

Article

Machine Learning Approaches for Streamflow Modeling in the Godavari Basin with CMIP6 Dataset

Subbarayan Saravanan ¹, Nagireddy Masthan Reddy ¹, Quoc Bao Pham ², Abdullah Alodah ^{3,*},
Hazem Ghassan Abdo ⁴, Hussein Almohamad ⁵ and Ahmed Abdullah Al Dughairi ⁵

¹ Department of Civil Engineering, National Institute of Technology, Tiruchirappalli 620015, India; ssaravanan@nitt.edu (S.S.); masthanreddy64@gmail.com (N.M.R.)

² Faculty of Natural Sciences, Institute of Earth Sciences, University of Silesia in Katowice, Będzińska Street 60, 41-200 Sosnowiec, Poland; quoc_bao.pham@us.edu.pl

³ Department of Civil Engineering, College of Engineering, Qassim University, Buraydah 51452, Saudi Arabia

⁴ Geography Department, Faculty of Arts and Humanities, Tartous University, Tartous P.O. Box 2147, Syria; hazemabdo@tartous-univ.edu.sy

⁵ Department of Geography, College of Arabic Language and Social Studies, Qassim University, Buraydah 51452, Saudi Arabia; h.almohamad@qu.edu.sa (H.A.); adgerie@qu.edu.sa (A.A.A.D.)

* Correspondence: aalodah@qu.edu.sa

Abstract: Accurate streamflow modeling is crucial for effective water resource management. This study used five machine learning models (support vector regressor (SVR), random forest (RF), M5-pruned model (M5P), multilayer perceptron (MLP), and linear regression (LR)) to simulate one-day-ahead streamflow in the Pranhita subbasin (Godavari basin), India, from 1993 to 2014. Input parameters were selected using correlation and pairwise correlation attribution evaluation methods, incorporating a two-day lag of streamflow, maximum and minimum temperatures, and various precipitation datasets (including Indian Meteorological Department (IMD), EC-Earth3, EC-Earth3-Veg, MIROC6, MRI-ESM2-0, and GFDL-ESM4). Bias-corrected Coupled Model Intercomparison Project Phase 6 (CMIP6) datasets were utilized in the modeling process. Model performance was evaluated using Pearson correlation (R), Nash–Sutcliffe efficiency (NSE), root mean square error (RMSE), and coefficient of determination (R²). IMD outperformed all CMIP6 datasets in streamflow modeling, while RF demonstrated the best performance among the developed models for both CMIP6 and IMD datasets. During the training phase, RF exhibited NSE, R, R², and RMSE values of 0.95, 0.979, 0.937, and 30.805 m³/s, respectively, using IMD gridded precipitation as input. In the testing phase, the corresponding values were 0.681, 0.91, 0.828, and 41.237 m³/s. The results highlight the significance of advanced machine learning models in streamflow modeling applications, providing valuable insights for water resource management and decision making.

Keywords: streamflow; CMIP6; machine learning; RF; SVR; MLP; water



Citation: Saravanan, S.; Reddy, N.M.; Pham, Q.B.; Alodah, A.; Abdo, H.G.; Almohamad, H.; Al Dughairi, A.A. Machine Learning Approaches for Streamflow Modeling in the Godavari Basin with CMIP6 Dataset. *Sustainability* **2023**, *15*, 12295. <https://doi.org/10.3390/su151612295>

Academic Editors: Stefano Morelli, Veronica Pazzi and Mirko Francioni

Received: 26 June 2023

Revised: 3 August 2023

Accepted: 10 August 2023

Published: 11 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In order to better plan and control water use, accurate predictions using streamflow models are essential. Water availability for different uses like drinking water supply, irrigation, and hydroelectric power generation may be predicted by studying the effects of changes in many random variables such as land use and climate using stream and river flow models developed by hydrologists and engineers [1]. Streamflow modeling is also useful for predicting extreme events (e.g., floods and droughts) for better planning and evaluating the effectiveness of flood protection and water management systems [2]. Precipitation, topography, evapotranspiration, and human activities are only a few of the many random elements that can affect streamflow, making it difficult to precisely predict future streamflow. Thus, it is a highly nonlinear and complex hydrologic cycle that has always attracted serious research attention. The three main types of streamflow models are

the physical model, the conceptual model, and the black-box model. However, to provide accurate estimates of hydrologic variables, like runoff, physical models need a great deal of physical data and a detailed mathematical description of the hydrologic structure.

Unlike physical hydrological models, data-driven models may accurately anticipate streamflow without describing the actual mechanics of many hydrological processes. AI methods have been developed to deal with non-stationary and nonlinear streamflow discharge data. More importantly, models based on artificial neural networks (ANNs) were proven to accurately predict streamflow discharge. ANNs or “black-box models” could provide results approximating the desired ones by tweaking their internal settings smartly. Consequently, ANN has the capability to make predictions based on novel and unfamiliar inputs due to the parameterization of the connection between input and output within the structural framework of the model [3]. ANN models could identify the complicated pattern with only a few inputs, such as rainfall and streamflow. The catchments’ spatial and temporal variability makes monitoring these variables exceptionally challenging [4,5]. Rainfall–runoff modeling, streamflow prediction, reservoir inflow forecasting, rainfall forecasting, river sediment modeling, and hydraulic energy estimates have all benefited from the use of ANNs in hydrological research [6–11]. Several studies (e.g., [12–14]) have investigated the effectiveness of using ANNs for streamflow estimation and have concluded that they yield acceptable outcomes. Ninety percent of hydrological applications have employed a traditional feedforward neural network, such as MLP trained using the backpropagation technique [15,16]. Similarly, support vector machines (SVM) are commonly utilized for hydrological prediction and management [17]. For example, the SVM model predicted China Huaxi station’s monthly river flow accurately, according to [18]. Sedighi et al. [19] used the ANN model and SVM built on MODIS image data from 2003–2005 to forecast streamflow in the Roodak region northeast of Tehran. Ghorbani et al. [20] used SVM and ANN to estimate the daily water flow in Cypress, Texas, to evaluate their ability in terms of river flow prediction. They came to the conclusion that the SVM provided more accurate results than the ANN. Ghorbani et al. [21] tested hybrid artificial intelligence models to estimate the monthly flow in Turkey’s Igdir river and found that the firefly algorithm combo model performed best. Also, [21,22] compared SVM and ANN models to predict the Zarineh-rood river’s discharge in Iran and found that the former was more accurate. Alizadeh et al. [23] tested the hybrid wavelet SVM model’s capacity to estimate daily US streamflow and found it to be very accurate. Several instances of SVM’s use in streamflow modeling could be found in the works of Ghorbani et al. [24], Lin et al. [25], and Seyam et al. [26]. Recently, many machine learning models have been adopted to simulate streamflow across the globe, e.g., RF [27,28], MLP [29,30], SVM [25,31], M5P [32,33], LR [34,35], and much more. A comprehensive examination of the applications of data-driven models in hydrologic processes can be found in the following publications: Fahimi et al. [36]; Hadi and Tombul [16].

According to Quinlan et al. [37], the M5 algorithm is categorized as a type of tree-based structure that incorporates multiple linear regression models within its components. Consequently, these model trees can be likened to piecewise linear functions. Although the M5 model tree is a recent development in water resources, its usage in actual occurrences has shown it to be fairly reliable. For instance, when it was applied to the water level–discharge relationship by Bhattacharya and Solomatine [38], it was noticed that M5 had a similar degree of prediction accuracy to an ANN created using the same data. M5 handles jobs with very high dimensions and learns effectively [39]. Sihag et al. [40] examined the optimum sediment estimation model utilizing M5P and RF regression and indicated that the M5P-based model showed the best performance. In the Koyna River basin in India, Bajirao et al. [32] evaluated the viability of many data-driven strategies for runoff forecasting, including ANN, SVM, RF, and M5P models. Machine learning algorithms were used by Reddy et al. [41] to forecast monthly surface runoff in the tropical Kallada River Basin. They discovered that machine learning algorithms can effectively simulate the rainfall–runoff process. Singh et al. [42] investigated the accuracy of the empirical Kostiaikov

model and the ANN, MLR, RF, and M5P prediction models to investigate the infiltration process. They discovered that the ANN, MLR, RF, and M5P models outperformed the empirical Kostiaikov model in terms of performance. In their assessment of the RF model's potential for daily streamflow forecasting in several watersheds, Pham et al. [43] found that RF can generate precise short-term streamflow forecasts for all examined watersheds.

Climate extremes are projected to increase in frequency and severity as global temperatures rise, posing significant challenges for vulnerable communities, particularly in developing economies with limited capacity for adaptation [44]. Streamflow modeling plays a crucial role in mitigating the impacts of climate change on water resources. To address uncertainties in weather and climate systems, the use of global circulation models (GCMs) is essential for collecting large-scale geographical and temporal data [45]. GCMs offer valuable insights into the climate system, complementing observational data for streamflow modeling and enhancing the applicability of strategies for mitigation and adaptation to changing climatic conditions [46].

Water resource management is of paramount importance for sustaining life, ecosystems, and various human activities. Accurate streamflow forecasting plays a crucial role in effective water resource planning, enabling stakeholders to make informed decisions and mitigate risks associated with water availability and flood control, especially considering the increasing impact of climate change and anthropogenic activities on hydrological processes. In this study, our focus is on forecasting one-day-ahead streamflow in the Pranhita subbasin (Wairagarh station), a vital part of the Godavari basin in India. To achieve this, the application of several advanced machine learning models, namely SVR, RF, M5P, MLP, and LR, as traditional hydrological models may have limitations in capturing the complex and nonlinear relationships between hydrological variables. Leveraging various precipitation datasets, including the IMD and bias-corrected CMIP6 (EC-Earth3, EC-Earth3-Veg, MIROC6, MRI-ESM2-0, and GFDL-ESM4) datasets, and incorporating lag in streamflow, to estimate streamflow one day in advance using maximum and minimum temperatures. The delay in rainfall and streamflow is assessed through correlation attribute evaluation and pairwise correlation attribute evaluation, utilizing a dataset spanning 7064 days from 1993 to 2014 for modeling. The study's innovative approach employs bias-corrected CMIP6 precipitation and IMD gridded data, providing a more accurate streamflow forecast with fewer inputs compared to traditional methods. These findings can offer valuable insights for water resource management and informed decision making, benefiting policymakers and stakeholders in coping with water-related challenges while ensuring the sustainable use of water resources in the Godavari basin and similar hydrological contexts worldwide.

2. Study Area

The present research was performed in the Pranhita subbasin of the Godavari River basin in the Indian state of Maharashtra. The research region has a total drainage area of 2600 km² and is located between the longitudes 80°5' E–80°40' E and latitudes 20°20' N–20°47' N in Maharashtra and a small area in Chhattisgarh. According to the digital elevation models (DEM) produced by the Shuttle Radar Topography Mission (SRTM), the elevation of the research region varies from its highest point, which is 660 m, to its lowest position, which is 208 m. Figure 1 shows the map of the research region, along with the IMD gridded stations, the Wairagarh Streamflow station, the stream network, and the DEM. The average annual rainfall in the study area is 1421 mm, while temperatures range from 20.75 °C to 33.33 °C. Geology in the study area is dominated by Dongargarh Granite and little traces of Wairagarh metasediments [47]. This study area comprises 76.01% deciduous broadleaf forest, 22.72% cropland, and less than 1% shrubland and mixed forest [48]. Since the city of Gadchiroli is located downstream of this research region, accurate streamflow modeling of this study area will assist in managing water resources and developing policies to reduce the risk of flooding.

