

Seasonal Forecast of Monsoon Rainfall over Bangladesh using Machine Learning

Apurbo Roy Chowdhury^{1*}, Md. Abdus Samad¹, S M Quamrul Hassan² and M. A. K Mallik²

¹*Department of Applied Mathematics, University of Dhaka, Dhaka, Bangladesh*

²*Bangladesh Meteorological Department, Dhaka, Bangladesh*

*Corresponding author's email: apurboroy73@gmail.com

Abstract

Seasonal forecasting plays a critical role in Bangladesh's socioeconomic development, as the country heavily relies on agriculture and is susceptible to the impacts of climate variability. Traditional forecasting methods often struggle to provide accurate and timely predictions, particularly for the dynamic monsoon season. This study investigates the utilization of machine learning methodologies for seasonal forecasting across Bangladesh, aiming to support climate resilience and facilitate informed decision-making. Drawing upon a comprehensive array of datasets including historical climate records, and atmospheric reanalysis data, machine learning models and ensemble methods are developed to effectively discern and predict the intricate spatiotemporal patterns characterizing seasonal climate variability within the region. Through rigorous analysis and evaluation of forecast outputs, the study demonstrates the efficacy of machine learning approaches in capturing seasonal patterns, predicting extreme weather events, and informing decision-making processes across various sectors, including agriculture, water management, disaster preparedness, and infrastructure development. Encouraging outcomes are observed, during the monsoon season 2023 spanning from June to September (JJAS), wherein the models exhibit commendable performance largely attributable to the pronounced influence of precipitation variables. Future directions surround the fine-tuning of machine learning algorithms to better encapsulate weather dynamics, the enhancement of data accessibility and integrity, and the cultivation of interdisciplinary collaboration to tailor forecasting outputs to sector-specific demands. Through these efforts, machine learning-based seasonal forecasting holds potential to enhance climate resilience and support sustainable development in Bangladesh.

Keywords: Seasonal forecasting, Machine Learning, Ensemble Methods.

1. Introduction

Seasonal forecasting is the prediction or estimation of atmospheric and climate conditions over a specific region for a time period, typically spanning a few months (known as seasonal timescales). However, it is not possible to predict day-to-day weather in detail when the timescale is beyond about a week. It is different from the weather forecast that we use in everyday life, and forecasts for small spatial regions are not available. A seasonal forecast aims to find the long-term average of the weather conditions. It is possible to forecast the general trend of weather in a vast region with the help of seasonal forecasting.

Traditionally, seasonal forecasts have relied on statistical and dynamical methods, utilizing historical climate data, common climate indices such as the El Niño-Southern Oscillation (ENSO) and the Madden-Julian Oscillation (MJO), and sophisticated computer models to predict future weather patterns. The majority of Bangladesh's efforts to predict the seasonal climate have relied on statistical and empirical forecasting techniques, such as regression models of the relationships between rainfall and different predictors like sea-surface temperature (SST; Hossain et al., 2019) or Auto-Regressive Integrated Moving Average (ARIMA) models for temperature and rainfall prediction (Bari et al., 2015; Mahmud et al., 2017; Mohsin et al., 2012). Mahmud et al. (2017) have reported that ARIMA models have been utilized to forecast rainfall with lead times of up to 12 months; nevertheless, Dalale and Singh (1993) have noted that the lack of statistical significance of year-to-year autocorrelation can result in restricted forecasting abilities.

The use of machine learning in the field of seasonal forecasts shows promise in both accuracy and in terms of bridging the shortcomings of the traditional methods (Gibson et al. [4]). Most of the global producing centers (GPCs) are recently training ML models to enhance forecasts. The ML models have lower computational costs as compared to the traditional methods. Random forests, support vector machines (SVM), artificial neural networks (ANN), gradient boosting machines (GBM), extreme gradient boosting (XGBoost), extreme learning machine (ELM), and long short-term memory (LSTM) are some of the existing methods used in machine learning for weather forecasting. Machine learning algorithms, including neural networks, and ensemble methods, have shown

promise in capturing nonlinear relationships and patterns in large and complex datasets, thereby improving forecast skill and reducing uncertainty. We have used Python as the programming language to imply the methods. A relatively new climate forecasting tool, XCast, has made working with gridded datasets in Python much easier in the field of meteorology. Xcast is a python library that is created especially for climate forecasting. It is based on the Xarray library and offers a suite of tools to facilitate the work with gridded climate data. As, Xarray is a well-known library in python, similarity with this makes the learning curve of Xcast much easier.

