



Gene expression programming models for liquefaction-induced lateral spreading

A. Keshavarz^{a,*} and H. Tofighi^b

a. School of Engineering, Persian Gulf University, Bushehr, Iran.

b. Civil Engineering Group, Islamic Azad University, Bushehr, Iran.

Received 27 December 2017; received in revised form 19 September 2018; accepted 13 November 2018

KEYWORDS

Liquefaction;
 Lateral spreading;
 Gene expression programming;
 Free face;
 Gentle slope.

Abstract. Lateral spreading is one of the most significant destructive and catastrophic phenomena associated with liquefaction caused by earthquake and it can cause very serious damage to structures and engineering facilities. The aim of this study is to evaluate liquefaction-induced lateral spreading and find new relations using Gene Expression Programming (GEP), which is a new and developed generation of genetic algorithms approaches. Since there are complicated, nonlinear, and higher-order relationships among many factors affecting the lateral spreading, GEP was assumed to be capable of finding complex and accurate relationships among the involved factors. This study includes three main stages: (i) compiling available database (484 data); (ii) dividing data into training and testing categories; and (iii) building new models and proposing new relationships to predict ground displacement in free face, gentle slope, and general ground conditions. The results of modeling each of these different ground conditions were presented in the form of mathematical equations. At the end, the final GEP models for 3 different cases of ground conditions were compared with Multiple Linear Regression (MLR) and other published models. The statistical parameters indicated the higher accuracy of GEP models over other relations.

© 2020 Sharif University of Technology. All rights reserved.

1. Introduction

Liquefaction is one of the most interesting and sophisticated seismic geotechnical issues. This phenomenon and various types of failure related to it (such as flow liquefaction and lateral spreading) can impose tremendous damages and losses on infrastructures, buildings and structures, lifelines, and buried structures. In general, liquefaction is considered one of the major causes of ground movement due to earthquake and

occurs as a result of rapid loss of shear strength of the soil due to increasing pore water pressure in saturated soils subjected to static or dynamic loads. Lateral spreading is the most common type of liquefaction-induced ground failure. During this phenomenon, blocks of intact surficial soil are displaced along a shear zone that has formed within the liquefied layer. Due to earthquake and gravitational forces, surficial blocks are transported downslope or in the direction of free face upon reaching mobilization. Lateral spreading can cause horizontal ground displacement from a few centimeters to several meters [1]. This phenomenon can be observed in fields with gentle slopes ranging from 0.3 to 5% and in fields with free face like stream channels and trenches. Lateral spreading depends on several factors: physical and mechanical properties of the soil layers, depth of groundwater, intensity and

*. Corresponding author. Tel./Fax: +98 7733440376
 E-mail addresses: keshavarz@pgu.ac.ir, and
 amin_keshavarz@yahoo.com (A. Keshavarz);
 tofighifarhad@gmail.com (H. Tofighi)

duration of ground shaking, distance from source of seismic energy, and seismic attenuation properties of in situ soil. Due to the participation of a large number of parameters affecting this phenomenon, estimating the liquefaction-induced lateral spreading is one of the most complicated issues of geotechnical engineering [2].

Estimating the deformations and displacements caused by liquefaction is part of soil liquefaction engineering. Unlike the phenomenon of liquefaction, no specific mechanism is recognized in liquefaction-induced lateral spreading. Several factors including seismic characteristics, soil specification and geology, topographical characteristics, etc. are involved in occurrence of lateral spreading. Generally, various methods that have been applied to predicting the magnitude of liquefaction-induced lateral spreading can be divided into 4 categories, as shown in Table 1.

In methods based on Newmark’s sliding block [3–6], according to the simplifications made in modeling the behavior of liquefied soil and considering soil as a rigid body, the obtained equations do not enjoy high accuracy of predicting lateral displacement. In models based on minimum potential energy [7,8], the liquefied soil is assumed to behave as liquid, and these models have received less attention than the sliding block methods.

In order to model all the details of liquefaction-induced lateral spreading, numerical models [9–16] should be able to simulate seismic excitation, softening of the soil due to increase in pore water pressure, rapid decrease in shear strength, and displacement continuation of soil after loading and reconsolidation due to the drainage of additional pore water pressure [17]. Due to the nonlinear behavior of soil and nonlinear relationships between participating parameters, the application of finite element method to modeling this phenomenon is very complicated [18].

Methods based on experimental results [12,14–16,19,20] also require a large number of the data obtained from tests. In these methods, preparation of high-quality specimens and simulation of shear strains

are of utmost importance and in practice, it is difficult to provide all the in-situ conditions at the laboratory.

Given the large number of parameters including seismological, topographical, and geotechnical parameters and nonlinear relationships among them, researchers have vastly used empirical and parametric methods. Empirical methods [1,21–33] are used based on the collected data after earthquakes. Because of their simplicity, they are widely used in predicting the lateral spreading. Most of these relations have been obtained using Multiple Linear Regression (MLR) method and they usually offer separate relations for gentle slope and free face ground conditions [17].

In complex problems where the relationship between variables is unknown, optimization algorithms such as machine learning-based methods are very powerful predictive tools for solving the problems as long as they can simulate the very nature of the problem. Lateral spreading is one of the complex issues of engineering problems and a large number of parameters are involved in its occurrence. Therefore, Artificial Neural Networks (ANNs) [18,34–36], Genetic Programming (GP) [2,37,38], and other machine learning techniques can be significantly viable tools for solving these problems. Table 2 lists a number of relationships between empirical and machine learning-based methods.

Due to the ability of finding complex relationships among multivariate problems, Gene Expression Programming (GEP) can be a useful tool for geotechnical issues. Johari et al. [39] applied GEP to predicting effective stress parameter of unsaturated soils. Keshavarz and Mehrmiri [40] utilized GEP to model the normalized shear modulus and damping ratio of sands. This method was also used to predict the maximum lateral displacement of retaining wall [41] and soil-water characteristic curve [42].

In this paper, a new approach was presented to evaluate the liquefaction-induced lateral spreading using GEP. Modeling liquefaction-induced lateral displacement by GEP method was carried out using

Table 1. Available methods of lateral spreading predictions.

Analytical methods	Based on Newmark’s sliding block [3–6]
	Based on minimum potential energy [7,8]
Numerical methods	Based on finite element methods [9–12]
Empirical and semi-empirical methods	Based on in situ collected data [1,21–29]
	Based on laboratory methods [12,19,20]
Machine learning approaches	Artificial neural networks [18,34]
	Genetic programming [2,37,38]

Table 2. Proposed equations to predict liquefaction-induced lateral spreading.

