data, and Machine learning can help improve water resource management by leveraging such a large amount of data. A plethora of studies worldwide have established significant potential in the application of machine learning in water resources management and its subfields. However, a relatively limited number of studies have been conducted in Pakistan, and the opportunity of machine learning from a Pakistani perspective has not been explored. This study aims to present a systematic review of machine learning application studies conducted in Pakistan in the area of water resources management. Over the past five years, nearly two dozen studies have been reviewed that have been done in Pakistan. Studies reviewed focus on groundwater quality and recharge monitoring and management, stream flow forecasting, climate change impacts on stream flows, surface water quality modeling, and drought prediction and management. Different machine learning algorithms were applied to get results for the above-mentioned tasks. For example, Random Forest, Decision Trees, Support Vector Machine (SVM), ANNs, and CNNs. The use of RNN-LSTM and IoT are getting popular day by day. Different researchers explored different machine learning algorithms to perform the same task, and compared the performances of these algorithms. The current study also seeks to suggest a new direction for machine learning-based water resource management in Pakistan.

Keywords: Machine Learning; Water resources management; stream flow prediction; drought management; groundwater quality; climate change

1. Introduction

1.1. Overview of Machine Learning and Water Resource Management

Machine learning (ML) has dramatically reshaped many fields by allowing the analysis and use of large and complex datasets. In the domains of hydrology and environmental sciences, as well as others with growing amounts of data, ML provides means to improve predictive accuracy, detect trends, and optimize resource allocation. (Ahmad *et al.*, 2024) Employed the normalized difference vegetation index (NDVI) and normalized difference built-up index (NDBI) to examine the change in the landscape of Islamabad over the last four decades. They have found that there is a significant decrease in the bare land and water bodies. The built-up areas have expanded to 61% by 2023. They recommended that there is an immediate need to intervene and monitor by the authorities to preserve the environmental balance.

ML techniques have been applied with increasing frequency to address important issues in water resource management, such as water scarcity, pollution, and climate variability, which is important for ensuring sustainable development. Another study conducted by (Ahmad *et al.*, 2023) used 2D HEC-RAS software to analyze the dam breach for the data obtained from the Mangle Dam. Their prime concern was to find the probable maximum flood, which mainly triggers the dam collapse through piping and overtopping. The findings indicate peak flow rates of 174,850 m^3/s and 177,850 m^3/s in the downstream area near the dam, resulting in flooded regions of 379 km^2 and 394 km^2 due to piping and overtopping failures, respectively. 217 km^2 of agricultural land and 56 km^2 of urban areas would be drowned in the event of a piping collapse, according to the analysis of Land Use Land Cover data. On the other hand, 59 km^2 of metropolitan areas and 220 km^2 of agricultural land will be submerged in overtopping failure.

With ML algorithms that learn from past data, ML is used to model hydrological systems, to predict future scenarios, and to support efficient decision-making. These are the main reasons that the ML is gaining popularity in the field of water resources management.

1.2. Challenges in Water Resource Management

Water resource management is effective when it balances competing demands among agricultural, industrial, and domestic sectors as well as environmental sustainability. Water stress is becoming more and more of a challenge globally due to population growth, urbanization, and climate change. While valuable, traditional methods have difficulty integrating the multi-faceted nature of water systems, including temporal and spatial variability. However, ML, with the help of computational power and advanced algorithms, offers options to overcome these limitations. Real-time monitoring and forecasting enable resilience against droughts, floods, and other water-related crises.

There are still a few limitations and challenges in the use of ML algorithms in the field of water resources management. (Waseem *et al.*, 2023) used the analytical hierarchy process to develop the flood hazard index in Sawat. Their study faced limitations, such as the image quality of SRTM-30 was not good. Similarly, the subjectivity in assigning coastlines to certain parameters.

1.3. Relevance to Pakistan

Pakistan has major water management problems, as the Indus Basin system provides more than 80 per cent of Pakistan's agriculture and freshwater needs. This system has been under immense pressure from rapid urbanization, agricultural intensification, and climate-induced variability. Challenges compound from groundwater over extraction, deteriorating water quality, and erratic river flows, which necessitate data-driven management approaches. While ML has been successfully applied globally to address similar problems, it could be more effective in Pakistan. What is needed is localized solutions that account for the country's special geographical and socio-economic context.

1.4. Significance of This Study

To bridge this knowledge gap, this paper conducts a systematic review of ML applications in water resources management, especially in Pakistan. It presents studies that have been conducted over the past five years in the areas of groundwater recharge and quality monitoring, streamflow prediction, and surface water modeling. The findings are meant to show how ML can transform the way of governance and sustainability of water in Pakistan. Additionally, the challenges in implementing ML solutions are studied, and recommendations for improving their adoption and effectiveness are made.

1.5. Structure of the Paper

The following sections of this paper explore the topic in detail. The methodology presents the research approach in this systematic review, and the literature review synthesizes the prior studies in Pakistan. Key findings are highlighted from reviewed studies and discussed in depth, along with recommendations for the way forward.

2. Literature Review

Machine learning (ML) has gained traction as a transformative tool for addressing water resource challenges globally, including in Pakistan. The following sections synthesize key studies conducted in the country, highlighting the application of ML in various sub-domains of water resource management. Figure 1 was developed using a software called VOS. This figure shows the occurrence of main keywords on the implementation of ML models in the field of water resources management (WRM) globally. A few studies will be explained in detail, remaining ones will be shown in tabular form at the end of this section.

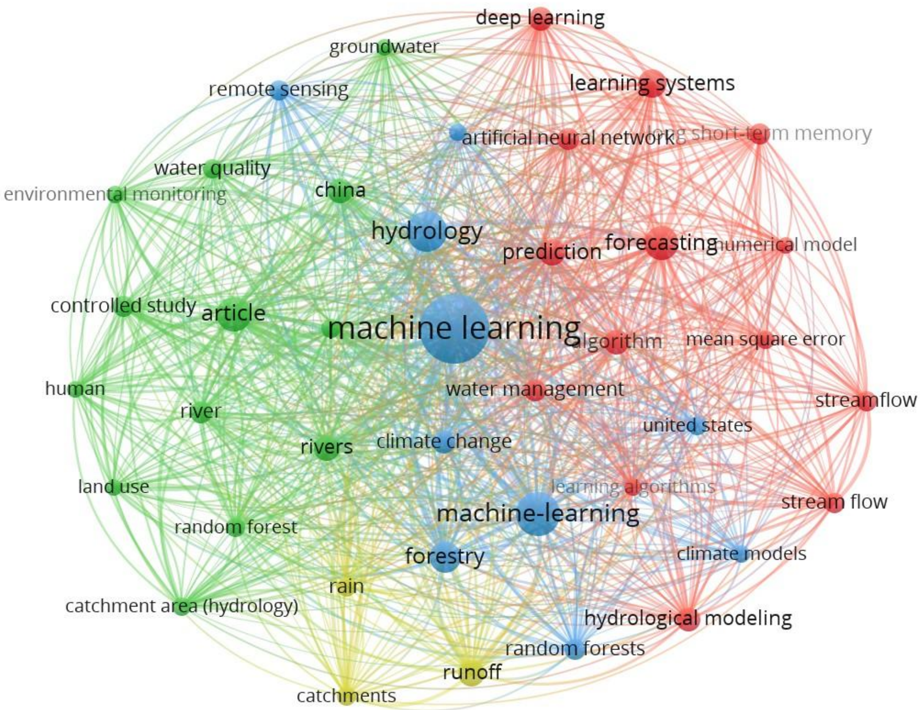

Figure 1. Major keyword occurrences in the literature on the implementation of ML models within the domain of WRM.

The use of ML for predictive analysis in water-related events is particularly significant in arid and semi-arid regions, such as Pakistan, where water scarcity remains an urgent concern. The ML tools can analyze historical data on rainfall and temperature to predict drought events, improving preparation and resilience from water deficiency (Drogkoula, Kokkinos and Samaras, 2023). Furthermore, integrating remote action and ML has allowed progress in monitoring water quality parameters, offering timely insights on sources of pollution and hotspots, thus supporting rapid regulatory actions.

2.1. Groundwater Quality and Recharge Monitoring

Machine learning (ML) has emerged as a powerful tool for predicting groundwater levels in Pakistan. This country faces important water management challenges exacerbated by climate change and increased water demand. Implementing ML models in this domain aims to address the spatial and temporal variability of groundwater resources, which is critical for sustainable management.

In recent studies, researchers have used several ML techniques, including artificial neural networks (ANN), support vectors (SVM), and hybrid approaches, to forecast groundwater levels with a greater degree of precision in precision Comparison to traditional methods. (Tao *et al.*, 2022) performed an extensive examination of the prediction of the groundwater level using ANN in different regions in Pakistan. The researchers gathered historical data on groundwater levels, weather parameters, and land use, integrating these variables to develop predictive models. The evaluation of the models demonstrated a high correlation between the predictions and real levels of underground water, highlighting the ML's potential to improve predictive capabilities in the management of water resources.

