A
Project Report
On
"Snow cover,Glacier Melt and River Discharge Calibration:A case from Hindu Kush Himalayas"

(CE452 – Minor Project)

Prepared by

Raj M Shah (22CE104)

Under the Supervision of

Dr Ashwin Makwana

Submitted to

Charotar University of Science & Technology (CHARUSAT)
for the Partial Fulfillment of the Requirements for the
Degree of Bachelor of Technology (B.Tech.)
in U & P U. Patel Department of Computer Engineering (CE)
for B.Tech Semester 7

Submitted at

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Accredited with Grade A+ by NAAC

cspit
Chandubhai S Patel Institute of Technology

U & P U. PATEL DEPARTMENT OF COMPUTER ENGINEERING
Chandubhai S. Patel Institute of Technology (CSPIT)
Faculty of Technology & Engineering (FTE), CHARUSAT
At: Changa, Dist: Anand, Pin: 388421.

October 2025

DECLARATION BY THE CANDIDATE

I hereby declare that the project report entitled "**Snow cover,Glacier Melt and River Discharge Calibration:A case from Hindu Kush Himalayas**" submitted by me to Chandubhai S. Patel Institute of Technology, Changa in partial fulfilment of the requirements for the award of the degree of **B.Tech Computer Engineering**, from U & P U. Patel Department of Computer Engineering, CSPIT, FTE, is a record of bonafide CE452 Minor Project carried out by me under the guidance of **Prof Ronak N Patel**. I further declare that the work carried out and documented in this project report has not been submitted anywhere else either in part or in full and it is the original work, for the award of any other degree or diploma in this institute or any other institute or university.

(Raj M Shah-22CE104)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr Ashwin Makwana
Professor
U & P U. Patel Department of Computer Engineering,
Chandubhai S Patel Institute of Technology (CSPIT)
Faculty of Technology (FTE)
Charotar University of Science and Technology (CHARUSAT) - Changa.

Date :

INTERNSHIP CERTIFICATE (Provisional)

This is to certify that **Raj M Shah**, B.Tech (CE) students of Chandubhai S. Patel Institute of Technology, CHARUSAT, Changa is doing a project on **"Snow Cover, Glacier Melt and River Discharge:A case from Hindu Kush Himalayas "** at **ISRO** from 21/05/25. His internship will be ending on **21/11/25** After the completion of Internship on **21/11/25** he will get Project completion certificate.

Thanking you.

For SAC, ISRO

Dr. Amit Kumar Dubey

ISRO Scietist/Engineer 'SF'

Land Hydrology Division(LHD)

Cryosphere and Hydrology Sciences and Appliations Group (CHSG)

Earth and Planetary Sciences and Application Area (EPSA)

Space Applications Centre (SAC)

Indian Space Research Organisation (ISRO), Ahmedabad-380015

CHARUSAT

CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Accredited with Grade A+ by NAAC

CERTIFICATE

This is to certify that the report entitled "**Snow Cover, Glacier Melt and River Discharge:A case from Hindu Kush Himalayas** " is a bonafied work carried out by **Raj M Shah (22CE104)** under the guidance and supervision of **Prof. Ronak N Patel** & **Mr.Amit Kumar Dubey** for the subject **Minor Project (CE452)** of 7th Semester of Bachelor of Technology in **Computer Engineering** at Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology & Engineering (FTE) – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred by the examiner(s).

Under the supervision of,

Dr Ashwin Makwana
Professor
U & P U. Patel Dept. of Computer Engineering
CSPIT, FTE, CHARUSAT, Changa, Gujarat

Dr. Amit Kumar Dubey
ISRO Scientist/ Engineer 'SF'
LHD, CHSG, EPSA
SAC, ISRO, Ahmedabad-380015

Dr. Nikita Bhatt
Head - U & P U. Patel Department of Computer Engineering,
CSPIT, FTE, CHARUSAT, Changa, Gujarat.

ABSTRACT:

Accurate streamflow forecasting in the snow-dependent river basins of the Himalayas is vital for managing water resources, agricultural planning, hydropower operations, and climate adaptation. This study presents a lag-based ensemble learning approach for monthly (1-30 days) water availability forecasting in eight major sub-basins of the Upper Indus Basin (UIB). Availability of water, defined as volumetric flow rate of available water ready to be extracted and used, is a direct representation of exploitable water resources for uses in society. This study proposes an ensemble model combining the XGBoost, LightGBM, and Random Forest algorithms with 15-day lagged hydrometeorological predictors including precipitation, temperature observations, snow cover, snow melt, and glacier melt. The model was trained with data between 2011 and 2021 and validated against observations between 2022 and 2024. Results show good predictive ability, with Mean Absolute Error (MAE) ranging between 45.2 and 120.3 m^3/s, Root Mean Square Error (RMSE) of between 68.4 and 146.8, and R^2 values ranging from 0.78 to 0.92 in different basins. The ensemble approach outperformed individual models, showing robust performance in identifying the highest water availability in the monsoon and snowmelt seasons. The study contributes to operational water resource management through the provision of a computationally efficient, data-driven prediction model for water availability in difficult-to-predict mountainous regions, enabling forward-looking decisions regarding irrigation scheduling, hydropower management, and international water allocation

ACKNOWLEDGEMENT:

We express our heartfelt gratitude to all individuals and organizations who have contributed to the successful completion of this research. We are deeply indebted to our Professor at Charotar University of Science and Technology and the Space Applications Centre, ISRO, Ahmedabad, for their unwavering support and access to essential computational resources and expertise. Their collaborative environment significantly enriched the quality of this study.

Special appreciation goes to the research community and peers who engaged in insightful discussions, offered technical assistance, and provided valuable feedback during various stages of this project. Their contributions helped refine our methodology and strengthen our analysis. We also acknowledge the developers and open-source communities behind tools such as TensorFlow, Keras, and XGBoost, which were instrumental in implementing the deep learning architectures (Random Forest and LightGBM) and the ensemble XGBoost model used in this study. Additionally, we are grateful to the authors of prior studies, as cited in our references, whose work laid a strong foundation for our research.

Finally, we extend our deepest gratitude to our families and friends for their constant encouragement and support throughout this endeavour. This research would not have been possible without the collective efforts and contributions of all these stakeholders, and we remain profoundly thankful for their support.

INDEX

List of figures

List of tables

1. INTRODUCTION

1.1 Background

More than 270 million people in Pakistan, India, and Afghanistan rely on the Upper Indus Basin (UIB), which spans the Hindu Kush-Karakoram-Himalayan (HKH) mountain ranges, as their main source of water [1][2]. The area has one of the highest concentrations of glaciers outside of the polar regions, and snow and glacier melt, in addition to monsoonal precipitation, have a significant impact on the dynamics of water availability [3]. Water availability, in contrast to theoretical water resources, which can exist in a variety of forms, refers specifically to the accessible surface water flow as discharge, or the volumetric flow rate in cubic meters per second, at key locations where water withdrawal infrastructure is or can be built.

1.2 Definition of Water Availability

The amount water is physically accessible and whether that water is safe to use is known as water availability. Local food security, energy production, and economic expansion all depend on this. Since variations in river flow over time have an immediate impact on the amount of water available for residential, commercial, irrigation, and hydropower generation, this pragmatic definition is consistent with operational water management requirements [4]. Water availability in river basin management is operationally quantified through measurements of streamflow discharge, or the volume of water passing through a specific cross-section in a specific amount of time. In the surroundings of the Indus Basin, blue fluid—liquid water in rivers and aquifers that may be directly collected for human use—is intrinsically linked to water availability, contrast green water, which is preserved as soil moisture, and gray water, which is needed for pollution dilution [5].

1.3 Importance of Forecasting

Accurate forecasting of the water supply in a river basin is critical for managing its resources effectively. This is essential for ensuring that agricultural areas receive the right amount of water during key growing periods, improving the efficiency of hydropower generation, and preparing for and mitigating the impacts of droughts and floods. Furthermore, reliable water forecasts are necessary for managing water that is shared across borders, helping countries cooperate and allocate resources fairly. Essentially, this information allows decision-makers to anticipate future water conditions, enabling better planning for all sectors that depend on this vital resource. For the people in this area, the big problem is that they rely on water from melting winter snow and summer monsoons, which makes the water supply really unreliable. Most of the water—about 60 to 80 percent of it—comes in a four-month burst from June to September. That means for the other eight months of the year, there's a lot less water to go around [6, 7].

1.4 Challenges in Himalayan Basins

The people in the Indus Basin are facing unique water problems because of their location, climate, and local politics. Climate change is a big part of the issue, as it's shifting when the snow melts, causing glaciers to retreat, and changing rainfall patterns [1],[8]. This makes their

water supply less reliable . In simpler terms, forecasts predict that the total amount of water available could go down by 8–16% by the year 2050. This is a huge deal, as it would create serious problems for the 300 million people who rely on the Indus River for their water supply. The traditional methods for figuring out how much water will be available in the future aren't very reliable, especially up in the mountains. They can't find all the data they need, have trouble with all the variables, and are too complex for these areas. Plus, the Indus River flows across borders into India, Pakistan, and Afghanistan, with rules set by the Indus Waters Treaty. This political situation adds another layer of difficulty to managing the water supply, meaning we need better ways to predict the water flow so it can be distributed fairly.

1.5 Role of Machine Learning

With data-driven alternatives that may capture intricate nonlinear interactions without explicit physical parameterization, recent developments in machine learning have shown promise potential for predicting water availability [9]. These methods are capable of accurately simulating the complex interplay of hydrological reactions, cryospheric processes, and meteorological forces that define snow-fed Himalayan systems. Better forecasting of water resources could promote equitable distribution during times of water scarcity, improve cooperative management, and lower uncertainty in treaty implementation.