There are different sources to get observational seasonal data. Observational networks, satellite data, climate model outputs are typical sources of weather dataset. The international research institute for climate and society also known as IRI has a large library of climate dataset easily available for seasonal forecasts. IRI website has different models observational, forecast and hindcast dataset. We have used Climate Hazards Group InfraRed Precipitation with Station data also known as CHRIPS data as observational data for seasonal forecasting. For forecast and hindcast we have used different climate models such as the Climate Forecast System (CFS), the Center for Ocean-Land Atmosphere Studies Rosenstiel School of Marine and Atmospheric Science (Cola RSMAS CCSM4), the Canadian Seasonal to Inter-annual Prediction System (CanSIPS- IC3) etc.

2. Data and Methodology

2.1 Data Used

For the seasonal forecast of monsoon rainfall over Bangladesh using machine learning, multiple sources of climate data were utilized. The primary datasets are

Model Acronym	Institution	Type
CHRIPS	The Climate Hazards Group (CHG) at the University of California, Santa Barbara (UCSB)	Observational
NCEP-CFSv2	NOAA's Centers for Environmental Prediction	Hindcast and Forecast
COLA-RSMAS-CCSM4	The Center for Ocean-Land-Atmosphere Studies	Hindcast and Forecast

2.2 Model Development: Extreme Learning Machine

The primary machine learning model used in this study is the Extreme Learning Machine (ELM), known for its efficiency and effectiveness in handling large datasets with nonlinear patterns. The ELM model was trained using the hindcast data. Hindcast data provide a historical context, allowing the model to learn patterns and relationships between different climatic variables and monsoon rainfall. To assess the performance of the Extreme Learning Machine (ELM) model, various skill metrics were utilized, providing a comprehensive evaluation of forecast accuracy and reliability. These metrics are Pearson correlation, Kling-Gupta Efficiency, Receiver Operating Characteristic etc. This training phase aimed to optimize the model's parameters to accurately capture these patterns. After training and evaluating the model, we ran the ELM model to generate forecasts for the 2023 monsoon rainfall season over Bangladesh. The forecasted results were then validated against the actual observed rainfall data for 2023.

The primary machine learning model used in this study is the Extreme Learning Machine (ELM), known for its efficiency and effectiveness in handling large datasets with nonlinear patterns. Extreme learning machine is a feed-forward neural network which is first introduced by Huang at 2006 (Huang *et al.*). Soon after the method became popular due to its easy to implementation, ability to generalization, and fast learning capabilities. The term 'Extreme' comes from the extremely fast learning speed of ELM. ELM's distinct training strategy sets it apart from conventional feedforward neural networks. The ELM model was trained using the hindcast data. Hindcast data provide a historical context, allowing the model to learn patterns and

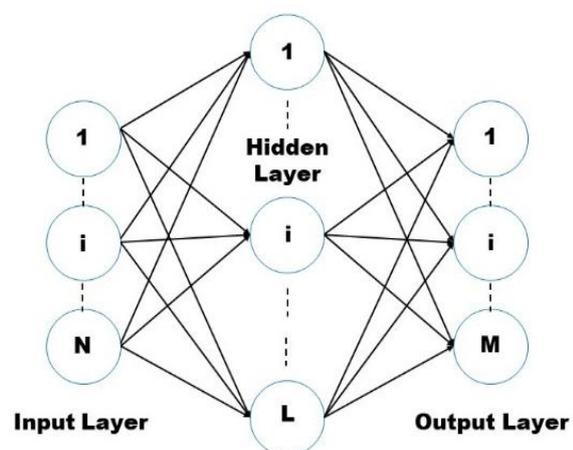


Fig. 1: Extreme Learning Machine

relationships between different climatic variables and monsoon rainfall.

