International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

Kernel And Tree-Based Time Series Forecasting Of Monthly Rainfall Using Large-Scale Climate Indices

Ayush Vashisth[1], Mohammed Sharif [2], Mohammed Shakeel [3]
[1,2,3] Department of Civil Engineering, Jamia Millia Islamia, New Delhi, India

Abstract
Rainfall prediction is vital for various economic activities, including agriculture, sustainable water management, and disaster risk reduction. The objective of the present research is to develop machine learning (ML) models for monthly rainfall forecasting using a combination of large-scale climate indices and local climatic variables. Predictors included the Indian Ocean Dipole (IOD), El Nino Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), North Atlantic Oscillation (NAO), and local maximum and minimum temperatures, while rainfall was as the target variable. Four ML techniques: Multivariate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), Gaussian Process Regression (GPR), and Random Forest (RF) were used to analyse and predict rainfall in Farukh Nagar station in the province of Haryana, India. The rainfall data consisted of 468 monthly observations split into 80% training and 20% testing subsets. To capture nonlinear rainfall climate linkages, Radial Basis Function (RBF) and Pearson VII kernel functions were used in SVM and GPR. Among the models, SVM with the RBF kernel outperformed others, achieving correlation coefficient of 0.811, mean absolute error of 20.44 mm, root mean square error of 49.19 mm, and scattering index of 1.15 during training. The corresponding statistics during testing were 0.718, 24.45 mm, 46.86 mm, and 1.19. The Taylor diagram further confirmed the superiority of SVM-RBF in representing rainfall variability compared with RF, GPR, and MARS. The integration of global climate indices with local temperature variables enhanced model robustness and improved the rainfall forecasting performance. The results highlight the significance of advanced ML approaches, particularly SVM-RBF, in improved rainfall forecasting leading to better agricultural planning and climate-resilient decision-making.
Keyword: *rainfall, prediction, climate, indices, machine, learning*

INTRODUCTION

Rainfall prediction plays a pivotal role in proper planning and management of water resources system. The dynamics nature of climate indices the atmospheric process is nonlinear and spatially diverse for rainfall prediction. The growing dynamic interaction between the climate indices and time lagged are frequently not captured by statistical methods. (Chattopadhyay et al., 2010; Mishra et al., 2020). Traditional statistical methods often fail to capture the dynamic interactions between rainfall and climatic variables, especially when time-lagged responses and large-scale teleconnection patterns are involved.
Rainfall variability is significantly influenced by PDO IOD, NAO, and ENSO (Ropelewski & Halpert, 1987; Saji et al., 1999). Regional rainfall prediction can result from anomalies of ENSO alter atmospheric circulation (Kumar et al., 2006; Wang et al., 2014). Dimri et al. (2023) states that the Bhagirathi River Basin's seasonal variations in water availability are being exacerbated by climate change, placing a strain on adaptive management. Dhillon et al. (2023) suggested that large-scale climatic effects the seasonal precipitation forecasts improve readiness and encourages more effective water management and irrigation planning. Sea surface temperature anomalies like ENSO can shift circulation patterns and cause precipitation changes months later, and these indices reflect ocean atmosphere linkages that affect rainfall through lagged effects (Kumar et al., 2006; Wang et al., 2014). This advanced machine learning combines with climate indices and climatology and metrological parameter to find correlations between conventional and statistical approach towards water management system. The need for accurate rainfall forecasts is highlighted by the growing frequency and intensity of climate-related events, which makes this research not only relevant but also used to transformed the ML techniques such as RF, GPR, SVM, and MARS. These machine learning models use enormous information from climate indices to uncover complex connections between rainfall and atmospheric conditions. According to studies, machine learning can perform better than traditional statistical techniques, improving predicting accuracy and facilitating more efficient agricultural planning and water resource management. Several studies have applied regional frequency analysis (RFA) using L-moments to capture spatial variability in extreme rainfall. Nain et al., 2021). The use of hyperparameter-optimized DNNs can effectively model both glass and carbon FRP beams, outperforming conventional nonlinear regression models (Sharif, M. 2025) and

also implementing optimized DNNs enables reliable prediction of compressive strength while reducing the need for labour intensive testing (Naved et al., 2024). Estimating the shear capacity of FRP-reinforced beams using conventional design equations and comparative studies indicate that international standards such as ACI, CSA, and JSCE differ in accuracy, with JSCE providing more consistent and moderately conservative predictions for both carbon and glass FRP beams. (Sharif, M. 2025). The use of advanced optimization techniques, such as the Adam optimizer, further enhances the predictive accuracy and practical applicability of DNN models in hydraulic engineering (Asim et al., 2022).

The nonlinear relationship between the large-scale predictor, local climate parameter and rainfall as the output for making strong prior assumptions, machine learning (ML) has become a potential way to get around the drawbacks of traditional models in recent years (Gardner & Dorling, 1998; Abrahart et al., 2008). In a water resource modelling problem, machine learning techniques such as Random Forests (RF), Multivariate Adaptive Regression Splines (MARS), and Support Vector Machines (SVM) have proven to be successful. For efficient regression, SVM uses kernel functions to map input data into higher-dimensional spaces (Smola & Scholkopf, 2004; Breiman, 2001; Friedman, 1991).

The major objective of this research is to estimate monthly rainfall in Farukh Nagar, Haryana, India, this study combines a number of large-scale climate indices, such as IOD ,ENSO,PDO, AMO, and NAO, along with local climatic driving parameters, such as maximum and minimum temperatures: applicability of machine learning techniques such as MARS, SVM, GPR, and RF in an effort to capture the monthly lagged dynamics of large scale climate indices. To develop novel tree-based approach techniques like RF and Kernel based approach as SVM and GPR and MARS to develop planning of water resources systems.

Modelling Techniques

Gaussian Process Regression (GPR)

Gaussian Process (GP) Regression is a approach related to machine-based learning which is widely used for prediction of hydrological modelling of space, making it a probabilistic and non-parametric approach. GPR is essentially a generalization of the Gaussian distribution, where the function is mean vector, covariance matrix, representing the statistical properties of the distribution. One of the key advantages of GPR is that it does not require explicit assumptions about the form of the target function, due to its ability to encode prior beliefs through the covariance function (kernel). GPR is capable of producing not only point predictions but also uncertainty estimates for these predictions, making it highly suitable for probabilistic forecasting (Rasmussen and Williams, 2006). A GP is fully described by two functions: the mean function, which represents the average trend of the data, and the covariance function (or kernel), which defines how strongly different points are related or correlated with each other.

Random Forest Regression (RFR)

Random Forest is a popular ensemble learning algorithm that builds a set of decision trees to increase prediction accuracy and robustness for both linear and nonlinear relationships. To reduce overfitting, each tree is trained using a bootstrapped sample of the dataset to minimize overfitting, and the outputs are aggregated using bootstrap aggregation or bagging (Breiman, 1996; 2001).The final prediction is obtained by averaging the outputs of all trees, which effectively maintaining low bias (Liaw & Wiener, 2002).The number of features (mm) that are randomly chosen at each split and the number of trees (kk) in the forest are the two main factors that affect the overall performance of output. This randomization enhances the model's ability to capture intricate connections. In this research, robust and accurate rainfall forecasting is achieved by applying RF regression with feature randomness.

Multivariate Adaptive Regression Splines (MARS)

Multivariate Adaptive Regression Splines is a flexible, non-parametric regression technique that models complex, nonlinear relationships without assuming a predefined functional form. The technique divides the predictor space into smaller regions and fits simple linear regressions within each, with division points known as knots. These knots generate basis functions (BFs), as:

$$BF(x) = \max(0, x - k) \tag{1}$$

where x is the predictor and k is the knot location. The overall MARS model is represented as:

$$y = f(a) = x + \sum(xi \cdot BF\ (a)) \tag{2}$$

where y is the predicted output, x the intercept, x_i coefficients, and BF (a) the basic functions.

The algorithm works in two phases: (i) a forward pass, where basis functions is introduced to improve fit, and (ii) a backward pruning, where insignificant terms are eliminates using Generalized Cross-Validation (GCV). This stepwise approach enables MARS to capture nonlinear dependencies while controlling overfitting and maintaining interpretability.

Support Vector Regression (SVR)

Support Vector Machines (SVMs), introduced by Vapnik in the late 1990s (Vapnik, 1998), are powerful machine learning techniques rooted in statistical learning theory. Its aims to construct an optimal hyperplane that best fits the data, where prediction error is quantified by the distance of observations from this hyperplane. To capture complex nonlinear model, SVR employs kernel functions, with commonly used options including polynomial, radial basis function (RBF), Pearson VII universal kernel (PUK), and normalized polynomial (Demirci, 2019). In rainfall prediction models, choosing an appropriate kernel function and tunning of hyperparameters such as γ (kernel width) and C (penalty term) are critical to model performance. Although research suggests that the type of kernel has only a minor influence, but the parameter values particularly γ and C have a decisive role in determining predictive performance.

Study Area Description

This study is centred on Farrukh Nagar, located in the Gurugram district of Haryana, India, at 28.44° N latitude and 76.83° E longitude. It is situated in northwest part of Indo Gangetic climate plain. The majority of the yearly precipitation occurs during the monsoon season, which may extend from late June to early October. The highest rainfall usually occurs in July and August. Restoring groundwater supplies, maintaining agriculture, assisting water resource management and groundwater recharge,

Climate Indices and Data Sources

Monthly time series of large-scale climate indices including IOD, PDO, AMO, ENSO, and NAO were sourced from the National Oceanic and Atmospheric Administration (NOAA) via the Physical Sciences Laboratory portal (https://psl.noaa.gov/data/climateindices/). These indices represent major ocean–atmosphere interactions across different ocean basins and are known to significantly influence rainfall variability over the Indian subcontinent. In addition, local climatic variables, specifically maximum and minimum temperatures, were included as predictors, while rainfall was considered the target variable.

The correlation matrix (Figure 1.) highlights that rainfall is more strongly linked with local temperatures than with large-scale climate indices. Rainfall shows a moderate positive correlation with minimum temperature (0.45) and a weaker positive correlation with maximum temperature (0.21), indicating that precipitation variability is closely tied to local thermal dynamics. Among the climate indices, NINO3.4 demonstrates moderate positive correlations with PDO (0.43) and IOD (0.32), suggesting interconnections among Pacific and Indian Ocean signals. AMO exhibits modest associations with minimum (0.25) and maximum (0.17) temperatures, whereas NAO shows generally weak and negative relationships. The strongest overall correlation is observed between minimum and maximum temperatures (0.92), reflecting their inherent interdependence. These findings demonstrate that large-scale indices exert only weak direct control on rainfall in the study region, while local thermal variability emerges as the dominant driver. The weak linear linkages highlight the complexity of teleconnections and reinforce the necessity of advanced nonlinear and lag-sensitive modelling approaches to better capture regional rainfall dynamics.

