



Prediction of Geotechnical Properties of Soil using Artificial Intelligence Framework

Jitendra Khatti, Kamaldeep Singh Grover

Abstract: The present research work is carried out to predict the geotechnical properties (consistency limits, OMC, and MDD) of soil using AI technologies, namely regression analysis (RA), support vector machine (SVM), Gaussian process regression (GPR), artificial neural networks (ANNs), and relevance vector machine (RVM). The models of machine learning (SVM, GPR), hybrid learning (RVM), and deep learning (ANNs) are constructed in MATLAB R2020a with different configurations. The models of RA are built using the Data Analysis Tool of Microsoft Excel 2019. The input parameters of AI models are gravel, sand, silt, and clay content. The correlation coefficient is calculated for pair of soil datasets. The correlation shows that sand, silt, and clay content play a vital role in predicting soil's liquid limit and plasticity index. The performance of constructed AI models is compared to determine the optimum performance models. The limited datasets of soil are used in this study. Therefore, artificial neural networks and relevance vector machines could not perform well. Based on the performance of AI models, the Gaussian process regression outperformed the RA, SVM, ANNs, and RVM AI technologies. Hence, the GPR AI approach can predict the geotechnical properties of soil by gravel, sand, silt, and clay content. The Monte-Carlo global sensitivity analysis is also performed, and it is observed that the prediction of geotechnical properties of soil is affected by sand and clay content

Keywords: Consistency limits, Deep Learning, Geotechnical Properties, Hybrid Learning, Machine Learning,

I. INTRODUCTION

The properties of soil play an important role in any Civil Engineering project. The soil properties are specific gravity (G_s), plasticity index (PI), liquid limit (LL), plastic limit (PL), maximum dry density (MDD), optimum moisture content (OMC), shear strength, etc. The pycnometer and Casagrande test apparatus are used to determine the liquid limit of soil. MDD and OMC are the compactive parameters of soil that are experimentally determined by the most popular apparatus, namely standard proctor test and modified proctor test. The modified and standard proctor tests are heavy and light compaction tests, respectively [2]. The laboratory procedures of determining the geotechnical properties of soil are time-consuming, lengthy, and require human resources.

Manuscript received on November 16, 2021.

Revised Manuscript received on November 22, 2021.

Manuscript published on November 30, 2021.

* Correspondence Author

Jitendra Khatti*, PhD Fellow, Department of Civil Engineering, Rajasthan Technical University, Kota (Rajasthan) India. E-mail. jitendrakhatti197@gmail.com

Dr. Kamaldeep Singh Grover, Professor, Department of Civil Engineering, Rajasthan Technical University, Kota (Rajasthan) India. E-mail. ksgrover@rtu.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Many investigators applied various methodologies, namely regression analysis, and artificial intelligence, to determine or predict the geotechnical properties of soil. The simple and multiple regression analysis is the statistical tool used to draw correlations between the pair of datasets and predictions. The correlation between clay and the permeability of soil gives unsatisfactory results. The satisfactory results can be achieved by drawing a correlation between fine content (silt + clay) and permeability of the soil [1]. The index properties and compaction properties play an essential role in predicting the CBR of soil [28] [10] [21] [12] [6]. The CBR of soil has a perfect correlation with the plasticity index of soil [16]. The linear, polynomial, logistic, and exponential analyses are the type of regression analysis. Polynomial regression analysis can achieve a good prediction of compaction parameters, strength parameters, and permeability of the soil [7]. The internal friction angle and cohesion of soil are affected by plastic limit and natural water content [13]. The preliminary prediction of LL and PI can be achieved using multiple regression analysis [11].

The artificial neural network, support vector machine, Gaussian regression, decision tree are the AI approaches used to solve regression and classification problems in a different field. The ANNs have the potential to predict the CBR of soil using soil properties [28]. The ANNs also compute the permeability, compaction parameters, and internal friction angle with high accuracy [27]. The quality and quantity of datasets play a vital role in the prediction. The best prediction using an artificial neural network can be achieved using large datasets [25]. The artificial neural network predicts swelling properties of soil in acceptable limits [20]. The well-established training dataset predicts the CBR of soil with high accuracy [10] [5]. The artificial neural network has the potential to predict the G_s , OMC, and MDD of soil using the consistency limits, coefficient of uniformity & curvature of soil [29]. Many researchers mapped the comparison between statistical methods and AI approaches based on their performances and results. The multiple regression analysis outperformed the ANNs in predicting the CBR of soil using index properties, density, and other properties [28] [10]. The ANNs outperformed the empirical [25] and ordinary least squares [20] methods in predicting compaction parameters. The preliminary prediction of Atterbergs limits can be carried out by applying SVM [22]. The SVM predicted consistency limits of soil precisely than the multiple regression analysis [11]. Most of the studies were carried out to predict soil CBR, compaction, and strength parameters using consistency limits, particle content, and other properties of soil.

Prediction of Geotechnical Properties of Soil using Artificial Intelligence Framework

The published work shows that the consistency limits and compactive parameters of soil are affected by gravel, sand, silt, and clay content. The literature also shows that artificial intelligence approaches have not been frequently applied to predict consistency limits, OMC, and MDD using gravel, sand, silt, and clay content. The present study has the following aims –

- Draw the correlation between input variables (gravel, sand, silt, and clay content) and output variables (LL, PL, PI, OMC, and MDD).
- Employ the SVM, GPR, ANN, and RVM models in MATLAB R2020a and RA in Microsoft Excel 2021 to predict LL, PL, PI, OMC, and MDD.
- Draw the performance comparison between machine learning, hybrid learning, and deep learning models for obtaining the optimum performance model.
- Compare the performance of employed AI models with published models in the literature.
- Perform the Monte-Carlo global sensitivity analysis to study the effect of input variables on output variables.

II. METHODOLOGY

In the present research work, the machine learning, hybrid learning, and deep learning models have been developed to predict the LL, PL, PI, OMC, and MDD of soil. Deep learning is a subset of machine learning, and hybrid learning is the advanced subset of machine learning and deep learning. The Support Vector Machine and Gaussian Process Regression is machine learning; the artificial neural network is deep learning, and the relevance vector machine is hybrid learning approaches. A brief of constructed models is discussed below.

