BILINGUAL PUBLISHING GROUP

Research in Ecology
https://journals.bilpubgroup.com/index.php/re

ARTICLE

Model Agreement and Disagreement in Rainfall Classification: A Region-Dependent Comparative Analysis of ARIMA and Random Forest across Punjab Districts, India

Vanshita Arora [1], ***Mohammad Shahfaraz Khan*** [2], ***Imran Azad*** [2], ***Amir Ahmad Dar*** [1*], ***Amit Kishore Sinha*** [3], ***Mohammed Wamique Hisam*** [4]

[1] *School of Chemical Engineering, Lovely Professional University, Phagwara 144411, India*

[2] *College of Economics and Business Administration, University of Technology and Applied Sciences-Salalah, Salalah 211, Oman*

[3] *Department of Commerce, Faculty of Business & Commerce, Chandigarh University, Uttar Pradesh Campus, Unnao 209801, India*

[4] *College of Commerce and Business Administration, Dhofar University, Salalah 211, Oman*

ABSTRACT

For agricultural planning and management in monsoon-dependent regions, accurate classification of annual rainfall into risk categories is crucial. The predictive effectiveness of several models for rainfall forecasting has been compared in numerous studies, but the diagnostic significance of model disagreement has not been examined. This study aimed to analyse whether there is a link between the disagreement between autoregressive integrated moving average (ARIMA) and Random Forest (RF) and the difference in the accuracy of the predictions given by the two models in seven districts of Punjab, India (1970–2024). Prior to modelling, a pre-specified, data-driven district selection protocol was used, which was based on the coefficient of variation (CV) of rainfall, avoiding outcome-driven selection bias. The data were discretized into three categories (bins) with training-set bin edges to avoid data leakage. Fisher's exact test, Wilson score confidence intervals, and block permutation testing were used for statistical inference. In the five reliable districts (cells $n \geq 3$), the prediction accuracy was significantly greater when the model disagreed with the data than when it agreed

*CORRESPONDING AUTHOR:
Amir Ahmad Dar, School of Chemical Engineering, Lovely Professional University, Phagwara 144411, India; Email: sagaramir200@gmail.com

ARTICLE INFO
Received: 27 April 2026 | Revised: 21 May 2026 | Accepted: 26 May 2026 | Published Online: 3 August 2026
DOI: https://doi.org/10.30564/re.v8i4.13458

CITATION
Arora, V., Khan, M.S., Azad, I., et al., 2026. Model Agreement and Disagreement in Rainfall Classification: A Region-Dependent Comparative Analysis of ARIMA and Random Forest across Punjab Districts, India. Research in Ecology. 8(4): 200–214. DOI: https://doi.org/10.30564/re.v8i4.13458

(0.83 vs. 0.52; Fisher's exact $p = 0.022$; block permutation $p = 0.038$). However, the Mansa district (dis.: 0.25; agree: 0.33) had a reverse pattern, indicating that the relationship was not pervasive. The findings are empirically consistent with the heuristic intuition of ensemble diversity theory but are not derivable from known theorems. The disagreement between ARIMA and RF can be used as a realistic uncertainty indicator for real-time rainfall monitoring. The findings are exploratory and hypothesis-generating, as there are small test sets per district.

Keywords: Rainfall Classification; Model Disagreement; ARIMA; Random Forest (RF); Ensemble Diversity; Punjab Hydrology; Coefficient of Variation; Southwest Monsoon (SWM)

1. Introduction

Punjab, the breadbasket of India, is one of the most agriculturally productive states in India. The spatial and temporal distribution of monsoon rainfall is necessary for the economy and food security of India. Most of the annual precipitation in the region is due to the southwest monsoon (SWM), which is normally experienced between June and September. In Punjab, the monsoon season of 77 days extending from July, August and September, receives rainfall at an average rate of 6 mm/day [1]. There is a great deal of spatial and temporal variability in SWM rainfall throughout the Indian subcontinent. Kumar et al. affirmed a considerable spatio-temporal fluctuation of southwest monsoon rainfall by district in Punjab over 1981–2020 and stressed the prerequisite of robust district-level rainfall forecasting models [2]. A complex system can be termed a climate, which influences the atmosphere, hydrosphere, lithosphere, and cryosphere [3]. Moreover, climate change and unpredictable monsoons may also be significant water stressors in Punjab [4]. Both droughts and extreme wet years are deviations from expected rainfall that have direct effects on crop yields, groundwater recharge, and irrigation demand. Extreme events have been continuously increasing because of human-induced activities that have led to climate change [5]. The variability is further enhanced at the regional level, whereby local climatic processes and land-atmosphere interactions result in large inter-district variability in rainfall distribution. In data-constrained environments where rainfall data are available only as annual totals, prediction of rainfall is not the only task; it is also necessary to understand the variability of rainfall to use that variability directly to decide. The nonlinearity, randomness, and regional heterogeneity of rainfall patterns, which are becoming increasingly extreme due to climate change, make accurate forecasting difficult [6]. Here, consistent categorisation of rainfall into risk levels (low, medium, or high) is a decision-relevant activity in agricultural planning, flood management, and drought preparedness. The division of rainfall into discrete risk levels gives an oversimplified but useful depiction of climatic conditions and is coherent with the output of models and their usage in agriculture and water management.

Rainfall forecasting in South Asian settings has been undertaken with a broad array of techniques, including classical time-series techniques to current machine learning models. ML is an Artificial Intelligence (AI) branch that induces regularities and patterns to offer easier implementation at a low computation cost, quicker training, validation, testing, and evaluation with higher performance than physical models, and relatively less complexity [7]. The measured climate indices and hydro-meteorological parameters are assimilated in data-driven approaches to offer more insight [8]. ARIMA has long been a benchmark for hydrological forecasting, which provides a parameterisation that is interpretable and which has theoretical underpinnings in time-series analysis. In modern contexts, ensemble methods like Random Forest (RF) have been used to solve hydrology problems by taking advantage of non-linear relationships and feature interactions unavailable to ARIMA. As Cheng et al. suggest, the traditional forecasting methods, including ARIMA methods, cannot reflect the development of such systems in terms of forecasting accuracy, computation load, and their sensitivity to the amount and quality of the a priori input data [9]. Tien Bui et al. also compared the performances of ANN, SVM, and RF to general application to floods, where RF performed the best [10]. This increasing methodological diversity is the subject of most comparative studies, which concern which model is more accurate. When model dis-

agreement is viewed as a nuisance, and not potentially informative. Variability between model predictions may be utilized to anticipate uncertainty, particularly in circumstances with little data, according to recent machine learning attempts [11]. Furthermore, two modeling approaches typically reflect opposing aspects of the complex system; this finding suggests that conflict may be the result of significant data rather than simple error [12].

By investigating the diagnostic potential of the discrepancy between ARIMA and RF in the context of rainfall category prediction, this study closes a gap in the literature. Two structurally dissimilar models may agree because of the true signal or because of a shared bias brought on by the sparse data. When there is disagreement, at least one model identifies a pattern that the other does not, and this variety can increase the possibility that one of the two models will predict the right category. The machine learning theory's ensemble diversity principle is heuristically related to this technique. Crucially, the current ensemble diversity theorems [13,14] are based on an ensemble's total error rather than per-case accuracy, which is dependent on whether two members agree with one another. In the current work, disagreement is seen as an empirical indicator rather than a formal theoretical deduction, and this intuition is interpreted in a disaggregated manner.

