KEC Conference 2025

ISSN: 3021-940X (Print) DOI: https://doi.org/10.3126/injet.v2i2.78616

Modeling Precipitation: A Statistical and Machine Learning Approach

Narayan Sapkota[1, 2, *], Khim Bahadur Khattri[1], Divyashwori Aryal[3]

[1]*Department of Mathematics, School of Science, Kathmandu University, Dhulikhel, Kavre, Nepal, ernarayan92@gmail.com*
[2]*Everest Engineering College, Pokhara University, Sanepa, Lalitpur, Nepal, khimkhattri@ku.edu.p*
[3]*School of Environmental Science and Management, Pokhara University, Kathmandu, Nepal, aryal.divya53@gmail.com*

Abstract

Precipitation modeling is indispensable for understanding hydrological processes, optimizing water resource management, and mitigating precipitation-induced geophysical mass flows such as debris and mud flow risks. This study employed both statistical and machine learning techniques to model precipitation patterns in Kathmandu, Nepal. Specifically, it compares the performance of the autoregressive integrated moving average (ARIMA) model and the Support Vector Regressor (SVR), using five years of historical meteorological data from the Department of Hydrology and Meteorology, Babarmahal. Stationarity tests confirmed the dataset's stationarity, allowing the use of ARIMA without differencing. Initial ARIMA parameters were selected using ACF and PACF plots and refined through grid search with AIC scores. Performance evaluation demonstrated that SVR outperformed ARIMA in both in-sample and out-of-sample predictions, as indicated by lower RMSE and MAE and higher Nash-Sutcliffe Efficiency (NSE) values, suggesting a better data fit and generalization. Furthermore, the Durbin-Watson statistic confirmed SVR's superior handling of autocorrelation, reinforcing its effectiveness in precipitation forecasting. However, both models faced challenges in accurately predicting extreme precipitation events. Future research should explore a hybrid approach integrating statistical models, machine learning, and deep learning to enhance predictive performance. Additionally, further consideration of meteorological and geographical parameters may improve model accuracy. These findings can support meteorologists, urban planners, disaster management authorities, and policymakers in strengthening early warning systems, improving water resource management, and enhancing resilience against climate-related challenges.

Keywords: Precipitation Prediction, ARIMA, SVR, Residual Analysis, Durbin-Watson Test, NSE

1. Introduction

Precipitation plays a pivotal role in shaping environmental and socio-economic landscape of Kathmandu, a city nestled within the complex topography of the Himalayan foothills. Governed primarily by the South Asian monsoon, Kathmandu experiences stark seasonal variability, with torrential precipitation between June and September often leading to urban flooding, landslides, and infrastructure disruptions (Anwar et al., 2021; Endalie et al., 2022; Kafle and Hooda, 2023; Nayak et al., 2024; Rahman et al., 2022; Singh et al., 2024). Moreover, Kathmandu's increasing urbanization and vulnerability to extreme weather events, accurate precipitation forecasting is of paramount importance for disaster preparedness, resource management, and sustainable urban planning. In recent years, the intensity and frequency of extreme precipitation events have surged, amplifying the risks associated with urban flooding. Government authorities have allocated significant portions of their budgets toward flood response, recovery, and compensation for affected communities. However, the economic burden of post-disaster recovery far outweighs the cost of proactive mitigation strategies. A robust precipitation forecasting system, integrated with an early warning mechanism, could significantly enhance disaster preparedness, allowing authorities to issue timely alerts, optimize evacuation plans, and minimize casualties(Akbar et al., 2024; Barrera-Animas et al., 2022; F. Abd-Elhamid et al., 2024a; Singh et al., 2024). Accurate precipitation forecasts assist water resource managers in optimizing water distribution, reducing the impact of droughts, and avoiding floods, thereby supporting more efficient and sustainable water management (Saleh et al., 2024; Singh et al., 2024). Consequently, developing an effective precipitation prediction model is imperative, not only to mitigate flood risks but also to optimize water resource management. Despite the pressing need for accurate precipitation predictions, the intricate interplay of atmospheric variables, topographic influences, and climate variability poses a formidable

**Corresponding Author*

challenge (Rahman et al., 2022; Singh et al., 2024). Conventional statistical model, for instance, ARIMA, have been extensively employed, for example (F. Abd-Elhamid et al., 2024a; Mishra et al., 2024; Sharma and Singh, 2024; Singh et al., 2024) for precipitation prediction, yet their reliance on linear assumptions limits their ability to capture the highly non-linear and chaotic nature of meteorological data. However, operation complexities and computational efficiency have limited the applicability of these models on a large scale (Di Nunno et al., 2022; Latif et al., 2023; Mishra et al., 2024; Saleh et al., 2024). Therefore, data-driven models should be a valid alternative, in particular, with limited time series data. The fundamental benefit of AI models is their ability to simulate non-linear and natural phenomena and/or events without requiring a thorough comprehension of their internal and or external mechanics (Saleh et al., 2024). Furthermore, an AI-based approach has proved to be the most reliable technique, allowing high computational speed without the need to define analytical relationships between the input data and the target variable (Di Nunno et al., 2022). This led the AI models to be widely used for the hydro-meteorological phenomena modeling, and presents a paradigm shift by leveraging multi-variable dependencies and intricate patterns within atmospheric conditions.

The extensive literature on various machine learning models for precipitation prediction from 2000 to 2023 can be found in (Saleh et al., 2024). Similarly, literature on precipitation prediction using remote sensing and machine learning from 2010 to 2023 is available in (Latif et al., 2023). In recent times, hybrid machine learning approaches, which combine two algorithms to enhance the model's predictive power, are gaining popularity, for instance, DWT-SVR-Prophet model (Li et al., 2023), and a hybrid M5P-SVR (Rahman et al., 2022), (Di Nunno et al., 2022) as well as a review of hybrid models is provided by (Dotse et al., 2024). However, a systematic, comparative evaluation of these methodologies, particularly within the context of Nepal's unique climatology, remains scarce in existing literature.

This study aims to address the limitations of traditional linear-statistical models and explore the predictive capabilities of non-linear machine learning models in the context of Kathmandu's precipitation. Specifically, it compares ARIMA and SVR models using five years of historical meteorological data. While several studies have applied hybrid or deep learning models in other regions, our work is novel in its comparative focus within Nepal's unique climatological setting. The analysis incorporates not only conventional evaluation metrics (RMSE, MAE) but also the NSE and residual analysis via the Durbin-Watson test, offering a more holistic model evaluation. This dual-evaluation framework, combining a detailed performance comparison, enables a more nuanced and interpretable assessment beyond traditional error metrics. The findings hold significant implications for meteorologists, urban planners, disaster management authorities, and policymakers, providing data-driven insights to enhance early warning systems, optimize water resource allocation, and bolster Kathmandu's resilience against climate-induced adversities.

