Machine learning algorithms for predicting rainfall in India

Sandip Garai[1,2], Ranjit Kumar Paul[3,*], Md. Yeasin[3], H. S. Roy[3] and A. K. Paul[3]

[1]The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi 110 012, India
[2]ICAR-Indian Institute of Agricultural Biotechnology, Ranchi 834 003, India
[3]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India

Due to the changing climate and frequent occurrence of extreme events, farmers face significant challenges. Precise rainfall prediction is necessary for proper crop planning. The presence of nonlinearity and chaotic structure in the historical rainfall series distorts the performances of the usual prediction models. In the present study, algorithms based on complete ensemble empirical mode decomposition with adaptive noise combined with stochastic models like autoregressive integrated moving average and generalized autoregressive conditional heteroscedasticity; machine learning techniques like random forest, artificial neural network, support vector regression and kernel ridge regression (KRR) have been proposed for predicting rainfall series. KRR has been considered to combine predicted intrinsic mode functions and residuals generated by various algorithms to capture the volatility in the series. The proposed algorithms have been applied for predicting rainfall in three selected subdivisions of India, namely, Assam and Meghalaya, Konkan and Goa, and Punjab. An empirical comparison of the proposed algorithms with the existing models revealed that the developed models have outperformed the latter.

Keywords: Climate change, crop planning, empirical comparison, machine learning, prediction, rainfall.

the phenomena of nonstationarity and nonlinearity distort the inherent functional form of the underlying series[9].

Therefore, several artificial intelligence (AI) techniques have been utilized to tackle such complexities in a dataset. The most widely used techniques are artificial neural networks (ANNs)[10,11] and SVM[3]. ANN has been proven to provide accurate prediction when used singly or combined with other methodologies. With a few exceptions, ANN-based hybrid models outperform the singular ANN models as they help reduce prediction failures in many real-world problems[12]. A combination of various techniques can be an alternative to achieve more efficiency in forecasting[13].

Wang and Ma[14] proposed two ensemble strategies, i.e. random subspace (RS) and bagging (B), and utilized support vector regression (SVR) as the base model to propose the RSB–SVR model for credit risk assessment. Most neural networks (NNs) do not possess a memory for the individual inputs fed to them, and there is no state-space maintained for them[15].

Changes in historical rainfall patterns may be due to climatic change, which alters cropping patterns, causes extreme weather behaviours like drought and flood, and is problematic for water resource management authorities[16]. So, rainfall risk mitigation is important for policymakers, hydrologists, and ultimately farmers. To improve water supply