Prediction Of Soil Compaction Parameters Using Machine Learning Approaches

MS GEOTECHNICAL ENGINEERING THESIS DISSERTATION

By

Muhammad Hasnain Ayub Khan

Reg No: 00000319587

Supervisor

Dr. Turab Haider Jafri

NUST Institute of Civil Engineering

School of Civil and Environmental Engineering

National University of Sciences and Technology, Islamabad, Pakistan

2022

This is to certify that thesis titled

Prediction Of Soil Compaction Parameters Using Machine Learning Approaches

Submitted by

MUHAMMAD HASNAIN AYUB KHAN

Fall 2019 MS-Geotechnical Engineering

00000319587

has been accepted towards the partial fulfilment of the requirements for the award of degree

of

Master of Science in Geotechnical Engineering

____

Dr. Badee Alshameri

HoD Geotechnical Engineering

NUST Institute of Civil Engineering (NICE)

School of Civil and Environmental Engineering (SCEE)

National University of Sciences and Technology (NUST), Islamabad, Pakistan

Thesis Acceptance Certificate

It is certified that final copy of MS thesis written by **Mr. Muhammad Hasnain Ayub Khan**, Registration No. **00000319587**, of MS Geotechnical Engineering 2019 Batch (NICE), has been vetted by undersigned, found completed in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of **MS degree in Geotechnical Engineering**.

Signature______________________________

Supervisor: Dr. Turab Haider Jafri

Date: ________________________________

Signature (**HoD**):______________________

Date: ________________________________

Signature (**Dean/Principal**): ____________

Date: _______________________________

ACKNOWLEDGEMENT

I want to express my sincere gratitude towards my research supervisor Dr. Syed Turab Haider Jafri who continuously and convincingly conveyed a spirit of hardworking and steadfastness to contrive and complete this project. Without his painstaking efforts, support and guidance, completion of this project would not have been possible.

I would also like to extend my thanks to Dr. Zamir Hussain and Dr. Aisha Shabbir. Their keen interest in my research work and their priceless efforts towards completion of my thesis work cannot be disremembered.

I am highly indebted to Dr. Badee Alshameri for all the inspiration and guidance I got from him during my studies. Last but not the least, I am extremely grateful to my parents for their love, support and hard work, thanking them for their endless patience and envouragement when it was needed the most.

ABSTRACT

The maximum dry density (MDD) and optimum moisture content (OMC) are the two important compaction parameters that are obtained using proctor tests in the laboratory, but they require energy and time. Therefore, extensive work has been done in the literature to predict these parameters rather than actually performing the proctor tests in the laboratory but either the developed models are applicable to specific soil type, specific compaction energy or the performance of the method, in terms of accuracy, is compromised when dealing with large dataset. In this study, three machine learning methods; Gene expression programming (GEP), Artificial Neural Network (ANN) and Gaussian Process Regression (GPR), were used to develop prediction models for soil compaction parameters; maximum dry density (MDD) and optimum moisture content (OMC) with higher accuracy. The database used to develop the prediction models was obtained from the literature. The dataset consists of both fine grained and coarse-grained soils; soils ranging from low plasticity to high plasticity; and compacted using different compaction energies. The performance of the developed models was evaluated based on coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) and a comparison was made with the Multi Expression Programing (MEP) model from the literature for the same database. It was found that all the new prediction models from this study performed better than the MEP model. In terms of R^2, ANN performed much better as compared to GEP and GPR. All the three developed models, in different forms, can be used to predict the compaction parameters for new datasets.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOTATIONS

ML	Machine Learning
AI	Artificial Intelligence
ANN	Artificial Neural Network
GEP	Gene Expression Programming
GPR	Gaussian Process Regression
R^2	coefficient of determination
RMSE	Root Mean Square Error
MAE	Means Absolute Error
MSE	Mean Squared Error
MDD	Maximum Dry Density
OMC	Optimum Moisture Content
CF	Fine content
CG	Gravel Content
CS	Sand Content
LL	Liquid Limit
PL	Plastic Limit
E	Compaction Energy

Chapter. 1

1. INTRODUCTION

1.1. General

The objective of this research is to present machine learning models to predict the soil compaction parameters, MDD and OMC, using artificial intelligence techniques i.e ANN, GEP and GPR with better accuracy. A compararison of the presented models with the MEP model in the literature has also been made for the same database based on the accuracy of the prediction of the models.

1.2. Background and Scope

Soil compaction is the densification of loose soil or a soil that has not been densified. Compaction of soil is an important process in construction in which the soil is compacted to reduce the voids through mechanical effort to improve its properties. Soils, when encountered as fill material, are prone to excessive settlements. The compaction process reduces the air void and thus densifies the soil. Soil compaction parameters, that are the MDD and OMC, are the two important parameters that are required to maintain the long-term performance of infrastructures i.e., building, roads etc. In the laboratory, the compaction parameters are determined by following ASTM D698 [1] for standard proctor energy and ASTM D1557 [2] for modified proctor energy. Figure 1 shows the proctor tests setup for the determination of compaction parameters.

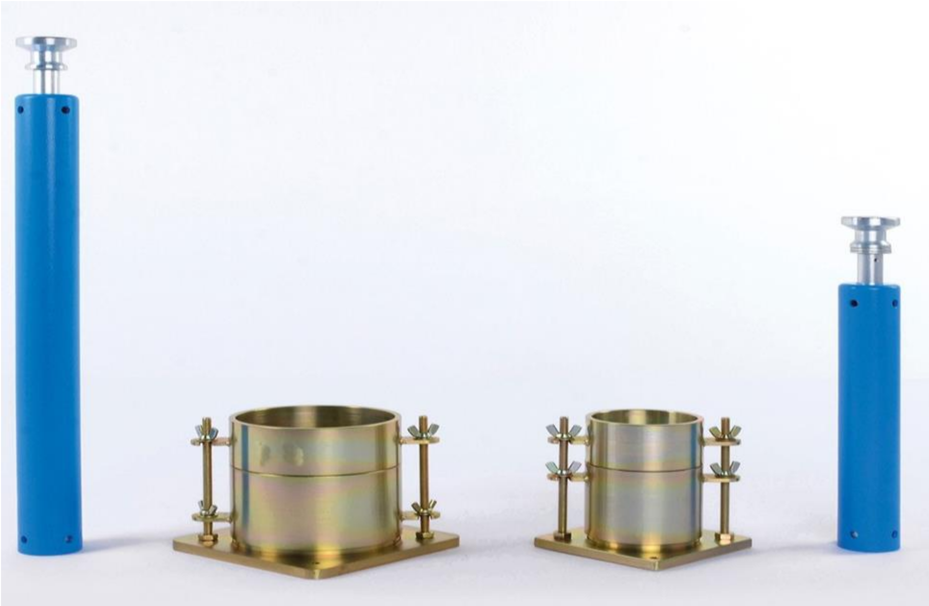


Figure 1. Standard and modified proctor tests setup

(https://www.utest.com.tr/en/23152/Proctor-Moulds-And-Rammers)

In the standard proctor test, the soil is compacted in three layers in a mould using a hammer that falls from a certain height. The diameter of the mould is 4 in and weight of the hammer is 5.5 lb. The hammer falls from a standard height of 12 in. A total of 25 number of blows from the hammer is given to each layer. The upper portion of the mold, which is detachable, is removed and the extra soil is trimmed and leveled with the mold as shown in Figure 2.

Figure 2. Trimming compacted soil in the standard proctor mold

The necessary calculations are made for the dry density and moisture content of the soil. The procedure is repeated by increasing the water content of the soil. A time comes when the further increase in the water content will result in no more increase of the dry density. If further increased, the water content will result in the decrease of the dry density. For each trail, the calculation for dry density and moisture content are made. Finally a graph is developed by plotting dry density values on vertical y-axis and their corresponding moisture content values on horizontal x-axis. The dot are joined to construct dry density-moisture content curve or compaction curve. The projection of summit point of the compaction curve on y-axis gives the value of MDD and the projection of summit point of the proctor curve on x-axis gives OMC of the soil. Typical compaction curves for standard proctor and modified proctor test are shown in Figure 3.

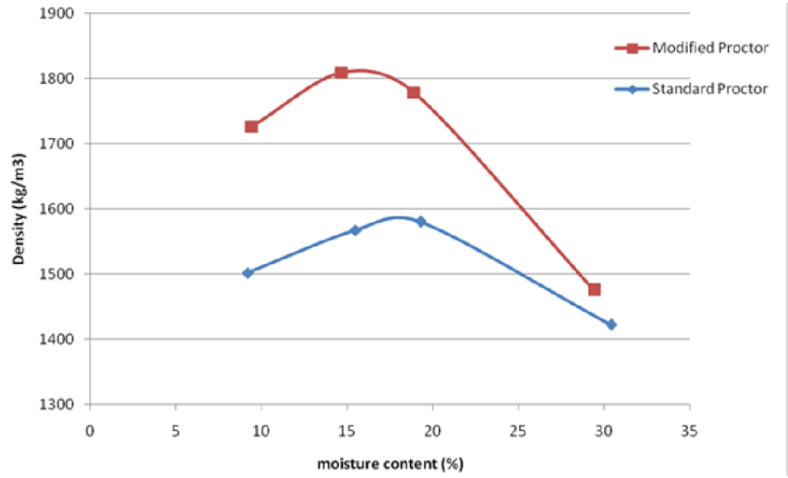


Figure 3. Compaction curve (Standard and Modified) [3]

In the modified proctor test, the compaction effort to carry out the test is more than the standard proctor test. The compaction energy is found out using Eq. (28).

$$E = (l \times b \times h \times f)/V \quad \textbf{(1)}$$

Where, l is the number of layers, b is the number of blows, h is the weight of the hammer, f is the height of fall, and V is the volume of the mold. Compaction energy, basically, is the energy consumed to carry out protor tests. Due to the requirement of more effort in modified proctor test as compared to standard proctor, it has more compaction energy than standard proctor test as shown in Figure 3. In the modified proctor test, the size of mould is 6 in. The soil is compacted in this mold in five layers. The number of blows to each layer is 25. The weight of the hammer is 10 lb. For each blow, this hammer falls from a height of 18 in. The calculation carried out for modified proctor is the same as standard proctor. Finally a compaction curve for modified proctor is obtained. The greater the compaction effort, the more is the MDD and the lesser is the OMC.

1.3. Problem Statement

The laboratory procedures to find compaction parameters at different compaction energies are time-taking and they also require extensive effort to be performed. Therefore, efforts have been made by researchers to present models to get the compaction parameters without actually performing the laboratory tests. Work has been done on both the conventional regression modeling and machine learning (ML) modeling to develop correlations between the index properties of soil and compaction parameters and predict these parameters with more accuracy. Lately, the focus has been on two things, the accuracy of the models and the types of soil on which the developed models are applicable. In this study, the database of soil was obtained from the literature [4]. The database contained 226 records of the soil properties. Models were developed using the said database and was compared with the multi expression programming (MEP) models for the same database in the literature.

1.4. Research Scope

The aim of this research is to develop models for the compaction parameters prediction. As discussed earlier, the standard laboratory procedures to find compaction parameters require both energy and effort. We have to perform proctor test with atleast five trials to construct a well defined compaction curve. There must be an easy way to determine theses parameters. Also, artificial interlligence (AI) is frequently being used in every field particulary in the engineering field. Engineers are confidently adopting the AI techniques in their respective fields. In the literaeture, the AI techniques have confidently been used for different geotechnical purposes. GPR has been used for the estimation of blast induced vibration, soil hydraulic properties, pile bearing capacity, tunnel geomechanical parameters etc. ANN techniques has been used for the estimation of UCS of soils, swell potention in clays, bearing capacity of soils etc. GEP has been used for the prediction of compression index of soils, swell potential, strength (UCS) of soils. So there is a confidence in using these methods, ANN, GPR, and GEP, in this research for MDD and OMC prediction.

Chapter. 2

2. LITERATURE REVIEW

2.1. General

Before the advent of machine learning approaches, work on developing prediction models for MDD and OMC were being carried out using conventional regression modeling. Conventional modeling refers to simple linear regression or use of excel to develop correlation between parameters. Conventional regression modeling has a disadvantage that as the number of records in a database increases, the accuracy of the developed models is compromised. To cater this, the focus has been on the machine learning approaches lately. The reliability on machine learning approaches is more as compared to conventional methods when dealing with large dataset and when the accuracy is of prime concern. Research carried out on developing prediction models for compaction parameters in the literature is discussed below for both the conventional regression modeling as well as the machine leaerning approaches.