The study makes three contributions: First, a pre-specified, data-driven strategy for district selection is used to suggest an objective regime categorization (structured, unstable, sparse) without model outputs (when using the coefficient of variation (CV) of rainfall). Second, a disagreement analysis methodology that assesses the correlation between prediction accuracy and inter-model disagreement was applied in seven districts of Punjab. Third, the statistical inference based on Fisher's exact test is pooled across all the districts, and this is more powerful than the test of the core hypothesis that would be supported by each of the districts separately.

2. Literature Review

2.1. ARIMA in Hydrological Forecasting

The most distinguished time series model, or more precisely, ARIMA [15], proposed by Box et al. has been extensively used for hydrometeorological time series [16]. Moreover, during the last several decades, ARIMA models have become instrumental tools for various meteorological tasks to understand the effects of precipitation and air temperature [17]. They are natural candidates with annual or seasonal rainfall data because of their capability to model autocorrelation structures in both stationary and differenced sequences. Several studies have used ARIMA to analyse Indian rainfall. Mishra and Desai were able to show how ARIMA can be useful in streamflow forecasting in Indian river basins [18], whereas Nirmala and Sundaram used ARIMA to predict seasonal rainfall in Tamil Nadu [19]. Sharma and Singh applied ARIMA to monthly rainfall data from eight stations in the district of Mandi, Himachal Pradesh [20]. Dahiya et al. utilized a seasonal ARIMA model on monthly rainfall in the Punjab and Haryana states from 1951 to 2020 and found that the ARIMA (1,1,2), (0,1,1) model fitted best to precipitation in Punjab, and forecasted precipitation patterns until 2030, showing the ongoing efficacy of ARIMA-based methods, such as SARIMA, have also been applied to study long-term rainfall trends in Punjab and Haryana, but point forecasting using classification is less widespread [21].

One drawback of ARIMA when it comes to the classification of rainfall is that it is more of a forecasting model: it generates continuous predictions, and these need to be broken into categories by a post-processing step. This brings a binning decision that influences the accuracy of classification independent of model skill. Because ARIMA generates continuous forecasts, converting their results into discrete formats creates another source of uncertainty that is not due to the performance of the underlying model [22]. Moreover, classic time-series models can have trouble with non-linear and complex patterns and tend to be poorer in highly variable climatic regimes [23]. The present study can help overcome this by applying identical, training-set-based bin edges to ARIMA and RF results, making a reasonable comparison.

2.2. RF for Rainfall Classification

The RF proposed by Breiman is a collection of decision trees trained on bootstrapped subsets of data with a random selection of features at each split [24]. The decorrelation of the trees was achieved by randomly selecting features at each split and minimizing overfitting [25]. Each tree generates

a predictor set of response values that are related to a set of independent values [26]. In addition, a collection of such trees chooses the most appropriate set of classes, which has been observed to be highly effective in hydrology applications because it is resilient to noisy data and nonlinear feature relationships. Shortridge et al. compared RF to ARIMA in streamflow prediction and discovered that RF was competitive, especially in non-stationary catchments [27]. Wani et al. compared RF with SVR, ANN, and KNN to predict rainfall across an altitudinal gradient in the North-Western Himalayas in data-limited scenarios [25]. RF can be used to complement the linear time structure of ARIMA in the context of rainfall classification because it can represent lagged relationships and nonlinear thresholds. Associating ARIMA with RF, Support Vector Machine, Artificial Neural Network, and XGBoost in seasonal rainfall forecasting over India, Mishra et al. discovered that "ARIMA models are likely to overfit the training information"; however, machine learning models better explained non-linear rainfall trends across regions, which justifies the dual-model design used here [28]. In addition, Alam and Majumder conducted a comparative study on ARIMA and statistical approaches to rainfall prediction in Kolkata and found that none of the models was always dominant under all conditions [29].

In the current analysis, RF was set to have one lag (Lag 1, the rainfall of the last year) because of the limited richness of the data and the fact that parity with the univariate ARIMA model was required. RF works with lag characteristics to model temporal relationships because it is inherently unable to remember past observations [28]. Although this is a bare-bones feature set, it is well suited for an exploratory comparison when the model structure and not predictive performance maximization is of interest.

2.3. Ensemble Diversity and Disagreement Theory

The literature on ensemble learning provides the theoretical core of this study's key hypothesis. Krogh and Vedelsby demonstrated that ensemble error could be broken down into average individual error, excluding ensemble diversity, which means that diverse models decrease aggregate error even when individual models are not perfect [14]. Dietterich generalizes this model further to validate that model disagreement, as caused by structural differences and not noise, acts as a diversification of error, and at least one diverse model will capture patterns that the other models will not [13]. Conflict among models has been employed as a criterion for detecting informative samples in active learning [30]. Wood et al. pioneered a cohesive theoretical framework of diversity in ensemble learning, formalising the circumstances in which model diversity can minimise aggregate prediction error to provide a contemporary theoretical foundation for the hypothesis that structurally different models (ARIMA and RF) will exhibit a complementary pattern of error [31]. In recent years, the connection between ensemble disagreement and predictive uncertainty has been revived, and ensemble disagreement has been described as indicative of epistemic uncertainty, with areas of high inter-model disagreement being interpreted as underrepresented and ambiguous inputs in the training distribution [32].

These theoretical verdicts are generally shown to be true when combined with ensembles (e.g., majority vote, averaging) rather than on the selective analysis of the agreement vs. disagreement cases. In the current study, this lens is used in a slightly unique way: rather than summing the results of the models, it questions whether, in situations where the two models are inconsistent, at least one of them is more likely to be accurate than in situations where both models agree. It is a disaggregated form of the diversity argument at the level of single test observations.

2.4. Rainfall Studies in Punjab, India

There has been extensive research on rainfall variability in Punjab. The non-stationarity of rainfall time series data is one of the leading problems. The traditional stationary assumptions of statistical models lose their significance when there is a sudden shift in the mean and variance of rainfall distributions caused by climatic variability and change [17]. Thus, examining rainfall trends across various spatial contexts is essential for comprehending the fluctuations in rainfall patterns within the disciplines of meteorology, hydrology, and climatology globally [33–35]. Kumar et al. analysed the fluctuations and trends of rainfall in Punjab from 1981 to 2020 using a modified Mann-Kendall test, revealing a significant increase in annual rainfall across the state [2]. Recent studies conducted

by Madane and Waghaye utilized the Mann-Kendall and innovative trend analysis across 20 districts in Punjab from 1951 to 2021, revealing a decline in annual rainfall trends in various southwestern districts, including Bathinda and Mansa, while the northwestern districts showed comparatively stable or rising trends [36].

These works all encourage the regime-based analytical framework that will be taken here: considering all Punjab districts as a homogeneous population would mask significant behavioural variations that impact model performance and disagreement.

