Comparing skill of historical rainfall data based monsoon rainfall prediction in India with NWP forecasts

Aastha Jain,[a] Apoorva Narula,[b] Jatin Batra,[c] M Rajeevan,[d] Sandeep Juneja [a]

[a] *Ashoka University*

[b] *Indian Institute of Technology Delhi*

[c] *Tata Institute of Fundamental Research*

[d] *Atria University*

Corresponding author: Aastha Jain, aasthajn25@gmail.com

ABSTRACT: The Indian summer monsoon is a highly complex and critical weather system that directly affects the livelihoods of over a billion people across the Indian subcontinent. Accurate short-term forecasting remains a major scientific challenge due to the monsoon's intrinsic nonlinearity and its sensitivity to multi-scale drivers, including local land-atmosphere interactions and large-scale ocean-atmosphere phenomena. In this study, we address the problem of forecasting daily rainfall across India during the summer months, focusing on both one-day and three-day lead times. We use Autoformers - deep learning transformer-based architectures designed for time series forecasting. These are trained on historical gridded precipitation data from the Indian Meteorological Department (1901–2023) at spatial resolutions of $0.25^\circ \times 0.25^\circ$. The models also incorporate auxiliary meteorological variables from ECMWF's reanalysis datasets, namely, cloud cover, humidity, temperature, soil moisture, vorticity, and wind speed. Forecasts are benchmarked against ECMWF's High-Resolution Ensemble System (HRES), widely regarded as the most accurate numerical weather predictor. We conduct both nationwide evaluations and localized analyses for major Indian cities. Our results indicate that transformer-based deep learning models consistently outperform HRES-NWP and other climatological baselines. Specifically, compared to our model, HRES forecasts exhibit approximately 22% higher error for one-day predictions, and 27% higher error for three-day forecasts. Persistence-based predictions show a 40% and 69% higher error for one-day and three-day forecasts, respectively. Moreover, our models demonstrate superior skill in balancing heavy rainfall detection with false positives. We also find that incorporating historical data up to 20 days prior significantly reduces forecast error, particularly in landlocked regions. Our findings suggest that NWP forecasts for the Indian monsoon can be substantially improved by integrating diverse, high-resolution observational data with carefully designed deep learning models tailored to monsoon dynamics.

1. Introduction

Accurate rainfall prediction in India during monsoons is crucial for a variety of reasons: agriculture planning, disaster management, day-to-day transportation planning, and so on. Anecdotally, it is well known that numerical weather prediction (NWP) does not perform well in the prediction of rainfall for India (Rajeevan 2023). It is also conjectured that during monsoons, rainfall data across India have spatial-temporal memory so that information on rainfall early in neighboring parts may be useful for future rainfall prediction (Goswami and Xavier 2003). In addition, rainfall has been shown to be also affected by a variety of other atmospheric, land, and ocean variables, such as temperature, wind, soil moisture, etc. (Prasad and Singh 1988).

In this paper, we consider daily gridded precipitation data from India Meteorological Department (IMD) (Pai et al. 2014) available from 1901–2023, at a spatial resolution of $0.25^\circ \times 0.25^\circ$[1]. We use this to predict rainfall for all of India, one day and three days in the future. We also use daily atmospheric and land data as additional covariates in an attempt to improve our forecasts. We compare our performance with operational NWP forecasts from the High-Resolution Integrated Forecast System (HRES-IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF (2021)). HRES is widely regarded the leading operational weather forecasting system in the world (Lam et al. (2023)).

Several attempts have been made to predict rainfall using machine learning (ML) techniques. For long-range forecasting of monsoon rainfall in India, Rajeevan et al. (2000, 2007) used a host of methods such as multivariate principal component regression, simple neural networks, linear discriminant analysis, ensemble multiple linear regression, and projection pursuit regression. They used multimodal data such as air temperature, sea surface temperature, rainfall, air pressure etc. These developments helped support IMD's two-stage monsoon forecasting system with the first stage forecast given in mid April and an updated second stage forecast given at the end of June.

More recently, deep learning and machine learning approaches have been explored for short-range rainfall prediction. Kumar et al. (2021, 2022) conducted a comparative analysis of Long-Short-Term-Memory (LSTM) and ConvLSTM models trained using ground-based IMD rainfall data and satellite data for Indian summer monsoon rainfall. They showed correlation coefficients between

[1] In the more elaborate arxiv version of the paper, we include comparisons with NCEP-NWP as well. We also conduct comparisons for coarser $1^\circ \times 1^\circ$ grids and observe similar results.

the observed and predicted rainfall of 0.67 for 1 day and 0.42 for 2 day lead time, respectively, indicating reasonable skill in short-range precipitation forecasting. However, their results also highlighted that model efficiency quickly drops after 2 days lead time, pointing to a limitation in capturing longer temporal dependencies. Similarly, Praveen et al. (2020) analyzed rainfall trends and forecasting using machine learning. They mainly focus on predicting long-term trends. Jose et al. (2022) developed ensemble predictions of daily precipitation and temperature using machine learning. Their analysis was limited to coarser spatial scales (1° resolution). Miao et al. (2019) used a combined CNN-LSTM neural network to improve the prediction of monsoon precipitation. The last three references relied on reanalysis data for ground truth. Typically, both the NWP based prediction models (such as HRES) and machine learning based prediction models, such as GraphCast (Lam et al. 2023) and ClimaX (Nguyen et al. 2023), also rely on reanalysis data for initialization and, in case of ML models, training. In our work, we instead use the IMD data for training, which is shown to be better representative of the ground truth (Kishore et al. 2016).
Chen et al. (2023) provide a comprehensive survey of machine learning methods in weather and climate applications and highlight persistent challenges such as the underprediction of extreme rainfall events and the need for better integration with physical models.

Our work adds to the growing literature on the use of ML for short-term rainfall prediction, primarily using historical IMD data and benchmarking against NWP forecasts. We compare and contrast the deep learning-based forecasts generated by autoformers (Wu et al. 2021) using historical rainfall data from IMD (referred to as DL-HD forecasts) and the forecasts generated using IMD rainfall data and additional covariates (called DL-HD+Covariates), with the HRES-NWP forecasts. In an attempt to arrive at improved forecasts, we also combine the NWP forecasts with DL-HD+Covariates using a simple neural network, to generate ensemble forecasts. Much of the rainfall in India occurs during the four monsoon months of June, July, August and September (JJAS), and hence, to build a more useful forecasting system, we restrict our forecasts to JJAS.
We find that for forecasting precipitation one and three days into the future, forecasts obtained by DL-HD+Covariates are substantially more accurate compared to NWP forecasts as well as forecasts based on climatological baselines. We also discuss the improvement in performance of our forecasts when they are combined with NWP forecasts. We further observe that using

autoformers, data up to 20 days in the past is useful in reducing errors of one and three day forecasts.

2. Data and Experiments

a. Data Sources

1. **IMD Ground Truth**: We use daily gridded precipitation data obtained from IMD spanning the period from 1901 to 2023, at a spatial resolution of $0.25° \times 0.25°$ (Pai et al. 2014). At this resolution, the geographical extent of India is discretized into 12,422 grids. This forms the ground truth dataset against which our predictions and other models are compared.

2. **Additional weather variables**: Apart from precipitation, we also use daily atmospheric and land data at 0.25° resolution provided by ECMWF as part of their reanalysis products (Copernicus Climate Change Service 2019). These variables include: horizontal and vertical components of wind at 10m, temperature, soil moisture, cloud cover, vorticity at 850hPa, humidity, and divergence at 700hPa. These are the lower tropospheric pressure levels, indicative of cloud development and rainfall processes. The data is available from 1950 onwards.

3. **NWP forecasts**: We assess performance relative to HRES, a commonly adopted NWP benchmark. The HRES daily forecasts are obtained from ECMWF (ECMWF 2021) for all years 2011 onwards, at a resolution of 0.25°, for both 1 and 3 days into the future. The dataset is downloaded only for the JJAS months.

b. Dataset Preparation

We compare the 06:00 UTC daily HRES forecasts for lead times of 1 and 3 days with the corresponding deep-learning-based forecasts. All data are aligned to 06:00 IST to ensure proper temporal consistency with the ground truth. The HRES data provide cumulative precipitation forecasts over 24-hour and 72-hour periods, which are directly comparable to our 1-day and 3-day predictions, respectively.

For DL-HD and DL-HD+Covariates training, the IMD dataset is partitioned into training (1901–2011) and test (2012–2023) subsets. Training samples are constructed using a time window

approach, where each input consists of rainfall data from all grid points over d contiguous days. The model is trained to predict the cumulative rainfall for the $(d+1)^{\text{th}}$ day at the same grid points.

c. Experiments

Below we outline how forecasts are generated using different models for lead times of 1 and 3 days.

1. **DL-HD:** We generate forecasts for all n grids across India using historical rainfall data from IMD, utilizing varying lengths of past information, spanning from 3 to 20 days (d). The input dimensions for the models are structured as $n \times d$, capturing the historical rainfall data for all grids over the specified timeframe. The output dimension is n, representing the forecasted rainfall for the subsequent day at each grid point.

   This is implemented using the Autoformer architecture, an advanced variant of the Transformer model specifically designed for time series forecasting. Autoformer incorporates two core components that make it particularly effective for modeling the complex spatiotemporal dynamics of rainfall data:

   (a) Series decomposition: This mechanism decomposes the input time series into trend and seasonal components at multiple stages within the model. The trend component captures long-term variations in rainfall, while the seasonal component isolates shorter-term, repeating fluctuations.