2.2. Conventional Regression Modeling

Bera, A., & Ghosh, A. (2011) [5]

Bera, A., and Ghosh, A. predicted the compaction parameters of fine-grained soils using log linear regression modelling for a total of 5 soil samples. The R^2 values of the developed equations were 0.98 and 0.95 for MDD and OMC respectively.

Patra, C., Sivakugan, N., & Das, B. (2010) [6]

Patra, C., Sivakugan, N., and Das, B correlated the relative density (D_r) and medium grain size (D_{50}) using 55 samples of coarse-grained soil against standard proctor and modified proctor compaction energies. The R^2 values were found to be 0.964 for standard proctor energy and 0.946 for modified proctor energy

Farooq, K. et al. (2016) [7]

Farooq, K. et al. developed correlations between the compaction parameters and consistency limits for fine grained soils using 105 soil samples. They used Statistical Product and Service Solution (SPSS) software for modeling and achieved R^2 values of 0.89 and 0.88 for MDD and OMC respectively.

Blotz, L. R et al. (1998) [8]

Blotz, L. R et al. predicted the MDD and OMC of fine-grained soil for a total of 22 dataset and the R^2 values of his models ranged from 0.88 to 1.

Al-Khafaji (1993) [9]

Al-Khafaji estimated the compaction parameters of fine-grained soil using Atterberg limits with the help of curve fitting method.

Günaydın, O. (2009) [10]

Günaydın, O. performed simple regression modeling and multi regression modeling on 126 datasets of soil, containing both the fine grained soil and coarse grained soil. The accuracy (R^2) of the presented models ranged from 0.64 to 0.82.

Mujtaba, H et al. (2013) [11]

Mujtaba, H et al. developed equations for compaction parameters of sandy soil using 150 datasets of soil and the accuracy (R^2) of the developed models were 0.81 and 0.7 for MDD and OMC respectively.

Gurtug, Y. and Sridharan, A. (2004) [12]

Gurtug, Y. and Sridharan, A. used graphical method to develop linear equations for compaction parameter on 86 datasets of fine-grained soil. The R^2 ranged 0.92 to 0.99 and 0.75 to 0.99 for MDD and OMC against different compaction energies.

Di Matteo, L et al. (2009) [13]

Di Matteo, L et al. performed regression analysis on 71 datasets of fine-grained soil. The OMC was best correlated with liquid limit (LL) and specific gravity (Gs), while MDD was best correlated with plasticity index (PI) and OMC.

Sridharan, A., and Nagaraj, H. B. (2005) [14]

Sridharan, A., and Nagaraj, H. B. developed linear equations for compaction parameters using graphical method for fine-grained soils using 64 datasets and achieved accuracy, in terms of R^2, of 0.93 and 0.99 for MDD and OMC.

Saikia, A et al. (2017) [15]

Saikia, A et al. performed single regression analysis and multi regression analysis on a total of 40 dataset of fine-grained soils. The R^2 values of the presented models ranged from 069 to 0.90.

Omar, M. et al. (2003) [16]

Omar, M. et al. carried out multi regression analysis on 311 datasets of coarse-grained soil. The R^2 values of the resulting equations were found to be 0.816 and 0.68 for MDD and OMC.

2.3. Machine Learning/Artificial Intelligence Techniques

Günaydın, O. (2009) [Error! Bookmark not defined.**]** [10]

Günaydın, O. used Artificial Neural Network (ANN) to predict the compaction parameters for 126 datasets of soil containing both the fine-grained soil and coarse-grained soil. The R^2 of the best models was found to be 0.836 and 0.893 for MDD and OMC.

Jalal, F. et al. (2021) [17]

Jalal, F. et al. predicted the compaction parameters of fine-grained soil for 195 datasets using Gene Expression Programming (GEP) and Multi Expression Programming (MEP). GEP outperformed the MEP for the MDD and OMC in terms of accuracy (R^2).

Khuntia, S et al. (2015) [18]

Khuntia, S et al. used three machine learning approaches: Multivariate Adaptive Regression Splines (MARS), Artificial Neural Network (ANN) and Least Square Support Vector Machine (LS-SVM) for the prediction of compaction parameters for coarse-grained soil on 110 datasets. MARS performance was the best among the three methods with R^2 of 0.88 and 0.81 for MDD and OMC.

Sinha, S. K. and Wang, M. C. (2008) [19]

Sinha, S. K. and Wang, M. C. predicted the compaction parameters using Artificial Neural Network (ANN) on 55 datasets containing both fine-grained soil and coarse-grained soil. The R^2 for the developed models ranged from 0.92 to 0.97.

Kurnaz, T. F., and Kaya, Y. (2020) [20]

Kurnaz, T. F., and Kaya, Y. used Group Method of Data Handling (GMDH)–type neural network, support vector machine (SVM), Bayesian regularization neural network (BRNN), and extreme learning machine (ELM) to predict compaction parameters on 415 datasets on soil, both fine-grained and fine-grained. ELM outperformed the rest of the methods with R^2 of 0.89 and 0877 for MDD and OMC respectively.

Ahangar-Asr, A et al. (2011) [21]

Ahangar-Asr, A et al. developed Evolutionary Polynomial Regression (EPR) models for 57 datasets of coarse grained soil and fine grained soil to predict MDD and OMC. The R^2 was found to be 0.96 and 0.94 for MDD and OMC.

Ardakani, A., and Kordnaeij, A. (2019) [22]

Ardakani, A., and Kordnaeij, A. used Group method of data handling (GMDH)-type neural network to predict MDD and OMC for 212 datasets of fine grained soil and coarse grained soils. The R^2 of MDD was found to be 0.81 and 0.86 for training and testing and that for OMC was 0.82 and 0.92 for training and testing respectively.

Wang, H. L., and Yin, Z. Y. (2020) [4]

Wang, H. L., and Yin, Z. Y. predicted the compaction parameters using multi expression programming (MEP) for a total of 226 datasets. The datasets cover both the fine grained soil and coarse grained soil, compacted using different compaction efforts, and soils with low to high plasticity. The developed models have the R^2 values of 0.872 and 0.858 for MDD training and validation, and 0.916 and 0.923 for OMC training and validation.

Chapter. 3

3. METHODOLOGY & RESEARCH WORK

3.1. General Overview Of The Machine Learning Approaches

3.1.1. Gene Expression Programming

Gene expression programming (GEP) is inspired by biological evolution and is a type of evolutionary algorithm that was, for the first time, presented by Candida Ferreira in 2002 [23]. Just like a living organism, a GEP model learn as it keeps on changing its size and shape. GEP is preferred over genetic programming (GP) as it does not preasume any relationship [24].

In GEP, a gene stores the genetic information and the complex trees behave and evolve according to the stored genetic information in a gene. GEP is divided into the different parts such as; function, terminal , fitiness, control and condition for termination. Chromosomes in GEP are of fixed lengths containing genes. Gene in chromosome comprises head and tail. These chromosomes evolve into expression trees [25]. Multiple trees are developed with varying size and shape and these trees lead to deriving mathematical expression for prediction. An expression tree is shown in Figure 4. In the tree, c represents constants generated during modeling while d represents the input variables. The function node consists of function set while terminal node contact of variables and contstants. The functions connect the genes in a GEP chromosome forming complex expression trees and these genes define the complexity of the trees [26]. Eq. (2) is the equation derived from the expression tree in Figure 4.

$$Y = \frac{d0 + c1}{d3 + c2} + \sqrt{d2} + \sqrt{d0} \tag{2}$$

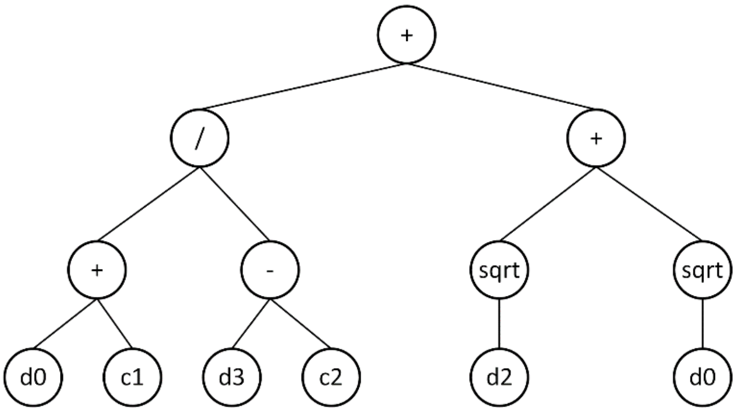


Figure 4. Expression tree (ET)

The learning in GEP starts as the chromosomes are created according to the initial population that are later on expressed as expression trees. The expression trees are then executed for each individual to check the fitness of the population. The reproduction process chooses the best individual. Different genetic operator are used for population change. The iterations keep going until the optimal solution is achieved.

3.1.2 Artificial Neural Network (ANN)

Artificial neural network is a special computing system that works just like a human brain and is based on biological neural networks [27]. The ANN comprises of three layers; the input, hidden and the output layer [28]. The input layer consists of variables that are used to develop the model, hidden layers has neurons and this layer make use of activation function to apply weights to inputs and identifies the hidden connection between the input and output layers, while the output layer consists of target variables for prediction (Figure 5)[29]. The performance of ANN is governed by the type of activation function [30] and it help in the non-linear translation of inputs to outputs [31]. The transfer function: sigmoid and hyperbolic tangent are used more often while rectified linear unit (ReLU) is the default choice of the people working in the field of machine learning [32]. These functions are given in Eq. (3) to Eq. (5).

$$Sigmoid: \sigma_{(z)} = \frac{1}{1+e^{-z}} \quad \textbf{(3)}$$

$$Hyperbolic\ tangent: \sigma_{(z)} = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \quad \textbf{(4)}$$

$$Rectified\ linear\ unit: \sigma_{(z)} = max\ (0, z) \quad \textbf{(5)}$$

Jalal, F. et al. (2021) [31] used TRANSIG and PURELIN functions to predict the swelling potential and unconfined compressive strength of expansive soils because these functions enhance the statistical indices of training dataset although they lower the validation and testing dataset accuracy [33].

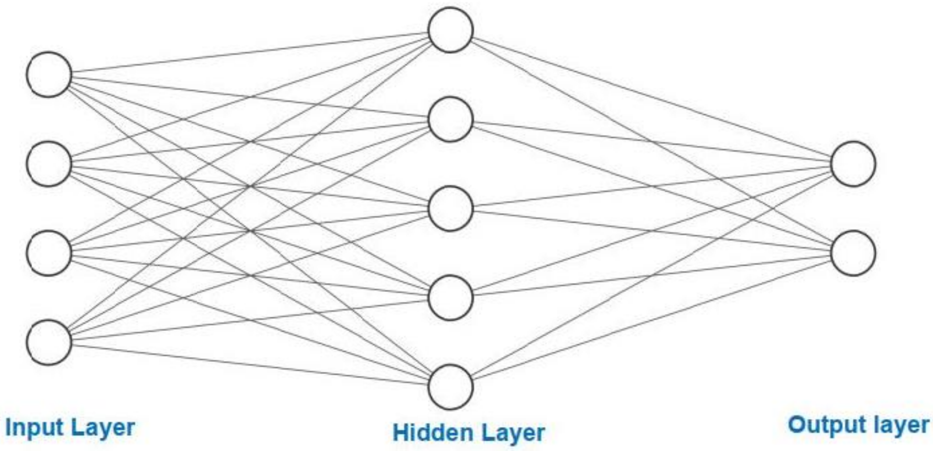


Figure 5. ANN model architecture

Feed forward neural network (FNN) based on back propagation and recurrent neural network are the two types of ANN and the former is observed to perform better as compared to the later [34]. Thus, the feed forward neural network was adopted in this research to develop the ANN-based prediction model for compaction parameters.

3.1.3 Gaussian Process Regression (GPR)

3.1.3.1 Gaussian Process Model

A Gaussian process (GP) is a non-parametric and stochastic process having a set of random variables and any finite set of those random variables holds a joint gaussian or normal distribution [35]. Gaussian Process (GP) is used to perform classification and regression [36]. A GP $t(x)$ has two functions; mean function $m(x)$ and covariance/kernel $k(x, x')$ function which are defined in the following equations.

$$m(x) = E(t(x)) \quad \textbf{(6)}$$

$$Cov\big(t(x), t(x')\big) = k(x, x'; \theta) = E((t(x) - m(x))(t(x') - m(x'))) \quad \textbf{(7)}$$

Where θ represents the hyperparameters of the model. So, a GP is given by the distribution:

$$t(x) \sim GP(m(x), k(x, x')$$

A kernel/covariance function is usually defined as "exponential squares" as given in Eq. (8) and the role of a kernel/covariance function is to develop relation between the observations [37].

$$k(x, x') = \sigma f^{2(\frac{-(x-x')^2}{el^2}} \quad \textbf{(8)}$$

A higher covariance value for the functions $h(x)$ and $h(x')$ reflects a good correlation between the functions and is obtained when $x \approx x'$while a lower covariance value shows a bad correlation between functions and it happen when x and $x'$are far from each other [36].