2.1 Regression Analysis

Regression analysis is a set of processes that computes the relationship between dependent variables and one or more independent variables. Linear regression is the primary form of regression analysis. The linear regression analysis is carried out for two variables, and multiple regression analysis is carried out for more than two variables. The regression analysis is used for predicting or forecasting in machine learning. The regression analysis is also used for mapping the relationship between dependent and independent variables. The compactive parameters and consistency limits have been predicted using regression analysis in the present study. The gravel, sand, silt, and clay content are used as input parameters in predicting soil's properties. The multiple regression models of liquid limit, plastic limit, plasticity index, OMC, and MDD are designated as MLR_LL, MLR_PL, MLR_PI, MLR_OMC, and MLR_MDD, respectively. The multiple regression models have been developed using the Data Analysis Tool of Microsoft Excel 2021 in this study. The following equations have been derived while developing the MLR models. Equation for liquid limit

$$MLR_{LL} = 2.9037 + 0.1187 * G + 0.3331 * S + 0.0135 * M + 0.9897 * C \quad (1)$$

Equation for plastic limit

$$MLR_{PL} = 1.5223 + 0.1235 * G + 0.285 * S + 0.098 * M + 0.2832 * C \quad (2)$$

Equation for plasticity index

$$MLR_{PI} = 1.375 - 0.005 * G + 0.04818 * S - 0.084 * M + 0.7066 * C \quad (3)$$

Equation for optimum moisture content

$$MLR_{OMC} = 24.509 - 0.053 * G - 0.182 * S - 0.104 * M + 0.099 * C \quad (4)$$

Equation for maximum dry density

$$MLR_{MDD} = 1.9445 - 0.0001 * G + 0.0003 * S + 0.0002 * M - 0.007 * C \quad (5)$$

The equations 1 to 5 have been used to predict the geotechnical properties, namely LL, PL, PI, OMC, and MDD of soil.

2.2 Support Vector Machine

The support vector machine is based on supervised learning used to solve the problem of classification and prediction [8]. The SVM approach is based on kernel function, namely Gaussian, linear, quadratic, and cubic kernel. The SVM models of LL, PL, PI, OMC, and MDD have been constructed using the Regression Learner Tool of MATLAB R2020a. The configurations of constructed SVM models are given in Table 1.

Table 1 – Configuration of constructed SVM models

Configuration	Status	LL	PL	PI	OMC	MDD
Kernel Functions	Auto	Auto selection for the best prediction				
Box Constraint	Auto	16.8 64	7.9 69	13.3 43	7.65 4	0.23 4
Kernel Scale	Auto	1	1	1	1	1
Epsilon	Auto	1.68 6	0.7 97	1.33 4	0.76 5	0.02 3
Standardize Data	Enable	YES				
Optimizer	Enable	Bayesian Optimizer				
Acquisition Function	Enable	Expected improvement per second plus				
Iterations	Default	Default (30)				
Max. Training time (sec.)	Default	Default (300)				
Number of Grid Divisions	Default	Default (10)				

The SVM models of liquid limit, plastic limit, plasticity index, OMC, and MDD are SVM_LL, SVM_PL, SVM_PI, SVM_OMC, and SVM_MDD, respectively.

2.3 Gaussian Process Regression

The Gaussian process regression (GPR) is the stochastic or random process used to solve the problem of probability over function in Bayesian interference [19] [14]. The GPR approach is based on covariance function: Linear, White Gaussian Noise, Ornstein-Uhlenbeck, Squared Exponential, Rational Quadratic, Periodic, Matern, etc. [19]. The GPR models of LL, PL, PI, OMC, and MDD have been constructed using the Regression Tool of MATLAB R2020a and designated as GPR_LL, GPR_PL, GPR_PI, GPR_OMC, and GPR_MDD, respectively. The configurations of constructed GPR models are given in Table 2.

Table 2 – Configuration of constructed GPR models

Configuration	Status	LL	PL	PI	OMC	MDD
Basic Function	Auto	Auto (Zero, Constant, Linear)				
Covariance Functions	Auto	Auto selection for the best prediction				
Kernel Scale	Auto	12.467	12.467	12.467	15.402	15.402
Signal Standard Deviation	Enable	12.786	5.258	10.182	4.646	0.145
Sigma	Auto	12.786	5.258	10.182	4.646	0.145
Standardize Data	Enable	YES				
Optimize Num. Parameter	Enable	YES				
Optimizer	Enable	Bayesian Optimizer				
Acquisition Function	Enable	Expected improvement per second plus				
Iterations	Default	Default (30)				
Max. Training time (sec.)	Default	Default (300)				
Number of Grid Divisions	Default	Default (10)				

2.4 Artificial Neural Networks

The artificial neural network is a computation system that is inspired by a biological neural network. The artificial neural network is a collection of input, hidden, and output layers connected with neurons. Each layer filters the information during training or validation for achieving better results. The hidden and output layers consist of activation functions, namely, linear activation and non-linear activation function. The ANN models of liquid limit, plastic limit, plasticity index, OMC, and MDD are designated as ANN_LL, ANN_PL, ANN_PI, ANN_OMC, and ANN_MDD, respectively, and constructed in MATLAB R2020a. The configurations of constructed GPR models are given in Table 3.

Table 3 – Configuration of constructed ANN models

Configurations	LL	PL	PI	OMC	MDD
Backpropagation Algorithm	Levenberg – Marquardt				
Number of Hidden Layers & Neurons	One Hidden Layer, Ten Neurons				
Normalization Function	Min-Max, Log Function				
Activation Function at Layers	Sigmoid, Linear				
Training: Validation Ratio	80: 20				
Epochs	Default (1000)				
Network Class	Multilayer Perceptron Class				
Network Type	Feed-Forward Backpropagation				
Mu, Max Fail, Min Gradient	0.001, 6, 10E-7				

2.5 Relevance Vector Machine

The relevance vector machine is a hybrid learning approach of artificial intelligence used to solve regression and classification problems [26]. The relevance vector machine

is also based on the kernel functions. The kernel functions are Gaussian, Linear, Polynomial Sigmoid, Laplacian. The genetic algorithm optimized single kernel function-based RVM models have been constructed in MATLAB R2020a in this study. The RVM models of liquid limit, plastic limit, plasticity index, OMC, and MDD are RVM_LL, RVM_PL, RVM_PI, RVM_OMC, and RVM_MDD, respectively. The configurations of constructed RVM models are given in Table 4.

Table 4 – Configuration of constructed RVM models

Configurations	LL	PL	PI	OMC	MDD
Free Basis	Enable				
Kernel Function	Gaussian				
Maximum Iterations	1000				
Type of Model	Single Kernel Based				
Optimization Method	Genetic Algorithm				
Target Kernel Type	Single Kernel				
Ib, uB	2 ⁻⁶				
Number of Variable	1				
Max. Iteration Optimizer	100				
K-folds	Default (5)				

III. DATA ANALYSIS

The datasets of soil have been collected from the published articles to carry out the present research work. A total of 166 datasets of soil have been used to employ the LL, PL, PI, OMC, and MDD models. These datasets consist of gravel content (GC), sand content (SC), silt content (MC), clay content (CC), LL, PL, PI, OMC, and MDD of soil. The details of the datasets are given in Table 5.