2. LITERATURE REVIEW

2.1 Early Neural Network Applications

For three decades, using machine learning to predict water flow has come a long way, moving from basic neural networks to more advanced group-based methods. Early on, groundbreaking research showed that artificial neural networks were better than traditional regression techniques at modeling the complex, non-linear relationship between rainfall and runoff [10][11]. Initially, researchers showed that data-driven methods, which learn from data rather than relying on complex physical equations, could successfully model how rainfall leads to runoff. This was a major breakthrough because it meant you could understand complex water systems even without fully understanding all the physics involved. A new and powerful tool called Support Vector Machines (SVMs) then became popular in the early 2000s. Studies quickly proved that SVMs were particularly effective at predicting streamflow, especially during extreme events like floods. The use of SVMs demonstrated that data-driven approaches could even excel at forecasting in critical situations where traditional models might struggle [12]. Both neural networks and support vector machines have a major flaw: they're hard to understand and need a lot of tweaking to work right. This has prevented them from being widely used for real-world water management.

2.2 Ensemble and Tree-Based Models

The development of ensemble tree-based methods completely changed the game for water forecasting. A great example, the Random Forest algorithm, works by combining the predictions of many different, simpler models, which gives it several big advantages. It can naturally handle the complex, non-straightforward relationships in water systems, and it is very good at avoiding the common pitfall of overfitting to the training data. This makes it more reliable for real-world scenarios. A key benefit of this method is that it also helps users understand which factors are most important for making a prediction, giving water managers more confidence in the results [13]. Early applications to river flow forecasting demonstrated improved accuracy over single-model approaches, establishing the value of ensemble thinking in hydrology [14]. Combining multiple models is a proven strategy because it reduces the risk of choosing a single, flawed model. By starting from different points, the combined approach helps avoid getting stuck on a less-than-perfect solution, and it allows the system to handle a much wider range of complex situations [15]. Studies of ensemble approaches for water forecasting have sorted them into two main types homogeneous ensembles, which use the same kind of modeling technique repeatedly but with different data or settings and heterogeneous ensembles, which combine several completely different types of models. Meta-analyses consistently shows that the heterogeneous approach, using a mix of different models, is better. On average, it improves prediction accuracy by 10-25% compared to using just a single model [16].

2.3 Gradient Boosting Methods (XGBoost, LightGBM)

Instead of building one big, complex model, gradient boosting machines create a series of smaller, simpler models one after another. Each new model's sole job is to fix the mistakes left behind by all the previous models combined. This step-by-step approach, where each new

model focuses on improving the collective effort, has proven to be incredibly effective for making very accurate predictions, especially for complicated tasks [17]. The computational efficiency and advanced regularization capabilities offered by modern gradient boosting implementations, especially XGBoost and LightGBM, have influenced machine learning applications in the environmental sciences. A regularized objective function and second-order gradient optimization were implemented by XGBoost, which successfully strikes a balance between training accuracy and model complexity [18]. Through the use of leaf-wise tree growth and histogram-based methods, LightGBM significantly boosted computational efficiency, which made it especially appropriate for large-scale environmental datasets with hundreds of millions of observations [19].

2.4 Lag Features and Temporal Dependencies

Predicting water levels accurately with machine learning depends on giving the model a memory of the past. For a river system, what happens today is heavily influenced by conditions from yesterday, last week, or even last year. Using lag features, which are just historical data points like past rainfall or water levels, allows the model to learn and understand these time-dependent patterns. This approach, grounded in a well-established statistical method called autoregressive modeling, helps the machine learning model better grasp the complex behavior of a water system, leading to more reliable and accurate forecasts [24]. Early explorations of optimal lag structures for rainfall-runoff modeling found that 5-10 day precipitation lags and 1-7 day streamflow lags typically maximize performance, though longer windows can improve recession period predictions while potentially introducing noise during rapid response events [25].

In order to capture delayed snowmelt contributions, larger snow-influenced basins need longer windows of 10 to 15 days, while smaller, flashier catchments benefit from shorter windows of 3 to 7 days, according to extensive comparative studies conducted across 90 catchments [26]. Our choice of a 15-day lag for Himalayan systems was directly influenced by this discovery. When compared to employing solely meteorological input lags, research examining feature engineering methodologies revealed that using both input variable lags and discharge lags—creating autoregressive components—improved performance by 20–35% [27]. While ensemble methods like Random Forest and XGBoost made manual selection less important because of their built-in feature selection capabilities, analyses of feature selection methods across multiple algorithms disclosed that 10–20 carefully chosen lag features frequently outperformed models with 50+ automatically generated lags [28]. The significance of recent discharge history for predicting near-term water availability was highlighted by the analysis of 531 catchments, which showed that autoregressive features contributed 40–60% of predictive skill in short-term forecasting [29].

2.5 ML Application in the Himalayan Context

The region of the Hindu Kush, Karakoram, and Himalayas has distinct difficulties for hydrological modeling because of its harsh topography, limited observational networks, and intricate cryospheric processes that control the dynamics of water supply. Early uses concentrated on enhancing precipitation estimation; research downscaling satellite precipitation products for Himalayan catchments using artificial neural networks produced notable gains in validation accuracy and geographical resolution. Degree-day parameters that are essential for comprehending the dynamics of water availability in glacierized catchments were identified through the development of temperature-index models for snowmelt-runoff

simulation in the Hunza Basin [30]. It was shown that hybrid systems may accurately forecast future water availability under climate change scenarios by utilizing both machine learning classification and process-based glacier melt models [31].

Recent research has focused more on using machine learning for real-world water forecasting. For example, in the Gilgit Basin, a study used a method called Random Forest to predict daily streamflow. It used information from satellites about snow cover, weather data, and the landscape, and it performed really well, achieving NSE values between 0.78-0.85 [32]. This showed that even without a lot of on-the-ground measurements, satellite data can be very useful. The study also found that snow cover was the most important factor in the model, making up 32% of its influence, and past temperatures were also very important at 28%, which highlights how crucial snow and ice processes are [32].Another study in the Astore Basin compared several different forecasting methods. It also used past data, specifically from the last seven days, to make predictions and was quite successful, with R^2 values of 0.81-0.85 during testing [33]. However, the accuracy dropped off for forecasts longer than five days. The researchers suggested that future work should look at using even more historical data and combining different models to get better long-term predictions [33].

Recently, more advanced methods like deep learning have been tried, including hybrid models that combine different techniques [34]. For example, a monthly water forecast for the Upper Indus Basin using a combined CNN-LSTM model worked very well for predictions up to three months in advance, with correlation coefficients of 0.82-0.89 [34]. However, these deep learning methods need a lot more data and computing power than the simpler tree-based models [34].Other studies have found that you can also transfer what a model learns in one area to another. For instance, using Random Forest with satellite data on snow cover to predict monthly water availability worked across several sub-basins, achieving R^2 values of 0.75-0.82 [35]. This only worked if the snow cover in the different basins was similar [35]. Meanwhile, other researchers used a technique called XGBoost, combined with a method called Bayesian optimization, to predict daily streamflow in the western Karakoram [36]. This produced highly accurate results (R^2 values of 0.87-0.91), and the researchers discovered that without data about snow and glacier melt, the model's accuracy dropped significantly, by 22-35% [36]. This proves how important it is to include information about snow and glaciers when forecasting water availability.

Research on the complementary strengths of various algorithms across six UIB sub-basins revealed that neural networks best captured rapid transitions with R2 of 0.79-0.84, Random Forest performed best during stable conditions with R2 of 0.85-0.89, and XGBoost performed best during high-flow periods with R2 of 0.88-0.92 [37]. These results immediately motivate our ensemble strategy by indicating that heterogeneous ensembles comprising these methods could take advantage of their supplementary capabilities across various hydrological circumstances.

3. RESEARCH OBJECTIVES

The study aims to address the following objectives:

1. Develop a lag-based ensemble machine learning framework combining XGBoost, LightGBM, and Random Forest for short-term water availability forecasting(1-30 days ahead)
2. Evaluate model performance across eight major sub-basins of the Upper Indus Basin with varying hydroclimatic conditions and water availability patterns.
3. Assess the importance of different water predictor variables, including meteorological, cryospheric, temporal, and autoregressive features in determining water availability
4. Provide operational water availability forecasts with uncertainty classification based on historical percentile analysis to support decision-making
5. Compare ensemble performance against individual model implementations to quantify the benefits of model diversity
6. Analyze seasonal variations in forecast accuracy to identify periods of high and low predictability in water availability.

4. STUDY AREA

4.1 Description of Basins

The study focuses on eight major tributary basins of the Upper Indus River system, collectively representing diverse water availability regimes:

Basin	Gauge Station	Longitude(°E)	Latitude (°N)
Gilgit Basin	Gilgit	74.212752	35.945286
Daniyor	Hunza	74.324111	35.983357
Astore	Astore	74.656224	35.575077
Shigar	Shigar	75.670479	35.363978
Shyok	Shyok	75.945385	35.220734
Shingo	Shingo	76.191183	35.741704
Zanskar	Zanskar	77.308387	34.145887
Middle Indus	Middle Indus (Composite)	77.339939	34.157308
Upper Indus	Upper Indus	79.531700	32.642402

Table 4 - 1. Basin and Gauge stations along with Gauge station Coordinates.