3.1.3.2 Gaussian Process Regression

Gaussian process regression's (GPR) main function is to perform probability distribution over all the observations. It basically is a Bayesian type non-linear regression. The regression model developed using GPR results in a Gaussian noise in addition to a function as given in Eq. (9).

$$y = f(x) + \varepsilon \quad \textbf{(9)}$$

Where y is the target output, the function $f(x)$ has the input variable(s) and ε is the Gaussian noise, where

$$\varepsilon \sim N(0, {\sigma_n}^2)$$

Thus, the distribution can be represented as [38]:

$$y \sim (0, Kf(x,x) + {\sigma_n}^2 I_n)$$

Where y represents the observation vector$[y_1, y_2, y_3, \dots, y_n]$, the input series $[x_1, x_2, x_3, \dots, x_n]$ is represented by x, ${\sigma_n}^2 I_n$ is the noise covariance matrix with I_n being the identity matrix and $Kf(x,x)$ is a symmetric of n dimension given by Eq. (10) and Eq. (11).

$$Kf(x,x) = (k_{ij})_{n \times n} \quad \textbf{(10)}$$

$$k_{ij} = \sigma f^{2(\frac{-(x-x')^2}{el^2}} \quad \textbf{(11)}$$

The value of k_{ij}, that is any element of the matrix $Kf(x,x)$, depends on x_i and x_j as discussed earlier.

Now, let y is the original output, x^* represents a new dataset and y^* the model prediction of output. The joint prior distribution of $y^*$and y for x^* is given by [38]:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N[0, \begin{bmatrix} Kf(x,x) + {\sigma_n}^2 I_n & Kf(x,x^*) \\ Kf(x,x^*)^T & Kf(x^*,x^*) \end{bmatrix}$$

Where T is the transpose of the matrix. Further details about the GPR technique can be found in the literature [36]–[38].

3.2 Machine Learning (ML) Model Development

3.2.1 Database

In this research, a comprehensive dataset of soil from the literature [4] was used for modelling. The database consists of 226 records of index properties of soils. The database includes both the coarse grained and fine grained soils and the soils ranging from low plasticity to high plasticity. Table 1 describes the statistical properties of the soil database used in this research.

Table 1. Statistical properties of the soil parameters

Parameters	Maximum	Minimum	Standard deviation	Mean
CF (%)	100	8.6	29.9	63.1
CS (%)	89	0	23.3	29.5
CG (%)	67.1	0	14.5	7.5
PL (%)	48.3	6.1	7.4	22
LL (%)	608	16	163.9	108.7
E (Kj/m^3)	2755	154.5	733.9	893.8
MDD (Mg/m^3)	2.33	1.09	0.2	1.75
OMC (%)	43.7	5.3	6	17.5

CF is the fine content (%), CS is the sand content (%), CG is the gravel content (%), PL represents plastic limit (%), LL represents liquid limit in (%), E is the compaction energy in $\mathrm{Kj/m^3}$, MDD represents the maximum dry density in $\mathrm{Mg/m^3}$ and OMC represents the optimum moisture content in %. The results of this research are compared with the results of multi expression programming (MEP) model in the literature [4] specifically as the authors also used the same database to develop prediction models. The reliability of the presented models was checked through determination coefficient (R^2), RMSE and MAE values and the mathematical functions that are used to determine these values are given in the equations Eq. (12) to Eq. (14), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x_i})^2} \quad \textbf{(12)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad \textbf{(13)}$$

$$MAE = \frac{\sum_{i=1}^{n}|x_i - y_i|}{n} \quad \textbf{(14)}$$

In the above equations, x_i is the value of original output for ith output; y_i is the value of the output predicted by the model for ith output; $\overline{x_i}$ is the average of original outputs and n represents the number of outputs. The value of R^2 lies in the range 0 – 1. The closer the value of R^2 to unity, the more accurate the model is and vise-versa. In case of RMSE and MAE, the lower value indicates the high accuracy of the models.

The correlation of individual input parameters with the output parameters must be known so that to know how strong or weak is the correlation between them. In this regard, a Person correlation matrix was developed (Table 2). Pearson correlation matrix shows the correlation of input parameters with each other and with the output parameter. The matrix is basically the representation of correlation coefficient or "R". The lesser the value of R the weaker is the correlation and vice versa. The range of R is from -1 to +1. The value "±1" shows that a very strong correlation exists between the parameters while "0" shows that no correlation exits at all between the parameters. The negative R value is the indication that the parameters are inversely related while the positive R value shows a direct relation between the parameters

[39].

Table 2.Correlation between parameters using Pearson correlation matrix

	CG	CS	CF	LL	PL	E	OMC	MDD
CG	1							
CS	0.208	1						
CF	-0.648	-0.88	1					
LL	-0.217	0.05	0.067	1				
PL	-0.283	-0.113	0.225	0.621	1			
E	0.09	0.117	-0.135	0.361	0.092	1		
OMC	-0.513	-0.569	0.692	0.405	0.694	-0.164	1	
MDD	0.553	0.515	-0.67	-0.493	-0.736	0.176	-0.953	1

A good correlation exits between the input parameters CG, CS, CF, LL, PL, and the output parameters OMC and MDD. PL has a very strong correlation among the input parameters with the OMC and MDD (R=0.694 for OMC and R=-0.736 for MDD). The correlation of E with output parameters is comparatively weak (R=-0.164 for OMC and R=0.176 for MDD). Also, CS and E are negatively correlated with the OMC while for MDD negative correlation exists for CF, LL, PL, and OMC. The order for is PL>CF>CS>CG>E and for MDD it is OMC>PL>CF>CG>CS>E.

3.2.2 Gene Expression Programming

The GEP models for soil compaction parameters' prediction were developed in GeneXproTool 5.0 [25]. GeneXproTool is a convenient modelling tool used for multiple purposes like regression, logistic regression, classification, time series prediction, and logic synthesis. It helps in modelling of large data as well as gives an option to view and access the code of the developed model in several programing languages [40]. Multiple models were developed using different parametric setting. The best model was chosen based on the accuracy i.e., R^2, RMSE and MAE. The optimal parametric setting for the GEP models is presented in Table 3.

Table 3. Parametric setting for GEP algorithm

GEP parameters	**GEP settings for MDD and OMC**
General	
Training records	158
Validation/Testing records	68
Number of chromosomes	100
Head size	8
Number of genes	5
Linking function	Addition
Function set	+, -, ÷, ×, exp, x^2
Numerical constants	
Constants per gene	10
Data type	Floating point
Ephemeral random constant	[-10,10]
Genetic operators	
Mutation	0.00138
Inversion rate	0.00138
IS transposition rate	0.00138
RIS transposition rate	0.00138
One-point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene recombination rate	0.00277

Uniform recombination	0.00755

For modelling, the dataset was partitioned into 70-30; 70% for training and 30% for validation. The data was then randomly shuffled. The models were developed individually for OMC and MDD with the same setting as shown in the Table 3. The input variables were CF, CS, CG, PL, LL, and E while the MDD and OMC were kept as output individually. Gravel content is an important parameter affecting different geotechnical properties of soil but it was not taken into consideration in the MEP models presented in the literature for the same database. Considering the importance of gravel content specially its effect on the soil compaction parameters, it was taken as an input parameter in this research. The expression trees (ET's) developed after modeling are presented in Figure 6 and Figure 7. The expression trees were later on converted to mathematical functions which are provided in Eq. (15) to Eq. (19) for MDD and Eq. (21) to Eq. (25) for OMC.

MDD:

ET-1: $$T_1 = CF + 91.84 \quad \textbf{(15)}$$

ET-2: $$T_2 = e^{\frac{-LL}{11.69}} + (CS^2 + e^{5.46}) - 7.244 \quad \textbf{(16)}$$

ET-3: $$T_3 = \frac{0.146}{PL} + CS + CG - (e^{-8.953} \times PL \times CF) \quad \textbf{(17)}$$

ET-4: $$T_4 = \frac{-0.627 \times CS \times CF}{E} - LL - 7.599 \quad \textbf{(18)}$$

ET-5: $$T_5 = LL - \frac{CG}{-5.5(CF)^2 + 5.07(CG)} \quad \textbf{(19)}$$

Thus

$$MDD = T_1 + T_2 + T_3 + T_4 + T_5 \quad \textbf{(20)}$$

OMC:

ET-1: $$T_1 = \frac{-34.49(PL+CS)}{-525-E} - \frac{PL}{3.57(CS+1.43)}$$ **(21)**

ET-2: $$T_2 = \frac{(7.956\times CG+CF)(CF)(LL)}{2.06E(E+CG)}$$ **(22)**

ET-3: $$T_3 = \frac{CS}{(CG+3.235)(CF)} + CF + 8.4$$ **(23)**

ET-4: $$T_3 = \frac{CF}{PL} - \frac{LL}{240.6} - CF - 4.25$$ **(24)**

ET-5: $$T_5 = \frac{CF(PL)}{CF+PL+38.43}$$ **(25)**

Thus

$$OMC = T_1 + T_2 + T_3 + T_4 + T_5$$ **(26)**

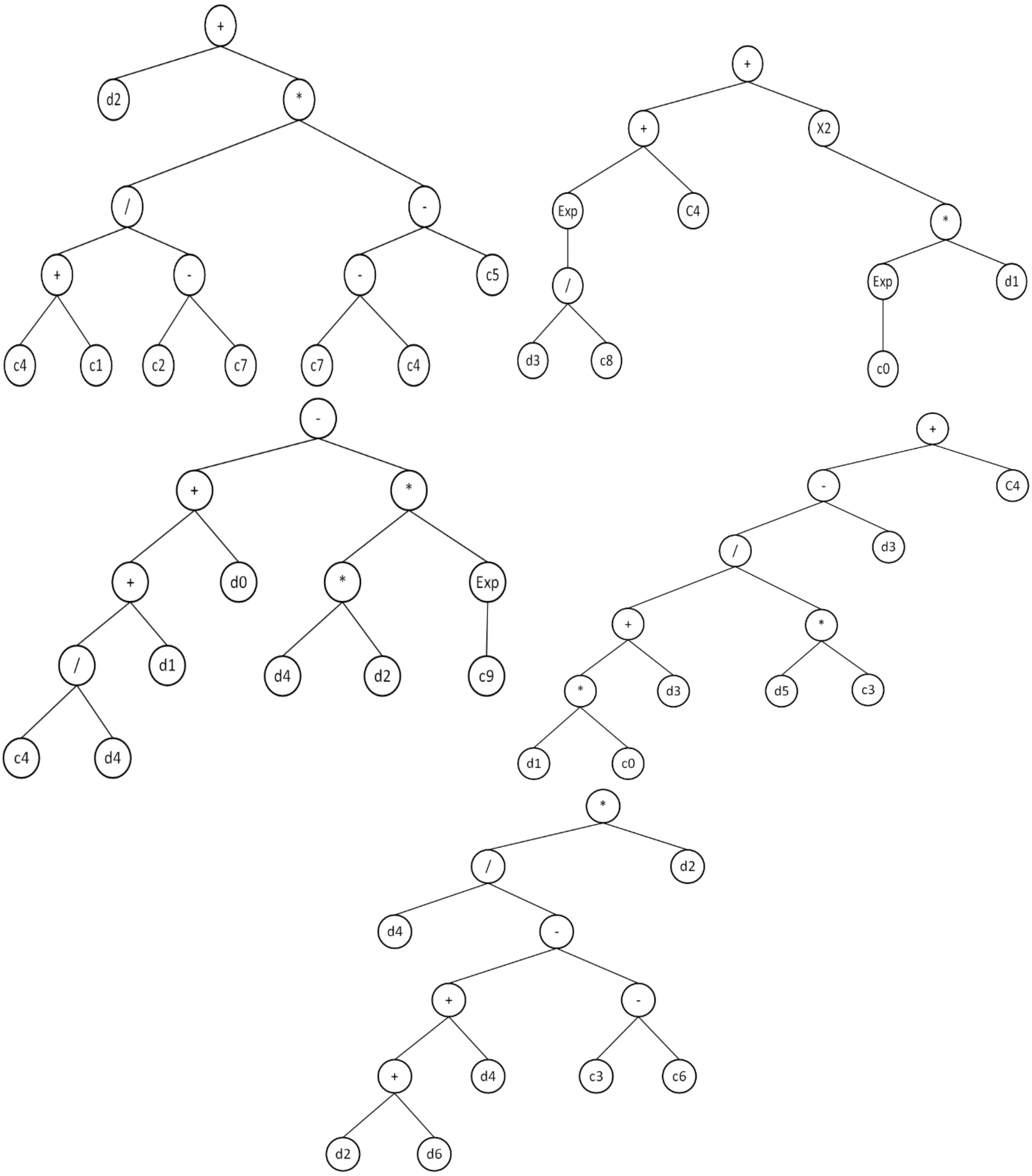


Figure 6: ET's for MDD prediction

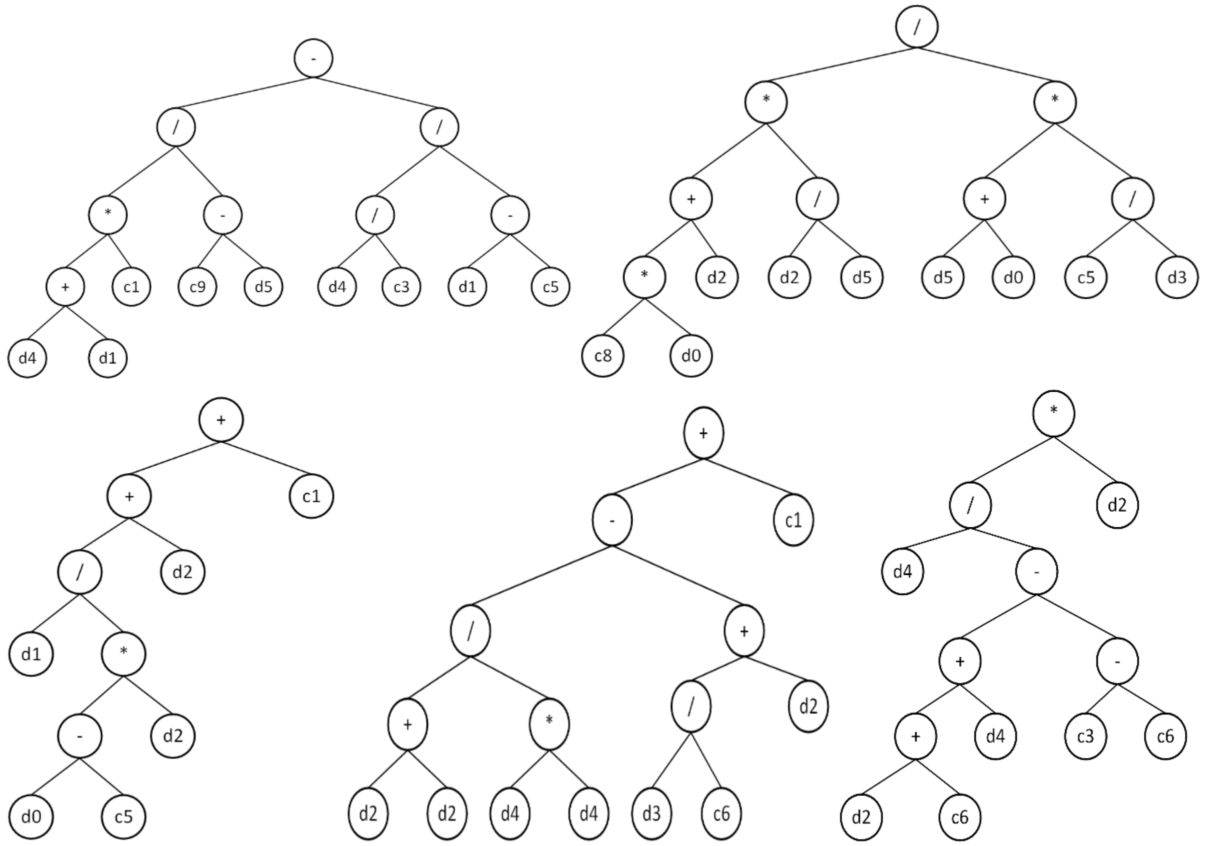


Figure 7: ET's for OMC prediction

3.2.3 Artificial Neural Network (ANN)

The ANN models were developed in MATLAB R2021a using Neural Network toolbox. 70% of the data was used to train the model while 15% was used for validation and 15% was used to test the trained model. The input layer consisted of six nodes, one for each of the input i.e. CF, CS, CG, PL, LL, and E. In the hidden layer, the total number of neurons was kept 15. MDD and OMC were kept as output in the output layer individually to create ANN models. Levenberg-Marquardt algorithm was used for modelling. The developed model architecture for MDD and OMC in ANN is shown in Figure 8. The type of network to develop prediction model was feed forward neural network with back propagation. The parametric setting of ANN models is provided in the Table 4. The model was continuously retrained until a better regression plot (R^2) and performance (minimum MSE) was achieved.

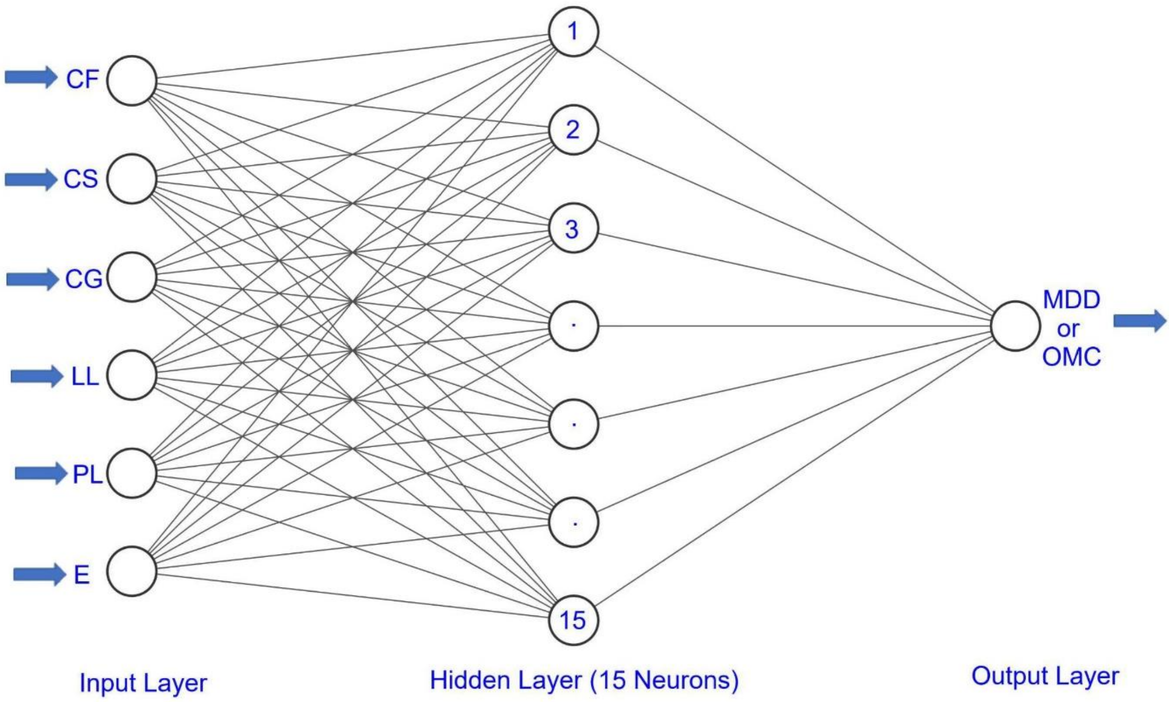


Figure 8. ANN model architecture for MDD and OMC prediction

Table 4. Parametric setting for ANN model

Parameter	Setting/Value
Dataset	
Training dataset (70%)	158
Validation dataset (15%)	34
Testing dataset (15%)	34
Network properties	
Network type	Feed-forward back-propagation
Training function	TRAINLM
Adaptation learning function	LEARNGMD
Performance function	MSE
Data division	Random basis

Number of layers	2
No of neurons in layer 1 (Hidden layer)	15
Transfer function for the hidden layer	TANSIG
Transfer function for the output layer	TANSIG
Training algorithm	Levenberg-Marquardt
Number of epochs	1000

3.2.4 Gaussian Process Regression

Modelling in GPR follows the flowchart as shown in the Figure 9. The very first step is to obtain data. Then the training dataset is employed to develop GPR model according to the discussion in 3.1.3. The testing dataset is then used to validate the trained GPR model and to check the reliability and performance of the developed GPR model in terms of accuracy.

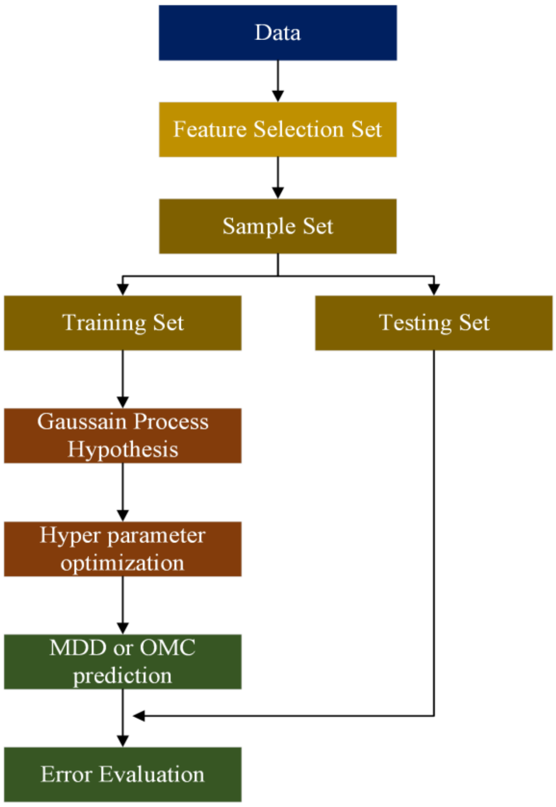


Figure 9. GPR modeling flowchart for compaction parameter prediction.

The hyper-parameters, that are the mean function and covariance function, of the GPR model govern the performance of the model. So, the training of these functions is an important part of the modeling. The maximum likelihood method, which helps in increasing the log-likelihood function, can be implemented for hyper-parameter set, Θ=[σ_n, l, σ_f], determination [41] and is given in Eq. (27) [38].

$$L = logp(y|x, \Theta) = -\frac{1}{2}log\left(det\left(K_f(x,x) + {\sigma_n}^2 I_n\right)\right) - \frac{1}{2}y^T\left[K_f(x,x) + {\sigma_n}^2 I_n\right]^{-1}y - \frac{n}{2}log2\pi \qquad \textbf{(27)}$$

The conjugate gradient method can be used to enhance the log-likelihood which is based on the gradient optimization algorithm and thus optimal solution can be obtained [38], [42].

The GPR models were developed using MATLAB R2021a for MDD and OMC. Multiple covariance-based GPR models models were developed in MATLAB R2021a i.e., Rational quadratic, Exponential, Squared exponential, Matern 5/2 and Optimizable GPR. 30-fold cross-validation was used for all the GPR model. 70% of the data was used to train the models and 30% was used to test the models. The reliability of the models, in terms of accuracy, was checked based on the R^2, MSE, MAE, and RMSE values and the model with the highest R^2 value was chosen as the representative model of GPR to predict the compaction parameters.