$D = 0.75\sqrt{H}\sqrt[3]{\theta}$	Hamada et al. [21]
$\log LSI = -3.49 - 1.86 \log R + 0.98M_w$	Youd and Perkins [22]
Free face:	
$\log D_H = -16.366 + 1.178M - 0.927 \log R - 0.013R + 0.657 \log W$ $+ 0.348 \log T_{15} + 4.527 \log(100 - F_{15}) - 0.922D50_{15}$	Bartlett and Youd [1,23]
Gentle slope:	
$\log D_H = -15.787 + 1.178M - 0.927 \log R - 0.013R + 0.429 \log S$ $+ 0.348 \log T_{15} + 4.527 \log(100 - F_{15}) - 0.922D50_{15}$	
Free face:	
$\log D_H = -16.713 + 1.532M - 1.406 \log R^* - 0.012R + 0.592 \log W$ $+ 0.540 \log T_{15} + 3.413 \log(100 - F_{15}) - 0.795 \log(D50_{15} + 0.1 \text{ mm})$	Youd et al. [27]
Gentle slope:	
$\log D_H = -16.213 + 1.532M - 1.406 \log R^* - 0.012R + 0.338 \log S$ $+ 0.540 \log T_{15} + 3.413 \log(100 - F_{15}) - 0.795 \log(D50_{15} + 0.1 \text{ mm})$	
Free face:	
$D_{hc} = -\frac{163.1}{M^2} + \frac{57}{RF_{15}} - \frac{0.0035T_{15}^2}{WD50_{15}^2} + \frac{0.02T_{15}^2}{F_{15}D50_{15}^2} - \frac{0.026T_{15}^2}{F_{15}^2} - 0.006T_{15}^2$ $- 0.0013W^2 + 0.0002M^2WT_{15} + 3.7$	Javadi et al. [2]
Gentle slope:	
$D_{hc} = -0.8\frac{F_{15}}{M} + 0.0014F_{15}^2 + 0.16T_{15} + 0.112S + 0.04\frac{ST_{15}}{D50_{15}}$ $- 0.026RD50_{15} + 1.14$	
Free face:	
$D_H = -\frac{2.1414\sqrt{RW}}{M^2\sqrt{D50_{15}}} - \frac{0.061863T_{15}F_{15}}{\sqrt{MW}} - \frac{11.1201M^2}{RF_{15}\sqrt{W}} + \frac{0.0017573M^2T_{15}\sqrt{W}}{D50_{15}\sqrt{F_{15}}} + 1.9671$	
Gentle slope:	
$D_H = -\frac{1.6941\sqrt{T_{15}F_{15}}}{M^2\sqrt{D50_{15}}} - \frac{0.78905T_{15}\sqrt{RSF_{15}}}{M^2} + \frac{0.036036MT_{15}\sqrt{S}}{\sqrt{D50_{15}}}$ $- \frac{2.254210^{-12}T_{15}^2D50_{15}^2\sqrt{M}}{F_{15}^2\sqrt{RS}} + 0.85441$	Rezania et al. [37]

484 datasets, compiled by Youd et al. [27] for gentle slope, free face, and general ground conditions and for each of these different ground conditions, a single mathematical expression was established. The results of the suggested relationships were compared to those obtained by MLR and other methods and the benefits of the proposed method were discussed in detail.

2. Gene Expression Programming (GEP)

GEP was fabricated by Ferreira [43]. GEP is a type of evolutionary algorithms inspired by biological systems. Like Genetic Algorithms (GAs) and GP, using population of individuals, selecting them according to fitness, and using genetic operators create new offspring in the search space and they move toward points with better fitness values [43]. Compared with GAs and GP, individuals in GEP are more developed. GA individuals are composed of linear strings with fixed lengths (chromosomes). Despite GAs, individuals (chromosomes) in GP are more complex and nonlinear

entities and are of different shapes and sizes (parse trees) [43]. Entities in GAs and GP work both as genotype and phenotype simultaneously. GEP combines simple linear chromosomes of fixed lengths, like the entities used in GAs, branched structures (Expression Trees (ETs)) with different shapes and sizes like GP, and parse trees to create a more complete program with high similarity to natural biological systems and better performance. Chromosomes and ETs are the main agents in GEP. ETs function as phenotypes and are encoded in fixed-length linear chromosomes (genotype). Selection is based on fitness of ETs. According to fitness of ETs, chromosomes will be selected to create new offspring using genetic modifications. Although selection of individuals is based on fitness of ETs, it is the chromosomes, not ETs, that are reproduced [43].

Information is encoded in linear entities of fixed length (chromosomes) and is decoded into ETs due to the translation process. Genetic codes in chromosomes are composed of members of functions and terminals sets. Each symbol in the chromosome (genetic code)

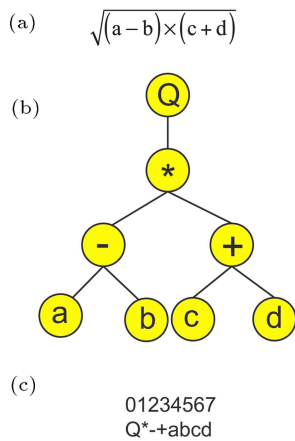


Figure 1. The expression tree and open reading frame of a mathematical expression [44].

forms a part of spatial organization of the corresponding ETs due to the translation process [44]. GEP chromosomes are usually composed of multiple genes of equal length. Each gene forms a sub ET, and the sub ETs together form a more complex multi-subunit ET.

Understanding the structural organization of GEP genes requires deeper familiarity with Open Reading Frames (ORFs). An algebraic or mathematical expression (Figure 1(a)) can be expressed in the form of an ET (Figure 1(b)). As mentioned, ET is in fact the phenotype of GEP individuals. In addition, through straightforward reading of the ET levels from left to right and from top to bottom, the ORF (genotype) can be inferred easily (Figure 1(c)). Expression presented in Figure 1(c) is an ORF, starting at “Q” and terminating at “d”. An ORF in GEP language is called K-expression [43]. Expressing an ORF into an ET is also very simple and straightforward. To express an ORF to an ET, the rules governing the spatial distribution of functions and terminals are dominant. Consider Figure 1 in reverse order; the start position of the ORF (Figure 1(c)) forms the root of the ET at the first line of the ET. As a rule, depending on the number of arguments of each element in a line (functions have different number of arguments, while terminals have no arguments), nodes will be formed in the next lines and by reading ORF from left to right, the new nodes are filled consecutively with the elements of the ORF. This process will be terminated when a line containing only terminals is formed (Figure 1(b)) [44].

GEP chromosomes are usually composed of more than one gene of equal length. Each gene codes for a sub-ET and the sub-ETs interact with one another, thus forming a more complex multi-subunit ET (Figure 2) [43]. It should be noted that genes in a chromosome have equal length, but multigenic chromosomes have different ORFs that are compared with each other. The chromosomes of GEP contain several ORFs, each ORF coding for a structurally and

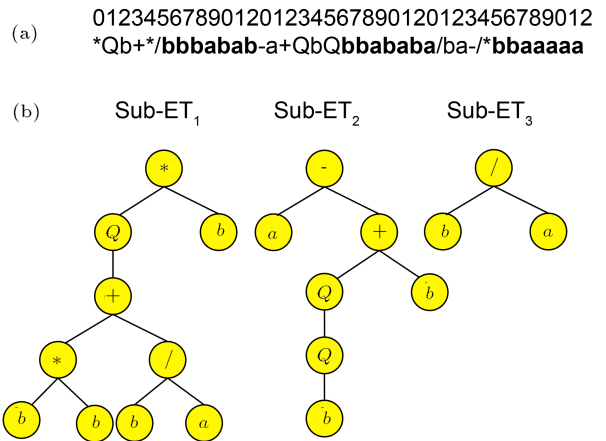


Figure 2. The three-genic chromosome: (a) The open reading frame and (b) the expression tree of each gene. The tail of each gene is shown in bold [44].

functionally unique sub-ET [44]. Complete expression of the genetic information requires the interaction of these sub-ETs with one another. One of the simplest interactions is the linking of sub-ETs by a particular function. In GEP, linking functions include addition (+) and multiplication (*) for algebraic expressions and If and OR for Boolean expressions [43].

