Deep learning for short-range monsoon rainfall forecast using ground truth rainfall data

Aastha Jain
Columbia University

Jatin Batra
Tata Institute of Fundamental Research

Apoorva Narula
Georgia Institute of Technology

Madhavan Nair Rajeevan
Atria University

Sandeep Juneja
Ashoka University

Abstract

The Indian summer monsoon is a highly complex and critical weather system that directly affects the livelihoods of about a billion and a half people across the Indian subcontinent. Accurate short-term forecasting remains a major scientific challenge due to the monsoon's 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. Forecasts are benchmarked against ECMWF's High-Resolution Ensemble System (HRES), widely regarded as the most accurate numerical weather predictor. **Our results provide the first evidence that a machine learning model can outperform HRES for short-term monsoon forecasts.** Specifically, compared to our model, forecasts from HRES model have about 22% higher error, for a single day prediction, and over 27% higher error, for a three day prediction. Such enhanced forecast accuracy translates into tangible climate adaptation benefits, enabling earlier flood warnings and helping farmers protect crops from unexpected downpours. We also find that incorporating historical data up to 20 days prior reduces forecast error, particularly in landlocked regions.

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. This need is especially critical as climate change amplifies monsoon variability and extreme precipitation Turner and Annamalai [2012], Roxy et al. [2017], Clemens et al. [2021]. Without accurate forecasts, sudden downpours can inundate cities and catch farmers off-guard causing devastating floods and crop losses, which underscores the urgency of improved predictive tools. It is well known that numerical weather prediction (NWP) does not perform well in the prediction of rainfall for India Rajeevan [2023].

In this paper, we consider $0.25^\circ \times 0.25^\circ$ daily gridded precipitation data from India Meteorological Department (IMD) Pai et al. [2014] available from 1901–2023 (each degree roughly corresponds to 111 km). 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 including HRES-IFS (High Resolution Integrated Forecast System) from ECMWF (European Centre for Medium-Range Weather Forecasts)

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2025.

ECMWF [2021]. HRES is widely regarded as the top operational weather forecasting system in the world Lam et al. [2023].

Prior work on ML-based monsoon forecasting can be broadly grouped into two categories. First, for medium- to long-range predictions, data-driven models using IMD observations have demonstrated improved skill over traditional models [Bach et al., 2024, Narang et al., 2024]. Second, for short-range weather prediction, advanced ML models have begun to outperform leading NWP systems in global settings (e.g., GraphCast, ClimaX, Pangu-Weather, Aardvark)[Lam et al., 2023, Nguyen et al., 2023, Bi et al., 2023, Allen et al., 2025]. However, these short-range successes relied on reanalysis Copernicus Climate Change Service [2019] data and were not validated against ground-truth Indian monsoon observations.

Our results. We compare the deep learning-based forecasts generated by autoformers Wu et al. [2021] using historical rainfall data from IMD (which is shown to be better representative of the ground truth Kishore et al. [2016]), with the NWP (HRES) forecasts for the four monsoon months of June, July, August and September (JJAS). We show the first results that an ML model trained on ground truth monsoon rainfall can surpass a state-of-the-art NWP (HRES) for short-range forecasting (1-3 days) across spatial scales, lead times and rainfall intensities.

2 Data

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 n=12,422 grids.
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 10 m, temperature, soil moisture, cloud cover, vorticity at 850 hPa, humidity, and divergence at 700 hPa.The data is available from 1950 onwards.
3. **NWP forecasts:** HRES-NWP daily forecasts are obtained from ECMWF [2021] for all years 2011 onwards, at a resolution of $0.25°$, for both 1 and 3 days into the future.

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. From the training period, 10% of the data is further held out as a validation set for hyperparameter tuning.