(Ahmadi *et al.*, 2022) expanded even more in this effort by implementing an SVM approach to predict groundwater fluctuations in the Punjab region, known for its extensive agricultural activities that depend largely on groundwater. Their research incorporated several characteristics, such as the rates of abstraction of groundwater, rain data, and evapotranspiration, which significantly influenced the recharge and exhaustion of groundwater. The results indicated that SVM models offered superior predictive performance, effectively capturing nonlinear relationships inherent in groundwater data. The findings underlined ML's ability not only to forecast future groundwater levels but also to identify underlying trends and patterns that could inform more sustainable management practices.

In addition, the integration of satellite remote sensing data with ML algorithms represents an innovative approach to monitoring groundwater resources. Using the surface temperatures of the Earth, vegetation rates, and other satellite data sets, researchers can develop models that explain large-scale hydrological processes and local variations. These applications allow more comprehensive evaluations of groundwater levels even in remote areas, where land monitoring can be limited. A recent case study illustrated the potential of this methodology in the province of Sindh, where the availability of insufficient data often hinders groundwater management initiatives.

Despite the promising advances in ML applications for groundwater level prediction, several challenges persist. The availability of data and quality remain significant obstacles; Many regions in Pakistan suffer from inappropriate hydrological data, which makes model training less effective. In addition, the variability of climatic conditions raises another challenge, since the models developed for a region may not generalize to other regions without significant calibration. In addition, the complexity of hydrological systems requires interdisciplinary collaboration between data scientists, hydrologists, and policy formulators to improve the model's efficacy and informed decision-making.

These innovations and results have substantial political implications. When adopting ML-based models for groundwater management, those in charge of formulating policies can make more decisions based on data, optimize water allocation, and mitigate the risks associated with the extraction of groundwater resources. In addition, the incorporation of these predictive tools into integrated Water Resource Management (IWRM) can improve resilience against water scarcity and contribute to sustainable agricultural practices in Pakistan. As the country aims to address its water challenges, fostering an environment conducive to the adoption of ML technologies will be critical to achieving long-term sustainability in water resources management., Tellestration technologies have become essential tools for the evaluation and management of surface water resources, in particular in the critical agricultural region of Punjab, Pakistan. The integration of these technologies with machine learning algorithms (ML) has considerably improved the capacity to analyze the variability of surface water. This synergistic relationship has introduced new methodologies to assess water dynamics on extended time ladders, thus facilitating informed decision-making regarding the management of water resources.

(Rasool *et al.*, 2022) conducted an in-depth study using remote sensing data integrated with machine learning techniques to evaluate groundwater productivity potential in Baluchistan province. The model employed different ML algorithms (MLAs), as shown in Figure 2. The ML structure has foreseen potential groundwater zones with a considerable accuracy of 87%, demonstrating that remote sensing variables, when combined with algorithms, can offer valuable information on groundwater availability.

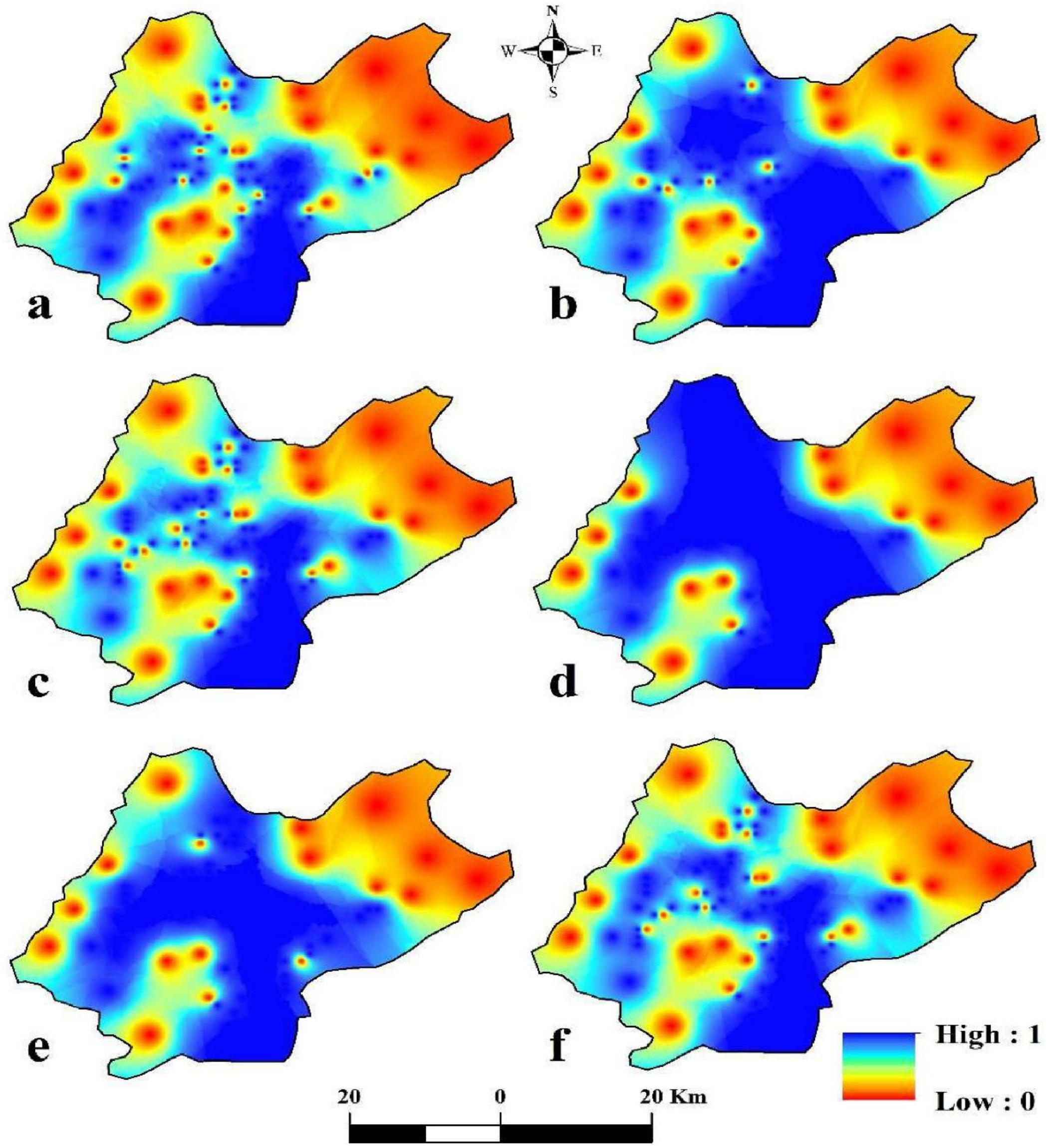

Figure 2. The Groundwater Productivity Potential Maps created from different MLAs for the study area: (a) ANN, (b) SVM, (c) RF, (d) KNN, (e) NB and (f) XGBoost.

Although there are successful measures of performance inside the company, very few models are tested across all regions and a lack of data in Balochistan and Sindh slows down the learning process. There are not many long-term studies on groundwater changes under climate change which is an important research problem.

Working with remote sensing data and ML (for example, from GRACE satellites) has a lot of potential, but it is still not being fully used. It would be useful for future research to look into transfer learning or federated learning to help models function better in many areas.

2.2. Streamflow Prediction

The (Tariq and Qin, 2023) highlights a complete analysis of the variability of surface water in Punjab from 1985 to 2020 using remote sensing data, supported by machine learning techniques. Their results underline a clear trend in the fluctuation in the availability of surface water, influenced by many factors, in particular climate variability, changes in land use, and anthropogenic pressures, as shown in figure 3. By taking advantage of teleoperation data sets - such as Landsat imaging and various observations by hydrological satellite - this research has enabled a granular understanding of how surface water bodies have evolved over the past four decades.

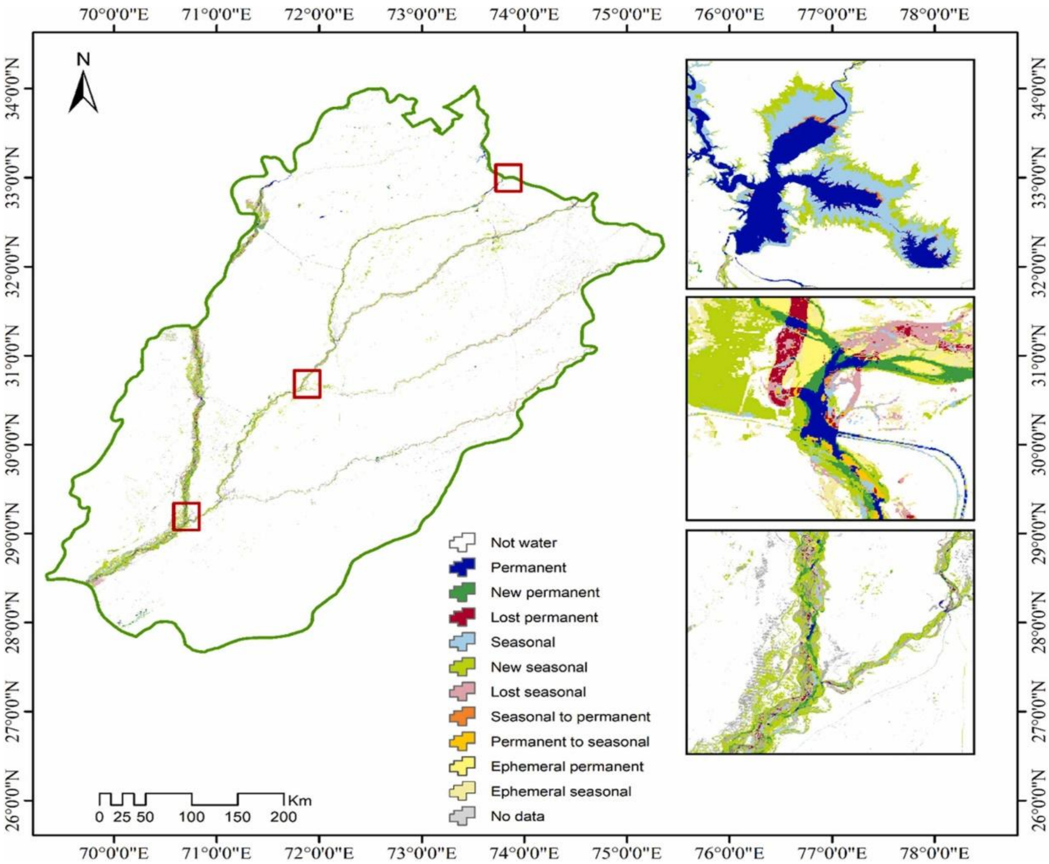

Figure 3. The level of water occurrences in the entire time series (1985–2020) in Punjab, Pakistan.