Chapter. 4

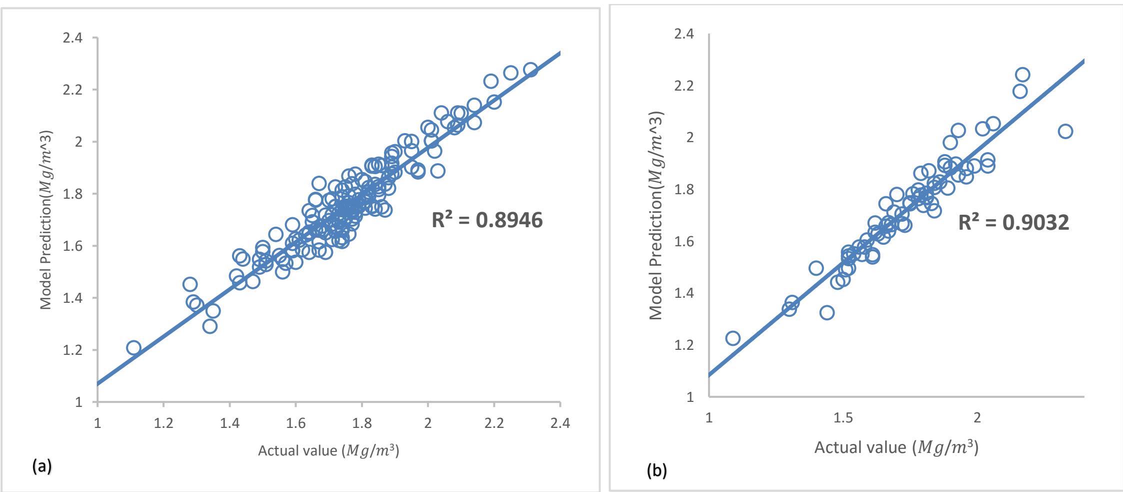
4. RESULTS AND DISCUSSION

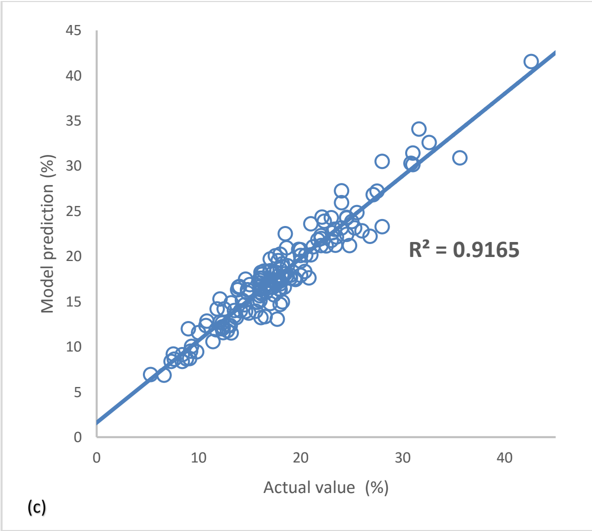
4.1. General

This study presents three machine learning (ML) models to predict the soil compaction parameters; MDD and OMC. The ML prediction models were developed using two softwares; MATLAB R2021a (for ANN and GPR) and Genexprotool 5.0 (for GEP). The performance, reliability and accuracy of the developed models, in terms of different error functions, will be discussed in this section.

4.2. Gene Expression Programming

The equations for MDD and OMC derived from the expression trees are given in the equations Eq. 14 to Eq. 25. The accuracy of GEP models is given in Figure 10. The R^2 values for training dataset and validation dataset of MDD were found to be 0.8946 and 0.9032 respectively. While for OMC, the R^2 values for training dataset and validation dataset were found to be 0.9156 and 0.9252. GEP predicted the MDD and OMC with better accuracy.





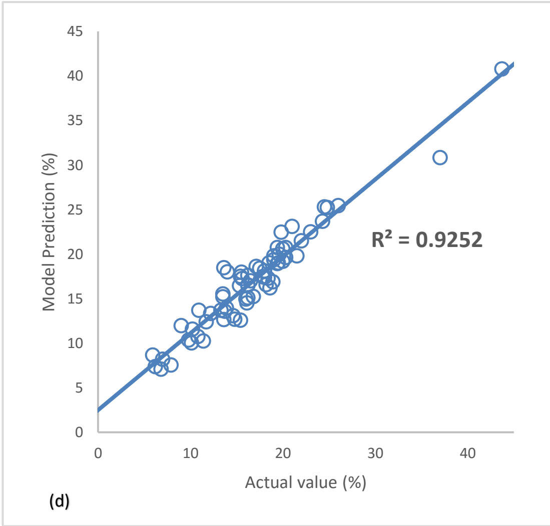


Figure 10. Accuracy of the GEP models for (a) MDD Training, (b) MDD Validation, (c) OMC Training and (d) OMC Validation.

4.3. Gaussian Process Regression

In training the GPR models, a total of 158 datapoints (70%) were used while the remaining 68 datapoints (30%) were used to validate the trained models. Multiple GPR models were developed based on the covariance function. It was found that the Optimized GPR model performed well as compared to the rest of the covariance-based GPR models i.e., Rational quadratic, Exponential, Squared exponential, and Matern 5/2. The performance of different models based on covariance is summarized in Table 5 and Table 6 in which R^2 is the coefficient of determination, MSE is the mean squared error, MAE is the mean absolute error and RMSE is the root mean square error. The Optimized GPR model outperformed the rest of the models based on the aforementioned performance evaluation parameters. Figure 12 presents the plot of the actual value of the output versus the value of the output predicted by the Optimized GPR model exported from MATLAB R2021a.

Table 5. Accuracy of the GPR models for MDD prediction

Model	Covariance function	R^2		MSE		MAE		RMSE	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	Rational quadratic	0.90	0.92	0.00354	0.00355	0.04679	0.04115	0.05950	0.0596
2	Exponential	0.89	0.93	0.00399	0.00319	0.04951	0.04167	0.06318	0.0565
3	Squared exponential	0.91	0.92	0.00346	0.00381	0.04619	0.04193	0.05886	0.0617
4	Matern 5/2	0.90	0.92	0.00349	0.00364	0.04649	0.04089	0.05912	0.0603
5	Optimized	0.93	0.94	0.00259	0.00260	0.03981	0.03953	0.05090	0.0510

Table 6. Accuracy of the GPR models for OMC prediction

Model	Covariance function	R^2		MSE		MAE		RMSE	
		Training	Validation	Training	Validation	Training	Validation	Training	Validation
1	Rational quadratic	0.88	0.92	4.2457	2.9119	1.5289	1.2515	2.0605	1.7064
2	Exponential	0.89	0.91	3.9126	3.396	1.4503	1.3243	1.978	1.8428
3	Squared exponential	0.87	0.92	4.3441	3.0328	1.551	1.2806	2.0842	1.7415
4	Matern 5/2	0.88	0.92	4.1691	2.9927	1.5239	1.2655	2.0418	1.7299
5	Optimized	0.91	0.93	3.2545	2.5014	1.3802	1.1439	1.804	1.5816

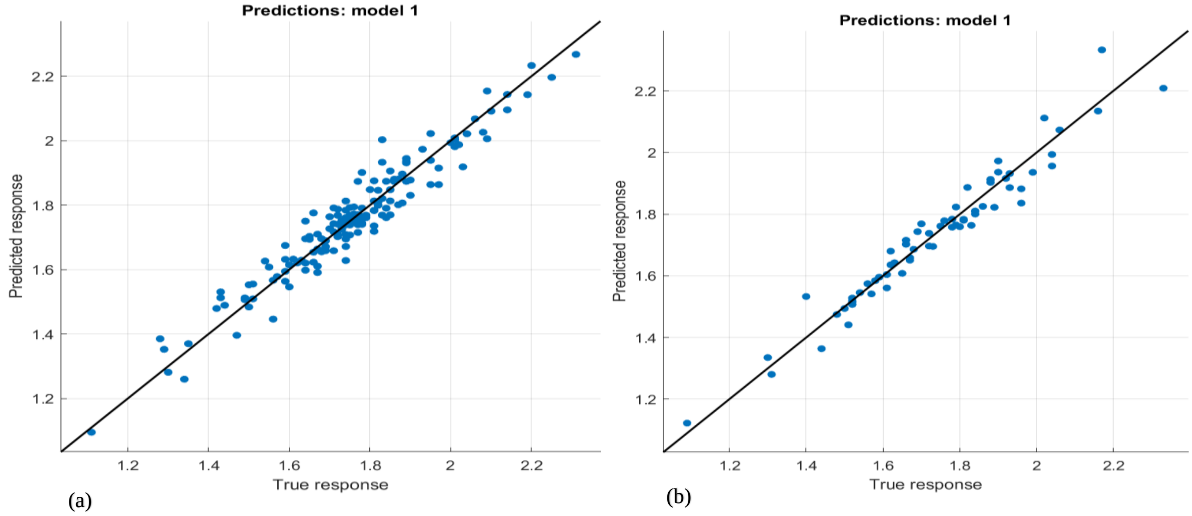


Figure 11. Predicted MDD response of the GPR model vs true response for (a) Training set (b) Validation set

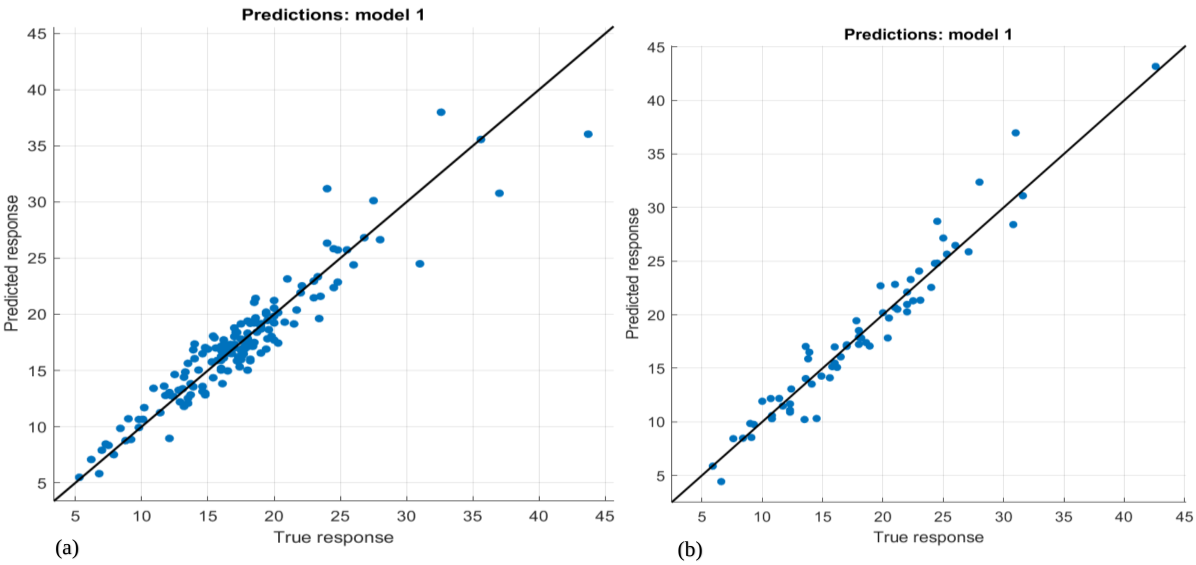


Figure 12. Predicted OMC response of the GPR model vs true response for (a) Training set (b) Validation set.

The minimum mean squared error (MSE) for optimized GPR models is shown in Figure 13 and Figure 14. The light blue points show the estimated minimum MSE at each iteration. The dark blue points represent the observed minimum MSE at each iteration, the squared red shape shows the best point hyper-parameters, and the yellow points show minimum error hyper-parameters. Bayesian optimizer was used to optimize the models. For OMC prediction as shown in Figure 13, the value of sigma was 0.035708 and zero basis function, Nonisotropic Matern 3/2 Kernel function, and 50 iterations were used which resulted in lower MSE for the prediction. For MDD prediction as shown in Figure 14, the value of sigma was 0.00013469 and zero basis function, Nonisotropic Matern 5/2 Kernel function, and 50 iterations were used which resulted the lower MSE for the prediction.

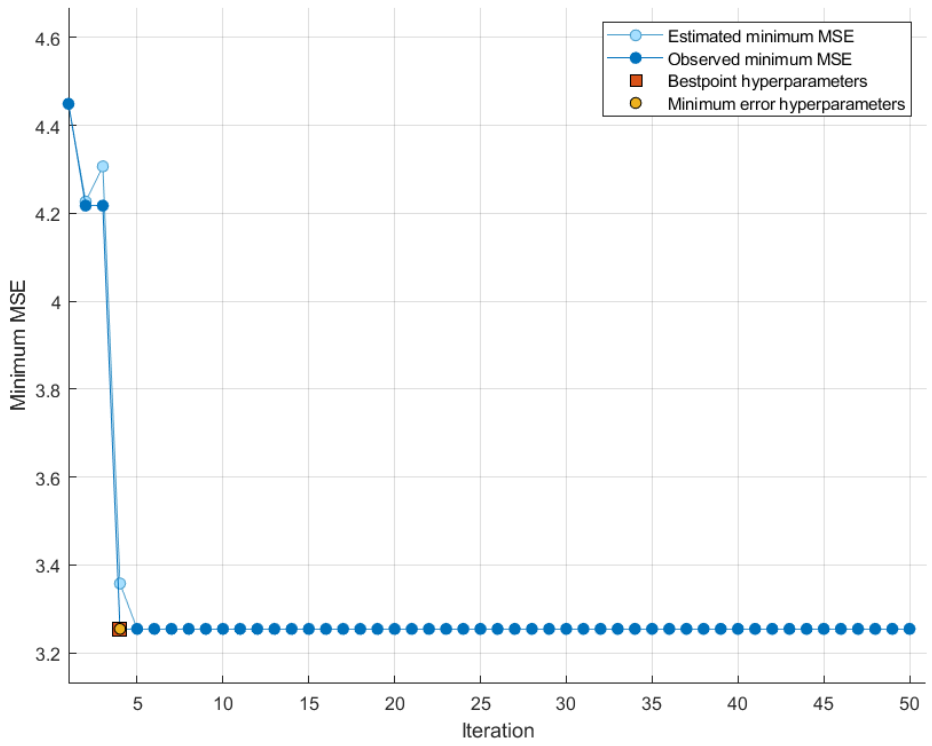


Figure 13. Minimum MSE plot for OMC (Optimized GPR model)