3. Study Area and Data

3.1. Study Area

Punjab, located in the northwestern part of India (29.5° N–32.5° N, 73.9° E–76.9° E), encompasses an area of approximately 50,362 km^2 and is divided into 23 administrative districts. There are roughly 2.77 million people living there (Statistical Abstract of Punjab, 2020). With a distinct monsoon season that lasts from June to September and provides 70–80% of the yearly rainfall, the climate is semi-arid to sub-humid. Significant gradients linked to the Himalayan range and the Bay of Bengal monsoon are implied by the average rainfall amounts, which are noticeably high and exhibit a wide range (ranging from 180 mm in the state's southwestern region, like Mansa, to more than 950 mm in the northeastern foothills, like Gurdaspur). The northwest of Punjab receives abundant rainfall of up to 900 mm, whereas the southwestern part of the state gets much less at 360 mm [37].

Seven districts were incorporated in this study: Amritsar, Ludhiana, Jalandhar, Gurdaspur, Hoshiarpur, Mansa, and Bathinda. These districts span the full north-south extent of the state and represent the range of rainfall regimes present in Punjab, as shown in **Figure 1**.

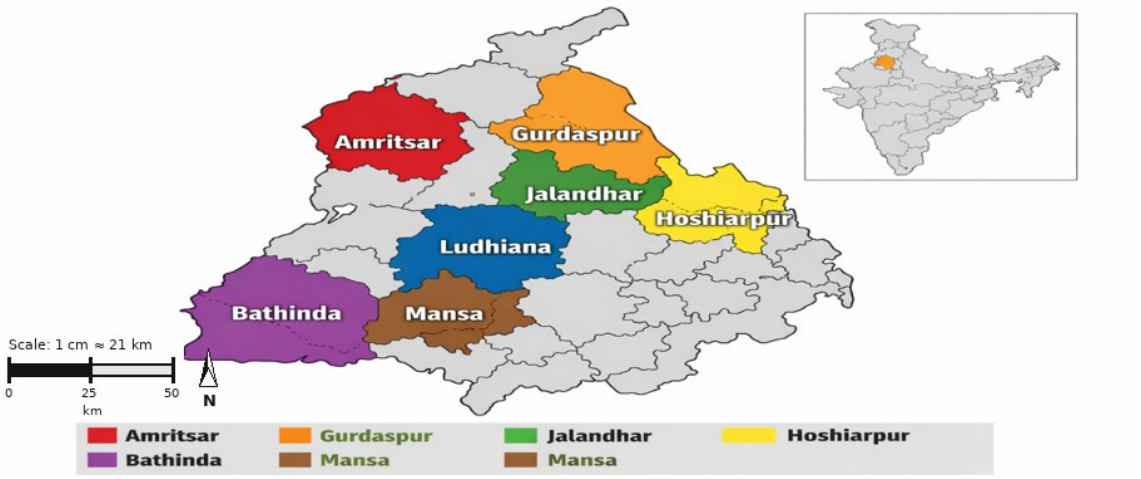

Figure 1. Map of Punjab State, India.

3.2. Dataset

Annual rainfall data for seven Punjab districts were obtained for the period 1970–2024 from the Open Government Data (OGD) platform released under the National Data Sharing and Accessibility Policy (NDSAP), yielding a primary dataset of approximately 55 years per district. In order to use the India Meteorological Department (IMD) operational category thresholds in the sensitivity analysis and to confirm district-level information, the Indian Meteorological Department's Rainfall Information of India [38] was utilized as a secondary source. There are acknowledged gaps in the data; for example, several districts are missing from the years 1971–1973 and 2012 since the data was not available at that time. These gaps were analyzed rather than imputed. There are no administrative rainfall records for Mansa previous to 1992 because the district was formally created on April 13, 1992, from the pre-existing Bathinda district. The Bathinda and Mansa districts are located in Punjab's south-western agro-climatic zone, which is known for brackish groundwater, an average annual rainfall of about 400 mm, and an agro-landscape

that is primarily canal-irrigated [39]. This is accurately captured in the dataset, since there are missing data between 1970 and 1991 and only post-1992 Mansa data (around 32 years) were utilized in all analyses. No proxy imputation was done during the period before 1992.

To make the models predictive, each district's isolated missing years from the training period were linearly interpolated, and test-set years with missing values were eliminated. Depending on the distribution of gaps, each district can use 32–53 annual observations from the final dataset.

Figure 2 presents the annual rainfall time series for all seven districts from 1970 to 2024, clearly illustrating the inter-district variation in mean rainfall levels and inter-annual variability that motivates the regime-based analytical approach adopted here.

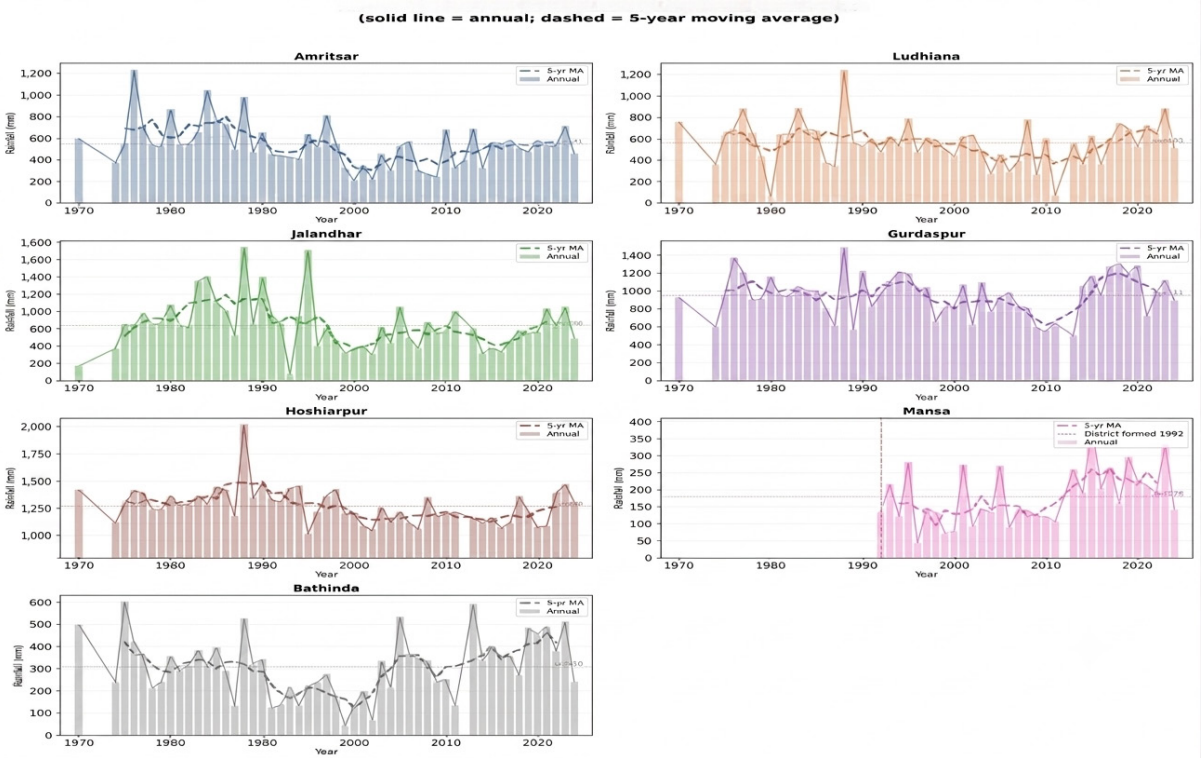

Figure 2. Annual rainfall (mm) for seven Punjab districts, 1970–2024, with a 5-year moving average (dashed).
Note: Horizontal grey line = district mean. Vertical dotted line in the Mansa panel = 1992 (district formation date).