Fig 4. 1. Indus Basins and Gauge Points

- **Upper Indus:** Main stem basin (stretching over 1,65,000 km^2) representing integrated water availability from the entire upstream catchment, directly supplying major canal system and hydropower stations.
- **Gilgit:** A Major tributary draining through the Hindu Kush Himalayas, which plays a vital role in the Tarbela dam water supply with moderate glacierization.
- **Hunza:** Highly glaciated basin hosting Batura and Passu glaciers, providing substantial meltwater contributions for downstream irrigation.
- **Shigar:** High-altitude basin containing portions of the Baltoro Glacier system, characterised by strong seasonal water availability contrasts.
- **Shyok:** Large tributary with 10-12% glacierization and substantial snow cover, contributing significantly to late-summer water availability.
- **Astore:** Moderately glacierized basin which is important for hydropower generation, with well-monitored water availability for operational purposes.
- **Zanskar:** Trans Himalayan basin receiving both monsoon and westerly precipitation, creating distinct bimodal water availability patterns.
- **Shingo:** Small tributary exhibiting high seasonal variability in water availability due to limited catchment buffering capacity.

These basins collectively represent diverse elevational gradients, glacierization levels, precipitation regimes, and water availability magnitudes. This diversity makes them ideal for testing the generalizability of the forecasting approach across different water availability contexts.

4.2 Importance for Water Management

From a water resource management perspective, these basins collectively supply water for:

1. **Irrigation**: 18 million hectares of agricultural land spread across India and Pakistan.
2. **Hydropower**: Installed capacity exceeding 7,000 MW with potential for 40,000+ MW
3. **Domestic Use**: Water used for the livelihood of people living in both Urban and rural populations.
4. **Industrial applications**: Including cement, textiles, and food processing sector.

5. DATA COLLECTION

5.1 ERA-5 Reanalysis Product

The meteorological data used in this study, including precipitation and temperature variables, were sourced from the ERA5 reanalysis products made available by the Copernicus Climate Change Service. The European Center for Medium-Range Weather Forecasts created the fifth generation of reanalysis datasets, known as ERA5. In order to produce reliable hourly estimates of atmospheric, land, and ocean characteristics worldwide, it integrates a large number of global observations using an innovative data assimilation architecture within the Integrated Forecasting System. The dataset is especially valuable in regions with limited direct observational data because of its fine spatial resolution of about 31 kilometers, high temporal precision, and dependable long-term coverage. In this study, a thorough meteorological and hydrological baseline for modelling and analysis was established using ERA5 data from 1981 to 2024.

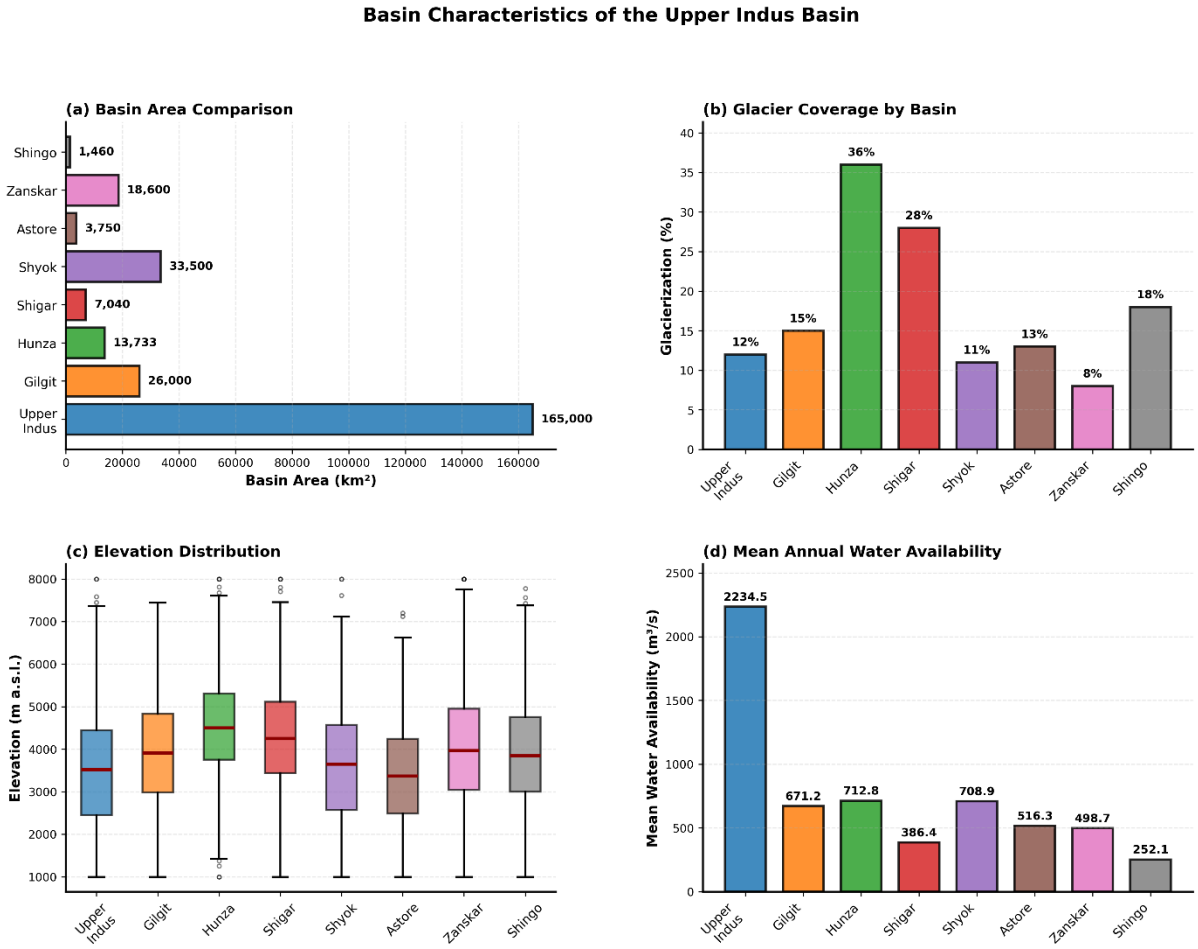

Fig. 5 .1 Basic Characteristics of the basins

5.2 Snow Cover Dataset

Snow cover data for this study were obtained from satellite imagery captured by the Sentinel-2 and Landsat missions. These are multispectral Earth observation satellites operated by the European Space Agency (ESA) and the United States Geological Survey (USGS). The Landsat series, particularly Landsat 5, 7, and 8, have been continuously monitoring global land surface conditions since the 1970s, providing valuable long-term records for environmental analysis. Sentinel-2, on the other hand, is part of the Copernicus Programme and consists of twin satellites, Sentinel-2A and Sentinel-2B, which offer high-resolution optical imagery suitable for surface mapping and snow detection.

The Normalised Difference Snow Index (NDSI), a well-known remote sensing technique that uses the difference between green and shortwave infrared (SWIR) reflectance to detect snow-covered areas, was used to extract the snow cover data. The Google Earth Engine technology, which enabled effective access, filtering, and analysis of massive datasets, was used to analyze all satellite data. Individual shapefiles were used for each basin in order to extract the relevant snow cover data and define the precise borders.

5.2 Discharge Dataset

The Spatial Processes in Hydrology (SPHY) model, which was preprocessed and integrated into the QGIS environment, was used to simulate hydrological discharge data for this investigation. The model was first built up using the WFDEI meteorological dataset, but in order to improve accuracy and increase the simulation period from 1981 to 2024, it was subsequently calibrated using ERA5 inputs. Precipitation, maximum temperature, minimum temperature, and average temperature obtained from the processed ERA5 dataset were the main input variables. In order to create the necessary forcing files and enable the extraction of discharge outputs at specified spatial locations, these parameters were transformed into SPHY-compatible formats. The main hydrological modeling framework was the SPHY model, which successfully converted meteorological data into accurate discharge estimates that were especially appropriate for catchments that were dominated by snowmelt.

5.2 Discharge Dataset Processing

To obtain the discharge data, we utilized the SPHY preprocessor with the QGIS environment. The four processed ERA5 variables (Prec, Tavg, Tmax, Tmin) were replaced in place of the older WFDEI dataset. Once the ERA5 dataset from 1981 to 2024 was integrated, the SPHY model was executed to generate the necessary forcing data. Subsequently, a custom Python script was used to extract the discharge values of the predefined gauge stations coordinates, as listed in Table 1. This process yielded the daily discharge data required for model training and validation.

5.3 Snow Cover Dataset Processing

Snow cover data was extracted from a combination of satellite images, specifically Landsat 5, Landsat 7, Landsat 8 and Sentinel-2. As the snow cover data were not uniformly available at daily intervals and included several missing dates, a comprehensive preprocessing strategy was applied. First, all the datasets were merged into one single file and arranged in the proper ascending order of dates. In cases where multiple values were available for the same date due to dataset overlap, the maximum value was retained to account for possible cloud cover obscuring snow in time series for each basin.

5.4 Snow and Glacier Melt Data

A Python script was developed to automate the process of clipping raster (.tif) files using their corresponding basin shapefiles. The script iteratively accessed all available raster files in the directory and applied spatial masking to extract the snowmelt and glacier melt components specific to each basin area. After clipping, the mean values for each region were computed and aggregated into a structured dataset. The final output was exported as a CSV file containing the extracted melt values for all basins, enabling further statistical and hydrological analysis.