GEP uses mutation, transposition, and recombination (crossover) to make variations in individuals and find the solutions. Due to the unconstrained genotype/phenotype mapping of GEP, several genetic operators can be easily implemented; Ferreira [43] used seven operators: mutation, three kinds of transposition (Insertion Sequence, IS; Root Insertion Sequence, RIS; and gene transposition), and three kinds of recombination (one-point, two-point, and gene recombination).

Ferreira [45] investigated genetic operators and found that mutation had the greatest impact by far among other genetic operators and RIS, IS, two-point recombination, one-point recombination, and gene recombination had the highest impact on the performance of chromosomes.

3. Data collection

Finding an empirical or semi-empirical model requires compiling valid and high-quality data. Concerning liquefaction-induced lateral spreading, researchers have used various data and parameters and Youd et al. [27] compiled the most reliable and frequently used datasets. Bartlett and Youd [1] compiled case histories of lateral spreading consisting of data obtained from different earthquakes. Based on Standard Penetration Test (SPT) and by using MLR, they provided a relatively better relationship between variables than existing relationships. Youd et al. [27] compiled a dataset consisting of 484 data from 11 different earth-

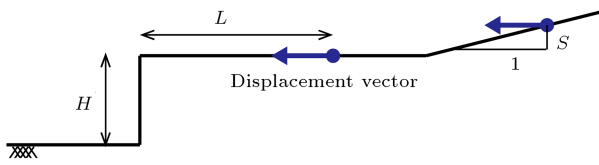


Figure 3. Parameters associated with free face and gentle slope conditions.

quakes by revising the previous dataset. Using this collection, they improved their previous equations.

In this study, the dataset and variables compiled by Youd et al. [27] were used to estimate the liquefaction-induced lateral spreading. For each measured displacement (D_{hc}) in the dataset, there are seven specifications obtained from tests and known as independent variables. Variables can be divided into three categories: seismological, topographical, and geotechnical variables. Seismological variables are associated with the intensity and duration of strong ground motion including: the moment magnitude of the earthquake (M_w), the nearest distance to the seismic energy source (R (km)), and the modified source distance (R^*). Topographical variables include free face ratio (W (H/L (%)), L : horizontal distance from the toe of free face to the displacement vector (m), H : height of the free face (m)), and the ground slope (S (%)). These variables were used to determine the location of lateral spreading and the boundary conditions of the ground (Figure 3). Geotechnical variables include cumulative thickness of saturated cohesionless soil layers with corrected SPT number (N_1)₆₀ less than 15 (T_{15} (m)), the average fines content for granular materials included within T_{15} (F_{15}), and the average mean size of granular materials within T_{15} ($(D_{50})_{15}$).

Statistical characteristics of variables (such as minimum, maximum, and average values) were used for data evaluation. Figure 4 shows the distribution graph of variables. As seen earlier, in most cases, lateral spreading occurred by earthquakes of moment magnitude ranging from 6.5 to 7.5 and at a distance less than 25 km away from the source of seismic energy. It was shown that most of the lateral spreading events occurred on a slope less than 3% and a free face ratio less than 10%. Moreover, the measured ground displacements (D_{hc}) for most of the case histories were below 2 m. The average value of each variable is written in the corresponding chart, and the dominant cases in each chart imply that most of the observed displacements have such specifications. Given that searching for solutions is carried out in wide spaces with many specifications and since the aim of this study is to find a proper relationship between D_{hc} and other independent variables, we expect the results to be more accurate and reliable in regions with dominant specifications.

Table 3. Classification of training and testing data in different ground conditions.

Ground condition	Number of case histories	Training data	Testing data
Free face	229	190 (82.97%)	39 (17.03%)
Gentle slope	255	198 (77.65%)	57 (22.35%)
General	484	388 (80.17%)	96 (19.83%)

In machine learning techniques, the available data are divided into two categories of training dataset and testing dataset. The training dataset was used as the selection environment and by using this set of data, the algorithm would attempt to adapt to the problem environment as well as learn and discover the relationships between variables. This dataset should be chosen in such a way that it can be representative of the whole problem space. Obviously, with a larger number of members of the training dataset and less dispersal of the data in training dataset, better relationships can be identified by the algorithm. The testing dataset is also used for validation of the relations obtained by the training dataset. In this study, as indicated in Table 3, the data were separated into three different types of ground conditions: gentle slope grounds, grounds with free face, and grounds with a general state.

The values of the Minimum (Min), average (Avg), Standard Deviation (SD), and Maximum (Max) of all independent and dependent variables in both of the training and testing sets are shown in Tables 4, 5, and 6 for the general, free face, and gentle slope ground conditions, respectively. The comparison shows the fact that the values of these statistical parameters are close in both training and testing datasets.

4. Modeling using GEP

The aim of modeling in GEP is to find and build different relationships between variables. Necessary parameters for modeling can be divided into three categories as follows:

- Population parameters including the number of individuals (chromosomes) in population and the characteristics of chromosomes (number of genes of each chromosome, the size of the head of genes, the gene's linking function, and the number of numerical constants of each gene);
- Control parameters including termination criterion of the program and appropriate fitness function;