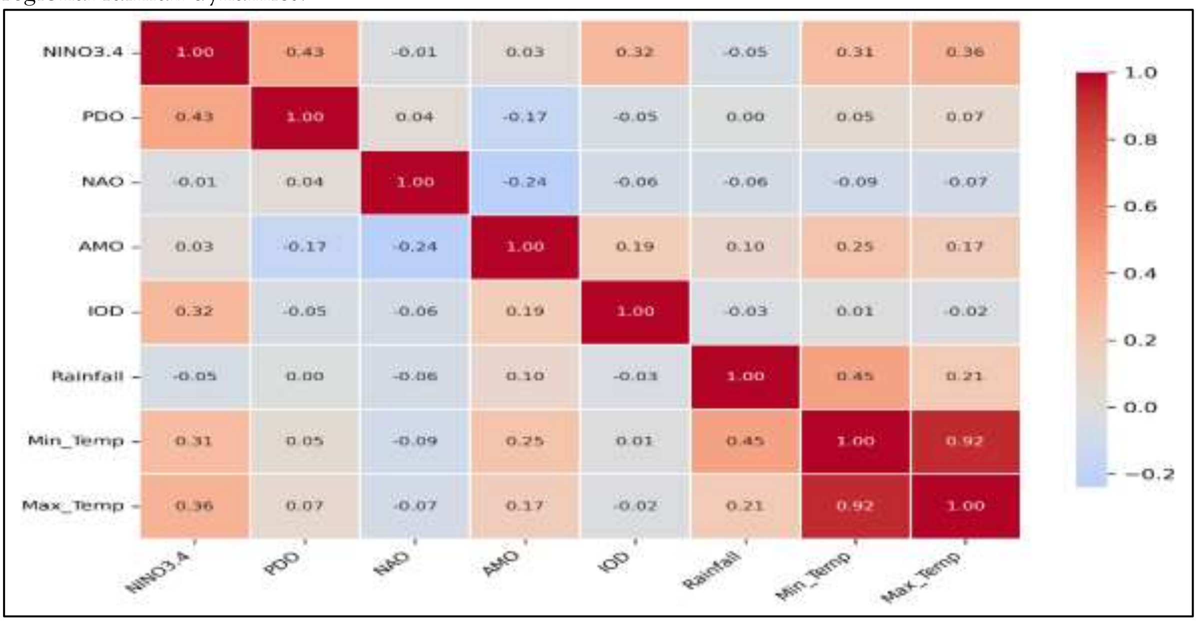

Figure 1. Cross-correlation matrix showing the relationships between climate indices, temperatures, and rainfall.

Kernel Function Details

Kernel functions play a critical role in both Support Vector Machine (SVM) and Gaussian Process regression (GPR) approaches by converting input data into higher dimensional feature spaces, enabling the modelling of intricate nonlinear patterns. The choice of kernel function significantly impacts on models' performance and generalization ability. In this study, two kernel functions were chosen for the frameworks of SVM and GPR regression models. These kernels were selected based on their suitability for time-series prediction and their proven efficacy in previous hydrological and climatic studies. The following kernel functions are used:

(a) Pearson VII Universal Kernel (PUK):

$$K_{PUK}(x_i, x_j) = \left[1 + \left(\frac{2\sqrt{\|x_i - x_j\|^2}}{\sigma\sqrt{2^{1/\omega} - 1}}\right)^2\right]^{-\omega}$$

(b) Radial Basis Function (RBF) Kernel:

$$K_{RBF}(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$$

In both cases, the kernel parameters (σ, ω, and γ) play a crucial role in shaping the covariance structure and directly influence the predictive strength of the Gaussian Process. To ensure proper tuning of hyperparameters, it is necessary to achieve accurate regression model to enhance model generalization, and maintain stability. These selected kernel functions were assessed in terms of predictive accuracy using the established performance metrics (R, RMSE, MAE, and SI). Their implementation and tuning were integral to the overall modelling strategy, particularly in improving the generalization capability of the SVM and GP models.

Kernel Parameter Optimization

The performance of Gaussian Process (GP) and Support Vector Machine (SVM) regression models is highly sensitive to the proper tunning of kernel specific hyperparameters. The key parameters include the regularization constant (C), Gaussian noise level, and kernel-related hyperparameters such as γ (gamma), σ (sigma), and ω (omega). These parameters govern the model's complexity, generalization capability, and fitting accuracy. The optimized primary hyperparameters of the GPR and SVM models are listed in Table 1. In this research, a physically guided method was adopted for the selection of these key parameters.

The tuning procedure was designed to minimize the Root Mean Square Error (RMSE) and maximize the correlation coefficient (CC) during model validation. For consistency and fairness in model comparison, the same kernel-specific parameters were applied in both GP and SVM regression models.

Table 1. Optimized primary hyperparameters of GPR and SVM Models

Approach	**Primary hyperparameters**
SVM with RBF Kernel	C = 2, γ = 1
SVM with PUK Kernel	C = 2, ω = 1, σ = 1
GP with RBF Kernel	Gaussian noise = 0.01, γ = 1
GP with PUK Kernel	Gaussian noise = 0.01, ω = 1, σ = 1

Model Evaluation

Evaluating the predictive accuracy of machine learning models is an essential component of any data-driven study. In this study, rainfall prediction models developed using Support Vector Machines (SVM), Multivariate Adaptive Regression Splines (MARS), Gaussian Process Regression (GPR) and Random Forest (RF) are assessed through a comparison of predicted rainfall values against observed values. To ensure a robust assessment, four widely used statistical performance measures were employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Scattering Index (SI) and the Correlation Coefficient (CC) measures how well a regression model performs.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|PR_i - OR_i| \tag{5}$$

Root Mean Square Error (RMSE), which penalizes larger errors more heavily than MAE, is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(PR_i - OR_i)^2} \tag{6}$$

Coefficient of Correlation (CC) quantifies the strength and direction of the linear relationship between observed and predicted values:

$$R = \frac{\sum_{i=1}^{N}\left(PR_i - \underline{OR}\right)(OR_i - \underline{OR})}{\sqrt{\sum_{i=1}^{N}(PR_i - \underline{OR})^2 \sum_{i=1}^{N}(OR_i - \underline{OR})^2}} \tag{7}$$

Scattering index (SI): $SI = \frac{RMSE}{\underline{OR}}$ (8)

Where N is number of observations, OR is actual rainfall values, PR is predicted rainfall values, $\_OR$ is the average of actual rainfall values

The ideal model is expected to produce lower values of MAE and RMSE, indicating minimal deviation between predicted and observed values, and a higher CC value (closer to 1.0), indicating strong correlation and reliability of the model (Smola & Scholkopf, 2004; Friedman, 1991; Breiman, 2001).

These criteria were applied to assess the relative performance of each model in both the training and validation phases, ensuring a thorough evaluation of generalization ability across datasets.

Rationale for Input Data Selection

The selection of input variables is a critical determinant of the accuracy and generalizability of predictive models. To enhance robustness in model development, thirteen distinct input feature combinations were systematically designed, as summarized in Table 2. These combinations, labelled N_1 through N_{13}, were formulated based on their observed linkages with monthly rainfall patterns, while accounting for both the physical relevance and statistical significance of each variable. The same input datasets were systematically applied during the training and evaluation of rainfall prediction models for Farukh Nagar, Haryana, India. This consistent framework enabled a comprehensive assessment of how different model configurations of climate indices and local climatic variables influence model performance, thereby enhancing the reliability of rainfall predictability in the study region. In addition to large-scale climate indices, local maximum and minimum temperatures were included as predictor variables to capture regional variability with predicting rainfall as the target variable.

Table 2. Structure of input sets (N_1–N_{13}) comprising different combinations of climate indices, rainfall, and temperatures used in model training and testing**.**

Model	Input Combination
N_1	$NI_{(t-1)}$, $PD_{(t-1)}$, $NAO_{(t-1)}$, $AMO_{(t-1)}$, $IOD_{(t-1)}$, $R_{(t-1)}$, T_{max}, T_{min}
N_2	$NI_{(t-2)}$, $NI_{(t-1)}$, $NAO_{(t-2)}$, $NAO_{(t-1)}$, $PD_{(t-2)}$, $PD_{(t-1)}$, $AMO_{(t-2)}$, $AMO_{(t-1)}$, $IOD_{(t-2)}$, $IOD_{(t-1)}$, $R_{(t-2)}$, $R_{(t-1)}$, T_{max}, T_{min}
N_3	$NI_{(t-3)}$, $NI_{(t-2)}$,$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$, T_{max}, T_{min}
N_4	$NI_{(t-3)}$,NI(t-2),$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$,$R_{(t-4)}$, T_{max}, T_{min}
N_5	$NI_{(t-3)}$,$NI_{(t-2)}$,$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$,$R_{(t-4)}$, $R_{(t-5)}$, T_{max}, T_{min}
N_6	$NI_{(t-3)}$,$NI_{(t-2)}$,$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$,$R_{(t-4)}$, $R_{(t-5)}$,$R_{(t-6)}$, T_{max}, T_{min}
N_7	$NI_{(t-3)}$,$NI_{(t-2)}$,$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$,$R_{(t-4)}$, $R_{(t-5)}$, $R_{(t-6)}$, $R_{(t-7)}$, T_{max}, T_{min}
N_8	$NI_{(t-3)}$,$NI_{(t-2)}$,$NI_{(t-1)}$,$NAO_{(t-3)}$,$NAO_{(t-2)}$,$NAO_{(t-1)}$,$PD_{(t-3)}$, $PD_{(t-2)}$,$PD_{(t-1)}$,$AMO_{(t-3)}$,$AMO_{(t-2)}$, $AMO_{(t-1)}$,$IOD_{(t-3)}$,$IOD_{(t-2)}$, $IOD_{(t-1)}$,$R_{(t-3)}$, $R_{(t-2)}$,$R_{(t-1)}$,$R_{(t-4)}$, $R_{(t-5)}$, $R_{(t-6)}$, $R_{(t-7)}$, $R_{(t-8)}$, T_{max}, T_{min}