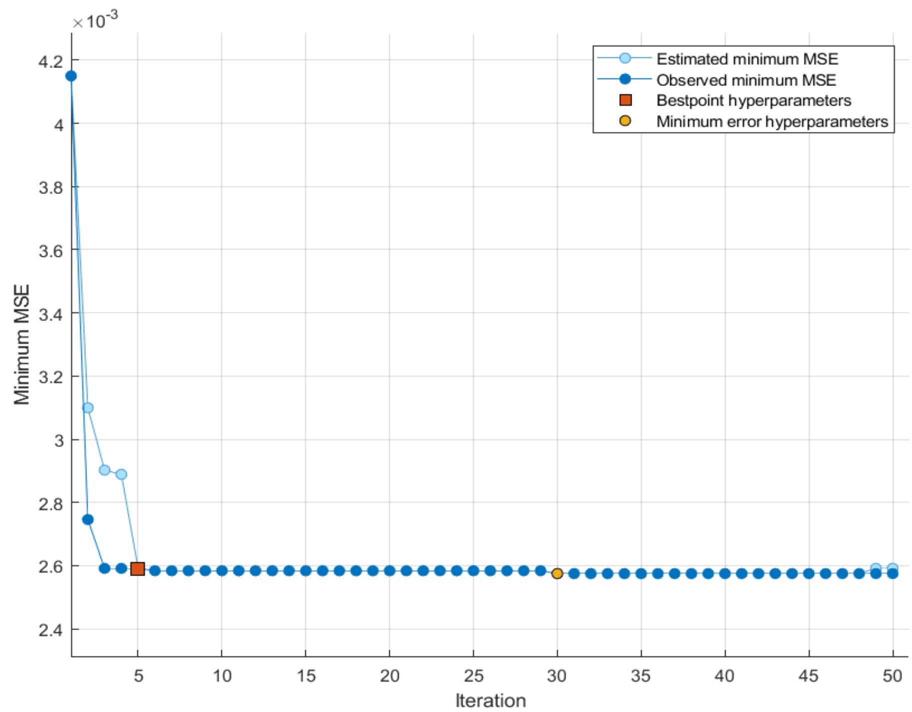


Figure 14. Minimum MSE plot for MDD (Optimized GPR Model)

4.4. Artificial Neural Network

For the prediction of compaction parameters in ANN, 158 datapoints (70%) were used to train the models while 34 datapoints (15%) were used to validate the trained models and 34 datapoints (15%) were used to test the trained models. The plots of the predicted versus the actual compaction parameter for ANN models, as exported from MATLAB R2021a are illustrated in Figure 16. For MDD prediction in ANN, the R^2 was found to be 0.9687, 0.9494, 0.94239, and 0.9620 for training, validation, testing, and all as shown in Figure 16-(a). For OMC prediction in ANN, the R^2 was found to be 0.9615, 0.96016, 0.9409, and 0.95619 for training, validation, testing, and all as shown in Figure 16-(b).

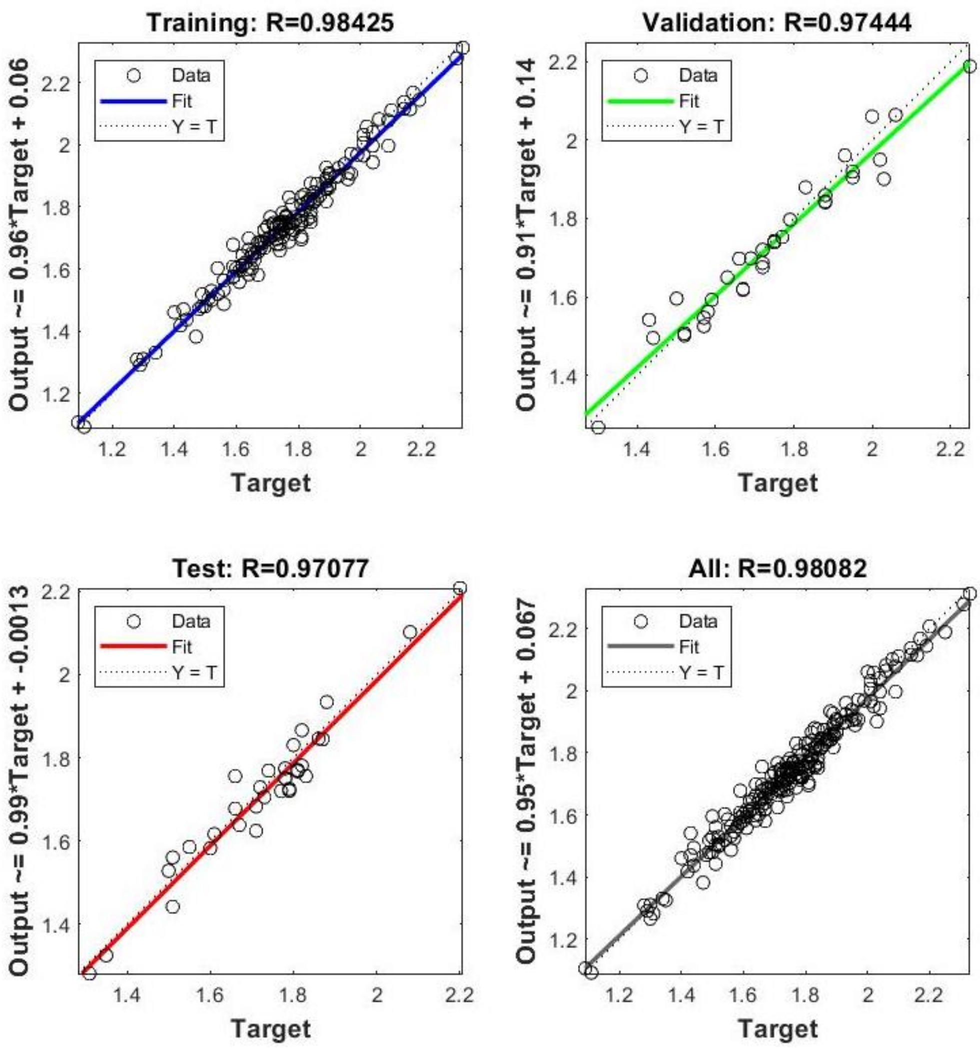


Figure 15. ANN model prediction of MDD

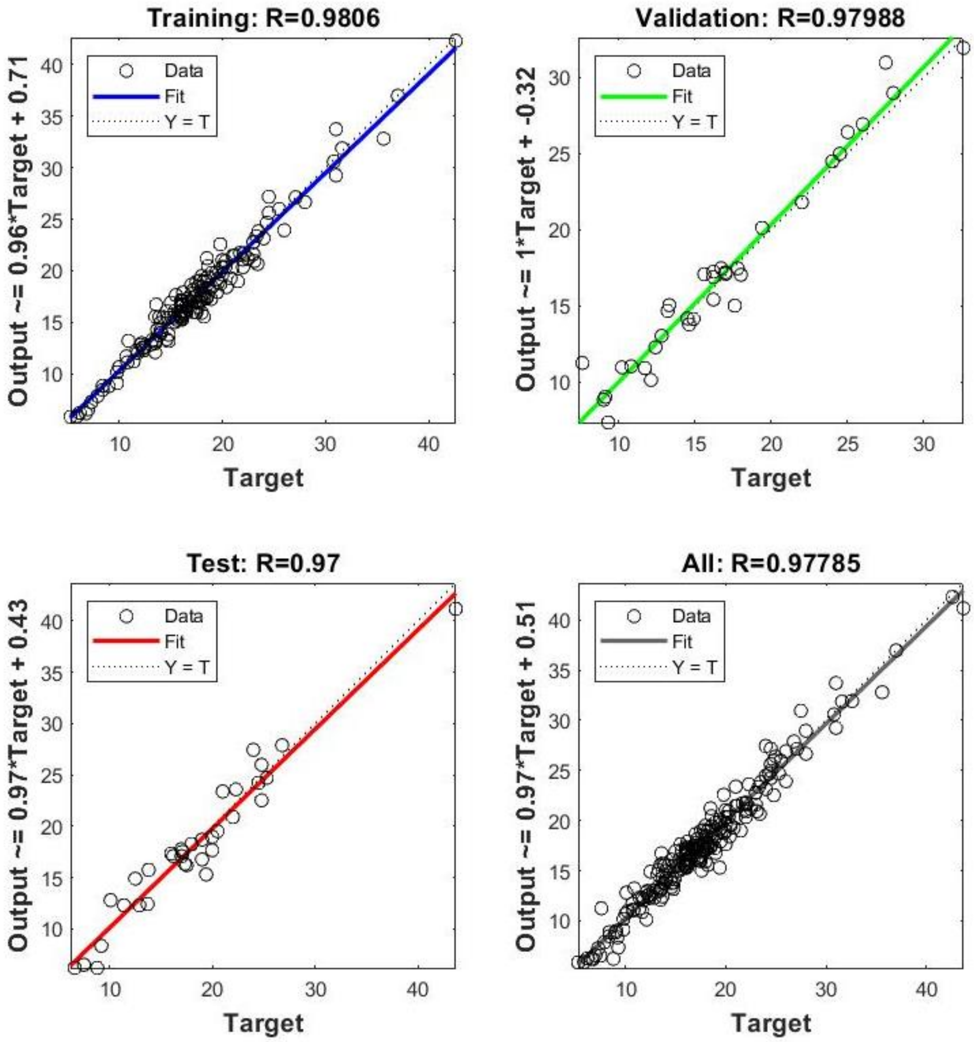


Figure 16. ANN model prediction of OMC

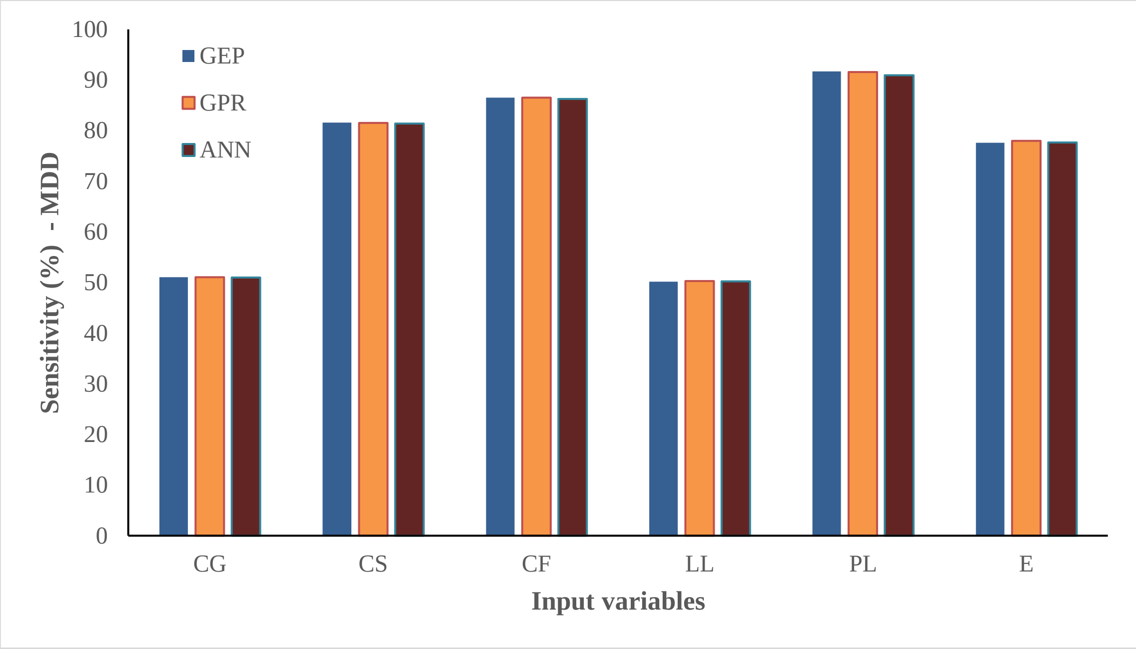
4.5 Sensitivity Analysis

To check the robustness of the presented models, and to evaluate the impact of the change in certain input parameter on the predicted output parameter, sensitivity analysis is performed. The sensitivity analysis shows how the developed model behaves when certain input parameter changes and it ranks the input parameters according to their importance with respect to the output parameter. The sensitivity analysis (SA) was performed using Eq. (28) [4], [22].