2. Related work

The work by Yulianto et al. (2020) explored precipitation forecasting using SVR and Particle Swarm Optimization (PSO). SVR is effective for predicting non-linear precipitation data, but selecting the right parameters is crucial for accuracy. The study highlighted that the ANOVA RBF Kernel is the most effective choice for precipitation forecasting. In a similar vein, Liyew and Melese (2021) focused on daily precipitation intensity forecasting, comparing Multiple Linear Regression (MLR), Random Forest, and XGBoost. The key atmospheric features are identified through the Pearson correlation coefficient with a threshold of 0.20. The findings revealed that year, month, day, and wind speed had minimal impact, relative humidity and daily sunshine were strongly correlated with precipitation prediction. Based on the results, XGBoost was the most effective algorithm for predicting daily precipitation. Further, Ban et al. (2022) conducted a comparison of GRU and XGBoost algorithms in predicting precipitation in Yibin City, Sichuan Province. The results demonstrate that the XGBoost model outperforms the GRU model, with lower prediction errors.

Similarly, Barrera-Animas et al.(2022) compared several AI models, including LSTM, Stacked-LSTM, Bidirectional-LSTM, XGBoost, and an ensemble of Gradient Boosting Regressor, SVR, and Extra-Trees Regressor, for hourly precipitation forecasting using climate data from five UK cities (2000–2020). The findings suggest that LSTM-based models with fewer hidden layers are more efficient and cost-effective for

precipitation forecasting, making them ideal for budget-conscious applications. In a similar context, Di Nunno et al. (2022) focused on precipitation forecasting in the tropical monsoon climate of northern Bangladesh using M5P and SVR along with a novel hybrid model combining both algorithms. The results showed that the hybrid M5P-SVR model provided the best predictions. The research work by Rahman et al.(2022) introduced a novel real-time precipitation prediction system for smart cities using a machine learning fusion technique. Given the challenges of accurate precipitation prediction due to extreme climate variations, the system combines four widely used supervised machine learning techniques: decision tree, Naïve Bayes, K-nearest neighbors, and support vector machines. To enhance prediction accuracy, fuzzy logic is incorporated to integrate the results of these techniques. The results show that the fusion-based machine learning approach outperforms other prediction models.

The study Akbar et al.(2024) compared the accuracy of monthly precipitation predictions using Seasonal ARIMA and LSTM models in Indonesia. SARIMA performs well with linear time series data, capturing trends and seasonality, but struggles with extreme fluctuations. In contrast, LSTMs are better suited for complex, non-linear relationships and can model both general trends and more intricate variations. Similarly, F. Abd-Elhamid et al. (2024b) utilized the ARIMA model to assess precipitation trends in Syria. Their findings revealed a decreasing trend in precipitation, which could significantly impact water resources, agriculture, and food security. Mishra et al. (2024) compared statistical models ARIMA with machine learning approaches (SVR, ANN, Random Forest, and XGBoost) to forecast precipitation patterns across different seasons in India. Their results indicated that XGBoost outperformed statistical models, particularly in capturing complex, non-linear precipitation patterns, while ARIMA tended to overfit the data. This highlights XGBoost's effectiveness for accurate precipitation forecasting.

Furthermore, the work by Nayak et al. (2024) investigated hydrometeorological processes and analyzed temporal precipitation variability to predict precipitation variations in the Udupi district, Karnataka, India. Using the SVR model, they validated their results against actual precipitation data and found that SVR provided reliable predictions, particularly for agricultural applications. The work by Sharma and Singh (2024) employed the ARIMA model to predict monthly precipitation time series for eight localities in the Mandi district of Himachal Pradesh over 30 years. Their study applied the Augmented Dickey–Fuller test to check for stationarity before fitting the ARIMA model and selected the best models using several criteria, including the Akaike Information Criterion and Bayesian Information Criterion. The model's accuracy was validated using the Ljung-Box test. The forecasted data's mean and standard deviation closely matched the actual data, indicating that the ARIMA model produced reliable monthly precipitation predictions. Similarly, Singh et al. (2024) also utilized the ARIMA model for precipitation forecasting, reinforcing its effectiveness in predicting seasonal precipitation patterns.

In the context of Nepal, Paudel and Yogi (2020) conducted a comparative analysis of three machine learning algorithms: Random Forest, Decision Tree, and Support Vector Machine (SVM) for the prediction of precipitation. The daily precipitation data were collected via multiple weather stations across Nepal from the period 2012 to 2013. The study included various attributes such as the date, station location, and temperature. The findings revealed that the Random Forest algorithm outperformed the other algorithms, achieving the highest scores in accuracy (80.56%), precision (74.50%), recall (76.50%), and F-measure (75.50%). This work highlighted the effectiveness of the ensemble method, Random Forest, in modeling non-linear relationships and capturing the variability in precipitation patterns, especially in Nepal's diverse topography. Another contribution, Kafle and Hooda (2023) focused on forecasting average annual precipitation in Bharatpur, Nepal. By comparing the performance of ARIMA and Exponential Smoothing models using adjusted annual precipitation data from 1990 to 2021, they aimed to identify the most reliable model for long-term precipitation prediction. The study concluded that ARIMA outperformed Exponential Smoothing, making it a more suitable model for predicting annual precipitation trends in the region. This research forms a valuable foundation for future studies on precipitation prediction models in various parts of Nepal, including Kathmandu. Most recently, Pujara and Paudel (2024) utilized historical meteorological data from the Department of Hydrology and Meteorology in Babarmahal, Kathmandu, to forecast precipitation in the Kathmandu metropolitan area. Their study employed two deep learning models, LSTM and Gated Recurrent Unit (GRU). The results indicated that the GRU model outperformed LSTM, achieving an RMSE of 2.31, an

MAE of 1.51, and an R^2 of 0.95, suggesting a better fit and more accurate precipitation predictions. Despite the GRU's lower RMSE and MAE, the visual inspection of its output revealed some inconsistencies in its predictions, raising questions about the model's reliability for real-world forecasting applications.

Despite the significant studies applying statistical, machine learning, and deep learning, there remains a notable gap in comparative studies that assess the performance of ARIMA and SVR models on local meteorological data within Nepal. Most prior work either focuses on machine learning models or deep learning without examining how simpler models perform under limited data conditions. This study aims to fill that gap by benchmarking these two distinct approaches, statistical vs. machine learning, on a localized dataset from Kathmandu.

Deep learning architectures such as LSTM and GRU are powerful in capturing temporal dependencies in large and complex datasets, however, they are computationally intensive. Provided that limited five-year dataset and emphasis on model interpretability, we focused on ARIMA and SVR models, which are more suitable for smaller datasets and require fewer computational resources.