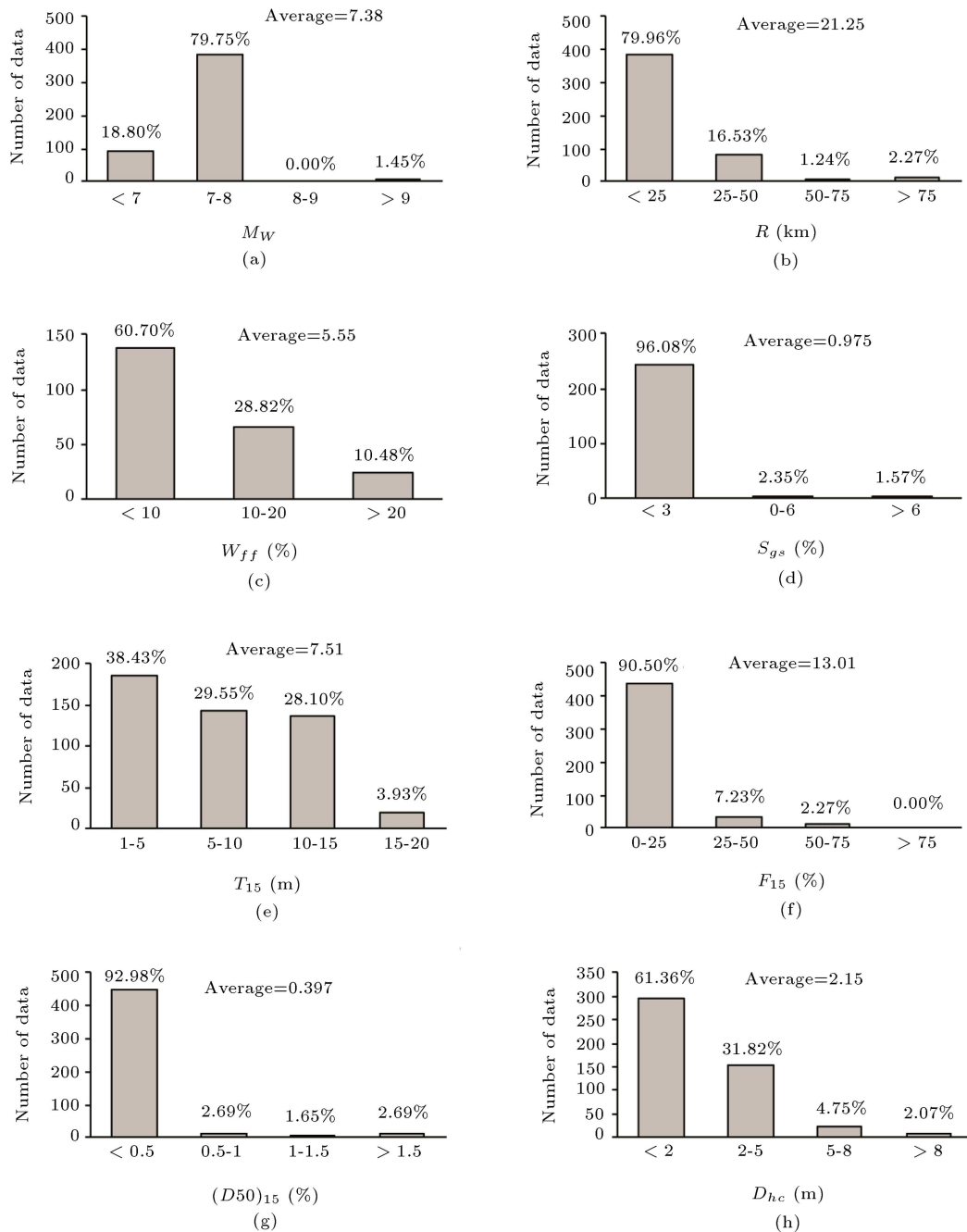


Figure 4. Distribution graph of variable values.

- Operating parameters including the rate of genetic operators, mathematical and logical functions, etc. to create a function set.

Conceptually, modeling can be divided into two parts: genetic modeling and mathematical modeling. Genetic modeling is that part of the modeling carried out by the user. Mathematical modeling is the result of genetic modeling, and the function of each genetic model is evaluated on the basis of its mathematical model. In GEP, mathematical models are built based on input

data and according to the values determined for genetic parameters. The function of GEP in dealing with a problem depends on the genetic model and the values selected for genetic parameters. Therefore, selecting the most appropriate and optimal values for genetic parameters is the main issue that we are dealing with GEP modeling, because the program may not reach the desired result using a set of these values, or it may take much longer for the process of the program to bear fruit. The best model in GEP was obtained by comparing different built mathematical models with a

Table 4. Comparison of statistical parameters for training and testing data sets in general ground conditions.

General	Training data				Testing data			
	Min	Avg	SD	Max	Min	Avg	SD	Max
$\log(D_{hc})$	-2.0	0.16	0.48	1.01	-2.0	0.17	0.47	0.94
M	6.4	7.36	0.45	9.2	6.4	7.45	0.47	9.2
R	0.2	20.61	12.72	100	0.2	23.83	17.11	100
$\log(R^*)$	0.13	1.38	0.41	2.66	0.23	1.46	0.4	2.66
$\log(W_{ff})$	0.0	0.44	0.51	1.75	0.0	0.35	0.47	1.62
$\log(S_{gs})$	-1.3	-0.11	0.28	1.04	-1.0	-0.14	0.28	1.04
$\log(T_{15})$	-2	0.75	0.38	1.22	-0.3	0.8	0.32	1.29
$\log(100 - F_{15})$	1.48	1.93	0.08	2.0	1.61	1.94	0.07	2.0
$\log(D50_{15} + 0.1)$	-0.87	-0.39	0.22	1.08	-0.79	-0.39	0.19	0.57

Table 5. Comparison of statistical parameters of training and testing data sets in grounds with free face conditions.

Free face	Training data				Testing data			
	Min	Avg	SD	Max	Min	Avg	SD	Max
$\log(D_{hc})$	-2.0	0.18	0.53	1.01	-2.0	0.14	0.6	0.94
M	6.4	7.19	0.51	9.2	6.4	7.34	0.6	9.2
R	0.5	17.52	14.26	95.0	0.5	22.56	18.7	100
$\log(R^*)$	0.21	1.24	0.49	2.62	0.21	1.39	0.48	2.66
$\log(W_{ff})$	0.0	0.9	0.34	1.75	0.31	0.86	0.33	1.61
$\log(T_{15})$	-0.7	0.81	0.38	1.22	-0.3	0.84	0.33	1.19
$\log(100 - F_{15})$	1.48	1.91	0.09	2.0	1.72	1.92	0.06	1.98
$\log(D50_{15} + 0.1)$	-0.87	-0.42	0.26	0.32	-0.77	-0.45	0.19	0.2

Table 6. Comparison of statistical parameters of training and testing data sets in grounds with gentle slope conditions.

Gentle slope	Training data				Testing data			
	Min	Avg	SD	Max	Min	Avg	SD	Max
$\log(D_{hc})$	-2.0	0.15	0.43	0.65	-1.22	0.18	0.36	0.73
M	6.4	7.53	0.29	9.2	6.4	7.53	0.33	9.2
R	0.2	23.57	10.19	100	0.2	24.69	15.87	100
$\log(R^*)$	0.13	1.52	0.25	2.66	0.13	1.5	0.34	2.66
$\log(S_{gs})$	-1.3	-0.21	0.36	1.04	-1.0	-0.24	0.33	1.04
$\log(T_{15})$	-2.0	0.69	0.38	1.15	0.0	0.77	0.32	1.29
$\log(100 - F_{15})$	1.51	1.95	0.07	2.0	1.61	1.95	0.07	2.0
$\log(D50_{15} + 0.1)$	-0.8	-0.37	0.18	1.08	-0.79	-0.35	0.17	0.57

model with lower error rate, and higher fitness was considered to be a better model. Therefore, having the best genetic model (the most optimal parameter values), the mathematical relationship between the variables was determined.

As mentioned previously, modeling involves determining the optimal values of genetic parameters and by using these parameters, the program moves forward in an evolutionary process. Then, the generated genetic models are expressed in mathematical terms, and fitness of each model is evaluated on the basis

of generated mathematical relations. In other words, modeling can be introduced as an attempt to find the best values of genetic parameters in a step-by-step process.

In this research, the modeling of lateral spreading was carried out using GEP in three separate phases. There are different steps in each phase, and each step involved determining the optimal value of one of the genetic parameters. Thus, in each step, the value of one of the parameters was considered as a variable and by changing the value of that variable, different models

Table 7. The values of genetic parameters of selected models.

