National Institute of Technology Rourkela

Departmental Seminar

Seminar Title	: Conference Return Seminar : MACHINE LEARNING-BASED GROUNDWATER LEVEL PREDICTION FOR SUSTAINABLE WATER RESOURCE MANAGEMENT IN AN INDUSTRIAL CATCHMENT
Speaker	: Dr. Ratnakar Swain
Supervisor	: Dr. M. Gattu
Venue	: CE Seminar Hall
Date and Time	: 17 Mar 2025 (430 pm)
Abstract	: Groundwater is a vital source of freshwater, playing a crucial role in agriculture and economic development. Accurate groundwater level (GWL) prediction is essential for sustainable water resource management. As soft computational data-driven technologies have advanced in recent years, numerous machine learning models have been created and are being used for GWL forecasting. ML models are more desirable than physically based and numerical methods because they can simulate and predict GWL without needing a thorough understanding of the underlying topographical and hydro-geophysical characteristics. An efficient substitute that does not call for precise and particular physical factors and attributes are machine learning (ML) techniques. Some of the widely used ML models in GWL Prediction are Support Vector Machine (SVM), Artificial Neural Networks (ANN), Extreme learning machine (ELM), Fuzzy logic, Adaptive Network-based Fuzzy Inference System (ANFIS). For many hydro-meteorological applications, hybrid machine learning and genetic models have been attempted and tested in numerous researches. The literature evaluation indicates that these hybrid machine learning models perform better in terms of prediction than many solo models. The ability of hybrid models to identify intricate mathematical nonlinear interactions between the independent and dependent parameters is one of its primary benefits. The objective of this study is to downscale Monthly GWL from quarterly GWL data using the Kalman Filter algorithm; To predict GWL using rainfall, evapotranspiration, mean temperature, and infiltration data using three ML methods – Decision tree, Random Forest and Bagged Tree; To compare and analyze the performance of the three models. Talcher is one of the 4 sub-divisions of Angul district in the state of Odisha. The latitude and longitude of Talcher are N 20° 56' 57.372", E 85° 14' 0.744". It has a savanna climate which is known as tropical wet and dry climate with the Classification: Aw. The quantity of rainfall during summers surpasses that of winters. The area's yearly temperature is 30.53ºC (or) 86.95ºF and is 4.56% higher than India's average. Talcher typically receives about 131.22 mm of precipitation and has 134.21 rainy days annually that is36.77% of the time.Brahmani River flows throughTalcher. The Brahmani River and its major tributaries such as NandiraJhor, SingharaJhor, TikraJhor, Samakoi, Nigra,Gambhira, etc. run through the Angul district. For the training stage the mean value of GWL is 3.90 m, with a maximum value of 7.45 m and minimum value of 0.22m. From the graphs and the performance metrics, it can be observed that for the training stage the Random Forest model is best among the three models used with r2= 0.82, RMSE = 0.62, MSE = 0.38 and NSE = 81.77. Decision tree models, with r2 around 0.68, has less accurate results than other models. Cross validation seems to not increase accuracy in prediction for the decision tree model during the training stage. The Bagged Tree model has an r2 value of 0.7, RMSE value of 0.85, MSE value of 0.76 and NSE value of 63.34. From the observed GWL versus predicted GWL graph of Bagged tree, it can be observed that around half of the values are over estimated and half of the values are underestimated. Smaller GWL values are over estimated and larger GWL are under estimated by the Bagged Tree model. The mean value of GWL for the testing stage is 3.94 m with maximal and minimal values of GWL for the testing stage being 8.83 m and 1.22 m, respectively. It can be observed that for the testing stage, from the graphs and the performance metrics, the Random Forest model is best among the three models used with r2= 0.80, RMSE = 0.61, MSE = 0.37 and NSE = 94.53. For decision tree, the cross validated models have shown an increase in accuracy. The 10 fold cross validated decision tree model has an r2 value of 0.70, RMSE value of 0.72, MSE value of 0.53 and NSE value of 73.70. The performance of Decision tree and Bagged tree models increased in the testing stage from the training stage. But the performance of Random Forest decreased slightly in the testing stage. Similar to the training stage values, the Bagged tree has overestimated smaller values of GWL and under estimated larger values of GWL. As in the training stage, here too the ensemble models have better performance than Decision tree. In this study, multiple models were built for groundwater level prediction and compared to explore potential knowledge of GWL predictions. Forecasting of GWL is essential for sustainable management of ground water resources. GWL datasets from 2003 -2020 from the Talcher region, Odisha were collected and used to train and test three ML models for long-term prediction. Rainfall, temperature, calculated infiltration and calculated evapotranspiration values were used for the prediction of GWL. The Evapotranspiration values were calculated using the FAO-24 Radiation equation. Kalman filter was used to downscale GWL data. The performance of three machine learning models - Decision Tree, Random Forest and Bagged Tree in predicting GWL was evaluated. Each model is evaluated using the mean square error (MSE), root mean square error (RMSE), coefficient of determination (R2), and Nash–Sutcliffe efficiency (NSE). The Random Forest model is seen to have best accuracy in both the training (r2 = 0.82) and testing stages (r2 = 0.80).