3. Methodology

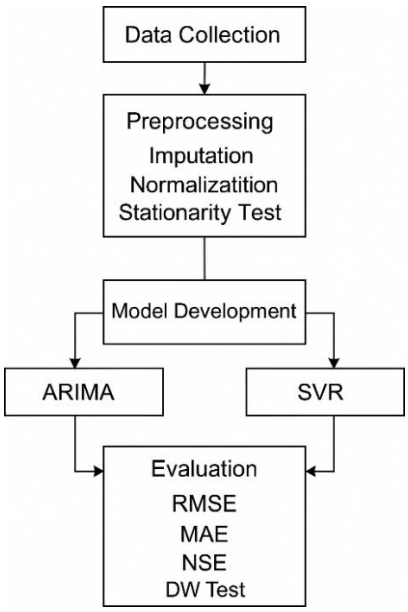

Figure 1. Conceptual Flowchart

Figure illustrates the conceptual framework adopted in this study. The process begins with the collection of meteorological data, followed by preprocessing steps such as handling missing values and normalizing inputs. The refined dataset is subsequently used to train and test two predictive models: ARIMA and SVR. Finally, the model outputs are evaluated using performance metrics including RMSE, MAE, and NSE, alongside residual analysis with the Durbin-Watson test. This structured workflow ensures a transparent, repeatable methodology and supports comparative analysis between statistical and machine learning approaches.

3.1 Data Collection

The historical meteorological data were collected from the Department of Hydrology and Meteorology, Babarmahal, Kathmandu. The dataset includes records from Kathmandu Airport, located in the Kathmandu District. The station index for this location is 1030, and its geographical coordinates are Latitude: 27.703825 and Longitude: 85.35624722. The elevation of the station is 1337 meters above sea level. The data spans from January 1, 2020, to December 31, 2024, a time series of meteorological measurements, including precipitation and other relevant weather variables, as shown in

Table 1.

Table 1. Weather Features used as Predictors

Feature Name	Description	Units
Time	Timestamp for each data entry (in YYYY-MM-DD HH:MM: SS format).	No unit
Precipitation	The amount of precipitation recorded during each period.	Millimeters(mm)
Max Temperature	The maximum temperature recorded.	Degrees Celsius (°C)
Min Temperature	The minimum temperature recorded.	Degrees Celsius (°C)
Relative Humidity	The percentage of moisture in the air.	Percentage (%)
Pressure	The atmospheric pressure.	Hectopascals (hPa)
Wind Direction	The direction from which the wind is blowing.	Degrees Celsius (°C)
Wind Speed	The speed at which the wind is blowing.	Meters per second (m/s)

The selection of meteorological features was based on domain knowledge and literature consensus indicating their influence on precipitation. However, no automated feature selection technique (e.g., Recursive Feature Elimination or feature importance via model weights) was applied in this study. Although the five-year dataset provides valuable insights into recent precipitation patterns, the relatively short temporal span may limit the model's ability to capture long-term trends. Extending the dataset or incorporating additional stations across varying topographies could enhance model robustness and generalizability.

3.2 Data Preprocessing

3.2.1 Imputation

The data preprocessing phase began with a thorough examination for any missing or NaN values within the dataset. To address this issue, *linear interpolation* was applied to fill in the missing values or NaN. Linear interpolation estimates missing data points by using the values of neighboring data points, ensuring that the dataset remains continuous and complete for further analysis. This method was chosen because it preserves the overall trend and structure of the data, making it suitable for time series data.

3.2.2 Standardization (Z-score Normalization)

To ensure that all features in the dataset contribute equally to the model and prevent the model from being biased toward variables with larger scales, we applied Z-score normalization to the data. This technique transforms the features such that they have a mean of 0 and a standard deviation of 1. The formula used for Z-score normalization is:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Where Z is the standardized value, X is the original value, μ is the mean of the feature(s), $\sigma$is the standard deviation of the feature(s).

3.2.3 Stationarity Test

After completing the normalization process, we conducted the Augmented Dickey-Fuller Test (ADF) (Mushtaq, 2011), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test (Syczewska, 2010), and Phillips-Perron (PP) Test (Breitung and Franses, 1998) for stationarity of the time series data. This step of preprocessing is applied only to the ARIMA model, not the SVR model.

3.2.4 Train-Test Split

With the data now prepared, it was divided into two distinct sets: a training set, containing data from January 1, 2020, to December 31, 2023, and a testing set, consisting of data from January 1, 2024, onward. This split

ensures that the model is trained on past data and evaluated on unseen data, enabling us to assess its performance and ability to generalize to future conditions.

3.3 Learning Algorithms

3.3.1 Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA model (Mishra et al., 2024; Singh et al., 2024) is a widely used method for forecasting non-stationary time-series data, particularly when data exhibit patterns such as trends, seasonality, and noise. It combines three components, namely *Autoregressive (AR), Integrated (I), and Moving Average (MA).* The AR component models the relationship between an observation at time *t* and several lagged observations. Essentially, it reflects the ***'memory'*** of the series; how much past data influences current data. The AR(p) model is defined as:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \tag{2}$$

Where Y_t is the observed value of the time series at time t, $\phi_1, \phi_2, \ldots, \phi_p$ are the AR coefficients (parameters), p is the number of lagged observations (AR order), ϵ_t is the error term (white noise) at time t.

The I component is used to make a time series stationary by differencing it. Stationary data have a constant mean and variance over time. The Integrated part involves differencing the time series to make it stationary, which removes trends and seasonality. $(1-B)^d Y_t$ is the d^{th} difference of Y_t with the backshift operator B. The MA component models the relationship between an observation at time *t* and the residual errors from a moving average model applied to lagged observations and is given by:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \tag{3}$$

Where $\theta_1, \theta_2, \ldots, \theta_q$ are the MA coefficients (parameters), q is the number of lagged forecast errors in the MA model. This model captures the dependency of Y_t on past terms.

The general form of the ARIMA (p, d, q) model combines all three components into a single equation:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right)(1-B)^d Y_t = \mu + \epsilon_t(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \tag{4}$$

Where $\phi_1, \phi_2, \ldots, \phi_p$ are the AR parameters, $\theta_1, \theta_2, \ldots, \theta_q$ are the MA parameters, d is the degree of differencing. After differencing, model Eq. (4) can be rewritten as:

$$Y_t' = \mu + \phi_1 Y_{t-1}' + \phi_2 Y_{t-2}' + \cdots + \phi_p Y_{t-p}' + \epsilon_t + \theta_1 \epsilon_{t-1} + +\theta_2 \epsilon_{t-2} + \cdots + +\theta_q \epsilon_{t-q} \tag{5}$$

where $Y_t' = (1-B)Y_t$ is the differenced series and ensures stationarity by removing trends and seasonality. Also, Eq. (5) explicitly shows how the current value of the differenced series. Y_t' depends on its past values and past errors.