Table 5 – Detail of datasets of soil used in this study

S. No	Author(s)	Title of Article	No. of Datasets
Consistency limits (LL, PL, PI) of Soil			
1	Benson Craig H. et al., 1994 [3]	Estimating Hydraulic conductivity of thirteen compacted clays Liners	55
2	Benson Craig H. et al., 1995 [4]	Hydraulic conductivity of thirteen compacted clays	13
3	Najjar M. Y. et al., 1996 [17]	Utilizing computational neural networks for evaluating the permeability of compacted clay liners	47
4	Nagaraj H. B. et al., 2014 [15]	Correlation of compaction characteristics of natural soils with modified plastic limit	42



Prediction of Geotechnical Properties of Soil using Artificial Intelligence Framework

5	NG K. S. et al., 2015 [18]	Estimating maximum dry density and optimum moisture content of compacted soils	9
Compactive Parameters (OMC & MDD) of Soil			
1	Benson Craig H. et al., 1994 [3]	Estimating Hydraulic conductivity of thirteen compacted clays Liners	55
2	Najjar M. Y. et al., 1996 [17]	Utilizing computational neural networks for evaluating the permeability of compacted clay liners	47
3	Sinha K Sunil et al., 2008 [23]	Artificial neural network prediction models for soil compaction and permeability	55
4	NG K. S. et al., 2015 [18]	Estimating maximum dry density and optimum moisture content of compacted soils	9

For the rescaling of the datasets, the normalization technique has been used. The min-max and log normalization functions have been used to normalize input (G, S, M, C content) and output (LL, PL, PI, OMC, MDD) soil variables.

3.1 Descriptive Statistics of Dataset

The descriptive statistics of GC, SC, MC, CC, PI, PL, LL, OMC, and MDD of soil have been calculated using the Data Analysis Tool of Microsoft Excel 2019 to study the behavior of the datasets. The minimum, maximum, mean, standard deviation (StDev), and confidence level at 95% (CL) statistical parameters for total, training, and testing datasets have been calculated, as shown in Table 5.

Table 5 – Descriptive statistics of datasets used in the present study

Parameters	Total Datasets					Training Datasets					Testing Datasets				
	Min	Max	Mean	StDev	CL	Min	Max	Mean	StDev	CL	Min	Max	Mean	StDev	CL
<i>Consistency Limit Datasets (LL, PL, PI)</i>															
G	0.00	26.00	2.17	4.05	0.62	0.00	26.00	2.28	4.20	0.68	0.00	5.00	1.01	1.62	0.90
S	0.00	80.00	22.38	16.45	2.52	0.00	80.00	22.20	16.29	2.62	0.00	76.00	24.23	18.48	10.24
M	6.00	76.00	39.32	12.51	1.92	6.00	76.00	39.69	12.30	1.98	6.00	60.00	35.54	14.35	7.95
C	3.00	78.00	36.23	17.25	2.64	3.00	75.30	35.91	17.08	2.75	6.00	78.00	39.37	19.25	10.66
LL	19.00	115.00	46.79	18.08	2.77	19.00	115.00	46.65	18.08	2.91	27.00	87.00	48.17	18.66	10.33
PL	8.00	45.30	22.05	7.38	1.13	8.00	45.30	22.19	7.44	1.20	12.00	37.80	20.68	6.88	3.81
PI	3.70	75.60	24.73	14.26	2.19	3.70	75.60	24.46	14.40	2.32	11.40	57.00	27.49	12.87	7.13
<i>Compaction Parameters Datasets (OMC & MDD)</i>															
G	0.00	26.00	4.73	6.25	0.96	0.00	26.00	4.91	6.42	1.03	0.00	10.00	2.89	3.80	2.11
S	0.00	100.00	21.75	19.35	2.97	0.00	100.00	21.97	19.98	3.21	0.00	51.00	19.53	11.36	6.29
M	0.00	83.00	37.18	15.75	2.41	0.00	83.00	36.60	16.02	2.58	28.00	70.00	43.00	11.58	6.42
C	0.00	84.00	36.32	18.65	2.86	0.00	84.00	36.52	19.19	3.09	7.00	55.30	34.31	12.28	6.80
OMC	8.00	38.10	19.97	6.41	0.98	8.00	38.10	20.06	6.57	1.06	12.00	28.00	19.13	4.62	2.56
MDD	0.82	2.12	1.69	0.20	0.03	0.82	2.12	1.69	0.21	0.03	1.46	1.95	1.72	0.15	0.08

3.2 Pearson's Product Moment Correlation Coefficient

The correlation coefficient is calculated to determine the relationship between the pair of datasets. The methods for determining the correlation coefficient are the Scatter Diagram Method, Karl Pearson's Coefficient of Correlation, Spearman's Rank Correlation Coefficient, and Method of Least-Squares. The Karl Pearson's Coefficient of Correlation or Pearson's Product Moment Correlation Coefficient method is used in the present study to determine the relationship between input variables (GC, SC, MC, CC) and output variables (LL, PL, PI, OMC, MDD). The value of the coefficient of correlation more than 0.81, between 0.61-0.80, between 0.41-0.60, between 0.21-0.40, and less than 0.2, show the very strong, strong, moderate, weak, and no correlation, respectively [9]. The Pearson's correlation coefficient for pair of datasets of consistency limits has been calculated using Microsoft Excel 2019, as shown in Table 6.

Table 6 – Pearson's correlation coefficient (r) for datasets of consistency limits

(r)	GC (%)	SC (%)	MC (%)	CC (%)	LL (%)	PL (%)	PI (%)
GC (%)	1.00	-	-	-	-	-	-
SC (%)	0.15	1.00	-	-	-	-	-
MC (%)	0.13	0.36	1.00	-	-	-	-
CC (%)	0.30	0.72	0.34	1.00	-	-	-
LL (%)	0.21	0.38	0.42	0.71	1.00	-	-
PL (%)	0.05	0.07	0.30	0.17	0.67	1.00	-
PI (%)	0.24	0.51	0.37	0.82	0.92	0.33	1.00

The graphical presentation of the correlation coefficient for consistency limits (LL, PL, PI) of soil with input variables is shown in Figure. 1.



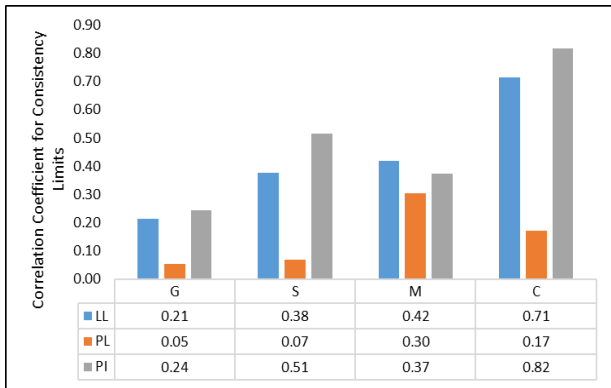


Figure 1. Correlation coefficient for LL, PL, and PI of soil

Table 7 – Pearson’s correlation coefficient (r) for datasets of OMC & MDD

(r)	GC (%)	SC (%)	MC (%)	CC (%)	OMC (%)	MDD (gm/cc)
GC (%)	1.00	-	-	-	-	-
SC (%)	0.15	1.00	-	-	-	-
MC (%)	0.13	0.36	1.00	-	-	-
CC (%)	0.30	0.72	0.34	1.00	-	-
OMC (%)	0.21	0.38	0.42	0.71	1.00	-
MDD (gm/cc)	0.05	0.07	0.30	0.17	0.67	1.00

The graphical presentation of the correlation coefficient for compaction parameters (OMC & MDD) of soil with input variables is shown in Figure. 2.