4. Methodology

4.1. Pre-Specified District Selection

To prevent the outcome-driven district selection that can result in artificially favourable outcomes, the three districts to be focused on to analyse the regime were selected by a pre-specified, data-driven protocol applied prior to any modelling. The coefficient of variation (CV = σ/μ) of the annual rainfall of all seven districts was calculated based on the complete historical record. CV ranked the districts in terciles. There were three types of regimes:

1. Structured regime: The district having the lowest CV in the bottom tercile (Gurdaspur, CV = 0.246), which is the most predictable pattern of rainfall.
2. Unstable regime: The district with the highest CV in the top tercile (Mansa, CV = 0.521), which reflects the most unstable inter-annual variability.
3. Sparse regime: The district with the lowest average yearly precipitation irrespective of CV (Bathinda, mean = 316.9 mm), which is a low-rainfall, and possibly a data-sparse region.

This choice was fixed before any model fitting or examination of results. The other four districts (Amritsar, Ludhiana, Jalandhar, Hoshiarpur) were retained in the

dataset and incorporated in all pooled analyses, which added some extra statistical power to the testing of the core hypothesis. **Table 1** shows the CV-based ranking of all seven Punjab districts.

Table 1. CV-based ranking of all seven Punjab districts.

District	Regime	Mean (mm)	Std Dev	CV	N (Years)	Selected
Gurdaspur	Structured (low CV)	959.6	236.2	0.246	52	Yes
Hoshiarpur	**—**	**775.4**	**276.3**	**0.356**	**52**	**No**
Ludhiana	—	569.6	211.3	0.371	52	No
Amritsar	**—**	**547.4**	**206.7**	**0.378**	**53**	**No**
Bathinda	Sparse (lowest mean)	316.9	142.3	0.449	52	Yes
Jalandhar	**—**	**642.1**	**298.0**	**0.464**	**52**	**No**
Mansa	Unstable (high CV)	187.5	97.6	0.521	32*	Yes

Note: Mansa data span 1992–2024 due to administrative formation date. Bolded rows indicate the three districts selected for regime-based analysis. CV = coefficient of variation (σ/μ). *: Due to data availability, only 32 years of observations were available for Mansa.

4.2. ARIMA Model

In each district, the time series was annualized and resampled to an annual start frequency (asfreq('YS')), and any remaining missing values in the training period were handled by linear interpolation. All the observations except the last ten years constituted the training set and the remaining formed the test set. The training series was tested via an Augmented Dickey-Fuller (ADF) test to identify the correct difference order: d = 0 when the series was at the 5% level stationary (ADF $p < 0.05$), and d = 1 when the series was not. Maximum likelihood estimation was then used to fit an ARIMA (1,d,1) model. Recursive forecasting was applied to the model to produce out-of-sample forecasts for the ten test years.

Continuous forecast values were then discretised into Low, Medium or High categories based on bin edges calculated only using the training-set rainfall distribution (binning with equal widths over the training range, 1% boundary padding to accommodate edge cases). This was done to make sure that test-set rainfall values were not used during the bin boundaries determination to avoid data leakage.

The results of the ADF test statistic, *p*-value, selected differencing order, and Ljung-Box test results for residual diagnostics are presented in **Table A1** (**Appendix A**) for each district. Five districts were stationary at the 5% level (d = 0): Amritsar (ADF $p = 0.015$), Jalandhar ($p < 0.001$), Gurdaspur ($p = 0.028$), Mansa ($p < 0.001$), and Bathinda ($p < 0.001$). Ludhiana ($p = 0.964$) and Hoshiarpur ($p = 0.189$) required first differencing (d = 1). Six of the seven districts had adequate fit, according to Ljung-Box residual diagnostics ($p > 0.05$ at both lag 5 and lag 10). The residual autocorrelation of the Ludhiana district at lag 10 ($p = 0.011$) warrants attention, since the ARIMA (1,1,1) model might not be the best fit for the data from this district. The Akaike Information Criterion (AIC)-based robustness checks (**Table A2**), which allow district-specific order selection, provide a complementary perspective on this limitation, which is further discussed.

4.3. RF Classifier

An RF classifier with 100 trees and a fixed random seed (42) was trained for each district. The input feature was only Lag 1, which was the amount of rainfall in the previous year. This sparse set of features was chosen to preserve structural equivalence with the univariate ARIMA configuration and to point towards the fact that annual-resolution data has only narrow information. The classification criterion was the equivalent three-category (Low, Medium, High) of the annual rainfall with the training-set bin edges applied to ARIMA classification. The training and test sets were constant in terms of time length as the ARIMA.

To perform a sensitivity check, Lag 2 (rainfall two years prior) was added as the second RF feature. This finding is not an artifact of the single-feature design of Lag 1, since the directional pattern (disagree accuracy ≥ agree accuracy) was preserved for the same set of reliable districts.

4.4. Disagreement Analysis

In each test-year observation, both models made a categorical prediction (Low, Medium, or High). A disagreement indicator was described as:

Disagreement = 1 if the ARIMA prediction ≠ RF prediction, 0 otherwise.

The empirical indicator was whether either or both two models predicted the actual amount of rainfall correctly. This is the true definition, which reflects the diagnostic power of model diversity: as the disagreement increases, the likelihood that one model is correct is greater, and it becomes an error-diversification mechanism. The accuracy was calculated on the subset of test years where the models agreed and, on the subset, where the models disagreed.

4.5. Statistical Inference

Fisher's exact test (one-sided, H_1: agreement accuracy < disagreement accuracy) was used on the cross-district data and within the individual districts. The pooled test added all agree case and disagree case observations in all seven districts to a 2 × 2 contingency table that facilitated stronger inference as distinguished from each district analysis. The primary confirmatory analysis consisted of a sensitivity analysis that was restricted to four districts with sufficient cell sizes (n ≥ 3 or more in agreement and disagreement states).

Wilson (1927) score 95% confidence intervals were computed for all reported accuracy proportions, using the formula:

$$CI = (\hat{p} + z^2 / 2n \pm z\sqrt{[p(1-p)/n + z^2/4n]}) / (1 + z^2/n), \text{where } z = 1.96.$$

Wilson CIs are preferred over normal approximation CIs for small cell sizes, as they remain valid when $\hat{p} = 0$ *or* $\hat{p} = 1$.

To address the independence assumption of Fisher's exact test, a block permutation test (N = 10,000) was performed, permuting agreement labels independently within each district block. This retains a within-district temporal autocorrelation and detaches a cross-district association for the test. The permutation *p*-value is a more conservative estimate of significance that does not assume independence between observations.