3 Methodology

Our forecasts are generated using different models for lead times of 1 and 3 days as follows:

1. **DL-HD (Deep Learning-Historical Data):** 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). We train the Autoformer model (see Section A.2) using data from 1901 to 2011 and generate test forecasts for the years 2012 to 2023.
2. **DL-HD + Covariates:** This is an extension of the above model where we also use past 3 days of reanalysis data. As stated earlier, the additional covariates include wind speed, temperature, soil moisture, cloud cover, vorticity, humidity, and divergence.
3. **NWP+:** We combine the NWP forecasts at the target grid and its 4 neighboring grids using a deep neural network, which is trained to minimize the error between the forecast and the IMD ground truth for that grid. The resulting forecast is called the NWP+ prediction.

For a thorough comparison, we also train *Ensemble* of DL-HD+Covariates and HRES (using the target grid and its four neighbors), *AR(1) model* (using only the previous day's rainfall and *AR(5) model +Spatial Grids* that uses the previous 5 days of rainfall together with its 4 neighbors.
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 (see e.g. Figure 4). To address this, we propose the following *peak-biased loss function*:

Table 1: Loss India 1-Day

Model	**Peak-biased Loss** ($\mathrm{mm}^{1.5} + \mathrm{mm}$)	**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

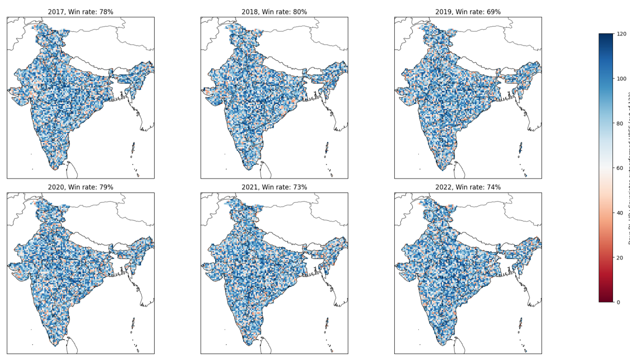


Figure 1: Annual count of grid points where the DL model produces lower daily forecast errors than NWP during JJAS from 2017 to 2022

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

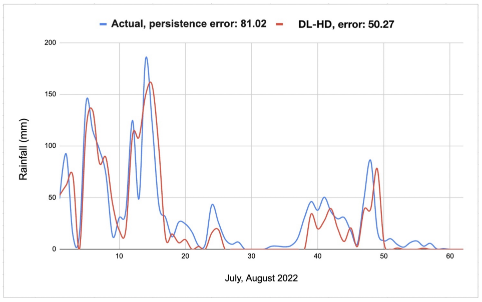
where r_t and $\hat{r}_t$ denote the observed and predicted rainfall at time t, respectively, and $\alpha > \beta$ ensures that *underestimation is penalized more heavily than overestimation* (for short-range prediction, underestimating extreme rainfall has greater human cost than overestimating it). We use $\alpha = 1.5$, $\beta = 1.0$, selected empirically to optimize standard skill metrics of rainfall forecasting (see Appendix A.1).

4 Results

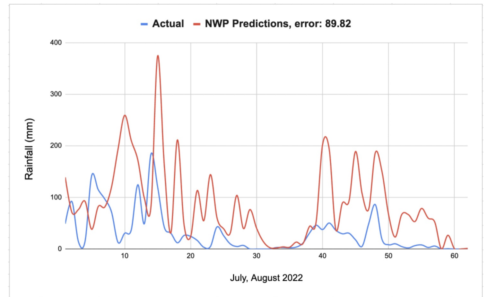
The predictions based on the models specified in Section 3 are compared with the ground truth daily rainfall data from IMD, along with climatological baselines, the **Rolling Mean (20 days), Climatological Mean, Persistence**. We show results both in a spatiotemporal average sense (Figure 1, Table 1 and Table 2), and specifically for Mumbai [1] for chosen dates (Figure 2). **DL-HD + Covariates** outperforms all other models in accuracy, and also in several other performance metrics (Subsection A.1) across rainfall intensities. We also find that incorporating historical data up to 20 days prior reduces forecast error, particularly in landlocked regions (Figure 3), adding to recent evidence of monsoon memory Mitra et al. [2018a,b].