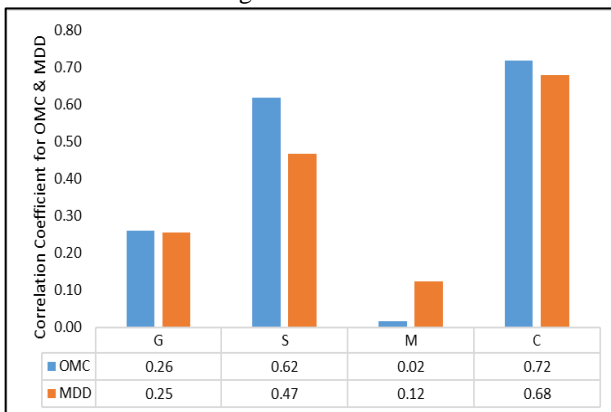


Figure 2. Correlation coefficient for OMC and MDD of soil

Figure 2 shows that the gravel content has a weak correlation with OMC and MDD of soil. Silt content shows no correlation with OMC and MDD of soil. Sand content has a strong and moderate correlation with OMC and MDD of soil, respectively. Clay content shows a strong and moderate correlation with OMC and MDD of soil. Therefore, it can be written that the prediction of OMC and MDD is strongly influenced by clay content, as shown in Figure 2.

3.3 Training, Testing and Validation Dataset

The datasets of soil have been divided for constructing the regression, support vector machine, Gaussian process regression, artificial neural network, and relevance vector machine model. The division and subdivision of soil datasets are shown in Table 8.

Figure 1 shows that the gravel, sand, and clay content have no correlation with the plastic limit. Sand and silt content has a weak to moderate correlation with LL and PI. Clay content shows a strong and very strong correlation with liquid limit and plasticity index, respectively. Therefore, it can be stated that the prediction of liquid limit and plasticity index is strongly influenced by clay content, as shown in Figure 1.

Similarly, the Pearson's correlation coefficient for pair of datasets of compaction parameters has been calculated using Microsoft Excel 2019, as shown in Table 7.

Table 8 – Divisions and subdivision of datasets of soil

Model	Training Data		Validation Data		Testing Data	
	LL, PL, PI	OMC, MDD	LL, PL, PI	OMC, MDD	LL, PL, PI	OMC, MDD
MLR	151	151	-	-	15	15
SVM	151	151	-	-	15	15
GPR	151	151	-	-	15	15
ANN	121	121	30	30	15	15
RVM	151	151	-	-	15	15

IV. RESULTS & DISCUSSIONS

Multiple regression, support vector machine, Gaussian regression process, artificial neural network, and relevance vector machine models have been constructed in this research work to predict the consistency limits and compaction parameters of soil. The performance of constructed models has been calculated in terms of R, RMSE, and MAE. The mathematical formula of R, RMSE, and MAE is –

$$R = \frac{\sum_{i=1}^N (T_i - \bar{T})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (T_i - \bar{T})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i - P_i)^2} \quad (7)$$

$$MAE = \frac{1}{N} (\sum_{i=1}^N abs(T_i - P_i)) \quad (8)$$



Prediction of Geotechnical Properties of Soil using Artificial Intelligence Framework

The coefficient of determination (R^2) has been calculated by mapping predicted vs actual plot of soil properties. The coefficient of determination value less than 0.2, between 0.2 to 0.8, and greater than 0.8 shows the weak, good and strong correlation between pairs of datasets [24]. The predicted results of consistency limits and compaction parameters of soil are discussed below.

4.1 Prediction of Liquid Limit

The liquid limit of fifteen soil specimens has been predicted using MLR_LL (Eq. 1), SVM_LL, GPR_LL, ANN_LL, and RVM_LL models. The performance of MLR_LL, SVM_LL, GPR_LL, ANN_LL, and RVM_LL models are shown in Figure 3.

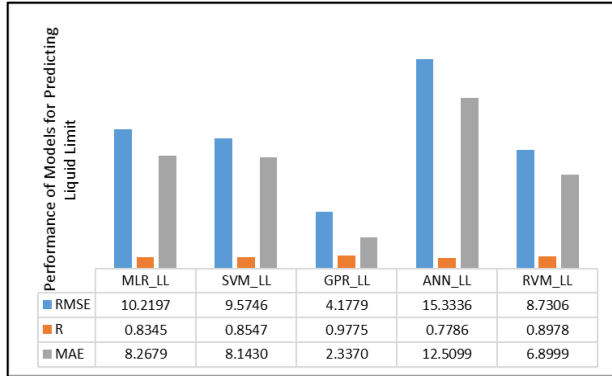


Figure 3. Performance of models for the liquid limit of soil

From Figure 3, it has been observed that the Gaussian process regression model has predicted LL of soil with $R = 0.9775$, which is higher than the MLR_LL, SVM_LL, ANN_LL, and RVM_LL.

Therefore, the GPR_LL model of LL has been identified as the optimum performance model. The performance curve of the GPR_LL model is shown in Figure 4.

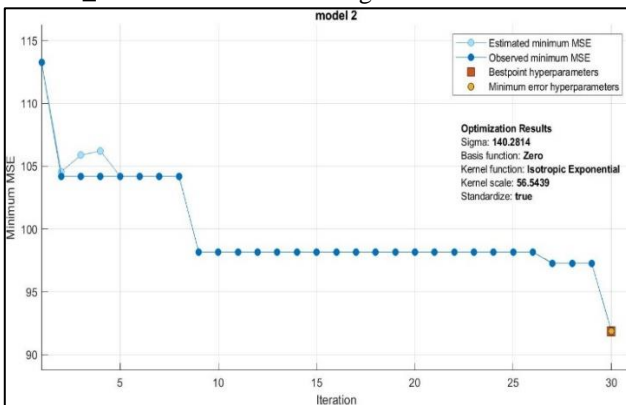


Figure 4. Performance curve of GPR_LL model

Figure 4 shows that the best prediction of LL can be achieved by selecting the isotropic exponential covariance function in the GPR model.

The GPR_LL model has achieved the best prediction of LL at the 30th iteration with the least RMSE = 4.1779 and MAE = 2.3370. The predicted vs Actual LL plot has also been drawn, as shown in Figure 5.

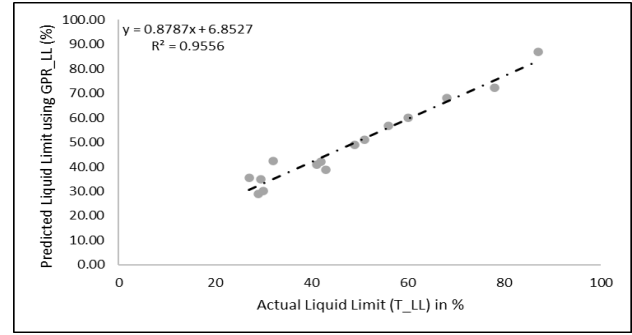


Figure 5. Predicted vs actual LL plot for GPR_LL model

From Figure 5, it has been observed that the GPR_LL model has predicted LL with $R^2 = 0.9556$, which shows a strong correlation with the actual LL of soil. Hence, it can be stated that the GPR_LL model can predict the liquid limit of soil.