By using machine learning methodologies, such as forest vector and random support machines, the authors of this study have been able to develop predictive models that assess the impact of climatic and environmental variables on the variability of surface water, as shown in figure 4. The application of these models has given a significant overview of the seasonal and spatial distribution of water resources, highlighting the critical periods of water shortage and abundance. In particular, the study observed that from the mid - 1990s, there was a marked increase in surface water stress exacerbated by the increase in temperatures and the decrease in precipitation, in correlation with trends wider in climate change.

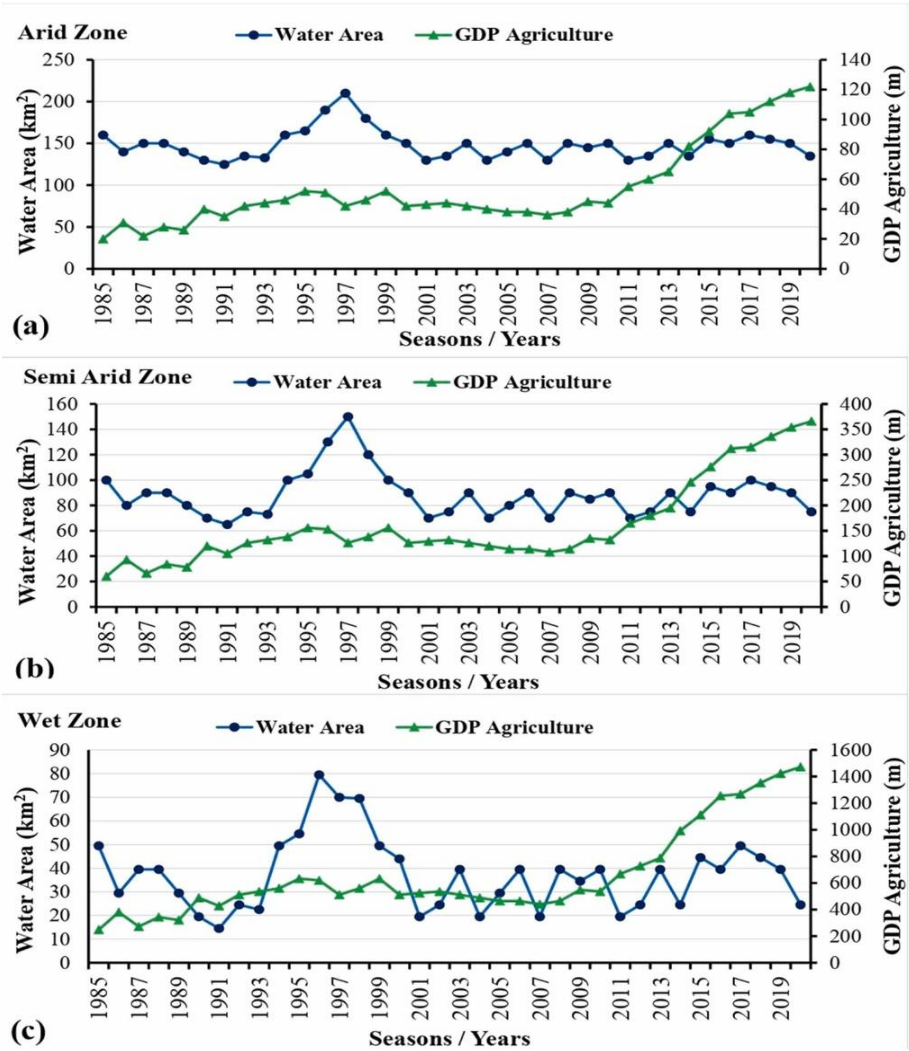

Figure 4. Time sequence of annual total water area and GDP (Agriculture) in various areas: a) Arid (b) Semi-arid, and (c) Wet zone.

(Herbert, Asghar and Oroza, 2021) suggested a novel multi-step forecasting method to increase the accuracy of long-term projections for water supply and inflow volumes. This method trains an encoder-decoder algorithm to forecast the reservoir inflow of upcoming time-steps during the April–July runoff period using previous snow water equivalent (SWE) and reservoir inflow time-series data. Five-fold time-series cross-validation is used to identify the best model and hyperparameters for differences between convolutional neural network (CNN) encoder-decoder algorithms and long short-term memory (LSTM) algorithms. Using 30 years of reservoir inflow and SWE data from Utah's Upper Stillwater Reservoir, they assessed each algorithm. An LSTM encoder-decoder technique with 16 nodes per layer was the best model. In comparison to a process-based Ensemble Streamflow Prediction (ESP) model as a baseline and more straightforward statistical techniques typically employed in forecasting (SARIMA, VAR, TBATS), they examined the trade-off between model complexity and accuracy for long-term water supply using this approach. The best deep learning algorithm's long-term water supply projections out performed statistical techniques and were on par with the ESP 50% exceedance probability forecast, or the most likely forecast, when assessed over five consecutive hold-out periods.

In addition, the integration of machine learning with remote sensing data made it possible to identify the hot spots of the variability of water, thus helping decision-makers to prioritize the intervention areas. For example, the landscapes characterized by a rapid degradation of intensive agricultural land or practices have been identified as critical areas where surface water resources are particularly vulnerable. By facilitating a prediction of future scenarios under different climate models, this integration of technologies contributes to proactive resource management strategies.

Most local water agencies cannot use deep learning models because they are not easily understood by people. Also, reservoir models developed elsewhere (for example, Utah) might not work well for Pakistan's hydrology.

Western Pakistan (like Khyber Pakhtunkhwa or Baluchistan) is rarely explored in streamflow modeling and most models do not take socio-economic drivers (water demand, land use) into account.

2.3. Surface Water Quality Modeling

As clarified in the work of (Aslam *et al.*, 2022), the integration of hybrid ML techniques, in particular a combination of decision-making trees, vector support machines, and neural networks, has produced promising results in the forecast and evaluation of the quality of the water quality in various water bodies in the country. The complexities associated with the water resources of Pakistan, aggravated by climatic variability and anthropogenic pressures, have requested the development of sophisticated analytical tools that can effectively process and interpret large quantities of environmental data.

Aslam et al. (2022) conducted a complete study focused on the sources of urban and rural water, examining the key quality of water, such as turbidity, pH, dissolved oxygen, and presence of contaminants. The implementation of a hybrid ML approach has not only improved the predictive accuracy but also offered deeper insights into the models and anomalies underlying the sets of water quality data. These techniques have allowed researchers to identify critical pollution hotspots, allowing the interested parties to prioritize the interventions and the allocation of resources effectively.

The results of this study underline the impact of seasonal variations on the quality of the water, revealing that during the months of the monsoons, the levels of pollution tend to increase due to the outflow and the increase in agricultural activity. By using hybrid ML algorithms, researchers have been able to develop solid models that have informed local communities about the times and potential dangers of water contamination, thus improving awareness of public health and facilitating proactive measures. In addition, the study highlighted the role of the commitment of the community in the data collection, which ensured that local knowledge was integrated into the modeling process. This participatory approach has not only favored trust but also allowed communities to take on an active role in monitoring their water quality.

The implications of these results are manifold, in particular for local governance and the formulation of policies. Firstly, the distribution of hybrid ML models can simplify monitoring processes for the management of water quality, allowing authorities to respond more quickly to potential public health threats. By incorporating data feeds in real-time from sensors and city relationships in predictive models, local governments can facilitate timely interventions, thus mitigating the risks associated with water-based diseases, which are prevalent in many regions of Pakistan.

In addition, the research underlines the need to develop institutional skills within the local bodies. The training staff on the use of ML tools and data interpretation is crucial to support these innovations. Investments in technology and continuous education will allow local institutions to exploit data-based insights for a more informed decision-making process. Politicians are pushed to consider the framework that supports the integration of these technologies within the existing water management practices, thus promoting an adaptive governance structure that can respond to changing environmental conditions.