N_9	$NI_{(t-3)}, NI_{(t-2)}, NI_{(t-1)}, NAO_{(t-3)}, NAO_{(t-2)}, NAO_{(t-1)}, PD_{(t-3)}, PD_{(t-2)}, PD_{(t-1)}, AMO_{(t-3)}, AMO_{(t-2)}, AMO_{(t-1)}, IOD_{(t-3)}, IOD_{(t-2)}, IOD_{(t-1)}, R_{(t-3)}, R_{(t-2)}, R_{(t-1)}, R_{(t-4)}, R_{(t-5)}, R_{(t-6)}, R_{(t-7)}, R_{(t-8)}, R_{(t-9)}, T_{max}, T_{min}$
N_{10}	$NI_{(t-3)}, NI_{(t-2)}, NI_{(t-1)}, NAO_{(t-3)}, NAO_{(t-2)}, NAO_{(t-1)}, PD_{(t-3)}, PD_{(t-2)}, PD_{(t-1)}, AMO_{(t-3)}, AMO_{(t-2)}, AMO_{(t-1)}, IOD_{(t-3)}, IOD_{(t-2)}, IOD_{(t-1)}, R_{(t-3)}, R_{(t-2)}, R_{(t-1)}, R_{(t-4)}, R_{(t-5)}, R_{(t-6)}, R_{(t-7)}, R_{(t-8)}, R_{(t-9)}, R_{(t-10)}, T_{max}, T_{min}$
N_{11}	$NI_{(t-3)}, NI_{(t-2)}, NI_{(t-1)}, NAO_{(t-3)}, NAO_{(t-2)}, NAO_{(t-1)}, PD_{(t-3)}, PD_{(t-2)}, PD_{(t-1)}, AMO_{(t-3)}, AMO_{(t-2)}, AMO_{(t-1)}, IOD_{(t-3)}, IOD_{(t-2)}, IOD_{(t-1)}, R_{(t-3)}, R_{(t-2)}, R_{(t-1)}, R_{(t-4)}, R_{(t-5)}, R_{(t-6)}, R_{(t-7)}, R_{(t-8)}, R_{(t-9)}, R_{(t-10)}, R_{(t-11)}, T_{max}, T_{min}$
N_{12}	$NI_{(t-3)}, NI_{(t-2)}, NI_{(t-1)}, NAO_{(t-3)}, NAO_{(t-2)}, NAO_{(t-1)}, PD_{(t-3)}, PD_{(t-2)}, PD_{(t-1)}, AMO_{(t-3)}, AMO_{(t-2)}, AMO_{(t-1)}, IOD_{(t-3)}, IOD_{(t-2)}, IOD_{(t-1)}, R_{(t-3)}, R_{(t-2)}, R_{(t-1)}, R_{(t-4)}, R_{(t-5)}, R_{(t-6)}, R_{(t-7)}, R_{(t-8)}, R_{(t-9)}, R_{(t-10)}, R_{(t-11)}, R_{(t-12)}, T_{max}, T_{min}$
N_{13}	$NI_{(t)}, PD_{(t)}, NAO_{(t)}, AMO_{(t)}, IOD_{(t)}, R_{(t-1)}, R_{(t-2)}, R_{(t-3)}, R_{(t-4)}, R_{(t-5)}, R_{(t-6)}, R_{(t-7)}, R_{(t-8)}, R_{(t-9)}, R_{(t-10)}, R_{(t-11)}, R_{(t-12)}, T_{max}, T_{min}$

Model Development

To assess the predictive performance of the developed models, four widely accepted statistical performance metrics were employed: the correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), and scatter index (SI). These metrics provide a comprehensive evaluation of model accuracy, robustness, and error distribution. The model development process followed an iterative trial-and-error approach. Optimal values for user-defined hyperparameters were done independently to identified a successive refinement based on the performance observed during each iteration. This process was conducted separately for each model type, as different algorithms require distinct tuning strategies to achieve the best results. This methodology ensured that the predictive models were not only statistically sound but also practically reliable for monthly rainfall forecasting.

The adopted methodology is illustrated in Figure 2, represents approach used for model construction, training, and testing.

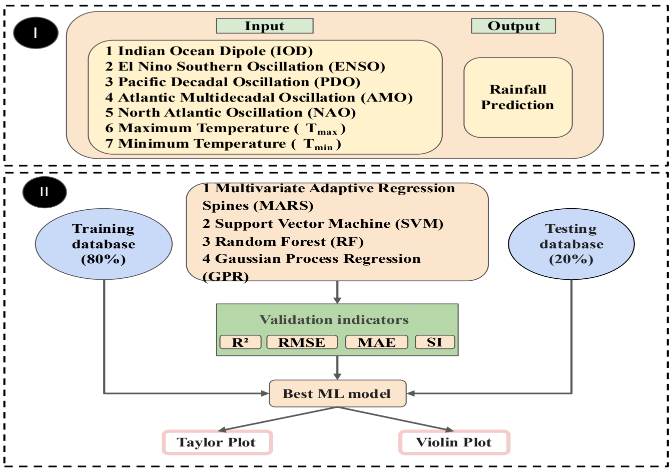

Figure 2. Flowchart represents the methodology adopted for rainfall prediction.

Result of SVM Models

Support Vector Machine (SVM) models were implemented in WEKA 3.9 using both radial basis function (RBF) and polynomial kernels. To enable a fair comparison between the two kernels, parameter tuning was performed through grid search combined with cross-validation. The optimized hyperparameters were $\gamma = 1$ and $C = 2$ for the RBF kernel, and degree = 3 with $C = 2$ for the polynomial kernel. The predictive performance of SVM-based rainfall prediction models using lagged climate indices is summarized in Table 3, where key metrics including MAE, RMSE, SI, and CC are reported. As illustrated in Figure 4a & 4b, the SVM-RBF model consistently outperformed the polynomial kernel in both training and testing phases. For the training dataset, the SVM-RBF achieved MAE = 20.4360, RMSE = 49.1941, CC = 0.8108, and SI = 1.1468, while for the testing dataset it yielded MAE = 24.4519, RMSE = 46.8608, CC = 0.7184, and SI = 1.1873. These findings indicate that the RBF kernel provides superior generalization ability and robustness in modelling nonlinear relationships between lagged climate indices and rainfall compared to the polynomial kernel as shown in Figure 3a & 3b.

Table 3. Performance evaluation of Support Vector Machine Regression (SVR) models with Pearson VII universal kernel (PUK) and radial basis function (RBF) kernels for monthly rainfall prediction.

Models	CC	MAE	RMSE	SI	CC	MAE	RMSE	SI
	Training				Testing			
SVM PUK Z1	0.9672	4.3253	21.7604	0.5073	0.6416	30.1696	51.4146	1.3027
SVM PUK Z2	1.0000	0.5933	0.6433	0.0150	0.6144	31.8809	52.0264	1.3182
SVM PUK Z3	1.0000	0.5757	0.6221	0.0145	0.5567	32.1812	54.6347	1.3843
SVM PUK Z4	1.0000	0.5498	0.6021	0.0140	0.5531	31.9019	54.7475	1.3871
SVM PUK Z5	1.0000	0.5890	0.6363	0.0148	0.5539	31.8523	54.7026	1.3860
SVM PUK Z6	1.0000	0.5841	0.6300	0.0147	0.5539	31.7758	54.6958	1.3858
SVM PUK Z7	1.0000	0.5504	0.5976	0.0139	0.5562	31.4963	54.5892	1.3831
SVM PUK Z8	1.0000	0.5647	0.6185	0.0144	0.5532	31.6778	54.7325	1.3868
SVM PUK Z9	1.0000	0.5498	0.6010	0.0140	0.5516	31.9092	54.9118	1.3913
SVM PUK Z10	1.0000	0.6018	0.6499	0.0151	0.5464	32.2175	55.1830	1.3982
SVM PUK Z11	1.0000	0.5580	0.6085	0.0142	0.5623	31.9124	54.3272	1.3765
SVM PUK Z12	1.0000	0.5792	0.6287	0.0147	0.5764	31.3686	53.5896	1.3578
SVM PUK Z13	1.0000	0.6047	0.6495	0.0151	0.5855	32.7878	54.2046	1.3734
SVM RBF Z1	0.8108	20.4360	49.1941	1.1468	0.7184	24.4519	46.8608	1.1873
SVM RBF Z2	0.8836	12.7358	39.6396	0.9240	0.6047	30.8292	52.2245	1.3232
SVM RBF Z3	0.9473	6.3706	27.5299	0.6417	0.5311	34.3158	56.5764	1.4335
SVM RBF Z4	0.9487	6.1127	27.1188	0.6322	0.5219	35.0538	57.1212	1.4473
SVM RBF Z5	0.9495	5.9953	26.9159	0.6274	0.5278	34.3283	56.6896	1.4364
SVM RBF Z6	0.9504	5.8040	26.6086	0.6203	0.5235	34.5111	56.6979	1.4366
SVM RBF Z7	0.9530	5.5071	25.8506	0.6026	0.5227	34.5976	56.9722	1.4435
SVM RBF Z8	0.9572	5.2112	24.8398	0.5790	0.5105	35.6324	57.7529	1.4633
SVM RBF Z9	0.9602	4.8476	23.8068	0.5550	0.4981	36.1706	58.7464	1.4885
SVM RBF Z10	0.9652	4.2444	22.3021	0.5199	0.4857	36.5557	59.4025	1.5051
SVM RBF Z11	0.9803	3.3343	16.9395	0.3949	0.4877	37.0898	59.3737	1.5044
SVM RBF Z12	0.9836	2.5901	15.0519	0.3509	0.4920	35.5581	59.0266	1.4956
SVM RBF Z13	0.9158	10.1687	33.1563	0.7729	0.5758	30.1309	54.1436	1.3718

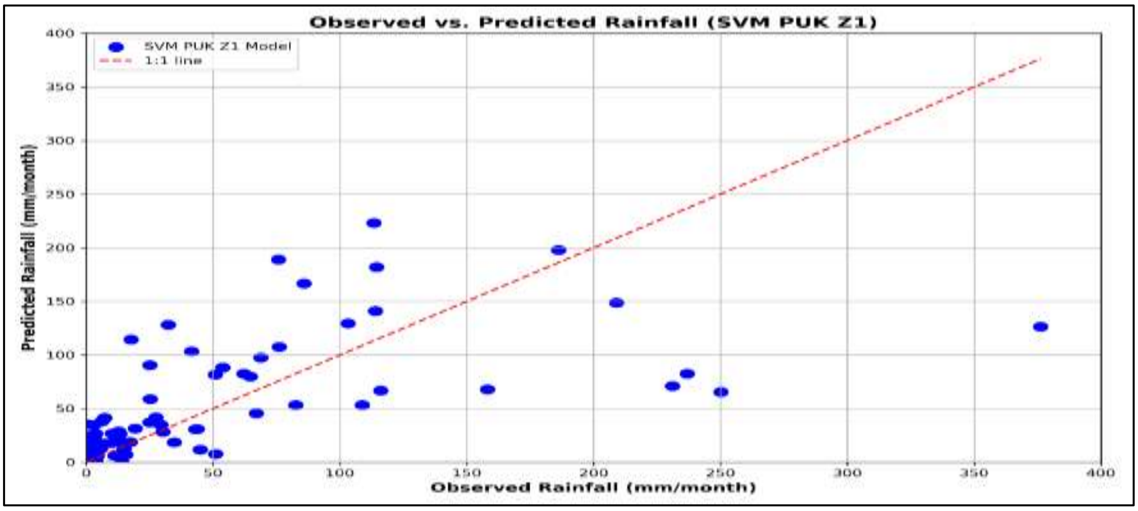

Figure 3. (a) Scatter plot of model predictions using the testing dataset for SVM-PUK at Z1 for Farukh Nagar, Haryana (all datasets)