3.3.2 Support Vector Regression (SVR)

Support vector Regression algorithm is a supervised learning model. SVR has demonstrated itself as highly reliable and effective for regression tasks (Di Nunno et al., 2022). SVR builds a hyperplane in high-dimensional space and can precisely distinguish objects from Kernel functions in linear or nonlinear data. The objective of the SVR is to find a function f(x) whose deviation from the target values y_i remains below a predefined threshold ε. Given the training dataset: $\{(x_i, y_i), i = 1, \ldots, l\} \subset X \times R$ where X indicates the input space, the objective is to minimize the Euclidean norm $||w||^2$. This is achieved by solving a constrained convex optimization problem to find a linear function, $f(x) = \langle w, x \rangle + b$, where $b \in R$ and $w \in X$. Additionally, slack variables are introduced to accommodate deviations beyond ε, ensuring the greater flexibility in the model. The optimization can be expressed as:

$$\text{Minimize}: \frac{1}{2}||\mathrm{w}||^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{l}}(\xi_\mathrm{i} + \xi_\mathrm{i}^*) \tag{6}$$

Subject to:

$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i$$
$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*$$

where deviation and function flatness depend on the constant C, which is greater than 0. The effectiveness of SVR also depends on the selection of the kernel function, which defines the feature space, and of its parameters.

3.4 Evaluation Metrics

3.4.1 RMSE

It quantifies the average magnitude of prediction errors by taking the square root of the mean of the squared differences between observed and predicted values. It is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{7}$$

3.4.2 MAE

MAE is another widely used metric that measures the average magnitude of prediction errors without considering their direction. MAE provides a straightforward interpretation by indicating the average absolute deviation between observed and predicted values. Unlike RMSE, MAE treats all individual differences equally. It is formulated as:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_t| \tag{8}$$

3.4.3 Nash-Sutcliffe Efficiency (NSE)

NSE is a normalized statistical metric commonly used to evaluate the predictive skill of hydrological models. It compares the predictive performance of a given model to the mean of the observed data and is defined as (McCuen et al., 2006):

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{9}$$

The NSE value ranges from $-\infty$ to 1, where an efficiency of 1 indicates a perfect match between predictions and observations, while an efficiency of 0 suggests that the model performs no better than the mean of the observed data. A negative NSE value implies that the model is performing worse than simply using the mean of observed values as predictions.

3.4.4 The Durbin-Watson (DW) Tests of Residuals

The Durbin-Watson (DW) test is used to detect autocorrelation in the residuals. Autocorrelation occurs when the residuals (errors) are correlated with each other, which violates the assumption of independence required for models. The DW test statistic ranges from 0 to 4. A value of 2 indicates no autocorrelation in the residuals, suggesting that they are independent (Bartels and Goodhew, 1981). If the DW statistic is less than 2, it implies positive autocorrelation, meaning the residuals are positively correlated. Conversely, if the statistic is greater than 2, it indicates negative autocorrelation, suggesting that the residuals are inversely correlated. The Durbin-Watson statistic d is calculated using the following formula (Bartels and Goodhew, 1981):

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \tag{10}$$

where e_t is the residual at time t, and n is the number of observations. A statistic close to 2 indicates no significant autocorrelation, making it an important diagnostic tool in regression analysis.

3.5 Model Selection

Two different algorithms, ARIMA and SVR, were employed to train the models on the training dataset. For the ARIMA (p, d, q) model, the hyperparameters to be tuned include p (the number of lag observations), d (the degree of differencing), and q (the size of the moving average window). The optimal combination of these hyperparameters was determined using the Akaike Information Criterion (AIC) score (Vrieze, 2012). For the SVR algorithm, hyperparameter tuning was performed using grid-search cross-validation with 5-fold cross-validation. The key hyperparameters tuned were C (the regularization parameter), epsilon (the margin of tolerance), and gamma (the influence of individual training points). This approach involved an exhaustive search across multiple combinations of hyperparameters, with model performance evaluated through cross-validation. The best-performing combination was then selected. Finally, the residuals from both in-sample and out-of-sample predictions were used to compare the performance of these two models.

4. Results and Discussion

4.1 Exploratory Data Analysis (EDA)

The EDA provided a comprehensive understanding of the meteorological dataset used. It highlighted the distribution of key weather variables, their relationships, and temporal trends. The time-series analysis (see Figure) of precipitation from 2020 to 2025 highlighted clear seasonal variations, with distinct wet and dry periods. Peaks in precipitation indicate extreme precipitation events that occur periodically, potentially driven by monsoonal patterns or other environmental factors. Extended dry periods further emphasize the cyclic nature of precipitation in Kathmandu. To summarize, the presence of extreme weather outliers suggested the need for further investigation into climate trends and predictive modeling to anticipate future variations effectively.

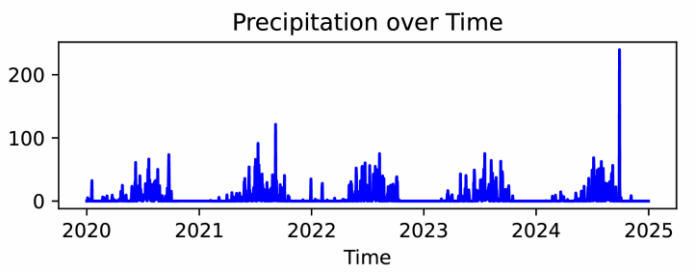

Figure 2. Precipitation from 2020-01-01 to 2024-12-31

The correlation matrix (see Figure) further explained the interactions between meteorological variables. Precipitation exhibited a weak correlation with most variables but showed a moderate correlation with relative humidity (0.37) and minimum temperature (0.32), suggesting these features may be useful predictors in non-linear models such as SVR. The positive correlation with humidity is expected, as higher moisture content in the air contributes to cloud formation and precipitation. Additionally, the relationship with minimum temperature may indicate that warmer nights enhance atmospheric moisture retention, supporting precipitation events. Moreover, precipitation had a moderate negative correlation with pressure (-0.32), which aligns with the well-established link between low-pressure systems and increased precipitation. Lower pressure often leads to atmospheric instability, rising air masses, and enhanced cloud formation, all of which contribute to precipitation. This relationship is consistent with previous meteorological studies that associate cyclonic activity with significant precipitation patterns.

These results suggest that other factors, such as topography, proximity to water bodies, convection, cloud formation, human activities, and solar radiation, may play a significant role in shaping precipitation patterns. For example, urbanization and land-use changes can alter local microclimates, influencing precipitation variability. Additionally, proximity to large water bodies may contribute to localized precipitation patterns through increased evaporation. Future studies incorporating these elements, along with high-resolution spatial and temporal datasets, could provide a more comprehensive understanding of precipitation variability.

Further research comparing these findings with different climatic zones could also help assess the generalizability of the observed correlations and improve predictive models for precipitation forecasting.