To assess the performance of the Extreme Learning Machine model, various skill metrics were utilized, providing a comprehensive evaluation of forecast accuracy and reliability. These metrics are Pearson Correlation, Kling-Gupta Efficiency (KGE), Receiver Operating Characteristic, Rank Probability Skill Score, etc. This training phase aimed to optimize the model's parameters to accurately capture these patterns. After training and evaluating the model, we ran the ELM model to generate forecasts for the 2023 monsoon rainfall season over Bangladesh. The forecasted results were then validated against the actual observed rainfall data for 2023.

Ensemble methods play a pivotal role in enhancing the accuracy and reliability of seasonal forecasts by combining predictions from multiple individual models. These methods combine forecasts from various sources to reduce the biases and uncertainties present in forecasts derived from a single model. Ensemble approaches increase forecast accuracy and offer probabilistic data that is essential for making decisions in many industries depending on seasonal climate projections (Singhratna *et al.*). In general, performance of ensemble methods outperforms the individual models.

2.3 Training and Validation Procedures

In order to ensure that forecasting models work reliably in a variety of scenarios and effectively generalize to previously unseen data, training and validation are crucial phases in the model-development process. The dataset is usually divided into two parts (training and testing dataset). The model is developed using the training dataset and then we see how well the machine scores on the test data. This process can be done using multiple techniques. We have used cross-validation (leave-one-out cross-validation) (Xiang *et al.*). The results show us how the model performs on unseen data. The evaluation metrics of the validation is discussed in the next section. Evaluation metrics play a pivotal role in quantifying the accuracy, reliability, and skill of seasonal forecasts. Some of the well-known metrics used to evaluate seasonal models are Pearson correlation, Kling-Gupta Efficiency (KGE), Rank Probability Skill Score (RPSS) etc.

These evaluation metrics provide comprehensive insights into forecast performance across different forecasting horizons, enabling model comparison and selection based on forecast accuracy, reliability, and skill.

Skill Score	Formula
Pearson Correlation Coefficient: ρ_{xy}	$\rho_{xy} = \frac{Cov(x, y)}{St(x) \times St(y)}$
RPS: Rank Probability Score	$RPSS = 1 - \frac{RPS_{exp}}{RPS_{ref}}$
KGE: Kling Gupta Efficiency MAPE: Mean Absolute Percentage Error	$KGE = \frac{MAPE_{benchmark}}{MAPE_{model}} \times 100\%$

3. Results and Discussion

By integrating historical climate data and machine learning models have demonstrated the capability to capture complex spatiotemporal patterns of seasonal climate variability. Deterministic and probabilistic forecasts are two distinct approaches in weather and climate prediction. Deterministic forecasts provide a single, specific prediction of future conditions based on initial conditions and model dynamics. This approach aims to predict a precise outcome, such as the exact amount of rainfall at a given time and place.

The results of the seasonal forecasting of monsoon rainfall over Bangladesh using the Extreme Learning Machine (ELM) model are discussed in this section. Using the climate monsoon data of the last 41 years (1982–2022), we forecast the monsoon, pre-monsoon and post-monsoon seasons of 2023. Figures 2 illustrate the deterministic forecast of precipitation patterns for June to September (JJAS), 2023. The latitude and longitude used in this forecast are 18°N–28°N and 85°E–95°E. The deterministic forecast suggests moderate to heavy precipitation in almost all parts of Bangladesh except the north-east parts (Sylhet region), where the precipitation is close to normal. The precipitation scale is taken from -4 to 4 and is divided into 8 parts.