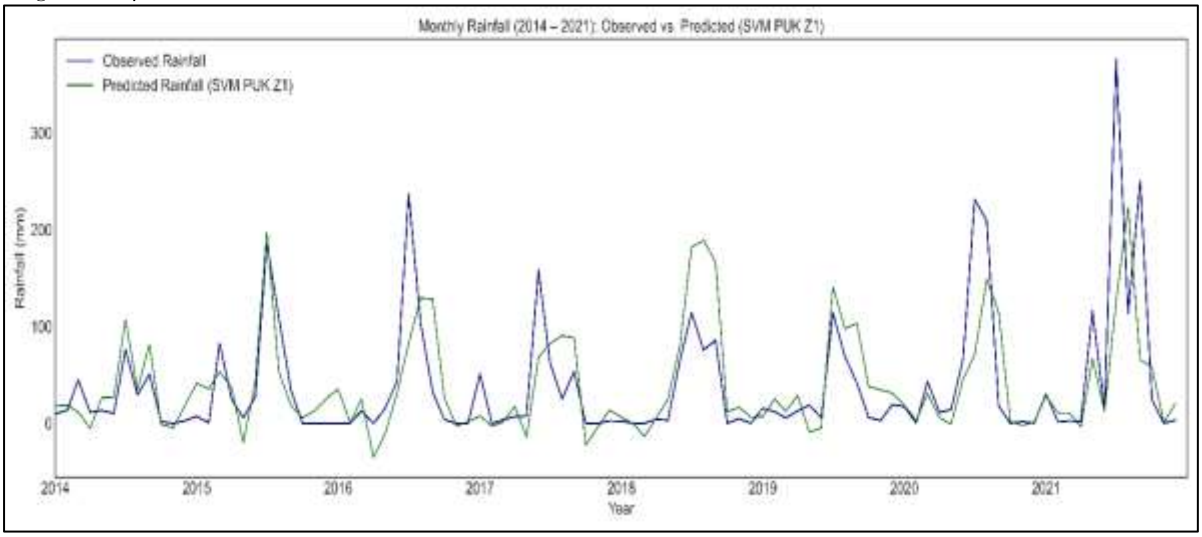

Figure 3. (b) Hydrograph showing model predictions using SVM-PUK at Z1 for Farukh Nagar, Haryana (all datasets)

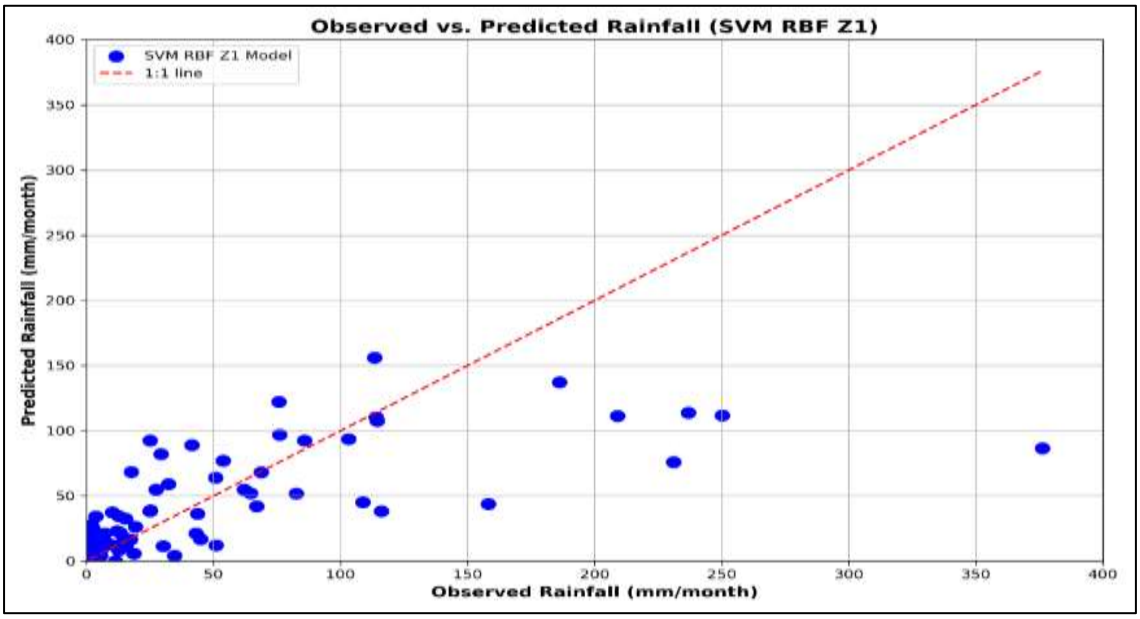

Figure 4. (a) Scatter plot of model predictions using the testing dataset for SVM-RBF at Z1 for Farukh Nagar, Haryana (all datasets)

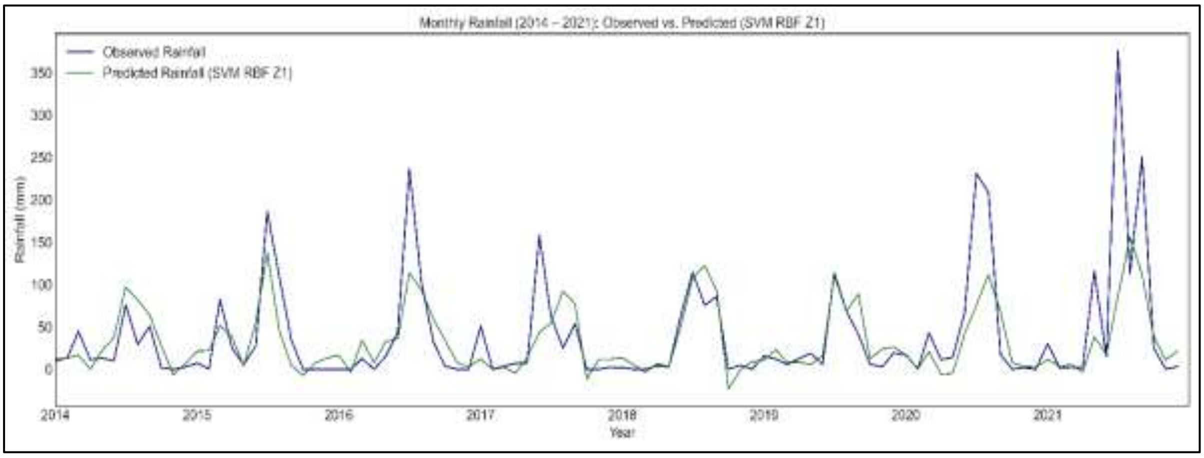

Figure 4. (b) Hydrograph showing model predictions using SVM-RBF at Z1 for Farukh Nagar, Haryana (all datasets)

Results of GP Models

In this research, Gaussian Process (GP) models utilizing the Pearson VII kernel (PUK) and the Radial Basis Function (RBF) kernel were implemented through the WEKA 3.9 platform. To ensure fair model comparison, the Gaussian noise level was held constant at 0.01 for both kernels. The optimal hyperparameters were identified for each model: PUK kernel was configured with $\omega = 1$ and $\sigma = 1$, whereas the RBF kernel achieved its best performance with $\gamma = 1$. As summarized in Table 4, the Gaussian Process (GP) model with the RBF Z1 kernel consistently outperformed its PUK-based counterpart. Specifically, the GP RBF Z1 model achieved MAE, RMSE, SI, and CC values of 30.1617, 51.2365, 1.1944, and 0.7550 on the training dataset, and 31.1363, 48.5957, 1.2313, and 0.6858 on the testing dataset as shown in Figure 5a &5b. These metrics highlight the model's superior generalization ability and predictive efficiency compared to the PUK-based GP model.

Table 4. Performance evaluation of Gaussian Process Regression (GPR) models with Pearson VII universal kernel (PUK) and radial basis function (RBF) kernels for monthly rainfall prediction.

	CC	MAE	RMSE	SI	CC	MAE	RMSE	SI
Models	Training				Testing			
GP PUK Z1	1.0000	0.0532	0.0979	0.0023	0.6317	34.5316	53.0679	1.3446
GP PUK Z2	1.0000	0.0109	0.0183	0.0004	0.6149	31.5407	51.9663	1.3167
GP PUK Z3	1.0000	0.0069	0.0122	0.0003	0.5572	32.0137	54.5960	1.3833
GP PUK Z4	1.0000	0.0068	0.0121	0.0003	0.5538	31.8109	54.7134	1.3863
GP PUK Z5	1.0000	0.0067	0.0121	0.0003	0.5548	31.7240	54.6482	1.3846
GP PUK Z6	1.0000	0.0067	0.0120	0.0003	0.5550	31.6747	54.6376	1.3844
GP PUK Z7	1.0000	0.0066	0.0119	0.0003	0.5570	31.4529	54.5451	1.3820
GP PUK Z8	1.0000	0.0065	0.0118	0.0003	0.5540	31.6730	54.6961	1.3858
GP PUK Z9	1.0000	0.0063	0.0114	0.0003	0.5520	31.9310	54.8840	1.3906
GP PUK Z10	1.0000	0.0061	0.0111	0.0003	0.5467	32.2153	55.1628	1.3977
GP PUK Z11	1.0000	0.0058	0.0104	0.0002	0.5626	31.9065	54.3063	1.3760
GP PUK Z12	1.0000	0.0056	0.0100	0.0002	0.5768	31.3014	53.5663	1.3572
GP PUK Z13	1.0000	0.0096	0.0160	0.0004	0.5856	32.6028	54.1892	1.3730
GP RBF Z1	0.7550	30.1617	51.2365	1.1944	0.6858	31.1363	48.5957	1.2313
GP RBF Z2	0.8009	28.8626	46.8730	1.0927	0.6055	36.8206	55.6046	1.4089
GP RBF Z3	0.8345	26.6474	43.2534	1.0083	0.5999	37.8331	56.9713	1.4435
GP RBF Z4	0.8364	26.3199	43.0318	1.0031	0.5915	38.8269	57.5730	1.4587