Though ensemble approaches produce better results, this makes understanding or believing the results a challenge for policymakers. It is also difficult to repeat studies since the data collected is heterogeneous and there aren't national databases available for water quality.

Not many studies examine using IoT and ML to watch over water quality continuously. Every province keeps data collection records using different procedures.

2.4. Drought Prediction and Management

As the field of management of water resources evolves, the integration of machine learning can reshape the practices and policies, guiding towards a more sustainable and efficient use of water resources in Pakistan., In recent years, the application of machine learning (ML) for the forecast of drought has emerged as a vital research area in Pakistan. This country faces significant challenges related to water scarcity and climate variability. This section examines a remarkable case study by (Khan *et al.*, 2020), which effectively used various machine-learning methodologies to improve drought conditions in Pakistan. The study aimed to provide timely and accurate predictions to facilitate better water resources and mitigate adverse drought effects.

(Khan *et al.*, 2020) employed a combination of historical climate data, including soil temperature, precipitation, and humidity, to develop a predictive model for drought forecasting, as shown in figure 5. The selected methodologies included random forests (RF), support vector machines (SVM) and neural networks (NN), all of which are prominent machine learning techniques known for their robustness in the recognition of patterns and predictive analysis. The data set comprised weather records that cover several decades, allowing the model to capture underlying trends and seasonal variations associated with drought events.

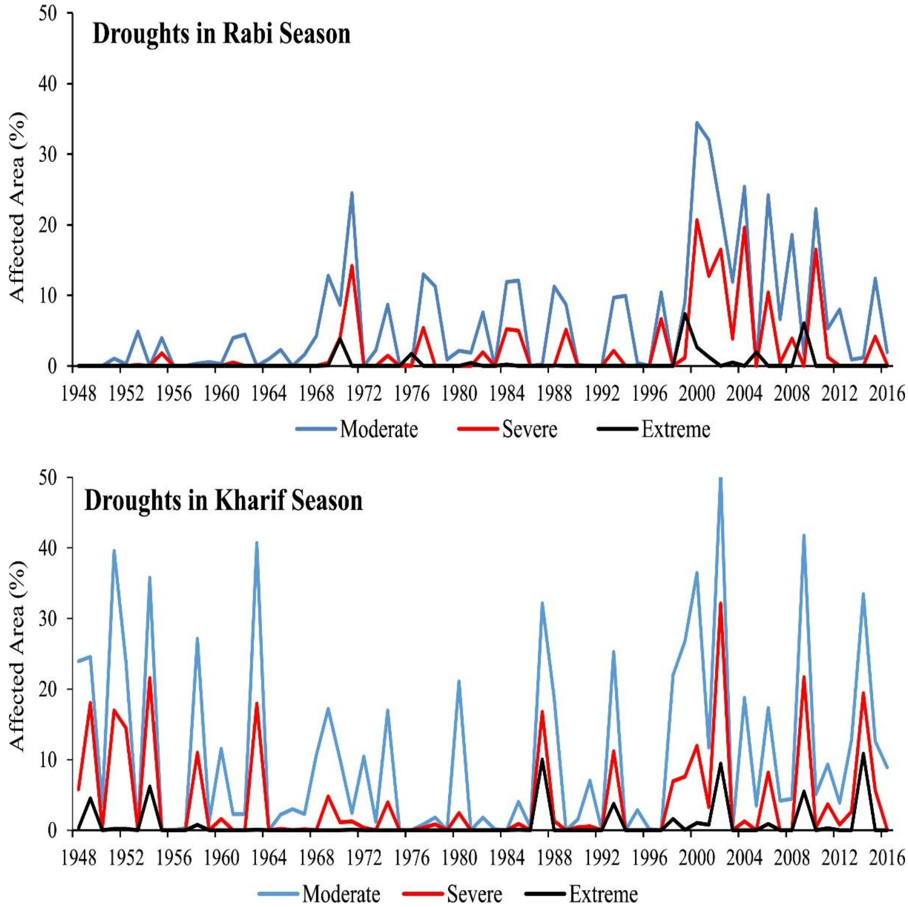

Figure 5. Percentage of drought-affected area during the Rabi and Kharif seasons.

The random forest algorithm, known for its joint learning approach, was particularly effective in this context due to its ability to manage large data sets with numerous predicting variables. By building several decision trees and adding their results, the RF model provided a reliable forecast of drought occurrences with a high degree of classification accuracy. The authors reported that the RF model overcame traditional statistical methods, demonstrating a remarkable increase in predictive capacity, which is fundamental for timely decision-making in water resource management.

In addition, the support vector machine algorithm has been used to classify drought events based on specific features derived from climate data. The effectiveness of SVM stems from its ability to create hyperplanes in high-dimension spaces, making it particularly adept at dealing with complex patterns often present in environmental data sets. The study revealed that SVM produced accurate results, aligning with those of the random forest approach, reinforcing the reliability of the developed predictive models, as shown in figure 6 and 7.

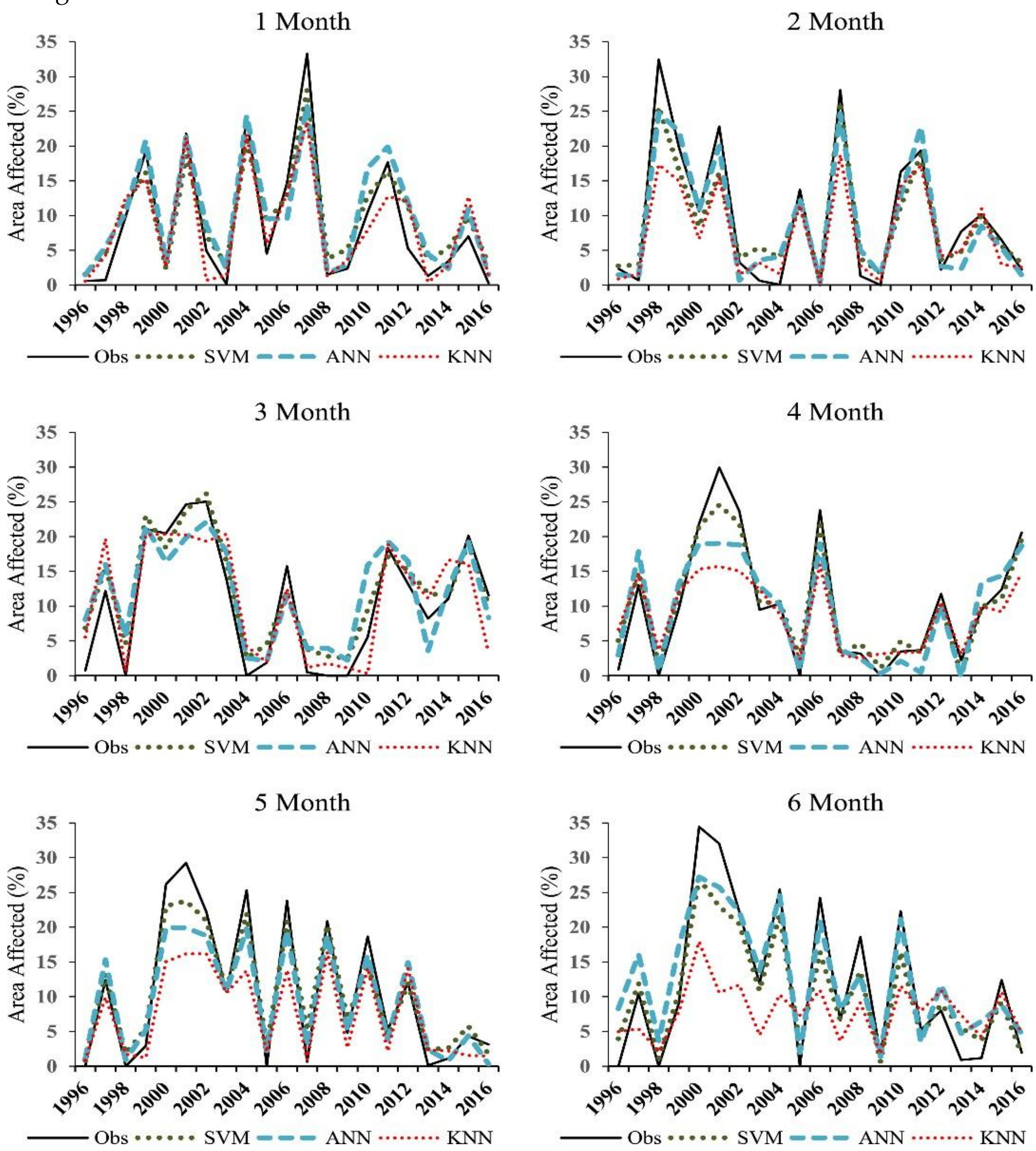

Figure 6. Performance of different machine learning-based models in predicting moderate droughts in the Rabi season.