   (b) Auto-correlation mechanism: Replacing the standard self-attention mechanism, Autoformer employs an auto-correlation approach that identifies repeating patterns over different time horizons. This helps capture periodic dependencies and reduces computational complexity from quadratic to near-linear in sequence length.

   We train the Autoformer model using data from 1901 to 2011 and generate test forecasts for the years 2012 to 2023. More details on the Autoformer architecture we use in our experiments is given in Appendix b.

2. **DL-HD + Covariates:** This is an extension of the above model where we use past d days of precipitation data ($3 \leq d \leq 20$) from IMD, and past 3 days of reanalysis data. As stated earlier, the additional covariates include wind speed, temperature, soil moisture, cloud cover,

vorticity, humidity, and divergence. The choice of these variables is justified later in Section 2e. The number of past days for reanalysis data were fixed to 3 as we did not observe any significant reduction in errors when incorporating information beyond 3 days. The input dimensions are now structured as $n \times d \times v$, where v is the number of covariates. The output is again n-dimensional, forecasting the rainfall value at each grid.

3. **NWP:** HRES daily forecasts are available for lead times of 1 and 3 days. These are compared with forecasts made using IMD data.

4. **NWP+:** We combine the HRES forecasts at the target grid and the 4 neighbouring grids using a deep neural network, which is trained to minimize the error between the forecast and IMD ground truth for the particular grid. The resulting forecast is called NWP+ prediction. The model here is trained from 2011 to 2020 and test forecasts are generated for 2021 - 2023. While using 4 surrounding grids improves the forecasts somewhat, we did not see further improvement with a higher number of grids.

5. **Ensemble:** We combine the DL-HD + Covariates forecasts, and the HRES forecasts of the 5 grids, to generate an ensemble forecast for each grid. This is done using a deep neural network. The models here are trained from 2011-2020, and forecasts are generated for 2021-2023.

We also use the following simpler baselines to benchmark our models against:

1. **Persistence:** This is a naive forecast which estimates the rainfall on day $d+1$ and the average of rainfall in days $d+1, d+2$ and $d+3$ as the observed rainfall on day d for each grid. This is reported for the period 2012-2023.

2. **Climatological mean:** This baseline estimates rainfall by computing the mean rainfall for the same calendar day across all previous years (1901–2011). It captures long-term seasonal trends and is used to benchmark model performance against historical averages.

3. **Rolling mean (20 Days):** This baseline forecasts rainfall as the average rainfall over the preceding 20 days for each grid point. It captures recent trends and smooths short-term variability but does not incorporate spatial or seasonal context.

4. **AR(1) - Temporal linear model:** This autoregressive baseline uses only the previous day's rainfall value at each grid point to forecast the next day's rainfall. A simple linear regression model is fit using this single temporal lag as input.

5. **AR(5) + spatial grids - spatiotemporal linear model:** This extends the autoregressive approach by incorporating the past 5 days of rainfall data for each grid point along with its four immediate neighboring grid points (up, down, left, right). A linear regression model is trained using this spatiotemporal input.

d. Loss function

Since rainfall forecasting is a regression task, the mean squared error (MSE) is the conventional choice of loss function. However, during training, we found that models optimized with MSE tend to produce overly smooth predictions that fail to capture extreme rainfall events. This behavior is particularly evident in Figures 1a and 1b, where MSE-trained models (dotted line) track the long-term average but significantly underestimate peaks for Mumbai and Ahmedabad, respectively. This limitation stems from MSE's symmetric treatment of over- and underestimation errors, which biases the model toward minimizing large deviations without prioritizing rare but high-impact events.

To address this, we propose a *peak-biased loss function* that places greater emphasis on underestimation. This reflects the practical importance of accurately forecasting extreme rainfall, where missing a peak can have far more serious consequences than a false alarm. The loss function is defined as:

$$L = \frac{1}{N}\sum_{t=1}^{N}\left[\mathbb{I}(\hat{r}_t < r_t)\cdot|r_t-\hat{r}_t|^{\alpha} + \mathbb{I}(\hat{r}_t > r_t)\cdot|\hat{r}_t-r_t|^{\beta}\right], \tag{1}$$

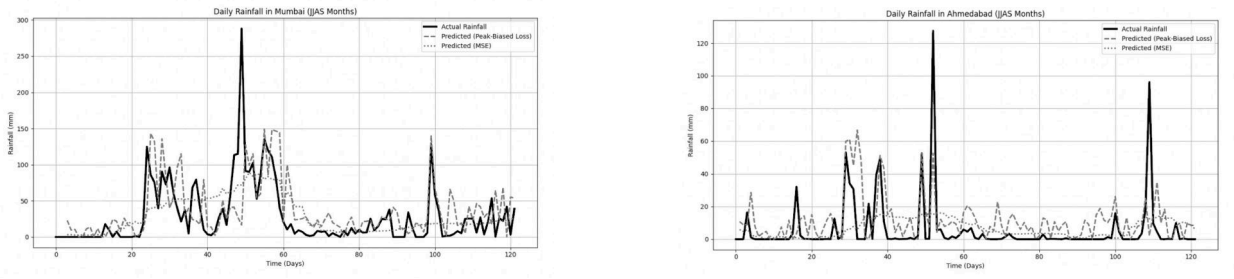
where r_t and $\hat{r}_t$ denote the observed and predicted rainfall at time t, respectively, and $\alpha > \beta$ ensures that underestimation is penalized more heavily than overestimation. We use $\alpha = 1.5$, $\beta = 1.0$, selected empirically to optimize forecast quality.

To validate this choice, we conduct a sensitivity analysis across various values of α and β. Results across all percentiles (Tables A1–A4) consistently show $\alpha = 1.5, \beta = 1.0$ achieving optimal balance between detection accuracy and false alarm rates.

This design is also supported by prior work addressing rare event prediction. For instance, Shi et al. (2017) and Xu et al. (2024) use such custom weighted loss functions to improve performance on infrequent but critical events.

The advantage of our loss function is also visually evident. In the case of Mumbai (Figure 1a), the model trained with the peak-biased loss (dashed line) captures multiple significant rainfall events missed by the MSE-trained model, particularly between days 20-40 and near day 50. In Ahmedabad (Figure 1b), the peak-biased model estimates the ~130mm rainfall at day 50 more accurately, while the MSE model significantly underpredicts it.

For completeness, we report both the proposed peak-biased loss L and MSE on the test set (Tables 2 and 3). The trends are consistent across both metrics: lower MSE corresponds to lower L, with peak-biased models outperforming in both average error and peak detection.



(a) (b)

FIG. 1: Plots comparing the predictions generated by the same Autoformer model under MSE and the proposed *peak-biased loss* in (a) Mumbai and (b) Ahmedabad respectively.

e. Feature selection

We adopt a sequential feature selection approach to identify input variables for forecasting rainfall. This greedy algorithm adds one variable at a time, selecting at each step the variable that yields the greatest reduction in forecasting error, as measured by the peak-biased loss function.

The process begins by evaluating each variable individually. The variable with the lowest standalone error is selected first. At each subsequent step, the model assesses the marginal improvement from adding each remaining variable to the current set and selects the one that most improves performance. This continues until no further variable leads to a significant reduction in error.

Table 1 outlines the progression of this selection. Past precipitation is chosen first, achieving the

lowest initial error. It is followed by cloud cover, vorticity, humidity, soil moisture, wind, and finally temperature, each selected for their incremental contribution to minimizing the loss. This method prioritizes immediate gains in predictive accuracy rather than exploring all possible feature combinations.

TABLE 1: Peak-biased loss ($mm^{1.5} + mm$) recorded for different combinations of input variables using a greedy approach

Variables	**Cloud Cover (CC)**	**H Wind (HW)**	**V Wind (VW)**	**Temperature (T)**	**Humidity (H)**	**Soil Moisture (SM)**	**Vorticity (Vo)**
Precipitation (Ppt)	**19.95**	21.13	21.48	21.01	20.71	20.52	20.39
Ppt + CC	-	19.63	19.88	20.20	20.01	**19.38**	19.47
Ppt + CC + Vo	-	19.73	19.71	19.36	**19.07**	19.33	-
Ppt + CC + Vo + H	-	18.91	18.96	18.99	-	**18.56**	-
Ppt + CC + Vo + H + SM	-	**18.42**	18.59	18.53	-	-	-

f. Model Configuration and Training

All experiments are conducted in Python, utilizing the TensorFlow and Keras libraries. DL-HD and DL-HD+Covariates forecasts are made using the Autoformer architecture. This is a recurrent model and we use past d days of v variables as input, where d ranges from 3 to 20. The number of parameters depend roughly linearly on d, with the number of parameters being approximately 200M for $d = 12$. This architecture is adopted from Wu et al. (2021). Since there are fewer data points, the models for NWP+ and Ensemble are trained using smaller feed-forward neural networks, with 2 hidden layers. All models are trained using the Adam optimizer, optimizing the peak-biased loss specified in (1). More details on the architecture of these models are given in the Appendix b. To ensure robustness of our models, each experiment is conducted across 10 independent runs, employing randomly generated seeds to initialize neural network parameters differently. Performance metrics are reported on the average prediction obtained from these 10 runs.