GP RBF Z5	0.8385	26.0102	42.7924	0.9975	0.5909	39.1394	57.6626	1.4610
GP RBF Z6	0.8398	25.8503	42.6432	0.9941	0.5892	39.1793	57.8548	1.4659
GP RBF Z7	0.8419	25.6187	42.3995	0.9884	0.5894	39.2638	57.9283	1.4677
GP RBF Z8	0.8437	25.5104	42.1909	0.9835	0.5886	39.5704	58.0917	1.4719
GP RBF Z9	0.8472	25.2115	41.7707	0.9737	0.5866	39.7144	58.3214	1.4777
GP RBF Z10	0.8537	24.9231	40.9794	0.9553	0.5835	40.2384	58.4916	1.4820
GP RBF Z11	0.8653	24.4018	39.5135	0.9211	0.5967	39.0501	56.8378	1.4401
GP RBF Z12	0.8767	23.5100	38.0216	0.8863	0.5821	39.5112	57.5579	1.4584
GP RBF Z13	0.8410	26.5318	42.5178	0.9911	0.6126	32.8850	53.0134	1.3432

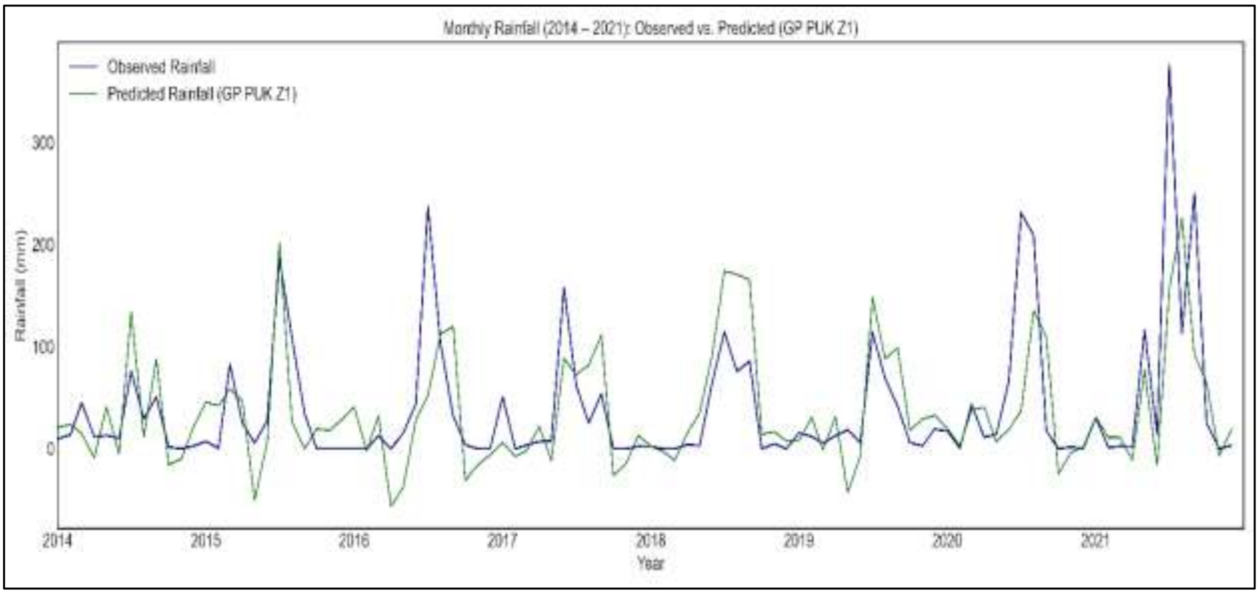

Figure 5. (a) Hydrograph showing model predictions using GP-PUK at Z1 for Farukh Nagar, Haryana (all datasets).

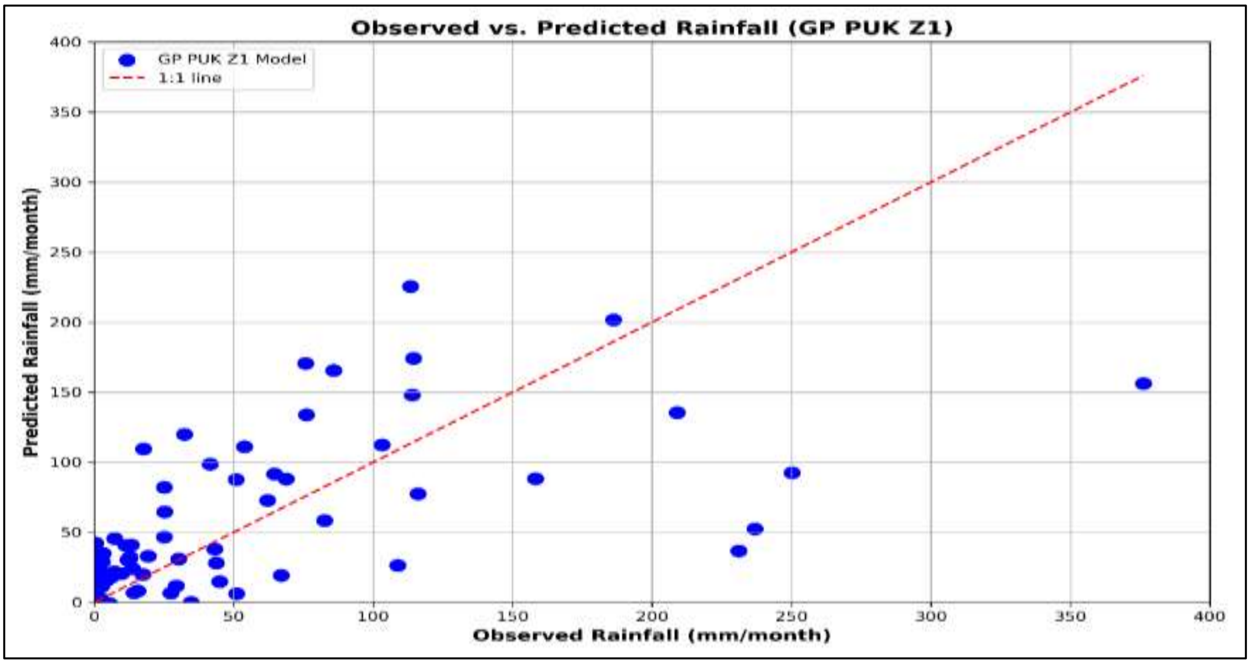

Figure 5. (b) Scatter plot of model predictions using the testing dataset for GP-PUK at Z1 for Farukh Nagar, Haryana (all datasets).

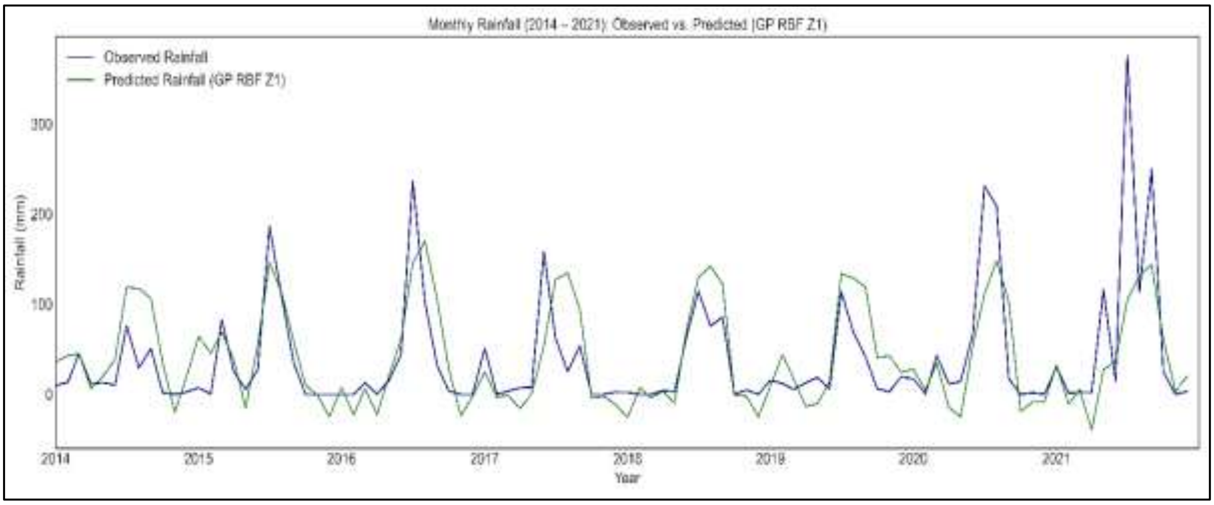

Figure 6. (a) Hydrograph showing model predictions using GP-RBF at Z1 for Farukh Nagar, Haryana (all datasets).

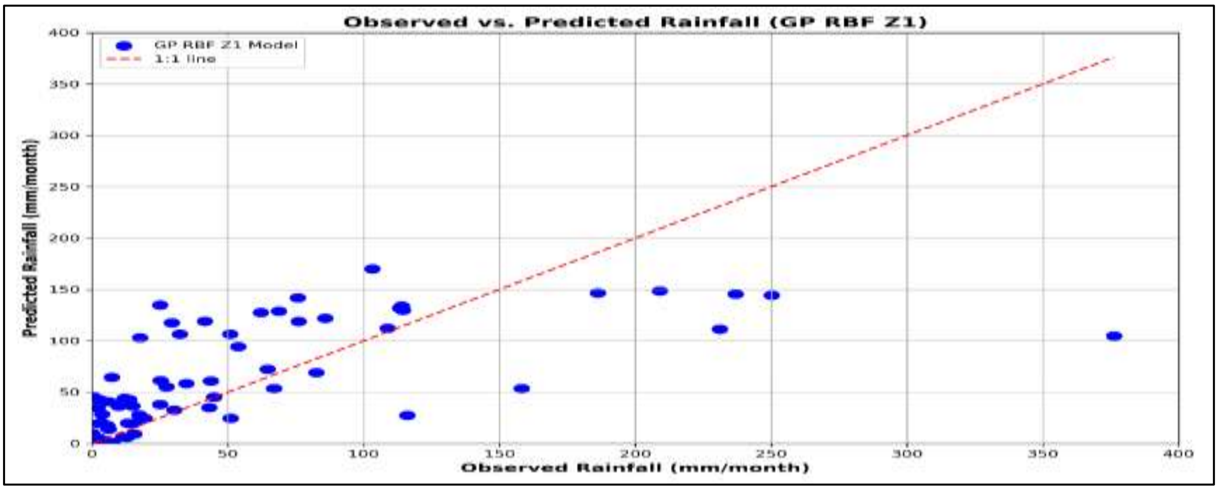

Figure 6. (b) Scatter plot of model predictions using the testing dataset for GP-RBF at Z1 for Farukh Nagar, Haryana (all datasets).