	Specification of genetic parameters	Free face	Gentle slope	General
1	Fitness function	RMSE	RMSE	RMSE
2	Number of generations	100000	100000	100000
3	Number of chromosomes	90	70	50
4	Number of genes	13	12	13
5	Head size	14	5	13
6	Linking function	+	+	+
7	Mutation rate	0.03	0.025	0.005
8	Inversion	0.1	0.1	0.1
9	IS transposition	0.1	0.1	0.1
10	RIS transposition	0.1	0.1	0.3
11	One-point recombination	0.3	0.3	0.3
12	Two-point recombination	0.1	0.6	0.1
13	Gene recombination	0.1	0.1	0.1
14	Gene transposition	0.1	0.1	0.1
15	Random numerical constants per gene	$2 \in [-10, 10]$	$2 \in [-10, 10]$	$2 \in [-10, 10]$

were built; finally, through the comparison of fitness and statistical errors of the models, the best model of that step and thus, the most optimal value of that variable could be identified. Evolutionary algorithms were of random-evolutionary nature and a program might not evolve appropriately over certain generations. To avoid this situation, the chosen model of each step was executed 5 times again, and the average value of fitness and its error were calculated. At the next step, these mean values were compared with the previously built models again, and the models with the values close to the mean values were also applied 5 times again; then, the average values of the models were compared. The aim of following this trend is to reduce the impact of random-evolutionary nature of algorithm on the results. By determining the value of a variable at each step, the process goes through the next step to determine the optimum value of the next variable and therefore, at the end of each phase, the optimum values of all variables and thereby, the optimized model were identified using the stepwise method. At the end of each phase, the variables with no effect on the results would be removed from the list of variables of the next phase. It should be noted that the model chosen in every step was considered as a prototype for the next step and also, the final model of each phase was sent to the next phase as a prototype. More information about how to model and find the optimal values of population, control, and operating parameters and how they will affect the results are available in [43,44].

Based on provided explanations, genetic modeling was carried out for three different types of ground conditions with emphasis on Root Mean Square Error (RMSE) as the fitness function. Number of generations is a termination criterion that allows populations to be built and evolved due to a certain number of generations. It is obvious that more generations provide more opportunity to evolve for individuals; however, due to time problems, it is set to 100000 in this study.

Modeling in the first phase is composed of 8 steps: determining the number of population individuals (chromosomes), the number of genes of each chromosome, the head size of each gene, the mutation operator rate, the rate of Root Insertion Sequence (RIS) operator, the rate of the two-point recombination operator, the rate of the Insertion Sequence of elements (IS) operator, and the rate of the gene recombination operator, respectively. By removing the step associated with determining the rate of IS operator, the second phase is reduced to 7 steps. Moreover, by eliminating the steps involved in determining the rate of two-point recombination and gene recombination operators, the third phase is also performed through 5 steps. Thus, at the end of the modeling process in the third phase, the genetic characteristics of the chosen models related to each of different ground conditions are given in Table 7. According to these genetic characteristics for each of different ground conditions, a final model is identified.

Table 8. The simplified form of the variables used in evaluation of liquefaction-induced lateral spreading.

Variable	<i>M</i>	<i>R</i>	$\log(R^*)$	$\log(S_{gs})$	$\log(W_{ff})$	$\log(T_{15})$	$\log(100 - F_{15})$	$\log(D50_{15} + 0.1)$
Abbreviated	<i>M</i>	<i>R</i>	<i>r</i>	<i>s</i>	<i>w</i>	<i>T</i>	<i>F</i>	<i>D</i>

5. Results

To avoid lengthy built relationships, a simple and concise form of the variables was used in relations, as shown in Table 8. The performance and efficiency of the obtained relations were compared with those of other equations using correlation coefficient (R^2) criteria, RMSE, and Mean Absolute Error (MAE), as defined below:

$$R_i = \frac{n \sum_{j=1}^n (T_j P_{ij}) - \sum_{j=1}^n T_j \sum_{j=1}^n P_{ij}}{\sqrt{\left[n \sum_{j=1}^n T_j^2 - \left(\sum_{j=1}^n T_j \right)^2 \right] \left[n \sum_{j=1}^n P_{ij}^2 - \left(\sum_{j=1}^n P_{ij} \right)^2 \right]}}, \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (P_{ij} - T_j)^2}, \tag{2}$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |P_{ij} - T_j|. \tag{3}$$

In these equations, *i* index represents the program number for which the fitness and errors are evaluated; P_{ij} is the value predicted by the individual program *i* for fitness case *j* (out of *n* fitness cases); and T_j is the target value for fitness case *j*.

5.1. Grounds with free face

As already noted, for the grounds with free face, 229 datasets were obtained from SPT, of which training and testing data constituted 82.97% (190 data) and 17.03% of the dataset (39 data), respectively. Following a considerable amount of time devoted to modeling and comparing developed models in terms of fitness values, complexity, and length of equations, the following equation was selected as the ultimate model for the grounds with free face:

$$\begin{aligned} \log D_{hc} = & - \frac{0.387FwR}{M^2} + \frac{(F - 1)^2}{M} - \frac{T}{M} \\ & + F^{1 - \frac{\sqrt{(w - 7.329)^2(F - T) - r}}{r}} \\ & + w \left(F - \sqrt{\frac{w}{M}(D - w - F - 9.588)} \right) \\ & + (F + R)^{w - \sqrt{5.031 + R^*r - T - F}} \end{aligned}$$

$$\begin{aligned} & + \left(\frac{r}{10^{\exp(0.035FD)}} \right)^{D+1.524} \\ & + 10^{\frac{1.616r - w - D - 3.212}{F} - T} \\ & - \exp(2 \exp(-1.811(w + r))) + \sqrt{T - D} \\ & + 2.927F - r + M - 12.406. \tag{4} \end{aligned}$$

Figure 5 shows the comparison of the correlation coefficient (R^2) of training data obtained using MLR method by Youd et al. [27] and the GEP method proposed in the present study. The obtained value of R^2 in training data was equal to 90.3% and 84.1% as calculated by GEP and MLR methods, respectively, reflecting the higher accuracy of the equation developed by GEP.

Figure 5 also shows the R^2 values in validation or testing data which have been estimated in GEP and MLR methods as 79.9% and 82.5%, respectively. Moreover, Figure 6 shows the comparison of the results of GEP model and those of the models of Rezania et al. [37] and Javadi et al. [2]. Referring to Table 9 and checking other statistical criteria indicate the higher accuracy of the equation obtained from GEP method than other methods. As is clear from Figure 5, the GEP results are characterized by significantly higher accuracy for the data up to 5 m, especially for the data up to 2 m. As previously noted, this result was predictable according to the frequency of the data collected in the mentioned range.

It should be noted that GEP managed to develop a model with much higher accuracy ($R^2 = 92.2\%$) in the training data. However, it was not briefed here due to the higher complexity of the obtained equation than the existing equation. Furthermore, in this study, only about 80% of the data were used to create the model and the other data were used for validation, whereas in MLR method, all data were used to create the model and validation was not performed using the other datasets.

5.2. Grounds with gentle slope

Of all the 255 data sets included in this category, 77.65% (198 data) and 22.35% (57 data) were considered as training and testing data, respectively. Following the comparison of the models developed in terms of fitness values, complexity, and length of equations, the following equation was given as the final model for the gentle slope ground conditions:

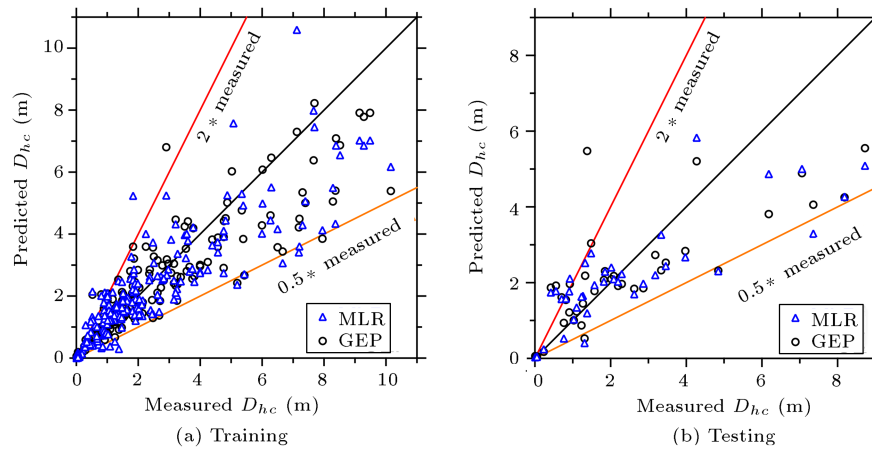


Figure 5. Comparison of the results of models created using Multiple Linear Regression (MLR) [27] and Gene Expression Programming (GEP) with the measured values of data in grounds with free face: (a) Training data and (b) testing data.