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. By providing earlier and more reliable warnings of heavy rainfall, our approach can empower diverse communities and stakeholders to mitigate losses due to variations in monsoon rainfall patterns because of climate change. Recent studies Kurz et al. [2024], Zheng et al. [2025] have shown that incorporating physical knowledge into machine learning-based weather models can further improve accuracy and efficiency.

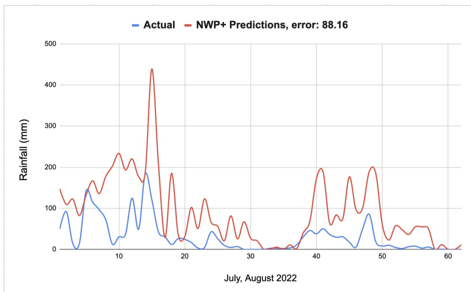
[1] selected due to being India's commercial capital and also its meteorological importance since it receives heavy rainfall



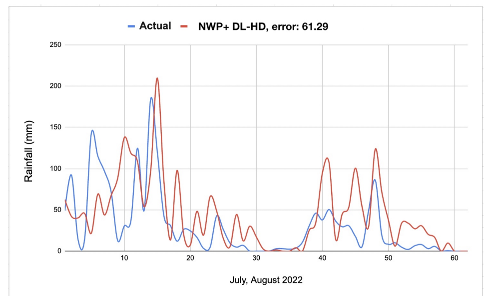
(a) DL-HD+Cov vs IMD



(b) HRES vs IMD



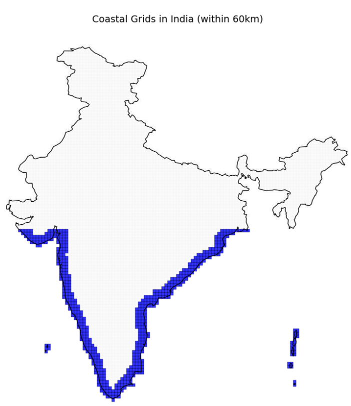
(c) HRES-NWP+ vs IMD



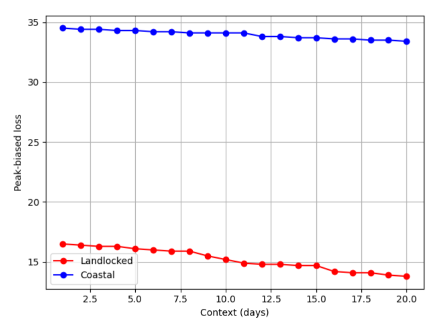
(d) Ensemble vs IMD

Figure 2: 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 shows significant improvement over NWP alone and captures most of the high rainfall events during this period.

Figure 3: 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)

In the future, developing such physics-informed deep learning models using IMD data is a promising direction for enhancing monsoon rainfall predictability.

References

Andrew G. Turner and H. Annamalai. Climate change and the south asian summer monsoon. *Nature Climate Change*, 2:587–595, 2012. doi: 10.1038/nclimate1495.

Mathew Koll Roxy, Subimal Ghosh, Amey Pathak, Athulya Radhakrishnan, Milind Mujumdar, Raghu Murtugudde, Pascal Terray, and M. Rajeevan. A threefold rise in widespread extreme rain events over central india. *Nature Communications*, 8:708, 2017. doi: 10.1038/s41467-017-00744-9.

Steven C. Clemens, Masanobu Yamamoto, Kaustubh Thirumalai, Liviu Giosan, Julie N. Richey, Katrina Nilsson-Kerr, Yair Rosenthal, Pallavi Anand, and Sarah M. McGrath. Remote and local drivers of pleistocene south asian summer monsoon precipitation: A test for future predictions. *Science Advances*, 7(23):eabg3848, 2021. doi: 10.1126/sciadv.abg3848.

Madhavan Nair Rajeevan. 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#`, 2023.

DS Pai, M Rajeevan, OP Sreejith, B Mukhopadhyay, and NS Satbha. Development of a new high spatial resolution ($0.25\times$ 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, 2014.

ECMWF. Open data, 2021. URL `https://www.ecmwf.int/en/forecasts/datasets/open-data`.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023.