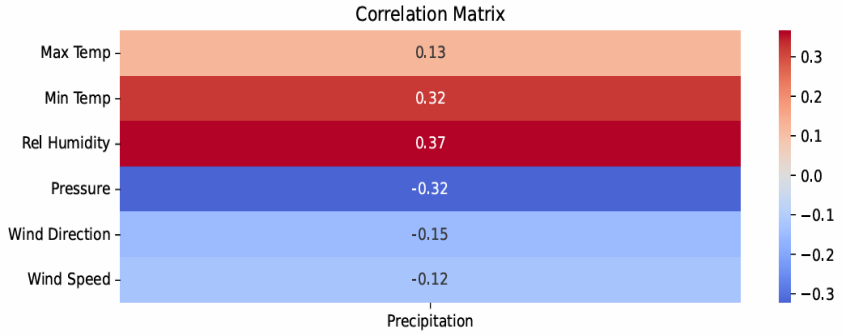

Figure 3. Correlation between independent variables and dependent variable

The box plots shown in **Error! Reference source not found.** provide a statistical summary of meteorological v ariables, revealing their distribution, variability, and presence of outliers. The precipitation box plot showed a highly skewed distribution with numerous outliers, indicating that most precipitation events were minimal, but occasional extreme precipitation events exceeded 200 mm. This suggested sporadic precipitation patterns, which could have been influenced by localized convection, monsoonal activity, or topographical effects. Similarly, maximum and minimum temperature distributions highlighted variations in daytime and nighttime temperatures. The maximum temperature remained relatively stable, ranging between approximately 10°C and 35°C, whereas the minimum temperature exhibited a wider interquartile range, indicating greater variability during nighttime cooling. These variations were likely due to seasonal changes, land surface properties, or urban heat effects.

Relative humidity was generally high, clustering between 40% and 100%, with some lower-end outliers indicating occasional dry conditions, possibly due to heatwaves or dry air advection. Atmospheric pressure remained relatively stable within a narrow range of approximately 855 hPa to 870 hPa, though some outliers suggested occasional fluctuations due to low-pressure systems or regional atmospheric disturbances. Wind direction was widely dispersed between 0° and 250°, reflecting varying wind patterns influenced by seasonal shifts, large-scale atmospheric circulation, or local geographical features. Wind speed, on the other hand, exhibited a highly skewed distribution with multiple outliers, signifying that while most days experienced low wind speeds, occasional strong wind events occurred, potentially associated with storms, monsoonal winds, or frontal systems.

These findings emphasized the significant variability in meteorological parameters, particularly precipitation and wind speed, which had vital implications for precipitation prediction and climate modeling. The presence of extreme precipitation events suggested the necessity for improved forecasting models to predict heavy precipitation episodes more accurately. Additionally, the variability in temperature, humidity, and pressure highlighted potential interactions between these factors that warranted further investigation. Future research should consider integrating additional factors such as topography, oceanic influences, and urbanization effects to develop a more comprehensive understanding of local weather dynamics and improve predictive capabilities.

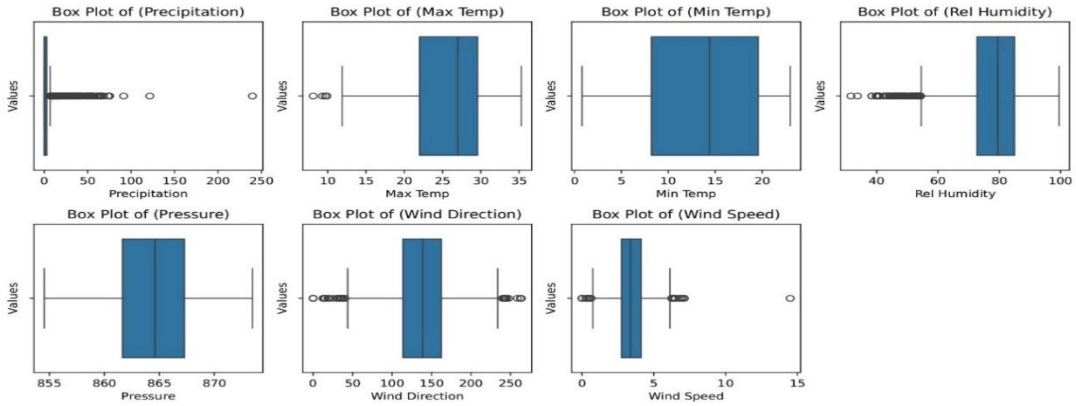

Figure 4. Box plot of dependent and independent variables

4.2 Stationarity Check

The results of the stationarity tests (see Table 2) were based on three different statistical tests: the ADF test, the KPSS test, and the PP test. The ADF test yielded a p-value of 0.001, which was below the 0.05 significance level. Additionally, the ADF statistic (-4.087) was more negative than the critical values at all significance levels (1%, 5%, and 10%). These results led to the rejection of the null hypothesis, indicating that the series was stationary. In contrast, the KPSS test produced a p-value of 0.1, which was greater than 0.05, suggesting that the null hypothesis could not be rejected. The KPSS statistic (0.124) was below the critical value at the 10% significance level (0.347), reinforcing the conclusion that the series was stationary. The PP test also yielded a p-value of 0.0, which was below 0.05. The PP statistic (-40.274) was much lower than typical critical values, leading to the rejection of the null hypothesis and further supporting the conclusion that the series was stationary.

Table 2. Stationarity Test Results

Test	**Statistic**	**p-value**	**Critical Values**		
			1%	**5%**	**10%**
ADF	-4.087	0.001	-3.434	-2.863	-2.568
KPSS	0.124	0.1	0.739	0.463	0.347
PP	-40.274	0.0		N/A	

4.3 Grid Search for ARIMA and SVR Model

Figure consisted of two plots: the ACF plot and the PACF plot, which were used to determine the optimal parameters for the ARIMA (p, d, q) model. The ACF plot exhibited a slow decay, indicating potential autocorrelation in the data and suggesting an initial estimate of q = 1 or 2. However, despite the ACF indicating autocorrelation, the ADF, KPSS, and PP tests confirmed that the data were stationary (see Section 0), implying that no differencing was required. Meanwhile, the PACF plot showed a sharp cutoff after lag 1, indicating a potential autoregressive (AR) process with p = 1. Since stationarity was confirmed, the degree of differencing was set to d = 0, leading to an initial ARIMA model of ARIMA (1,0,1) or ARIMA (1,0,2). However, through further model selection using grid search optimization and Akaike Information Criterion (AIC) comparisons, the optimal parameters were determined to be p = 1 and q = 18. This result suggests that, while the PACF indicated a low-order autoregressive component, the ACF captured longer-term dependencies, necessitating a higher-order moving average component for improved model performance.