3. Results

The predictions based on the models specified in Section 2c are compared with the ground truth daily rainfall data from IMD. We outline results both in an all-India average sense and separately for a set of key cities to capture regional performance differences.

a. Comparison for entire India

The average peak-biased loss over India for 1-day forecasts is presented in Table 2. We also compare the spatial distribution of prediction skill for 1 and 3 day prediction on July 15, 2022 in Figure 3. From the analysis, we make the following observations:

1. **DL-HD + Covariates** achieves the lowest error and outperforms all other models.
2. Among the climatological baselines, the **Rolling Mean (20 Days)** performs worst, exhibiting 81.68% higher error than DL-HD + Covariates, followed by the **Climatological Mean** with 59.02% higher error.
3. **Persistence** has relatively high error but performs better than the climatological baselines, with 39.36% higher MSE.
4. **HRES-NWP** performs better than persistence but still has 32.90% higher error relative to DL-HD + Covariates.
5. Pooling neighboring grids in **HRES-NWP+** slightly improves over HRES, yet remains 28.21% worse than DL-HD + Covariates.
6. The **Ensemble (NWP + DL-HD + Covariates)** demonstrates improvement over NWP alone, with an error 9.56% higher than DL-HD + Covariates.
7. The regression-based models, **AR(1)** and **AR(5) + Spatial Grids**, perform worse than DL-HD + Covariates, with MSEs 32.18% and 30.98% higher respectively, indicating the DL model's advantage in capturing non-linear patterns beyond lag-based predictors.

b. Comparison for key cities

We also analyzed the model performance separately for 20 of the most populated cities spread across India. These cities were selected to provide a representative distribution across coastal and landlocked regions, and to ensure significant rainfall during the monsoon months (JJAS), which is important for evaluating rainfall forecasting accuracy. A map showing the geographical distribution of these selected cities is included in Figure 2.

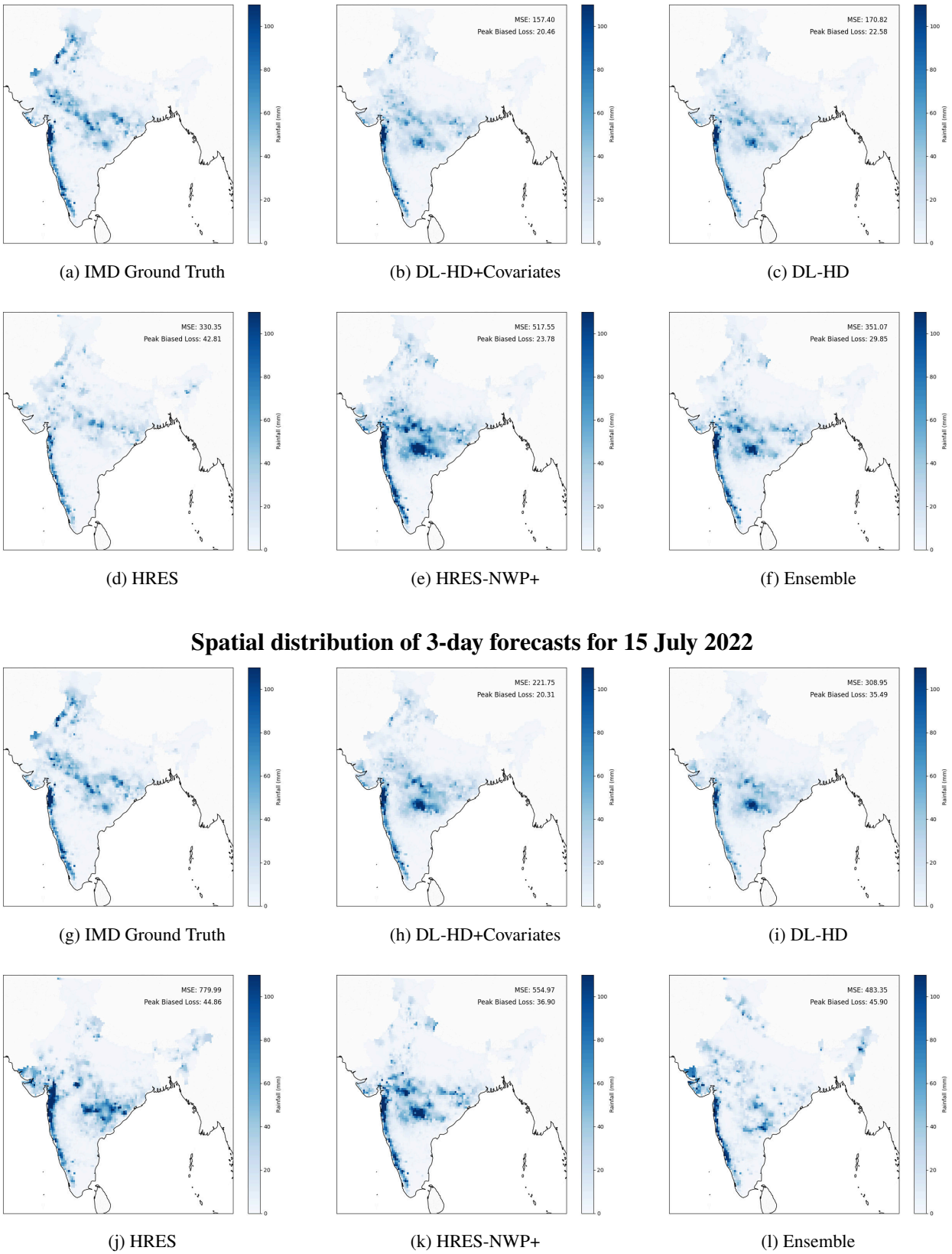
The average peak-biased loss for these cities is shown in Table 4 for 1-day forecasts, and in Table 5 for 3-day forecasts. The last row in both tables shows the excess error percentage of the

different forecasts compared to the DL-HD+Covariates forecast. We make similar observations here as for the whole of India.

Figures 4 to 7 graphically compare the different forecasts with the ground truth for the cities Mumbai, Bhopal, Ahmedabad, and Chennai, for the months of July and August in 2022, for 1-day forecasts. Similar comparisons for 3-day forecasts are shown in Figures 9 to 11.

It is clear from the figures that DL-HD+Covariates forecasts consistently outperform other methods in tracking actual rainfall.

FIG. 2: Study area map showing the selected 20 cities across India, including both coastal and landlocked locations. The cities marked with a cross are chosen for detailed analysis based on the rainfall during JJAS and the population in these cities.



(a) IMD Ground Truth (b) DL-HD+Covariates (c) DL-HD

(d) HRES (e) HRES-NWP+ (f) Ensemble

(g) IMD Ground Truth (h) DL-HD+Covariates (i) DL-HD

(j) HRES (k) HRES-NWP+ (l) Ensemble

FIG. 3: Spatial distribution of forecasts for 15 July 2022. Top: 1-day forecasts. Bottom: 3-day forecasts.

TABLE 2: Comparison of 1-day ahead precipitation forecasting performance over India at 0.25° resolution.

Model	**Peak-biased Loss** ($mm^{1.5} + mm$)	**MSE** (mm^2)	**% Higher Error vs DL-HD + Covariates**
DL-HD + Covariates	18.24	268.59	-
DL-HD	20.90	312.11	16.21
HRES-NWP	22.25	356.97	32.90
HRES-NWP+	22.13	344.18	28.21
Ensemble	18.96	294.25	9.56
Persistence	25.42	448.10	39.36
Climatological Mean	27.10	510.12	59.02
Rolling Mean (20 Days)	29.75	563.20	81.68
AR(1)	28.58	397.66	32.18
AR(5) + Spatial Grids	27.22	364.21	30.98

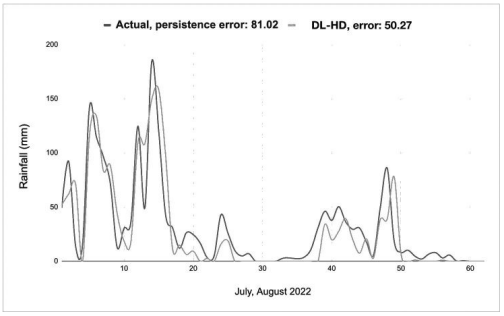
TABLE 3: Comparison of 3-day ahead precipitation forecasting performance over India at 0.25° resolution.

Model	**Peak-biased Loss** ($mm^{1.5} + mm$)	**MSE** (mm^2)	**% Higher Error than DL-HD + Covariates**
DL-HD + Covariates	67.28	2878.52	-
DL-HD	81.87	3752.44	30.37
HRES-NWP	85.59	4486.25	55.85
HRES-NWP+	84.15	3884.24	34.89
Ensemble	74.73	3019.81	4.91
Persistence	114.06	8300.21	188.34
Climatological Mean	120.50	8962.63	211.31
Rolling Mean (20 Days)	126.75	9445.16	228.16
AR(1)	106.42	5386.11	87.18
AR(5) + Spatial Grids	83.45	4707.47	63.44

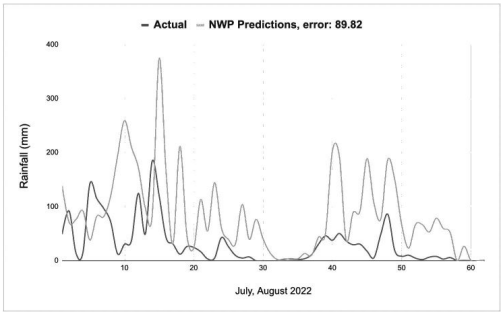
TABLE 4: Average peak-biased loss ($mm^{1.5} + mm$) for 1-day forecasts in grids corresponding to 20 major cities across India