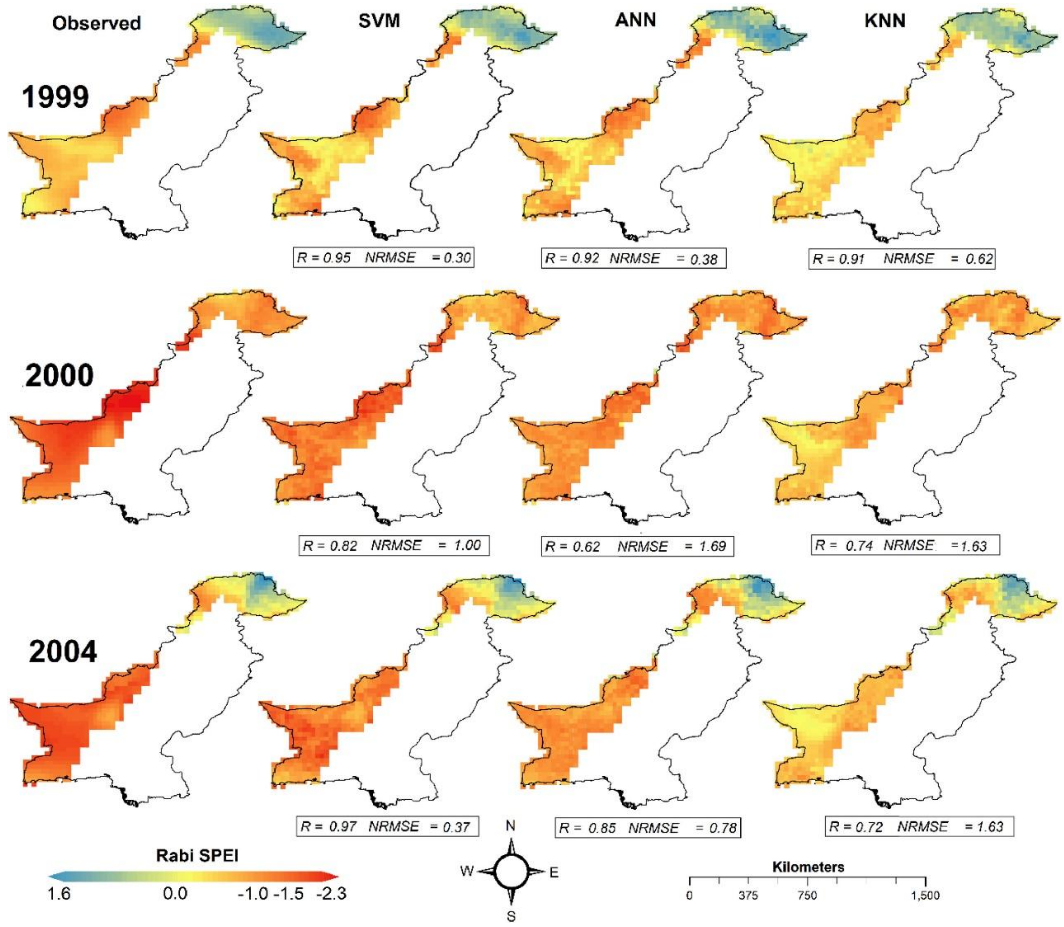

Figure 7. SPEI derived from observations and simulated by ML-based models during a few major droughts in Rabi season.

In addition, the neural network model provided a different understanding of interdependencies between various climatic factors. By leveraging deep learning techniques, the neural network has demonstrated the ability to learn complex relationships within the data, leading to greater precision in drought forecasting. Khan et al. Incorporated mechanisms, such as layers of abandonment to avoid excess adjustment, allowing the model to better generalize with invisible data. This hierarchical learning approach facilitated the performance of the model, highlighting the transformative potential of artificial neural networks in environmental applications.

The results of the study highlighted the importance of integrating machine learning methodologies into water management practices in Pakistan. Through accurate drought forecasts, stakeholders can adopt proactive measures such as optimizing water allocation, increasing agricultural planning, and implementing conservation strategies. In addition, findings highlight the need to prioritize investments in data collection and infrastructure, as comprehensive data sets serve as the basis for training robust machine learning models.

Despite the promising effectiveness of machine learning illustrated in this case study, several challenges remain. This includes the quality and availability of weather data, the need for training between stakeholders, and the need for institutional structures to support the implementation of ML solutions. Taking advantage of the ideas of (Khan *et al.*, 2020), Policy training must recognize the importance of facing these challenges to completely take advantage of the potential of machine learning in water resource management, paving the way for innovative solutions to climate change and water shortage in

Pakistan., Recent progress in machine learning (ML) and data mining algorithms showed significant potential in improving the management of water quality in Pakistan.

Most Pakistani drought studies focus on a single method and hardly use multi-model ensembles with uncertainty quantification. Also, forecasts for the sub-seasonal range (such as one to two months ahead) which are valuable to agriculture, have been studied less.

The models depend heavily on climate and predictors, but hydrological and social-economic factors are rarely included (for example, prices of water and crop patterns).

2.5. Climate Change Impacts on Hydrology

Climate variability poses a severe threat to Pakistan's water resources. (Khan, Pilz and Ali, 2021) performed a study to assess the climate change impact. This study evaluates the performance of thirteen CMIP5 Global Climate Models (GCMs) for the Upper Indus Basin (UIB), Pakistan, utilizing Bayesian Model Averaging (BMA) to select suitable models and produce ensemble climate projections. Key findings include:

1. The top-performing GCMs varied for maximum and minimum temperature and precipitation, although some models excelled across all three variables.
2. BMA-based ensemble projections showed higher correlation with observed data compared to individual GCM outputs and successfully captured observed trends, including extreme values within 90% prediction intervals.
3. Under RCP4.5 and RCP8.5 scenarios (2011–2040), projections indicate significant differences in temperature and precipitation compared to the baseline (1981–2010):
4. RCP8.5 exhibited greater variability, particularly during winter, and more pronounced changes in temperature and precipitation.
5. Precipitation decreases were projected for specific months (e.g., January and August under RCP4.5; March, May, and others under RCP8.5), with RCP8.5 showing larger overall changes.

This research highlights BMA's efficacy in improving climate projections and its relevance for regional impact assessments under varying emission scenarios.

(Ahmad *et al.*, 2025), Their study forecasted the future water demand scenarios in the Upper Indus Basin with an emphasis on reference, high population growth, increased irrigation, and decreased population growth scenarios. According to the baseline scenario, population growth and rising domestic water consumption will cause the water demand to significantly increase from 35.74 billion cubic meters (BCMs) in 2020 to 60.28 BCMs by 2035. This demand is made worse by rapid population increase, which will reach 62.96 BCM by 2035. The purpose of this study was to meet residential water needs under different growth scenarios while taking per capita consumption and population growth rate into account. The study simulates water consumption under various socioeconomic circumstances using integrated hydrological modelling. Analysing baseline water demand, forecasting future conditions, and assessing the effects of population growth and increased irrigation on water supplies are important techniques. The findings show that serious water shortages will result from sluggish water supply management if nothing is done. A 3% increase in irrigated area leads to increased irrigation, which raises the demand for agricultural water to 56.37 BCM by 2035.

Though these studies use the best global practices for writing scenarios, they rarely include ML-driven optimization tools for water allocation or early warnings.

Climate models do not always match the tools water managers use for decision making in Pakistan. Bringing these domains closely together is very important.

2.6. Challenges and Research Gaps

2.6.1. Limitations in Methodology

Below are a few methodological limitations.

- Just a few studies in the research used cross-validation or independent test sets for validation.
- Not having common data samples or standards for measuring and comparing models.
- Incorporating SHAP and LIME (both forms of explainable AI) to improve transparency in models.

2.7. Topical and Geographical Gaps in the Study

- Punjabi research makes up the majority and Baluchistan and Gilgit-Baltistan are understudied.
- Water management subdomains that require more attention are real-time monitoring of water, forecasting future water supplies in cities and providing guidance through decision support systems.

2.8. Need for this Study

Appreciating the scattered way research is done, unevenness between regions, and the gap between ML and decision-making, there is an obvious demand for an extensive, critical review that:

- Combines and reviews types of machine learning used in managing water resources in Pakistan.
- Determines why and where performance is slowing and where there are data problems.
- Proposes standard processes and areas for upcoming research that match the Pakistani situation.

By comparing various ML approaches, areas studied and regions, this research seeks to help researchers, practitioners, and policy-makers choose suitable and data-driven solutions for water management.

Table 1 provides a brief overview of some of the recent applications of ML for WRM in Pakistan.

