

PAPER • OPEN ACCESS

Application of artificial intelligence techniques in prediction of cyclic resistance ratio (CRR) of clean sands

To cite this article: V Akhila and S Adarsh 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **491** 012048

View the [article online](#) for updates and enhancements.



240th ECS Meeting ORLANDO, FL

Orange County Convention Center Oct 10-14, 2021



Abstract submission due: April 9

SUBMIT NOW

Application of artificial intelligence techniques in prediction of cyclic resistance ratio (CRR) of clean sands

V Akhila^{1*} and S Adarsh²

¹ Department of Civil Engineering, Indian Institute of Technology, Guwahati, Assam, India 781039

² Department of Civil Engineering, Thangal Kunju Musaliar College of Engineering, Kollam, Kerala, India 691005

*Corresponding author – Email: akhilavasudev@gmail.com

Abstract. Liquefaction study by in-situ tests like SPT and CPT are very complicated and time consuming. Cyclic Resistance Ratio (CRR) of a soil is controlled by various properties of the soil. Artificial intelligence techniques can identify relationship between various parameters which influence the liquefaction phenomenon from sufficiently large data set to generate models connecting those parameters. Models for prediction of cyclic resistance ratio (CRR) of clean sand is generated using MGGP, GPR and M5' model tree in the present study using data from cyclic triaxial test and cyclic direct shear test. Using 346 data points, divided in 50% train to 50% test ratio, sufficiently accurate models were generated through the algorithms considered. These algorithms were compared by means of the Root Mean Square Error (RMSE), Coefficient of correlation (R^2) and Maximum absolute Error in prediction (MAE). An equation connecting the CRR with other input parameters was developed using the MGGP algorithm, which also showed the maximum R^2 value of 0.96 for the test data. The AI algorithms were observed to satisfactorily model the relation between the input parameters and the CRR without any prior knowledge of the same.

Keywords: Cyclic Resistance Ratio, Multigene genetic programming, Gaussian Process Regression, M5' model tree

1. Introduction

The loss of shear strength of saturated cohesionless soils, triggered by earthquake motion is a complicated phenomenon. It makes the soil behave like water, making the pore pressure shoot up, reducing the effective stress excessively, and hence lowering the resistance to load over it. This process is therefore termed as liquefaction. Various indices and parameters have been shown to influence this occurrence through experiments as well as using artificial intelligence techniques. With sufficient data available, many studies have used machine learning techniques like ANN [1],[2] to simulate the process of liquefaction and to understand the parameters that influence the same. Liquefaction potential can be defined by a factor of safety defined by [3] as Equation (1)

$$\text{factor of safety} = \frac{CRR}{CSR} \quad (1)$$

Where CSR is the cyclic stress ratio which is the seismic load imposed on the ground by the earthquake waves and CRR is the cyclic resistance ratio, the resistance that the ground can provide towards this ground motion. At any depth, if CSR is greater than CRR, there is a possibility of



liquefaction. In-situ testing like SPT and CPT are used for the finding the values of these factors. Since SPT is difficult and not repeatable to precision, CPT results with overburden correction can be used for CRR value calculation in case of clean sand and silts [4]. SPT and CPT data are already used in various artificial intelligence techniques for prediction of liquefaction potential [5-9]. The understanding of liquefaction resistance of soil based on laboratory experiments is an area which has not been much explored in a machine learning framework. Few attempts using ANN [10-12] are reported. They have used data from laboratory tests like cyclic triaxial test, cyclic direct shear test and static triaxial test for learning and testing.

Other machine learning algorithms are also popular, which have not yet been used to study the laboratory data. Gaussian Process regression (GPR) is a technique which is effective and has been used in a few studies like [13-16]. Genetic programming is an algorithm that mimics the evolution process of species and has been proved effective in modelling in the studies of [8][9][17][18]. The M5 model tree is a decision tree algorithm that has good potential of learning large amount of data[19][20]. The study in this paper uses these three algorithms on cyclic triaxial test and cyclic direct shear test data for clean sands. Data for nine different types of sand, connecting number of cycles(N) with CRR is used, with other influencing parameters for training and testing the models. The comparative efficiency of models was studied.

