

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

CERTIFICATE

This is to certify that the revised thesis entitled "**PREDICTION OF THE SOAKED CBR OF FINE-GRAINED PLASTIC SOILS THROUGH MACHINE LEARNING ALGORITHMS**" is being submitted by **Mr. GAURAV VERMA** (Registration No. 16061007) in partial fulfillment for the award of Ph.D. in Department of Civil Engineering of the Indian Institute of Technology (Banaras Hindu University) Varanasi is a record of bonafide work carried out by him.

Date of Submission: 29.11.2022



Supervisor
Department of Civil Engineering
Indian Institute of Technology, (BHU)
Varanasi-221005
(Supervisor)

Professor

Department of Civil Engineering

Indian Institute of Technology (Banaras Hindu University) Varanasi

Forwarded by:



Head of Department

विभागाध्यक्ष/HEAD
(Signature with Seal) विभाग
Department of Civil Engineering
भारतीय प्रौद्योगिकी संस्थान (बी.एच.यू.)
Indian Institute of Technology, (BHU)
Varanasi-221005

CERTIFICATE

It is certified that the work contained in the thesis titled “*PREDICTION OF THE SOAKED CBR OF FINE-GRAINED PLASTIC SOILS THROUGH MACHINE LEARNING ALGORITHMS*” by *Mr. GAURAV VERMA* (Registration No. 16061007) has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.


10/04/2023

Dr. Brind Kumar

Supervisor

(Supervisor)

Department of Civil Engineering
Indian Institute of Technology, (BHU)
Varanasi-221005

Professor

Department of Civil Engineering

Indian Institute of Technology (Banaras Hindu University) Varanasi

DECLARATION BY THE CANDIDATE

I, **GAURAV VERMA**, certify that the work embodied in this thesis is my own bonafide work carried out by me under the supervision of **Dr. BRIND KUMAR** from July 2016 to May 2022 at the Department of Civil Engineering of the Indian Institute of Technology (Banaras Hindu University) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I also declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully lifted any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, thesis, etc., or available at websites and included them in this thesis and cited as my own work.

Date: 10/04/2023

Place: IIT (BHU) Varanasi



GAURAV VERMA


CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my knowledge.



Dr. Brind Kumar Supervisor
Professor Department of Civil Engineering
Indian Institute of Technology, (BHU)
Department of Civil Engineering, 221005

Indian Institute of Technology (Banaras Hindu University) Varanasi



Signature of Head of Department
विभागाध्यक्ष/HEAD

जानपद अभियांत्रिकी विभाग
Department of Civil Engineering
भारतीय प्रौद्योगिकी संस्थान (बी.एच.यू.)
Indian Institute of Technology, (BHU)
221005 Varanasi

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: "*PREDICTION OF THE SOAKED CBR OF FINE-GRAINED PLASTIC SOILS THROUGH MACHINE LEARNING ALGORITHMS*"

Name of the Student: **Mr. Gaurav Verma**

Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the **Doctor of Philosophy** degree.

Date: 10/04/2023

Place: IIT (BHU) Varanasi



(Gaurav Verma)

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

ACKNOWLEDGEMENT

Though only my name appears on the cover letter of this dissertation, a lots of wonderful people have contributed to reach me to my destination. I wish to owe my sincere appreciation to all those peoples who helped me to accomplish several achievements during this span.

Firstly, I would like to bow my head to this Holy Land of Varanasi, where the combined divine power of *Kashi Vishwanath*, *Maa Durga* and *Sankat Mochan* provided me good health, knowledge, strength and capability to work hard at vicious time period. Additionally, I would also like to thank “*Shrimad Bhagavad Gita*” who always boosted my moral toward my work on every occasion of negatory time period. Medical science has greatly improved their vision about many organs of the human body but study about the MIND, which control other organs of the body, is only taught in “*Shrimad Bhagavad Gita*”.

I also bow my head with great respect and gratitude in the pious feet of “*Bharat Ratna Mahamana Pandit Madan Mohan Malaviya Ji*” for establishing this great historical institute for the center of excellence and knowledge, Banaras Hindu University (now Indian Institute of Technology (BHU) Varanasi for engineering studies) who facilitated me the opportunity to enhance and expand the horizon of knowledge.