4.2 Prediction of Plastic Limit

The plastic limit of fifteen soil specimens has been predicted using MLR_PL (Eq. 2), SVM_PL, GPR_PL, ANN_PL, and RVM_PL models. The performance of MLR_PL, SVM_PL, GPR_PL, ANN_PL, and RVM_PL models are shown in Figure 6.

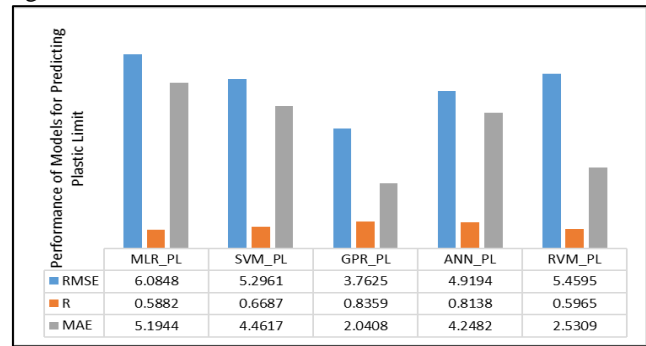


Figure 6. Performance of models for the plastic limit of soil

From Figure 6, it has been observed that the Gaussian process regression model has predicted PL of soil with $R = 0.8359$, which is higher than the MLR_PL, SVM_PL, ANN_PL, and RVM_PL. Therefore, the GPR_PL model of PL has been identified as the optimum performance model. The performance curve of the GPR_PL model is shown in Figure 7.

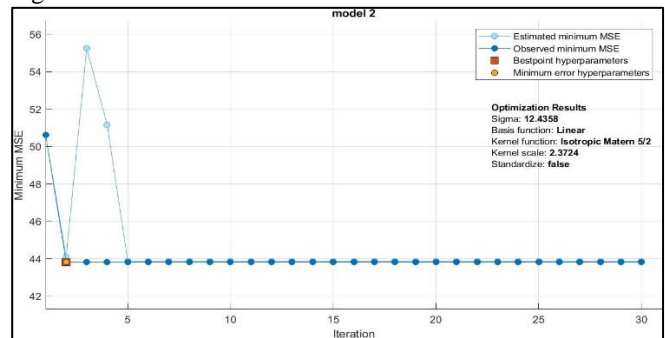


Figure 7. Performance curve of GPR_PL model

Figure 7 shows that the best prediction of PL can be achieved by selecting the isotropic Matern 5/2 covariance function in the GPR model. The GPR_PL model has achieved the best prediction of PL at the 1st iteration with the least RMSE = 3.7625 and MAE = 2.0408. The predicted vs Actual PL plot has also been drawn, as shown in Figure 8.

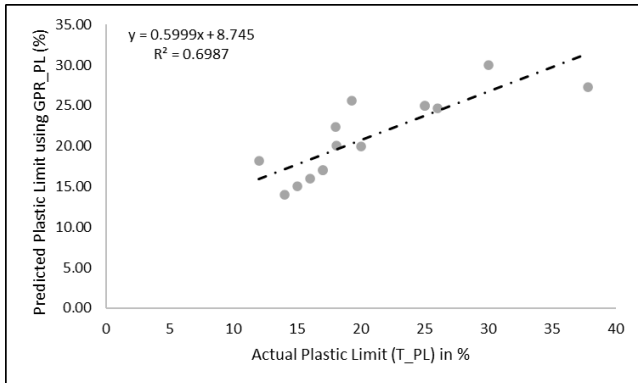


Figure 8. Predicted vs actual PL plot for GPR_PL model

From Figure 8, it has been observed that the GPR_PL model has predicted PL with $R^2 = 0.6987$, which shows a good correlation with the actual PL of soil. Hence, it can be stated that the GPR_PL model can predict the plastic limit of soil.

4.3 Prediction of Plasticity Index

The plasticity index of fifteen soil specimens has been predicted using MLR_PI (Eq. 3), SVM_PI, GPR_PI, ANN_PI, and RVM_PI models. The performance of MLR_PI, SVM_PI, GPR_PI, ANN_PI, and RVM_PI models are shown in Figure 9.

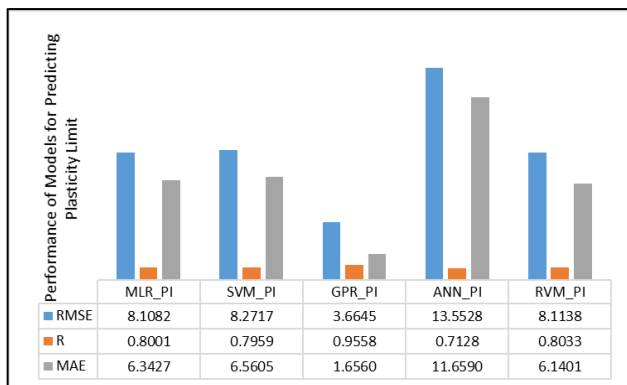


Figure 9. Performance of models for plasticity index of soil

From Figure 9, it has been observed that the Gaussian process regression model has predicted PI of soil with $R = 0.9558$, which is higher than the MLR_PI, SVM_PI, ANN_PI, and RVM_PI. Therefore, the GPR_PI model of PI has been identified as the optimum performance model. The performance curve of the GPR_PI model is shown in Figure 10.

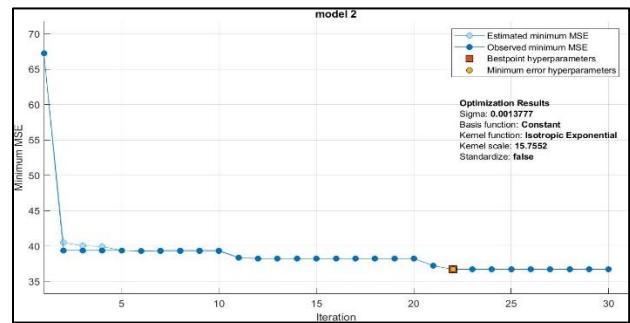


Figure 10. Performance curve of GPR_PI model

Figure 10 shows that the best prediction of PI can be achieved by selecting the isotropic exponential covariance function in the GPR model. The GPR_PI model has achieved the best prediction of PI at the 22nd iteration with the least RMSE = 3.6645 and MAE = 1.6560. The predicted vs Actual PI plot has also been drawn, as shown in Figure 11.

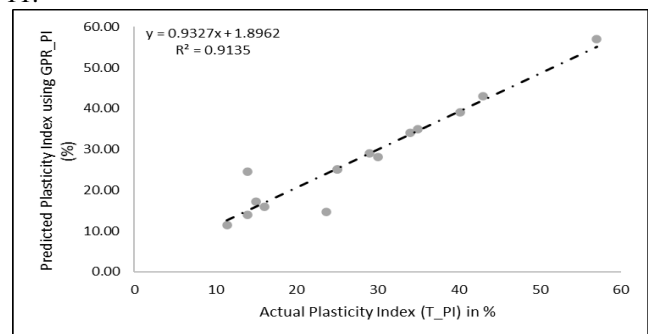


Figure 11. Predicted vs actual PI plot for GPR_PI model

From Figure 11, it has been observed that the GPR_PI model has predicted PI with $R^2 = 0.9135$, which shows a strong correlation with the actual PI of soil. Hence, it can be stated that the GPR_PI model can predict the plasticity index of soil.