2. Methodology

In-situ tests for liquefaction resistance assessment are difficult and lacks repeatability in certain conditions. The laboratory tests data on nine clean sands were divided into proper testing and training data sets. These data were used in GPR, M5 model tree and MGGP algorithms for generating prediction model for CRR. The predicted and actual values were compared on the basis of root mean square error (RMSE), coefficient of correlation(R^2) and maximum absolute error (MAE), from a statistical point of view. Root mean square error measures the deviation from the line of equality where the predicted value is equal to the actual value. The coefficient of correlation marks the extent to which the values predicted by the models can be correlated with the actual values.

$$\text{Coefficient of correlation } R = \frac{\sum_{i=1}^n (CRRp_i - \overline{CRRp})(CRRa_i - \overline{CRRa})}{\sqrt{\sum_{i=1}^n (CRRp_i - \overline{CRRp})^2} \sqrt{\sum_{i=1}^n (CRRa_i - \overline{CRRa})^2}} \quad (2)$$

$$\text{Root Mean Square Error RMSE} = \sqrt{\frac{\sum_{i=1}^n (CRRp_i - CRRa_i)^2}{n}} \quad (3)$$

where n is the number of observations, CRR_p the predicted value of cyclic resistance ratio, CRR_a is the actual value of cyclic resistance ratio.

2.1. Gaussian process regression (GPR)

The gaussian process regression (GPR) is a machine learning technique based on gaussian process. Gaussian process, according to [21] is a set of randomly considered variables, from which a finite number of variables will form a joint Gaussian distribution or normal distribution. Therefore, it is a Bayesian based approach which infers a probability distribution over all possible values. The training data is observed and according to this the distribution is modified to form posterior gaussian process. The training requires a covariance function also called a kernel function which will act as a prior function for the prediction. Kernels have hyperparameters associated with them which will be optimised based on training data. GPR is a non-parametric technique, implying that for each prediction the whole training set will be referred again and again, making it a bit computationally expensive. Mathematical details regarding the different kernel functions are available in [14].

2.2. Multi gene genetic programming (MGGP)

Genetic programming is an artificial intelligence technique, developed by [22] which can create a predictive relationship through training from a given data set. This evolutionary algorithm-based method generates a population of random programs from the data provided for training, which are further developed by applying mutation, cross over and elitism, generation after generation, to properly model the issue at hand. Multi gene genetic programming (MGGP) is used in this study to model liquefaction for the prediction of CRR. It is a tree-based approach resulting in a mathematical Equation which can predict the required value. The method was implemented in liquefaction potential prediction in [8][9][17] from standard penetration test (SPT) data and liquefaction index prediction from CPT data. Best models are selected based on the fitness value, which can be defined by error values like root mean square error (RMSE) or correlation coefficient (R2). In MGGP, the final model is represented as a combination of several trees of expressions which are called genes. These are obtained by least square method, as a linear combination of non linear expressions with constant coefficients. The model complexity and overfitting is controlled by the parameters Gmax and dmax which are the maximum number of genes in the tree and the maximum depth of the tree, resulting in accurate and compact models [23]. Figure 1 represents an MGGP tree of two genes and a tree depth of 3.

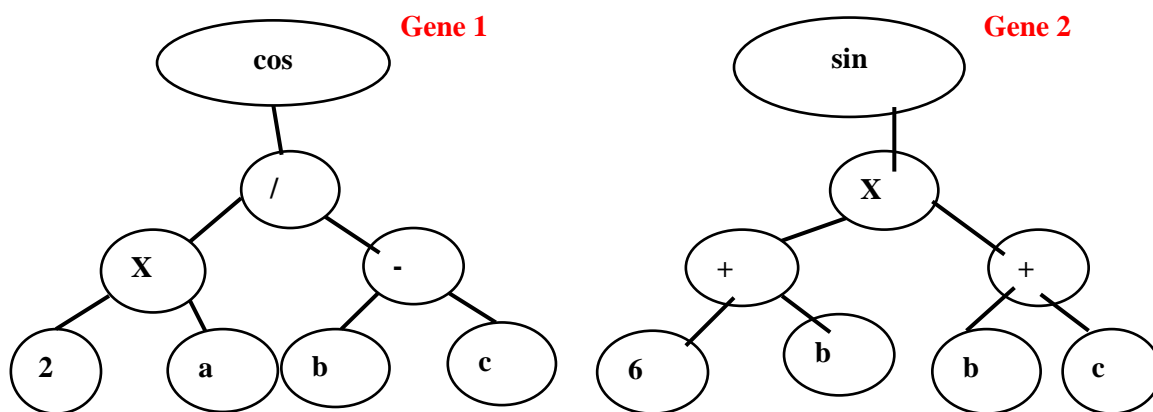


Figure 1 MGGP tree for the expression $Y=C_0+C_1 \cos(2a/(b-c))+C_2\sin((6+b)/(b+c))$

