मशीन लर्निंग एल्गोरिदम के माध्यम से महीन दाने वाली ढलनशील मृदा के भीगे हुए सीबीआर का अनुमान

PREDICTION OF THE SOAKED CBR OF FINE-GRAINED PLASTIC SOILS THROUGH MACHINE LEARNING ALGORITHMS

Thesis Submitted in partial fulfilment

for the Award of Degree

Doctor of philosophy

By

गौरव वर्मा

GAURAV VERMA

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY (BANARAS HINDU UNIVERSITY)

VARANASI – 221005 (INDIA)

Roll No. 16061007 **2022**

CHAPTER 5 SUMMARY CONCLUSIONS AND SCOPE FOR THE FUTURE STUDY

5.1 GENERAL

The present chapter summarizes the findings and conclusions from the current investigation. Moreover, the scope for future studies, which authors feel would help strengthen the predictive ability of the CBR prediction model.

5.2 SUMMARY AND CONCLUSION

The current study offers the novel application of various machine learning algorithms in predicting the soaked CBR of fine-grained plastic soils. For this purpose, 1011 in-situ soil samples were collected from an ongoing construction project work site, and laboratory tests were performed as per BIS specifications. The prepared dataset was analyzed using the frequency distribution plot and descriptive statistical analysis. The selection of input parameters for the model development was made through Pearson's correlation analysis and the information available for this type of study in the literature. The dataset was divided into TR and TS sets using the statistical, K-fold, and FCM divisional approaches. Two ML algorithms, MEP and XGBoost, were adopted to develop the CBR prediction model in soaked conditions. The prediction accuracy of the developed models was assessed through various statistical performance measurement parameters. To develop the models, the following conclusions are drawn:

1. Among numerous index and engineering properties of fine-grained plastic soil, CBR value was in good association with S, FC, PL, PI, MDD and OMC. It was also observed from the correlation analysis that the CBR value of fine-grained

plastic soil increases with an increase in the SC and MDD and a decrease in the FC, LL, PL, PI and OMC.

2. In the MEP algorithm, models developed through statistical, K-fold, and FCM approaches can explain a maximum variability of 63%, 63% and 62% in the CBR value through adopted input parameters. The maximum variability was found for the statistical approach followed by K-fold and FCM approaches. However, based on the overfitting and ranking analysis, it is concluded that the model developed through the K-fold approach established the highest accuracy followed by statistical and FCM approaches. The final order of accuracy achieved in the MEP algorithm is $MEP_K > MEP_S > MEP_F$.
3. The final selected MEP model is comprised of S, FC, PL, PI, MDD and OMC as input parameters and CBR as an output parameter. The MEP algorithm hyper-parameters value achieved corresponding to the final selected model were: number of subpopulation 20, population size 1000, chromosome length 50, number of generations 10000, the probability for mutation and crossover were 0.01 and 0.9, respectively, uniform crossover type and the adopted functional sets were +, -, ×, /, square root.
4. In the XGBoost algorithm, models developed through statistical, K-fold and FCM data divisional approaches can explain a maximum variability of 78%, 80%, and 81%, respectively, in the CBR value through adopted input parameters. This seems to be maximum for the FCM approach followed by K-fold and statistical approaches. Based on the overfitting and ranking analysis, it is concluded that the model developed through the K-fold approach demonstrates the highest accuracy, followed by FCM and statistical approaches. The final order of accuracy achieved in the XGBoost algorithm is $XGB_K > XGB_F > XGB_S$. Consequently, one can

easily understand that the accuracy of any predictive model is significantly influenced by the data divisional approaches i.e., data used to train the models.

5. The final selected XGBoost model is comprised of S, FC, PL, PI, MDD and OMC as input parameters and CBR as an output parameter. The values of XGBoost algorithms hyper-parameters achieved corresponding to the final selected model were: base score 0.5, gbtree booster type, column sample by level, node and tree was 1, 1 and 1, respectively, gamma value was 0, learning rate 0.161, maximum delta step 0, maximum depth 3, minimum child weight 1, n estimators 80, n jobs -1, number of parallel tree 1, subsample 0.5, regularization alpha and lambda were 0 and 1, respectively.
6. The results of the REC curve reveal that both algorithms are reliable in predicting the soaked CBR value of fine-grained plastic soils. However, a comparative analysis demonstrates that the XGBoost algorithm is considerably superior to the MEP algorithm. The accuracy analysis exhibit that R^2, R, MAE, RMSE, VAF, IOA, IOS, and a20-index value get improved by 27%, 13%, 23%, 65%, 27%, 7%, 2%, and 3% when XGBoost algorithm is adopted over the MEP algorithm. Moreover, from an error point of view, the MEP algorithm can predict almost 96% of observations within ±20%. In contrast, the XGBoost algorithm could predict 99% of observations which is substantially higher than the MEP algorithm. Conclusively, it was perceived that the soaked CBR value of fine-grained plastic soil predicted through the XGBoost algorithm is much more than the MEP algorithm. Therefore, one can understand that the predictive ability is prominently influenced by the type of ML algorithm used for developing the prediction model.

7. The internal validation of the present study models was done through the K-Fold cross-validation approach. With five-folds, the selected MEP_{K-1} and XGB_{K-1} models exhibit an R-value greater than 0.95, which means that both models are efficient in predicting the soaked CBR of fine-grained plastic soil samples. However, the selected models were processed for external validation through some literature datasets to identify their generalization capability. The results obtained from the analysis reveal that the developed models are insufficient in predicting the soaked CBR of fine-grained plastic soils collected from different locations. Nevertheless, for cross-checking the results, the external validation of the literature models was conducted using the present study datasets. Similarly, the literature models were also inadequate in predicting the soaked CBR value. Conclusively, it is established from these investigations that models developed for the soils collected from a particular location may only be suitable for that specific region.
8. In this study, a reliable GUI was designed in python language for the XGB_{K-1} model. The developed interface was designated as "*CBR Prediction tool for Fine-grained plastic soils*" where CBR referred to California Bearing Ratio (test method for which prediction is being performed). Therefore, the developed interface is beneficial for the researchers and user-friendly for the site engineers.

5.3 SCOPE FOR THE FUTURE STUDIES

In the present study, an effort has been made to bridge the gaps found in the existing literature studies. However, the following are a few gaps that authors feel necessary to be bridged to strengthen the generalization capability of the CBR prediction model.

1. This study attempted to predict the soaked CBR value of fine-grained plastic soils. Researchers in the future should attempt an investigation for other types of soil used in the subgrade or sub-base layer of the pavement.
2. An investigation should also be performed by adopting many other machine learning algorithms.
3. The present investigation should also be extended by incorporating many more datasets from other regions of India as well as from the world.
4. In this study, the prediction of CBR value was performed as the test method used to design the flexible pavement in India. Forthcoming, the prediction can also be performed for the Resistance Value (R-value) and Resilient Modulus (M_R) test methods which are recently being used in many developed countries.