Notification No: F.NO. COE/Ph.D./(Notification)/600/2026

Date: 26-05-2026

Name of the Scholar: Ayush Vashisth

Name of the Supervisor: Prof. Mohammed Sharif

Name of the Co-Supervisor: Prof. Mohammad Shakeel

Name of Department: Department of Civil Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia, New Delhi

Title of Thesis: Prediction of Rainfall Based on Large Scale Climate Indices

Findings

Accurate rainfall prediction plays a vital role in sustainable water resources management, agricultural planning, and disaster risk reduction, particularly in monsoon-dominated regions where rainfall exhibits substantial temporal and spatial variability. The present research developed and evaluated a comprehensive, lag-aware, and interpretable machine learning framework for monthly rainfall prediction by integrating large-scale climate indices with advanced data-driven modeling techniques. The analysis incorporated several influential large-scale climate indices, including the El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO), and North Atlantic Oscillation (NAO), together with local climatic variables such as maximum and minimum temperature. To adequately capture delayed atmospheric and oceanic influences on regional rainfall, lagged values of the climate indices were incorporated into the modeling framework. This lag-based approach enabled the models to account for temporal teleconnections that often influence precipitation patterns with a delayed response.

A diverse range of supervised machine learning algorithms was implemented to model the complex and nonlinear relationships between climatic predictors and monthly rainfall. These algorithms included Random Forest (RF), M5P Model Tree, REPTree, Artificial Neural Networks (ANN), Support Vector Machine (SVM) with Radial Basis Function (RBF) and Pearson VII Universal Kernel Function (PUK), Gaussian Process Regression (GPR) with RBF and PUK kernels, and Multivariate Adaptive Regression Splines (MARS). To ensure reliable model assessment and generalization capability, the available datasets were divided into training (80%) and testing (20%) subsets. Furthermore, extensive hyperparameter tuning was conducted for each algorithm to optimize predictive performance while minimizing the risks of overfitting and underfitting.

The comparative analysis revealed notable differences in the predictive capabilities of the implemented models. At the Farukh Nagar station, the SVM model with the RBF Z1 kernel emerged as the most effective predictor of monthly rainfall. During the training phase, the model achieved a correlation coefficient (R) of 0.811, a mean absolute error (MAE) of 20.44 mm, a root mean square error (RMSE) of 49.19 mm, and a scattering index of 1.15. During testing, the model maintained strong predictive performance, yielding an R value of 0.718, MAE of 24.45 mm, RMSE of 46.86 mm, and a scattering index of 1.19. These results demonstrate the ability of the SVM-RBF Z1 model to effectively capture nonlinear relationships between climatic predictors and rainfall, while also maintaining satisfactory generalization performance on unseen data.

For the Safdarjung station, the Random Forest model demonstrated superior predictive accuracy compared with all other machine learning algorithms. The RF model achieved correlation coefficients of 0.9853 and 0.6674 for the training and testing datasets, respectively. Corresponding MAE values were 15.16 mm during training and 33.70 mm during testing, while RMSE values were 25.12 mm and 44.07 mm, respectively. The outstanding performance of the RF model highlights the effectiveness of ensemble tree-based approaches in handling complex climate–rainfall interactions and extracting meaningful information from high-dimensional climatic datasets.

The superiority of the SVM-RBF Z1 model at Farukh Nagar and the RF model at Safdarjung was further validated through Taylor diagram analysis. The Taylor diagrams demonstrated that these models more accurately reproduced the observed rainfall variability compared with alternative machine learning approaches. Their closer proximity to the reference observation point indicated better agreement in terms of correlation, standard deviation, and overall predictive skill. This graphical validation strengthened the statistical findings and confirmed the robustness of the selected models.

From an application perspective, the findings provide valuable support for climate-informed decision-making. The identification of dominant climatic predictors and their relative contributions enhance model transparency and interpretability, thereby addressing one of the major challenges associated with machine learning applications in environmental sciences. The ability to explain model behaviour increases confidence among decision-makers, policymakers, and resource managers who rely on forecasting tools for operational planning.

The implications of this research extend to multiple sectors, including hydrology, agriculture, and climate risk management. Accurate rainfall forecasts can support proactive water allocation strategies, reservoir operation planning, flood preparedness, drought mitigation, and agricultural scheduling. Improved understanding of climate teleconnections can also contribute to the development of more effective early warning systems and climate adaptation strategies. By providing reliable and interpretable rainfall predictions, the proposed framework helps bridge the gap between advanced machine learning methodologies and real-world operational applications. Overall, the research contributes to improved understanding of climate rainfall linkages and provides a scientifically robust foundation for climate-resilient water resources management, agricultural planning, and disaster risk reduction.