4.6. Sensitivity to the Binning Scheme

An equal bin width over the range of the training-set rainfall was used for the main analysis to identify the Low, Medium and High categories. The analysis was replicated using a simplified three-class approximation of IMD operational thresholds [38]: Deficient (<80% Long Period Average (LPA)), Normal (80–120% LPA), and Above Normal (>120% LPA). Note that the official IMD scheme uses five categories with narrower bands (deficient: <90% LPA; near-normal: 96–104% LPA; excess: >110% LPA); the three-class simplification was adopted to maintain parity with the three-category equal-width binning scheme of the primary analysis. The no-leakage principle was maintained by calculating the LPA for each district using only the training series, without leakage. **Table A3** (**Appendix A**) reports per-district and pooled results under this alternative scheme, where the pooled result among reliable districts was significant ($p = 0.009$).

5. Results

5.1. Model Performance

Table 2 summarizes the ARIMA and RF accuracies, balanced accuracies, disagreement rates, and naive baseline accuracies for all seven districts. Whereas the RF models' accuracy varied from 0.20 (Mansa) to 0.70 (Hoshiarpur), the ARIMA models' accuracy ranged from 0.30 (Mansa) to 0.80 (Ludhiana). It's interesting to note that for all ten test years, Hoshiarpur had no conflict between the two models; that is, both ARIMA and RF predicted the same category in every year. Consequently, Hoshiarpur cannot be utilized in the disagree-state analysis despite its relative accuracy.

While RF performed at or above the majority-class baseline in four of the seven districts, ARIMA did so in all seven. Five of the seven districts in both models showed upward deviations from the persistent baseline, suggesting that the temporal structure is exploited for purposes other than simple prediction. The persistence baseline was between 0.20 and 0.60, and the majority-class baseline accuracy was between 0.30 (Mansa) and 0.80 (Ludhiana). The disagreement rate showed significant variations between 0.0 (Hoshiarpur) and 0.6 (Jalandhar and Gurdaspur), and

the highest disagreement rate appeared to be associated with districts having complex or variable rainfall.

Table 2. Per-district model accuracy, balanced accuracy (Bal.), disagree rate, and naive baselines.

District	N Test	ARIMA Acc.	RF Acc.	ARIMA Bal.	RF Bal.	Disagree Rate	Majority Class (Acc.)	Persistence Acc.
Amritsar	10	0.50	0.50	0.50	0.50	0.4	Low (0.50)	0.40
Ludhiana	10	0.80	0.50	0.33	0.21	0.3	Medium (0.80)	0.50
Jalandhar	10	0.50	0.50	0.50	0.50	0.6	Medium (0.50)	0.60
Gurdaspur	10	0.40	0.50	0.33	0.63	0.6	Medium (0.40)	0.50
Hoshiarpur	10	0.70	0.70	0.50	0.50	0.0	Low (0.70)	0.60
Mansa	10	0.30	0.20	N/A	N/A	0.4	Medium (0.30)	0.20
Bathinda	10	0.60	0.50	0.50	0.42	0.2	Medium (0.60)	0.60

Note: N/A = balanced accuracy is not computable due to single-class predictions. Majority class column shows predicted class with accuracy in parentheses.

5.2. Accuracy Conditioned on Agreement and Disagreement

The core findings are presented in **Table 3**: prediction accuracy given a model agreement/disagreement, along with the 95% CI of Wilson [40] and Fisher's exact test. The Amritsar, Ludhiana, Jalandhar, Gurdaspur, and Mansa districts had a sufficient number of cells ($n \geq 3$) in both states. For these five, four had an accuracy comparable to or higher than the agreed accuracy. Mansa was the exception: the accuracy was 0.33 when agreeing and 0.25 when disagreeing, which is in opposition to the main directional claim. This is not a fluke but a real discovery to be treated as such.

The two models showed no disagreement cases for the district of Hoshiarpur (n_disagree = 0) for all test years; therefore, a disagree-state accuracy could not be calculated. There were only two cases with disagreements in Bathinda (n_disagree = 2), making it difficult to draw reliable inferences. The two districts were not included in the primary analysis; they were included for completeness only.

Table 3. Accuracy conditioned on model agreement vs. disagreement, Wilson 95% CIs, and Fisher's exact test (one-sided, H_1: agree acc. < disagree acc.).

District	Reliable	n Agree	Acc. (Agree)	95% CI	n Disagree	Acc. (Disagree)	95% CI	Fisher's *p*
Amritsar	Yes	6	0.50	[0.19, 0.81]	4	1.00	[0.51, 1.00]	0.167
Ludhiana	Yes	7	0.71	[0.36, 0.92]	3	1.00	[0.44, 1.00]	0.467
Jalandhar	Yes	4	0.50	[0.15, 0.85]	6	1.00	[0.61, 1.00]	0.133
Gurdaspur	Yes	4	0.50	[0.15, 0.85]	6	0.83	[0.43, 0.97]	0.333
Hoshiarpur	No	10	0.70	[0.40, 0.89]	0	N/A	—	N/A
Mansa	Yes*	6	0.33	[0.10, 0.70]	4	0.25	[0.05, 0.70]	0.833
Bathinda	No	8	0.62	[0.30, 0.86]	2	0.50	[0.09, 0.91]	0.867
Pooled (5 reliable districts)	**—**	**27**	**0.52**	**—**	**23**	**0.83**	**—**	**0.022**
Pooled (all 7 districts)	—	*45*	*0.58*	—	*25*	*0.80*	—	*0.051*

Note: Mansa meets cell-size criteria but contradicts the directional claim. Cell size n < 3 or n = 0; results are not interpretable. Bolded row = primary pooled result (5 reliable districts). Italics row = all-7-district robustness check. *: Meets the cell-size criterion for reliability but contradicts the directional hypothesis.

5.3. Pooled Analysis

The 2 × 2 table for the five reliable districts had 27 observations that agreed with the state and 23 disagreement observations (accuracies of 0.52 and 0.83, respectively). The *p*-value of Fisher's exact test was 0.022, which was statistically significant at an α value of 0.05. In the all-district pooled analysis (n = 70), the *p*-value was 0.051, which is just above the conventional value, due to the zero-disagreement data from Hoshiarpur and Mansa's contradictory results.

The Fisher combined *p*-value method for the seven individual *p*-values yielded a non-significant result (χ^2 =

11.97, df = 14, $p = 0.609$). The correct interpretation is that no single district test provides significant evidence by itself; therefore, the statistically significant pooled result is completely dependent on the aggregation across districts. It is important to note that this is an effect of the combined dataset and should not be interpreted as a per-district effect that can be replicated.

To have a valid check on the main fixed-specified analysis (ARIMA (1,d,1)), the orders of ARIMA were re-specified by AIC grid search (p, q ∈ {0–3}) for each district, which consequently gave each district a different order instead of fixing it at (1,d,1). The orders chosen by the AIC are listed in **Table A2**. The pooled disagree accuracy was also higher than the agree accuracy for reliable districts (Fisher's exact $p = 0.095$; agree = 12/23 = 0.522 vs. disagree = 20/27 = 0.741), but not as significant as conventional ($p = 0.05$). Both types of specifications are reported transparently; the direction finding remains valid under flexible ordering of specifications, but it is less statistically safe.

