



Hybrid deep learning algorithms on the dimensionally reduced dataset with optimized parameters for high-precision predictions of rainfall in Chhattisgarh State

NISHA THAKUR and SANJEEV KARMAKAR

Bhilai Institute of Technology, Durg, Chhattisgarh, India

(Received 9 April 2023, Accepted 23 November 2023)

e mails : *nishathakur.india@gmail.com; dr.karmakars@gmail.com

सार- अनुक्रमिक हाइब्रिड मॉडल का उपयोग करके बहु-परिवर्त वर्षा डेटा का समय श्रृंखला पूर्वानुमान किया गया। इस मॉडल में, मूल जानकारी की न्यूनतम कमी के साथ डेटासेट के आयाम को कम करने के लिए प्रमुख घटक विश्लेषण (पीसीए) का उपयोग किया गया। डीप लर्निंग एल्गोरिदम (LSTM) में उपयोग की जाने वाली गवाक्ष आकार और लॉन्ग शॉर्ट टर्म मेमोरी (LSTM) इकाइयों की संख्या का अनुकूलित मान जेनेटिक एल्गोरिदम (GA) का उपयोग करके आकलित किया गया। 99 प्रतिशत मूल जानकारी को बनाए रखते हुए, इसके आयामों को कम करने के लिए मूल डेटासेट पर पीसीए लागू किया गया। इसके बाद, PCA का उपयोग करके प्राप्त डेटासेट को LSTM गवाक्ष आकार और इकाइयों की संख्या के अनुकूलित मान प्राप्त करने के लिए GA एल्गोरिथम में इनपुट किया गया। विभिन्न मॉडलों जैसे कि LSTM, PCA के LSTM में विलय (PCA-LSTM), GA के LSTM में विलय (GA-LSTM) तथा PCA के GA और LSTM में विलय (PCA-O-LSTM) से प्राप्त परिणामों का एक व्यापक, तुलनात्मक अध्ययन किया गया। LSTM का उपयोग पूर्वानुमान 90:10, 80:20 और 70:30 के प्रशिक्षण-परीक्षण अनुपातों के लिए किया गया, जहाँ 80:20 अनुपात ने बेहतर परिणाम प्रदान किए, इसलिए बाकी विश्लेषण के लिए 80:20 के अनुपात का उपयोग किया गया। परिणामों की बेहतर व्याख्या के लिए, प्रत्येक मॉडल को विभिन्न कालावधियों, जैसे 10, 20, 50, 100 और 200 के लिए चलाया गया। विभिन्न मॉडलों का उपयोग करके पूर्वानुमानों की गुणवत्ता का मूल्यांकन विभिन्न प्राचलों जैसे निर्धारण गुणांक (R²), माध्य वर्ग त्रुटि (MSE), मूल-माध्य-वर्ग त्रुटि (RMSE), माध्य निरपेक्ष त्रुटि (MAE), सामान्यीकृत त्रुटि (NORM), RMSE-प्रेक्षण मानक विचलन अनुपात (RSR) और कोसाइन समानता (CS) द्वारा किया गया। R² का मान (0.962874, 0.972276), (0.970131-0.955826) और (0.950982- 0.972991) की रैंज में पाया गया, जिसमें GA-LSTM, PCA-LSTM और PCA-O-LSTM के मामले में क्रमशः 200, 200 और 100 कालावधियों के लिए उक्त प्राचलों का सर्वोत्तम मान पाया गया। R² का सर्वोत्तम सम्भावित मान PCA-O-LSTM मॉडल के मामले में देखा गया, जिसमें GA के साथ-साथ कम आयामी डेटासेट को गवाक्ष आकार और इकाइयों की संख्या को अनुकूलित करने के लिए शामिल किया गया।

ABSTRACT. A Time series forecasting of multi-variant rainfall data was done using a sequential hybrid model. In this model, principal component analysis (PCA) was used to reduce the dimension of the dataset with minimal loss of the original information. The optimized value of window size and the number of Long Short Term Memory (LSTM) units to be used in the deep learning algorithm (LSTM) were estimated using the Genetic algorithm (GA). PCA was applied to the original dataset to reduce its dimensions, keeping 99 percent of the original information. Thereafter, the dataset retrieved using PCA was inputted to the GA algorithm to get the optimized values of LSTM window size and number of units. A comprehensive, comparative study of the results obtained from various models, such as LSTM, PCA merged to LSTM (PCA-LSTM), GA merged to LSTM (GA-LSTM), and PCA merged to GA and LSTM (PCA-O-LSTM) was carried out. The prediction using LSTM was carried out for training-testing ratios of 90:10, 80:20 and 70:30, where the 80:20 ratio provided better results therefore this ratio of 80:20 was used for the rest of the analysis. For a better interpretation of the results, each of the models was run for various epochs, like 10, 20, 50, 100 and 200. The quality of predictions using various models was evaluated by different parameters like using determination coefficient (R²), mean square error (MSE), root-mean-square error (RMSE), mean absolute error (MAE), Normalized error (NORM), RMSE-

observations standard deviation ratio (RSR) and cosine similarity (CS). The value of R2 was found in the range of (0.962874, 0.972276), (0.970131-0.955826) and (0.950982- 0.972991) with the best value of the said parameter for 200, 200 and 100 epochs in case of GA-LSTM, PCA-LSTM and PCA-O-LSTM, respectively. The best possible value of R2 was seen in the case of the PCA-O-LSTM model in which a dimensional-reduced dataset along with GA optimized the window size and numbers of units were incorporated.

Key words – Prediction, Precipitation, Genetic algorithm (GA), Principal component analysis (PCA), Long-Short-Term memory network (LSTM), Optimization method.