City	DL-HD+Covariates	DL-HD	HRES-NWP	HRES-NWP+	Ensemble
Ahmedabad	**16.23**	18.46	25.34	23.15	17.32
Bangalore	11.76	12.28	14.67	12.33	**11.12**
Bhopal	**24.35**	26.76	28.92	27.14	23.98
Bhubaneswar	**26.42**	29.11	28.04	27.91	27.74
Chandigarh	18.45	**18.35**	20.42	20.18	19.65
Chennai	**10.33**	11.15	12.40	12.26	12.18
Coimbatore	10.27	10.77	11.55	11.31	**10.09**
Delhi	**7.44**	7.86	12.79	11.77	8.85
Gangtok	**38.71**	41.47	42.29	41.93	39.33
Hyderabad	**18.55**	20.28	26.51	23.14	20.08
Indore	11.22	11.41	15.65	15.41	**10.40**
Kochi	19.26	22.77	26.13	25.42	**19.07**
Kolkata	**35.96**	37.58	41.75	40.59	36.86
Lucknow	11.86	**11.72**	16.46	16.11	11.99
Mumbai	**42.48**	47.50	66.13	63.81	45.05
Patna	12.71	13.33	15.58	15.51	**12.46**
Pune	**15.11**	16.85	25.46	25.29	16.78
Raipur	**26.38**	27.51	29.53	28.17	28.47
Shimla	**9.62**	11.77	12.21	11.61	9.82
Vishakhapatnam	23.49	26.75	27.21	27.18	**22.53**
Total Error	**390.60**	423.69	499.36	481.01	402.77
%age higher	0	8.38	27.77	23.05	3.08

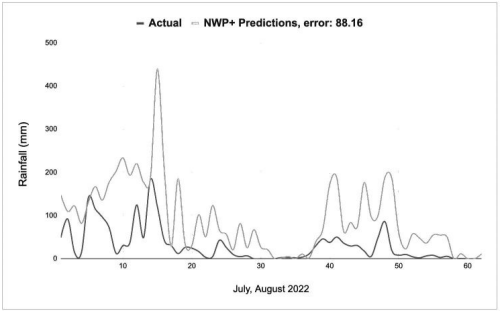
FIG. 4: 1-day forecasts for Mumbai in July and August 2022. DL-HD+Covariates predictions closely track the ground truth, while HRES predictions tend to over estimate the rainfall. The ensemble is a significant improvement over NWP alone, and can be seen to capture most of the high rainfall events during this period.



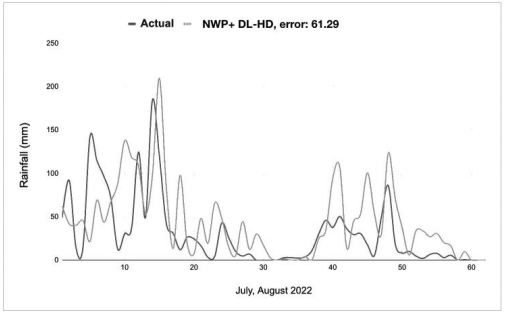
(a) DL-HD + Covariates vs IMD



(b) HRES vs IMD

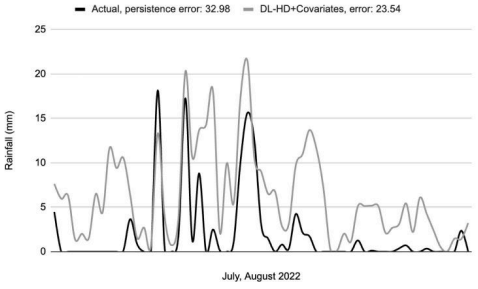


(c) HRES+ vs IMD

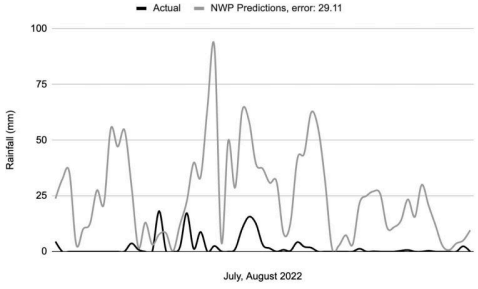


(d) Ensemble vs IMD

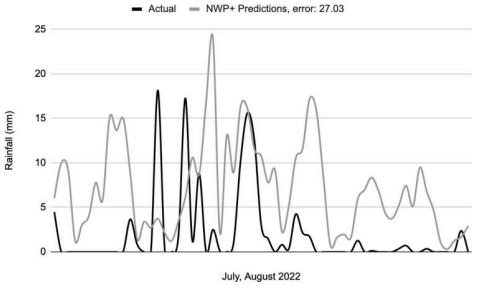
FIG. 5: 1-day forecasts for Bhopal in July and August 2022.



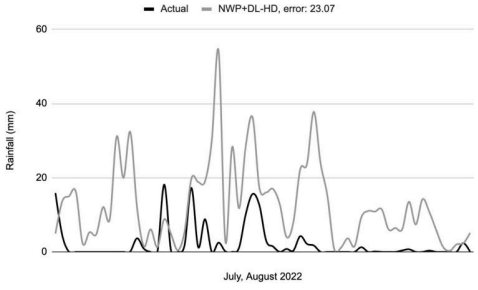
(a) DL-HD + Covariates vs IMD



(b) HRES vs IMD

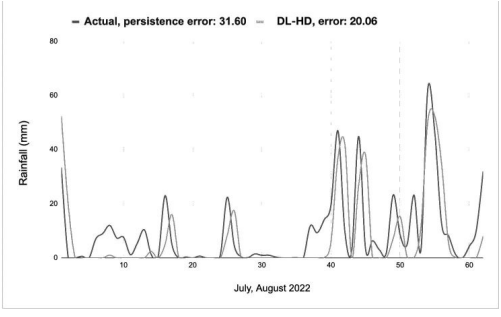


(c) HRES+ vs IMD

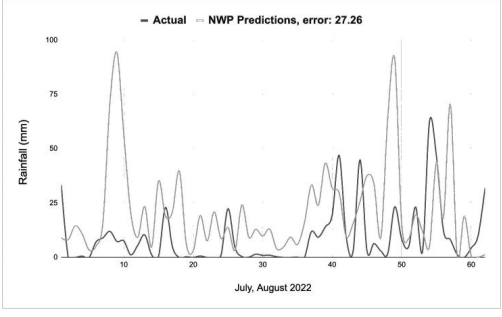


(d) Ensemble vs IMD

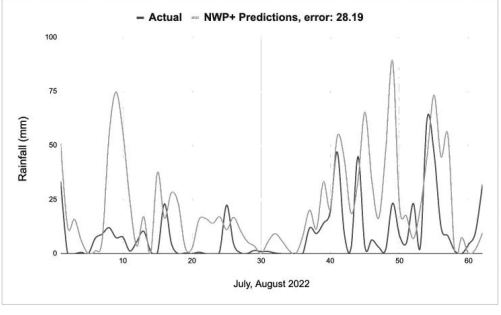
FIG. 6: 1-day forecasts for Ahmedabad in July and August 2022. DL-HD+Covariates closely track the ground truth, capturing most high rainfall events during this period. HRES predictions consistently overestimate the rainfall



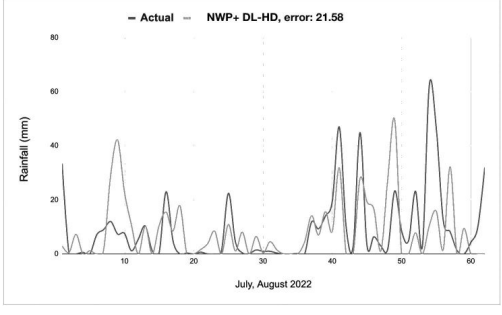
(a) DL-HD + Covariates vs IMD



(b) HRES vs IMD

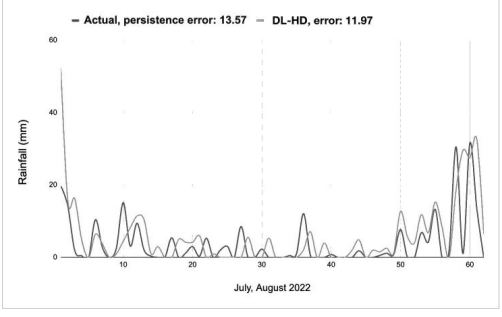


(c) HRES-NWP+ vs IMD

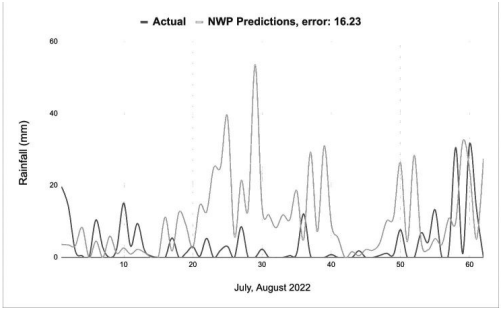


(d) Ensemble vs IMD

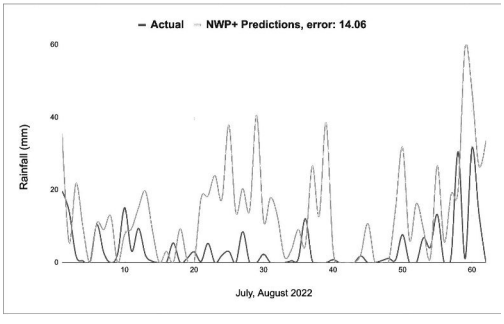
FIG. 7: 1-day forecasts for Chennai in July and August 2022. DL-HD+Covariates closely track the IMD ground truth. HRES predictions consistently overestimate the rainfall