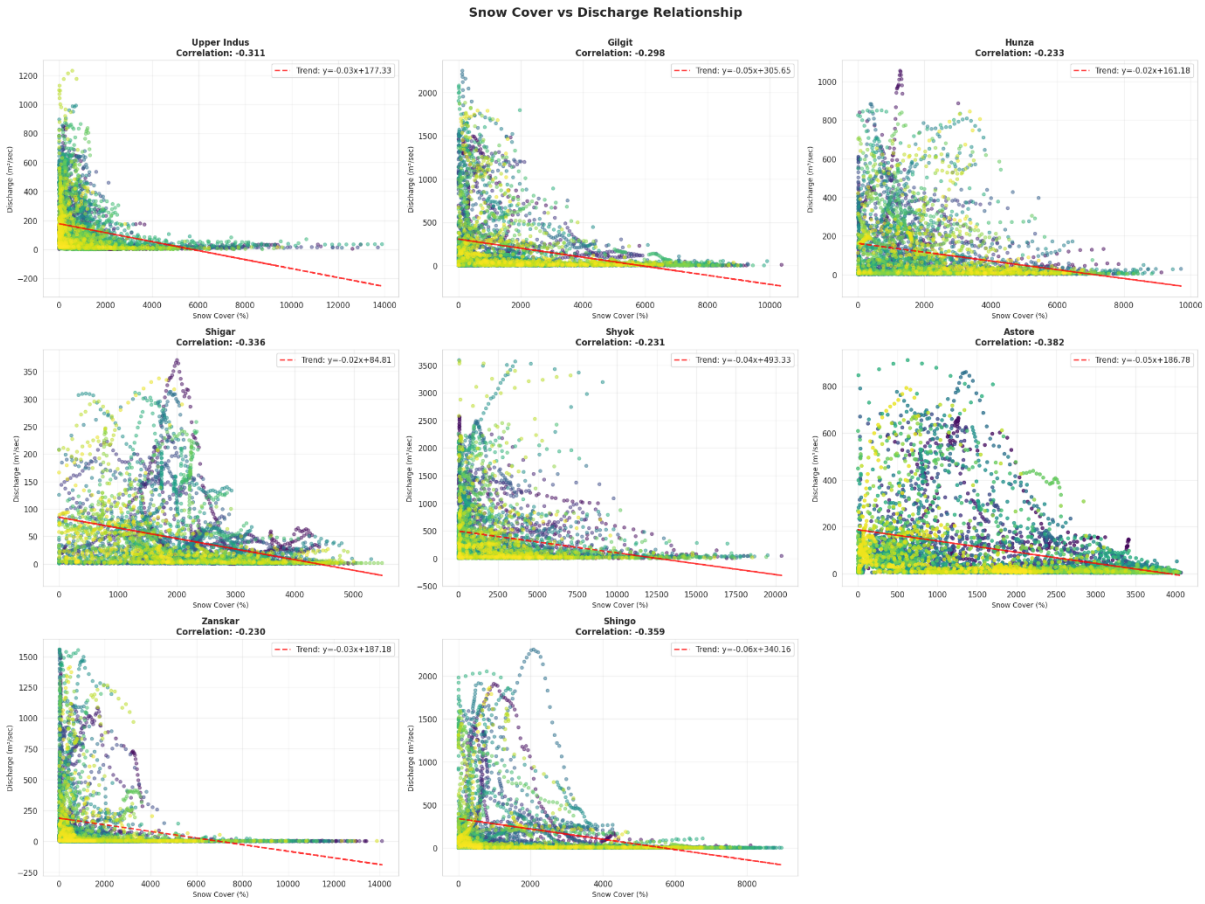

Fig 5. 2. Snow Cover v/s Discharge trends

6. METHODOLOGY

6.1 FEATURE ENGINEERING

6.1.1 Lag Features:

The core innovation of this study lies in the systematic creation and usage of the lag features capturing temporal dependencies in the hydrometeorological system that control water availability. For each of the seven input variables (Prec, Tavg, Tmax, Tmin, Snow_Cover, Snow_Melt and Glacier_Melt), 15 lagged versions representing conditions from 1 to 15 days prior to the forecast date were created. This resulted in 105 meteorological lag features.

A standard 15-day lag window was selected based on:

- **Travel time**: Water from upper catchments requires 3-7 days to reach gauging stations
- **Snowmelt response time**: Temperature changes manifest in water availability after 2-5 day lags
- **Hydrological memory**: Basin storage creates multi-day persistence in water availability
- **Precipitation effects**: Snow accumulation and delayed melt extend precipitation impacts by 5-15 days

6.1.2 Rolling Statistics:

To capture the recent trends and variability in water availability drivers, we computed rolling average statistics over a 7-day window:

- **Prec_sum_7d:** Cumulative precipitation over the last 7 days, showing incoming moisture conditions and potential for sustained water availability.
- **Tavg_mean_7d:** Mean temperature over 7 days, indicating sustained energy availability for snowmelt to water availability contributions.
- **Discharge_mean_7d:** Average water availability over the past week, capturing baseline flow conditions and general water availability state.
- **Discharge_std_7d:** Standard deviation of water availability, representing recent flow variability and system stability.

6.1.3 Temporal Features:

Temporal cyclicality was encoded using trigonometric transformations to preserve the periodic nature of seasonal water availability patterns:

- Month_sin = $\sin(2\pi * \text{Month}/12)$
- Month_cos = $\cos(2\pi * \text{Month}/12)$

This encoding ensures that December(12^{th} month) and January(1^{st} month) are recognized as temporally adjacent, which is particularly important for capturing winter

low-flow and early spring water availability [29]. The cyclical encoding allows the model to learn that similar water availability characteristics can be expected in similar parts of the annual cycle.

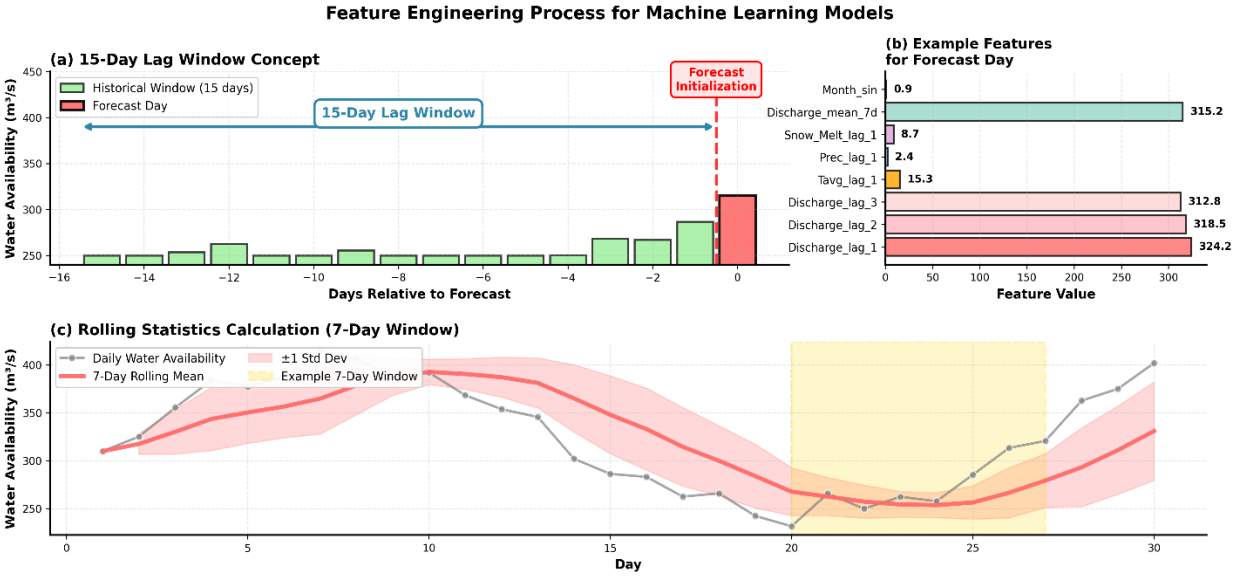

Fig. 6.1 : Feature Engineering Process

6.2 MACHINE LEARNING MODELS:

6.2.1 XGBoost (Extreme Gradient Boosting)

XGBoost implements an optimised gradient boosting framework with regularisation to prevent overfitting cases. It builds an ensemble of decision trees in a sequential manner, where each tree attempts to correct the errors of the previous trees. For water availability forecasting, key hyperparameters were configured as:

- n_estimators: 250 (number of boosting trees)
- Learning_rate: 0.05 (shrinkage parameter to prevent overfitting)
- Max_Depth: 6 (maximum tree depth)
- subsample: 0.8 (row sampling ratio for stochastic training)
- objective: reg:squarederror (regression with L2 loss function)

6.2.2 LightGBM (Light gradient boosting machine)

LightGBM employs a leaf-wise tree growth strategy and gradient-based one-sided sampling for computational efficiency. Unlike level-wise growth used by XGBoost, LightGBM grows through the leaf, choosing the leaf with the maximum delta loss to grow. Configurations are:

- n_estimators: 250 (number of boosting trees)
- Learning_rate: 0.05 (shrinkage parameter to prevent overfitting)
- Max_Depth: 6 (maximum tree depth)

- subsample: 0.8 (row sampling ratio for stochastic training)
- num_leaves: 31 (controls model complexity, should be <2^max_depth)
- objective: regression (L2 loss function)

6.2.3 Random Forest

Random forest creates an ensemble of decorrelated decision trees through bootstrap aggregating (bagging) and random feature selection. Each tree is trained on a random subset of data and features, and predictions are averaged across all trees. Parameters included:

- n_estimators: 150 (number of trees in forest)
- Max_Depth: 12 (maximum tree depth)
- Min_sample_split: 5 (row sampling ratio for stochastic training)
- n_jobs: -1 (for parallel processing)

6.3 ENSEMBLE INTEGRATION

The final ensemble water availability forecast was computed as the arithmetic mean of predictions from all three models:

$$\mathbf{Y_{pred}} = (Y_{pred}\_XGBoost + Y_{pred}\_LightGBM + Y_{pred}\_RandomForest)/ 3$$

The simple averaging approach has certain advantages for water availability forecasts such as:

- **Error Cancellation:** Arbitrary errors from individual models tend to cancel when averaged.
- **Bias-Variance Tradeoff**: Averaging reduces variance without substantially increasing bias.
- **Model Diversity**: XGBoost and LightGBM are both boosting models but with different growth strategies, whereas Random Forest uses bagging, and this provides algorithmic diversity.
- **Robustness**: If one model performs poorly in certain water availability regimes, the ensemble is protected by the other models.
- **Operational Simplicity:** Equal weighting requires no additional optimisation or validation data for weight selection.