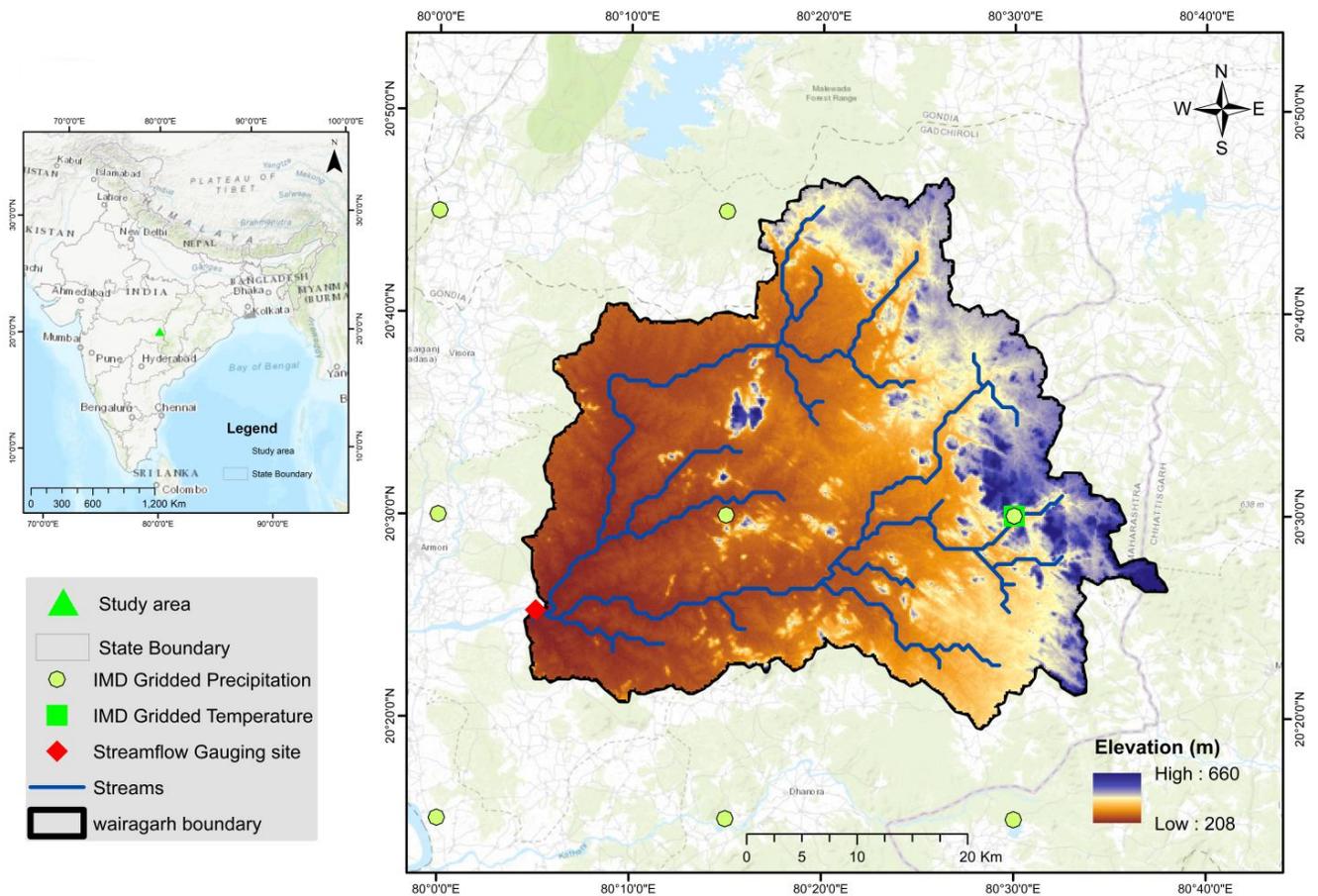


Figure 1. Location map of the study area.

3. Materials and Methods

This study aims to forecast the streamflow of the Indian Godavari River. To achieve this objective, the required data were collected and standardized. The scientific time series data for discharge, temperature, and precipitation were gathered on a daily basis. After organizing the data, the University of Waikato models were implemented using the Weka 3.8.6 application [49]. The software was utilized for two rounds of training and testing to determine the optimal combination for each model. The best model for predicting was chosen from among four machine learning models and linear regression developed in this work, utilizing the IMD and CMIP6 datasets as training data. This procedure aimed to choose the best model for machine learning to use for forecasting purposes using the IMD and CMIP6 datasets. The optimal AI model architecture was chosen by calculating the least value of RMSE while simultaneously maximizing the values of R^2 , NSE, and R. The entire methodology and procedures of this investigation are presented in Figure 2 in a flowchart format.

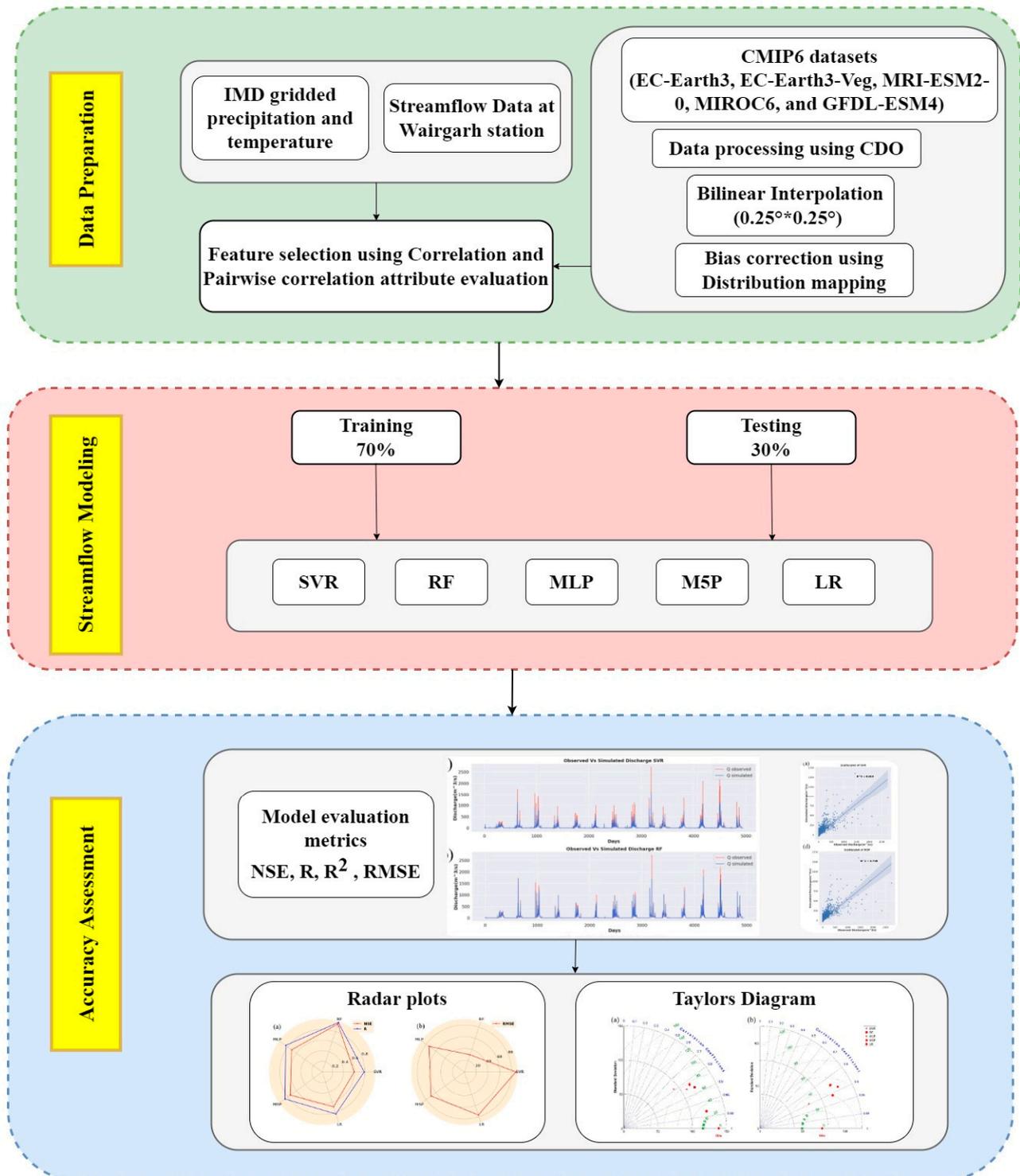


Figure 2. Flowchart of the methodology adopted in this study.

3.1. IMD Data

The gridded IMD dataset for precipitation and temperature, available from 1901 to 2021, was used in this study. This dataset provides spatial resolutions of 0.25° for precipitation and 1° for temperature. To create the dataset, IMD employed Shepard's interpolation method, utilizing data from 6695 gauges. It has been widely employed in

India as a reference for precipitation data to rectify biases in CMIP6 models. IMD generated a gridded precipitation dataset established on gauge observations [50,51].

3.2. CMIP6 Model Data

The five CMIP6 models that were employed in this study to assess the streamflow prediction are shown in Table 1. The Earth System Grid Federation (ESGF) archives, available for review at <https://esgf-node.llnl.gov/search/cmip6>, accessed on 15 July 2022, provide access to GCMs data. To ensure consistency, all GCMs data were spatially remapped to a standardized latitude and longitude grid of $0.25^\circ \times 0.25^\circ$ using a bilinear interpolation [52]. The selected datasets in this study, namely EC-Earth3, EC-Earth3-Veg, MRI-ESM2-0, GFDL-ESM4, and MIROC6, are renowned for their representation of extreme precipitation patterns in India [53]. EC-Earth3, EC-Earth3-Veg, MRI-ESM2-0, and GFDL-ESM4 are advanced Earth System Models from ECMWF, MRI, and GFDL, respectively, providing comprehensive representations of land–atmosphere interactions and atmospheric, oceanic, and land components. MIROC6, with high-resolution atmospheric and oceanic processes, is ideal for detailed regional climate simulations. These datasets enable a comprehensive assessment of their performance in streamflow forecasting and their relevance to water resource management in India.

Table 1. CMIP6 models used in the study.

Model	Atmospheric Resolution	Institution
EC-Earth3	$0.7^\circ \times 0.7^\circ$	EC-EARTH consortium
EC-Earth3-Veg	$0.7^\circ \times 0.7^\circ$	EC-EARTH consortium
GFDL-ESM4	$1.3^\circ \times 1^\circ$	Geophysical Fluid Dynamics Laboratory
MIROC6	$1.41^\circ \times 1.41^\circ$	JAMSTEC, AORI, NIES, and R-CCS
MRI-ESM2-0	$1.1^\circ \times 1.1^\circ$	Meteorological Research Institute

3.3. Streamflow Data

Daily streamflow data for the Wairagarh station were sourced from the India Water Resources Information System portal (<https://indiawris.gov.in/wris/#/> accessed on 10 April 2022) for the period spanning 1993 to 2014 [54].

3.4. Data Processing

The IMD provided gridded precipitation and temperature in NetCDF format. Data in NetCDF format were processed and extracted using Climate Data Operators (CDO) [55] and ArcGIS 10.3. When working with ArcGIS 10.3, the “make NetCDF table view” tool can be found in the “multi-dimension tools” section of the “Arc Toolbox”. This tool is used to extract grid-based data from NetCDF files [56]. After data extraction, there were 8 points of gridded precipitation data from an IMD in the research region. The average rainfall across the research region was estimated using the Thiessen polygon technique. Forecasting future streamflow is a dynamically evolving natural process, where the current response of any hydrologic process is shaped by the memory of past reactions stored within the hydrologic system. The CMIP6 precipitation datasets were downscaled using the distribution mapping method and the IMD dataset was used as a reference. To gain additional insights into the distribution mapping approach, the following literature may be helpful [57,58].

The current and past reactions to various hydrologic parameters, such as precipitation, runoff, and temperature, would determine the present and past streamflow response. Consequently, the selection of data inputs for forecasting streamflow is performed using a correlation attribute evaluation and pairwise correlation attribute evaluation; as seen in Table 2, the top 5 influencing factors were considered in this study, where St represents the current streamflow and St-1 indicates the precipitation from one day prior, similar to how Pt indicates present-day precipitation and Pt-1, Pt-2 reflects precipitation from the previous day, respectively. Of the data from 1993 to 2014, 70% (4944 days) were utilized for training, and 30% (2120 days) were used for testing, after the deletion of the missing

data. All inputs were normalized to a certain range between 0 and 1 for input data training purposes. In this study, input parameters were normalized using Equation (1) to eliminate their dimensionality and guarantee that all input variables were assigned sufficient weight during the training phase. It facilitates the construction of models by enabling the quick convergence of learning. It makes the model development more interpretable [59].

$$S_{norm_i} = \frac{S_i - S_{min}}{S_{max} - S_{min}}, i = 1, 2, 3, 4, \dots, n \quad (1)$$

where S_{norm_i} is the normalized value of any parameter, S_{min} and S_{max} are the minimum and maximum values of the datasets, and n is the total number of datasets used for training and testing.

Table 2. Correlation and pairwise correlation attribute evaluation.

Correlation Attribute Evaluation		Pairwise Correlation Attribute Evaluation	
Parameter	Score	Parameter	Score
Pt	0.678	St-1	9.6452
Pt-1	0.615	P	8.8522
St-1	0.611	Pt-1	8.3758
St-2	0.391	Pt-2	6.8578
Pt-2	0.371	St-2	6.8182
St-5	0.341	Pt-7	6.3642
St-4	0.34	Pt-4	6.3135
St-3	0.325	Pt-3	6.3092
St-6	0.323	Pt-6	6.3069
St-7	0.321	Pt-5	6.2276

3.5. SVR

SVR is a subclass of SVM designed specifically for tackling regression problems; it was developed by [60]. SVR is used to forecast continuous values as opposed to class labels, like SVM is used for classification [61]. The key to SVR's success is identifying the optimal border (or "hyperplane") that divides the data into distinct groups. The objective of SVR is to identify a boundary that keeps the data points within a specified distance of the hyperplane while maximizing the margin between the data points and the hyperplane (called the "epsilon-tube"). Because of this, SVR can better understand data with higher noise. It is effective in dealing with large dimensional datasets and may be utilized for both linear and nonlinear regression issues, making SVR a versatile tool. The SVM approach is described in great length in a number of different published works [62,63]. A schematic diagram of SVR can be seen in Figure S1. An SVR carries out two main tasks: (1) estimating training-time prediction errors and (2) calculating output values from weight, bias, and input data [64].

$$y = \sum_{l=1}^n (\alpha_l - \alpha_l^*) \cdot Kr(x_l, x_m) + c \quad (2)$$

where c represents the bias, α_l and α_l^* represent Lagrange multipliers, and $Kr(x_l, x_m)$ represents the kernel function, which is shown in Equations (3) and (4).