2.3. M5' Model Tree

The M5' Model Tree algorithm learns models that makes trees using multivariate linear models, without leading to any unusual computational growth. It is a binary decision tree with functions of linear regression at the terminal or leaf nodes, which is capable of predicting attributes which are continuous and numerical in nature. A two stepped process splitting and pruning is used in building the model. The instance/input is repeatedly split in order to build the regression tree, such that along the branches, the intra-subset variation from the root to the node is decreased. Splitting in the input space is terminated based on the standard deviation reduction (SDR). For M5' model tree, the standard deviation of a class at the nodal position is taken as error value at that node. Each attribute is tested at that node to find the one which can give maximum reduction in error. The resulting child nodes have lesser standard deviation than the parent nodes. A number of possible splits are checked and the one that yields maximum expected reduction in error is implemented. This continuous splitting at nodes result in overfitting, caused by large trees. This is avoided by the second step, which is pruning along with smoothing of such trees and replacing the pruned subtrees with linear regression functions[19, 20, 24] More details are available in [25].

3. Database used for development of models

The database reported by [10], consisting of 346 data points was used in this study, connecting the number of cycles (N) and Cyclic Resistance Ratio (CRR). These data were extracted from the cyclic triaxial and cyclic direct shear tests conducted on different types of sands in different initial conditions. Tests were conducted to produce sufficient shear distortion. Various input parameters considered are uniformity coefficient (U_c), average grain size (D_{50}), effective mean principal stress (σ_{mc}), initial static shear stress ratio (η_s), relative density ($D_r\%$), maximum void ratio (e_{max}), minimum void ratio (e_{min}), amplitude strain ($\epsilon_a\%$) and number of cycles to liquefaction (N). The algorithms used yielded best results at 50% train and 50% test ratio. Statistical parameters of the data are provided in the Table 1. The various parameters were kept almost consistent for the train and test data based on this ratio.

Table 1: Statistical details of parameters in the data

Data	Detail	U_c	D_{50} (mm)	σ_{mc} (kPa)	η_s	D_r (%)	e_{max}	e_{min}	ϵ_a (%)	N	CRR
Train Data	Maximum	2.4	3.3	1568	1	90	2.451	1.621	5	707.18	0.95
	Minimum	1.2	0.18	33	0	31	0.77	0.48	2.5	1	0.036
	Standard Deviation	0.392	0.761	275.101	0.379	15.485	0.413	0.280	1.235	93.751	0.199
	Mean	1.585	0.519	174.559	0.488	58.787	1.052	0.689	3.888	46.823	0.301
Test Data	Maximum	2.4	3.3	1568	1	90	2.451	1.621	5	707.18	1
	Minimum	1.2	0.18	33	0	31	0.77	0.48	2.5	1	0.05
	Standard Deviation	0.369	0.693	310.519	0.375	15.816	0.451	0.305	1.231	71.852	0.194
	Mean	1.623	0.492	203.386	0.483	59.369	1.067	0.695	3.702	31.246	0.280

4. Results and Discussion

The data base was trained and tested for different train to test ratios. Finally a ratio of 50% train to 50% test was observed to give sufficiently good results for the three methods adopted for comparison. The models were compared on the basis of errors and the plot of predicted Vs actual values of CRR were also plotted for each of the methods are explained below.

4.1. M5' Model Tree technique

M5' model tree algorithm was applied with the help of the graphical user interface (GUI) provided by the platform WEKA. Models were pruned and smoothed to control overfitting. The developed model has three instance and is based on if-then conditions. Model tree formed is shown in Figure 2. Table 2 gives linear model coefficients shown in the model tree. Notations in the tree are also explained in the same. The RMSE value of the M5' model tree is 0.064 for training data and 0.062 for testing data. A correlation coefficient of 0.948 is also observed which proves that the model is effective.