I express my sincere gratitude to my supervisor *Dr. Brind Kumar*, Professor, Department of Civil Engineering, Indian Institute of Technology (Banaras Hindu University) Varanasi. I am really fortunate to have a supervisor who gave me the liberty to express my own and at the same time the valuable guidance to recover when my steps faltered. I also want to acknowledge him for including me in many of the field visits which ultimately enhanced my knowledge and vision. I also learned from him that how

to take maximum advantages from fewer facilities. Without his support, this dissertation would not have been possible and I shall always be grateful to him for his guidance, valuable suggestions, and kind encouragement during my entire research work.

I am also grateful to my research committee members and Comprehensive examination member: **Prof. Krishna Kumar Pathak (Civil)** and **Prof. Sanjay Kumar Pandey (Mathematics)** for their instructions, valuable suggestions and support throughout the study.

I am much obliged to **Pankaj Sahoo**, engineer at National Highway Authority of India, for his continuous support and providing the laboratory facilities at Jaypee Construction site situated in Mau (Ghazipur), U.P. I am also thankful to the **Jaypee Group of construction** for their healthy support in the laboratory.

I am thankful to **Dr. Ankit Gupta**, Associate Professor, IIT (BHU) Varanasi, for providing the laboratory facilities. I am really thankful to **Dr. Abhisek Mudgal**, Assistant Professor, IIT (BHU) Varanasi, for commenting on my views and helping me enrich my ideas. My personal discussions with him regarding the statistical analysis and numerical simulation have helped me in enhancing my knowledge. I also want to express my gratitude to **Dr. Nikhil Saboo**, Assistant Professor, IIT Roorkee, for having a peaceful and energetic discussion on various issues in the laboratory during his professional experience at IIT (BHU) Varanasi. I feel my gratitude to **Dr. Agnivesh Pani**, Assistant Professor, IIT (BHU) Varanasi, for his valuable support.

I am obliged to **Prof. Sasankasekhar Mandal**, Head, Department of Civil Engineering, IIT (BHU) Varanasi, for providing the necessary facilities and a pleasant working atmosphere in the department. I would also like to thank and express my heartfelt thanks and indebtedness to other faculty members of the Department of Civil

Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, for their support and motivation.

I am sincerely thankful to my M. Tech. Supervisor **Dr. Som Nath Sachdeva**, Professor Department of Civil Engineering, National Institute of Technology Kurukshetra (currently Vice-Chancellor, Kurukshetra University), for encouraging me to do a Ph.D. I am also thankful to him for enhancing my knowledge to Highway Engineering.

I am highly obliged to **Dr. G.D. Ransinchung R.N.**, Professor, Department of Civil Engineering, IIT Roorkee for providing me an unforgettable opportunity at the beginning of my profession in the field of Highway Engineering.

In my journey, I have found a good teacher, a good friend and an elder brother, **Dr. Sandeep Arora**, Department of Geophysics, Institute of Science Banaras Hindu University who provide me moral support and inspirations throughout the completion of this degree.

I am also grateful to Library In-charge, Civil Engineering Library IIT (BHU) for allowing me to access vast literature resources. I am also grateful to Laboratory In-charge and technicians of the following labs of our institute, other institutes and the construction company for allowing and helping me in performing laboratory works.

1. Structural Engineering Laboratory, Civil Engineering Department IIT (BHU)
2. Geotechnical Engineering Laboratory, Civil Engineering Department IIT (BHU)
3. Environmental Engineering Laboratory, Civil Engineering Department IIT (BHU)
4. Hydraulic Engineering Laboratory, Civil Engineering Department IIT (BHU)
5. Transportation Engineering Laboratory, Civil Engineering Department IIT (BHU)
6. Workshop, Mechanical Engineering, IIT (BHU)

7. Central Instrument Facility, IIT (BHU)
8. National Highway Authority of India (NHAI), India
9. Jaypee Group of Engineering & Construction, India
10. Ministry of Education, India