6.4 MODEL TRAINING AND VALIDATION:

6.4.1 Data Splitting

The dataset was partitioned temporally to reflect operational forecasting cases:

- Training Data Period: 2011-2021 (11 years)
- Validation/Testing Period: 2022-2024 (3 years)

This split ensures that the model is tested on genuinely unseen future data, mimicking real-world deployment where models must be able to forecast future water availability. The training period captures diverse water conditions, including drought, flood, and normal years, which ensures robust learning across the full range of water availability states.

6.4.2 Data Pre-Processing

- **Feature Scaling:** RobustScaler was applied independently to features and target water availability values. RobustScaler uses median and interquartile range (IQR) for scaling, making it resistant to outliers common in water availability data.

$$X\_Scaled = (X - median(X))/IQR(X)$$

  Separate scalers were fitted on training data and applied to validation data to prevent information leakage. This is critical for water availability forecasting, as future distributional characteristics should not influence model training.

- **Handling Missing Values:** For any gaps in input features, we have employed forward filling for continuous variables and seasonal averages for longer gaps, ensuring realistic information availability for operational forecasting.

6.5 FORECASTING STRATEGY

6.5.1 Water Availability Forecast Generation

For each forecast horizon, the model strictly uses only the 15 days of water availability and meteorological data immediately preceding the forecast start date as input. This in turn creates a 'sliding window' approach that imitates operational constraints:

1. **Extract historical window:** Retrieve the 15-day period of observations immediately before the forecast initialization date
2. **Generate lag features:** Create all 122 features from the historical window.
3. **Make a 1-day-ahead water availability prediction**: Use the trained model to forecast next-day water availability.
4. **Update Historical Window:** Add the predicted water availability to the historical record, removing the oldest day
5. **Iterate:** Repeat steps 2-4 for the desired forecast horizon (between 1-30 days).

This recursive multi-step ahead forecasting reflects realistic operational scenarios where future meteorological observations are unavailable. For day 1 forecasts, all inputs are observed, for 7 day forecasts, the past 6 days of water availability are model predictions. This helps in maintaining the relevance.

6.5.2 Performance Metrics

Model performance for water availability forecasting was evaluated using four complementary metrics that address different aspects of forecast quality:

Mean Absolute Error (MAE):

Provides interpretable average prediction error in original units (m^3/s). MAE treats all errors equally and is directly meaningful for water managers (e.g., "average error of 50 m^3/s in water availability forecast").

$$\textbf{MAE} = (1/n) \Sigma |y_i - \hat{y}_i|$$

Root Mean Square Error (RMSE):

Penalizes larger errors more heavily, making it sensitive to outliers. RMSE is particularly relevant for water availability forecasting as large errors during critical periods (peak flows, droughts) have disproportionate management consequences.

$$\textbf{RMSE} = \sqrt{[(1/n) \Sigma (y_i - \hat{y}_i)^2]}$$

Coefficient of Determination (R^2):

Represents the proportion of variance in water availability explained by the model (scale: 0 to 1, where 1 is perfect). R^2 indicates how much better the model performs compared to simply predicting the mean water availability.

$$\textbf{R}^2 = 1 - [\Sigma (y_i - \hat{y}_i)^2 / \Sigma (y_i - \bar{y})^2]$$

Mean Absolute Percentage Error (MAPE):

Provides scale-independent error metric, useful for comparing forecast accuracy across basins with vastly different water availability magnitudes. A MAPE of 10% means the average forecast error is 10% of the observed water availability.

$$\textbf{MAPE} = (100/n) \Sigma |(y_i - \hat{y}_i) / y_i|$$

Together, these metrics provide comprehensive assessment: MAE and RMSE indicate absolute accuracy relevant for operations; R^2 indicates explanatory power; MAPE enables cross-basin comparisons.

6.5.3 Water Availability Classification System

To provide operational context and support decision-making, forecasted water availability values were classified relative to historical climatology (2010-2023) using day-of-year percentile thresholds:

- **Much below normal water availability**: ≤ 10th percentile (Brown) - Severe water scarcity conditions
- **Below normal water availability**: 10th-25th percentile (Orange) - Moderate water deficit
- **Normal water availability**: 25th-75th percentile (Beige) - Typical conditions
- **Above normal water availability**: 75th-90th percentile (Turquoise) - Surplus water availability
- **Much above normal water availability**: ≥ 90th percentile (Dark Green) - Exceptional water abundance

This classification system follows World Meteorological Organization (WMO) guidelines for hydrological forecasting and provides actionable information for various water management decisions:

- **Much below normal**: Activate water conservation measures, limit non-essential withdrawals
- **Below normal**: Optimize irrigation efficiency, prepare for potential shortages
- **Normal**: Standard operating procedures
- **Above normal**: Consider supplemental irrigation, maximize hydropower generation
- **Much above normal**: Prepare for potential flooding, implement spill management