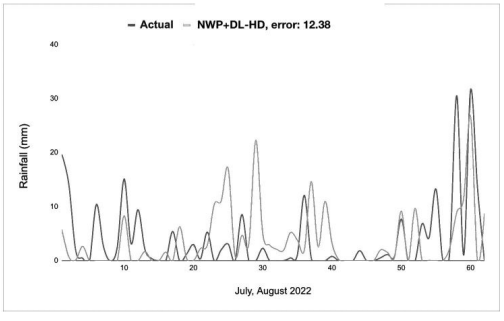
(a) DL-HD + Covariates vs IMD



(b) HRES vs IMD



(c) HRES-NWP+ vs IMD

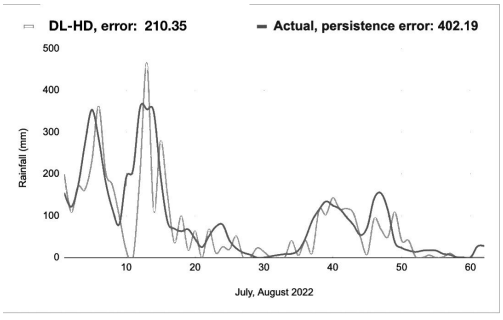


(d) Ensemble vs IMD

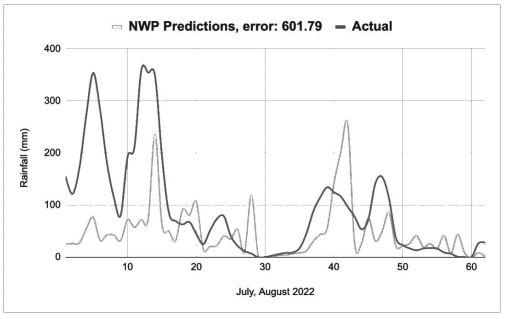
TABLE 5: Average peak-biased loss ($mm^{1.5} + mm$) for 3-day forecasts in grids corresponding to 20 major cities across India

City	DL-HD+Covariates	DL-HD	HRES-NWP	NWP+	Ensemble
Ahmedabad	**58.89**	65.22	66.29	65.96	59.88
Bangalore	40.39	44.73	49.11	48.47	**38.80**
Bhopal	**72.25**	79.16	82.43	81.57	75.49
Bhubaneswar	**56.76**	69.46	74.03	74.14	64.32
Chandigarh	**45.87**	50.25	55.27	53.34	48.19
Chennai	**33.53**	43.10	51.54	49.78	40.22
Coimbatore	**42.15**	47.36	50.96	50.48	46.14
Delhi	**32.16**	32.53	40.53	39.81	32.11
Gangtok	**79.56**	100.39	112.74	109.56	88.46
Hyderabad	**42.27**	50.27	54.13	54.22	44.91
Indore	57.42	57.29	62.94	60.41	**56.71**
Kochi	**59.49**	69.56	74.21	74.44	62.24
Kolkata	**94.26**	112.73	118.36	115.40	99.18
Lucknow	**30.71**	34.58	53.91	51.76	32.15
Mumbai	**153.56**	201.28	215.68	211.42	166.77
Patna	**29.93**	34.22	41.39	41.14	30.28
Pune	**40.75**	50.45	54.23	52.61	46.86
Raipur	**64.66**	73.29	81.64	80.11	70.13
Shimla	**22.42**	31.58	34.75	34.49	21.94
Vishakhapatnam	**72.48**	80.56	76.50	75.93	76.54
Total Error	**1130.85**	1331.49	1449.77	1423.93	1223.65
%age higher	0	17.62	28.51	26.23	6.37

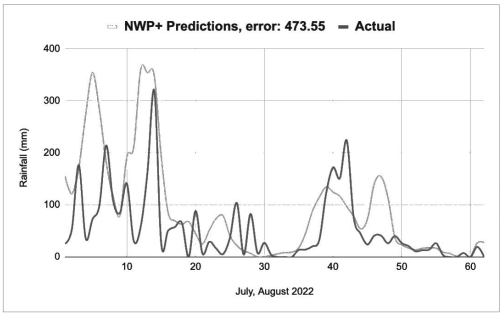
FIG. 8: 3-day forecasts for Mumbai in July and August 2022. None of the forecasts track the IMD ground truth well. However, DL-HD+Covariates capture some of the high and low rainfall events well, and we see a significant improvement using HRES-NWP+.



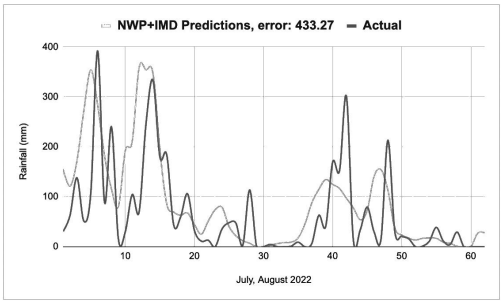
(a) DL-HD+Covariates vs IMD



(b) HRES vs IMD

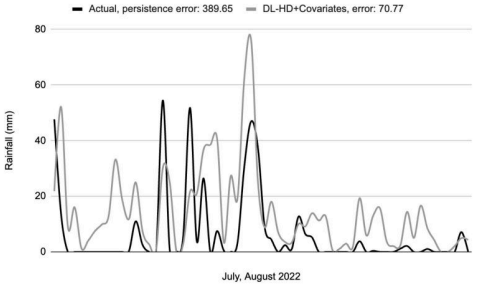


(c) HRES-NWP+ vs IMD

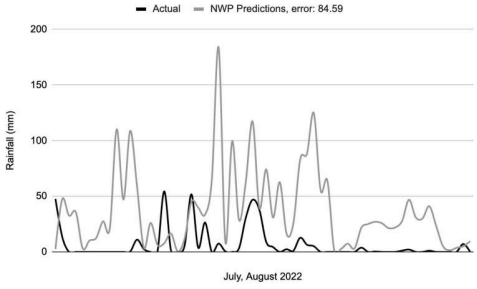


(d) Ensemble vs IMD

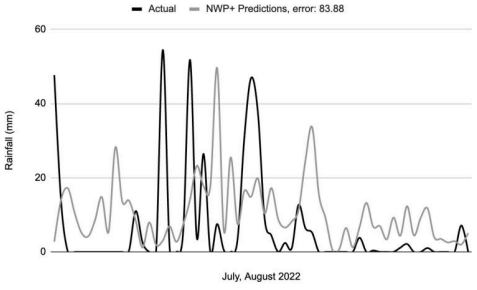
FIG. 9: 3-day forecasts for Bhopal in July and August 2022.



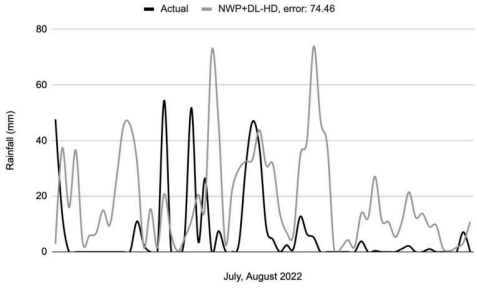
(a) DL-HD+Covariates vs IMD



(b) HRES vs IMD

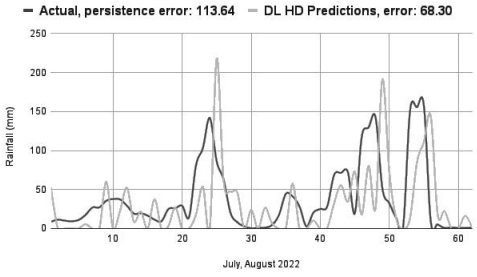


(c) HRES-NWP+ vs IMD

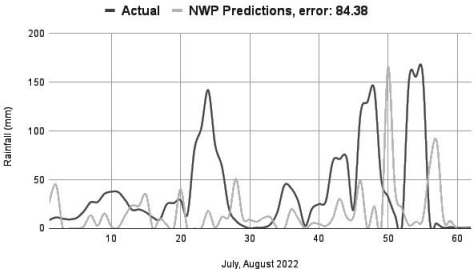


(d) Ensemble vs IMD

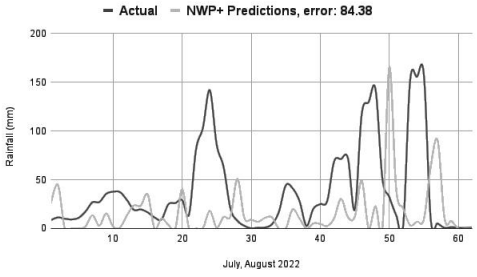
FIG. 10: 3-day forecasts for Ahmedabad in July and August 2022