Result of the Random Forest (RF) Model

The Random Forest (RF) model was developed in WEKA 3.9 using an optimal configuration of k = 1, m = 1, and I = 100 trees. Model performance was evaluated using standard statistical metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Scatter Index (SI), and the Correlation Coefficient (CC). During training, the model achieved MAE = 9.9992, RMSE = 18.8641, SI = 0.4397, and CC = 0.9769. In the testing phase, the corresponding values were MAE = 27.3642, RMSE = 50.1163, SI = 1.2698, and CC = 0.6603, as summarized in Table 5. The scatter plots in Figure 7a & 7b. demonstrate the strong agreement between observed and predicted compressive strength values for both training and testing datasets. Overall, the RF Z1 model emerged as the best-performing model, exhibiting superior predictive accuracy and generalization capability.

Table 5. Performance evaluation of Random Forest (RF) models for monthly rainfall prediction

	CC	MAE	RMSE	SI	CC	MAE	RMSE	SI
Models	Training				Testing			
RF Z1	0.9769	9.9992	18.8641	0.4397	0.6603	27.3642	50.1163	1.2698
RF Z2	0.9756	10.4300	19.7299	0.4599	0.6175	31.4052	52.1440	1.3212
RF Z3	0.9769	10.7461	20.1487	0.4697	0.5409	34.7510	57.3528	1.4532
RF Z4	0.9770	10.7522	19.8952	0.4638	0.5843	32.8147	54.4057	1.3785
RF Z5	0.9758	10.9485	20.2106	0.4711	0.5705	33.6584	55.4899	1.4060

RF Z6	0.9762	11.2307	21.0611	0.4910	0.5498	35.2810	55.6102	1.4090
RF Z7	0.9764	11.0440	20.4672	0.4771	0.5667	33.9365	55.2566	1.4000
RF Z8	0.9758	11.4965	20.9843	0.4892	0.5406	35.7917	57.0328	1.4451
RF Z9	0.9749	11.2641	21.1559	0.4932	0.5507	35.0686	55.8371	1.4148
RF Z10	0.9752	11.5841	21.5901	0.5033	0.5492	34.7081	55.9407	1.4174
RF Z11	0.9774	11.2721	20.7201	0.4830	0.5882	33.8079	53.7665	1.3623
RF Z12	0.9751	11.6001	21.2518	0.4954	0.5982	32.4018	52.9782	1.3423
RF Z13	0.9803	10.3070	18.9486	0.4417	0.6304	28.8830	50.9748	1.2916

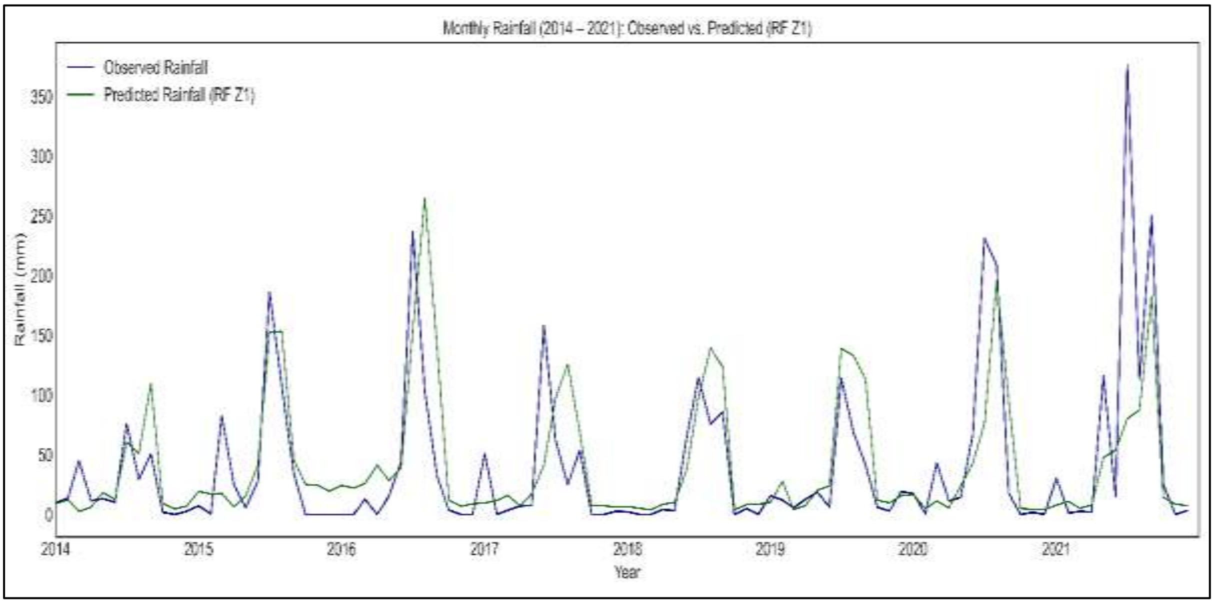

Figure 7. (a) Hydrograph showing model predictions using RF at Z1 for Farukh Nagar, Haryana (all datasets).

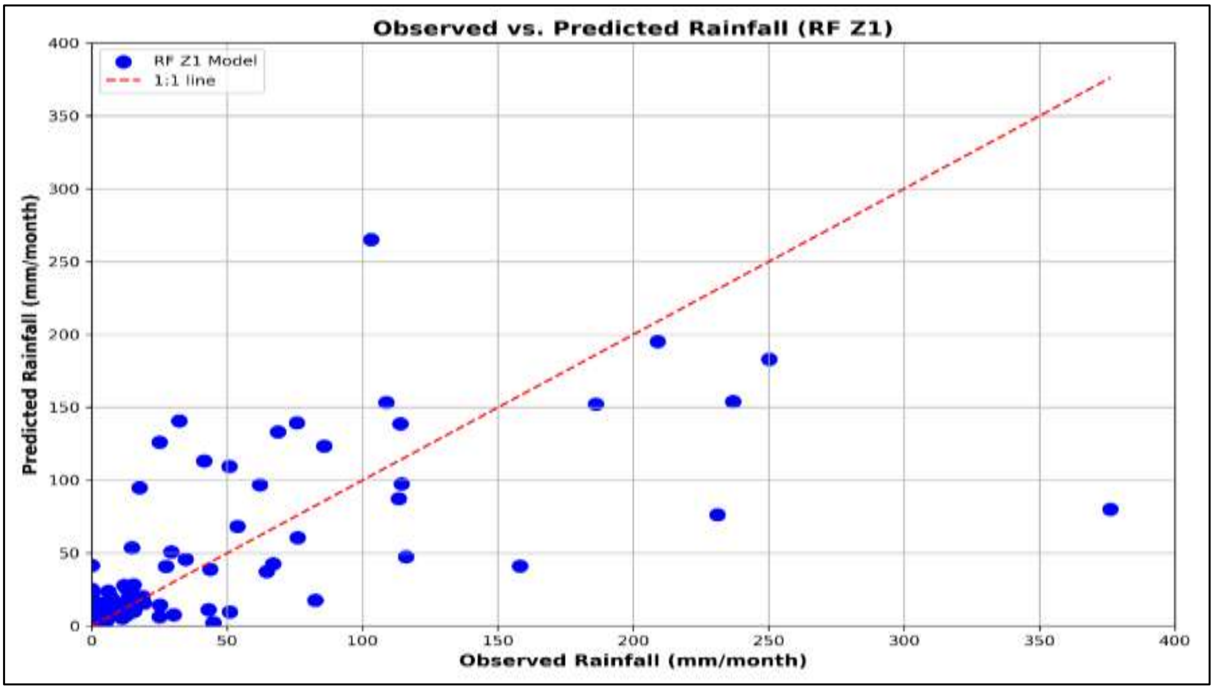

Figure 7. (b) Scatter plot of model predictions using the testing dataset for RF at Z1 for Farukh Nagar, Haryana (all datasets)

Results of the MARS Model

The Multivariate Adaptive Regression Splines (MARS) method was employed to predict the compressive strength of concrete. The modelling process began with the generation of initial basis functions, which were subsequently refined through a pruning procedure that retained the most significant basis functions to enhance model generalization. Pruning was performed using the Generalized Cross-Validation (GCV) criterion, yielding an optimal GCV score of 0.11. The structure of the final MARS model is presented in Table 6, while its predictive performance across both training and testing datasets is illustrated in Figure 8. These visualizations highlight the model's robustness and its ability to accurately capture the input–output relationships. During training, the model achieved a correlation coefficient (CC) of 0.8933, a mean absolute error (MAE) of 23.9939, and a root mean square error (RMSE) of 35.0962. In the testing phase, the model maintained strong predictive capability, with CC = 0.6199, MAE = 33.2684, and RMSE = 53.7740. The scatter index (SI) was 0.8181 for training and 1.3625 for testing, indicating a good agreement between predicted and observed values. Overall, the MARS Z12 model emerged as the best-performing model, demonstrating superior accuracy and generalization as shown in Figure 8a & 8b.

Table 6. Performance evaluation of Multivariate Adaptive Regression Splines (MARS) models for monthly rainfall prediction

	CC	MAE	RMSE	SI	CC	MAE	RMSE	SI
Models	Training				Testing			
MARS Z1	0.8619	26.0217	39.5926	0.9229	0.1908	70.9206	171.2349	4.3386
MARS Z2	0.8791	25.2543	37.2249	0.8677	0.0432	75.0626	345.7353	8.7600
MARS Z3	0.8782	25.4094	37.3561	0.8708	0.2654	53.0129	90.3225	2.2885
MARS Z4	0.8782	25.4094	37.3561	0.8708	0.2654	53.0129	90.3225	2.2885
MARS Z5	0.8730	27.1842	38.0826	0.8877	0.5159	42.3755	63.2477	1.6025
MARS Z6	0.8730	27.1842	38.0826	0.8877	0.5159	42.3755	63.2477	1.6025
MARS Z7	0.8762	25.4918	37.6400	0.8774	0.1319	80.5234	239.8892	6.0781
MARS Z8	0.8762	25.4918	37.6400	0.8774	0.1319	80.5234	239.8892	6.0781
MARS Z9	0.8796	24.7006	37.1528	0.8661	0.2431	61.1876	141.6284	3.5885
MARS Z10	0.8686	25.8968	38.7013	0.9022	0.3714	42.7698	68.1388	1.7264
MARS Z11	0.8708	27.7975	38.3893	0.8949	0.4862	44.6157	65.1817	1.6515
MARS Z12	0.8933	23.9939	35.0962	0.8181	0.6199	33.2684	53.7740	1.3625
MARS Z13	0.9131	22.4879	31.8395	0.7422	0.5600	31.4184	55.8182	1.4143