Polynomial Kernel:

$$Kr(x_l, x_m) = (x_l \cdot x_m)^d \quad (3)$$

Gaussian Radial Basis function:

$$Kr(x_l, x_m) = \exp\left(-\frac{\|x_l - x_m\|^2}{2\sigma^2}\right) \quad (4)$$

3.6. RF

RF is a type of ensemble learning method first presented by [65]. It is a slight modification of bagged decision trees that are created from a wide collection of uncorrelated

trees and requires the adjustment of only a few variables [66]. As a “supervised learning method”, RF draws conclusions about a given dataset by employing a collection of “decision trees” to draw such conclusions. By lowering precision, it creates trees whose growth is dependent on that of their neighbors. In a manner analogous to that of a “Decision Tree,” it is compatible with “classification” as well as “regression” models. A schematic representation is shown in Figure S2.

The training process for the random forest is accomplished by constructing a large number of decision tree models that are unconnected to one another $[h(X, \theta_k); k = 1, \dots]$. The modes of the data are the final result of the classification process, and each of these unique decision trees makes its own prediction on the classification of the sample. The efficacy of the random forest model is improved by the inclusion of additional training sets that are unrelated to one another. The output of the random forest based on the many classifications learned from training sets is decided by following Equation (5)

$$H(x) = \arg\max_z \sum_{i=1}^k I(h_i(x) = Z)I(\cdot) \quad (5)$$

where Z is the outcome variable and $I(\cdot)$ is the indicative function. Here, $H(x)$ is the RF model, and h_i is the single decision tree model. Random forests enhance accuracy in classification and regression issues while also reducing the likelihood of decisions being overly tailored to their context. In addition, data normalization is not required because the model is governed by a set of rules. However, in order to construct a large number of decision trees and obtain the output, a larger amount of processing power and training time is required. It is impossible to assess each variable’s relevance using the random forest classifier, and its interpretability is also compromised.

3.7. MLP

Inspired by the neurons in our brains, neural networks are a sort of algorithm. Its primary purpose is to find regularities in huge datasets. In the last several decades, ANNs have been more popular for dealing with hydrology-related issues due to their flexibility and effectiveness in simulating nonlinear and complex hydrologic processes [67–70]. The ANN technique differs from previous computing approaches because it operates in parallel. An ANN consists of many neurons organized into input, output, and hidden layers. The data signals are received and processed by the artificial input neurons, which then send the output to the remaining neurons in the system. Multilayer feedforward refers to the method of organizing layers and processing forward. The weighted linkages feed activations in the forward path from input to output. Adjusting the “weights” of the various connections between nodes trains a neural network to carry out a predetermined task [71]. The basic operation of an MLP neural network is shown in a simplified form in Figure S3. The neurons in MLP’s input, hidden, and output layers reveal the basic layout of the network. To generate an output, a transfer function is applied to the weighted sum of the inputs from outer space or the outputs of the preceding layer at each node in the hidden and output layers. Neuronal function is developed using Equation (6)

$$Y_j = \sum_{i=1}^n f(w_{ij}x_i + b_j) \quad (6)$$

Here, Y_j represents the output at node “ j ”, w_{ij} is the weight connecting node “ i ” and node “ j ” of the previous and current layer, x_i represents the sequence of inputs, and b_j represents bias at node “ j ”.

3.8. M5P

M5P is a decision tree technique that can perform both classification and regression; it was proposed by [37]. The “P” in M5P refers to “piecewise,” indicating that this is a variant of the M5 decision tree method. To provide more precise predictions, M5P employs linear regression models rather than a single constant value at the branch nodes of the

decision tree. The technique can also work with category variables and missing data. The splitting criteria are used to decide upon a characteristic by which to partition the training data into subsets T , of which each ultimately approaches a distinct node. Each feature is evaluated by computing the predicted reduction in error at a certain node, where the standard deviation of the class in T represents the error. At each node, the predicted error reduction is maximized by selecting the characteristic for a split that maximizes that reduction. For an estimate of the predicted error reduction, use Equation (7) to obtain the standard deviation reduction (*SDR*) [39].

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} * sd(T_i) \quad (7)$$

where T_i is the collection of attributes along which the node was divided when it was initially created. Continuous quantitative characteristics are predicted via linear regression models at the leaf level. They are like piecewise linear functions, but when you put them all together, you obtain a nonlinear function [38]. The goal is to build a model that predicts an output value based on the input attribute values of the training examples. In most circumstances, a model's quality will be determined by how well it can predict the values of unknown cases. When the remaining number of instances is small, or the standard deviation is just slightly smaller than the standard deviation of the original set, the splitting procedure ends.

3.9. LR

One of the fundamental challenges in statistical analysis is developing a model that accurately describes the connection between a dependent variable and a group of independent variables [72]. Simply put, it is a statistical method for examining the interplay between a number of predictor variables (or features) and a single dependent variable (also known as the response variable or outcome). MLR seeks to identify the optimal linear combination of predictor factors for a given response. It is similar to linear regression but uses several factors to draw conclusions. Fitting a linear function as a model for a quantitative connection is what linear regression is all about, and we see it in Equation (8):

$$y = \gamma_0 + \gamma_1x_1 + \gamma_2x_2 + \gamma_3x_3 + \dots + \gamma_nx_n \quad (8)$$

where y is the streamflow at Wairagarh, and x_1 to x_n are the independent variables such as lag in precipitation, streamflow, and temperature [73–75].

Tables 3–6 display the hyperparameters of the various methods employed in this original study model creation. Weka 3.8.6 was used to create many SVM, RF, MLP, and M5P models for this research.

Table 3. Hyperparameters used for SVR.

Parameter	Value
batchSize	100
C	1.0
filterType	Normalize training data
kernel	PolyKernel
numDecimalPlaces	2
cacheSize	250,007
exponent	1.0
regOptimizer	RegSMOImproved
epsilon	1×10^{-12}
epsilonParameter	0.001
seed	1
tolerance	0.001

Table 4. Hyperparameters used for RF.

Parameter	Value
bagSizePercent	100
batchSize	100
maxDepth	0
numDecimalPlaces	2
numExecutionSlots	1
numFeatures	0
numIterations	100
seed	1

Table 5. Hyperparameters used for MLP.

Parameter	Value
batchSize	100
hiddenLayers	5
learningRate	0.3
momentum	0.2
numDecimalPlaces	2
seed	0
trainingTime	500
validationSetSize	0
validationThreshold	20

Table 6. Hyperparameters used for M5P.

Parameter	Value
batchSize	100
minNumInstances	4.0
numDecimalPlaces	4

3.10. Model Evaluation Metrics

The Wairagarh station employs four commonly used evaluation metrics, namely R^2 , NSE, RMSE, and R, to analyze the daily streamflow measurements. NSE is a widely used statistical measure that quantifies the ratio of the residual variance to the variance of the observed data [51,54,76]. The NSE metric quantifies the level of agreement between observed streamflow and modeled streamflow data, as indicated by their alignment with the 1:1 line. The NSE ranges are explicitly specified in Table 7, accompanied by the corresponding formula [77]. The variable R serves as a measure of the degree of similarity between simulated data and observed data. RMSE is a commonly utilized statistical metric that is employed to quantify the disparity between the predicted values generated by a product and the corresponding actual values. R^2 quantifies the extent to which the observed data exhibits variability. Table 7 displays the expressions, parameter range, and performance value for evaluation metrics. In this table, S_O^i denotes the observed streamflow data, S_S^i denotes the simulated streamflow, and \bar{S}_O denotes the mean of the observed streamflow data.

Table 7. Model evaluation metrics.

Parameter	Expression	Range	Performance
Nash–Sutcliffe efficiency	$NSE = 1 - \frac{\sum_{i=1}^n (S_o^i - S_s^i)^2}{\sum_{i=1}^n (S_o^i - \bar{S}_o)^2}$	0.75 < NSE ≤ 1.00 0.65 < NSE ≤ 0.75 0.50 < NSE ≤ 0.65 0.4 < NSE ≤ 0.50 NSE ≤ 0.4	Very good Good Satisfactory Acceptable Unsatisfactory
Pearson correlation	$R = \left(\frac{n \sum_{i=1}^n (S_o^i S_s^i) - (\sum_{i=1}^n S_o^i)(\sum_{i=1}^n S_s^i)}{\sqrt{(n \sum_{i=1}^n (S_o^i)^2 - (\sum_{i=1}^n S_o^i)^2)} \sqrt{(n \sum_{i=1}^n (S_s^i)^2 - (\sum_{i=1}^n S_s^i)^2)}} \right)$	−1 to 1	-
Root means square error	$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_o^i - S_s^i)^2}{n}}$	0 to ∞	-
Coefficient of determination	$R^2 = \left(\frac{n \sum_{i=1}^n (S_o^i S_s^i) - (\sum_{i=1}^n S_o^i)(\sum_{i=1}^n S_s^i)}{\sqrt{(n \sum_{i=1}^n (S_o^i)^2 - (\sum_{i=1}^n S_o^i)^2)} \sqrt{(n \sum_{i=1}^n (S_s^i)^2 - (\sum_{i=1}^n S_s^i)^2)}} \right)^2$	0.7 < R ² ≤ 1 0.6 ≤ R ² < 0.7 0.5 < R ² < 0.6 0.0 ≤ R ² < 0.5	Very good Good Satisfactory Unsatisfactory

4. Results

In this current study, five models, namely SVR, RF, MLP, M5P, and LR, were used to predict one-day-ahead streamflow with two-day streamflow lag, maximum temperature, minimum temperature, and numerous precipitation datasets (such as IMD, EC-Earth3, EC-Earth3-Veg, MRI-ESM2-0, MIROC6, and GFDL-ESM4) with two-day lag. The models were also used to predict one-day-ahead streamflow; Table 8 presents the statistical characteristics of the information that was used. The generated models are simulated from the years 1993 to 2014. Table 8 demonstrates the data for streamflow, Tmin, Tmax, and different precipitation datasets. Streamflow and all precipitation datasets have considerably skewed distributions (in the range of 3.94 to 13.43). However, the data for Tmax and Tmin are symmetrical.

Table 8. Statistics of streamflow, IMD precipitation, maximum temperature, minimum temperature, and CMIP6 datasets.

Statistic	Streamflow (m ³ /s)	IMD (mm)	Tmin (°C)	Tmax (°C)	EC-Earth3 (mm)	EC-Earth3-Veg (mm)	MIROC6 (mm)	MRI-ESM2-0 (mm)	GFDL-ESM4 (mm)
Training									
Mean	40.91	4.15	20.57	33.12	3.67	3.68	3.88	2.91	2.73
Median	0.31	0.00	22.45	31.88	0.00	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	6.58	21.70	0.00	0.00	0.00	0.00	0.00
Maximum	2732.00	312.60	32.89	46.57	147.38	121.56	221.50	481.70	261.80
Standard Deviation	138.74	13.22	5.16	4.77	11.04	11.06	12.42	13.74	12.33
Skew	8.31	7.51	−0.46	0.76	4.79	4.42	5.70	13.43	9.28
Testing									
Mean	24.56	4.06	21.27	33.29	3.43	3.86	3.16	2.53	2.83
Median	0.46	0.00	22.94	31.90	0.00	0.00	0.00	0.00	0.00
Minimum	0.00	0.00	7.66	20.66	0.00	0.00	0.00	0.00	0.00
Maximum	1405.00	305.15	32.14	46.24	81.10	118.40	166.06	157.75	155.64
Standard Deviation	73.06	12.74	5.01	4.94	10.07	11.40	10.31	9.50	11.93
Skew	7.38	9.45	−0.48	0.73	3.94	4.26	5.89	5.74	7.21

Tables 9–14 illustrate the predictive performance of the five chosen models for stream-flow forecasting one day in advance.

Table 9. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using EC-Earth3 dataset.

EC-Earth3		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.356	0.604	0.365	111.327	0.539	0.749	0.562	49.572
RF	0.916	0.969	0.938	40.192	0.496	0.777	0.604	53.878
MLP	0.467	0.686	0.470	101.306	0.500	0.751	0.563	51.669
M5P	0.452	0.673	0.452	102.646	0.502	0.756	0.572	51.556
LR	0.400	0.633	0.400	107.426	0.484	0.722	0.521	52.478

Table 10. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using EC-Earth3-Veg dataset.

EC-Earth3-Veg		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.357	0.605	0.366	111.228	0.543	0.751	0.564	49.398
RF	0.917	0.967	0.936	39.988	0.406	0.748	0.560	56.278
MLP	0.405	0.698	0.488	107.021	0.108	0.783	0.612	69.001
M5P	0.453	0.673	0.453	102.604	0.493	0.754	0.568	52.019
LR	0.403	0.634	0.403	107.224	0.482	0.722	0.522	52.599

Table 11. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using GFDL-ESM4 dataset.

GFDL-ESM4		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.354	0.602	0.362	111.529	0.539	0.747	0.558	49.571
RF	0.917	0.970	0.940	39.859	0.441	0.754	0.568	54.594
MLP	0.470	0.693	0.481	100.943	0.579	0.779	0.607	47.390
M5P	0.452	0.672	0.452	102.698	0.493	0.752	0.565	51.991
LR	0.400	0.632	0.400	107.466	0.479	0.719	0.517	52.724

Table 12. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using IMD dataset.

IMD		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.604	0.787	0.619	87.321	0.796	0.892	0.796	33.027
RF	0.951	0.979	0.959	30.805	0.681	0.910	0.829	41.238
MLP	0.716	0.850	0.723	73.972	0.652	0.862	0.743	52.514
M5P	0.748	0.865	0.748	69.597	0.483	0.882	0.778	52.542
LR	0.692	0.832	0.692	76.938	0.491	0.851	0.725	52.098

Table 13. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using MIROC6 dataset.