I am also thankful to my seniors: *Dr. Kanakabandi Shalini, especially to Dr. Arunabh Pandey, especially to Dr. Jayvant Choudhary, Dr. Satyajit Mondal, Mr. Rahul Singh, especially to Dr. Asnit Gangwar (MET) and Dr. Shashi Kant Shah (BHU)* colleagues: *especially to Mr. Arunava Ray (MIN), Mr. Rajnish Upadhyay, Mr. Dheeresh Nayak, Mr. Abhishek Sharma, Mr. Amiya Abhash, Mr. Bhupendra Singh Rana, Mr. Nitesh Bonal, Mr. Gaurav Pandey, Dr. Ravi Ranjan, Mr. Ahin Banerjee, Mr. Ranjish, especially to Mr. Anupam Tiwari (ME), Mr. Jitendra Singh (ME), Dr. Akash Awale (ME), Dr. Sanjay Gupta (ME), Dr. Chandra Shekhar Kumar (MET), Dr. Shashank Singh (CSE), Dr. Ajay Singh (CSE), Dr. Tapas Das (CHE), Mr. Alok Tiwari (BME), Mr. Subham Shaw (MET), Mr. Aman Kumar Das Lal (MET), especially to Mr. Akshat Dimri (IIT KGP), and juniors: Mr. Gaurang Asthana, Mr. Deepak Prasad, Mr. Sarvesh Kumar, Mr. Mohit Chaudhary, Mr. Ashish Kumar Chouksey, Mr. Amar deep Pandey, Mr. Rohit Rodhia, Mr. Mayank Sukhija, Mr. Saroj Kanta Behera, Mr. Ashutosh Yadav, Mr. Abhinav Kumar, Mr. Nirmal Prasad, Mr. Ankit Kumar Singh, Mr. Vivek Wagh and Mr. Ravindra* for their constant support through many of the problems related to academics, research, COVID-19 pandemic situation, early morning exercises, Himalayas tracking, mess foods and many gossips during badminton in A.S.N. Bose hostel. Time spent with you all will always be laughing me. Thanks a lot to the god for having such wonderful friends like you in my life.

I express my appreciation to *Mr. Chintoo Kumar* (CSE, IIT (BHU) Varanasi), *Mr. Jatin Karwal* (IISc Bangalore) and *Mr. Rahul Karwal* (DIET, Delhi) for their continuous

support in programming and coding related problems. Without their support it couldn't be possible to reach my destination within a prescribed time period.

From my heart, I am also thankful to **Mr. Ajay** and **Mr. Sarvan**, guards in the A.S.N. Bose hostel, for constantly supplying the vegetables, fruits and other eating materials in one of the horrible situations of my life with COVID-19.

The acknowledgment is not completed till I express my gratitude to my family members. My father, **Lect. Hariom Verma** (Ret.) always taught me the lesson of honesty and punctuality throughout my life span. He always encourages me to perform exercises for physical and mental fitness. My mother, **Smt. Rani Verma** is the epitome of selfless care and is always ready to sacrifice anything for her family's happiness. She is the one who always prayed only one thing to Maa Durga that my son could complete his degree as soon as possible so she could live with me. She is the love of my life. My elder brother, **Mr. Manish Verma**, doing a Ph.D. from the National Institute of Technology Kurukshetra, and his wife, **Smt. Sushila Verma** always inspired and motivated me throughout this journey. My brother is the one who can fight with anyone for my happiness and a well-wisher from my childhood to now. My niece, **Rudranshi Verma** and **Kritika Verma**, has made me stronger, better and more fulfilled that I could ever imagine.



GAURAV VERMA

Dedicated To Highway and Airport Engineers

TABLE OF CONTENTS

CERTIFICATE	i
CERTIFICATE	ii
DECLARATION BY THE CANDIDATE	iii
COPYRIGHT TRANSFER CERTIFICATE	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	xiii
LIST OF TABLES	xvii
LIST OF FIGURES	xix
LIST OF EQUATIONS	xxiii
ABBREVIATIONS and SYMBOLS	xxvii
ABSTRACT	xxix
CHAPTER 1 INTRODUCTION	1
1.1 GENERAL	1
1.2 BACKGROUND	1
1.3 PROBLEM STATEMENT	2
1.4 OBJECTIVES OF THE STUDY	4
1.5 SCOPE OF THE STUDY	5
1.6 ORGANIZATION OF THE DISSERTATION	6
CHAPTER 2 LITERATURE REVIEW	9
2.1 GENERAL	9
2.2 LITERATURE REVIEW ON CBR OF SOIL	9
2.2.1 Factors influencing the shear strength of soil	9
2.2.2 Literature's correlation and estimation model for the CBR value	11
2.3 LITERATURE REVIEW ON MACHINE LEARNING (ML) ALGORITHMS	27
2.3.1 ML Algorithms	27
2.3.2 Application of ML Algorithms in Geotechnical Engineering	29
2.3.2.1 Genetic programming	30
2.3.2.1.1 Model development	32
2.3.2.1.2 Multi-expression programming (MEP)	39
2.3.2.2 Ensemble ML algorithm	42
2.3.2.2.1 Decision trees in ML	44
2.3.2.2.2 eXtreme gradient boosting (XGBoost) algorithm	46