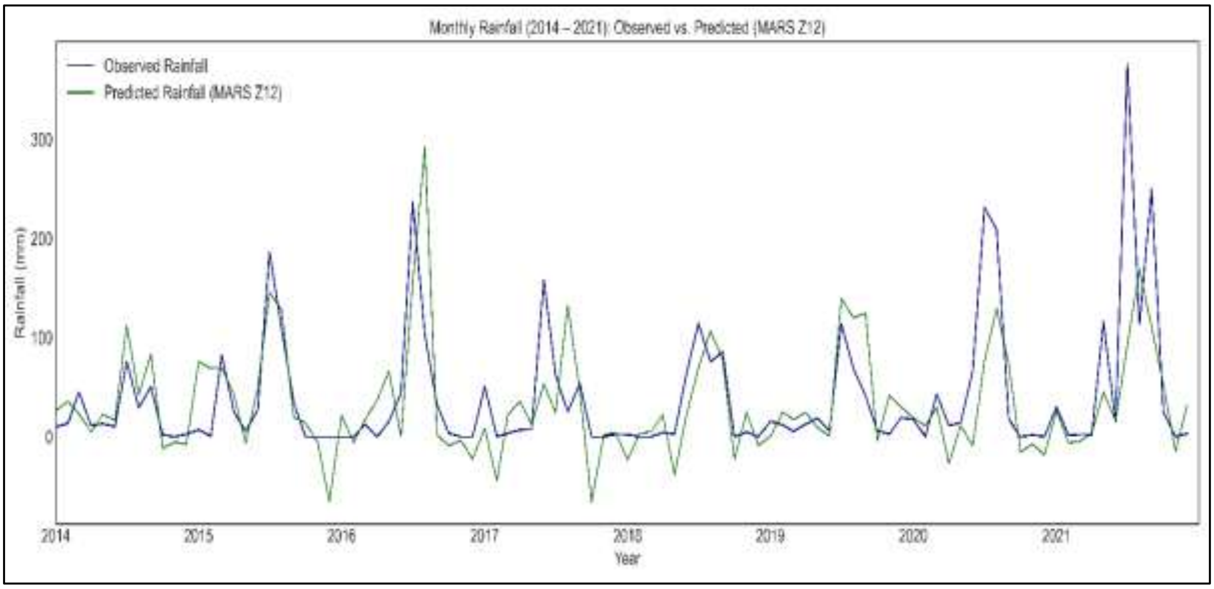

Figure 8. (a) Hydrograph showing model predictions using MARS at Z12 for Farukh Nagar, Haryana (all datasets).

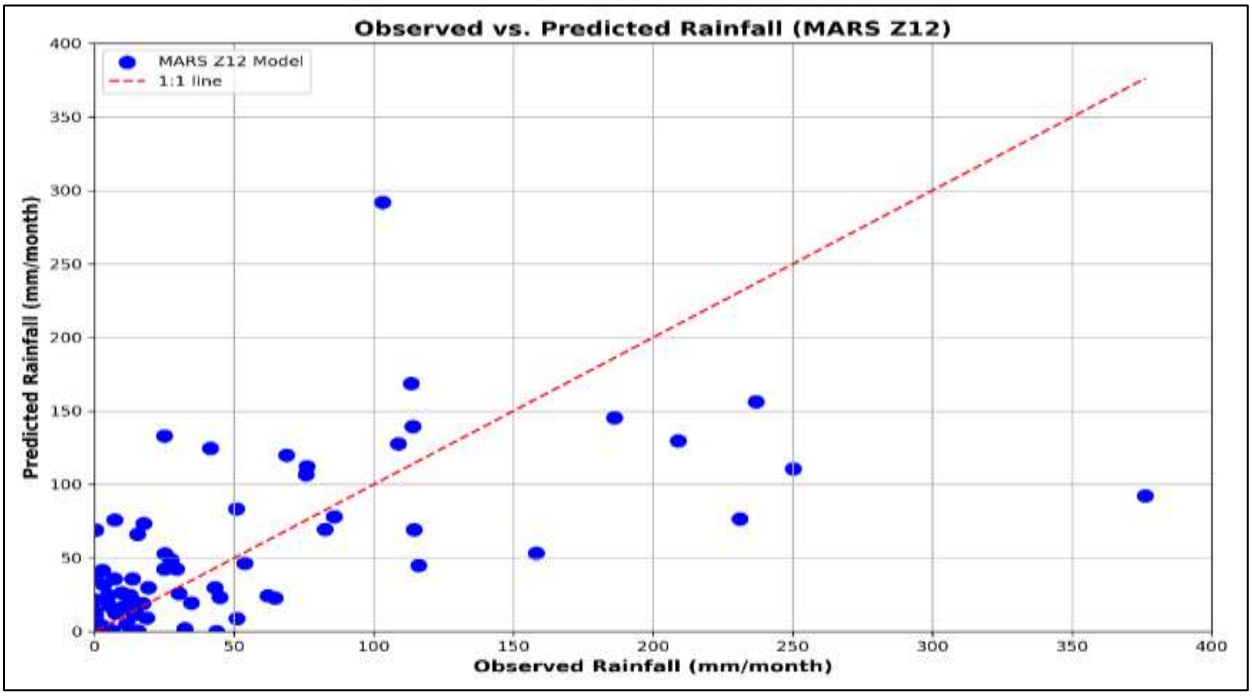

Figure 8. (b) Scatter plot of model predictions using the testing dataset for MARS at Z12 for Farukh Nagar, Haryana (all datasets)

DISCUSSIONS

In this research, Multivariate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), Gaussian Process Regression (GPR), and Random Forest (RF) models were employed to forecast rainfall in Farukh Nagar, Haryana, India. Figure 10 presents the agreement plots for the MARS Z12, SVM PUK Z1, SVM RBF Z1, GPR PUK Z1, GP RBF Z1, and RF Z1 models during the testing phase. The plots reveal that the predicted rainfall values from the SVM RBF Z1 model closely follow the line of perfect agreement, indicating superior predictive alignment. As summarized in Table 7, SVM models demonstrated strong performance during the training phase, as reflected by high correlation coefficients (CC) and low values of Scatter Index (SI), RMSE, and MAE. During testing, the SVM RBF Z1 model outperformed all other models for rainfall prediction in Farukh Nagar, achieving a CC of 0.7184, MAE of 24.4519, RMSE of 46.8608, and SI of 1.1873. Conversely, the MARS Z12 model exhibited the lowest predictive performance, with CC = 0.6199, MAE = 33.2684, RMSE = 53.7740, and SI = 1.3625 during the testing phase.

Figure 9 shows the hydrographs comparing predicted and observed rainfall for all models, highlighting that predictions from the SVM RBF Z1 model closely match the actual rainfall patterns. Additionally, the Taylor diagram in Figure 11 provides a comprehensive comparison of model performance, showing the SVM RBF Z1 as the best-performing model (red circle) followed by MARS Z12 (blue circle). Both of these lie close to the reference hollow black circle, while the MARS Z12 model (green circle) is positioned farthest from the reference, indicating the lowest predictive accuracy. The Taylor diagram (Taylor, 2001) is a widely adopted statistical tool that effectively summarizes model performance by integrating correlation, standard deviation, and root mean square error into a single plot. It is particularly valuable for evaluating complex models with multiple performance aspects. In the present study, Taylor diagrams were generated in R using the plotrix package (Figure 11) to compare different machine learning techniques for predicting the monthly rainfall. Results indicated that while all techniques performed satisfactorily during training, SVM RBF Z1 achieved the highest prediction accuracy in testing. Conversely, MARS Z12 was the least effective, showing a correlation of 0.66. Figure 12 shows the violin plots of residual errors between actual and predicted values for different models. The results indicate that the SVM RBF Z1 model performed the best, evidenced by its narrowest violin width.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

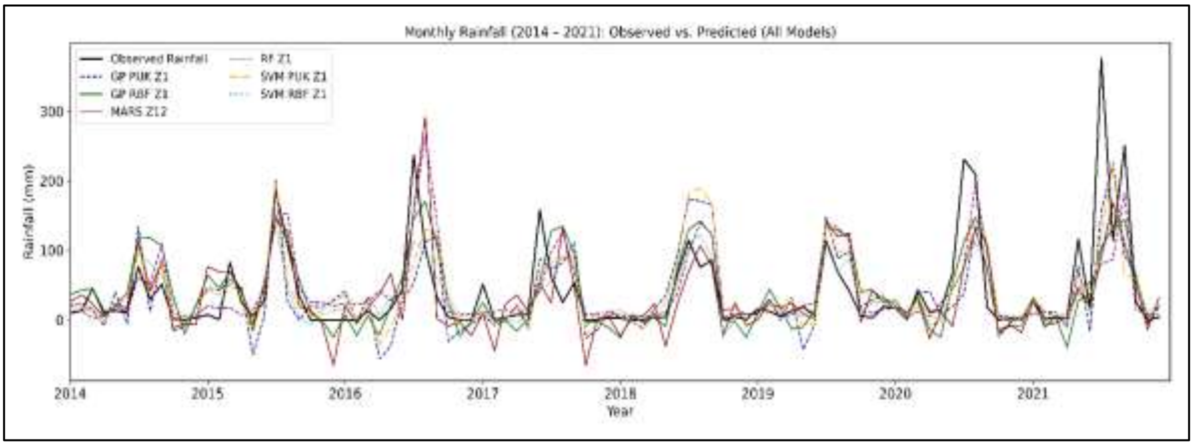

Figure 9: Comparison of observed and model-predicted rainfall at Farukh Nagar, Haryana using all implemented machine learning models.

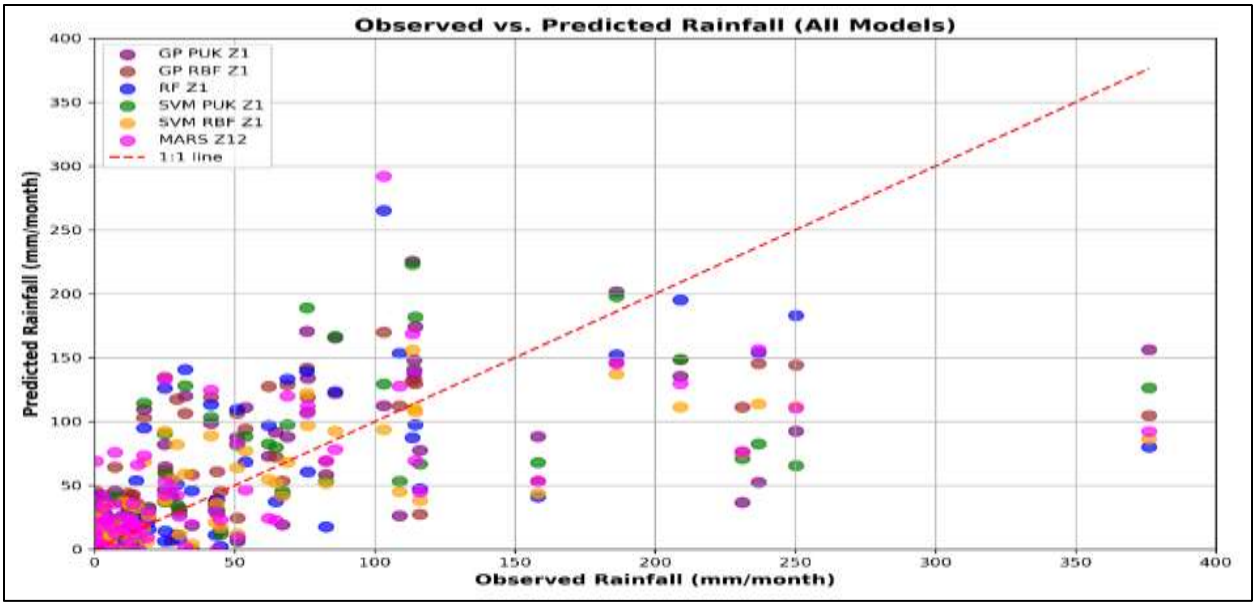

Figure 10: Agreement plots illustrating the performance of all implemented models during the testing phase.