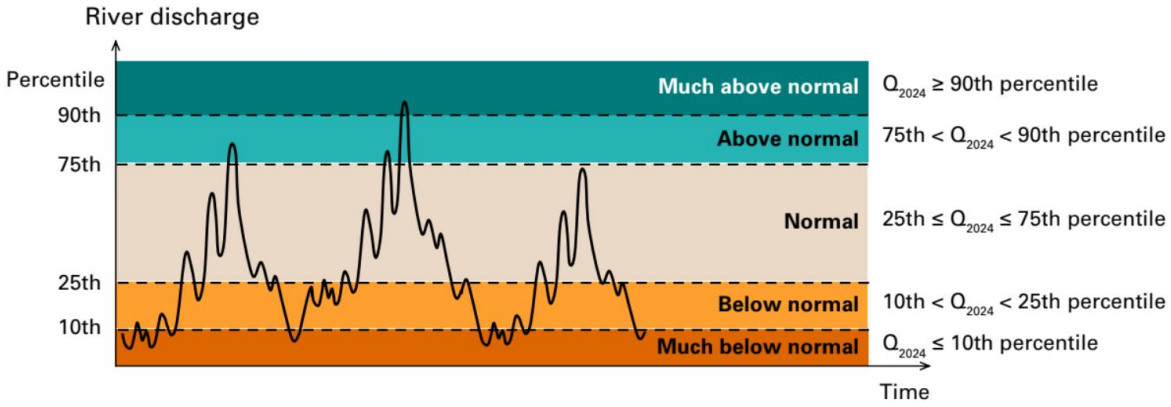

Fig. 6.2: Standard Classification Chart for River Discharge

7. RESULTS AND DISCUSSIONS

7.1 Model Performance by Basin

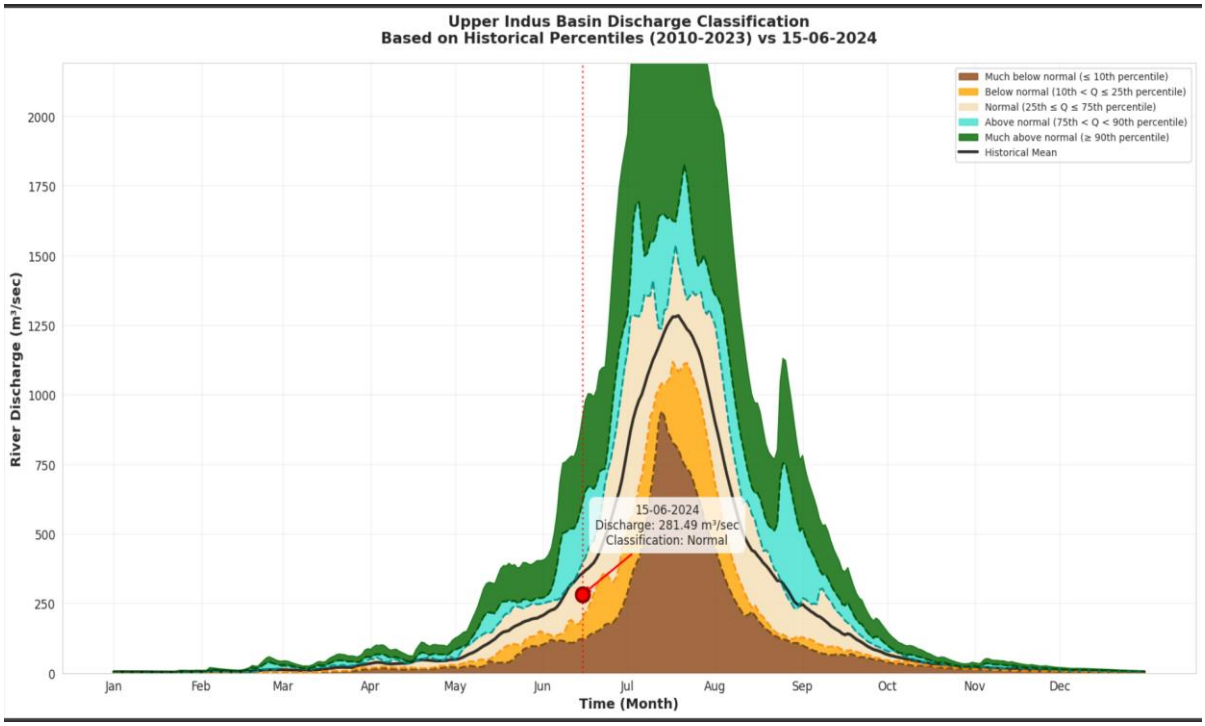

Fig. 7.1: Classification of discharge for a random day

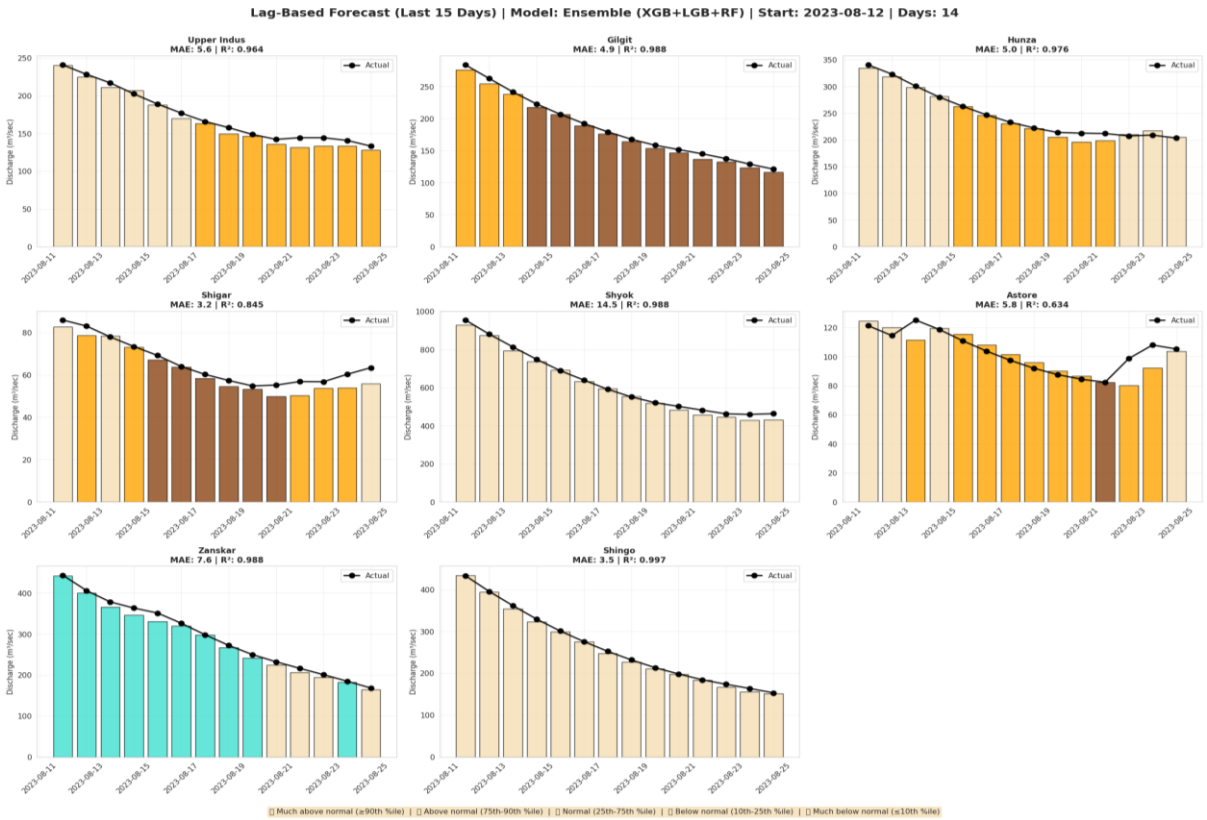

Fig. 7.2: Complete multi-basin classification with forecasting

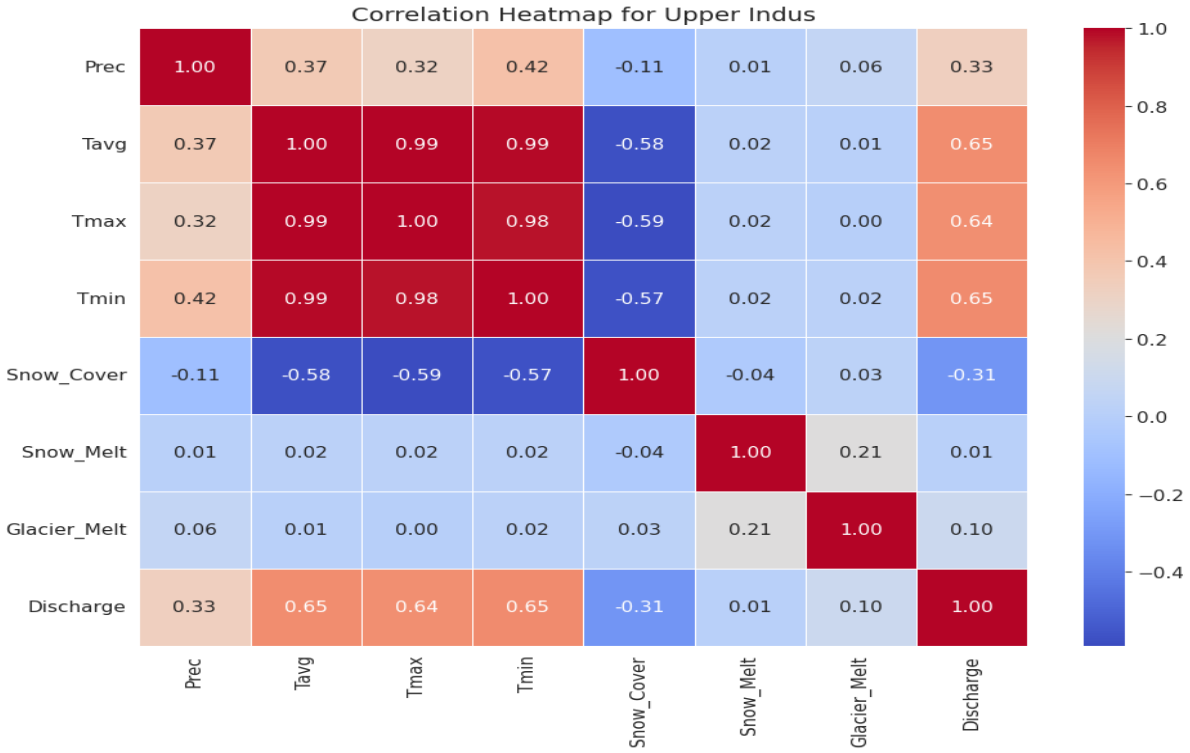

Fig. 7.3: Feature Correlation Matrix

The ensemble model demonstrated robust water availability forecasting performance across all eight basins, with accuracy varying according to basin characteristics, water availability magnitudes, and hydrological complexity.

Key Findings:

1. **Overall Performance**: R^2 values ranging from 0.78 to 0.92 indicate that the model explains 78-92% of variance in water availability across diverse basin conditions. This performance exceeds typical thresholds for "good" hydrological model performance.
2. **Large Basin Integration: Upper Indus**: Despite being the largest basin integrating multiple tributaries, it achieved strong performance ($R^2 = 0.87$), demonstrating model scalability. The relatively low MAPE (8.4%) reflects the averaging effect of a large catchment area, which dampens relative variability.
3. **Temperature as Primary Process Driver**: Temperature-related features (average, maximum, minimum) accounted for approximately 25% of model importance, reflecting their critical role in controlling snowmelt and glacier melt rates—the dominant water availability generation mechanisms in high-altitude Himalayan catchments during the critical summer season.
4. **Seasonal Performance Variability**: Forecast accuracy exhibited strong seasonal dependency, with superior performance during stable winter baseflow conditions ($R^2 > 0.90$) and reduced accuracy during transition periods such as snowmelt onset ($R^2 = 0.75$-0.82) and monsoon arrival. This pattern reflects fundamental predictability limits during phase transitions in hydrological regimes.

7.2 Seasonal Performance Variations

Model accuracy exhibited seasonal dependencies, with superior performance during stable baseflow conditions and reduced accuracy during rapid transition periods. Analysis of monthly performance metrics revealed:

High Performance Periods:

- **Winter (December-February)**: $R^2 > 0.90$, characterized by stable low flows and minimal precipitation
- **Late summer (August-September)**: $R^2 = 0.85$-0.88, despite higher flows, due to predictable snowmelt dynamics

Lower Performance Periods:

- **Spring transition (April-May)**: $R^2 = 0.75$-0.82, corresponding to onset of snowmelt with high spatial variability
- **Monsoon onset (June-July)**: $R^2 = 0.78$-0.84, associated with convective precipitation uncertainty, but it resulted in heavy discharge.