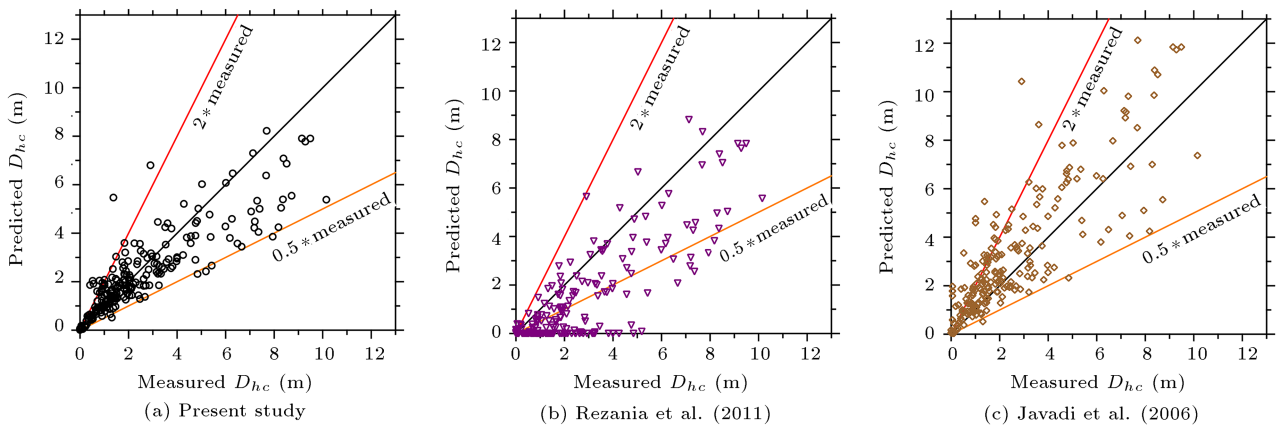


Figure 6. Comparison of the results of the model developed by Gene Expression Programming (GEP) in grounds with free face: (a) Present study, (b) model developed by Rezania et al. [37], and (c) model developed by Javadi et al. [2].

$$\log D_{hc} = \frac{1}{3.805R} - \frac{\sqrt{R}}{\exp F} - 10^{-3.058}(R + s) + \cos \cos \exp r + \cos F^F + \sqrt[3]{10^{10-1.992+s-D}} + 10^{\frac{M-T-6.271}{16}} + \sqrt{s - F + M} - \sqrt[3]{M} - MTF^{-7.708} - 0.492FD(M - 6.855) + T + 3F - 10.061. \tag{5}$$

Variables used in this equation are introduced in Table 8. A comparison between Eq. (5) and the equation obtained from MLR method of Youd et al. [27] was made in Figure 7. In the training dataset, GEP method created an equation with a correlation coefficient of 88.7%, while the correlation coefficient of the MLR equation was 83.7%. In addition, the correlation coefficients of the validation dataset for both GEP and MLR equations were 79.02% and 79.3%, respectively. In Figure 8, the results of the model developed by

GEP were shown along with the models developed by Rezania et al. [37] and Javadi et al. [2] for all the data in gentle slope ground conditions. The results presented in Table 9 indicate the better accuracy of the equation obtained by GEP method than the other methods.

In this case of ground conditions, GEP managed to develop a model with a correlation coefficient of 90.16% in training data, which is not briefed here due to the higher complexity and length of the equation.

5.3. General condition

Another option examined in this study is lateral spreading in grounds with both gentle slope and free face conditions. Therefore, all available data (484 data) were used to present a general equation that can be used both in gentle slopes and free faces. In this case of ground conditions called general condition, 388 (80.17% of) data were used as the training data and the other data (96 (19.83% of) data) were used for validation. Thus, with a trend similar to that for the grounds with free face and gentle slope, the

Table 9. Statistical comparison of different models results.

Model	Performance criteria								
	R^2 (%)			RMSE			MAE		
	Train	Test	Total	Train	Test	Total	Train	Test	Total
Free face									
MLR ¹	84.1	82.5	83.7	0.214	0.251	0.220	0.161	0.194	0.166
GP ²	73.7	69.3	72.6	1.781	1.566	1.746	1.382	1.269	1.362
EPR ³	70.0	60.5	51.2	1.776	1.996	2.799	1.374	1.488	2.128
GEP ⁴	90.3	79.9	88.11	0.168	0.272	0.190	0.123	0.202	0.137
Gentle slope									
MLR ¹	83.7	79.3	82.9	0.175	0.162	0.172	0.132	0.131	0.132
GP ²	60.5	58.4	59.8	0.527	0.697	0.621	0.470	0.562	0.490
EPR ³	59.5	59.7	59.4	0.624	0.689	0.640	0.496	0.548	0.508
GEP ⁴	88.7	79.02	87.0	0.145	0.165	0.149	0.118	0.132	0.122
General									
ANN ⁵	95	85	92	0.52	1.1	0.7	—	—	—
GEP ⁴	89.4	77.6	87.1	0.158	0.222	0.172	0.122	0.173	0.132

1: Youd et al. [27]; 2: Javadi et al. [2]; 3: Rezanian et al. [37]; 4: Present Study;
5: Baziar and Ghorbani [34].

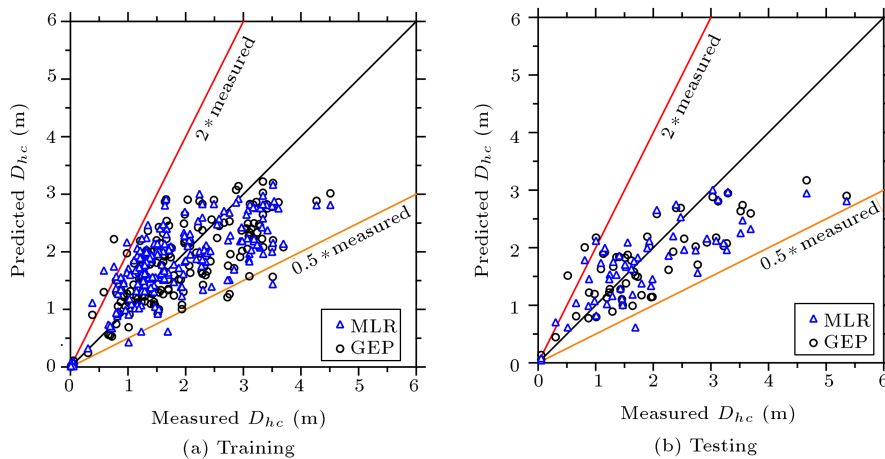


Figure 7. Comparison of the results of models created using Multiple Linear Regression (MLR) [27] and Gene Expression Programming (GEP) with the measured values of data in grounds with a gentle slope: (a) Training data and (b) testing data.