4.4 Prediction of Optimum Moisture Content

The optimum moisture content of fifteen soil specimens has been predicted using MLR_OMC (Eq. 4), SVM_OMC, GPR_OMC, ANN_OMC, and RVM_OMC models. The performance of MLR_OMC, SVM_OMC, GPR_OMC, ANN_OMC, and RVM_OMC models are shown in Figure 12.

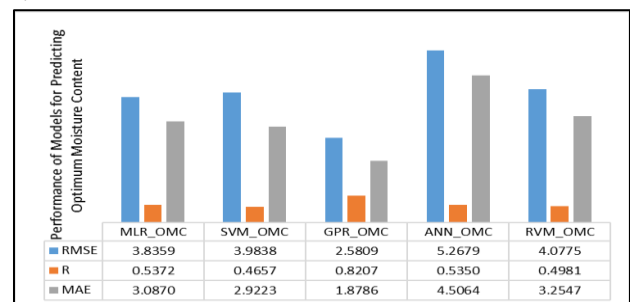


Figure 12. Performance of models for OMC of soil

Prediction of Geotechnical Properties of Soil using Artificial Intelligence Framework

From Figure 12, it has been observed that the Gaussian process regression model has predicted OMC of soil with $R = 0.8207$, which is higher than the other OMC models. Therefore, the GPR_OMC model of OMC has been identified as the optimum performance model. The performance curve of the GPR_OMC model is shown in Figure 13.

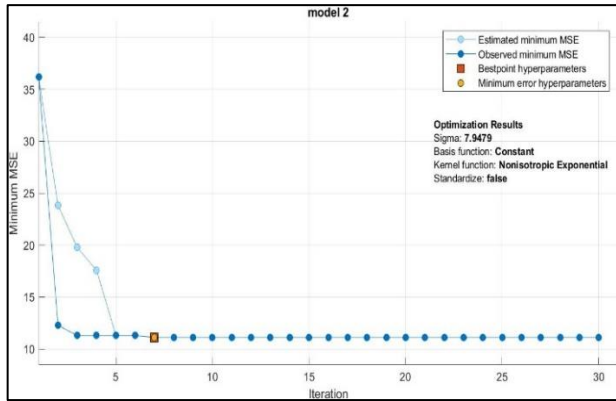


Figure 13. Performance curve of GPR_OMC model

Figure 13 shows that the best prediction of OMC can be achieved by selecting the non-isotropic exponential covariance function in the GPR model. The GPR_OMC model has achieved the best prediction of OMC at the 7th iteration with the least RMSE = 2.5809 and MAE = 1.8786. The predicted vs Actual OMC plot has also been drawn, as shown in Figure 14.

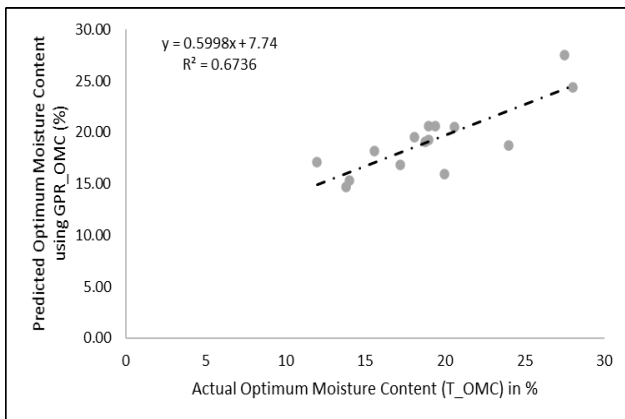


Figure 14. Predicted vs actual OMC plot for GPR_OMC model

From Figure 14, it has been observed that the GPR_OMC model has predicted OMC with $R^2 = 0.6736$, which shows a good correlation with the actual OMC of soil. Hence, it can be stated that the GPR_OMC model can predict the optimum moisture content of the soil.

4.5 Prediction of Maximum Dry Density

The maximum dry density of fifteen soil specimens has been predicted using MLR_MDD (Eq. 5), SVM_MDD, GPR_MDD, ANN_MDD, and RVM_MDD models. The performance of MLR_MDD, SVM_MDD, GPR_MDD, ANN_MDD, and RVM_MDD models are shown in Figure 15.

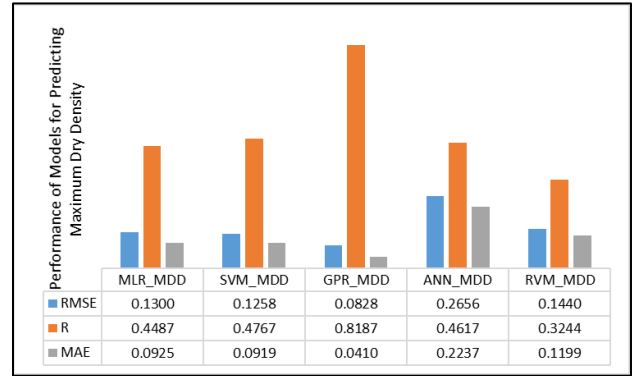


Figure 15. Performance of models for MDD of soil

From Figure 15, it has been observed that the Gaussian process regression model has predicted MDD of soil with $R = 0.8187$, which is higher than the other MDD models. Therefore, the GPR_MDD model of MDD has been identified as the optimum performance model. The performance curve of the GPR_MDD model is shown in Figure 16.

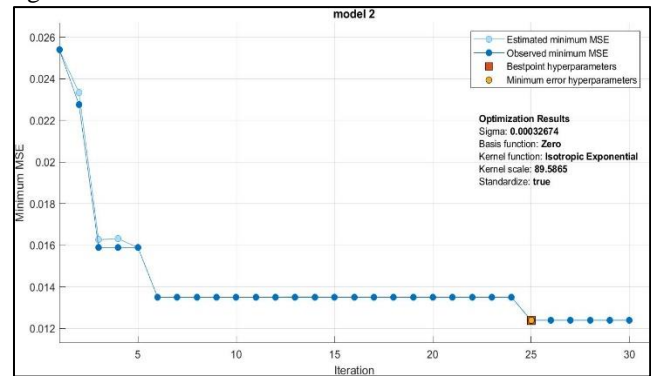


Figure 16. Performance curve of GPR_MDD model

Figure 16 shows that the best prediction of MDD can be achieved by selecting the isotropic exponential covariance function in the GPR model. The GPR_MDD model has achieved the best prediction of MDD at the 25th iteration with the least RMSE = 0.0828 and MAE = 0.0410. The predicted vs Actual MDD plot has also been drawn, as shown in Figure 17.