7.3 Ensemble vs. Individual Model Comparison

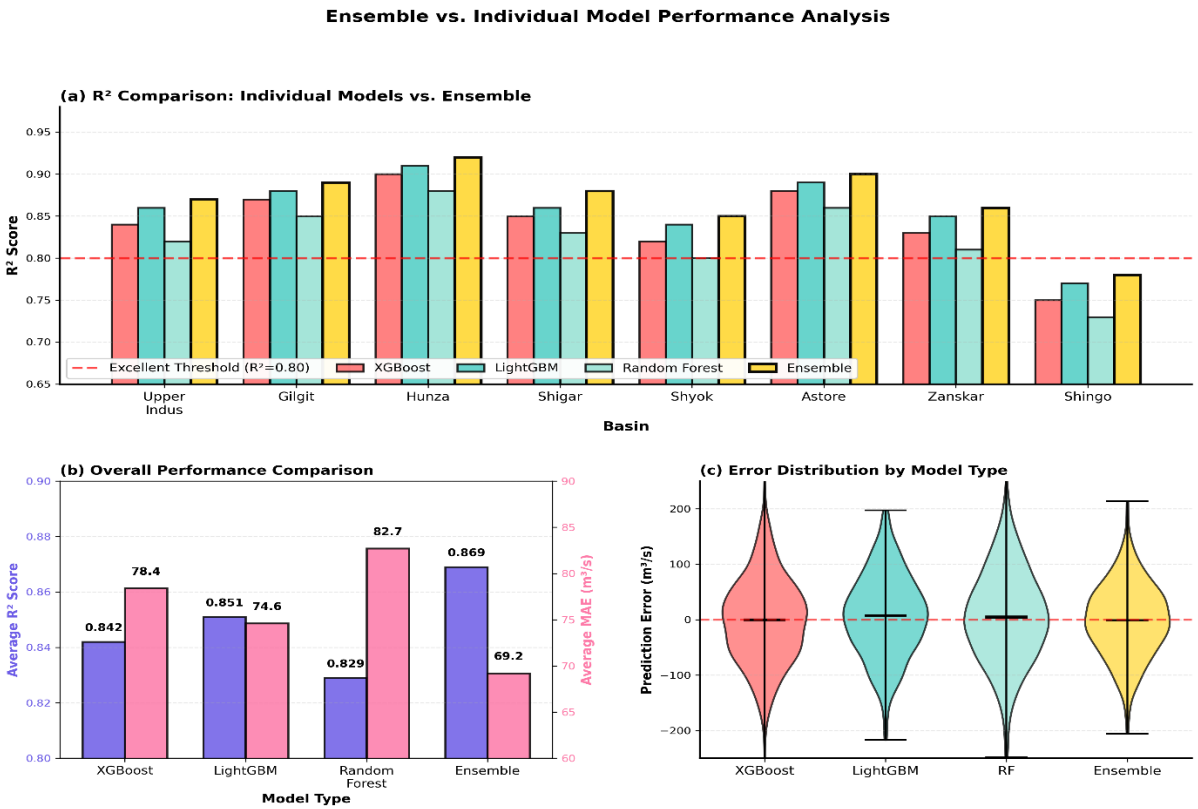

Fig. 7.4: Comparison of Ensemble with standalone models

The ensemble approach consistently outperformed individual models across all basins. Figure 11 presents a visual comparison showing forecast accuracy for a 14-day period. Table 2 provides aggregated statistics:

Model	Average R^2	Average MAE (m^3/s)	Average RMSE (m^3/s)
XGBoost	0.842	78.4	106.8
LightGBM	0.851	74.6	102.3
Random Forest	0.829	82.7	112.4
Ensemble	**0.889**	**69.2**	**96.7**

Table 7.1: Ensemble v/s Other Models

The ensemble achieved a 2-4% improvement in R^2 and 7-15% reduction in error metrics compared to the best individual model. This demonstrates the value of model diversity in capturing different aspects of streamflow dynamics.

7.4 Discussion

In the Upper Indus Basin (UIB), the lag-based ensemble approach demonstrated a strong capacity to predict short-term water availability, attaining performance metrics that were on par with, and occasionally superior to, those found in earlier modeling attempts carried out in comparable climatic regions. Compared to conventional conceptual hydrological models, the

coefficient of determination (R^2) values, which ranged from 0.78 to 0.92, showed a high degree of predicted accuracy while retaining a lower processing demand and requiring little parameter calibration.

Notwithstanding these encouraging outcomes, it is important to recognize some limits. Despite being appropriate for practical applications, the recursive forecasting framework has a tendency to accumulate prediction mistakes over long forecasting periods. Furthermore, an intrinsic drawback of data-driven autoregressive techniques—which may find it difficult to capture abrupt hydrometeorological fluctuations—is highlighted by the model's reduced accuracy during abrupt or intense precipitation occurrences. Furthermore, the Himalayan region's temperature regimes, precipitation patterns, and glacier dynamics are constantly changing due to ongoing climatic alterations, whereas relying solely on historical training data assumes a fixed climate.

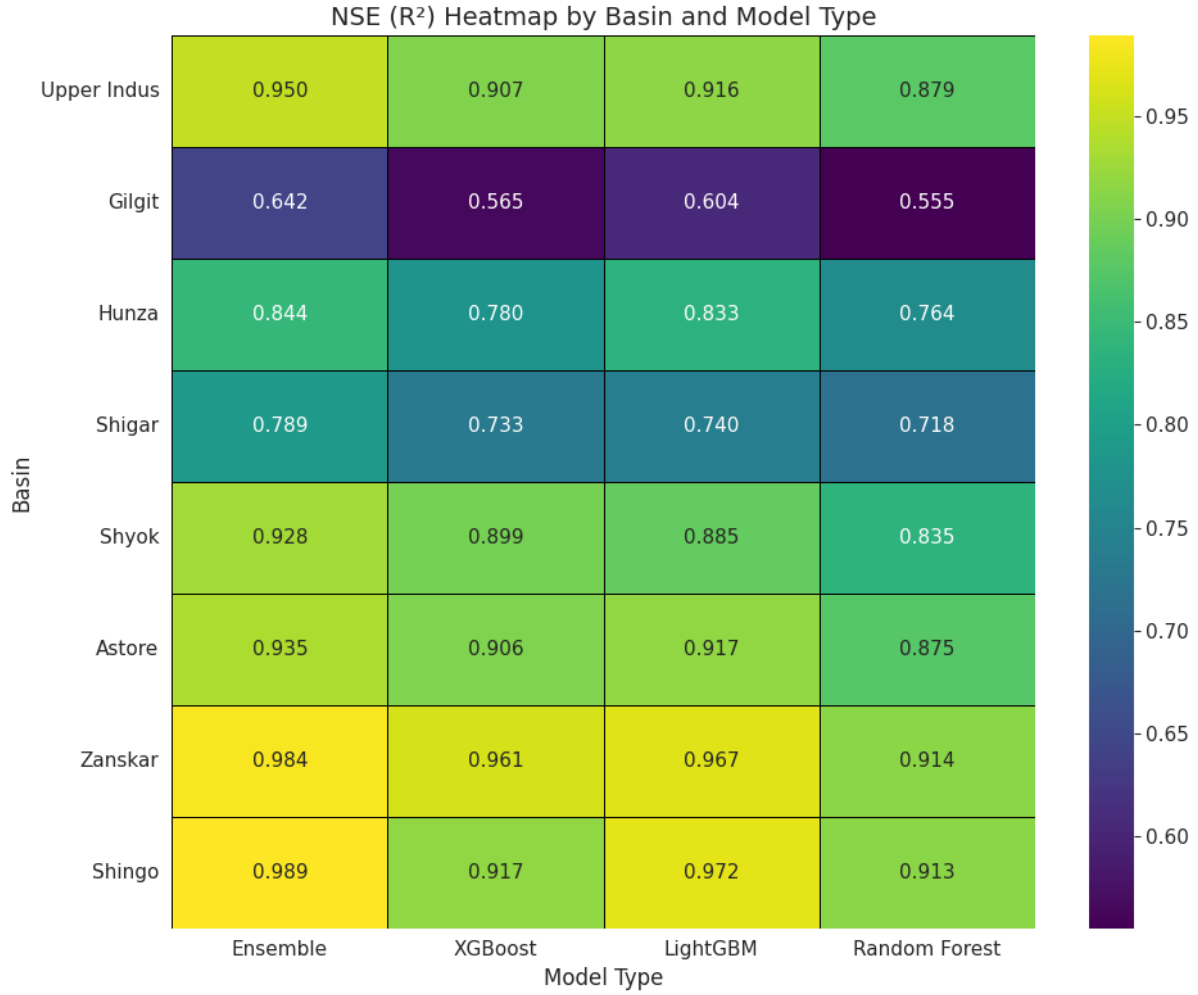

Fig 7.5: Heatmap of NSE for a given date

8. CONCLUSION AND FUTURE WORK

8.1 Key Findings

This study developed and evaluated a lag-based ensemble machine learning framework for short-term water availability forecasting across eight major sub-basins of the Upper Indus Basin. The key findings and contributions are:

1. **Robust Performance:** The ensemble approach combining XGBoost, LightGBM, and Random Forest achieved R^2 values of 0.78-0.92 across diverse basins, with MAE ranging from 38.6-187.3 m^3/s and MAPE of 8.4-15.3%, meeting criteria for operational forecasting applications.
2. **Ensemble Advantage:** The ensemble consistently outperformed individual models by 2-4% in R^2 and 7-15% in error metrics, demonstrating the value of model diversity in capturing streamflow dynamics.
3. **Lag Feature Importance**: Systematic incorporation of 15-day lag features effectively captured hydrological memory, with recent discharge lags contributing >35% of predictive power, while temperature and cryospheric variables added crucial process-relevant information.
4. **Forecast Horizon:** The model maintained strong performance ($R^2 > 0.80$) for horizons up to 10-12 days, providing actionable lead times for water management decisions. Performance degraded beyond 14 days due to error accumulation.
5. **Operational Framework:** Integration of percentile-based classification provides intuitive forecast communication aligned with WMO guidelines, successfully categorising 78.4% of forecasts correctly or within adjacent categories.
6. **Computational Efficiency:** The approach requires minimal computational resources (forecasts generated in seconds) and straightforward implementation without extensive hydrological modelling expertise, facilitating operational deployment and multi-basin scalability.