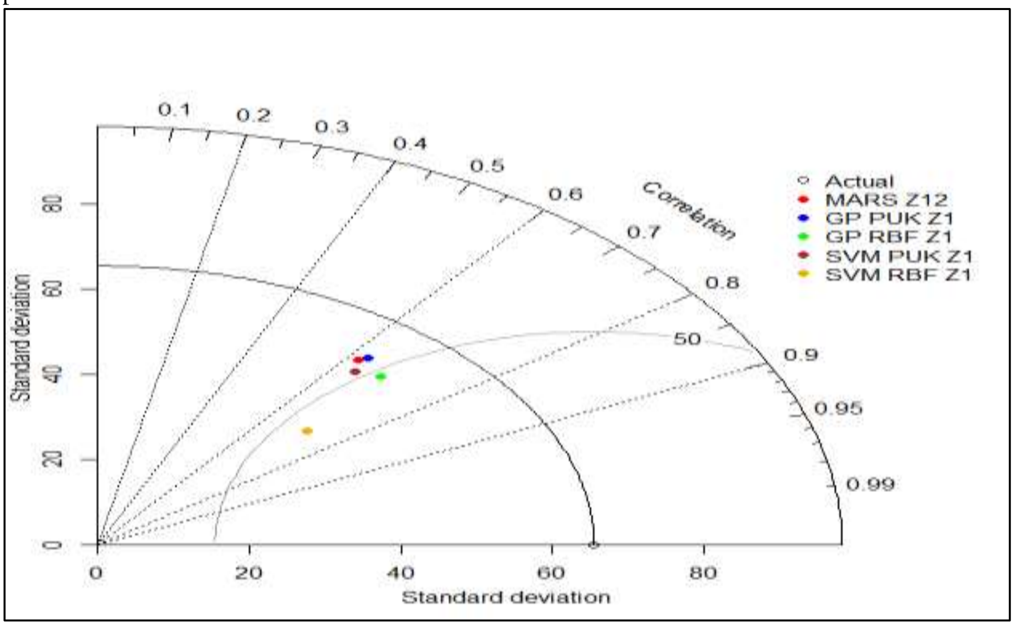

Figure 11: Taylor diagram for comparing the predictive performance of applied models for rainfall predication

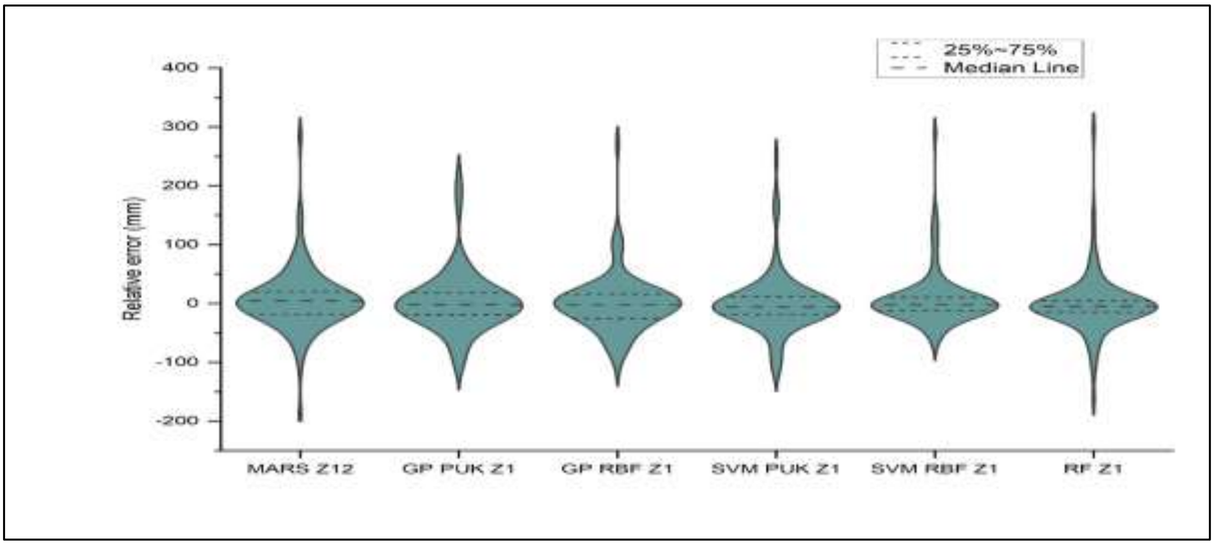

Figure 12: Violin plots of residual value comparing the performance of applied models

CONCLUSIONS

This research focused on kernel and tree-based machine learning approaches for multi-step monthly rainfall forecasting, emphasizing the integration of large-scale climate indices (IOD, ENSO, PDO, AMO, NAO) alongside local meteorological variables (maximum and minimum temperatures). A distinguishing feature of this research was the systematic optimization of model hyperparameters, ensuring that each model operated at its optimal configuration—a strategy rarely implemented in conventional rainfall prediction studies. This optimization significantly enhanced the predictive capabilities of the models.

The study employed four machine learning models such as Multivariate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), Gaussian Process Regression (GPR), and Random Forest (RF) and evaluated their performance using correlation coefficient (CC), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Scatter Index (SI) across both training and testing datasets. Among the models, SVM PUK N1 achieved the highest predictive skill (CC = 0.718, MAE = 24.45 mm, RMSE = 46.86 mm, SI = 1.19), highlighting its robustness in capturing rainfall variability at Farukh Nagar, Haryana, India. In contrast, the MARS Z12 model exhibited relatively lower accuracy, reflecting its limitations in modelling the complex, non-linear dynamics of rainfall patterns.

Key insights from the study include:

- SVM PUK demonstrated superior performance, underscoring the effectiveness of hyperparameter optimization in enhancing kernel and tree-based forecasting models.
- Integration of large-scale climate indices with local climatic variables improved model accuracy, although reliance on historical indices may reduce predictive reliability under extreme weather events or changing climate conditions.
- The proposed models offer practical benefits, providing faster, reliable, and less labour-intensive alternatives to traditional rainfall prediction methods.
- Future work should explore ensemble and hybrid modelling approaches, inclusion of real-time climate indicators, and multi-location studies to improve generalizability, robustness, and resilience under diverse climatic conditions.

The results of the present research demonstrates that carefully optimized kernel- and tree-based machine learning models, particularly SVM, can deliver reliable rainfall forecasts, supporting improved water resource management, agricultural planning, and early warning systems in the context of climate variability.

REFERENCES

1. Abrahart, R. J., See, L. M., & Kneale, P. E. (2008). *Neural networks for hydrological modeling*. Taylor & Francis.
2. Asim, M., Rashid, A., & Ahmad, T. (2022). Scour modeling using deep neural networks based on hyperparameter optimization. *ICT Express*, *8*(3), 357-362.
3. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
4. Chattopadhyay, S., Chattopadhyay, G., & Jain, R. (2010). Multivariate statistical and artificial neural network techniques in rainfall forecasting: A comparative study. *Computers & Geosciences, 36*(2), 216–225. https://doi.org/10.1016/j.cageo.2009.07.008

5. C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
6. Dhillon, M. S., Sharif, M., Madsen, H., & Jakobsen, F. (2023). Seasonal precipitation forecasting for water management in the Kosi Basin, India using large-scale climate predictors. *Journal of Water and Climate Change, 14*(6), 1868–1880. https://doi.org/10.2166/wcc.2023.479
7. Dimri, T., Ahmad, S., & Sharif, M. (2023). Impact of climate change on water availability in Bhagirathi River Basin, India. *ISH Journal of Hydraulic Engineering, 29*(5), 642–651.
8. Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics, 19*(1), 1–67. https://doi.org/10.1214/aos/1176347963
9. Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmospheric Environment, 32*(14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0
10. Kumar, K. K., Rajagopalan, B., Hoerling, M., Bates, G., & Cane, M. (2006). Unraveling the mystery of Indian monsoon failure during El Niño. *Science, 314*(5796), 115–119. https://doi.org/10.1126/science.1131152
11. Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R. News 2 (3), 18–22.
12. Mishra, A., Singh, V. P., & Jain, S. K. (2020). *Hydrological modeling and the water cycle*. Springer. https://doi.org/10.1007/978-3-030-29820-4
13. Nain, M., & Hooda, B. K. (2021). Regional frequency analysis of daily maximum rainfall in Haryana. *Mausam*, *72*(4), 835-846.
14. Naved, M., Asim, M., & Ahmad, T. (2024). Prediction of concrete compressive strength using deep neural networks based on hyperparameter optimization. *Cogent Engineering*, *11*(1), 2297491.
15. Ropelewski, C. F., & Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review, 115*(8), 1606–1626. https://doi.org/10.1175/1520-0493(1987)115
16. Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature, 401*(6751), 360–363. https://doi.org/10.1038/43854
17. Sharif, M. (2025). Prediction of Shear Capacity of FRP Reinforced Concrete Beams Using Deep Neural Networks Based on Hyperparameter Optimization. *Journal of Soft Computing in Civil Engineering*, *9*(3), 117-136.
18. Sharif, M. (2024). Shear capacity analysis of fiber reinforced polymer concrete beams. *Discover Civil Engineering*, *1*(1), 121.
19. Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88
20. Vapnik V.," Statistical Learning Theory", Wiley, New York, 1998.
21. Wang, C., Deser, C., Yu, J. Y., DiNezio, P., & Clement, A. (2014). El Niño and Southern Oscillation (ENSO): A review. In P. W. Glynn, D. P. Manzello, & I. C. Enochs (Eds.), *Coral reefs of the eastern tropical Pacific* (pp. 85–106). Springer. https://doi.org/10.1007/978-94-017-7499-4_4