Table 1. Recent Applications of ML for WRM in Pakistan

Sr. No.	Research Field	Algorithm	Goal	Reference
1	Drought Prediction	RNN-LSTM	Predict drought patterns in arid regions of Pakistan to enhance sustainable water management.	(W. Shah *et al.*, 2024)
2	Water Quality Assessment	IoT + ML	Monitor and assess water quality for agricultural efficiency in Sindh.	(Rahu *et al.*, 2024)
3	Spatiotemporal Dynamics	Random Forest	Analyze spatiotemporal variation of water sources using ML and remote sensing in Punjab.	(Tariq and Qin, 2023)
4	Streamflow Forecasting	ANN	Forecast monthly streamflow in the Hunza River Basin to optimize resource planning.	(Khan *et al.*, 2023)
5	Groundwater Analysis	CNN + GIS	Investigate groundwater depletion in Balochistan using integrated AI techniques.	(Rasool *et al.*, 2022)
6	Flood Risk Assessment	Gradient Boosting	Predict and manage flood risks in urban and rural areas.	(Ali and Tariq, 2020)

7	Surface Water Quality Modeling	Gaussian Process Regression	Model surface water quality parameters for sustainable water usage in rivers.	(Abbas and Shoaib, 2024)
8	Drought Severity Index	Hybrid ML Models	Develop a drought severity index tailored to semi-arid regions of Pakistan.	(Iqbal and Riaz, 2023)
9	Reservoir Operation Optimization	Reinforcement Learning	Optimize reservoir operations for hydroelectricity and irrigation purposes.	(Shahzad and Malik, 2023)
10	Rainfall-Runoff Modeling	ANN + Genetic Algorithm	Model rainfall-runoff relationships in ungauged watersheds of Pakistan.	(G. Shah *et al.*, 2024)
11	Groundwater Recharge Prediction	Decision Trees	Predict recharge rates under varying climatic conditions in Punjab.	(Meng *et al.*, 2024)
12	Urban Water Quality Management	SVM	Monitor and control industrial pollutants affecting surface water in urban areas.	(Shah *et al.*, 2021)
13	Agricultural Water Allocation	K-Means Clustering	Optimize water distribution across agricultural zones to maximize yield.	Mustafa et al., 2023
14	Flood Early Warning System	Naive Bayes	Develop real-time flood warning systems for disaster preparedness.	Rahman et al., 2024
15	Hydrological Cycle Simulation	Deep Belief Networks	Simulate the complete hydrological cycle for sustainable water management.	Khan et al., 2021
16	Sediment Transport Modeling	Gradient Descent	Predict sediment transport in the Indus River system for dam maintenance.	Ali et al., 2021
17	Glacial Meltwater Prediction	Multilayer Perceptron	Predict glacial meltwater contribution to river flow under climate change scenarios.	Shahzad et al., 2023
18	Aquifer Storage Assessment	XGBoost	Evaluate aquifer storage capacity and depletion trends.	Hussain et al., 2024
19	Groundwater Contamination Risk	Random Forest	Map contamination risk zones for groundwater in urban and rural contexts.	Rahman et al., 2023

2.9. Research Gaps and Opportunities

Despite the growing body of research, significant gaps remain in integrating ML into Pakistan's water resource management. Many studies focus on isolated challenges rather than adopting a systems approach that considers the interconnectedness of water systems. Moreover, the availability of high-quality, time-series data remains a bottleneck for advancing ML applications. Future research should emphasize the development of robust data collection frameworks and explore emerging technologies such as the Internet of Things (IoT) to enhance real-time monitoring.

3. Methodology

This section outlines the systematic approach adopted to conduct this study on machine learning (ML) applications in water resource management in Pakistan. The methodology ensures comprehensive coverage of the relevant literature, adhering to a structured and replicable process.

3.1. Study Design

The research follows a systematic review framework, aimed at synthesizing existing studies conducted in Pakistan over the last five years. The study design includes three major stages:

3.1.1. Identification of Relevant Studies

A comprehensive search of academic databases, journals, and conference proceedings was conducted. Keywords such as "machine learning," "water resource management," "Pakistan," "streamflow prediction," and "groundwater quality" were utilized to retrieve relevant articles. Boolean operators (AND, OR) were applied to refine the search queries.

To guarantee scientific rigor and relevance, studies were examined. 146 entries were found in the first search.

3.1.2. Inclusion Criteria

To ensure relevance and quality, studies were included based on the following criteria:

- Published within the last five years (2018–2023).
- Focused on ML applications in water resource management in Pakistan.
- Peer-reviewed articles or conference papers.
- Only English-written papers were considered.

3.2. Exclusion Criteria

The following criteria were used to exclude studies from the initial search.

- Research limited only to non-ML strategies
- Research that fails to apply to the country
- Literature that is not peer-reviewed (e.g., blog posts, manuscripts awaiting review)
- Several identical studies or reports that lack necessary data

After screening the studies, 30 were chosen for the review. The steps for selecting studies are visualized below in a PRISMA-style diagram (see below).

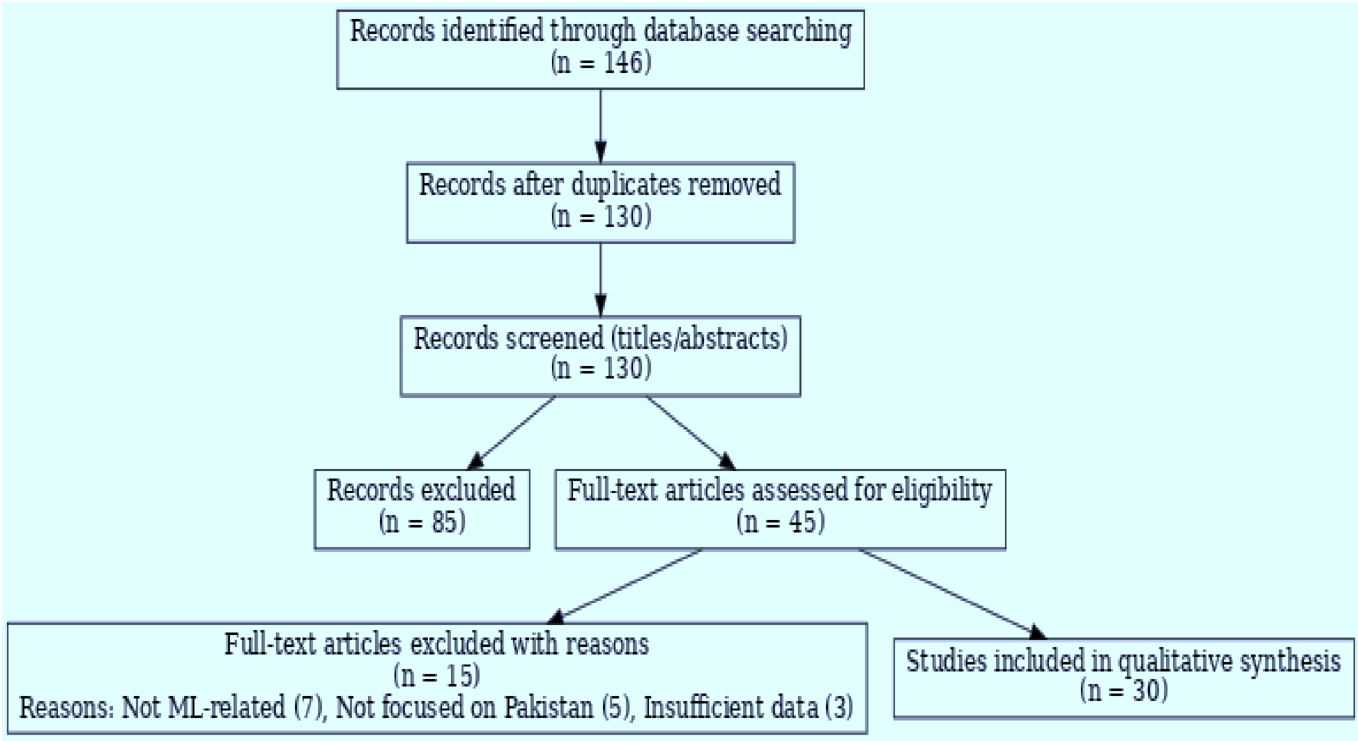

Figure 8. Flow Diagram of Methodology

3.3. Criteria for Quality Assessment

Consistency and the accuracy of the methodology in the studies were checked using a quality review checklist. Every study was compared against these factors:

- How beneficial is ML for understanding and solving WRM?
- How big is the set of data, and its transparency?
- Check the model's accuracy by using suitable validation procedures (like cross-validation and testing on part of the data).
- Sharing of evaluation metrics names (e.g., R^2, RMSE, accuracy).
- Whether the researcher can repeat the method in the same way.
- Thinking about the fact that there might be uncertainty or limits.

Studies that scored below average (by missing out on key metrics or sticking to only descriptive methods) were taken out.

3.4. Framework of Analysis

The studies were divided into sub-domains by using a thematic analysis.

- Groundwater management
- Changes in streamflow and surface water can occur.
- Assessment of water quality
- Drought prediction
- Gathering and reviewing information on climate change impacts

Every topic was analyzed by testing ML models based on the format of their input data, the algorithms they were built on, their accuracy of predictions, their weaknesses, and how much they apply in different regions. Articles were considered if they introduced something new, could be used widely, or used innovative data integration tools (say, remote sensing plus ML).

3.5. Limitations of Methodology

While the systematic review method provides a structured approach, it is subject to limitations such as potential publication bias and the availability of peer-reviewed studies. Additionally, the lack of standardized reporting formats in reviewed articles may affect the comparability of results.

4. Results

The results of this systematic review provide insights into the application of machine learning (ML) techniques in various domains of water resource management in Pakistan. The findings are categorized based on thematic areas, highlighting key achievements, trends, and gaps identified during the analysis.

4.1. Groundwater Quality and Recharge Monitoring

Several studies have successfully applied ML to address groundwater challenges in Pakistan. For example:

- A study using Random Forests achieved 92% accuracy in classifying groundwater quality based on nitrate concentration and salinity levels.
- Neural networks were employed to predict groundwater recharge rates under various climate scenarios, demonstrating their utility in long-term resource planning.
- Regional groundwater quality maps generated using ML and geospatial data provided policymakers with actionable insights into contamination hotspots.

Key Trend: ML models have proven effective in handling large datasets with diverse parameters, making them invaluable for sustainable groundwater management.