Eviatar Bach, V. Krishnamurthy, Safa Mote, Jagadish Shukla, A. Surjalal Sharma, Eugenia Kalnay, and Michael Ghil. Improved subseasonal prediction of south asian monsoon rainfall using data-driven forecasts of oscillatory modes. *Proceedings of the National Academy of Sciences*, 121(15): e2312573121, 2024. doi: 10.1073/pnas.2312573121.

Udit Narang, Kushal Juneja, Pankaj Upadhyayay, Tanmoy Chakraborty Popat Salunke, Swadhin K. Behera, Saroj K. Mishra, and Akhil Dev Suresh. Artificial intelligence predicts normal summer monsoon rainfall for india in 2023. *Scientific Reports*, 14:1495, 2024. doi: 10.1038/s41598-023-44284-3.

Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*, 2023.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619:533–538, 2023. doi: 10.1038/s41586-023-06185-3.

Anna Allen, Stratis Markou, Will Tebbutt, James Requeima, Wessel P. Bruinsma, Tom R. Andersson, Michael Herzog, Nicholas D. Lane, Matthew Chantry, J. Scott Hosking, and Richard E. Turner. End-to-end data-driven weather prediction. *Nature*, 641:1172–1179, 2025. doi: 10.1038/s41586-025-08897-0.

Copernicus Climate Change Service. Era5-land hourly data from 1950 to present, 2019. URL `https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac`.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34:22419–22430, 2021.

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

Adway Mitra, Amit Apte, Rama Govindarajan, Vishal Vasan, and Sreekar Vadlamani. A discrete view of the indian monsoon to identify spatial patterns of rainfall. *Dynamics and Statistics of the Climate System*, 3(1):dzy009, 2018a.

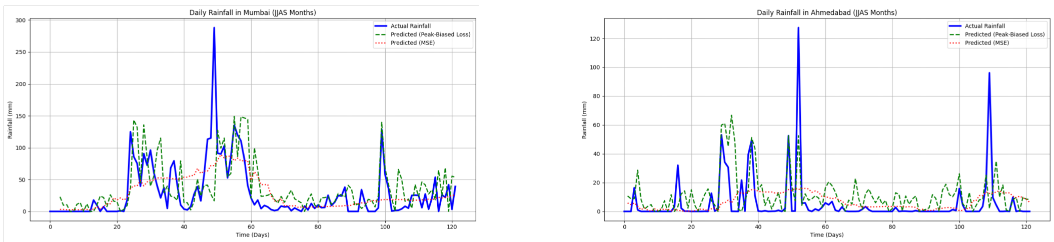
Adway Mitra, Amit Apte, Rama Govindarajan, Vishal Vasan, and Sreekar Vadlamani. Spatio-temporal patterns of daily indian summer monsoon rainfall. *Dynamics and Statistics of the Climate System*, 3(1):dzy010, 2018b.

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

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

A Appendix

In the Appendix, we give the following supporting material: justification for peak-biased loss and performance comparisons for 3 day forecasts in Figure 4 and Table 2, respectively, additional performance comparisons using standard skill metrics in rainfall forecasting in Subsection A.1, and details of Autoformer training in Subsection A.2.



(a) (b)

Figure 4: 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.

Table 2: Comparison of 3-day ahead precipitation forecasting performance over India at 0.25° resolution.

Model	**Peak-biased Loss** ($\mathrm{mm}^{1.5} + \mathrm{mm}$)	**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

A.1 Additional performance comparisons

In this section, we compare the performance of the DL-HD + Covariates model and HRES using confusion matrices (Figures 5–8) 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 $\mathrm{POD} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$, 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 $\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}}$; 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 $\mathrm{POFD} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$, 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 $\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$, 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 4. 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 3) 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)

Figure 5: 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)

Figure 6: 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)

Figure 7: Threshold = 50th Percentile (14 mm)

A.2 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.

A.2.1 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.

A.2.2 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.

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)

Figure 8: Threshold = 75th Percentile (26 mm)

Table 3: 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 4: 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

- **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.

A.2.3 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.

A.2.4 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.