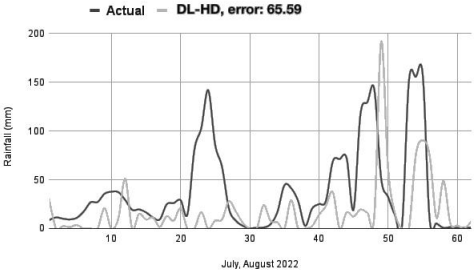
(a) DL-HD+Covariates vs IMD



(b) HRES vs IMD

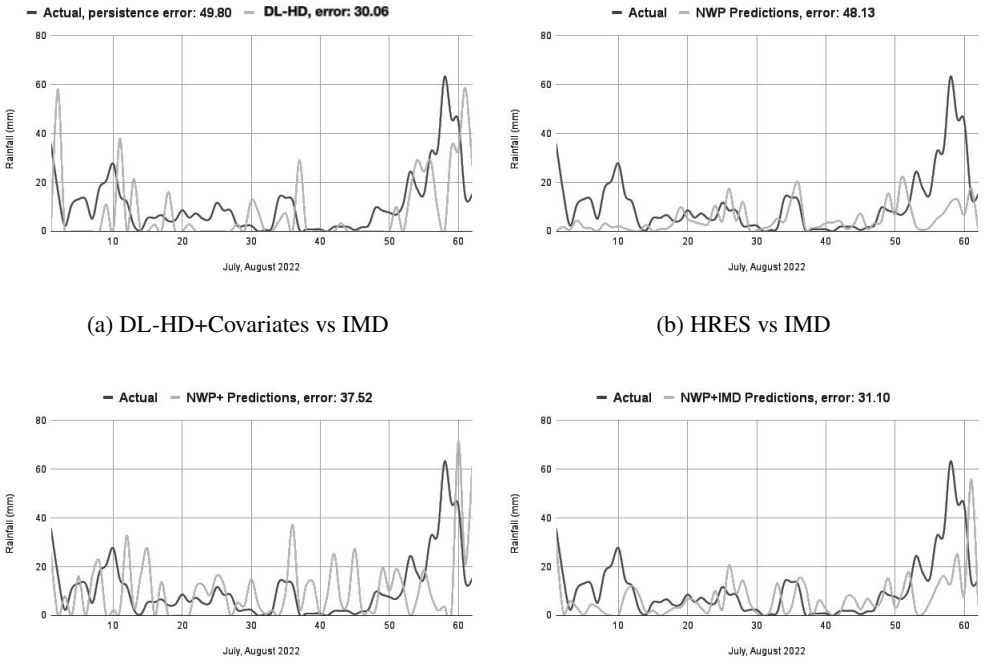


(c) HRES-NWP+ vs IMD



(d) Ensemble vs IMD

FIG. 11: 3-day forecasts for Chennai in July and August 2022



(a) DL-HD+Covariates vs IMD

(b) HRES vs IMD

(c) HRES-NWP+ vs IMD

(d) Ensemble vs IMD

c. Additional performance comparisons

In this section, we compare the performance of the DL-HD + Covariates model and HRES using confusion matrices (Figures 12–15) computed across multiple rainfall thresholds, for the period 2022-2023. Specifically, we analyze confusion matrices in the 0th, 25th, 50th, and 75th rainfall percentiles to capture the behavior of the model over a wide range of rainfall intensities. These matrices provide detailed information on each model's ability to correctly classify rainfall occurrences at varying thresholds.

To quantify classification performance, we report standard metrics derived from the confusion matrix: **Probability of Detection (POD)**, **False Alarm Ratio (FAR)**, **Probability of False Detection (POFD)**, and the **Critical Success Index (CSI)**. TP and FP denote true and false positives, respectively, and TN and FN denote true and false negatives. The POD measures the fraction of actual rainfall events that were correctly predicted as rain, and is computed as $\text{POD} = \frac{\text{TP}}{\text{TP+FN}}$, with higher values indicating better sensitivity to rainfall occurrences. The FAR quantifies the proportion of predicted rainfall events that did not actually occur, and is given by $\text{FAR} = \frac{\text{FP}}{\text{TP+FP}}$; a lower

FAR implies improved precision by reducing the number of false alarms. The POFD captures the fraction of actual dry days that were incorrectly classified as rainy, calculated as POFD = $\frac{\text{FP}}{\text{FP+TN}}$, and is especially important for operational relevance, as a low POFD reduces unnecessary alerts. Finally, the CSI reflects the overall accuracy of rainfall predictions, penalizing both missed events and false alarms. It is defined as CSI = $\frac{\text{TP}}{\text{TP+FP+FN}}$, with higher values indicating more skillful and balanced classification performance.

In addition to classification skill, we also report the **Correlation Coefficient (CC)** between the predicted rainfall and the IMD ground truth across all grid points and time steps in Table 7. It measures the linear relationship between the predicted and observed rainfall values, and is defined as: $CC = \frac{\sum_i (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_i (P_i - \bar{P})^2}\sqrt{\sum_i (O_i - \bar{O})^2}}$, where P_i and O_i denote the predicted and observed rainfall at index i, and $\bar{P}$ and $\bar{O}$ represent their respective means. CC values closer to 1 indicate stronger positive correlation, i.e., better agreement between the predicted and observed rainfall.

Our results, again for the period 2022-2023 (Table 6) demonstrate that DL-HD + Covariates consistently outperforms HRES at all examined rainfall percentile thresholds, both in confusion matrix statistics and derived skill scores. DL-HD+Covariates shows higher POD across all thresholds, reflecting better ability to detect rainfall events. The DL-HD+Covariates model also achieves lower FAR, indicating greater reliability in rain predictions. It also more effectively avoids false detection of rain during dry periods. Finally, higher CSI values demonstrate better overall classification performance when accounting for hits, misses, and false alarms.

DL-HD+Covariates	**Actual >0 mm**	**Actual ≤0 mm**
Predicted >0 mm	136,394 (TP)	15,154 (FP)
Predicted ≤0 mm	15,154 (FN)	1,348,776 (TN)

HRES	**Actual >0 mm**	**Actual ≤0 mm**
Predicted >0 mm	128,500 (TP)	18,200 (FP)
Predicted ≤0 mm	22,100 (FN)	1,355,000 (TN)

FIG. 12: Threshold = 0th Percentile (0 mm)

DL-HD+Covariates	**Actual >3.5 mm**	**Actual ≤3.5 mm**
Predicted >3.5 mm	21,170 (TP)	6,109 (FP)
Predicted ≤3.5 mm	6,109 (FN)	1,475,096 (TN)

HRES	**Actual >3.5 mm**	**Actual ≤3.5 mm**
Predicted >3.5 mm	19,800 (TP)	7,400 (FP)
Predicted ≤3.5 mm	8,200 (FN)	1,480,500 (TN)

FIG. 13: Threshold = 25th Percentile (3.5 mm)

DL-HD+Covariates	Actual >14 mm	Actual ≤14 mm
Predicted >14 mm	12,500 (TP)	3,200 (FP)
Predicted ≤14 mm	4,800 (FN)	1,512,000 (TN)

HRES	Actual >14 mm	Actual ≤14 mm
Predicted >14 mm	11,300 (TP)	4,100 (FP)
Predicted ≤14 mm	5,600 (FN)	1,518,000 (TN)

FIG. 14: Threshold = 50th Percentile (14 mm)

DL-HD+Covariates	Actual >26 mm	Actual ≤26 mm
Predicted >26 mm	8,400 (TP)	1,900 (FP)
Predicted ≤26 mm	2,300 (FN)	1,530,000 (TN)

HRES	Actual >26 mm	Actual ≤26 mm
Predicted >26 mm	7,800 (TP)	2,500 (FP)
Predicted ≤26 mm	3,100 (FN)	1,538,000 (TN)

FIG. 15: Threshold = 75th Percentile (26 mm)

TABLE 6: Comparison of classification metrics at multiple rainfall thresholds for DL-HD+Covariates and HRES.

Threshold	Model	POD	FAR	POFD	CSI
0 mm	DL-HD+Covariates	0.900	0.100	0.011	0.818
	HRES	0.853	0.124	0.013	0.761
3.5 mm	DL-HD+Covariates	0.776	0.224	0.004	0.634
	HRES	0.707	0.272	0.005	0.559
14 mm	DL-HD+Covariates	0.723	0.204	0.002	0.610
	HRES	0.669	0.266	0.003	0.538
26 mm	DL-HD+Covariates	0.785	0.184	0.001	0.667
	HRES	0.716	0.243	0.002	0.582

TABLE 7: Correlation coefficient (CC) of predicted rainfall with IMD ground truth for the period 2022–2023.

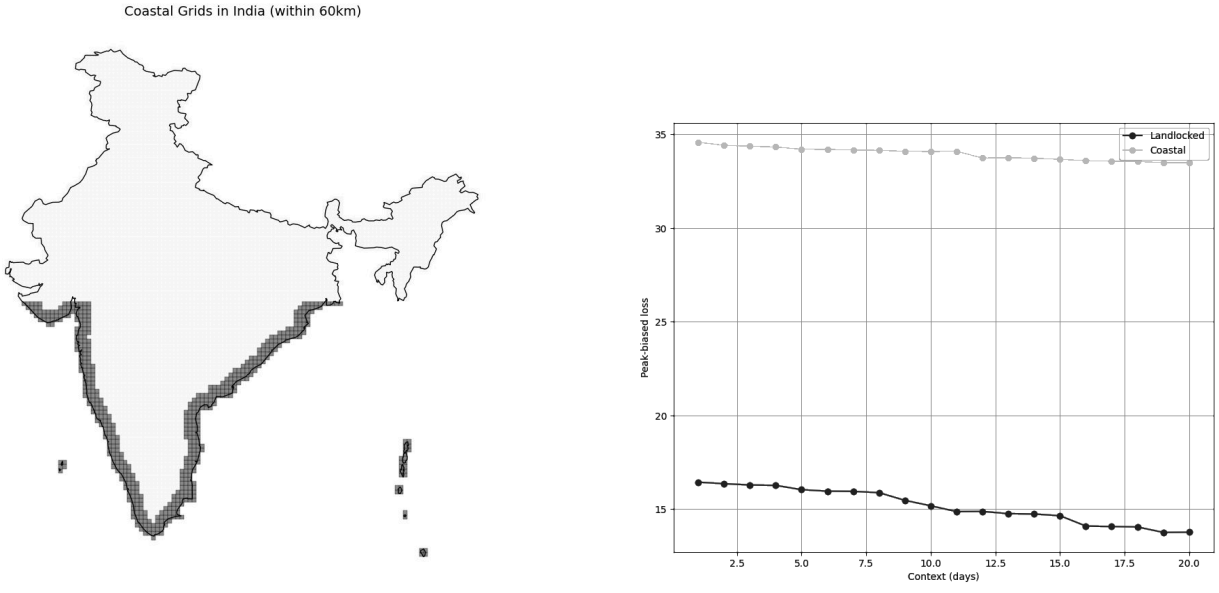
Model	Correlation Coefficient (CC)
DL-HD + Covariates	**0.82**
DL-HD	0.75
HRES	0.69
NWP+	0.62
Ensemble	0.81
Persistence	0.49

d. Spatio-temporal information in rainfall observations across India

To examine the impact of historical context on forecast accuracy, we conducted experiments using input lags ranging from 3 to 20 days. We observe a pattern of small but consistently decreasing errors with increasing context length, indicating the presence of long-term memory in the data. To

further analyze regional variation, we compare the aggregated performance across landlocked and coastal regions in Figure 16b. While increased historical context improves forecast accuracy in both regions, the gains are more pronounced in landlocked areas. This difference may be attributed to the availability of richer surrounding data in landlocked regions, whereas coastal areas are adjacent to oceanic regions where IMD precipitation data is unavailable. Incorporating oceanic rainfall data could potentially enhance forecast performance in coastal zones.

