ABSTRACT

In the current scenario, India's infrastructure is rapidly growing as the Ministry of Road Transport and Highways (MoRTH) has planned to construct many new Expressways and National Highways (NH) throughout the country. Additionally, numerous roads are being widened from two-lane to four-lane and six-lane through the Engineering Procurement Construction (EPC) model infrastructure development plan. The designed thickness for constructing various layers of these roads depends on the California bearing ratio (CBR) value of the soils/materials used in the subgrade layer.

CBR has been a widely used test for the design of road and airfield pavements for the last few decades. A considerable quantity of soil is obtained after testing for their suitability to construct embankments for such structures. CBR test is also among the selection criteria test. This test method is known for some rigorous and time consuming effort in the laboratory, including soaking for 96 hours on many occasions. This may sometimes become an issue when the soil properties frequently change in short stretches and the results are expected quickly by the management from the field unit.

Laboratories are often facing long queue of materials for testing. Therefore, preserving such a massive quantity of soil awaiting its turn for testing makes it more laborious and time-consuming. Furthermore, the testing at the laboratory includes a testing fee, material transportation, its processing and finally, discarding the tested materials in a dump yard. All this adds to the final cost of the project.

To overcome such problems, this study attempts to predict the soaked CBR value of fine-grained plastic soils through some machine learning algorithms. A quick and reliable estimation of CBR value is likely to help select the materials from borrow areas, providing the requisite input for pavement design with a good measure of assurance,

faster cost estimation of embankment and subgrade along with the various pavement layers for finalizing the funding proposal of the project within a short period.

In the present study, in-situ soil samples were collected from an ongoing National Highway project (NH-29, Varanasi to Gorakhpur) entailing 4-laning from the existing 2-lane. As a result, 1011 soil samples were collected from different chainage along the length of the road and brought to the laboratory for assessing the geotechnical parameters comprising Atterberg's limits, sieves analysis, modified Proctor compaction parameters, and CBR in accordance with respective BIS specifications.

These parameters were examined for the application of Multi Expression Programming (MEP), a variant of Genetic Programming (GP) which is itself a subset of Genetic Algorithm (GA), and eXtreme Gradient Boosting (XGBoost) algorithms for predicting the soaked CBR value. Additionally, three data divisional approaches, statistical, K-Fold and Fuzzy C-Means (FCM) clustering, were also analyzed on each of the adopted machine learning algorithms. Numerous performance measurement parameters such as coefficient of determination (R^2), coefficient of correlation (R), mean absolute error (MAE), root mean square error (RMSE), variance account for (VAF), Willmott's index of agreement (IOA), index of scattering (IOS) and a20-index were used to identifying the predictive and generalization capability of the developed models.

Based on the overfitting and ranking analysis performed on the values obtained for each of the performance measurement parameters mentioned above of the three data division approaches model, it is perceived that the CBR prediction model developed through the K-Fold approach demonstrates higher accuracy in both MEP and XGBoost algorithms. The selected MEP model comprises Sand Content (%), Fines Content (FC), Plasticity Index (PI) and Maximum Dry Density (MDD) as input parameters. In contrast, the selected model of the XGBoost algorithm consists of S, FC, PL, PI, MDD and Optimum Moisture Content (OMC). Therefore, these parameters were considered significant for developing the prediction model for the soaked CBR value of fine-grained plastic soils.

It was observed from the scatter plot that the closeness of data points towards the line of equality (also known as the 45° line) is maximum in the XGBoost algorithm's model as compared to the MEP algorithm. Furthermore, the CBR model developed through the MEP algorithm was found to predict almost 83% and 96% observations within ±10% and ±20% variations, respectively, whereas for XGBoost algorithm predicted 92% and 99% observations within ±10% and ±20% variations, respectively. Therefore, the XGBoost algorithm's model has significant advantage in predicting the soaked CBR value of soils close to the laboratory CBR value of the soils.

The Regression Error Characteristics (REC) curve demonstrates that both algorithms Area Under the Curve (AUC) value were higher than 0.9. Though MEP and XGBoost algorithms were sufficient to predict the soaked CBR value of fine-grained plastic soils, a comparative analysis shows that the XGBoost algorithm was superior to the MEP algorithm. Additionally, the accuracy analysis shows that R^2, R, MAE, RMSE, VAF, IOA, IOS and a20-index values were improved by 27%, 13%, 23%, 65%, 27%, 7%, 2%, and 3%, respectively when XGBoost algorithm was adopted over the MEP algorithm. The overall comparative analysis of both algorithms reveals that the CBR prediction model developed through the XGBoost algorithm was much more efficient than the MEP algorithm. Hence, it is understood from the above observations that the machine learning algorithms prominently influenced the predictive ability in developing the prediction model.

The results obtained for the internal validation, performed through the K-Fold cross-validation approach, on the developed model show that both MEP and XGBoost algorithm models efficiently predicted the soaked CBR of fine-grained plastic soils. However, the external validation results of the developed model and the models cited in the literature demonstrate that the location of the soil significantly influences the predictive ability of any model in the field of geotechnical engineering.

Lastly, a Graphical User Interface (GUI) was developed in Python for the final selected model i.e., the XGBoost model. The designed interface was named “CBR prediction tool for Fine-grained plastic soils” to benefit the field engineers.