MIROC6		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.354	0.602	0.363	111.484	0.539	0.747	0.559	49.603
RF	0.917	0.968	0.938	39.931	0.512	0.766	0.586	51.975
MLP	0.419	0.700	0.490	105.693	0.202	0.788	0.622	65.235
M5P	0.451	0.672	0.451	102.775	0.496	0.753	0.567	51.839
LR	0.399	0.632	0.399	107.528	0.480	0.720	0.518	52.674

Table 14. NSE, R, R², and RMSE for SVR, RF, MLP, M5P, and LR models using MRI-ESM2-0 dataset.

MRI-ESM2-0		Training			Testing			
Method	NSE	R	R ²	RMSE	NSE	R	R ²	RMSE
SVR	0.353	0.602	0.362	111.547	0.539	0.747	0.558	49.603
RF	0.918	0.969	0.939	39.693	0.430	0.755	0.569	55.144
MLP	0.385	0.701	0.491	108.814	0.137	0.768	0.589	67.874
M5P	0.581	0.764	0.584	89.746	0.467	0.755	0.570	53.323
LR	0.399	0.632	0.399	107.503	0.482	0.720	0.519	52.567

Table 9 represents the performance evaluation indices using the EC-Earth3 dataset; NSE, R, R², and RMSE values of the selected finest model RF were observed to be 0.916, 0.969, 0.938, and 40.192 m³/s, correspondingly, during training and 0.496, 0.777, 0.604, and 53.878 m³/s, correspondingly, during testing. Similar to the EC-Earth3 dataset, the EC-Earth3-Veg dataset was used as input in the place of precipitation, in which the NSE, R, R², and RMSE values of the selected best model RF were observed to be 0.917, 0.967, 0.936, and 39.988 m³/s during training and 0.406, 0.748, 0.560, and 56.278 m³/s during testing, as shown in Table 10. As shown in Table 11, EC-Earth3-Veg precipitation was replaced with GFDL-ESM4 to run all five models, and model evaluation metrics such as NSE, R, R², and RMSE for the RF model were seen to be 0.917, 0.970, 0.940, and 39.859 m³/s, respectively, during training and 0.44, 0.754, 0.568 and 54.594 m³/s, correspondingly, during testing. Table 13 shows MIROC6 as the input precipitation used where the evaluation metrics NSE, R, R², and RMSE were observed to be 0.917, 0.968, 0.938, and 39.931 m³/s while training and 0.512, 0.766, 0.586, and 51.975 m³/s while testing for the RF model. Table 14 indicates that the MRI-ESM2-0 was used as the input dataset, in which the evaluation metrics were NSE, R, R², and RMSE, which are 0.918, 0.969, 0.939, and 39.693 m³/s during training and 0.430, 0.755, 0.569, and 55.144 m³/s during testing.

The IMD gridded precipitation used by the five models is shown in Table 12. The values of the NSE, R, R², and RMSE of the chosen SVR model were found to be 0.604, 0.787, 0.619, and 87.321 m³/s during training, and 0.796, 0.892, 0.796, and 33.027 m³/s during testing. The best RF model was picked in the same way as SVR, utilizing quantitative statistical performance evaluation criteria. The results for the chosen RF model's NSE, R, R², and RMSE were found to be 0.951, 0.979, 0.959, and 30.805 m³/s during training, and 0.681, 0.910, 0.829, and 41.238 m³/s during testing. Statistical performance indicators were used to choose the optimal MLP model from among the several that had been built. The chosen MLP model was found to have NSE, R, R², and RMSE training values of 0.716, 0.850, and 73.972 m³/s and testing values of 0.652, 0.862, 0.743, and 52.514 m³/s. The optimum M5P model was also chosen through an iterative process of trial and error. The chosen M5P model had training-time NSE, R, R², and RMSE values of 0.748, 0.865, and 69.597 m³/s, and test-time values of 0.483, 0.882, and 52.542 m³/s. To the same effect, a process of trial and error was used to determine which LR model performed the best. It was found that the training NSE, R, R², and RMSE values of the chosen M5P model were 0.692, 0.832, 0.692, and 76.938 m³/s, whereas the testing values were 0.491, 0.851, 0.725, and 52.098 m³/s. Based on training and testing performance using IMD gridded precipitation, the RF model was shown to be better capable of simulating one-day-ahead runoff time series compared to SVR, RF, MLP, M5P, and LR. Training and testing results showed that RF models had the best prediction performance, followed by SVR, MLP, M5P, and LR models. IMD gridded precipitation performed exceptionally well in terms of model assessment criteria compared to other climate datasets.

Time series and scatter plots of predicted vs. actual streamflow were used to qualitatively compare the performance of various models' predictions. Here, the assessment was carried out visually by comparing the predicted and actual hydrographs. Figures 3 and 4 represent the time series plots of all five models during training and testing using IMD

gridded precipitation as input. Figures 5 and 6 represent the scatterplot of all the models during training and testing using IMD gridded precipitation as input.

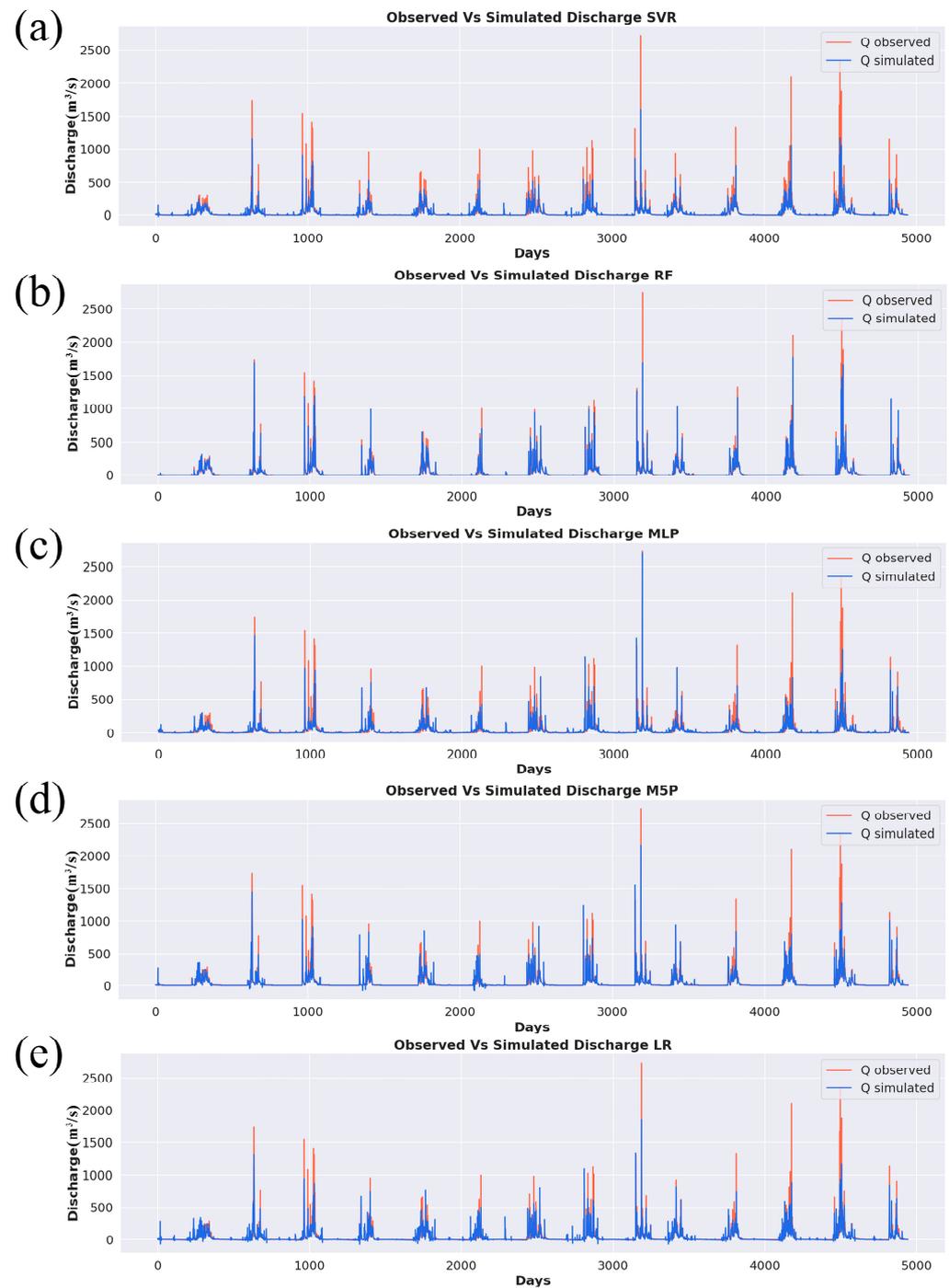


Figure 3. Line plot for observed vs. simulated streamflow for (a) SVR, (b) RF, (c) MLP, (d) M5P, and (e) LR during training.

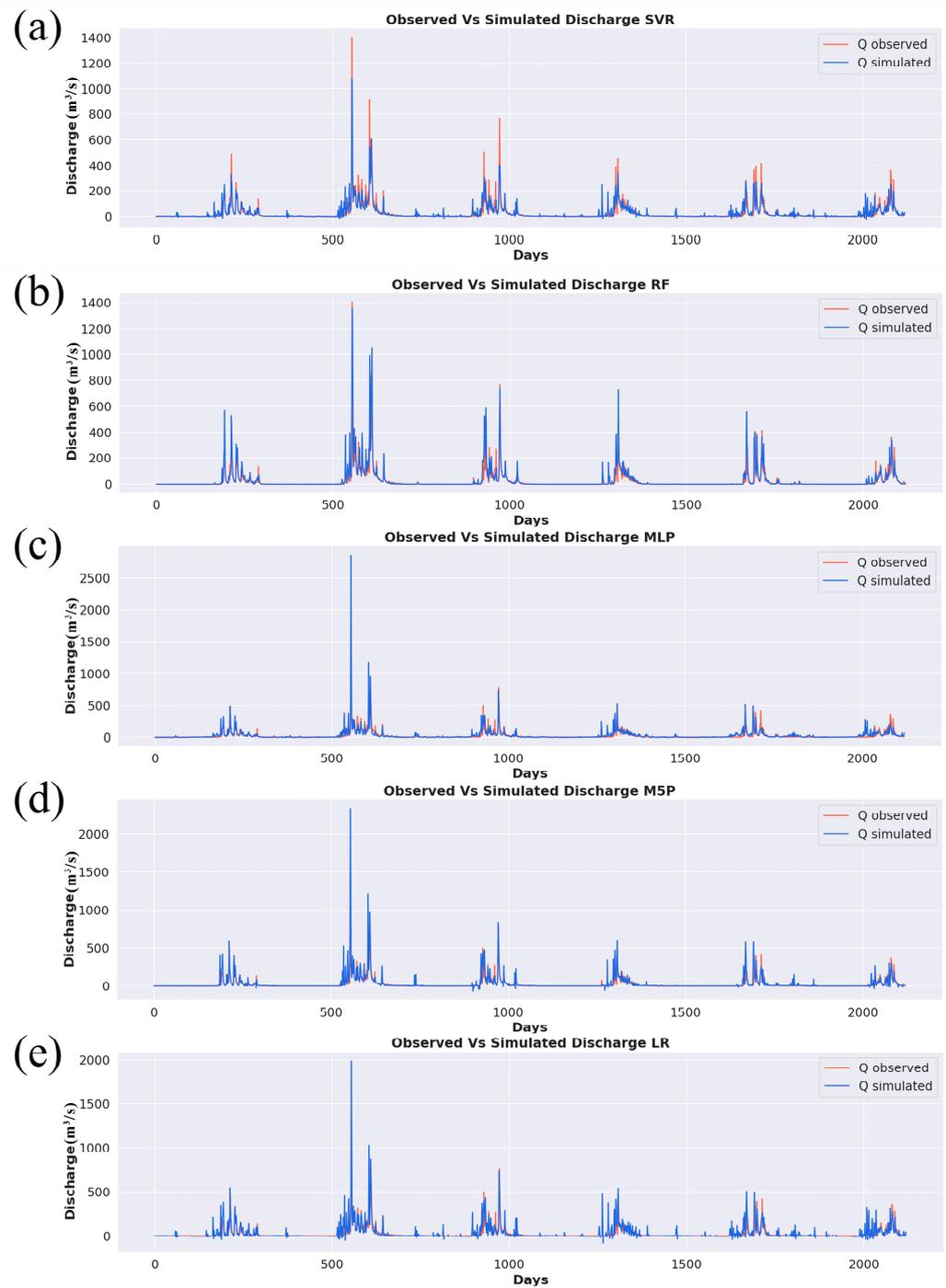


Figure 4. Line plot for observed vs. simulated streamflow for (a) SVR, (b) RF, (c) MLP, (d) M5P, and (e) LR during testing.