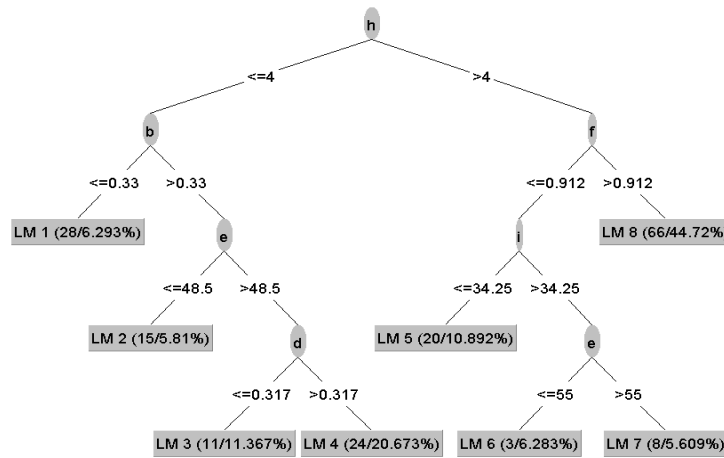


Figure 2. M5' model tree

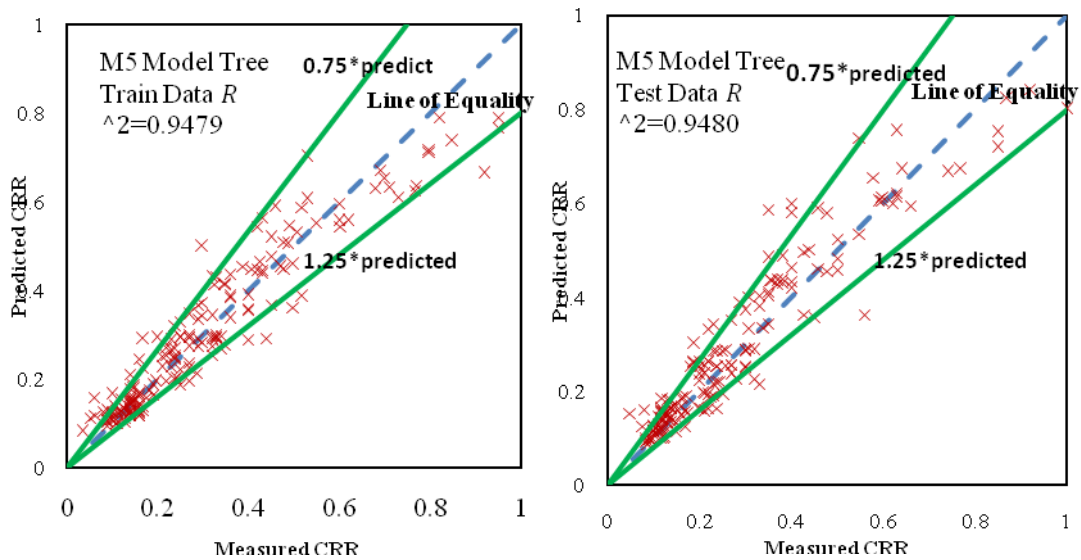


Figure 3: Comparison of predicted Vs actual CRR value for M5' model tree

Table 2: Linear model coefficients shown in the model tree

Notation	Parameter	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8
a	Uc	-0.0253	0.2786	0.4472	0.6847	-0.2634	-0.2634	-0.2634	-0.0395
b	D50 (mm)	0.1498	0.0991	0.0991	0.0991	0	0	0	0
c	σ_{mc} (kPa)	-0.0001	-0.0001	-0.0001	-0.0001	0	0	0	0
d	η_s	0.0247	0.0158	0.0562	0.0477	0.0827	0.0827	0.0827	0.4053
e	Dr(%)	0.0012	0.0023	0.0023	0.003	0.0047	0.0039	0.0038	0.0045
f	e_max	0.0356	0.0356	0.0356	0.0356	0.162	0.162	0.162	0.1085
g	e_min	0	0	0	0	0	0	0	0
h	ϵ_a (%)	0.0093	0.0093	0.0093	0.0093	0.0078	0.0078	0.0078	0.0078
i	N	-0.0005	-0.0004	0.0004	-0.0003	-0.0023	-0.0003	-0.0003	-0.0012
Constant		0.023	-0.4978	-0.7593	-1.1324	0.1998	0.2105	0.2233	0.0299

4.2. GPR method

Gaussian process regression was implemented through WEKA GUI. The covariance or kernel function used for training is the radial basis kernel function (RBF). It can be represented as $K(x, y) = e^{-\gamma \|x-y\|^2}$. Gamma is taken as 5 in this study. Figure 4 shows the predicted vs actual value of CRR from the GPR model. A correlation value of 0.952 was observed for the train data and 0.953 for the test data. The RMSE is 0.006 for train and 0.063 for the test data. The train and test data was observed to behave in a similar way resulting in maximum correlation and very less error.