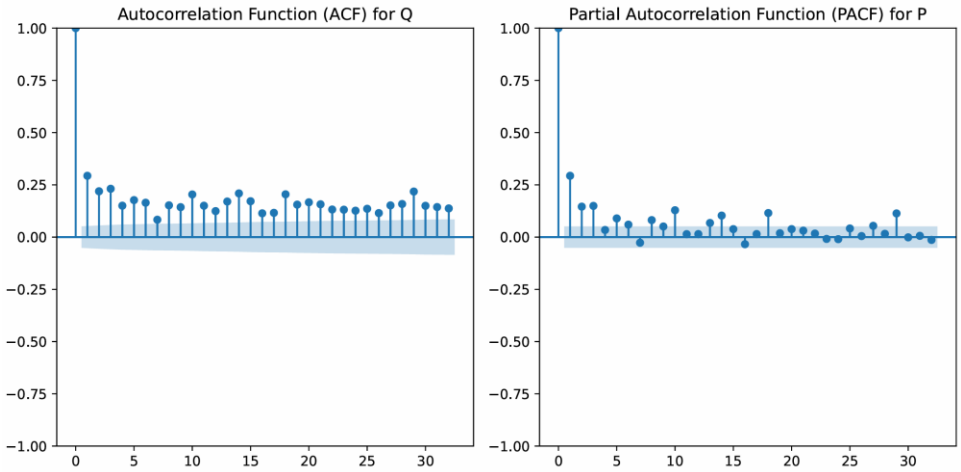

Figure 5. ACF and PACF plot of Precipitation to determine optimal parameters of p and q

In case of the SVR model, the optimal hyperparameters were determined as follows: C = 1, epsilon = 0.1, kernel = rbf, and gamma = 0.1. The C parameter, set to 1, indicated a balanced trade-off between minimizing error on the training data and ensuring the model's generalization to unseen data. A smaller C would have

likely resulted in underfitting by increasing bias, while a larger value could have caused overfitting. The epsilon = 0.1 parameter specified a margin of tolerance for errors, implying that deviations of up to 0.1 from the actual values were acceptable without penalty. This choice helped the model remain flexible while controlling complexity. The RBF (Radial Basis Function) kernel was selected, which was particularly effective in capturing non-linear relationships within the data. By mapping input features into a higher-dimensional space, the kernel enabled the model to represent complex, non-linear patterns that a linear model could not. The gamma = 0.1 parameter indicated that the influence of each training point was spread over a larger region. This choice helped avoid overfitting by reducing the model's sensitivity to noise in the training data. Together, these hyperparameters facilitated a good balance between bias and variance, allowing the SVR model to make accurate predictions while avoiding overfitting. The results suggested that the chosen parameters effectively captured the underlying trends in the data for precipitation prediction, leading to a model that generalized well to new, unseen data.

4.4 *Model Evaluation*

The results presented in Table 3 and

Table 4 compare the performance of ARIMA and SVR models for both in-sample and out-of-sample predictions. The evaluation metrics used include RMSE, MAE, NSE, and the Durbin-Watson statistic. In the in-sample prediction results (Table 3), the SVR model demonstrates superior performance compared to ARIMA, with lower RMSE (7.11454 vs. 10.3908) and MAE (3.56349 vs. 5.7384), as well as a significantly higher NSE (0.862 vs. 0.29807), indicating a better fit to the data. Additionally, the Durbin-Watson statistic suggests that SVR exhibits slightly better handling of autocorrelation. Similarly, in the out-of-sample prediction results (

Table 4), SVR continues to outperform ARIMA, with lower RMSE (8.93584 vs. 16.56769) and MAE (2.38322 vs. 7.0984), and a substantially higher NSE (0.75620 vs. 0.13702), suggesting that SVR generalizes better to unseen data. The Durbin-Watson statistic remains within an acceptable range for both models, indicating no severe autocorrelation issues. To summarize, these findings suggest that SVR provides more accurate and reliable predictions for precipitation forecasting compared to ARIMA. SVR's superior performance may stem from its ability to capture non-linear relationships between precipitation and predictors such as relative humidity, pressure, and temperature. Unlike ARIMA, which assumes linearity and stationarity, SVR leverages kernel methods to model more complex interactions, particularly relevant in a geographically and climatologically dynamic region like Kathmandu.

Table 3. In-sample prediction

Metric	**ARIMA**	**SVR**
RMSE	10.3908	7.11454
MAE	5.7384	3.56349
NSE	0.29807	0.862
Durbin-Watson Statistic	1.87043	2.0130

Table 4. Out-of-sample prediction

Metric	**ARIMA**	**SVR**
RMSE	16.56769	8.93584
MAE	7.0984	2.38322
NSE	0.13702	0.75620
Durbin-Watson Statistic	1.40943	2.087880

Although SVR outperformed ARIMA in this study, both models struggled to accurately predict extreme precipitation events. This challenge has also been reported in prior works using traditional statistical or single-model machine learning approaches. Recent studies have demonstrated that deep learning models like LSTM and ensemble methods like XGBoost may provide improved performance in capturing abrupt changes due to their ability to model temporal dependencies or learn complex feature interactions. For example, (Pujara and Paudel, 2024) showed that GRU outperformed LSTM in Kathmandu, achieving an RMSE of 2.31 and R^2 of 0.95, while (Mishra et al., 2024) found that XGBoost outperformed ARIMA across Indian regions with high variability. However, such models often require larger datasets and more hyper-parameter tuning. In contrast, our SVR model, despite its limitations, provided a reasonable trade-off between complexity and performance for smaller datasets. This suggests that hybrid approaches such as ARIMA-SVR or SVR-LSTM may offer improved results by leveraging the strengths of each model class.

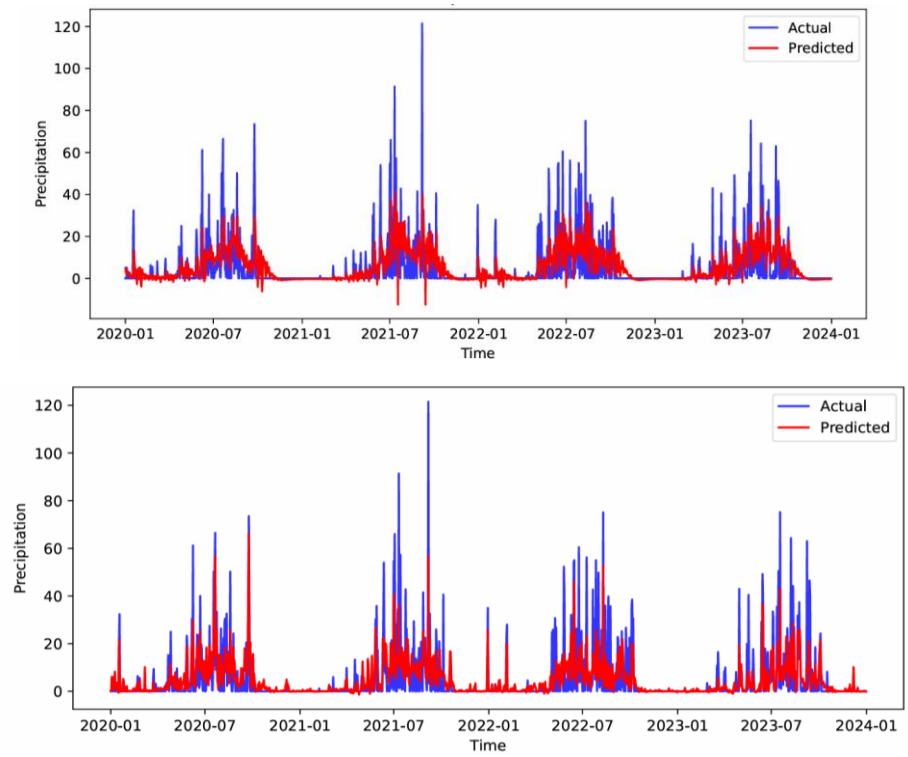

Figure 6. (Top) In-sample prediction of ARIMA and (Bottom) In-sample Prediction of SVR