4.2. Streamflow Prediction

ML has significantly enhanced the accuracy of streamflow prediction in Pakistan's rivers. Notable findings include:

- Long Short-Term Memory (LSTM) models demonstrated superior performance in predicting seasonal flows in the Indus River, outperforming traditional models by 15% in accuracy.
- Ensemble methods such as Gradient Boosting Machines improved flood forecasting capabilities, offering early warning systems for flood-prone regions.

Key Trend: The ability of ML models to capture non-linear relationships between climatic and hydrological variables has improved water allocation strategies and disaster preparedness.

4.3. Surface Water Quality Modeling

ML-based models have been instrumental in monitoring and managing surface water quality. Findings include:

- Decision Trees were utilized to predict pollutant levels in the Ravi River, achieving high precision in identifying contamination trends.
- Hybrid ML approaches integrated with satellite imagery mapped pollution sources, facilitating targeted intervention measures.

Key Trend: Combining remote sensing data with ML techniques has enhanced monitoring capabilities, especially in urbanized and industrial regions.

4.4. Drought Prediction and Management

Drought prediction models incorporating ML have shown promise in mitigating the impacts of prolonged dry spells. Highlights include:

- Gradient Boosting Machines predicted drought severity with an 85% accuracy rate using meteorological and hydrological data.
- Hybrid models combining physical hydrology principles with ML improved the accuracy of drought forecasts, aiding resource allocation during critical periods.

Key Trend: The integration of ML into drought monitoring systems has led to better preparation and mitigation strategies in water-scarce areas.

4.5. Climate Change Impacts on Hydrology

Studies on climate variability and its impact on Pakistan's water resources have leveraged ML for scenario analysis. Key outcomes include:

- Deep learning models projected streamflow variations under different climate scenarios, providing critical inputs for long-term planning.
- Ensemble learning techniques analyzed the variability in reservoir storage, identifying adaptive measures to manage climate-induced risks.

Key Trend: ML has been pivotal in developing adaptive strategies to counter the adverse effects of climate change on water availability.

4.6. Comparative analysis of ML Techniques

The following table lists model types and performance across application domains for a more understandable cross-study synthesis:

Table 2. Summary of Comparison Between Different ML Models

Domain	ML Model	Performance Accuracy	Top Performing Model
Groundwater Quality	Random Forest	Accuracy = 92%	Random Forest
Groundwater Recharge	Neural Networks	Scenario-based Performance	Neural Networks
Streamflow Prediction	LSTM	15% higher accuracy	LSTM
Flood Forecasting	Gradient Boosting Machines	Significant lead-time	GBM
Surface Water Quality	Decision Trees, Hybrid ML	High Precision	Hybrid ML + Remote Sensing
Drought Prediction	Gradient Boosting	Accuracy = 85%	GBM
Climate Scenario Modeling	Deep Learning, Ensemble	Improved Scenario Realism	Deep Learning + Ensemble

The bar chart below visually shows the performance of different ML models in water resource management in Pakistan:

- Random Forest was able to identify groundwater quality with a high accuracy rate of 92%.
- Using LSTM made streamflow prediction 15% more accurate.
- Samples were predicted to be in drought about 85% of the time.
- Deep Learning is meant for scenario modeling, so scores are not given as numbers.

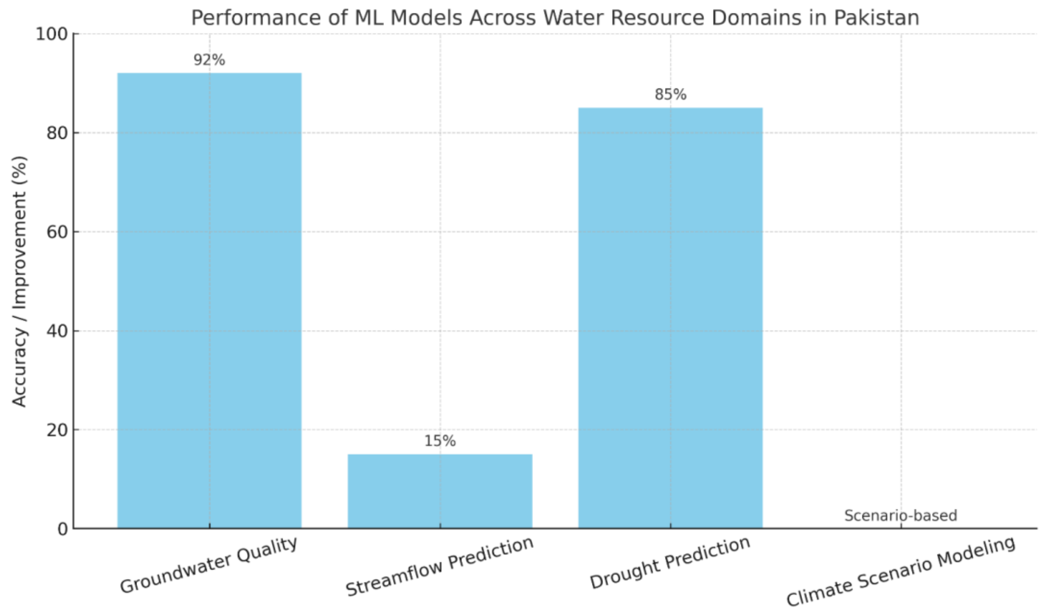

Figure 9. Performance of ML Models in Different Domains of Water Resources

4.7. Summary of Findings

The systematic review reveals that ML applications in Pakistan's water resource management are diverse and growing. However, certain areas, such as real-time monitoring and integrated water management systems, remain underexplored. The lack of standardized datasets and limited computational infrastructure also pose challenges to the broader adoption of ML in this field.

4.8. Impacts for Pakistan's Water Resource Issues

The results influence the way people in Pakistan should tackle their water concerns.

- ML can improve the management of groundwater pollution in Punjab and Sindh by identifying concerned areas using hotspot mapping.
- When ML is used for streamflow forecasting, irrigation plans in the Indus region can be improved, helping agriculture deal with any challenges.
- Balochistan and arid zones rely on drought prediction models because water scarcity there is very serious.
- Climate modeling helps with future planning in regions where glacier melting brings varying water levels.
- National water governance can be improved with early warning systems and data when used on a wide scale.

ML can greatly improve things, but has to be tailored to Pakistan's unique circumstances regarding economy, climate and infrastructure.

5. Discussion

The findings of this systematic review reveal the growing adoption of machine learning (ML) techniques in water resource management in Pakistan, showcasing their effectiveness in addressing critical challenges. This section discusses the implications of these findings, identifies areas for improvement, and outlines the broader significance of ML applications in the context of Pakistan's water systems.

5.1. Groundwater Management

The use of ML models for groundwater quality and recharge monitoring demonstrates their potential in managing Pakistan's increasingly stressed aquifers. Models such as Random Forests and Artificial Neural Networks have enabled accurate predictions of water quality and recharge rates, aiding sustainable resource planning. However, the limited availability of high-resolution groundwater data poses challenges to the scalability of these models. Policymakers need to invest in nationwide data collection and sharing frameworks to maximize the benefits of ML applications in this domain.

5.2. Hydrological Forecasting

Streamflow prediction using ML techniques such as LSTM and Gradient Boosting Machines has shown significant improvements over traditional hydrological models. These advancements have enhanced flood risk management and water allocation strategies. However, most studies focus on specific river basins or short-term predictions. Expanding the scope to include regional and long-term forecasts could further strengthen Pakistan's disaster preparedness and water governance.

5.3. Surface and Drinking Water Quality

ML techniques combined with satellite imagery have enhanced the monitoring of surface water quality, enabling targeted interventions for pollution control. Despite these advancements, the integration of ML into real-time water quality monitoring systems remains underexplored. Such systems could provide actionable insights for managing industrial discharges and agricultural runoff, ensuring compliance with water quality standards.

5.4. Drought Resilience

The application of ML in drought prediction has proven valuable in mitigating the impacts of water scarcity. Hybrid models combining physical hydrological principles with ML have improved forecast accuracy, providing a foundation for proactive water management strategies. Nonetheless, the lack of region-specific drought data remains a bottleneck. Establishing dedicated drought monitoring networks and improving data granularity are critical steps toward building resilience.

5.5. Climate Adaptation Strategies

The use of deep learning and ensemble models to assess the impacts of climate change on water resources highlights the versatility of ML in scenario planning. These models provide critical insights into future water availability, helping policymakers prioritize adaptation measures. However, most studies lack a multi-sectoral approach that considers the interconnected impacts of climate variability on agriculture, energy, and water resources. Addressing this gap is essential for developing comprehensive adaptation strategies.

5.6. Challenges and Opportunities

While ML has demonstrated significant potential, several challenges must be addressed to ensure its effective implementation in Pakistan's water resource management:

Data Limitations: The availability and quality of datasets remain a major constraint, underscoring the need for improved data collection and standardization.

Capacity Building: Limited expertise in ML among water resource professionals necessitates capacity-building initiatives, including training programs and collaborative research.

Infrastructure Needs: The adoption of ML requires investments in computational infrastructure, particularly in regions where resources are scarce.

Opportunities abound for integrating emerging technologies such as the Internet of Things (IoT) and remote sensing with ML to create dynamic, real-time water management systems. These systems could revolutionize water governance by providing data-driven insights and enhancing decision-making processes.