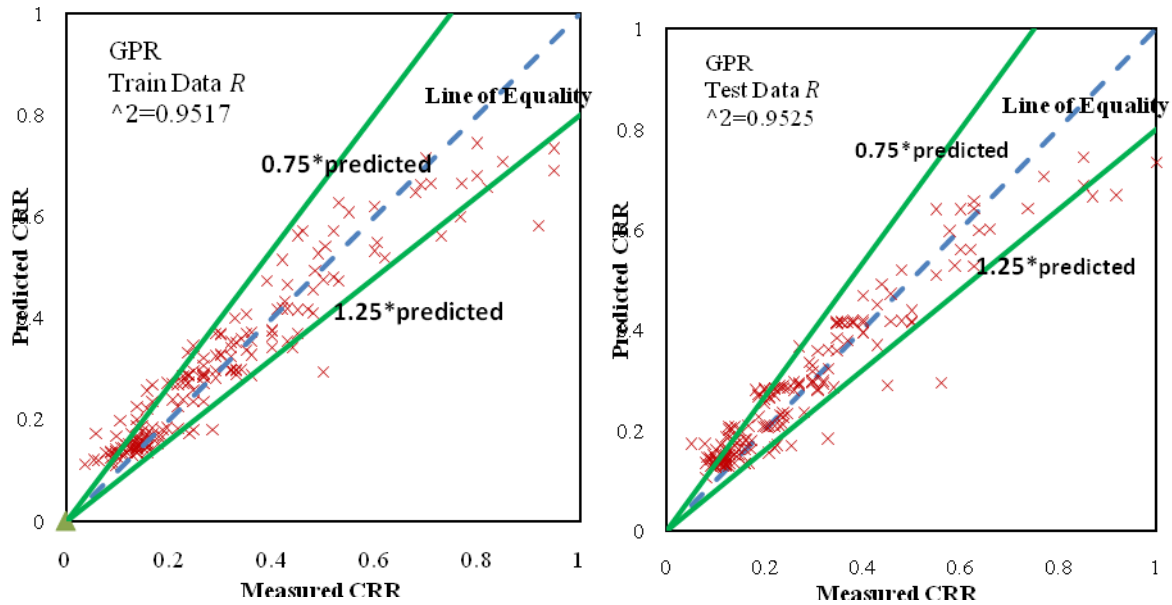


Figure 4: Comparison of predicted Vs actual CRR value for GPR method

4.3. Modelling using MGGP

Multi Gene Genetic programming was applied using GPTips in MATLAB. The configuration script file was modified by changing values of population, number of generations, maximum tree depth, size of tournament for parent selection, number of genes, etc. The Population size was varied from 100 to 1000, the number of generations from 25 to 500, maximum number of genes was varied from 2 to 6 and tournament size was varied from 2 to 7. The best values for these parameters were obtained by trial and error. The parameters giving best model are as given in Table 3. Equation 4 is the model developed using MGGP. The model excluded e_{max} term. This might be because it is already accommodated in the relative density (D_r).

$$CRR = C_1 D_r EXP(\eta_s^{0.5}) EXP(\cos(\sigma_{mc})) (D_{50} + \varepsilon_a) + C_2 e_{min} (2\varepsilon_a - 354) (\varepsilon_a - \eta_s + \sin(\varepsilon_a)) + C_3 D_r \varepsilon_a^2 EXP(\cos(\sigma_{mc})) \ln(N) (U_c + \eta_s) + C_4 \quad (4)$$

The RMSE value for train data using MGGP is 0.052 and the same for test data in 0.051. The MGGP model gave high correlation coefficient for both train and test data which is 0.94 and 0.96 respectively. The predicted Vs actual value plot for this model is given as Figure 5. The comparison of error values of different models is given in Table 4

Table 3: Run parameter values for the best model

Run parameter	Value
Population size	250
Max. Generations	100
Tournament size (Parent selection)	4
Elite fraction	0.15
Max. Genes	3
Max. tree depth	4
Crossover Probability	0.84
Mutation Probability	0.14
Function set	TIMES MINUS PLUS SQRT SQUARE SIN COS EXP ADD3 MULT3 LOG