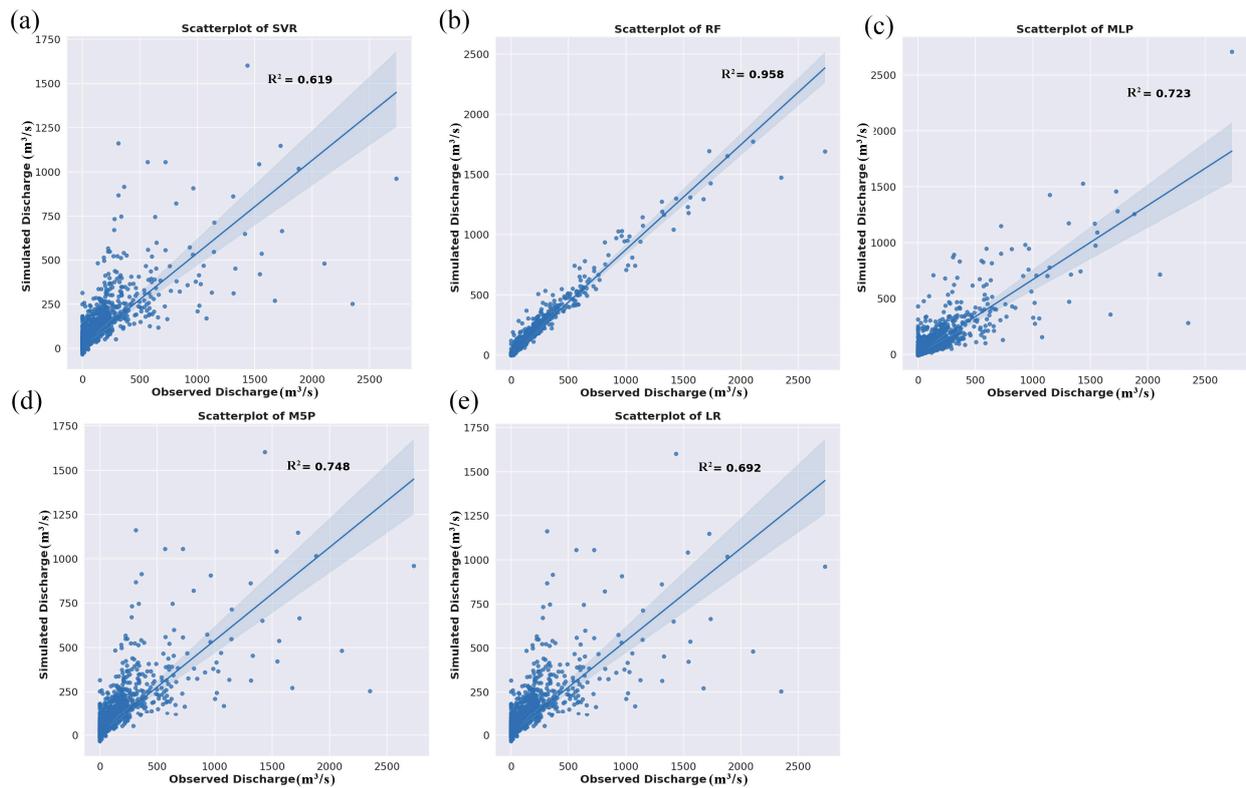


Figure 5. Scatter plot for observed vs. simulated streamflow for (a) SVR, (b) RF, (c) MLP, (d) M5P, and (e) LR during training.

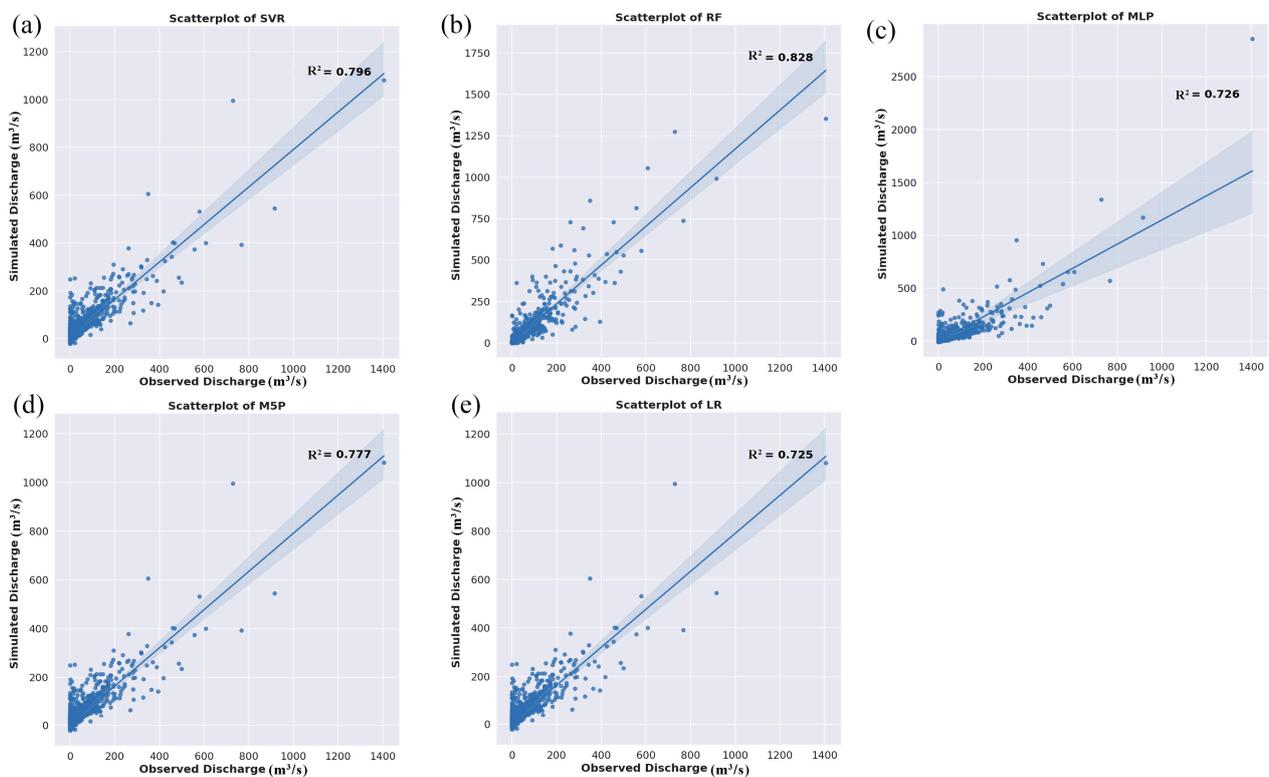


Figure 6. Scatter plot for observed vs. simulated streamflow for (a) SVR, (b) RF, (c) MLP, (d) M5P, and (e) LR during testing.

As seen in Figures 3 and 4, RF performed the best in matching the hydrograph pattern, especially in the remaining testing models, i.e., SVR underestimated peak flows, and MLP, M5P, and LR overestimated the peak flows. Still, RF captures all the peak flows similarly to the observed hydrograph. Similarly, Figures 5 and 6 represent RF performing outstandingly in capturing the streamflow with an R^2 of 0.959 and 0.829 during training and testing. In training, RF is the best model, followed by M5P, MLP, LR, and SVR, with an R^2 of 0.748, 0.723, 0.692, and 0.619. Even during testing, RF is best-performing model in terms of R^2 followed by SVR, M5P, MLP, and LR, with values of 0.796, 0.778, 0.743, and 0.725.

Figures 7 and 8 represent the radar chart during training and testing using IMD gridded precipitation as input data. In Figures 7a and 8a, both NSE and R are mapped; in Figures 7b and 8b, RMSE is plotted in a radar chart. Figure 7a clearly demonstrates RF performing best, with a maximum value of NSE and R compared to other models. Figure 7b shows that a minimum RMSE was observed in the RF model, with a value of $30.805 \text{ m}^3/\text{s}$. Similarly, during testing, Figure 8a,b exhibit both RF and SVR performing better in terms of NSE, R, and RMSE. RMSE is $41.237 \text{ m}^3/\text{s}$ in RF and $33.027 \text{ m}^3/\text{s}$ in SVR in testing.

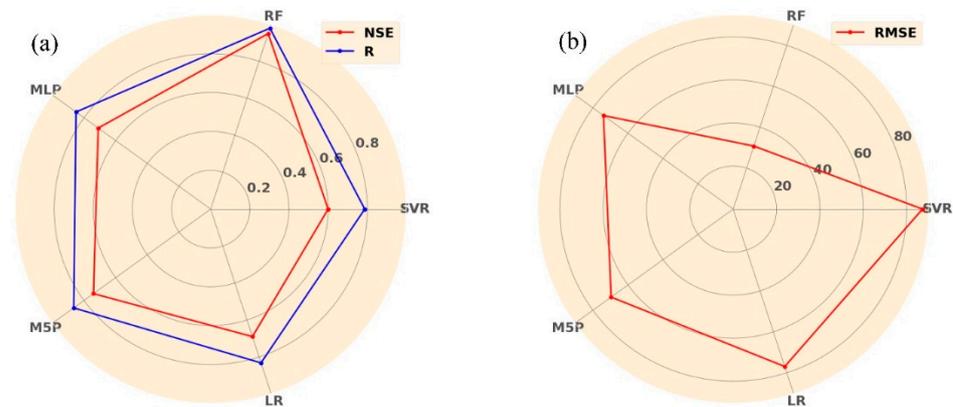


Figure 7. Radar plot during training (a) NSE and R, (b) RMSE.

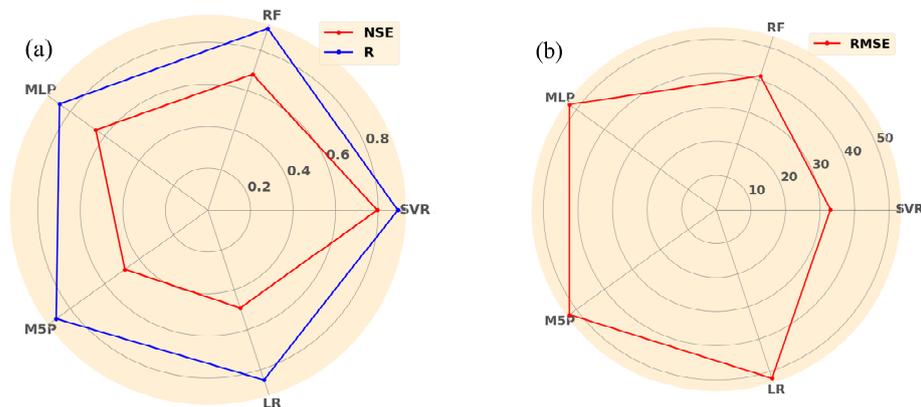


Figure 8. Radar plot during testing (a) NSE and R, (b) RMSE.

The violin plots seen in Figure 9a,b were designed for both training and testing using IMD gridded precipitation as input. For each model, violin plots were created for the interquartile range that was less than 95%, with the higher extreme flow values left out. RF was the best model in which the simulated streamflow displayed flow behavior that was more similar to the flow data of the actual streamflow than the other four models.

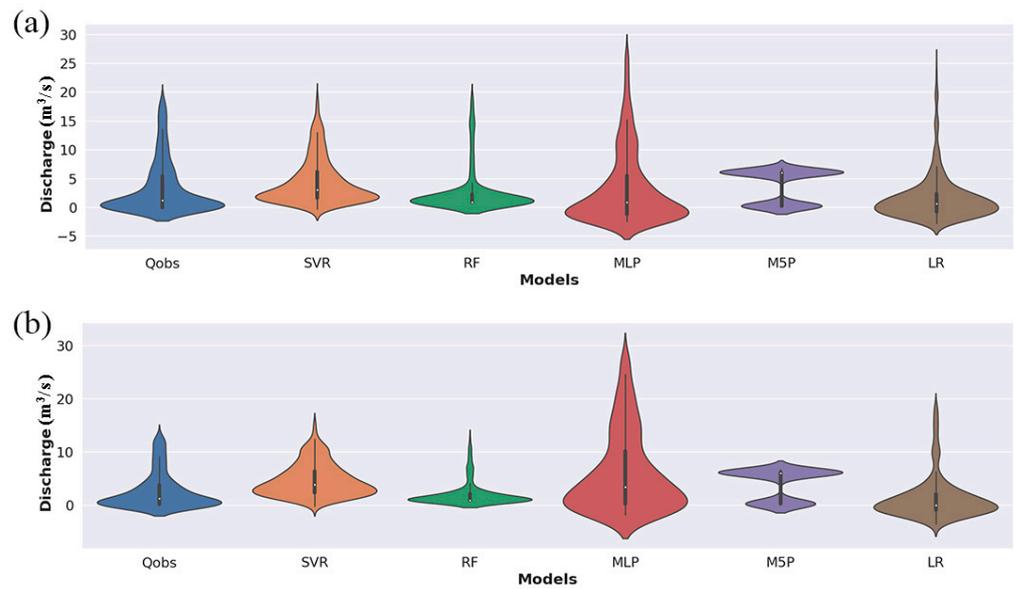


Figure 9. Violin plots during (a) training and (b) testing periods.

Figures 10–15 represent the Taylor diagrams of all five models using different precipitation datasets, i.e., EC-Earth3, EC-Earth3-Veg, GFDL-ESM4, IMD, MIROC6, and MRI-ESM2-0. It is abundantly evident in the Taylor diagrams that the results mentioned before are validated. The training and testing results indicate that RF is the model that performs the best in all scenarios. IMD is the best-performing precipitation dataset compared to the other CMIP6 datasets, making it the ideal choice for modeling streamflow.