Figure presents the in-sample prediction results for precipitation using the ARIMA model (top panel) and the SVR model (bottom panel). The performance of each model was assessed by visually inspecting its ability to capture the temporal variations and magnitudes of precipitation events over the study period (2020–2024). The ARIMA model (top panel) demonstrated a relatively good fit to the observed data, effectively capturing general trends and seasonal precipitation patterns. However, it tended to underestimate peak precipitation values, particularly during high-intensity events. This limitation was common in linear time series models, where sudden spikes in precipitation, likely due to extreme weather conditions, were challenging to predict accurately. Additionally, ARIMA performed well in periods of low or moderate precipitation but struggled to replicate the volatility and sharp fluctuations in precipitation intensity. In contrast, the SVR model (bottom panel) exhibited a different prediction behavior. While it captured some of the fluctuations in precipitation, it appeared to provide a more conservative prediction, with reduced variability compared to actual observations. The SVR model produced relatively smoother predictions, which could be attributed to its non-linear nature and kernel-based regression. Although SVR effectively captured complex patterns, its ability to predict extreme precipitation values remained limited, as seen in its underestimation of peak precipitation intensities

Figure illustrates the out-of-sample (testing dataset) predictions of precipitation using ARIMA (top panel) and SVR (bottom panel). The ARIMA model showed reasonable forecasting performance for moderate

precipitation levels, with predicted values aligning well with actual observations in many instances. The confidence interval provided insight into the model's uncertainty, and it remained relatively narrow in stable periods but expanded as precipitation variability increased. However, the ARIMA model struggled significantly in capturing extreme precipitation events, as evidenced by its failure to predict the sharp peak observed around September 2024. While the model followed the general trend of precipitation fluctuations, its inability to predict extreme values was a key limitation, likely due to the model's assumption of linear dependencies and stationarity in the time series. In contrast, the SVR model demonstrated better adaptability in capturing moderate fluctuations in precipitation. Compared to ARIMA, SVR exhibited greater accuracy in matching observed precipitation patterns, especially during periods of gradual change. However, like ARIMA, SVR failed to predict the extreme peak in September 2024, indicating its limitations in handling highly volatile and non-linear weather patterns. The SVR model provided a relatively smooth approximation, which, while useful in reducing noise, may have led to an underestimation of extreme precipitation values.

This comparative analysis of in-sample and out-of-sample predictions revealed distinct strengths and limitations of ARIMA and SVR models in precipitation forecasting. ARIMA, as a statistical model, proved effective in capturing periodic trends, but struggled with handling abrupt changes, which are critical for forecasting extreme precipitation events. In contrast, SVR, utilizing a machine learning approach, provided a more generalized fit, yet faced challenges in accurately predicting high-magnitude variations. Given both models' underperformance in extreme rainfall prediction, ensemble or hybrid approaches, such as ARIMA-SVR or SVR combined with LSTM, could be explored. These methods can potentially blend the temporal robustness of ARIMA with the non-linear flexibility of SVR to better capture sudden spikes.

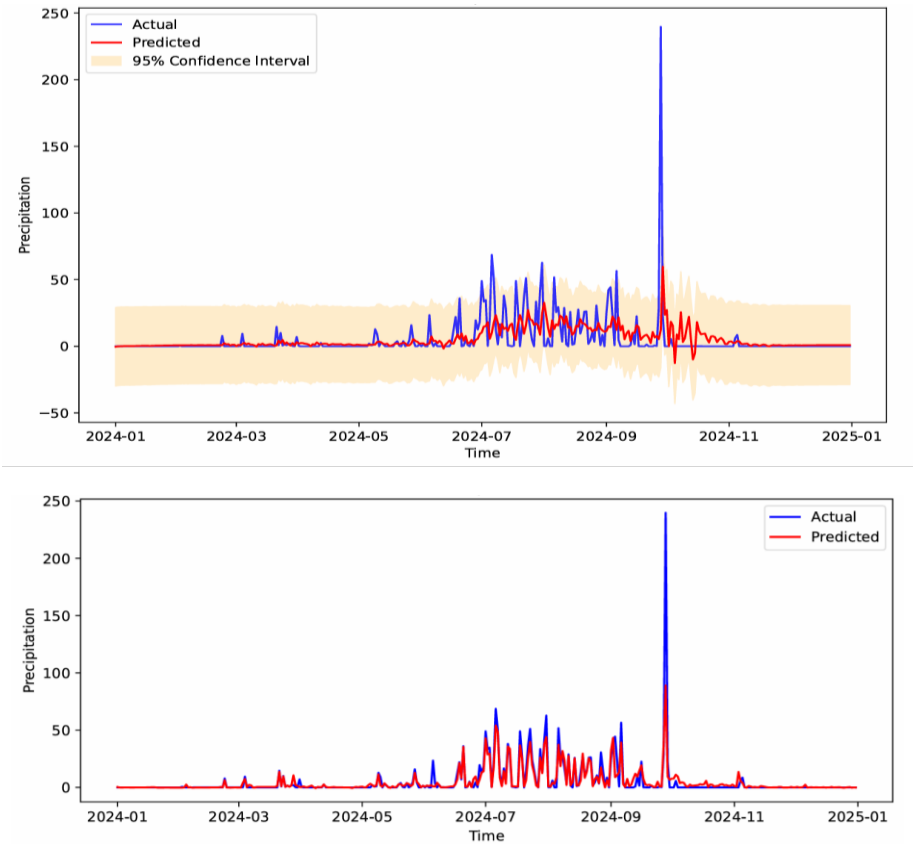

Figure 7. (Top) Out-of-sample prediction of ARIMA and (Bottom) Out-of-sample Prediction of SVR

5. Conclusions and Future Work

The analysis and comparison of ARIMA and SVR models for precipitation forecasting provided valuable insights in the context of Nepal. The stationarity tests confirmed that the dataset was stationary, allowing for the direct application of ARIMA without the need for differencing. The ACF and PACF plots served as a foundation for selecting the initial ARIMA parameters, which were further refined using a grid search with

AIC scores. In terms of model performance, the SVR model significantly outperformed ARIMA in both in-sample and out-of-sample predictions. SVR demonstrated superior accuracy, as indicated by lower RMSE and MAE values and a higher NSE, suggesting that it provided a better fit to the data and generalized more effectively to unseen observations. Furthermore, residual analysis using the Durbin-Watson statistic confirmed that SVR handled autocorrelation more effectively than ARIMA, further reinforcing its superior performance in precipitation forecasting.