2.4	LITERATURE ON DATA ANALYSIS AND DATA DIVISIONAL APPROACHES	51
2.4.1	Data Analysis	51
2.4.1.1	Statistical analysis of the dataset	52
2.4.1.1.1	Frequency distribution/ Histogram plot	52
2.4.1.1.2	Pearson's cross-correlation (R) Analysis	52
2.4.1.2	Model Performance Measurement Parameters	54
2.4.1.2.1	Coefficient of Determination (R^2)	54
2.4.1.2.2	Mean absolute error	55
2.4.1.2.3	Root mean square error	55
2.4.1.2.4	Variance accounted for	56
2.4.1.2.5	Willmott's index of agreement (IOA)	56
2.4.1.2.6	Index of scattering (IOS)	56
2.4.1.2.7	a20-index	57
2.4.2	Data Divisional Approaches	58
2.4.2.1	Statistical approach	58
2.4.2.2	K-fold cross validation approach	59
2.4.2.3	Fuzzy C-means (FCM) clustering approach	60
2.5	LITERATURE GAP	63
CHAPTER 3	LABORATORY EXPERIMENTS AND PREPARATION OF DATASETS	65
3.1	GENERAL	65
3.2	COLLECTION OF TEST SAMPLES	65
3.3	LABORATORY EXPERIMENTS	66
3.3.1	Atterberg's Limits test	67
3.3.1.1	Liquid limit	67
3.3.1.2	Plastic limit	68
3.3.1.3	Plasticity index	69
3.3.2	Sieve analysis test	69
3.3.3	Proctor compaction test	70
3.3.4	California bearing ratio test	72
3.3.4.1	Laboratory CBR testing	72
CHAPTER 4	DEVELOPMENT OF CALIFORNIA BEARING RATIO PREDICTION MODEL	79
4.1	GENERAL	79
4.2	PREPARATION OF DATASET	79

4.2.1	Frequency distribution plot for all geotechnical parameters	79
4.2.2	Descriptive statistic ranges for all geotechnical parameters	80
4.2.3	Selection of input parameters and correlation analysis	81
4.2.4	Division of dataset	83
4.2.4.1	Statistical approach	83
4.2.4.2	K-fold cross-validation approach	84
4.2.4.3	Fuzzy C-means clustering approach	85
4.3	DEVELOPMENT OF CBR PREDICTION MODEL	87
4.3.1	Multi-expression programming (MEP) model development	87
4.3.2	eXtreme gradient boosting (XGBoost) model development	88
4.4	RESULTS AND DISCUSSION	89
4.4.1	Performance of MEP model	89
4.4.1.1	Statistical details of the MEP model	89
4.4.1.2	Selection of MEP model	91
4.4.1.2.1	Ranking analysis (RA)	91
4.4.1.2.2	Overfitting analysis	94
4.4.2	Performance of XGBoost model	96
4.4.2.1	Statistical details of the XGBoost model	96
4.4.2.2	Selection of XGBoost model	99
4.4.2.2.1	Ranking analysis	99
4.4.2.2.2	Overfitting analysis	101
4.4.3	Comparative performance of selected MEP and XGBoost model	102
4.4.3.1	Visual interpretation of the results	102
4.4.3.1.1	Trend and error plot for the selected models	102
4.4.3.1.2	Regression error characteristics (REC) curve	107
4.4.3.1.3	Accuracy analysis	108
4.4.4	Validation of present and literature study models	109
4.4.4.1	Internal validation of present study models through the K-Fold CV approach	110
4.4.4.2	External validation of present study models through the literature dataset	110
4.4.4.3	External validation of literature models on present study datasets	113
4.4.4.4	Influence of soil origin on the predictive ability of various models	114
4.5	GRAPHICAL USER INTERFACE (GUI) FOR PREDICTING THE SOAKED CBR VALUE	115