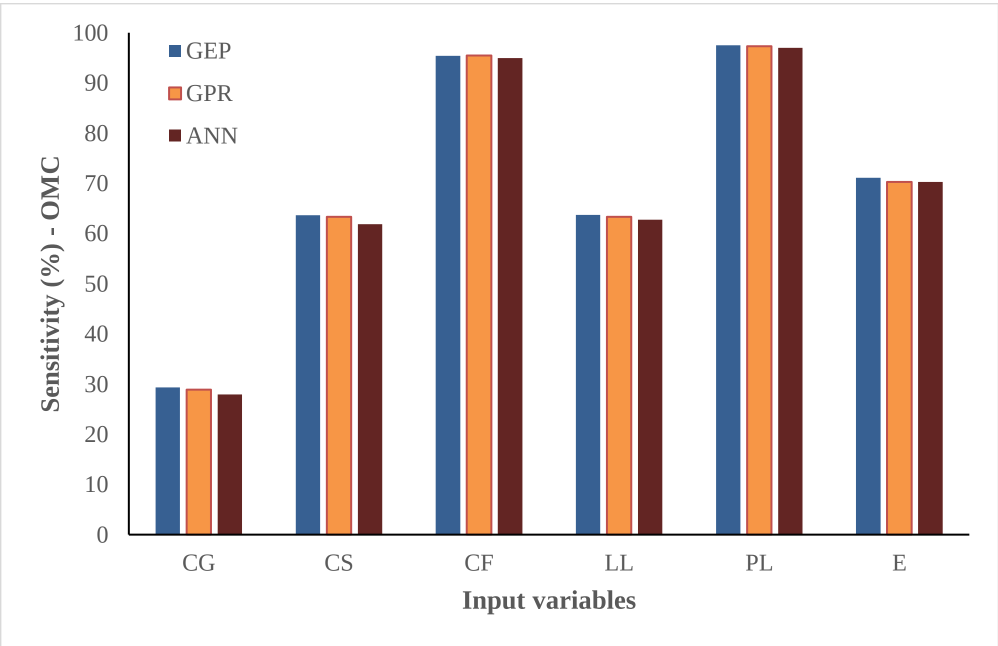
$$SA(\%) = \frac{\sum_{i=1}^{n}(h_i k_{i)}}{\sqrt{\sum_{i=1}^{n} {h_i}^2 \times \sum_{i=1}^{n} {k_i}^2}} \times 100 \qquad \textbf{(28)}$$

Where h_i is the input parameter for which SA is being calculated, k_i is predicted response of the model, n is the total number tests or records (226). The value of the sensitivity ranges between 0 and 100%. For the SA value nearer to 100% for certain input parameter, the input parameter affects the predicted output parameters significantly, while input parameter has the least effect on predicted output parameters if the SA value is close to zero. Figure 17 shows the sensitivity of input parameter for GEP, GPR and ANN prediction of MDD and OMC, respectively.

In the MDD prediction, the sensitivity of all the input variables is more than 50% in case of GEP, GPR, and ANN. While in OMC prediction, the sensitivity of all the input variables, except gravel content (CG), is more than 50% in case of GEP, GPR, and ANN. Sensitvity analysis for MDD prediction, as in Figure 17-(a) shows that, liquid limit has the least effect on the prediction while plastic limit affects the prediction significantly. The sensitivity analysis for OMC prediction is given in Figure 17-(b) that indicates that the gravel content has the least influence on the prediction while plastic limit is the most sensitive parameter and affects the prediction significantly. The ranking of the input parameters for MDD and OMC prediction, considering their influence on prediction, would be PL>CF>CS>E>CG>LL and PL>CF>E>LL>CS>CG, respectively.



(a)



(b)

Figure 17. Sensitivity analysis of input parameters for GPR and ANN prediction of (a) MDD and (b) OMC

4.6 Comparison Of The Machine Learning Methods

The comparison of the methods used in this research is provided in Table 7. For both the training and validation, ANN outperformed GEP and GPR. The ranking of the methods, based on their accuracy, would be ANN>GPR>GEP. The comparison of the methods used in this research with the MEP method is shown in Figure 18, Figure 19, and Figure 20 for R^2, MAE and RMSE, respectively.

Table 7. Comparison of the machine learning methods

		R^2		MAE		RMSE	
Method	Parameter	Training	Validation	Training	Validation	Training	Validation
MEP	MDD	0.872	0.858	0.050	0.057	0.069	0.077
	OMC	0.916	0.923	1.206	1.383	1.574	1.78
GEP	MDD	0.8945	0.9032	0.0496	0.04977	0.0618	0.0696
	OMC	0.9165	0.92519	1.2995	1.3059	1.6873	1.7367
GPR	MDD	0.93	0.94	0.03809	0.03816	0.05336	0.0507
	OMC	0.91	0.93	1.33	1.1439	1.7341	1.5816
ANN	MDD	0.9687	0.9494	-	-	-	-
	OMC	0.9615	0.9606	-	-	-	-

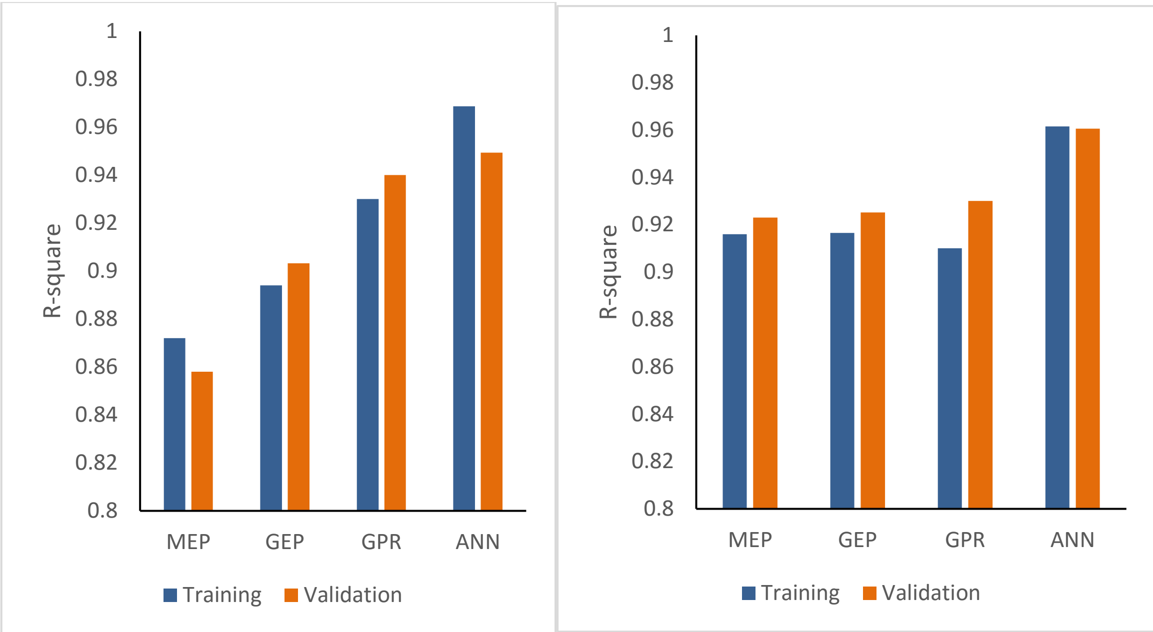


Figure 18. Comparison of the methods, based on R^2, used in this research with the MEP model for (a) MDD (b) OMC

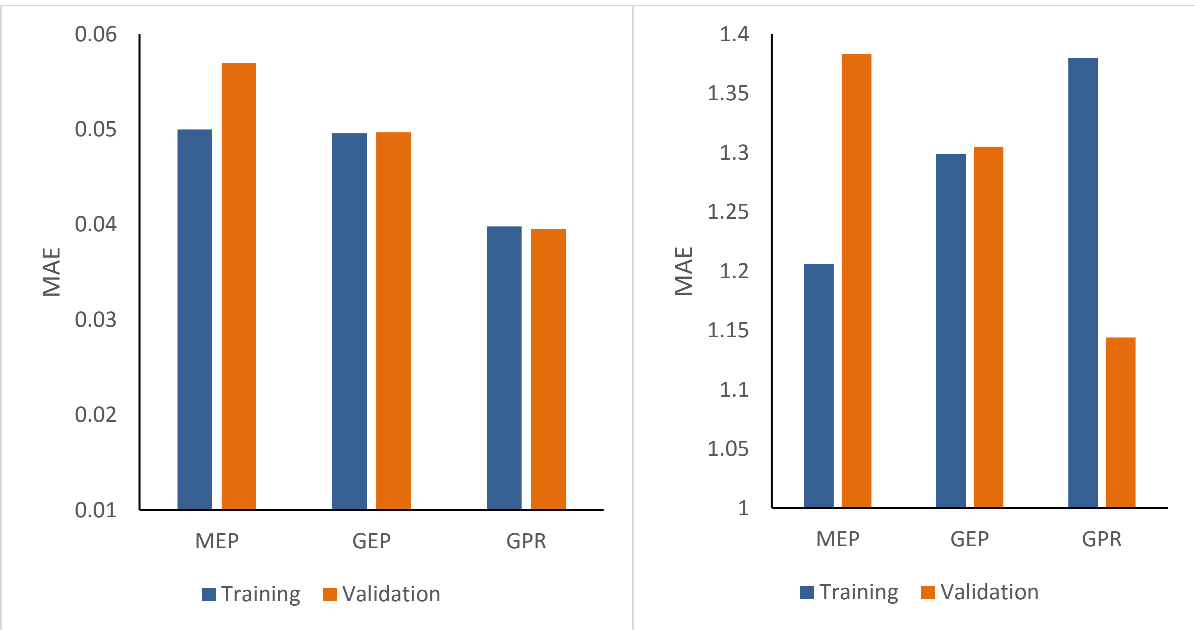


Figure 19. Comparison of the methods, based on MAE, used in this research with the MEP model for (a) MDD (b) OMC

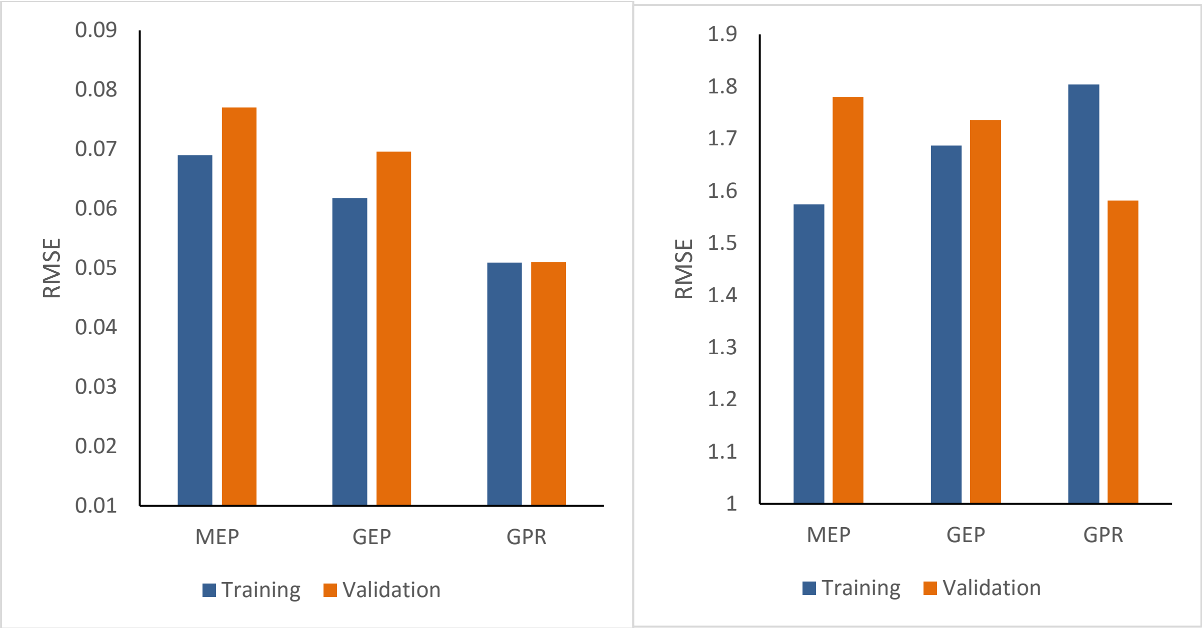


Figure 20. Comparison of the methods, based on RMSE, used in this research with the MEP model for (a) MDD (b) OMC

It can be seen that all the models performed well in term of R^2, MAE and RMSE for training and validation of MDD. But for OMC, the MAE and RMSE values increased for training and decreased for validation for GEP and GPR. The R^2 value of GEP for OMC training and validation was almost the same as MEP model in the literature and for GPR the training R^2 was slightly less than the MEP model while validation R^2 value was slightly more than it. ANN outperformed both the GEP and GPR models of this paper and MEP model in the literature in term of R^2.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion/ Summary

To save the time and effort of the practioners and engineers, an effort is made in this research to provide models to get the MDD and OMC values rather than actually performing the laboratory test. In this research, three machine learning approaches were adopted to develop models for soil compaction parameters prediction; ANN, GEP, and GPR. The database was obtained from the literature where Multi Expression Programming (MEP) models have been presented for the same database. The input parameters to develop the models were CF in %, CS in %, CG in %, PL in %, LL in %, and E in $\mathrm{Kj/m^3}$. The database consisted of 226 records of the geotechnical properties of soils. Out of 226 datapoints, 158 datapoints (70%) were used to training and 68 datapoints (30%) were used to validate the trained models for GEP and GPR. For ANN, 70% of the datapoints were used for training, 15% of the datapoints were used for validation and 15% percent of the datapoints were used testing. For ANN, the MDD and OMC were kept as output simultaneously as well as individually to develop multiple models. The performance of the developed models was checked for accuracy and reliability using R^2, MAE and RMSE.