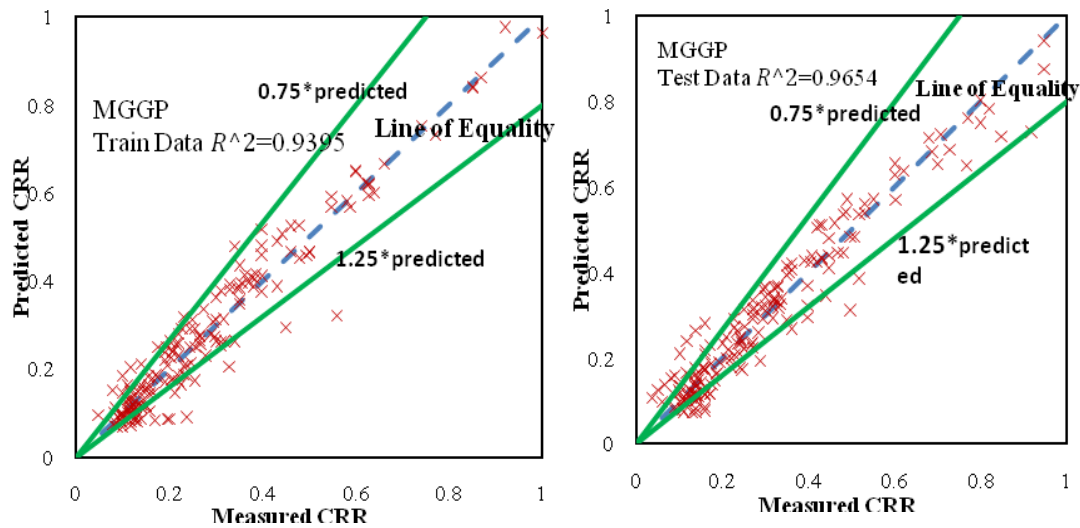


Figure 5. Predicted Vs actual CRR for MGGP Model

Even though the correlation coefficient is high in the testing data, we observe an underprediction in CRR value towards the higher side. Since a soil is considered to have a possibility to liquify if $CRR < CSR$, this underprediction is on the conservative side.

Table 4: Comparison of performance of AI methods for CRR prediction

Algorithms	Train			Test		
	R^2	RMSE	MAE	R^2	RMSE	MAE
MGGP	0.940	0.052	0.011	0.965	0.051	0.231
M5 Model tree	0.948	0.064	0.043	0.948	0.062	0.197
GPR	0.952	0.066	0.005	0.953	0.063	0.265

5. Conclusion

The study aimed at developing models for the prediction of CRR from laboratory test data using artificial intelligence techniques. The advantage with these methods is that once they are trained

properly, they can be used for predicting the CRR value of any test data without any manual work. 346 data points from cyclic triaxial test and cyclic direct shear test were used in MGGP, GPR and M5' model tree, at 50% test to 50%train ratio. All the methods showed effective modelling capabilities. MGGP developed a prediction Equation while M5' model tree algorithm generated a model tree with if then conditions incorporating linear modelling expressions for each condition. All the methods generated models with high correlation coefficient value, with maximum of 0.96 for MGGP. From the predicted Vs actual value plots, it was seen that there are more underprediction possibility than over prediction for higher values of CRR, which is to the conservative side as far as the factor of safety of liquefaction is concerned. Therefore, the present study demonstrates the use of non-mechanistic methods to understand the response of soil, without knowing the fundamental relation between various parameters involved in a traditional, mechanistic point of view.