FIG. 16: Comparison of average peak biased loss ($mm^{1.5} + mm$) for coastal vs landlocked regions. In (a) the shaded region represents the grids spanning up to 60km from the coastline. (b) compares the error reduction with context for the different regions.



(a) (b)

e. Spatial performance and consistency

To assess spatial and annual consistency, we compute the number of grid points across India where the DL-HD+Covariates model produced lower daily forecast errors than the NWP baseline during the JJAS monsoon months for each year from 2017 to 2022. For each year, we calculate a *win rate* by identifying, at each grid point, whether the DL-HD+Covariates model had a lower mean daily error compared to NWP. The total number of such grid points is then plotted annually.

Figure 17 illustrates this win count for each monsoon season. We observe that the DL-HD+Covariates model wins in a majority of grid points across the country in every year, demonstrating its spatial robustness and consistent outperformance of NWP.

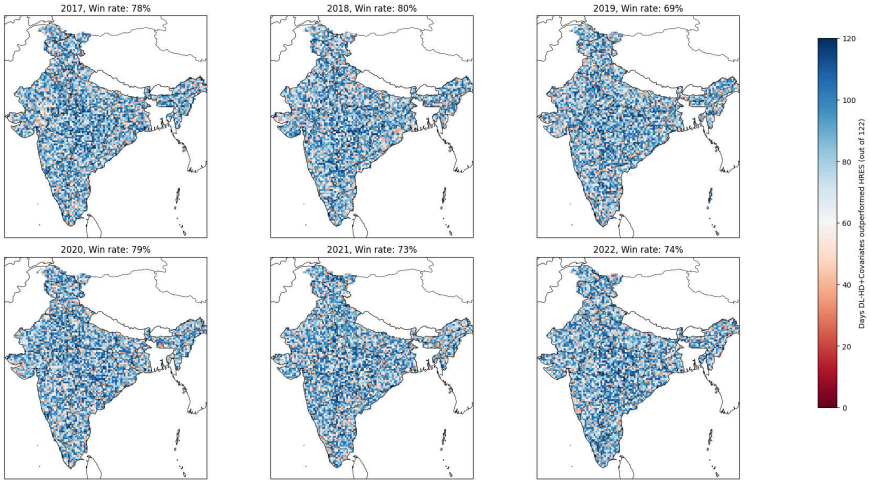


FIG. 17: Annual count of grid points where the DL model produces lower daily forecast errors than NWP during JJAS from 2017 to 2022

4. Discussion and Conclusion

In this study, we demonstrated that deep learning models can consistently outperform traditional NWP systems, specifically the HRES model, in forecasting monsoon rainfall over India, across spatial scales, lead times, and rainfall intensities. The DL-HD+Covariates model achieves a peak-biased loss of **18.24** and an MSE of **268.59**, surpassing HRES-NWP's peak-biased loss of **22.25** (+22% higher) and MSE of **356.97** (+32.9% higher). For a lead time of 3 days, the gap widens: DL's MSE of **2,878.52** contrasts sharply with HRES-NWP's **4,486.25** (+55.85% higher). These margins are consistent nationwide, and across key cities in India. The strong performance of deep learning (DL) models in rainfall forecasting can be explained by several physical factors. Traditional NWP models use physical equations to simulate the atmosphere. However, many small-scale processes, such as convection and cloud formation, occur at scales too fine to be directly resolved, so they are handled using simplified parameterizations. These approximations can introduce significant errors, especially over the Indian subcontinent, where weather patterns are highly complex and variable (Randall et al. 2003; Stensrud 2007). DL models, in contrast, do not rely on such physical approximations. Instead, they learn directly from historical data, identifying patterns that improve

forecasts without needing to explicitly simulate physical processes. Another advantage of DL models is that they do not require precise initial conditions. NWP forecasts are highly sensitive to their initial inputs, and even small errors, particularly in regions with limited observations, such as oceans or mountainous areas, can grow quickly and reduce forecast accuracy (Lorenz 1963; Kalnay 2003). DL models avoid this issue by using past sequences of observations to make predictions, making them more robust in data-sparse settings. Finally, the atmosphere behaves in a non-linear and sometimes chaotic way, which is difficult for traditional models to capture. DL architectures, especially those designed for spatiotemporal data are well-suited to handle this complexity. They can learn to represent these chaotic patterns, leading to more accurate and stable forecasts in challenging conditions.

a. Possible future research directions

Recent works (Kurz et al. 2024; Zheng et al. 2025) have shown that incorporating physics into machine learning-based weather models can lead to models that are cheaper to train, more accurate, and faster to run. An exciting direction to extend the work of this paper is to develop such models for India using IMD data.

A key direction suggested by our analysis is that more and diverse data relevant to monsoon prediction, including radar- and satellite-based data (see Espeholt et al. (2022)), combined with carefully selected neural network architecture, are likely to substantially improve existing NWP forecasts. However, radar-based data for public use are limited in India, and satellite data tend to be bulky and require substantially more computation. We leave incorporating radar-based data and satellite data into deep learning-based predictions using IMD data to a more expansive future project.

Acknowledgments. We would like to thank Sandip Trivedi, TIFR for motivating us to work on weather modelling and for related discussions. We thank Debasis Sengupta, ICTS, Rama Govindrajan, ICTS and Partha Mukhopadhyay, IITM for many helpful discussions. We also acknowledge the use of the High-Performance Computing (HPC) facility at TIFR for training large scale models required for this research.

Data availability statement. The data utilized in this paper are publicly available and have been cited in Section 2a.

APPENDIX

a. Sensitivity Analysis

To validate the design of our peak-biased loss function, we conducted a sensitivity analysis across different values of the parameters α and β. Tables A1-A4 report the model performance for different values of α, and $\beta = 1$, across 25-percentile rainfall thresholds.
The results indicate that $\alpha = 1.5$ and $\beta = 1.0$ provide the best trade-off. This analysis supports the empirical choice of asymmetry used in the main experiments and demonstrates the robustness of the proposed loss formulation.

TABLE A1: Confusion matrices for rainfall threshold ≥ 0 mm (0th percentile) with different α values ($\beta = 1$)

$\alpha = 1.0, \beta = 1$	**Actual** ≥ 0 **mm**	**Actual** $= 0$ **mm**
Predicted ≥ 0 **mm**	132,450 (TP)	17,820 (FP)
Predicted $= 0$ **mm**	19,098 (FN)	1,341,110 (TN)
$\alpha = 1.25, \beta = 1$	**Actual** ≥ 0 **mm**	**Actual** $= 0$ **mm**
Predicted ≥ 0 **mm**	134,625 (TP)	16,438 (FP)
Predicted $= 0$ **mm**	16,923 (FN)	1,342,492 (TN)
$\alpha = 1.5, \beta = 1$	**Actual** ≥ 0 **mm**	**Actual** $= 0$ **mm**
Predicted ≥ 0 **mm**	136,394 (TP)	15,154 (FP)
Predicted $= 0$ **mm**	15,154 (FN)	1,343,776 (TN)
$\alpha = 1.75, \beta = 1$	**Actual** ≥ 0 **mm**	**Actual** $= 0$ **mm**
Predicted ≥ 0 **mm**	135,840 (TP)	16,225 (FP)
Predicted $= 0$ **mm**	15,708 (FN)	1,342,705 (TN)

TABLE A2: Confusion matrices for rainfall threshold ≥ 3.5 mm (25th percentile) with different α values ($\beta = 1$)

$\alpha = 1.0, \beta = 1$	**Actual** ≥ 3.5 **mm**	**Actual** < 3.5 **mm**
Predicted ≥ 3.5 **mm**	19,875 (TP)	7,230 (FP)
Predicted < 3.5 **mm**	7,404 (FN)	1,475,969 (TN)
$\alpha = 1.25, \beta = 1$	**Actual** ≥ 3.5 **mm**	**Actual** < 3.5 **mm**
Predicted ≥ 3.5 **mm**	20,635 (TP)	6,750 (FP)
Predicted < 3.5 **mm**	6,644 (FN)	1,476,449 (TN)
$\alpha = 1.5, \beta = 1$	**Actual** ≥ 3.5 **mm**	**Actual** < 3.5 **mm**
Predicted ≥ 3.5 **mm**	21,170 (TP)	6,109 (FP)
Predicted < 3.5 **mm**	6,109 (FN)	1,477,090 (TN)
$\alpha = 1.75, \beta = 1$	**Actual** ≥ 3.5 **mm**	**Actual** < 3.5 **mm**
Predicted ≥ 3.5 **mm**	20,843 (TP)	6,875 (FP)
Predicted < 3.5 **mm**	6,436 (FN)	1,476,324 (TN)