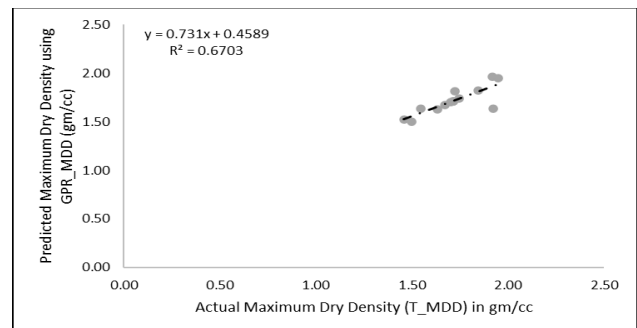


Figure 17. Predicted vs actual MDD plot for GPR_MDD model

From Figure 17, it has been observed that the GPR_MDD model has predicted MDD with $R^2 = 0.6703$, which shows a good correlation with the actual MDD of soil. Hence, it can be stated that the GPR_MDD model can predict the maximum dry density of soil.

4.6 Classification of Soil

The Gaussian process regression models of consistency limits, namely GPR_LL, GPR_PL, and GPR_PI, have been identified as the optimum performance model. The fifteen soil specimens have been classified based on the liquid limit and plasticity index using the Casagrande plasticity chart. The classification of soil by laboratory methods and prediction methods have been compared, as shown in Figure 18.

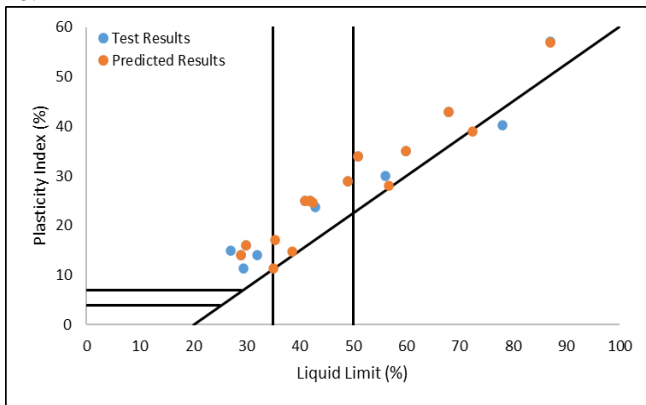


Figure 18. Classification of soil using test and predicted results

From Figure 18, it has been observed that the fifteen soil specimens have been classified with an accuracy of 86.67%. Hence, GPR_LL and GPR_PI models can predict the LL and PI of soil and classify the soil as per the Casagrande plasticity chart.

V.SENSITIVITY ANALYSIS

The sensitivity analysis is performed to determine the impact of the input variable on the output variable. The sensitivity analyses are local sensitivity analysis and global sensitivity analysis. In the present study, the Monte – Carlo global sensitivity analysis has been performed. The Monte – Carlo sensitivity analysis is a better way of describing uncertainty in variables. The results of Monte – Carlo analysis for consistency limits of soil is shown in Figure 19.

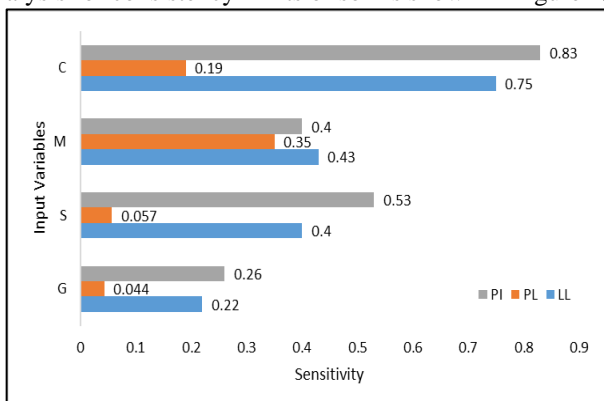


Figure 19. Sensitivity analysis for consistency limits

Figure 19 shows that the clay content is more sensitive in predicting the liquid limit and plasticity index. The gravel, sand, silt, and clay contents are less sensitive in predicting the plastic limit of soil. Similarly, the sensitivity analysis has been performed for compaction parameters, as shown in Figure 20.

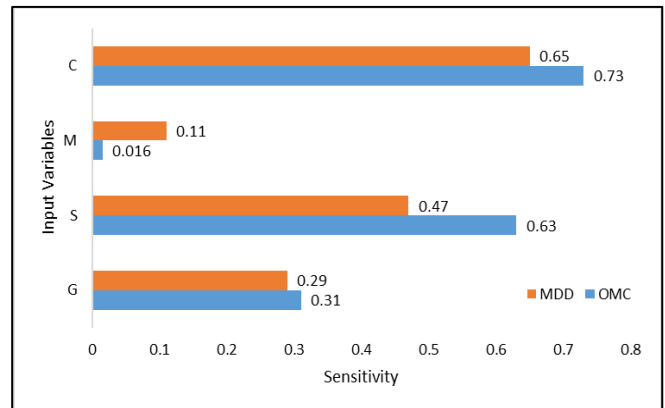


Figure 20. Sensitivity analysis for compaction parameters

Figure 20 shows that the OMC and MDD of soil are highly influenced by sand and clay content compared to gravel and silt content. From the sensitivity analysis, it may be stated that the geotechnical properties of soil are influenced by sand and clay content.

VI.CONCLUSIONS

The MLR, SVM, GPR, ANN, and RVM models have been developed to predict the LL, PL, PI, OMC, and MDD of soil. The results and performance of developed models have been compared to achieve an optimum performance model. The following conclusions are mapped based on the comparison of models and statistical study.

- The liquid limit, plasticity index, OMC, and MDD of soil are more influenced by sand and clay content. The sand and clay content are less effective in the prediction of the plastic limit of soil.
- The machine learning models outperformed the hybrid learning and deep learning models. Also, the deep learning model outperformed the hybrid learning model.
- The GPR models of geotechnical properties of soil achieved a performance of more than 0.8, which shows a strong correlation with laboratory results.
- The GPR_LL and GPR_PI models predicted the classification of soil with an accuracy of 86.67%.
- Monte – Carlo sensitivity analysis shows that the geotechnical properties of soil are susceptible to sand and clay content.

From the comparison of the performance of the model, it has been observed that the GPR model outperformed the other models. Hence, the GPR models can be used to predict the geotechnical properties of soil.