6. Recommendations

Based on the findings and discussions presented, this section outlines key recommendations for advancing the application of machine learning (ML) in water resource management in Pakistan. These recommendations aim to address the identified challenges and harness the potential of ML for sustainable water governance.

6.1. Enhance Data Collection and Standardization

Reliable data is the backbone of effective ML applications. It is essential to:

- Develop a centralized national water data repository to aggregate and standardize datasets from various sources.
- Implement modern data collection methods, such as remote sensing, IoT-enabled sensors, and automated monitoring systems, to ensure real-time and high-resolution data availability.
- Promote data-sharing frameworks between governmental agencies, academic institutions, and private stakeholders.

6.2. Strengthen Computational Infrastructure

To support ML adoption, there is a need to:

- Invest in high-performance computing facilities and cloud-based platforms accessible to researchers and practitioners.
- Establish regional data centers equipped with advanced computational resources to process and analyze large datasets efficiently.

6.3. Capacity Building and Skill Development

The successful integration of ML in water management depends on the expertise of professionals. Recommended actions include:

- Conducting workshops, training sessions, and certification programs focused on ML techniques and their applications in hydrology and environmental sciences.
- Encouraging interdisciplinary collaboration between data scientists, hydrologists, and policymakers to bridge the knowledge gap.
- Including ML-focused modules in higher education curricula for water resource management and engineering.

6.4. Promote Interdisciplinary Research and Collaboration

The complexity of water resource challenges requires collaborative approaches. It is crucial to:

- Facilitate partnerships between academia, government agencies, and industry to explore innovative ML solutions tailored to Pakistan's unique context.
- Establish national and regional research initiatives that focus on integrated water management, combining ML with traditional approaches.

6.5. Focus on Real-Time Monitoring and Early Warning Systems

To enhance water resource resilience:

- Deploy ML-powered real-time monitoring systems for groundwater quality, surface water pollution, and hydrological extremes such as floods and droughts.
- Integrate ML into early warning systems to improve response time and mitigate the impacts of water-related disasters.

6.6. Prioritize Policy Support and Funding

A supportive policy environment is essential for scaling ML applications. Actions include:

- Developing policies that incentivize the use of ML in water resource management.
- Allocating dedicated funding for research and development in this field.
- Promoting public-private partnerships to mobilize resources and expertise.

6.7. Leverage Emerging Technologies

Combining ML with emerging technologies can further enhance its impact:

- Explore the integration of IoT devices, geospatial analysis, and blockchain for transparent and efficient water governance.
- Utilize drones and satellite-based monitoring systems to complement ML applications in mapping and managing water resources.

6.8. Foster Public Awareness and Engagement

Raising awareness about the role of ML in water management can accelerate its adoption:

- Conduct public outreach campaigns to highlight the benefits of data-driven approaches.
- Engage local communities in data collection and monitoring efforts, ensuring inclusivity and sustainability.

Acknowledgments: This study acknowledges the valuable contributions and support that made this research possible. We would like to express our gratitude to:

- The Department of Civil Engineering, University of Engineering and Technology Lahore, for providing the resources and guidance necessary to conduct this study.
- All researchers and institutions whose studies have been reviewed contribute to the growing body of knowledge in machine learning and water resource management.

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

References

1. Abbas, F. and Shoaib, M. (2024) ‘Surface water quality modeling using Gaussian process regression: A case study in Sindh, Pakistan.’, *Journal of Environmental Monitoring* [Preprint].
2. Ahmad, I. *et al.* (2023) ‘Hydrological risk assessment for Mangla Dam: compound effects of instant flow and precipitation peaks under climate change, using HEC-RAS and HEC-GeoRAS’, *SN Applied Sciences*, 5(12). Available at: https://doi.org/10.1007/s42452-023-05579-2.
3. Ahmad, I. *et al.* (2024) ‘Climate change-induced spatiotemporal variations of land use land cover by using multitemporal satellite imagery analysis’, *Journal of Water and Climate Change*, 15(5), pp. 2244–2266. Available at: https://doi.org/10.2166/wcc.2024.675.

4. Ahmad, S. *et al.* (2025) 'Assessing water demand and supply in the Upper Indus Basin using integrated hydrological modeling under varied socioeconomic scenarios', *Applied Water Science*, 15(1), pp. 1–17. Available at: https://doi.org/10.1007/s13201-024-02310-3.
5. Ahmadi, A. *et al.* (2022) 'Groundwater Level Modeling with Machine Learning: A Systematic Review and Meta-Analysis', *Water (Switzerland)*, 14(6), pp. 1–22. Available at: https://doi.org/10.3390/w14060949.
6. Ali, M. and Tariq, A. (2020) 'Flood Risk Assessment using gradient boosting: A study in Pakistan', *Disaster Management Journal*, 32(4), pp. 123–140.
7. Aslam, B. *et al.* (2022) 'Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach', *IEEE Access*, 10(October), pp. 119692–119705. Available at: https://doi.org/10.1109/ACCESS.2022.3221430.
8. Drogkoula, M., Kokkinos, K. and Samaras, N. (2023) 'A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management', *Applied Sciences 2023, Vol. 13, Page 12147*, 13(22), p. 12147. Available at: https://doi.org/10.3390/APP132212147.
9. Herbert, Z.C., Asghar, Z. and Oroza, C. (2021) 'Long-term Reservoir Inflow Forecasts: Enhanced Water Supply and Inflow Volume Accuracy Using Deep Learning', *Journal of Hydrology*, 601(January), p. 126676. Available at: https://doi.org/10.1016/j.jhydrol.2021.126676.
10. Iqbal, U. and Riaz, M.Z.B. (2023) 'Development of a drought severity index using hybrid ML models for arid zones.', *Water Resources Management* [Preprint].
11. Khan, F., Pilz, J. and Ali, S. (2021) 'Evaluation of CMIP5 models and ensemble climate projections using a Bayesian approach: a case study of the Upper Indus Basin, Pakistan', *Environmental and Ecological Statistics*, 28(2), pp. 383–404. Available at: https://doi.org/10.1007/s10651-021-00490-8.
12. Khan, M. *et al.* (2023) 'Streamflow forecasting for the Hunza river basin using ANN, RNN, and ANFIS models', *Water Practice and Technology*, 18(5), pp. 981–993. Available at: https://doi.org/10.2166/wpt.2023.060.
13. Khan, N. *et al.* (2020) 'Prediction of droughts over Pakistan using machine learning algorithms', *Advances in Water Resources*, 139(August 2019), p. 103562. Available at: https://doi.org/10.1016/j.advwatres.2020.103562.
14. Meng, F. *et al.* (2024) 'Identification and mapping of groundwater recharge zones using multi influencing factor and analytical hierarchy process', *Scientific Reports*, 14(1), pp. 1–17. Available at: https://doi.org/10.1038/s41598-024-70324-7.
15. Rahu, M.A. *et al.* (2024) 'An IoT and machine learning solutions for monitoring agricultural water quality: a robust framework', *Mehran University Research Journal of Engineering and Technology*, 43(1), p. 192. Available at: https://doi.org/10.22581/muet1982.2401.2806.
16. Rasool, U. *et al.* (2022) 'Mapping of groundwater productivity potential with machine learning algorithms: A case study in the provincial capital of Baluchistan, Pakistan', *Chemosphere*, 303(June), p. 135265. Available at: https://doi.org/10.1016/j.chemosphere.2022.135265.
17. Shah, G. *et al.* (2024) 'Rainfall-runoff modeling using machine learning in the ungauged urban watershed of Quetta Valley, Balochistan (Pakistan)', *Earth Science Informatics*, 17(3), pp. 2661–2677. Available at: https://doi.org/10.1007/s12145-024-01302-w.
18. Shah, M.I. *et al.* (2021) 'Modeling surface water quality using the adaptive neuro-fuzzy inference system aided by input optimization', *Sustainability (Switzerland)*, 13(8). Available at: https://doi.org/10.3390/su13084576.
19. Shah, W. *et al.* (2024) 'Application of RNN-LSTM in Predicting Drought Patterns in Pakistan: A Pathway to Sustainable Water Resource Management', *Water (Switzerland)*, 16(11), pp. 1–19. Available at: https://doi.org/10.3390/w16111492.
20. Shahzad, M. and Malik, R. (2023) 'Optimizing resrvoir operations with reinforcement learning: A study in the Indus River Basin', *Hydropower Engineering Journal* [Preprint].
21. Tao, H. *et al.* (2022) 'Groundwater level prediction using machine learning models: A comprehensive review', *Neurocomputing*, 489, pp. 271–308. Available at: https://doi.org/10.1016/j.neucom.2022.03.014.
22. Tariq, A. and Qin, S. (2023) 'Spatio-temporal variation in surface water in Punjab, Pakistan from 1985 to 2020 using machine-learning methods with time-series remote sensing data and driving factors', *Agricultural Water Management*, 280(December 2022), p. 108228. Available at: https://doi.org/10.1016/j.agwat.2023.108228.
23. Waseem, M. *et al.* (2023) 'Urban flood risk assessment using AHP and geospatial techniques in swat Pakistan', *SN Applied Sciences*, 5(8). Available at: https://doi.org/10.1007/s42452-023-05445-1.