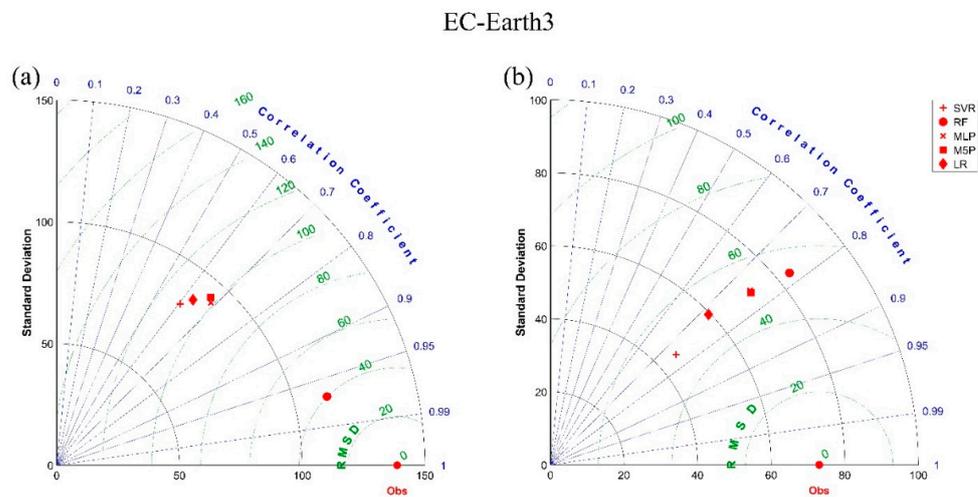


Figure 10. Taylor diagrams during (a) training and (b) testing using EC-Earth3 dataset.

EC-Earth3-Veg

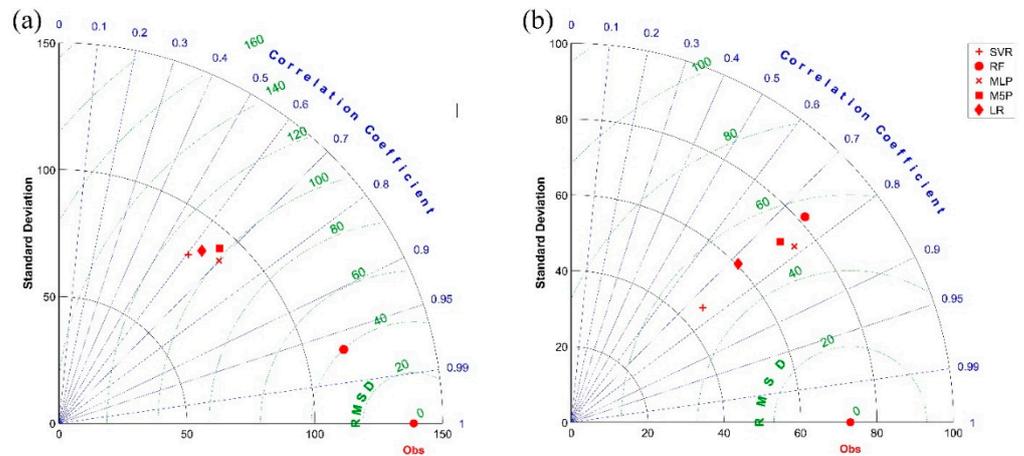


Figure 11. Taylor diagrams during (a) training and (b) testing using EC-Earth3-Veg dataset.

GFDL-ESM4

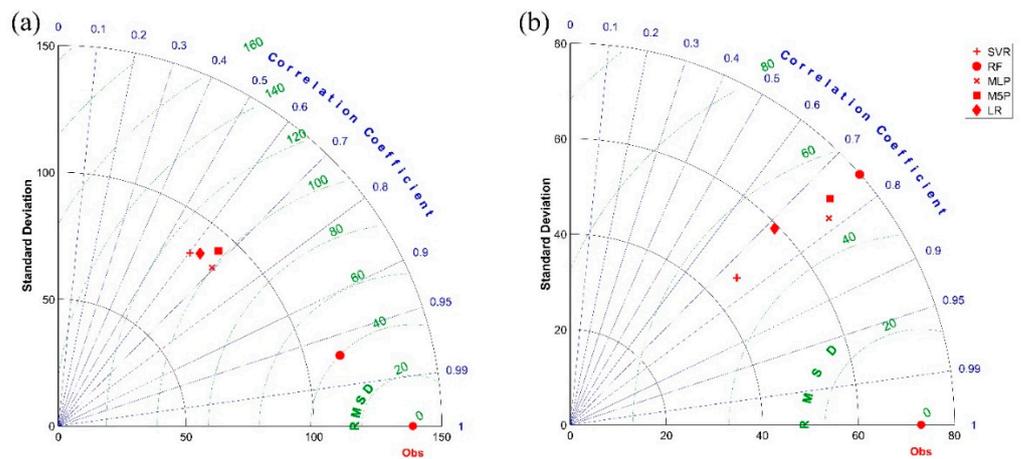


Figure 12. Taylor diagrams during (a) training and (b) testing using GFDL-ESM4 dataset.

IMD

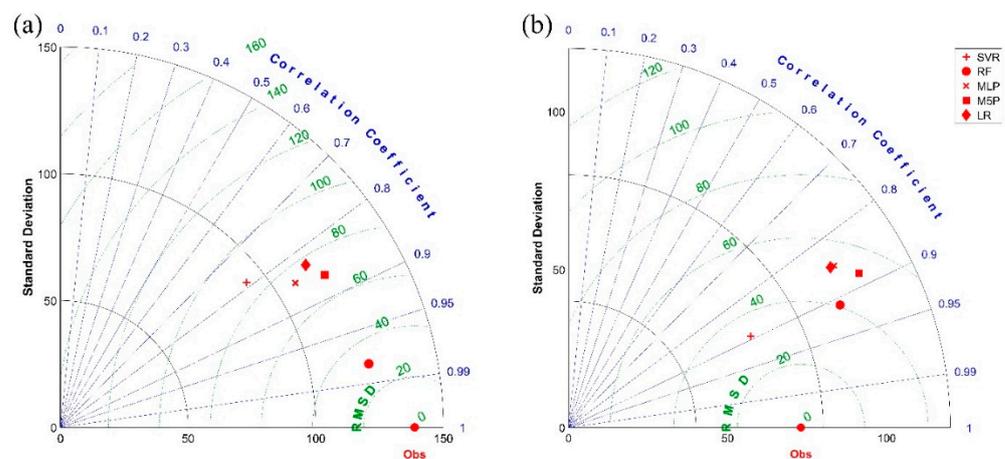


Figure 13. Taylor diagrams during (a) training and (b) testing using IMD dataset.

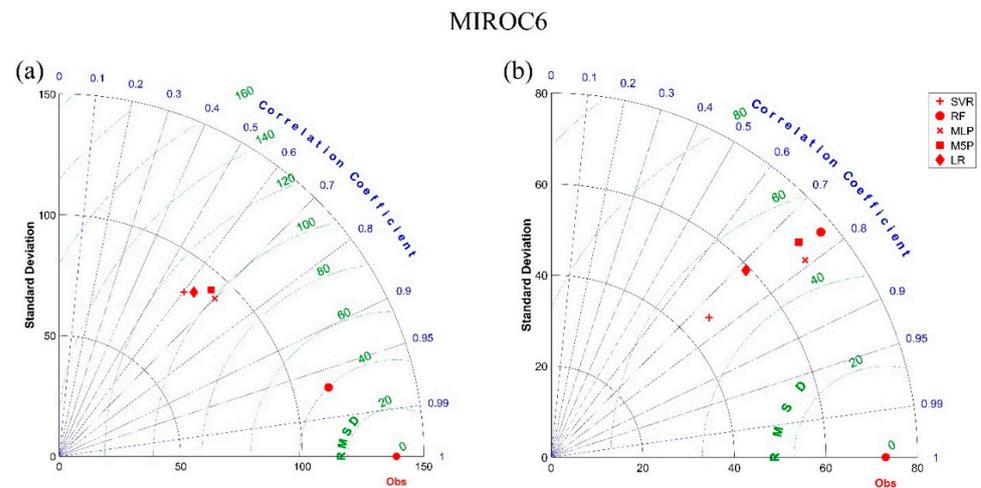


Figure 14. Taylor diagrams during (a) training and (b) testing using MIROC6 dataset.

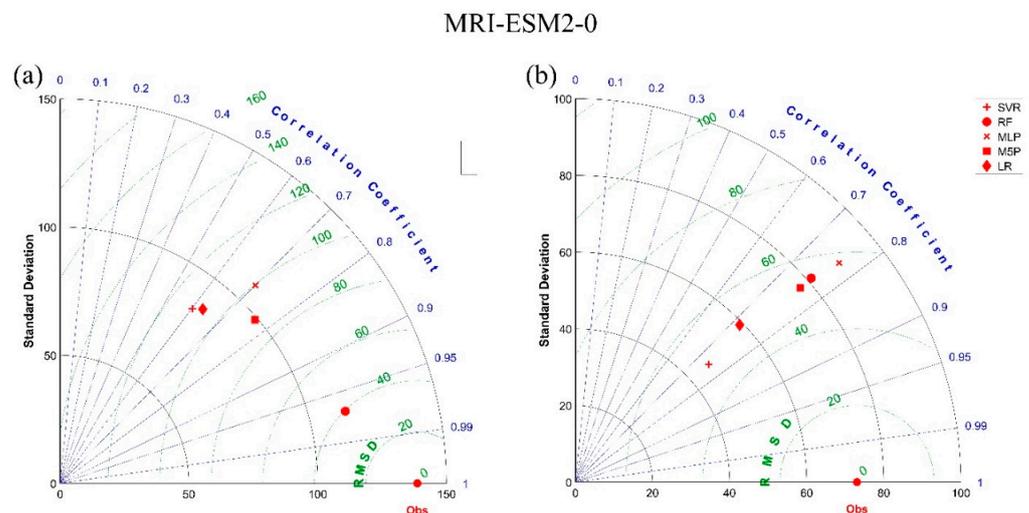


Figure 15. Taylor diagrams during (a) training and (b) testing using MRI-ESM2-0 dataset.

5. Discussion

In this study, the applicability of CMIP6 precipitation datasets for simulating streamflow were assessed with the IMD using five different models, i.e., SVR, RF, MLP, M5P, and LR. During the training and testing phases, time-lagged streamflow observations, lagged precipitation datasets, minimum temperature, and maximum temperature were used as model inputs, and each method was analyzed for its efficiency. In most cases, the error variance between the observed and simulated values was used to evaluate the correctness of the model using metrics like R^2 , NSE, RMSE, R, MAE, MBE, and so on, as utilized in earlier research [68,69,71,78]. From previous studies, only precipitation data as input are insufficient to simulate streamflow. Therefore, the present study included a lag in the streamflow and temperature [68,79,80]. Compared to all the CMIP6 datasets, IMD performs best in terms of all evaluation metrics. When considering models, RF best predicted 1-day streamflow simulation in both CMIP6 and IMD datasets. Metrics such as NSE, R, R^2 , and RMSE were observed to be 0.95, 0.979, 0.937, and 30.805 m^3/s and 0.681, 0.91, 0.828, and 41.237 m^3/s during training and testing using IMD gridded precipitation dataset as input for RF model development. These findings agree with many other studies found: In general, RF has superior performance. [28,32]. A similar type of result was obtained in previous studies on Indian river basins by Kumar et al. [81], concluding that RF and LR outperform other models, such as MLP and ANN, in simulating river discharge

prediction. Hussain and Khan [78] conducted a study in Pakistan to simulate monthly streamflow forecasts and concluded that RF outperformed SVR and MLP. A study carried out by Essam et al. [82] over various river basins in Malaysia identified that ANN performs best in predicting daily streamflow values when compared to SVM and LSTM. One more study conducted in Malaysia by Muhammed et al. [83] concluded that RF-based models performed the best compared to LS-SVM and other M5P models, which supports the results obtained in this study. As part of their investigation on streamflow forecasting, Gianni Vesuviano et al. [84] conducted a study in the Wairagarh catchment using a lumped sub-catchment modeling approach with a single parameter set, which resulted in an NSE value of 0.172 and an R of 0.472. In contrast, our study implemented five machine learning models (SVR, RF, M5P, MLP, and LR) for one-day-ahead streamflow forecasting, with the RF model utilizing IMD gridded precipitation data as input. Our developed RF model demonstrated significantly improved performance, with an NSE value of 0.95 and R of 0.979. These results highlight the superiority of our machine learning models over the lumped sub-catchment modeling approach, offering more accurate and reliable streamflow predictions for the Wairagarh station.

Even for long-term datasets, RF performs far better than ANN, SVM, and boosted tree regression (BTR) [85]. At the same time, compared to conceptual hydrological models (AWBM and Sacramento), AI models perform best in predicting daily streamflow [54]. In addition, Contreras et al. [86] employed RF for 4, 12, and 24 h, and they said that the proposed RF models achieved an excellent result in discharge forecasting with minimal statistical errors. Their discoveries have the potential to be helpful in the development of fully operational early warning devices. Also, the results of this study correlate with those found by Peng et al. [87], who revealed that RF outperformed the BP neural network and the SVM in terms of accurate prediction and computation time while working with complicated and nonlinear hydrological models. Our results, supported by Li et al. [27], explain that RF captures peak flows better than other machine learning models such as ELM-kernel, BPNN, and SVR.