REFERENCES

- Aronovici V. S. (1946), The mechanical analysis as an index of subsoil permeability, Soil Science Society Proceedings, pp. 137-141.
- Arora K. R. (2004), Soil mechanics and foundation engineering, Standard Publisher Distributors, New Delhi, India
- Benson C. H. (1994), Estimating Hydraulic Conductivity of Compacted clay liners, Journal of Geotechnical Engineering, 120, pp. 366-387, [https://doi.org/10.1061/\(ASCE\)0733-9410\(1994\)120:2\(366\)](https://doi.org/10.1061/(ASCE)0733-9410(1994)120:2(366)).
- Benson C. H. and Trast J. M. (1995) Hydraulic conductivity of thirteen compacted clays. Clays and Clay Minerals, 43, No. 6, 669-681, <https://doi.org/10.1346/CCMN.1995.0430603>
- Bourouis Mohammed el Amin, Zadjajou Abdeldjalil, Djedid Abedkader, Bensenouci Abderrahmen (2017), Design of neural networks by using genetic algorithm for the prediction of immersed CBR index, Long-Term Behaviour and Environmentally Friendly Rehabilitation Technologies of Dams, pp. 331-338, doi:10.3217/978-3-85125-564-5-046.
- Chandrakar Vishal, Yadav R K (2016), Study of correlation of CBR value with engineering properties and index properties of coarse-grained soil, International Research Journal of Engineering and Technology, 3(11), pp. 772-778.
- Chenari Reza Jamshidi, Tizpa Parichehr, Rad Mohammad Rasool Ghorbani, Machado Sandro Lemos, Frad Mehran Karimpour (2014), The use of index parameters to predict soil geotechnical properties, Arab Journal of Geosciences, 8, pp. 4907-4919, doi: 10.1007/s12517-014-1538-0.
- Cortes, Corinna; Vapnik, Vladimir N. (1995), "Support vector networks", Machine Learning, 20 (3), pp. 273-297, doi:10.1007/BF00994018. S2CID 206787478.
- Hair Jr. J, Wolfnibarger M. C., Ortinau, D. J. & Bush, R. P. 2013, Essentials of Marketing, New York, Mc Graw Hill
- J. K. Pradeep Kumar, M. Y. Harish Patel (2016), Soft computing technique for prediction of CBR from index properties of subgrade soil, International Journal of Innovative Research in Science, Engineering and Technology, 5(7), pp. 13852-13860, doi:10.15680/IJIRSET.2016.0507214.
- J. Khatti, Prof. Kamaldeep Singh Grover (2021), Prediction of index properties of soil using AI approaches, Proceedings of International Conference on Advancements in Computer Applications 2021, FISAT, pp. 8-14, ISBN: 978-93-5473-978-1, https://www.researchgate.net/publication/354378693_Prediction_of_Index_Properties_of_Soil_using_AI_Approaches
- Janjua Zohib Shahzad, Chand Jagdish (2016), Correlation of CBR with index properties of soil, International Journal of Civil Engineering and Technology, 7(5), pp. 57-62, <http://iaeme.com/Home/issue/IJCIET?Volume=7&Issue=5>
- Jasim Mustafa M., Al-Khaddar Rana M., Al-Rumaithi Ayad (2019), Prediction of bearing capacity, angle of internal friction, cohesion, and plasticity index using ANN (Case study of Baghdad, Iraq), International Journal of Civil Engineering and Technology, 10(1), pp 2670-2679, <http://www.iaeme.com/ijciat/issues.asp?JType=IJCIET&VType=10&IType=1>
- Liu, W.; Principe, J.C.; Haykin, S. (2010), Kernel Adaptive Filtering: A Comprehensive Introduction, John Wiley, ISBN 978-0-470-44753-6.
- Nagaraj H. B., Reesha B., Sravan M. V., Suresh R. M., (2015), Correlation of compaction characteristics of natural soils with modified plastic limit, Transportation Geotechnics, 2, pp 65-77, <http://dx.doi.org/10.1016/j.trgeo.2014.09.002>
- N. Srinivasa Reddy, N. Ruchita, Sharma Pankaj, Satyanarayana Sathiraju Venkata (2019), Prediction of California bearing ratio through empirical correlations of index properties for tropical Indian soils, International Journal of Innovations in Engineering and Technology, 15(1), pp. 67-77, <http://dx.doi.org/10.21172/ijiet.1511.09>.
- Najjar, Y.M., Basheer, I.A., 1996. Utilizing computational neural networks for evaluating the permeability of compacted clay liners. J. Geotech. Geol. Eng. 14, 193-212, <https://doi.org/10.1007/BF00452947>.
- NG, K.S., Chew, Y.M., Osman, M.H., Mohamad Ghazali, S.K. (2015). Estimating Maximum Dry Density and Optimum Moisture Content of Compacted Soils, International Conference on Advances in Civil and Environmental Engineering 2015, 1-8.
- Rasmussen, C.E.; Williams, C.K.I (2006), Gaussian Processes for Machine Learning, MIT Press, ISBN 978-0-262-18253-9.
- Rassoul Ajallooian, Mojtaba Kiani (2015), Predicting maximum dry density, optimum moisture content and California bearing ratio (CBR) through soil index using ordinary least squares (OLS) and artificial neural networks (ANNs), International Journal of Innovative Technology and Exploring Engineering, 5(3), pp. 1-5.
- Rehman Z. U, Khalid U., Farooq K., Mujtaba H. (2017), Prediction of CBR value from index properties of different soils, Technical Journal, University of Engineering and Technology, 22(2), pp. 17-26.
- Sherzoy Mohammad Murtaza (2017), Atterberg limits prediction comparing SVM with ANFIS model, Journal of Geoscience, Engineering, Environment, and Technology, 2(1), pp. 20-30.
- Sinha, S.K., Wang, M.C. (2008) Artificial Neural Network Prediction Models for Soil Compaction and Permeability. Geotech Geol Eng, 26,47-64, doi: 10.1007/s10706-007-9146-3
- Smith G. N. (1986), Probability and statistics in civil engineering – An introduction, Collins, London.
- Tenpe Ashwini, Kaur Suneet (2015), Artificial neural network modeling for predicting compaction parameters based on index properties of soil, International Journal of Science and Research, 4(7), pp. 1198-1202, Paper ID: SUB156564.
- Tipping, Michael E. (2001), "Sparse Bayesian Learning and the Relevance Vector Machine", Journal of Machine Learning Research, 1, pp. 211-244.
- Tizpa Parichehr, Chenari Reza Jamshidi, Frad Mehran Karimpour, Machado Sandro Lemos (2014), ANN prediction of some geotechnical properties of soil from their index parameters, Arab Journal of Geosciences, 8, pp. 2911-2920, doi: 10.1007/s12517-014-1304-3.
- Venkatasubramanian C, Dhinakaran G. (2011), ANN model for predicting CBR from index properties of soils, International Journal of Civil and Structural Engineering, 2(2), pp. 605-611.
- Vetrivel N., Shivakumar N. V. (2019), Prediction on soil index properties by artificial neural network in East part of Bangalore City, Karnataka, International Journal of Basic and Applied Research, 9(4), pp. 280-287.

AUTHORS PROFILE



Mr. Jitendra Khatti is pursuing PhD at Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan. He completed his Master's from Rajasthan Technical University, Kota. He published many research articles in International and National Journals/ Conferences. The research areas of Mr. Jitendra Khatti are Applications of Artificial Intelligence in Civil Engineering, Stabilization of Soil, Pavement Engineering, Use of Waste in Concrete and Tall Structures.



Dr. Kamaldeep Singh Grover is Professor at Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan. Dr. Kamaldeep Singh Grover is specialized in Geotechnical Engineering and Structural Engineering. The research areas of Dr. Kamaldeep Singh Grover are Artificial Intelligence, Concrete Technology, Tall Structures, Pavement Engineering and Stabilization of Soil.