References

- [1] Samui, P., and T. G. Sitharam. Machine learning modelling for predicting soil liquefaction susceptibility. *Natural Hazards and Earth System Sciences* 11, no. 1 (2011): 1-9.
- [2] Juang, C. H, Yuan H, Lee D-H, and Lin Ping-Sien. Simplified cone penetration test-based method for evaluating liquefaction resistance of soils. *Journal of Geotechnical and Geo-environmental engineering* 129, no. 1 (2003): 66-80.
- [3] Seed HB, and Idriss IM. Simplified procedure for evaluating soil liquefaction potential. *Journal of Soil Mechanics & Foundations Div* 97(9) (1971), 1249–1273.
- [4] Robertson PK, and. Wride CE. Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian geotechnical journal* 35, no. 3 (1998): 442-459.
- [5] Goh ATC. Probabilistic neural network for evaluating seismic liquefaction potential. *Canadian Geotechnical Journal* 39, no. 1 (2002): 219-232.
- [6] Rezania M, Javadi AA, and Giustolisi O. Evaluation of liquefaction potential based on CPT results using evolutionary polynomial regression. *Computers and Geotechnics* 37, no. 1-2 (2010): 82-92.
- [7] Kayadelen C. Soil liquefaction modeling by genetic expression programming and neuro-fuzzy. *Expert Systems with Applications* 38, no. 4 (2011): 4080-4087.
- [8] Muduli PK., and Das SK. Evaluation of liquefaction potential of soil based on standard penetration test using multi-gene genetic programming model. *Acta Geophysica* 62, no. 3 (2014): 529-543.
- [9] Muduli PK, and Das SK. CPT-based seismic liquefaction potential evaluation using multi-gene genetic programming approach. *Indian Geotechnical Journal* 44, no. 1 (2014): 86-93.
- [10] Kim Young-Su, and Kim Byung-Tak. Use of artificial neural networks in the prediction of liquefaction resistance of sands. *Journal of Geotechnical and Geo-environmental Engineering* 132, no. 11 (2006): 1502-1504.
- [11] Sharafi H, and Jalili S. Assessment of Cyclic Resistance Ratio (CRR) in Silty Sands Using Artificial Neural Networks. *Open Journal of Civil Engineering* 4, no. 03 (2014): 217.
- [12] Sabbar A S, Chegenizadeh A, and Nikraz H. Prediction of Liquefaction Susceptibility of Clean Sandy Soils Using Artificial Intelligence Techniques. *Indian Geotechnical Journal* 49, no. 1 (2019): 58-69.
- [13] Pal M, and Deswal S. Modelling pile capacity using Gaussian process regression. *Computers and Geotechnics* 37, no. 7-8 (2010): 942-947.
- [14] Karthikeyan, J, and Samui P. Determination of Strain Energy for Triggering Liquefaction Based on Gaussian Process Regression. *Engineering Journal* 17, no. 4 (2013): 71-78.
- [15] Samui P. Utilization of Gaussian process regression for determination of soil electrical resistivity. *Geotechnical and Geological Engineering* 32, no. 1 (2014): 191-195.
- [16] He Peng, Li Shu-cai, Xiao Jie, Zhang Qian-qing, Xu Fe, and Zhang J. Shallow sliding failure prediction model of expansive soil slope based on Gaussian process theory and its engineering application. *KSCE Journal of Civil Engineering* 22, no. 5 (2018): 1709-1719

- [17] Muduli PK, and Das SK. Evaluation of Liquefaction Potential of Soil Based on Shear Wave Velocity Using Multi-Gene Genetic Programming. In *Handbook of Genetic Programming Applications*, pp. 309-343. Springer, Cham, 2015.
- [18] Javadi A A, Rezanian M, and Nezhad MM. Evaluation of liquefaction induced lateral displacements using genetic programming. *Computers and Geotechnics* 33, no. 4-5 (2006): 222-233.
- [19] Samadi, M, Jabbari E, and Azmathulla H. M. Assessment of M5' model tree and classification and regression trees for prediction of scour depth below free overfall spillways. *Neural Computing and applications* 24, no. 2 (2014): 357-366.
- [20] Avval, YJ, and Derakhshani A. New formulas for predicting liquefaction-induced lateral spreading: model tree approach." *Bulletin of Engineering Geology and the Environment* 78, no. 5 (2019): 3649-3661.
- [21] Williams, CKI, and Rasmussen CE. *Gaussian processes for machine learning*. Vol. 2, no. 3. Cambridge, MA: MIT press, 2006.
- [22] Koza, JR. *Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems*. Vol. 34. Stanford, CA: Stanford University, Department of Computer Science, 1990.
- [23] Searson, DP., David E. Leahy, and Mark J. Willis. GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In *Proceedings of the International multiconference of engineers and computer scientists*, vol. 1, pp. 77-80. Hong Kong: IMECS, 2010.
- [24] Pal, M, and Deswal. M5 model tree based modelling of reference evapotranspiration." *Hydrological Processes: An International Journal* 23, no. 10 (2009): 1437-1443.
- [25] Quinlan, John R. "Learning with continuous classes." In *5th Australian joint conference on artificial intelligence*, vol. 92, pp. 343-348. 1992.