This is supported by the fact that the RF performed better in both of these methods. The model assessment results reveal that the RF performs significantly better in basins controlled by snowmelt than in basins driven by rainfall [88]. One more study by Singh et al. [89] supported that RF exhibits strong potential for simulating streamflow over the Himalayan catchment in India compared to MLR, MARS, and SVM. Even for medium- and long-term runoff forecasting, RF performs best compared to SVM and IARMA [90]. Compared to neural networks and SVM, the RF model offers greater prediction accuracy and requires less computation when working with highly nonlinear hydrological time series, when considering monthly streamflow simulations [87]. Not only for streamflow modeling, but RF has also been applied in various studies like predicting total nitrogen (TN), total suspended solids (TSS), total phosphorus (TP), and ortho-phosphorus (Ortho-P) EMCs in urban runoff [91].

There are several limitations attached to machine learning models. The location is a limitation of the above optimal model (the RF model). Since the RF model was trained using data from the Wairagarh catchment, it is more likely to produce correct findings when applied to other catchments. The significant degree of randomness in the streamflow pattern has necessitated the application of several machine learning algorithms in a variety of geographic areas to locate appropriate models for reliable forecasting. It is, therefore, a continuous challenge to investigate and build an expert model for use in hydrological modeling. If it is used for other catchments, it will need to be retrained on the past data of the concerning catchments.

6. Conclusions

In this study, five models, i.e., SVR, RF, MLP, M5P, and LR, were developed to simulate 1-day-ahead streamflow at Wairagarh station in the Pranhita subbasin (Godavari basin) of India. For this analysis, different precipitation datasets were considered. CMIP6 precipi-

tation datasets were downscaled using the distribution mapping method. Models were developed for 1993–2014, in which 70% of data were used for training, and the remaining 30% were used for testing, after excluding any missing data. The input parameters were chosen using correlation and pairwise correlation attribution evaluation methods. Important takeaways are outlined here:

Both CMIP6 and IMD performed better in streamflow forecasting using lagged data (precipitation and streamflow), minimum temperature, and maximum temperature as input.

Using CMIP6 datasets as input, RF and M5P performed very well according to different evaluation metrics. RF showed very good ($0.75 < \text{NSE} < 1$ and $0.7 < R^2 < 1$) performance in training and acceptable ($0.4 < \text{NSE} < 0.50$ and $0.5 < R^2 < 0.6$) performance in testing. Similarly, M5P represented a satisfactory ($0.4 < \text{NSE} < 0.50$ and $0.5 < R^2 < 0.6$) performance in both training and testing. For CMIP6 input precipitation dataset is found to be MRI-ESM2-0 for the M5P model and MIROC6 for the RF model.

Compared to downscaled CMIP6 precipitation datasets, IMD outperformed all the models in evaluation metrics. In comparison with all five models, RF outperformed the others, with NSE, R, R^2 , and RMSE values of 0.95, 0.979, 0.937, and $30.805 \text{ m}^3/\text{s}$ and 0.681, 0.91, 0.828, and $41.237 \text{ m}^3/\text{s}$ during training and testing, respectively. RF showed the best performance in evaluation metrics and in capturing peak flow events and hydrograph patterns in both training and testing.

Overall, the best-performing models in forecasting streamflow one day in advance when using IMD gridded precipitation as input are ranked in the following order: RF, SVR, M5P, MLP, and finally LR. However, the last two methods exhibited very poor performance for the chosen study area.

The findings of this study hold crucial implications for water resource management and hydrological research. The accurate streamflow forecasting models developed using advanced machine learning algorithms can empower decisionmakers with better water planning strategies, flood control, and drought management. Incorporating multiple gridded satellite precipitation datasets and bias-corrected CMIP6 data enhances the understanding of climate change impacts on hydrological processes. However, limitations exist, such as data availability, model generalization, and uncertainties in climate models. Future research can explore ensemble machine learning modeling, real-time streamflow predictions, and risk assessment studies. Additionally, efforts can be directed toward addressing hydrological complexities and refining model validation techniques. By overcoming these limitations and pursuing further research, the field of streamflow forecasting can advance, contributing to sustainable water management and preparedness for water-related challenges worldwide.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su151612295/s1>, Figure S1: Schematic diagram of SVR; Figure S2: Schematic diagram of RF; Figure S3: Schematic diagram of MLP.

Author Contributions: Conceptualization, S.S., N.M.R., Q.B.P., H.A. and A.A.A.D.; Methodology, S.S. and N.M.R.; Software, N.M.R. and Q.B.P.; Validation, S.S., N.M.R. and Q.B.P.; Formal analysis, S.S., N.M.R., Q.B.P. and A.A.; Investigation, H.G.A. and H.A.; Resources, H.G.A.; Data curation, H.G.A.; Writing—original draft, S.S.; Writing—review & editing, A.A., H.G.A., H.A. and A.A.A.D.; Visualization, A.A.; Supervision, A.A. and A.A.A.D.; Project administration, H.A.; Funding acquisition, A.A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: Researchers would like to thank the Deanship of Scientific Research, Qassim University for funding publication of this project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the author, Quoc Bao Pham, quoc_bao.pham@us.edu.pl, upon reasonable request.

Conflicts of Interest: There is no conflict of interest to declare.

References

- Xu, Z.P.; Li, Y.P.; Huang, G.H.; Wang, S.G.; Liu, Y.R. A multi-scenario ensemble streamflow forecast method for Amu Darya River Basin under considering climate and land-use changes. *J. Hydrol.* **2021**, *598*, 126276. [[CrossRef](#)]
- Brunner, M.I.; Slater, L.; Tallaksen, L.M.; Clark, M. Challenges in modeling and predicting floods and droughts: A review. *Wiley Interdiscip. Rev. Water* **2021**, *8*, e1520. [[CrossRef](#)]
- Abdulkadir, T.S.; Salami, A.W.; Anwar, A.R.; Kareem, A.G. Modelling of hydropower reservoir variables for energy generation: Neural network approach. *Ethiop. J. Environ. Stud. Manag.* **2013**, *6*, 310–316. [[CrossRef](#)]
- Nazmi, N.; Rahman, M.A.A.; Yamamoto, S.-I.; Ahmad, S.A. Walking gait event detection based on electromyography signals using artificial neural network. *Biomed. Signal Process. Control* **2019**, *47*, 334–343. [[CrossRef](#)]
- Ali, S.; Shahbaz, M. Streamflow forecasting by modeling the rainfall–streamflow relationship using artificial neural networks. *Model. Earth Syst. Environ.* **2020**, *6*, 1645–1656. [[CrossRef](#)]
- Bayram, A.; Kankal, M.; Tayfur, G.; Önsöy, H. Prediction of suspended sediment concentration from water quality variables. *Neural Comput. Appl.* **2014**, *24*, 1079–1087. [[CrossRef](#)]
- Sanikhani, H.; Kisi, O. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. *Water Resour. Manag.* **2012**, *26*, 1715–1729. [[CrossRef](#)]
- Minns, A.W.; Hall, M.J. Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.* **1996**, *41*, 399–417. [[CrossRef](#)]
- Kote, A.S.; Jothiprakash, V. Reservoir inflow prediction using time lagged recurrent neural networks. In Proceedings of the 2008 First International Conference on Emerging Trends in Engineering and Technology (IEEE), Nagpur, India, 16–18 July 2008; pp. 618–623.
- Cancelliere, A.; Giuliano, G.; Ancarani, A.; Rossi, G. A neural networks approach for deriving irrigation reservoir operating rules. *Water Resour. Manag.* **2002**, *16*, 71–88. [[CrossRef](#)]
- Uzlu, E.; Akpınar, A.; Kömürçü, M.İ. Restructuring of Turkey’s electricity market and the share of hydropower energy: The case of the Eastern Black Sea Basin. *Renew. Energy* **2011**, *36*, 676–688. [[CrossRef](#)]
- Kişi, Ö. Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *J. Hydrol. Eng.* **2009**, *14*, 773–782. [[CrossRef](#)]
- Shiri, J.; Kisi, O. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *J. Hydrol.* **2010**, *394*, 486–493. [[CrossRef](#)]
- Imrie, C.E.; Durucan, S.; Korre, A. River flow prediction using artificial neural networks: Generalisation beyond the calibration range. *J. Hydrol.* **2000**, *233*, 138–153. [[CrossRef](#)]
- Coulibaly, P.; Anctil, F.; Bobée, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J. Hydrol.* **2000**, *230*, 244–257. [[CrossRef](#)]
- Hadi, S.J.; Tombul, M. Forecasting Daily Streamflow for Basins with Different Physical Characteristics through Data-Driven Methods. *Water Resour. Manag.* **2018**, *32*, 3405–3422. [[CrossRef](#)]
- Latifoğlu, L. A novel approach for prediction of daily streamflow discharge data using correlation based feature selection and random forest method. *Int. Adv. Res. Eng. J.* **2022**, *6*, 1–7. [[CrossRef](#)]
- Huang, S.; Chang, J.; Huang, Q.; Chen, Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J. Hydrol.* **2014**, *511*, 764–775. [[CrossRef](#)]
- Sedighi, F.; Vafakhah, M.; Javadi, M.R. Rainfall–runoff modeling using support vector machine in snow-affected watershed. *Arab. J. Sci. Eng.* **2016**, *41*, 4065–4076. [[CrossRef](#)]
- Ghorbani, M.A.; Zadeh, H.A.; Isazadeh, M.; Terzi, O. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ. Earth Sci.* **2016**, *75*, 476. [[CrossRef](#)]
- Ghorbani, M.A.; Khatibi, R.; Karimi, V.; Yaseen, Z.M.; Zounemat-Kermani, M. Learning from multiple models using artificial intelligence to improve model prediction accuracies: Application to river flows. *Water Resour. Manag.* **2018**, *32*, 4201–4215. [[CrossRef](#)]
- Ghorbani, M.A.; Deo, R.C.; Karimi, V.; Yaseen, Z.M.; Terzi, O. Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 1683–1697. [[CrossRef](#)]
- Alizadeh, F.; Gharamaleki, A.F.; Jalilzadeh, M.; Akhoundzadeh, A. Prediction of river stage-discharge process based on a conceptual model using EEMD-WT-LSSVM approach. *Water Resour.* **2020**, *47*, 41–53. [[CrossRef](#)]
- Ghorbani, M.A.; Khatibi, R.; Goel, A.; FazeliFard, M.H.; Azani, A. Modeling river discharge time series using support vector machine and artificial neural networks. *Environ. Earth Sci.* **2016**, *75*, 685. [[CrossRef](#)]
- Lin, J.-Y.; Cheng, C.-T.; Chau, K.-W. Using support vector machines for long-term discharge prediction. *Hydrol. Sci. J.* **2006**, *51*, 599–612. [[CrossRef](#)]
- Seyam, M.; Othman, F.; El-Shafie, A. Prediction of stream flow in humid tropical rivers by support vector machines. *MATEC Web Conf.* **2017**, *111*, 1007. [[CrossRef](#)]
- Li, X.; Sha, J.; Wang, Z.-L. Comparison of daily streamflow forecasts using extreme learning machines and the random forest method. *Hydrol. Sci. J.* **2019**, *64*, 1857–1866. [[CrossRef](#)]