TABLE A3: Confusion matrices for rainfall threshold ≥ 14 mm (50th percentile) with different α values ($\beta = 1$)

$\alpha = 1.0, \beta = 1$	**Actual** ≥ 14 **mm**	**Actual** < 14 **mm**
Predicted ≥ 14 **mm**	5,892 (TP)	3,145 (FP)
Predicted < 14 **mm**	3,201 (FN)	1,498,240 (TN)
$\alpha = 1.25, \beta = 1$	**Actual** ≥ 14 **mm**	**Actual** < 14 **mm**
Predicted ≥ 14 **mm**	6,348 (TP)	2,752 (FP)
Predicted < 14 **mm**	2,745 (FN)	1,498,633 (TN)
$\alpha = 1.5, \beta = 1$	**Actual** ≥ 14 **mm**	**Actual** < 14 **mm**
Predicted ≥ 14 **mm**	6,735 (TP)	2,358 (FP)
Predicted < 14 **mm**	2,358 (FN)	1,499,027 (TN)
$\alpha = 1.75, \beta = 1$	**Actual** ≥ 14 **mm**	**Actual** < 14 **mm**
Predicted ≥ 14 **mm**	6,532 (TP)	2,634 (FP)
Predicted < 14 **mm**	2,561 (FN)	1,498,751 (TN)

b. Neural Network Hyperparameters

This section describes the design and training setup of the two main models used in this study: a transformer-based model (Autoformer) and the simpler neural networks used for NWP+ and Ensemble models. Our choices were driven by strong empirical performance and computational efficiency.

TABLE A4: Confusion matrices for rainfall threshold ≥ 26 mm (75th percentile) with different α values ($\beta = 1$)

$\alpha = 1.0, \beta = 1$	**Actual** ≥ 26 **mm**	**Actual** < 26 **mm**
Predicted ≥ 26 **mm**	3,106 (TP)	1,254 (FP)
Predicted < 26 **mm**	1,440 (FN)	1,504,678 (TN)
$\alpha = 1.25, \beta = 1$	**Actual** ≥ 26 **mm**	**Actual** < 26 **mm**
Predicted ≥ 26 **mm**	3,384 (TP)	1,097 (FP)
Predicted < 26 **mm**	1,162 (FN)	1,504,835 (TN)
$\alpha = 1.5, \beta = 1$	**Actual** ≥ 26 **mm**	**Actual** < 26 **mm**
Predicted ≥ 26 **mm**	3,728 (TP)	818 (FP)
Predicted < 26 **mm**	818 (FN)	1,505,114 (TN)
$\alpha = 1.75, \beta = 1$	**Actual** ≥ 26 **mm**	**Actual** < 26 **mm**
Predicted ≥ 26 **mm**	3,525 (TP)	936 (FP)
Predicted < 26 **mm**	1,021 (FN)	1,504,996 (TN)

1) AUTOFORMER CONFIGURATION

We use the Autoformer model (Wu et al. 2021), which is especially well-suited for making predictions over long time periods. It works by breaking down weather signals into different components and learning patterns over time using attention mechanisms. Key settings include:

- **Transformer Layers:** Two layers in both the encoder and decoder, each using 8 attention heads. This allows the model to capture complex rainfall patterns across different time scales without becoming too heavy.
- **Embedding Size** ($d_{\text{model}} = 512$)**:** This size balances detail and efficiency, for the 20-day input of 9 weather variables and helping the model learn interactions between them.
- **Feedforward Dimension** ($d_{\text{ff}} = 2048$)**:** A larger internal layer helps the model learn complex relationships in atmospheric data.
- **Decomposition Kernel (Size 25):** This setting helps the model separate short-term fluctuations (like storms) from longer seasonal trends, which is important for understanding monsoon behavior.

2) NWP+ CONFIGURATION

NWP+ is a basic multilayer perceptron (MLP) that uses weather data from a central grid point, along with the 4 neighboring grid cells for better spatial context.

- **Architecture:** A simple neural network with three layers, using 32, 16, and 1 neurons. This structure is compact because of limited data.
- **Activations:** ReLU is used in the hidden layers to handle spikes in rainfall, while the final output uses a sigmoid to keep predictions in a reasonable range after normalization.
- **Spatial Context:** Including nearby grid points helps improve the model's accuracy, especially for short-term predictions. Adding more distant points didn't help much, so we kept the neighborhood small.

3) Training Strategy

- **Batch Size:** We use a larger batch (64) for the transformer to fully utilize GPU resources, and a smaller one (24) for the MLP due to memory limits.
- **Learning Rate:** The transformer uses a smooth cosine decay schedule, while the MLP uses a step-wise decrease every 50 epochs to help stabilize learning.
- **Regularization:** To avoid overfitting and training issues, we use weight decay and clip gradients that grow too large.
- **Early Stopping:** We monitor performance on a validation set and stop training if there's no improvement after 10–20 epochs. We also limit the training to a maximum of 100–300 epochs.
- **Mixed Precision:** We train the Autoformer using half-precision (FP16), which speeds things up and reduces memory usage.

4) Input and Output Processing

- **Normalization:** Input features are standardized using data from 2017–2021. This keeps the data consistent while still allowing year-to-year differences to be learned.
- **Prediction Target:** The models forecast daily rainfall for the next 1 to 3 days.

References

Chen, L., B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang, 2023: Machine learning methods in weather and climate applications: A survey. *Applied Sciences*, **13 (21)**, 12 019.

Copernicus Climate Change Service, 2019: Era5-land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), URL https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac, https://doi.org/10.24381/CDS.E2161BAC.

ECMWF, 2021: Open data. ECMWF, URL https://www.ecmwf.int/en/forecasts/datasets/open-data, https://doi.org/10.21957/OPEN-DATA.

Espeholt, L., and Coauthors, 2022: Deep learning for twelve hour precipitation forecasts. *Nature communications*, **13 (1)**, 1–10.

Goswami, B. N., and P. K. Xavier, 2003: Potential predictability and extended range prediction of indian summer monsoon breaks. *Geophysical Research Letters*, **30 (18)**.

Jose, D. M., A. M. Vincent, and G. S. Dwarakish, 2022: Improving multiple model ensemble predictions of daily precipitation and temperature through machine learning techniques. *Scientific Reports*, **12 (1)**, 4678.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Kishore, P., S. Jyothi, G. Basha, S. Rao, M. Rajeevan, I. Velicogna, and T. C. Sutterley, 2016: Precipitation climatology over india: validation with observations and reanalysis datasets and spatial trends. *Climate dynamics*, **46**, 541–556.

Kumar, B., and Coauthors, 2021: Deep learning based forecasting of indian summer monsoon rainfall. *algorithms*, **9 (11)**, 13.

Kumar, B., and Coauthors, 2022: Deep learning based short-range forecasting of indian summer monsoon rainfall using earth observation and ground station datasets. *Geocarto International*, **37 (27)**, 17 994–18 021.

Kurz, T., A. Dombrowski, D. Nino, and et al., 2024: Climode: Climate and weather forecasting with physics-informed neural odes. *International Conference on Learning Representations (ICLR)*, URL https://openreview.net/forum?id=xuY33XhEGR.

Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382 (6677)**, 1416–1421.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, **20 (2)**, 130–141.

Miao, Q., B. Pan, H. Wang, K. Hsu, and S. Sorooshian, 2019: Improving monsoon precipitation prediction using combined convolutional and long short term memory neural network. *Water*, **11 (5)**, 977.

Nguyen, T., J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, 2023: Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*.

Pai, D., M. Rajeevan, O. Sreejith, B. Mukhopadhyay, and N. Satbha, 2014: Development of a new high spatial resolution (0.25× 0.25) long period (1901-2010) daily gridded rainfall data set over india and its comparison with existing data sets over the region. *Mausam*, **65 (1)**, 1–18.

Prasad, K., and S. Singh, 1988: Large-scale features of the indian summer monsoon rainfall and their association with some oceanic and atmospheric variables. *Advances in atmospheric sciences*, **5 (4)**, 499–513.

Praveen, B., S. Talukdar, Shahfahad, S. Mahato, J. Mondal, P. Sharma, A. R. M. T. Islam, and A. Rahman, 2020: Analyzing trend and forecasting of rainfall changes in india using non-parametrical and machine learning approaches. *Scientific reports*, **10 (1)**, 10 342.

Rajeevan, M., P. Guhathakurta, and V. Thapliyal, 2000: New models for long range forecasts of summer monsoon rainfall over north west and peninsular india. *Meteorology and Atmospheric Physics*, **73**, 211–225.

Rajeevan, M., D. Pai, R. Anil Kumar, and B. Lal, 2007: New statistical models for long-range forecasting of southwest monsoon rainfall over india. *Climate Dynamics*, **28**, 813–828.

Rajeevan, M. N., 2023: T.N. rains— Predicting extreme rainfall events using probabilistic forecasts. https://www.thehindu.com/sci-tech/science/predicting-extreme-rainfall-events-using-probabilistic-forecasts/article67666205.ece#.

Randall, D., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, **84 (11)**, 1547–1564.

Shi, X., Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in neural information processing systems*, **30**.

Stensrud, D. J., 2007: *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press.

Wu, H., J. Xu, J. Wang, and M. Long, 2021: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, **34**, 22 419–22 430.

Xu, W., K. Chen, T. Han, H. Chen, W. Ouyang, and L. Bai, 2024: Extremecast: Boosting extreme value prediction for global weather forecast. *arXiv preprint arXiv:2402.01295*.

Zheng, J., Q. Ling, and Y. Feng, 2025: Physics-assisted and topology-informed deep learning for weather prediction. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*, URL https://arxiv.org/abs/2505.04918, accepted.