Figure 3 illustrate the probabilistic forecast of precipitation patterns for June to September (JJAS), 2023. Probabilistic forecasts offer a range of possible outcomes and their associated probabilities, reflecting the uncertainty and variability in weather and climate systems. By generating multiple simulations through ensemble methods, probabilistic forecasts provide a more comprehensive picture of future conditions, helping to quantify risks and inform decision-making under uncertainty. The probabilistic forecast suggests high probability of above-

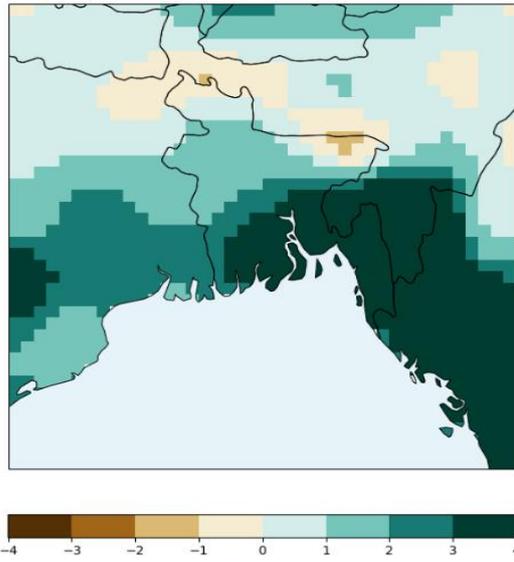


Fig. 2: Deterministic Forecast

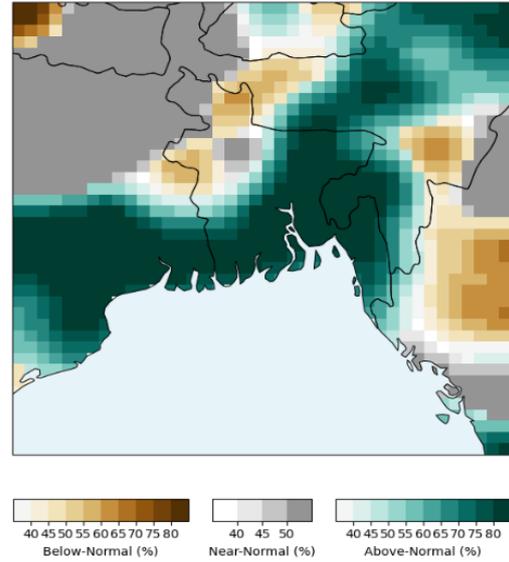


Fig. 3: Probabilistic Forecast

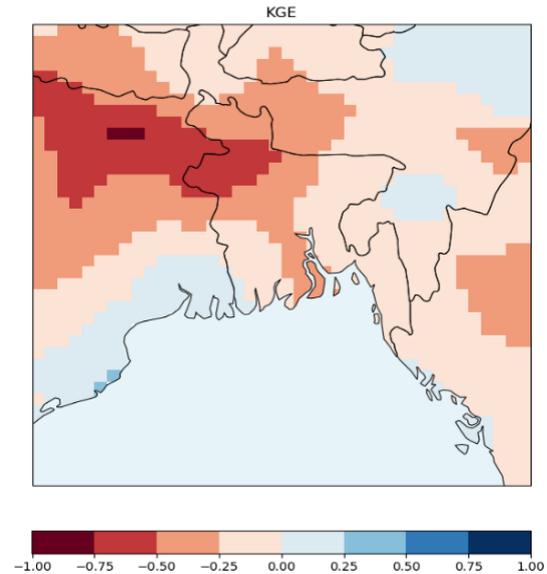
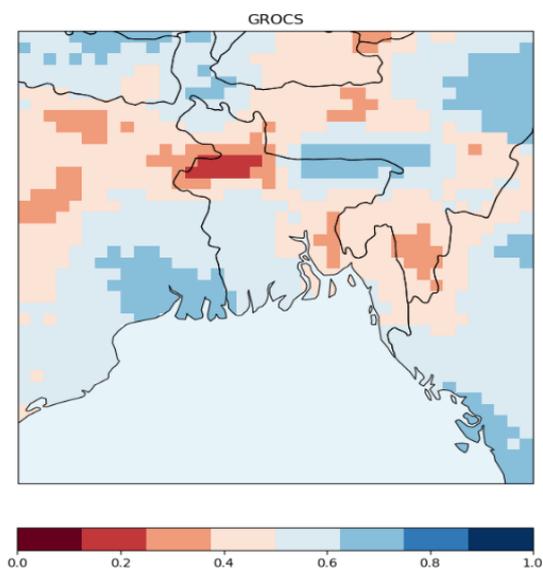
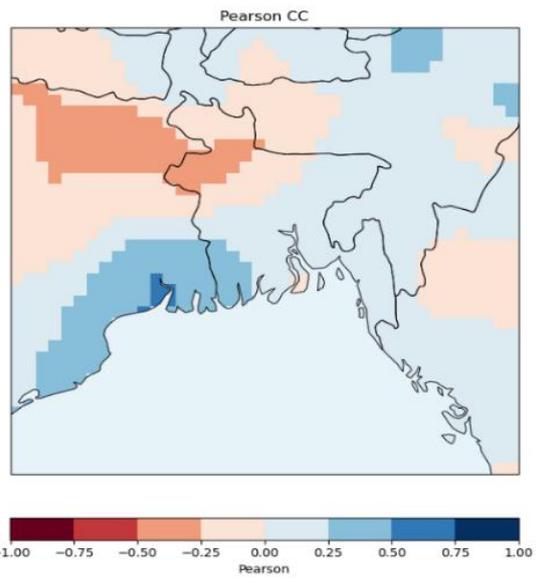
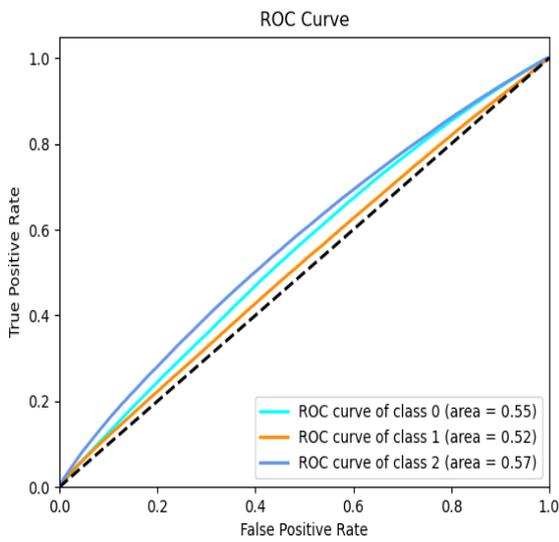


Fig. 4: Skill Score Metrics

normal precipitation in almost all parts of Bangladesh except the north-western parts (Rangpur region), where the probability of precipitation is near-normal. This approach is particularly valuable for long-term forecasts, such as seasonal and sub-seasonal predictions, where the inherent uncertainty is higher. Based on the comprehensive analysis and findings presented in the study, several promising avenues emerge for future research and development in the field of seasonal forecasting over Bangladesh using machine learning. The skill score of this model is determined using different metrics. The results are shown in figure below (Figure 4).

4. Conclusion

This study demonstrates the potential of machine learning models, specifically the Extreme Learning Machine (ELM), in improving the accuracy of seasonal forecasts of monsoon rainfall over Bangladesh. The model showed promising results during the monsoon season (June to September), capturing key precipitation patterns and achieving significant skill scores. Future research should focus on integrating more diverse climate variables to improve forecast accuracy, especially during non-monsoon periods. Advanced machine learning models and ensemble techniques should be explored to enhance the robustness and reliability of seasonal forecasts. The promising results during the monsoon season indicate the potential of machine learning models to revolutionize seasonal forecasting in Bangladesh. The findings underscore the necessity of further research to refine these models, integrate diverse datasets, and explore advanced machine learning techniques. By enhancing forecast accuracy, particularly during critical weather periods, this research has significant implications for agricultural planning, water management, and disaster preparedness in Bangladesh, ultimately contributing to the region's resilience against climate variability.