CHAPTER 5 SUMMARY CONCLUSIONS AND SCOPE FOR THE FUTURE STUDY	117
5.1 GENERAL	117
5.2 SUMMARY AND CONCLUSION	117
5.3 SCOPE FOR THE FUTURE STUDIES	120
REFERENCES	123
LIST OF PUBLICATIONS	131
Published/Accepted in Journals	131
Under Review in Journals	131
Conferences	131
Appendix A. Appendix	133

LIST OF TABLES

Table 1.1 Time taken by the geotechnical parameters in the laboratory testing	3
Table 2.1 Summary of literature model for various type of soil.....	21
Table 2.2 Basis function values for MARS-L and MARS-C model (Bardhan, Gokceoglu, et al., 2021).....	25
Table 2.3 Ideal values of different statistical performance measurement parameters	57
Table 3.1. A list of sample dataset.....	66
Table 3.2. Standard load at various depths of penetration.....	78
Table 4.1 Descriptive statistics value for the geotechnical parameters of all dataset.....	81
Table 4.2 Variation in geotechnical parameters based on soil group.	82
Table 4.3 Tentative combinations of input parameters for developing the CBR prediction model	83
Table 4.4 Descriptive statistics value for the selected input and output parameters of the TR dataset obtained through a statistical approach	84
Table 4.5 Descriptive statistics value for the selected input and output parameters of TS dataset obtained through a statistical approach.....	84
Table 4.6 Descriptive statistics value for the selected input and output parameters of the TR dataset obtained through the K-fold approach.....	84
Table 4.7 Descriptive statistics value for the selected input and output parameters of TS dataset obtained through the K-fold approach	85
Table 4.8 Descriptive statistics value for the selected input and output parameters of the TR dataset obtained through the FCM approach	86
Table 4.9 Descriptive statistics value for the selected input and output parameters of TS dataset obtained through the FCM approach	86
Table 4.10 Hyper-parameters setting for MEP algorithm	88

Table 4.11 Hyper-parameters setting for XGBoost algorithm code	89
Table 4.12 Statistical performance of various MEP models for TR dataset.....	90
Table 4.13 Statistical performance of various MEP models for TS dataset	91
Table 4.14 Rank analysis for selecting the MEP model from a statistical approach.....	92
Table 4.15 Rank analysis for selecting the MEP model from the K-fold approach	93
Table 4.16 Rank analysis for selecting the MEP model from the FCM approach	93
Table 4.17 Overfitting ratio values along with the rank analysis in MEP models	95
Table 4.18 Selected combination of MEP algorithm hyper-parameters.....	95
Table 4.19 Statistical performance of various XGBoost models for TR dataset.....	97
Table 4.20 Statistical performance of various XGBoost models for TS dataset	98
Table 4.21 Rank analysis for selecting the XGBoost model from the statistical approach	99
Table 4.22 Rank analysis for selecting the XGBoost model from the K-fold approach	100
Table 4.23 Rank analysis for selecting the XGBoost model from the FCM approach	100
Table 4.24 Overfitting ratio values along with the rank analysis in XGBoost models	101
Table 4.25 Selected combination of XGBoost algorithm hyper-parameters.....	102
Table 4.26 Accuracy of selected MEP_{K-1} and XGB_{K-1} models.....	109
Table 4.27 Details of the literature dataset used to validate the present study model ..	111
Table 4.28 Descriptive statistic details for literature dataset	111
Table 4.29 Comparative performance of proposed ML algorithm models on literature dataset	112
Table 4.30 Literature models validation on the present study datasets	113
Table 4.31 Comparative performance of literature models on present study datasets ..	114