equation identified for this state of the ground is given below. The parameters used in the following equation are described in Table 8:

$$\begin{aligned} \log D_{hc} = & -7.541 + \tan \tan \tan \left(\frac{0.0102}{\tan(F - R - 5.089)} \right) \\ & + \cos^{10} (1.577(r - M)) \\ & + \cos^5 (0.0120(M - 2w - R - 2 \times 0.0120^T) + 7.268) \\ & + \cos \left(7.731 - s - \frac{1}{MD + 7.210} \right)^{0.2} \end{aligned}$$

$$\begin{aligned} & + \cos \left(\sqrt[3]{(4.099 - R - w)^2 \sin w} - w \right)^{\frac{1}{12}} \\ & + \cos^{0.4} \left(\cos^{1.6} (\cos(F - D)^{-7.336} - M)^3 \right) \\ & + \sqrt{\sin \exp (R (T - 10^{4w} - \sin^2(D + 0.899)))} \\ & + \sin w - 0.17 (r + \cos(TsD - Tr)^2) \\ & + r^5 \exp(-\sin(s^2 - 2.163s) \sin F^3 - 4.36) + 2F. \quad (6) \end{aligned}$$

Since no equation was presented by MLR for this type of the ground condition, only the results of GEP

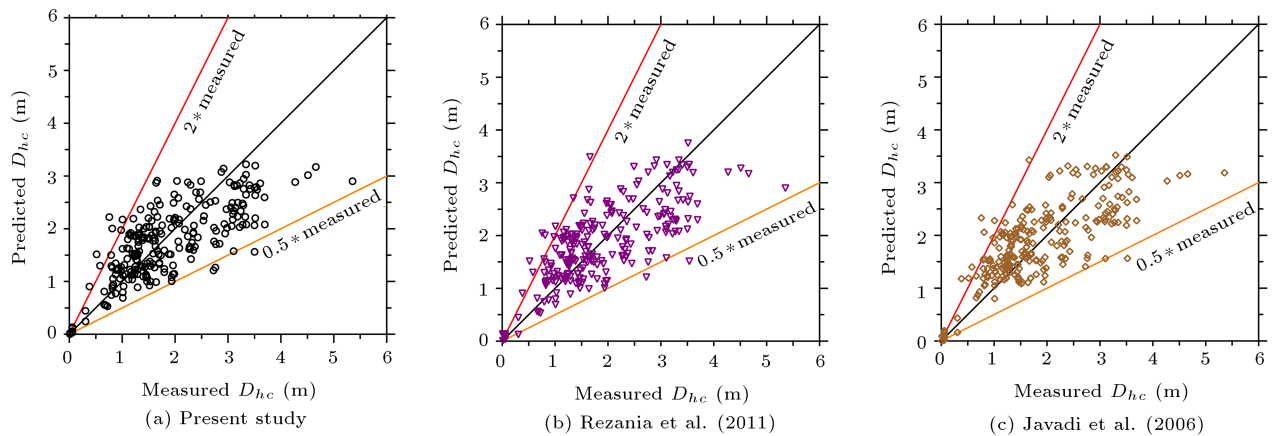


Figure 8. Comparison of the results of the model developed by Gene Expression Programming (GEP) in grounds with a gentle slope: (a) Present study, (b) model developed by Rezania et al. [37], and (c) model developed by Javadi et al. [2].

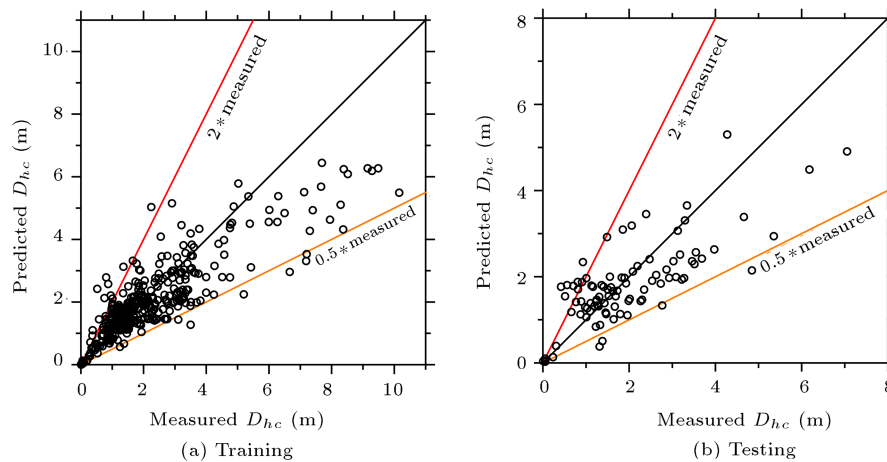


Figure 9. Comparison of the results of the model made by Gene Expression Programming (GEP) with the measured values of data in general case: (a) Training data and (b) testing data.

are shown in Figure 9. In Table 9, the equation developed for this case was compared to the equation developed by ANN approach and despite the lower value of R^2 , it can be seen that in other comparative criteria (RMSE, MAE), the equation obtained from GEP method provided much better results than ANN.

6. Discussion and review

The aim of this study is to provide a new method based on machine learning techniques to predict the extent of liquefaction-induced lateral spreading. The methodology used in this study is a subcategory of GA that searches within the solution space using the laws of evolution and Darwinian survival. In this method that is a more complete and more advanced generation of GAs than the previous ones, attempts were made to simulate the natural evolution and what would occur in natural systems, more effectively.

Thus, for datasets and the three different states

of the ground conditions, some models were developed using GEP. This method had a great ability to develop high-precision models and identify the complex relationships between variables. Unlike many of the artificial intelligence methods, it can offer the equations in a closed form. According to Figures 5, 7, and 9 and based on the comparison of the values predicted by GEP and MLR methods, this claim was investigated with the real data.

Although the equations obtained by GEP method were more complex than those obtained by the traditional methods, it is necessary to use methods that can simulate the complexity of these problems due to the complex nature of these problems. Therefore, in practical problems, GEP can analyze the problems more accurately and provide more accurate solutions.

Figure 4 shows that the data are more abundant in certain domains of geotechnical, topographical, and seismological characteristics and given the nature of GAs, the developed equations are more reliable within

these domains. Therefore, in order to obtain a general relationship that can be applicable to different ground conditions in different regions, it is necessary to compile more collections of data that can cover a wider range of specifications. By achieving this data set and using GEP, more accurate equations can be developed.

7. Conclusions

Lateral spreading is a type of ground failure caused by liquefaction due to earthquakes and unlike the liquefaction phenomenon, no clear mechanism has been characterized for it. This phenomenon includes fracturing of the surface layers on the liquefied soil and the movement of these pieces toward the downstream slope or the free face. Given that this phenomenon may occur again in grounds with free face, grounds with gentle slopes, and a combination of both ground conditions, all the three different ground states were studied in this research.

In order to study the phenomenon of lateral spreading, a dataset consisting of 484 data, compiled by Youd et al. [27], was used. In all artificial intelligence methods, the data should be categorized into two categories of training and testing data; the first category of the data was used for learning and developing equations, and the second category was used for verification of the developed equations. Therefore, in this research, for each of the different ground conditions, about 80% and 20% of the data were used as training data and testing data, respectively.