Acknowledgments

The authors thank Bangladesh Meteorological Department (BMD) for providing the support and guidance. The authors also express their gratitude to department of Applied Mathematics, University of Dhaka, for the help and cooperations they had extended. The observational, hindcast and forecast data were obtained from IRI website. Authors extend heartfelt gratitude to the esteemed researchers, scientists, and practitioners in the fields of climate science and machine learning, whose groundbreaking work has paved the way for the methodologies and analyses presented here.

References

- Ahasan, M.N., Chowdhary, M.A.M. & Quadir, D.A. (2010) Variability and trends of summer monsoon rainfall over Bangladesh. *Journal of Hydrology and Meteorology*, 7, 1–7.
- Acharya, N., Ehsan, M.A., Admasu, A., Teshome, A. & Hall, K.J.C. (2021) On the next generation (NextGen) seasonal prediction system to enhance climate services over Ethiopia. *Climate Services*, 24, 100272. Available from: <https://doi.org/10.1016/j.cliser.2021.100272>.
- Bari, S.H., Rahman, M.T., Hussain, M.M. & Ray, S. (2015) Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, 7, 69–77.
- Casanova, Sophie, and Bodo Ahrens. "On the weighting of multimodel ensembles in seasonal and short-range weather forecasting." *Monthly weather review* 137.11 (2009): 3811-3822.
- Cohen, Judah, et al. "S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts." *Wiley Interdisciplinary Reviews: Climate Change* 10.2 (2019): e00567.
- Dahale, S.D. & Singh, S.V. (1993) Modelling of Indian monsoon rainfall time series by univariate box-Jenkins type of models. *Advances in Atmospheric Sciences*, 10, 211–220.
- Ehsan, Muhammad Azhar, et al. "The ENSO fingerprint on Bangladesh summer monsoon rainfall." *Earth Systems and Environment* 7.3 (2023): 617-627.
- Hansen, J.W., Challinor, A., Ines, A., Wheeler, T. & Moron, V. (2006) Translating climate forecasts into agricultural terms: advances and challenges. *Climate Research*, 33, 27–41.
- Hossain, M., Kalam Azad, M., Karmakar, S., Islam Mondal, M., Das, M., Rahman, M.M. et al. (2019) Assessment of better prediction of seasonal rainfall by climate predictability tool using Global Sea surface temperature in Bangladesh. *Asian Journal of Advanced Research and Reports*, 4(4), 1–13. Available from: <https://doi.org/10.9734/ajarr/2019/v4i430116>.
- Huang, Guang-Bin, Qin-Yu Zhu, and Chee-Kheong Siew. "Extreme learning machine: theory and applications." *Neurocomputing* 70.1-3 (2006): 489-501.

- Lyu, Yang, et al. "Improving Subseasonal-To-Seasonal Prediction of Summer Extreme Precipitation Over Southern China Based on a Deep Learning Method." *Geophysical Research Letters* 50.24 (2023): e2023GL106245.
- Mahmud, I., Sheikh, H.B. & Rahman, M. (2017) Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. *Environmental Engineering Research*, 22(2), 162–168.
- Mohsin, M., Akhter, Y. & Begum, M. (2012) Modeling rainfall in Dhaka division of Bangladesh using time series analysis. *Journal of Mathematical Modelling and Application*, 1, 67–73.
- Singhrattna, Nkrintra, et al. "Seasonal forecasting of Thailand summer monsoon rainfall." *International Journal of Climatology: A Journal of the Royal Meteorological Society* 25.5 (2005): 649-664.
- Weigel, Andreas P., M. A. Liniger, and C. Appenzeller. "Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?" *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography* 134.630 (2008): 241-260.
- Xavier, Prince K., and Bhupendra Nath Goswami. "An analog method for real-time forecasting of summer monsoon subseasonal variability." *Monthly Weather Review* 135.12 (2007): 4149-4160.
- Xiang, Weiming, et al. "Verification for machine learning, autonomy, and neural networks survey." *arXiv preprint arXiv:1810.01989* (2018).