To test the independence assumption used in Fisher's exact test, a block permutation test (N = 10,000) was performed. Within each district block, agreement labels were permuted independently, such that there was a temporal structure within each block but no cross-block association between agreement status and accuracy. The difference between the observed statistic (agree accuracy − disagree accuracy = −0.222) was observed at the 3.8th percentile of the permutation distribution (i.e., $p = 0.038$) and was found to be statistically significant at the 5% significance level, given a test that does not assume independence across observations or districts. **Figure 3** shows the permutation distribution.

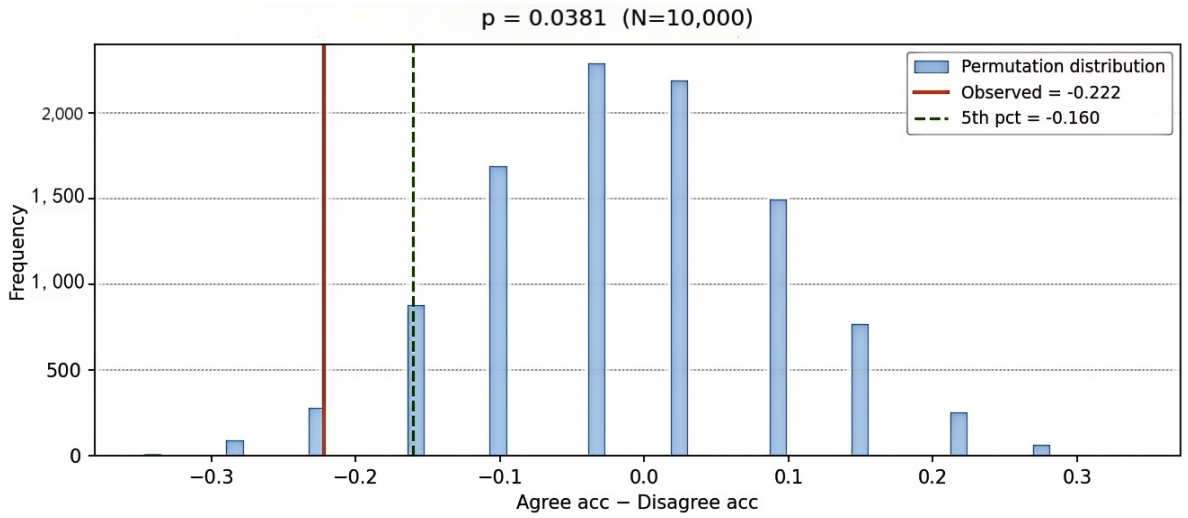

Figure 3. Block permutation test distribution (Data: 1970–2024, N = 10,000).

6. Discussion

6.1. Interpretation of the Core Finding

The significant differences in prediction accuracy obtained from the table of model agreement and the ability of the models to better predict the observed data (Fisher's exact test, $p = 0.022$; block permutation, $p = 0.038$) support the heuristic intuition behind ensemble diversity theory in the five reliable districts. Previous ensemble diversity theorems [13,14,31] focused on the ensemble's prediction error rather than the per-case accuracy given the agreement or disagreement of the two members. Thus, the present finding is an empirical extension of this intuition and not something that follows from the theory.

The two models (ARIMA and RF) use structurally different parts of the data; ARIMA uses temporal autocorrelation, and RF uses the non-linear relationship between the rainfall in the previous year and the category of the current year. In three out of four supporting districts (Amritsar, Ludhiana, and Jalandhar), the disagree accuracy was 1.00, indicating that one of the two models was always correct when the two models disagreed. The absolute improvement with model disagreement ranged from +0.33 (Gurdaspur: 0.50 → 0.83) to +0.50 (Amritsar, Jalandhar: 0.50 → 1.00) across

these four districts, with a pooled absolute improvement of +0.31 (0.52 → 0.83). In practical applications, years of conflicting models are the years when both models are most useful (not in terms of the greatest uncertainty).

The practical importance of this finding for agricultural early warning systems in Punjab is that when ARIMA–RF disagrees, it could be used as a real-time warning to produce an additional warning to prompt further investigation or consultation with other stakeholders without using another model or extra data. To realize such a system in practice, there is a natural classification framework, which is the operational category of the India Meteorological Department [38].

6.2. The Mansa Exception

The directionality of the claim is contradicted by a high cell size in only one district, Mansa (agree cell size 0.33, as opposed to disagree cell size 0.25). There are several reasons why this could occur. First, Mansa has only 32 years of data (the district was formed in 1992), the shortest record of any district studied. Second, for Mansa, the mean rainfall is extremely low (187.5 mm), and the coefficient of variance is relatively high (CV = 0.521), which might lead to a distribution of rainfall with such sharp classification boundaries that the choice of bin edges is even more critical. Third, the six agree state and four disagree state cases are still too small a sample to allow for conclusions of any kind; the reversal observed may be the result of random variation, not structural. The frank answer is that the main finding, that there is a disagree-accuracy advantage, is not found across the Punjab rainfall regime, and Mansa's result is a testament to that.

6.3. Limitations

There are several shortcomings that must be taken into consideration. To begin with, the test sets per district are small (n = 10 years each), which constrains the power of individual district tests and doesn't allow high power to be drawn at the district level. The pooled analysis addresses this to some degree but assumes that the relationship between agreement and accuracy is widely constant across districts—an assumption that might not be true in all hydrological situations.

Second, the RF model operates on only one feature (Lag 1), which can be a low estimate of its potential performance. Meteorological covariates (including temperature, humidity or large-scale climate variability indices, e.g., (El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD)) might be of significant value to RF accuracy and shift the pattern of disagreement. The existing configuration does not focus on RF performance optimisation but structural comparability to ARIMA.

Third, the best-fit AIC for each district was calculated, but each ARIMA model order from (1,0,0) to (3,1,3) across districts was pooled into a single Fisher's exact test to determine the overall significance, which gives $p = 0.095$, a result that is not significant at the conventional level ($\alpha = 0.05$). This finding is marginal even in the flexible, data-driven model selection, and with the small number of reliable districts (n = 4), it must be interpreted with caution.

Fourth, residual autocorrelation was observed for AR at lag 10 (Ljung-Box $p = 0.011$) in Ludhiana, which makes ARIMA (1,1,1) a suboptimal model for this district. Therefore, caution should be exercised when interpreting the results for Ludhiana.

Fifth, it is limited to seven districts in one state. Replication with independent data is needed to generalize to other Indian states or semi-arid regions internationally.

Sixth, the independence of test observations across the 70 observations in the pooled Fisher's exact test is not fully satisfied by the temporally autocorrelated series from several districts in one state. The block permutation test ($p = 0.038$) provides a more conservative estimation of the structure within the districts. The nominal Fisher p-value should be considered in conjunction with the block permutation p-value.