- It was observed that all the three methods from the current study performed better than MEP in the literature.
- In terms of R2, the performance of ANN was much better than GEP and GPR. The R2 values were found to be 0.9687 and 0.9494 for training and validation for MDD as an output while for OMC as an output, the R2 values were 0.9615 and 0.9606 for training and validation.
- The performance of GPR was better as compared to GEP and MEP with R^2 of 0.93

and 0.94 for training and validation of MDD and 0.91 and 0.93 for training and validation of OMC respectively.

- In GPR, the MAE values of training set and validation set were 0.03981 and 0.03953 and the RMSE of training set and validation set 0.0509 and 0.05104, respectively for MDD. For OMC, the MAE for training and validation were found to be 1.3802 and 1.1439 and the RMSE for training and validation were found to be 1.804 and 1.5816. So, for the OMC prediction in GPR, the MAE and RMSE for validation was good as compared to training while for the MDD prediction, both the MAE and RMSE of training and validation were better than the MEP model in the literature.

- GEP had the prediction accuracy (R^2) of 0.8945 and 0.9032 for training and validation of MDD, and 0.9165 and 0.92519 for training and validation of OMC. GEP outperformed the MEP model for MDD. While for OMC, GEP prediction accuracy was almost the same as MEP model with a slight change occurring in the 3rd decimal place.

- For the MAE and RMSE of the training and validation of MDD and OMC, the GEP had the same trend as GPR i.e, for training, the MAE and RMSE was good as compared to validation in the OMC prediction while in the MDD prediction, both the training and validation MAE and RMSE were better as compared to that of MEP model in the literature.

- Sensitivity analysis revealed that the plastic limit and fine contents have significant influence on the prediction of compaction parameters, while liquid limit and gravel contents have the least influence on the prediction.

5.2. Future Recommendations

i. There are other index properties of soil that are important, easy to find and can be used as input variables for modeling such as specific gravity, unit weight etc. These index properties were not considered in this study.

ii. Feature engineering is the method of creating useful features using the existing database and changing your data in such a form that it better relates with your target feature and thus improves the accuracy of the machine learning models. Feature engineering can be done to add additional variable(s) that can help in improving the accuracy of the models.

iii. The mathematical functions developed using GEP can easily be used for future prediction but the prediction of new dataset using ANN and GPR models presented in this research require MATLAB software. To make the models user-friendly and easy to use for the practitioners, mobile application of the models can be developed where the users can make prediction on just one touch.

6. REFERENCES

[1] ASTM, 2012a. Standard Test Method for Laboratory Compaction Characteristics of Soil Using Standard Effort. ASTM D698. ASTM, West Conshohocken, PA.

[2] ASTM, 2012b. Standard Test Method for Laboratory Compaction Characteristics of Soil Using Modified Effort. ASTM D1557. ASTM, West Conshohocken, PA

[3] M. A. Hezmi, R. Saari, M. Z. Zahari, R. A. Abdullah, N. Z. M. Yunus, and A. S. A. Rashid, "Soil Water Characteristic Curves Of Compacted Kaolin For Various Initial Moisture Content," *J. Teknol.*, vol. 76, no. 2, 2015.

[4] H.-L. Wang and Z.-Y. Yin, "High performance prediction of soil compaction parameters using multi expression programming," *Eng. Geol.*, vol. 276, p. 105758, 2020.

[5] A. Bera and A. Ghosh, "Regression model for prediction of optimum moisture content and maximum dry unit weight of fine grained soil," *Int. J. Geotech. Eng.*, vol. 5, no. 3, pp. 297–305, 2011.

[6] C. Patra, N. Sivakugan, and B. Das, "Relative density and median grain-size correlation from laboratory compaction tests on granular soil," *Int. J. Geotech. Eng.*, vol. 4, no. 1, pp. 55–62, 2010.

[7] K. Farooq, U. Khalid, and H. Mujtaba, "Prediction of compaction characteristics of fine-grained soils using consistency limits," *Arab. J. Sci. Eng.*, vol. 41, no. 4, pp. 1319–1328, 2016.

[8] L. R. Blotz, C. H. Benson, and G. P. Boutwell, "Estimating optimum water content and maximum dry unit weight for compacted clays," *J. Geotech. Geoenvironmental Eng.*, vol. 124, no. 9, pp. 907–912, 1998.

[9] A. N. Al-Khafaji, "Estimation of soil compaction parameters by means of Atterberg limits," *Q. J. Eng. Geol. Hydrogeol.*, vol. 26, no. 4, pp. 359–368, 1993.

[10] O. Günaydın, "Estimation of soil compaction parameters by using statistical analyses and artificial neural networks," *Environ. Geol.*, vol. 57, no. 1, p. 203, 2009.

[11] H. Mujtaba, K. Farooq, N. Sivakugan, and B. M. Das, "Correlation between gradational parameters and compaction characteristics of sandy soils," *Int. J. Geotech. Eng.*, vol. 7, no. 4, pp. 395–401, 2013.

[12] Y. Gurtug and A. Sridharan, "Compaction behaviour and prediction of its characteristics of fine grained soils with particular reference to compaction energy,"

Soils Found., vol. 44, no. 5, pp. 27–36, 2004.

[13] L. Di Matteo, F. Bigotti, and R. Ricco, “Best-fit models to estimate modified proctor properties of compacted soil,” *J. Geotech. geoenvironmental Eng.*, vol. 135, no. 7, pp. 992–996, 2009.

[14] A. Sridharan and H. B. Nagaraj, “Plastic limit and compaction characteristics of finegrained soils,” *Proc. Inst. Civ. Eng. Improv.*, vol. 9, no. 1, pp. 17–22, 2005.

[15] A. Saikia, D. Baruah, K. Das, H. J. Rabha, A. Dutta, and A. Saharia, “Predicting compaction characteristics of fine-grained soils in terms of Atterberg limits,” *Int. J. Geosynth. Gr. Eng.*, vol. 3, no. 2, pp. 1–9, 2017.

[16] M. Omar, A. Shanableh, A. Basma, and S. Barakat, “Compaction characteristics of granular soils in United Arab Emirates,” *Geotech. Geol. Eng.*, vol. 21, no. 3, pp. 283–295, 2003.

[17] F. E. Jalal, Y. Xu, M. Iqbal, B. Jamhiri, and M. F. Javed, “Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms,” *Transp. Geotech.*, vol. 30, p. 100608, 2021.

[18] S. Khuntia, H. Mujtaba, C. Patra, K. Farooq, N. Sivakugan, and B. M. Das, “Prediction of compaction parameters of coarse grained soil using multivariate adaptive regression splines (MARS),” *Int. J. Geotech. Eng.*, vol. 9, no. 1, pp. 79–88, 2015.

[19] S. K. Sinha and M. C. Wang, “Artificial neural network prediction models for soil compaction and permeability,” *Geotech. Geol. Eng.*, vol. 26, no. 1, pp. 47–64, 2008.

[20] T. F. Kurnaz and Y. Kaya, “The performance comparison of the soft computing methods on the prediction of soil compaction parameters,” *Arab. J. Geosci.*, vol. 13, no. 4, pp. 1–13, 2020.

[21] A. Ahangar-Asr, A. Faramarzi, N. Mottaghifard, and A. A. Javadi, “Modeling of permeability and compaction characteristics of soils using evolutionary polynomial regression,” *Comput. Geosci.*, vol. 37, no. 11, pp. 1860–1869, 2011.

[22] A. Ardakani and A. Kordnaeij, “Soil compaction parameters prediction using GMDH-type neural network and genetic algorithm,” *Eur. J. Environ. Civ. Eng.*, vol. 23, no. 4, pp. 449–462, 2019.

[23] C. Ferreira, “Gene expression programming in problem solving,” in *Soft computing and industry,* Springer, 2002, pp. 635–653.

[24] A. Johari, G. Habibagahi, and A. Ghahramani, “Prediction of soil–water characteristic curve using genetic programming,” *J. Geotech. Geoenvironmental Eng.*, vol. 132, no. 5, pp. 661–665, 2006.

[25] C. Ferreira, “Gene expression programming: a new adaptive algorithm for solving problems,” *arXiv Prepr. cs/0102027*, 2001.

[26] Z.-L. Cheng, W.-H. Zhou, and A. Garg, “Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree,” *Eng. Geol.*, vol. 268, p. 105506, 2020.

[27] P. G. Asteris and V. G. Mokos, “Concrete compressive strength using artificial neural networks,” *Neural Comput. Appl.*, pp. 1–20, 2019.

[28] N. N. Kourgialas, Z. Dokou, and G. P. Karatzas, “Statistical analysis and ANN modeling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed,” *J. Environ. Manage.*, vol. 154, pp. 86–101, 2015.

[29] T.-T. Le, B. T. Pham, V. M. Le, H.-B. Ly, and L. M. Le, “A robustness analysis of different nonlinear autoregressive networks using Monte Carlo simulations for predicting high fluctuation rainfall,” in *Micro-electronics and Telecommunication Engineering*, Springer, Singapore, 2020, pp. 205–212.

[30] Y. Koçak and G. Ü. Şiray, “New activation functions for single layer feedforward neural network,” *Expert Syst. Appl.*, vol. 164, p. 113977, 2021.

[31] F. E. Jalal, Y. Xu, M. Iqbal, M. F. Javed, and B. Jamhiri, “Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP,” *J. Environ. Manage.*, vol. 289, p. 112420, 2021.

[32] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, “Machine learning methods for optical communication systems and networks,” in *Optical fiber telecommunications VII*, Elsevier, 2020, pp. 921–978.

[33] M. Tahani, M. Vakili, and S. Khosrojerdi, “Experimental evaluation and ANN modeling of thermal conductivity of graphene oxide nanoplatelets/deionized water nanofluid,” *Int. Commun. Heat Mass Transf.*, vol. 76, pp. 358–365, 2016.

[34] D. Van Dao, H.-B. Ly, H.-L. T. Vu, T.-T. Le, and B. T. Pham, “Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete,” *Materials (Basel).*, vol. 13, no. 5, p. 1072, 2020.

[35] C. K. I. Williams and C. E. Rasmussen, “Gaussian Processes for Machine Learning (MA: 2. MIT press Cambridge),” 2006.

[36] A. Mahmoodzadeh *et al.*, “Tunnel geomechanical parameters prediction using Gaussian process regression,” *Mach. Learn. with Appl.*, vol. 3, p. 100020, 2021.

[37] M. Ebden, “Gaussian processes for regression: A quick introduction,” *Website Robot.*

Res. Gr. Dep. Eng. Sci. Univ. Oxford, vol. 91, pp. 424–436, 2008.

[38] D. Yang, X. Zhang, R. Pan, Y. Wang, and Z. Chen, “A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve,” *J. Power Sources*, vol. 384, pp. 387–395, 2018.

[39] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton, “Effective use of Pearson’s product–moment correlation coefficient,” *Anim. Behav.*, vol. 93, pp. 183–189, 2014.

[40] S. Hanandeh, A. Ardah, and M. Abu-Farsakh, “Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula,” *Transp. Geotech.*, vol. 24, p. 100358, 2020.

[41] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, “Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression,” *Microelectron. Reliab.*, vol. 53, no. 6, pp. 832–839, 2013.

[42] C. J. Moore, A. J. K. Chua, C. P. L. Berry, and J. R. Gair, “Fast methods for training Gaussian processes on large datasets,” *R. Soc. open Sci.*, vol. 3, no. 5, p. 160125, 2016.