Despite these promising results, both models underperformed in predicting extreme precipitation events. To address this limitation, future studies should prioritize hybrid models that integrate the strengths of statistical and machine learning techniques. For instance, combining ARIMA with SVR (ARIMA-SVR) or with LSTM (ARIMA-LSTM) may offer improved performance by capturing both linear trends and non-linear dynamics. Among these, ARIMA-SVR is recommended as a practical first step due to its lower data requirement and relative ease of implementation. On the data front, future research should incorporate more diverse and high-resolution inputs. Priority should be given to ERA5 reanalysis datasets, which offer comprehensive global meteorological fields with hourly temporal resolution and spatial granularity. Additionally, satellite-derived precipitation estimates (e.g., from TRMM or GPM missions) and remote sensing products such as soil moisture, NDVI, and land surface temperature can provide valuable context. These should be considered in conjunction with local meteorological data to enrich the feature space. Furthermore, the inclusion of topographic data, land-use patterns, distance to water bodies, and convection indices would allow models to account for the geographic and physical processes that influence precipitation. These variables are particularly important in regions like Kathmandu, where orographic effects and rapid urbanization significantly affect weather patterns. Finally, future research should also investigate longer-term datasets to analyze seasonal trends and climate anomalies. This would support the development of more generalized models, capable of contributing to climate resilience planning and long-term disaster risk reduction strategies.

Acknowledgements

We most cordially convey our gratitude to the reviewers for their careful reading, valuable comments and helpful suggestions which enable us to improve the presentation of this paper significantly.

References

Akbar, A.A., Darmawan, Y., Wibowo, A., Rahmat, H.K., 2024. Accuracy Assessment of Monthly Rainfall Predictions using Seasonal ARIMA and Long Short-Term Memory (LSTM). J. Comput. Sci. Eng. JCSE 5, 100–115.

Anwar, M.T., Winarno, E., Hadikurniawati, W., Novita, M., 2021. Rainfall prediction using extreme gradient boosting, in: Journal of Physics: Conference Series. IOP Publishing, p. 012078.

Ban, J.R., Gou, Q., Li, Y.S., 2022. Study on rainfall prediction of yibin city based on gru and xgboost, in: 2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC). IEEE, pp. 1–5.

Barrera-Animas, A.Y., Oyedele, L.O., Bilal, M., Akinosho, T.D., Delgado, J.M.D., Akanbi, L.A., 2022. Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Mach. Learn. Appl. 7, 100204. https://doi.org/10.1016/j.mlwa.2021.100204

Bartels, R., Goodhew, J., 1981. The robustness of the Durbin-Watson test. Rev. Econ. Stat. 136–139.

Breitung, J., Franses, P.H., 1998. On Phillips–Perron-type tests for seasonal unit roots. Econom. Theory 14, 200–221.

Di Nunno, F., Granata, F., Pham, Q.B., De Marinis, G., 2022. Precipitation Forecasting in Northern Bangladesh Using a Hybrid Machine Learning Model. Sustainability 14, 2663. https://doi.org/10.3390/su14052663

Dotse, S.-Q., Larbi, I., Limantol, A.M., De Silva, L.C., 2024. A review of the application of hybrid machine learning models to improve rainfall prediction. Model. Earth Syst. Environ. 10, 19–44. https://doi.org/10.1007/s40808-023-01835-x

Endalie, D., Haile, G., Taye, W., 2022. Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia. Water Supply 22, 3448–3461.

F. Abd-Elhamid, H., M. El-Dakak, A., Zeleňáková, M., O. K, S., Mahdy, M., H. Abd El Ghany, S., 2024a. Rainfall forecasting in arid regions in response to climate change using ARIMA and remote sensing. Geomat. Nat. Hazards Risk 15, 2347414. https://doi.org/10.1080/19475705.2024.2347414

F. Abd-Elhamid, H., M. El-Dakak, A., Zeleňáková, M., O. K, S., Mahdy, M., H. Abd El Ghany, S., 2024b. Rainfall forecasting in arid regions in response to climate change using ARIMA and remote sensing. Geomat. Nat. Hazards Risk 15, 2347414. https://doi.org/10.1080/19475705.2024.2347414

Kafle, S.C., Hooda, E., 2023. ARIMA and Exponential Smoothing Model to Forecast Average Annual Precipitation in Bharatpur, Nepal. BMC J. Sci. Res. 6, 113–125.

Latif, S.D., Alyaa Binti Hazrin, N., Hoon Koo, C., Lin Ng, J., Chaplot, B., Feng Huang, Y., El-Shafie, A., Najah Ahmed, A., 2023. Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. Alex. Eng. J. 82, 16–25. https://doi.org/10.1016/j.aej.2023.09.060

Li, D., Ma, J., Rao, K., Wang, X., Li, R., Yang, Y., Zheng, H., 2023. Prediction of Rainfall Time Series Using the Hybrid DWT-SVR-Prophet Model. Water 15, 1935. https://doi.org/10.3390/w15101935

Liyew, C.M., Melese, H.A., 2021. Machine learning techniques to predict daily rainfall amount. J. Big Data 8, 153. https://doi.org/10.1186/s40537-021-00545-4

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe Efficiency Index. J. Hydrol. Eng. 11, 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597)

Mishra, P., Al Khatib, A.M.G., Yadav, S., Ray, S., Lama, A., Kumari, B., Sharma, D., Yadav, R., 2024. Modeling and forecasting rainfall patterns in India: a time series analysis with XGBoost algorithm. Environ. Earth Sci. 83, 163. https://doi.org/10.1007/s12665-024-11481-w

Mushtaq, R., 2011. Augmented dickey fuller test.

Nayak, K., Nayak, S.K., Shivarama, S.B., 2024. Rainfall prediction using support vector regression in Udupi region Karnataka, India. TELKOMNIKA Telecommun. Comput. Electron. Control 23, 166–174.

Paudel, N., Yogi, T.N., 2020. Comparative study of machine learning algorithms for rainfall prediction-a case study in Nepal. Int. J. Adv. Res. Eng. Technol. IJARET 11, 1582–1591.

Pujara, M., Paudel, N., 2024. Rainfall Prediction using Long Short-Term Memory and Gated Recurrent Unit with Various Meteorological Parameters. Nepal. J. Stat. 8, 47–60.

Rahman, A., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M.A., Mosavi, A., 2022. Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors 22, 3504. https://doi.org/10.3390/s22093504

Saleh, Md.A., Rasel, H.M., Ray, B., 2024. A comprehensive review towards resilient rainfall forecasting models using artificial intelligence techniques. Green Technol. Sustain. 2, 100104. https://doi.org/10.1016/j.grets.2024.100104

Sharma, A., Singh, K., 2024. ARIMA-Based forecasting of monthly rainfall in Mandi district, Himachal Pradesh. Water Supply 24, 3226–3237.

Singh, P., Ramkumar, K.R., Hasija, T., 2024. Monsoon Rainfall Prediction for Punjab Using ARIMA Model: A Time Series Analysis, in: 2024 International Conference on Automation and Computation (AUTOCOM). IEEE, pp. 110–113.

Syczewska, E.M., 2010. Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol. Methods 17, 228.

Yulianto, F., Mahmudy, W.F., Soebroto, A.A., 2020. Comparison of regression, support vector regression (SVR), and SVR-particle swarm optimization (PSO) for rainfall forecasting. J. Inf. Technol. Comput. Sci. 5, 235–247.