Seventh, the overall-7-district Fisher p-value is 0.051, which is not significant at $\alpha = 0.05$, and is only considered marginal. The main confirmatory result was based on a reliable five-district subset ($p = 0.022$).

Eighth, the accuracy of the headline numbers are not independent of the binning scheme. The absolute accuracy values differed for equal-width binning and IMD operational categories, with directionally consistent results (**Table A3**). The disagree-accuracy advantage should be viewed in the context of the binning scheme used.

7. Conclusions

This paper examined the existence of inter-model uncertainty between ARIMA and RF in the classification of annual rainfall concerning increasing prediction accuracy across seven districts of Punjab. The main conclusions are as follows:

i. A pre-specified CV-based district selection protocol was able to recognise three different hydrological regimes (structured, unstable, sparse) based purely on raw data properties, without any model fitting. It is a strategy that can be directly applied to other multi-district studies and does not suffer from the outcome-driven selection bias that invalidates comparative studies.
ii. In four of the five districts with satisfactory cell sizes, there was a positive association between model disagreement and prediction accuracy (agree, 0.52; disagree, 0.83; Fisher's exact $p = 0.022$, block permutation $p = 0.038$). In contrast, the shortest data record and the most extreme low rainfall regime of Mansa showed an opposing pattern that was a genuine exception that qualified rather than invalidated the main finding.
iii. Fisher's combined *p*-value analysis was used to test the hypothesis that the results from individual districts do not provide significant evidence of the hypothesis, but rather that the statistically significant overall pooled result is dependent upon pooling across districts, and as such, the finding is limited in generalizability per district. This encourages further research with larger per-district test sets to determine whether the effect can be measured at the district level alone.

These are hypothesis-generating and exploratory findings. The framework should be expanded to cover all 23 Punjab districts; meteorological variables such as temperature, humidity, and ENSO indices could be added to the RF model; and the effect of disagreement-based flagging on the outcomes of an operational agricultural early warning system could be tested. The IMD operational binning scheme was used as a sensitivity study in the present study (Section 4.6, **Table A3**); further research could use it as the main classification criterion and test whether the thresholds based on the LPA provide more operationally sound thresholds than equal-width binning.

Author Contributions

Conceptualization, A.A.D.; methodology, M.S.K. and I.A.; validation, A.K.S.; formal analysis, V.A., M.S.K. and A.K.S.; investigation, V.A.; data curation, V.A.; writing—original draft preparation, V.A., A.A.D. and M.W.H.; writing—review and editing, M.S.K., I.A., A.A.D., A.K.S. and M.W.H.; supervision, I.A.; project administration, A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the findings of this study are publicly available and can be accessed from the Open Government Data (OGD) Platform India (https://www.data.gov.in/resource/district-wise-annual-average-rainfall-punjab-1970-2021) [41].

Conflicts of Interest

The authors declare no conflict of interest.

AI Use Statement

The authors used AI-assisted tools to refine and edit only the language and structure of this manuscript. The authors subsequently reviewed and edited the content as

necessary and take full responsibility for the final content of the published article.

Appendix A

Table A1. ARIMA Model Diagnostics per District.

District	ADF Stat	ADF *p*	Stationary	d	ARIMA Order	LB *p* (Lag 5)	LB *p* (Lag 10)
Amritsar	−3.298	0.015	Yes	0	(1,0,1)	0.785	0.551
Ludhiana	0.065	0.964	No	1	(1,1,1)	0.172	0.011
Jalandhar	−5.115	<0.001	Yes	0	(1,0,1)	0.995	0.935
Gurdaspur	−3.075	0.028	Yes	0	(1,0,1)	0.801	0.970
Hoshiarpur	−2.250	0.189	No	1	(1,1,1)	0.828	0.945
Mansa	−6.044	<0.001	Yes	0	(1,0,1)	0.410	0.646
Bathinda	−4.908	<0.001	Yes	0	(1,0,1)	0.606	0.593

Note: ADF test results, selected differencing order, fitted ARIMA specification, and Ljung-Box residual diagnostic *p*-values at lag 5 and lag 10. LB $p > 0.05$ indicates adequate residual fit.

Table A2. Robustness Check: AIC-Selected ARIMA Orders.

District	AIC Order	n Agree	Acc. Agree	n Disagree	Acc. Disagree	Fisher's *p*	Direction
Amritsar	(1,0,0)	6	0.50	4	1.00	0.167	Supports
Ludhiana	(3,1,3)	3	0.67	7	0.71	—	Supports
Jalandhar	(2,0,2)	5	0.60	5	1.00	—	Supports
Gurdaspur	(2,0,0)	3	0.67	7	0.71	—	Supports
Hoshiarpur	(0,1,1)	10	0.70	0	N/A	—	—
Mansa	(1,0,1)	6	0.33	4	0.25	—	Contradicts
Bathinda	(2,0,3)	8	0.62	2	0.50	—	—
Pooled (Reliable, n = 4)	—	23	0.52	27	0.74	0.095	Supports*

Note: Per-district results replicating the core disagree-versus-agree accuracy analysis using AIC-optimised ARIMA order selection. *: Pooled result ($p = 0.095$) does not reach significance at $\alpha = 0.05$, indicating sensitivity to model specification.

Table A3. Sensitivity Analysis: IMD Binning Scheme.

District	LPA (mm)	n Agree	Acc. Agree	n Disagree	Acc. Disagree	Fisher's *p*	Direction
Amritsar	536.5	2	0.50	8	1.00	N/A	Supports
Ludhiana	542.4	5	0.40	5	0.80	N/A	Supports
Jalandhar	628.8	3	0.67	7	0.86	N/A	Supports
Gurdaspur	901.5	7	0.29	3	1.00	N/A	Supports
Hoshiarpur	777.1	3	0.67	7	0.86	N/A	Supports
Mansa	150.8	5	0.00	5	0.40	N/A	Supports
Bathinda	296.2	8	0.25	2	0.00	N/A	Contradicts
Pooled (5 reliable, n ≥ 3)	—	18	0.44	32	0.81	0.009	Supports

Note: Per-district accuracy conditioned on agreement vs. disagreement using IMD operational categories (Deficient: <80% LPA; Normal: 80–120%; Above Normal: >120%). Pooled results among reliable districts (n ≥ 3 for both cells) confirm that the directional finding holds under this alternative scheme.

References

[1] Sandhu, S., Kaur, P., 2023. A case study on the changing pattern of monsoon rainfall duration and its amount during recent five decades in different agroclimatic zones of Punjab state of India. Mausam. 74(3), 651–662.

[2] Kumar, A., Giri, R.K., Taloor, A.K., et al., 2021. Rainfall trend, variability and changes over the state of Punjab, India 1981–2020: A geospatial approach. Remote Sensing Applications: Society and Environment. 23, 100595.

[3] Solomon, S., Qin, D., Manning, M., et al., 2007. IPCC Fourth Assessment Report (AR4): Climate

Change 2007: The Physical Science Basis. Cambridge University Press: Cambridge, UK.

[4] Khera, S., 2024. Analysis of Long-term Precipitation Trends in Punjab, India. Ecology, Environment & Conservation. 30, S44–S52.