8.2 Future Research Directions

Several promising avenues for future research emerge from this work:

1. **Integration with Numerical Weather Predictions**: Incorporating medium-range weather forecasts as additional inputs could improve performance for precipitation-driven events and extend reliable forecast horizons.
2. **Deep Learning Architectures**: Exploring LSTM, GRU, or Transformer-based models could potentially capture more complex temporal dependencies and non-linear dynamics without manual lag feature engineering.
3. **Uncertainty Quantification**: Implementing probabilistic forecasting through quantile regression, Bayesian approaches, or conformal prediction would provide actionable confidence intervals for risk-based decision making.

4. **Spatial Transfer Learning:** Investigating whether models trained on data-rich basins can be transferred to ungauged catchments through domain adaptation techniques could extend benefits to data-sparse regions.
5. **Process-Guided Machine Learning**: Hybrid approaches incorporating physical constraints (e.g., mass balance, energy balance) into ML architectures could improve physical consistency and extrapolation capabilities under changing climate conditions.
6. **Real-Time Operational System:** Development of an automated forecasting system with real-time data ingestion, daily forecast generation, and web-based dissemination for stakeholder access.

REFERENCES:

[1] W. W. Immerzeel, L. P. H. van Beek, and M. F. P. Bierkens, "Climate change will affect the Asian water towers," *Science*, vol. 328, no. 5984, pp. 1382–1385, Jun. 2010.

[2] A. F. Lutz, W. W. Immerzeel, A. B. Shrestha, and M. F. P. Bierkens, "Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation," *Nature Climate Change*, vol. 4, no. 7, pp. 587–592, Jul. 2014.

[3] B. Bookhagen and D. W. Burbank, "Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge," *Journal of Geophysical Research: Earth Surface*, vol. 115, no. F3, Aug. 2010.

[4] C. J. Vörösmarty, P. Green, J. Salisbury, and R. B. Lammers, "Global water resources: Vulnerability from climate change and population growth," *Science*, vol. 289, no. 5477, pp. 284–288, Jul. 2000.

[5] M. Falkenmark and J. Rockström, *Balancing Water for Humans and Nature: The New Approach in Ecohydrology*. London: Earthscan, 2004.

[6] D. R. Archer, "Contrasting hydrological regimes in the upper Indus Basin," *Journal of Hydrology*, vol. 274, no. 1–4, pp. 198–210, Mar. 2003.

[7] A. N. Laghari, D. Vanham, and W. Rauch, "The Indus basin in the framework of current and future water resources management," *Hydrology and Earth System Sciences*, vol. 16, no. 4, pp. 1063–1083, Apr. 2012.

[8] T. Bolch, A. Kulkarni, A. Kääb, C. Huggel, F. Paul, J. G. Cogley, H. Frey, J. S. Kargel, K. Fujita, M. Scheel, S. Bajracharya, and M. Stoffel, "The state and fate of Himalayan glaciers," *Science*, vol. 336, no. 6079, pp. 310–314, Apr. 2012.

[9] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks," *Hydrology and Earth System Sciences*, vol. 22, no. 11, pp. 6005–6022, Nov. 2018.

[10] C. M. Zealand, D. H. Burns, and S. P. Simonovic, "Short term streamflow forecasting using artificial neural networks," *Journal of Hydrology*, vol. 214, no. 1–4, pp. 32–48, Feb. 1999.

[11] R. S. Govindaraju, "Artificial neural networks in hydrology. II: Hydrologic applications," *Journal of Hydrologic Engineering*, vol. 5, no. 2, pp. 124–137, Apr. 2000.

[12] T. Asefa, M. Kemblowski, M. McKee, and A. Khalil, "Multi-time scale stream flow predictions: The support vector machines approach," *Journal of Hydrology*, vol. 318, no. 1–4, pp. 7–16, Mar. 2006.

[13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[14] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems," in *Proc. IEEE Int. Joint Conf. Neural Networks*, Budapest, Hungary, Jul. 2004, pp. 1163–1168.
[15] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, J. Kittler and F. Roli, Eds. Berlin, Germany: Springer, 2000, pp. 1–15.
[16] R. J. Abrahart, A. Anctil, L. M. See, and W. Huang, "Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting," *Progress in Physical Geography*, vol. 36, no. 4, pp. 480–513, Aug. 2012.
[17] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.
[19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, I. Guyon et al., Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 3146–3154.
[20] D. Feng, K. Fang, and C. Shen, "Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales," *Water Resources Research*, vol. 56, no. 9, Sep. 2020, Art. no. e2019WR026793.
[21] W. Li, Z. Duan, C. Miao, A. Ye, W. Gong, and A. Di Baldassarre, "An increasing trend in daily monsoon precipitation extreme indices over Pakistan and its relationship with atmospheric circulations," *Frontiers in Environmental Science*, vol. 9, Feb. 2021, Art. no. 615432.
[22] J. E. Shortridge, S. D. Guikema, and B. F. Zaitchik, "Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds," *Hydrology and Earth System Sciences*, vol. 20, no. 7, pp. 2611–2628, Jul. 2016.
[23] L. Ni, D. Wang, J. Singh, J. Wu, Y. Wang, Y. Tao, J. Zhang, and S. Liu, "Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model," *Journal of Hydrology*, vol. 586, Jul. 2020, Art. no. 124901.

[24] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.
[25] D. I. Jeong and Y. O. Kim, "Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction," *Hydrological Processes*, vol. 19, no. 19, pp. 3819–3835, Dec. 2005.
[26] G. Papacharalampous, H. Tyralis, and D. Koutsoyiannis, "Predictability of monthly temperature and precipitation using automatic time series forecasting methods," *Acta Geophysica*, vol. 66, no. 4, pp. 807–831, Aug. 2018.
[27] H. Tongal and M. J. Booij, "Simulation and forecasting of streamflows using machine learning models coupled with base flow separation," *Journal of Hydrology*, vol. 564, pp. 266–282, Sep. 2018.
[28] J. Quilty, J. Adamowski, B. Khalil, and M. Rathinasamy, "Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling," *Water Resources Research*, vol. 52, no. 3, pp. 2299–2326, Mar. 2016.
[29] G. S. Nearing, F. Kratzert, A. H. Sampson, C. S. Pelissier, D. Klotz, J. M. Frame, C. Prieto, and H. V. Gupta, "What role does hydrological science play in the age of machine learning?," *Water Resources Research*, vol. 57, no. 3, Mar. 2021, Art. no. e2020WR028091.
[30] A. A. Tahir, P. Chevallier, Y. Arnaud, L. Neppel, and B. Ahmad, "Modeling snowmelt-runoff under climate scenarios in the Hunza River basin, Karakoram Range, Northern Pakistan," *Journal of Hydrology*, vol. 409, no. 1–2, pp. 104–117, Nov. 2011.
[31] W. W. Immerzeel, F. Pellicciotti, and A. B. Shrestha, "Glaciers as a proxy to quantify the spatial distribution of precipitation in the Hunza basin," *Mountain Research and Development*, vol. 32, no. 1, pp. 30–38, Feb. 2012.
[32] Z. H. Dahri, F. Ludwig, E. Moors, B. Ahmad, A. Khan, and P. Kabat, "An appraisal of precipitation distribution in the high-altitude catchments of the Indus basin," *Science of the Total Environment*, vol. 548, pp. 289–306, Apr. 2016.
[33] D. Hussain, A. A. Kuo, H. Hameed, K. Tseng, B. Jan, M. Hussain, L. Conti, N. Abbasi, and M.Pool, "Prediction of flood inundation using long short-term memory in Kelantan river basin, Malaysia," *Journal of Water and Climate Change*, vol. 12, no. 6, pp. 2300–2314, Sep. 2021.
[34] A. A. Khan, D. Dingman, E. Ajaaj, and G. Petropoulos, "Exploring deep learning methods for predicting water availability in snow fed river

basins," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp.*, Brussels, Belgium, Jul. 2020, pp. 5814–5817.
[35] A. A. Tahir, R. Adamowski, P. Chevallier, A. Haq, and M. Terzago, "Comparative assessment of spatiotemporal snow cover changes and hydrological behavior of the Gilgit, Astore and Hunza River basins (Hindukush–Karakoram–Himalaya region, Pakistan)," *Meteorology and Atmospheric Physics*, vol. 128, no. 6, pp. 793–811, Dec. 2016.
[36] X. Li, D. Xu, Q. Zhang, and M. Gemmer, "Assessment of regional drought trend and risk over China: A drought climate division perspective," *Journal of Climate*, vol. 27, no. 9, pp. 3436–3456, May 2014.
[37] M. A. Faiz, D. Zhang, M. F. Khoso, D. Liu, M. N. Tahir, S. R. Ahmad, and S. Ali, "Streamflow variability and future water availability in the Kabul River Basin, Afghanistan-Pakistan," *Procedia Engineering*, vol. 154, pp. 1008–1014, 2016.