LIST OF FIGURES

Figure 1.1 Methodology adopted to develop the CBR prediction model	7
Figure 2.1 The relation of CBR against plasticity index and grading modulus (Kleyn, 1955).....	12
Figure 2.2 The relationship between CBR and plasticity index at various consistency indices (Black, 1962).....	13
Figure 2.3 The relationship between saturated and un-saturated CBR of soil at same moisture content (Black, 1962).....	14
Figure 2.4 PMGSY developed Nomograph for computing the soaked CBR value.	16
Figure 2.5 Classifications of machine learning algorithms	29
Figure 2.6 Genetic programming algorithm's flowchart	31
Figure 2.7 Genetic programming tree representation of function $5.8X - 75 + 25\sin Y$	32
Figure 2.8 Roulette wheel showing the fitness section for four different solutions	34
Figure 2.9 Roulette wheel showing the fitness section for before ranking (a) and after ranking (b).....	34
Figure 2.10 Crossover techniques in genetic algorithm	36
Figure 2.11 Tree encoding crossover.....	37
Figure 2.12 Encoded expression of MEP chromosome represented as a tree structure .	42
Figure 2.13 Different types of ensemble methods	43
Figure 2.14 Decision-tree structure	45
Figure 2.15 Graphical representation of correlation coefficient extreme cases (J. Verma, 2012)	53
Figure 2.16 Data splitting process in K-fold cross validation	60
Figure 3.1 Test methods conducted in the laboratory.....	66

Figure 3.2 Casagrande apparatus for liquid limit.....	67
Figure 3.3 Plastic limit of soil mixtures.....	69
Figure 3.4 Mechanical sieve shaker for particle size gradation analysis.	70
Figure 3.5 Wet sieve analysis through 0.075 mm sieve.	70
Figure 3.6 Modified Proctor compaction of soil samples.....	72
Figure 3.7 Preparation of materials for CBR specimen.....	73
Figure 3.8 Static compaction of the specimen through a compression machine.	74
Figure 3.9 Laboratory prepared remolded CBR specimen.	74
Figure 3.10 CBR test specimen in soaking condition and swell measurement.	76
Figure 3.11 A complete test setup of CBR loading frame.....	77
Figure 4.1 Frequency plot for all the geotechnical parameters obtained through laboratory experiments.....	80
Figure 4.2 Correlation of CBR value with gradational properties, plasticity properties and compaction parameters of all dataset.....	82
Figure 4.3 Correlation matrix for the geotechnical parameters of all dataset.....	82
Figure 4.4 Pie plot for the different soil's groups of fine-grained soil.	82
Figure 4.5 Silhouette value obtained for each of the cluster.....	86
Figure 4.6 Actual versus predicted CBR value for the TR datasets of the MEP_{K-1} model.....	103
Figure 4.7 Actual versus predicted CBR value for the TR datasets of the XGB_{K-1} model.....	103
Figure 4.8 Actual versus predicted CBR value for the TS datasets of the MEP_{K-1} model.....	103
Figure 4.9 Actual versus predicted CBR value for the TS datasets of the XGB_{K-1} model.....	103

Figure 4.10 Error distribution plot for the TR datasets of the MEP _{K-1} model	105
Figure 4.11 Error distribution plot for the TR datasets of the XGB _{K-1} model.....	105
Figure 4.12 Error histogram plot for the TR datasets of MEP _{K-1} model	105
Figure 4.13 Error histogram plot for the TR datasets of the XGB _{K-1} model	105
Figure 4.14 Error distribution plot for the TS datasets of MEP _{K-1} model	106
Figure 4.15 Error distribution plot for the TS datasets of the XGB _{K-1} model	106
Figure 4.16 Error histogram plot for the TS datasets of MEP _{K-1} model.....	106
Figure 4.17 Error histogram plot for the TS datasets of the XGB _{K-1} model.....	106
Figure 4.18 REC curve for MEP _{K-1} TR dataset, MEP _{K-1} TS dataset, XGB _{K-1} TR dataset and XGB _{K-1} TS dataset	107
Figure 4.19 Performance measure of MEP _{K-1} model through K-Fold CV approach ...	110
Figure 4.20 Performance measure of XGB _{K-1} model through K-Fold CV approach ...	110
Figure 4.21 Representation of scatter plot (a-b), error radar plot (c-d) and error histogram plot (e-f) for the MEP _{K-1} and XGB _{K-1} models on the literature datasets	113
Figure 4.22 Designed GUI for predicting the soaked CBR of fine-grained plastic soils	116