[5] Intergovernmental Panel on Climate Change, 2023. Technical summary. In: Masson-Delmotte, V., Zhai, P., Pirani, A., et al. (Eds.). Climate Change 2021: The Physical Science Basis. Cambridge University Press: Cambridge, UK. pp. 35–144.

[6] Nayak, G.H., Alam, W., Singh, K.N., et al., 2024. Modelling monthly rainfall of India through transformer-based deep learning architecture. Model Earth Systems and Environment. 10(3), 3119–3136.

[7] Mekanik, F., Imteaz, M.A., Gato-Trinidad, S., et al., 2013. Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes. Journal of Hydrology. 503, 11–21.

[8] Mosavi, A., Ozturk, P., Chau, K.W., 2018. Flood prediction using machine learning models: Literature review. Water. 10(11), 1536.

[9] Cheng, C., Sa-Ngasoongsong, A., Beyca, O., et al., 2015. Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. IIE Transactions. 47(10), 1053–1071.

[10] Tien Bui, D., Tuan, T.A., Klempe, H., et al., 2016. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides. 13(2), 361–378.

[11] Ovadia, Y., Fertig, E., Ren, J., et al., 2019. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In: Wallach, H., Larochelle, H., Beygelzimer, A., et al. (Eds.). Advances in Neural Information Processing Systems 32. Curran Associates, Inc.: New York, NY, USA.

[12] Willard, J., Jia, X., Xu, S., et al., 2020. Integrating physics-based modeling with machine learning: A survey. arXiv preprint. arXiv:2003.04919. DOI: https://arxiv.org/abs/2003.04919

[13] Dietterich, T.G., 2000. Ensemble methods in machine learning. In International Workshop on Multiple Classifier Systems (MCS 2000), Vol 1857: Lecture Notes in Computer Science. Springer: Berlin/Heidelberg, Germany. pp. 1–15.

[14] Krogh, A., Vedelsby, J., 1994. Neural network ensembles, cross validation, and active learning. In: Tesauro, G., Touretzky, D., Leen, T. (Eds.). Advances in Neural Information Processing Systems 7. MIT Press: Cambridge, MA, USA. pp. 231–238.

[15] Ali, S., 2013. Time series analysis of Baghdad rainfall using ARIMA method. Iraqi Journal of Science. 54(4. Appendix), 1136–1142.

[16] Box, G.E., Jenkins, G.M., Reinsel, G.C., et al., 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons: Hoboken, NJ, USA.

[17] Senjaliya, J., Vaghasia, V., Shah, S.M., 2026. A comprehensive review of machine learning and deep learning approaches for rainfall forecasting: Current progress, challenges, and future directions. Theoretical and Applied Climatology. 157(3), 177.

[18] Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment. 19(5), 326–339.

[19] Nirmala, M., Sundaram, S.M., 2010. A seasonal ARIMA model for forecasting monthly rainfall in Tamil Nadu. National Journal on Advances in Building Sciences and Mechanics. 43–47.

[20] Sharma, A., Singh, K., 2024. ARIMA-based forecasting of monthly rainfall in Mandi district, Himachal Pradesh. Water Supply. 24(9), 3226–3237.

[21] Dahiya, P., Kumar, M., Manhas, S., et al., 2024. Time series study of climate variables utilising a seasonal ARIMA technique for the Indian states of Punjab and Haryana. Discover Applied Sciences. 6(12), 650.

[22] Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice. OTexts: Melbourne, Australia.

[23] Nearing, G.S., Kratzert, F., Sampson, A.K., et al., 2021. What role does hydrological science play in the age of machine learning? Water Resources Research. 57(3), e2020WR028091.

[24] Breiman, L., 2001. Random forests. Machine Learning. 45(1), 5–32.

[25] Wani, O.A., Mahdi, S.S., Yeasin, M., et al., 2024. Predicting rainfall using machine learning, deep learning, and time series smodels across an altitudinal gradient in the North-Western Himalayas. Scientific Reports. 14(1), 27876.

[26] Yu, P.S., Yang, T.C., Chen, S.Y., et al., 2017. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. Journal of Hydrology. 552, 92–104.

[27] Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation. Hydrology and Earth System Sciences. 20(7), 2739–2758.

[28] Mishra, P., Al Khatib, A.M.G., Yadav, S., et al., 2024. Modeling and forecasting rainfall patterns in India: A time series analysis with XGBoost algorithm. Envi-

ronmental Earth Sciences. 83(6), 163.

[29] Alam, M.J., Majumder, A., 2022. A Comparative Analysis of ARIMA and other Statistical Techniques in Rainfall Forecasting: A Case Study in Kolkata (KMC), West Bengal. Current World Environment. 18(3).

[30] Settles, B. 2009. Computer Sciences Technical Report 1648: Active Learning Literature Survey. University of Wisconsin–Madison: Madison, WI, USA.

[31] Wood, D., Mu, T., Webb, A.M., et al., 2023. A unified theory of diversity in ensemble learning. Journal of Machine Learning Research. 24(359), 1–49.

[32] Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In: von Luxburg, C., Guyon, I., Bengio, S. (Eds.). Advances in Neural Information Processing Systems 30. Curran Associates Inc.: New York, NY, USA. pp. 6405–6416.

[33] Yang, P., Ren, G., Yan, P., 2017. Evidence for a strong association of short-duration intense rainfall with urbanization in the Beijing urban area. Journal of Climate. 30(15), 5851–5870.

[34] Xia, J., She, D., Zhang, Y., et al., 2012. Spatio-temporal trend and statistical distribution of extreme precipitation events in Huaihe River Basin during 1960–2009. Journal of Geographical Sciences. 22(2), 195–208.

[35] Rao, B.B., Chowdary, P.S., Sandeep, V.M., et al., 2014. Rising minimum temperature trends over India in recent decades: Implications for agricultural production. Global and Planetary Change. 117, 1–8.

[36] Madane, D.A., Waghaye, A.M., 2023. Spatio-temporal variations of rainfall using innovative trend analysis during 1951–2021 in Punjab State, India. Theoretical and Applied Climatology. 153(1), 923–945.

[37] Ministry of Environment, Forest and Climate Change, n.d. Environmental Information System (ENVIS). Available from: https://moef.gov.in/environmental-information-system-envis (cited 27 April 2026).

[38] Yadav, B.P., Saxena, R., Kr. Das, A., et al., 2022. Rainfall Statistics of India—2021. Report No. MoES/IMD/HS/Rainfall Report/02(2022)/60. Hydromet Division, India Meteorological Department (IMD), Ministry of Earth Sciences: New Delhi, India. Available from: https://www.scribd.com/document/697389656/Rainfall-Statistics-of-India-2021

[39] Sharma, Y., Sidana, B.K., Kumar, S., et al., 2023. Pre and post water level behaviour in Punjab: Impact analysis with DID approach. Sustainability. 15(3), 2426.

[40] Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 22(158), 209–212.

[41] Open Government Data (OGD) Platform India, 2021. District-wise annual average rainfall of Punjab (1970–2021). Available from: https://www.data.gov.in/resource/district-wise-annual-average-rainfall-punjab-1970-2021 (cited 27 April 2026).