28. Papacharalampous, G.A.; Tyralis, H. Evaluation of random forests and Prophet for daily~streamflow~forecasting. *Adv. Geosci.* **2018**, *45*, 201–208. [[CrossRef](#)]
29. Sammen, S.S.; Ehteram, M.; Abba, S.I.; Abdulkadir, R.A.; Ahmed, A.N.; El-Shafie, A. A new soft computing model for daily streamflow forecasting. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2479–2491. [[CrossRef](#)]
30. Mohammadi, B.; Ahmadi, F.; Mehdizadeh, S.; Guan, Y.; Pham, Q.B.; Linh, N.T.T.; Tri, D.Q. Developing Novel Robust Models to Improve the Accuracy of Daily Streamflow Modeling. *Water Resour. Manag.* **2020**, *34*, 3387–3409. [[CrossRef](#)]
31. Kambalimath S, S.; Deka, P.C. Performance enhancement of SVM model using discrete wavelet transform for daily streamflow forecasting. *Environ. Earth Sci.* **2021**, *80*, 101. [[CrossRef](#)]
32. Bajirao, T.S.; Elbeltagi, A.; Kumar, M.; Pham, Q.B. Applicability of machine learning techniques for multi-time step ahead runoff forecasting. *Acta Geophys.* **2022**, *7*, 757–776. [[CrossRef](#)]
33. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhau, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [[CrossRef](#)] [[PubMed](#)]
34. Jozaghi, A.; Shen, H.; Ghazvinian, M.; Seo, D.-J.; Zhang, Y.; Welles, E.; Reed, S. Multi-model streamflow prediction using conditional bias-penalized multiple linear regression. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2355–2373. [[CrossRef](#)]
35. Brown, J.D.; Wu, L.; He, M.; Regonda, S.; Lee, H.; Seo, D.-J. Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 1. Experimental design and forcing verification. *J. Hydrol.* **2014**, *519*, 2869–2889. [[CrossRef](#)]
36. Fahimi, F.; Yaseen, Z.M.; El-shafie, A. Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review. *Theor. Appl. Climatol.* **2017**, *128*, 875–903. [[CrossRef](#)]
37. Quinlan, J.R.; Adams, A.; Sterling, L. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; pp. 343–348.
38. Bhattacharya, B.; Solomatine, D.P. Neural networks and M5 model trees in modelling water level–discharge relationship. *Neurocomputing* **2005**, *63*, 381–396. [[CrossRef](#)]
39. Onyari, E.K.; Ilunga, F.M. Application of MLP neural network and M5P model tree in predicting streamflow: A case study of Luvuvhu catchment, South Africa. *Int. J. Innov. Manag. Technol.* **2013**, *4*, 11.
40. Sihag, P.; Sadikhani, M.R.; Vambol, V.; Vambol, S.; Prabhakar, A.K.; Sharma, N. Comparative study for deriving stagedischarge–sediment concentration relationships using soft computing techniques. *J. Achiev. Mater. Manuf. Eng.* **2021**, *104*, 57–76. [[CrossRef](#)]
41. Reddy, B.S.N.; Pramada, S.K.; Roshni, T. Monthly surface runoff prediction using artificial intelligence: A study from a tropical climate river basin. *J. Earth Syst. Sci.* **2021**, *130*, 35. [[CrossRef](#)]
42. Kumar, A.; Singh, R.; Jena, P.P.; Chatterjee, C.; Mishra, A. Identification of the best multi-model combination for simulating river discharge. *J. Hydrol.* **2015**, *525*, 313–325. [[CrossRef](#)]
43. Vojtek, M.; Vojteková, J.; Costache, R.; Pham, Q.B.; Lee, S.; Arshad, A.; Sahoo, S.; Linh, N.T.T.; Anh, D.T. Comparison of multi-criteria-analytical hierarchy process and machine learning-boosted tree models for regional flood susceptibility mapping: A case study from Slovakia. *Geomatics, Nat. Hazards Risk* **2021**, *12*, 1153–1180. [[CrossRef](#)]
44. Pörtner, H.-O.; Roberts, D.C.; Adams, H.; Adler, C.; Aldunce, P.; Ali, E.; Begum, R.A.; Betts, R.; Kerr, R.B.; Biesbroek, R.; et al. Climate change 2022: Impacts, adaptation and vulnerability. In *IPCC Sixth Assessment Report*; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2022.
45. Chinasho, A.; Yaya, D.; Tessema, S. The adaptation and mitigation strategies for climate change in pastoral communities of Ethiopia. *Am. J. Environ. Prot.* **2017**, *6*, 69. [[CrossRef](#)]
46. Stouffer, R.J.; Eyring, V.; Meehl, G.A.; Bony, S.; Senior, C.; Stevens, B.; Taylor, K.E. CMIP5 Scientific Gaps and Recommendations for CMIP6. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 95–105. [[CrossRef](#)]
47. Sashidharan, K.; Mohanty, A.K.; Gupta, A. A note on diamond incidence in Wairagarh area, Garhchiroli district, Maharashtra. *Geol. Soc. India* **2002**, *59*, 265–268.
48. Roy, P.S.; Meiyappan, P.; Joshi, P.K.; Kale, M.P.; Srivastav, V.K.; Srivasatava, S.K.; Behera, M.D.; Roy, A.; Sharma, Y.; Ramachandran, R.M.; et al. *Decadal Land Use and Land Cover Classifications across India, 1985, 1995, 2005*; ORNL DAAC: Oak Ridge, TN, USA, 2016.
49. Merufinia, E.; Sharafati, A.; Abghari, H.; Hassanzadeh, Y. On the simulation of streamflow using hybrid tree-based machine learning models: A case study of Kurkursar basin, Iran. *Arab. J. Geosci.* **2022**, *16*, 28. [[CrossRef](#)]
50. Pai, D.; Sridhar, L.; Rajeevan, M.; Sreejith, O.P.; Satbhai, N.S.; Mukhopadhyay, B. Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. *Mausam* **2014**, *65*, 1–18. [[CrossRef](#)]
51. Reddy, N.M.; Saravanan, S. Evaluation of the accuracy of seven gridded satellite precipitation products over the Godavari River basin, India. *Int. J. Environ. Sci. Technol.* **2022**. [[CrossRef](#)]
52. Almazroui, M.; Saeed, F.; Saeed, S.; Ismail, M.; Ehsan, M.A.; Islam, M.N.; Abid, M.A.; O'Brien, E.; Kamil, S.; Rashid, I.U.; et al. Projected Changes in Climate Extremes Using CMIP6 Simulations Over SREX Regions. *Earth Syst. Environ.* **2021**, *5*, 481–497. [[CrossRef](#)]
53. Reddy, N.M.; Saravanan, S. Extreme precipitation indices over India using CMIP6: A special emphasis on the SSP585 scenario. *Environ. Sci. Pollut. Res.* **2023**, *30*, 47119–47143. [[CrossRef](#)]

54. Reddy, N.M.; Saravanan, S.; Abijith, D. Streamflow simulation using conceptual and neural network models in the Hemavathi sub-watershed, India. *Geosyst. Geoenviron.* **2023**, *2*, 100153. [[CrossRef](#)]
55. Schulzweida, U.; Kronblueh, L.; Budich, R.G. CDO: Climate Data Operators: Version 1.8.1. 2019. Available online: <https://code.mpimet.mpg.de/news/369> (accessed on 25 June 2023).
56. Bandyopadhyay, A.; Nengzouzam, G.; Singh, W.R.; Hangsing, N.; Bhadra, A. Comparison of various re-analyses gridded data with observed data from meteorological stations over India. *Epic Ser. Eng.* **2018**, *3*, 190–198.
57. Smitha, P.S.; Narasimhan, B.; Sudheer, K.P.; Annamalai, H. An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. *J. Hydrol.* **2018**, *556*, 100–118. [[CrossRef](#)]
58. Chen, J.; Brissette, F.P.; Chaumont, D.; Braun, M. Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.* **2013**, *49*, 4187–4205. [[CrossRef](#)]
59. Ravansalar, M.; Rajaei, T. Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environ. Monit. Assess.* **2015**, *187*, 366. [[CrossRef](#)]
60. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Networks* **1999**, *10*, 988–999. [[CrossRef](#)]
61. Pinthong, S.; Ditthakit, P.; Salaeh, N.; Hasan, M.A.; Son, C.T.; Linh, N.T.T.; Islam, S.; Yadav, K.K. Imputation of missing monthly rainfall data using machine learning and spatial interpolation approaches in Thale Sap Songkhla River Basin, Thailand. *Environ. Sci. Pollut. Res.* **2022**, *Online ahead of print*.
62. Xu, C.; Dai, F.; Xu, X.; Lee, Y.H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145*, 70–80. [[CrossRef](#)]
63. Li, Y.H.; Xu, J.Y.; Tao, L.; Li, X.F.; Li, S.; Zeng, X.; Chen, S.Y.; Zhang, P.; Qin, C.; Zhang, C. SVM-Prot 2016: A web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PLoS ONE* **2016**, *11*, e0155290. [[CrossRef](#)]
64. Das, J.; Nanduri, U.V. Assessment and evaluation of potential climate change impact on monsoon flows using machine learning technique over Wainganga River basin, India. *Hydrol. Sci. J.* **2018**, *63*, 1020–1046. [[CrossRef](#)]
65. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
66. Boehmke, B.; Greenwell, B. *Hands-On Machine Learning with R*; CRC Press: Boca Raton, FL, USA, 2019.
67. Shiau, J.-T.; Hsu, H.-T. Suitability of ANN-based daily streamflow extension models: A case study of Gaoping River basin, Taiwan. *Water Resour. Manag.* **2016**, *30*, 1499–1513. [[CrossRef](#)]
68. Poonia, V.; Tiwari, H.L. Rainfall-runoff modeling for the Hoshangabad Basin of Narmada River using artificial neural network. *Arab. J. Geosci.* **2020**, *13*, 944. [[CrossRef](#)]
69. Bajirao, T.S.; Kumar, P.; Kumar, M.; Elbeltagi, A.; Kuriqi, A. Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theor. Appl. Climatol.* **2021**, *145*, 1207–1231. [[CrossRef](#)]
70. Sharma, P.; Machiwal, D. Streamflow forecasting: Overview of advances in data-driven techniques. *Adv. Streamflow Forecast.* **2021**, 1–50. [[CrossRef](#)]
71. Sharma, P.; Madane, D.; Bhakar, S.R. Monthly streamflow forecasting using artificial intelligence approach: A case study in a semi-arid region of India. *Arab. J. Geosci.* **2021**, *14*, 2440. [[CrossRef](#)]
72. Tabari, H.; Sabziparvar, A.-A.; Ahmadi, M. Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorol. Atmos. Phys.* **2011**, *110*, 135–142. [[CrossRef](#)]
73. Özbayoğlu, G.; Evren Özbayoğlu, M. A new approach for the prediction of ash fusion temperatures: A case study using Turkish lignites. *Fuel* **2006**, *85*, 545–552. [[CrossRef](#)]
74. Khazaei Poul, A.; Shourian, M.; Ebrahimi, H. A Comparative Study of MLR, KNN, ANN and ANFIS Models with Wavelet Transform in Monthly Stream Flow Prediction. *Water Resour. Manag.* **2019**, *33*, 2907–2923. [[CrossRef](#)]
75. Li, P.-H.; Kwon, H.-H.; Sun, L.; Lall, U.; Kao, J.-J. A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan. *Int. J. Climatol.* **2010**, *30*, 1256–1268. [[CrossRef](#)]
76. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
77. Faizollahzadeh Ardabili, S.; Najafi, B.; Alizamir, M.; Mosavi, A.; Shamshirband, S.; Rabczuk, T. Using SVM-RSM and ELM-RSM Approaches for Optimizing the Production Process of Methyl and Ethyl Esters. *Energies* **2018**, *11*, 2889. [[CrossRef](#)]
78. Hussain, D.; Khan, A.A. Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth Sci. Inform.* **2020**, *13*, 939–949. [[CrossRef](#)]
79. Almazroui, M.; Ashfaq, M.; Islam, M.N.; Rashid, I.U.; Kamil, S.; Abid, M.A.; O'Brien, E.; Ismail, M.; Reboita, M.S.; Sörensson, A.A.; et al. Assessment of CMIP6 Performance and Projected Temperature and Precipitation Changes Over South America. *Earth Syst. Environ.* **2021**, *5*, 155–183. [[CrossRef](#)]
80. Mutlu, E.; Chaubey, I.; Hexmoor, H.; Bajwa, S.G. Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. *Hydrol. Process. Int. J.* **2008**, *22*, 5097–5106. [[CrossRef](#)]
81. Kumar, M.; Elbeltagi, A.; Pande, C.B.; Ahmed, A.N.; Chow, M.F.; Pham, Q.B.; Kumari, A.; Kumar, D. Applications of Data-driven Models for Daily Discharge Estimation Based on Different Input Combinations. *Water Resour. Manag.* **2022**, *36*, 2201–2221. [[CrossRef](#)]
82. Essam, Y.; Huang, Y.F.; Ng, J.L.; Birima, A.H.; Ahmed, A.N.; El-Shafie, A. Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Sci. Rep.* **2022**, *12*, 3883. [[CrossRef](#)] [[PubMed](#)]

83. Muhammed, P.S.; Parveen, S.; Bin, S.A.; Balraj, S.; Bao, P.Q. Time-Series Prediction of Streamflows of Malaysian Rivers Using Data-Driven Techniques. *J. Irrig. Drain. Eng.* **2020**, *146*, 4020013.
84. Vesuviano, G.; Griffin, A.; Stewart, E. Flood Frequency Estimation in Data-Sparse Wainganga Basin, India, Using Continuous Simulation. *Water* **2022**, *14*, 2887. [[CrossRef](#)]
85. Tofiq, Y.M.; Latif, S.D.; Ahmed, A.N.; Kumar, P.; El-Shafie, A. Optimized Model Inputs Selections for Enhancing River Streamflow Forecasting Accuracy Using Different Artificial Intelligence Techniques. *Water Resour. Manag.* **2022**, *36*, 5999–6016. [[CrossRef](#)]
86. Contreras, P.; Orellana-Alvear, J.; Muñoz, P.; Bendix, J.; Célleri, R. Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment. *Atmosphere* **2021**, *12*, 238. [[CrossRef](#)]
87. Peng, F.; Wen, J.; Zhang, Y.; Jin, J. Monthly streamflow prediction based on random forest algorithm and phase space reconstruction theory. *J. Phys. Conf. Ser.* **2020**, *1637*, 12091. [[CrossRef](#)]
88. Pham, Q.B.; Pal, S.C.; Chakraborty, R.; Norouzi, A.; Golshan, M.; Ogunrinde, A.T.; Janizadeh, S.; Khedher, K.M.; Anh, D.T. Evaluation of various boosting ensemble algorithms for predicting flood hazard susceptibility areas. *Geomat. Nat. Hazards Risk* **2021**, *12*, 2607–2628. [[CrossRef](#)]
89. Singh, A.K.; Kumar, P.; Ali, R.; Al-Ansari, N.; Vishwakarma, D.K.; Kushwaha, K.S.; Panda, K.C.; Sagar, A.; Mirzania, E.; Elbeltagi, A.; et al. An Integrated Statistical-Machine Learning Approach for Runoff Prediction. *Sustainability* **2022**, *14*, 8209. [[CrossRef](#)]
90. Shijun, C.; Qin, W.; Yanmei, Z.; Guangwen, M.; Xiaoyan, H.; Liang, W. Medium- and long-term runoff forecasting based on a random forest regression model. *Water Supply* **2020**, *20*, 3658–3664. [[CrossRef](#)]
91. Behrouz, M.S.; Yazdi, M.N.; Sample, D.J. Using Random Forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff. *J. Environ. Manag.* **2022**, *317*, 115412. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.