LIST OF EQUATIONS

(2.1).....	13
(2.2).....	13
(2.3).....	13
(2.4).....	21
(2.5).....	21
(2.6).....	21
(2.7).....	21
(2.8).....	21
(2.9).....	21
(2.10).....	21
(2.11).....	22
(2.12).....	22
(2.13).....	22
(2.14).....	22
(2.15).....	22
(2.16).....	22
(2.17).....	22
(2.18).....	22
(2.19).....	22
(2.20).....	23
(2.21).....	23
(2.22).....	23
(2.23).....	23

(2.24).....	23
(2.25).....	23
(2.26).....	24
(2.27).....	24
(2.28).....	24
(2.29).....	24
(2.30).....	24
(2.31).....	24
(2.32).....	24
(2.33).....	24
(2.34).....	46
(2.35).....	46
(2.36).....	46
(2.37).....	47
(2.38).....	53
(2.39).....	55
(2.40).....	55
(2.41).....	55
(2.42).....	56
(2.43).....	56
(2.44).....	56
(2.45).....	57
(2.46).....	61
(2.47).....	61
(2.48).....	61

(2.49).....	61
(3.1).....	69
(3.2).....	73
(3.3).....	78
(4.1).....	94
(4.2).....	96
(4.3).....	108
(4.4).....	108

ABBREVIATIONS and SYMBOLS

Abbreviations /Symbols	Definitions
Abbreviations	
AASHTO	American Association of State Highway and Transportation Officials
ACO	Ant Colony Optimization
AI	Artificial Intelligence
ANFIS	Adaptive -Network-Based Fuzzy Inference System
ANN	Artificial Neural Networks
ANSI	Adaptive Neuro Swarm Intelligence
ASCE	American Society of Civil Engineers
AUC	Area Under the Curve
BBO	Biogeography-Based Optimization
BF	Basis Functions
BIS	Bureau of Indian Specifications
CART	Classification and Regression Tree
CBR	California Bearing Ratio
CGP	Cartesian Genetic Programming
CI	Consistency Index
CS	Coarse Sand
CV	Cross Validation
ELM	Extreme Learning Machine
FC	Fine Content
FCM	Fuzzy C-Means Clustering
FS	Fine Sand
GA	Genetic Algorithms
GADS	Genetic Algorithm for Deriving Software
GEP	Gene Expression Programming
GMDH	Group Method of Data Handling
GP	Genetic Programming
GUI	Graphical User Interface
IFF	Improved Firefly Algorithm
IOA	Index of Agreement
IOS	Index of Scatter
IPSO	Improved Particle Swarm Optimization
IRC	Indian Road Congress
IS	Indian Standards
LGP	Linear Genetic Programming
LL	Liquid Limits
LPR	Local Polynomial Regression
MAE	Mean Absolute Error
MARS	Multivariate Adaptive Regression Splines

MDD	Maximum dry density
MEP	Multi Expression Programming
MEP _F	Multi Expression Programming for FCM data divisional approach
MEP _K	Multi Expression Programming for K-Fold data divisional approach
MEP _S	Multi Expression Programming for Statistical data divisional approach
ML	Machine Learning
MLR	Multiple Linear Regression
MoRTH	Ministry of Road Transport and Highways
MPSO	Modified Particle Swarm Optimization
NCHRP	National Cooperative Highway Research Program
NH	National Highways
NHAI	National Highway Authority of India
NHDP	National Highway Development Project
OMC	Optimum Moisture Contents
PI	Plasticity Index
PL	Plastic Limit
PSD	Particle Size Distribution
PSO	Particle Swarm Optimization
RBN	Radial Basic Network
REC	Regression Error Characteristics
RF	Random Forest
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristics
SAS	Statistical Analysis System
SC	Sand Content
SLR	Simple Linear Regression
SPSS	Statistical Package for Social Sciences
SS _{res.}	Residual Sum of Square
SVM	Support Vector Machine
TPSO	Time varying acceleration coefficients PSO
TR	Training Dataset
TS	Testing Dataset
VAF	Variance Account For
XGB	eXtreme Gradient Boosting
XGB _F	eXtreme Gradient Boosting for FCM data divisional approaches
XGB _K	eXtreme Gradient Boosting for K-Fold data divisional approaches
XGB _S	eXtreme Gradient Boosting for Statistical data divisional approaches

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 $\pm 10\%$ and $\pm 20\%$ variations, respectively, whereas for XGBoost algorithm predicted 92% and 99% observations within $\pm 10\%$ and $\pm 20\%$ 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.