MISSING RAINFALL DATA ESTIMATION USING MACHINE LEARNING APPROACHES: A STUDY IN SRI LANKA

R.M.P.S. Rathnayaka, M.K.N. Kumari, K.G.S. Nirmanee, and M.H.J.P. Gunarathna

Department of Agricultural Engineering and Soil Science, Faculty of Agriculture, Rajarata University of Sri Lanka, Anuradhapura, Sri Lanka.

Establishing a robust method to estimate the missing rainfall is vital for many applications, including water resource planning and management. This study aimed to find the applicability of machine learning algorithms to estimate monthly missing rainfall data in Anuradhapura, Matara and Colombo districts of Sri Lanka with contrasting rainfall variations. Seven rain gauge stations belong to DL_{1b}, which has minor rainfall variation among the station (CV= 8.56), were selected from Anuradhapura district. Six rain gauge stations belong to wet and intermediate zones (IL_{1a}, WM_{1a}, and IL_{1b}) were selected from Matara District, which has relatively higher rainfall variation among the stations (CV= 33.76). Six rain gauge stations from Colombo district (WL_{1a} and WL_3), which has moderate rainfall variation among the stations (CV = 22.22) were selected for the study. Waikato Environment for Knowledge Analysis (WEKA) software was used to implement artificial neural networks, random forest, K-nearest neighbor, Support Vector Machine (SVM), and linear regression algorithms. The wrapper approach and best-first search method were performed to identify suitable stations for missing rainfall data estimations. The root mean square error of the estimations was used to compare the performances of different machine learning algorithms. The results revealed that the machine learning algorithms could estimate the missing rainfall data of Anuradhapura district satisfactorily, but not in Matara district. Colombo district showed moderate results compared to the other two districts. Among the machine learning methods, SVM showed better results compared to other methods. Hence, this study demonstrated that machine learning approaches effectively estimate missing rainfall data for the stations with minor rainfall variations. The SVM algorithm performed better compared to other selected machine learning algorithms.

Keywords: Agro-ecological zones, Linear regression, Root mean squared error, Support vector machine, WEKA