In this study, Gene Expression Programming (GEP) method was used to evaluate the liquefaction-induced lateral spreading. Being one of the methods associated with machine learning techniques, GEP is a new generation of genetic algorithms and is used as an optimization tool that searches the solution space to find the most optimal one using the laws of Darwinian evolution, survival, and reproduction.

Comparative criteria of R^2 , Root Mean Square Error (RMSE), and MAE and also the length of equations were used for choosing the best model, and all developed models were evaluated using these criteria. Eq. (4) was the best relationship obtained for the grounds with free face that would produce correlation coefficients of 90.3% and 79.9% for training and testing data sets, respectively. Moreover, in grounds with a gentle slope, the best developed equation was Eq. (5) with correlation coefficients of 88.7% and 79.02% in training and testing datasets, respectively. Moreover, the best general equation was Eq. (6) with correlation coefficients of 89.4% in training data and 77.6% in testing data. As previously mentioned, equations with higher accuracy were also made by GEP method. However, due to their level of complexity and lower length, these equations were offered as final equations.

In Table 9, the equations obtained by GEP were compared with other equations developed by other methods through statistical criteria and the results indicated the higher accuracy of developed equations than other equations. Furthermore, the comparison showed the substantial ability of GEP to diagnose relationships and building appropriate models.

Acknowledgement

The authors gratefully thank Prof. T.L. Youd for making his database available, which was used to develop the models in this research.

References

1. Bartlett, S.F. and Youd, T.L., *Empirical Analysis of Horizontal Ground Displacement Generated by Liquefaction-Induced Lateral Spreads*, US National Center for Earthquake Engineering Research (NCEER) (1992).
2. Javadi, A.A., Rezaei, M., and Nezhad, M.M. "Evaluation of liquefaction induced lateral displacements using genetic programming", *Comput. Geotech.*, **33**(4–5), pp. 222–233 (2006). <https://doi.org/10.1016/j.compgeo.2006.05.001>
3. Newmark, N.M. "Effects of earthquakes on dams and embankments", *Geotechnique*, **15**(2), pp. 139–160 (1965). <https://doi.org/10.1680/geot.1965.15.2.139>
4. Yegian, M.K., Marciano, E.A., and Ghahraman, V.G. "Earthquake-induced permanent deformations: probabilistic approach", *J. Geotech. Eng.*, **117**(1), pp. 35–50 (1991). [https://doi.org/10.1061/\(ASCE\)0733-9410\(1991\)117:1\(35\)](https://doi.org/10.1061/(ASCE)0733-9410(1991)117:1(35))
5. Baziar, M.H., Dobry, R., and Elgamal, A., *Engineering Evaluation of Permanent Ground Deformations Due to Seismically-Induced Liquefaction*, US National Center for Earthquake Engineering Research (NCEER) (1992).
6. Jibson, R.W. "Predicting earthquake-induced landslide displacements using Newmark's sliding block analysis", *Transp. Res. Rec.*, **1411**, pp. 9–17 (1993).
7. Towhata, I., Sasaki, Y., Tokida, K.I., et al. "Prediction of permanent displacement of liquefied ground by means of minimum energy principle", *Soils Found.*, **32**(3), pp. 97–116 (1992). https://doi.org/10.3208/sandf1972.32.3_97
8. Tokida, K., Matsumoto, H., Azuma, T., et al. "Simplified procedure to estimate lateral ground flow by soil liquefaction", In *Soil Dyn. Earthq. Eng. VI, WIT Trans. Built Environ.*, Brebbia ASC and CA, Editor., Elsevier, pp. 381–396 (1993).
9. Finn, W.D. "Assessment of liquefaction potential and post-liquefaction behavior of earth structures: developments 1981-1991", *Second Int. Conf. Recent Adv. Geotech. Earthq. Eng. Soil Dyn.*, University of Missouri (1991).

10. Finn, W.D., Ledbetter, R.H., and Wu, G. “Liquefaction in silty soils: design and analysis”, *Gr. Fail. under Seism. Cond.*, American Society of Civil Engineers (ASCE), pp. 51–76 (1994).
11. Gu, W.H., Morgenstern, N.R., and Robertson, P.K. “Postearthquake deformation analysis of Wildlife site”, *J. Geotech. Eng.*, **120**(2), pp. 274–289 (1994). [https://doi.org/10.1061/\(ASCE\)0733-9410\(1994\)120:2\(274\)](https://doi.org/10.1061/(ASCE)0733-9410(1994)120:2(274))
12. Yasuda, S., Nagase, H., Kiku, H., et al. “The mechanism and a simplified procedure for the analysis of permanent ground displacement due to liquefaction”, *Soils Found.*, **32**(1), pp. 149–160 (1992). <https://doi.org/10.3208/sandf1972.32.149>
13. Ghasemi-Fare, O. and Pak, A. “Numerical investigation of the effects of geometric and seismic parameters on liquefaction-induced lateral spreading”, *Soil Dyn. Earthq. Eng.*, **89**, pp. 233–247 (2016). <https://doi.org/10.1016/J.SOILDYN.2016.08.014>
14. de la Maza, G., Williams, N., Sáez, E., et al. “Liquefaction-induced lateral spread in Lo Rojas, Coronel, Chile: field study and numerical modeling”, *Earthq. Spectra*, **33**(1), pp. 219–240 (2017). <https://doi.org/10.1193/012015EQS012M>
15. Munter, S.K., Boulanger, R.W., Krage, C.P., et al. “Evaluation of liquefaction-induced lateral spreading procedures for interbedded deposits: Çark Canal in the 1999 M7.5 Kocaeli earthquake”, *Geotech. Front. 2017*, American Society of Civil Engineers, Reston, VA. pp. 254–266 (2017). <https://doi.org/10.1061/9780784480489.026>
16. Boulanger, R.W., Moug, D.M., Munter, S.K., et al. “Evaluating liquefaction and lateral spreading in interbedded sand, silt, and clay deposits using the cone penetrometer”, *Aust. Geomech. J.*, **51**(4), pp. 109–128 (2016).
17. Baziar, M.H. and Saeedi Azizkandi, A. “Evaluation of lateral spreading utilizing artificial neural network and genetic programming”, *Int. J. Civ. Eng.*, **11**(2), pp. 100–111 (2013).
18. Wang, J. and Rahman, M.S. “A neural network model for liquefaction-induced horizontal ground displacement”, *Soil Dyn. Earthq. Eng.*, **18**(8), pp. 555–568 (1999). [https://doi.org/10.1016/S0267-7261\(99\)00027-5](https://doi.org/10.1016/S0267-7261(99)00027-5)
19. Shamoto, Y., Zhang, J.-M., and Tokimatsu, K. “New charts for predicting large residual post-liquefaction ground deformation”, *Soil Dyn. Earthq. Eng.*, **17**(7), pp. 427–438 (1998). [https://doi.org/10.1016/S0267-7261\(98\)00011-6](https://doi.org/10.1016/S0267-7261(98)00011-6)
20. Zhang, G., Robertson, P.K., and Brachman, R.W.I. “Estimating liquefaction-induced lateral displacements using the standard penetration test or cone penetration test”, *J. Geotech. Geoenvironmental Eng.*, **130**(8), pp. 861–871 (2004). [https://doi.org/10.1061/\(ASCE\)1090-0241\(2004\)130:8\(861\)](https://doi.org/10.1061/(ASCE)1090-0241(2004)130:8(861))
21. Hamada, M., Yasuda, S., Isoyama, R., et al., *Study on Liquefaction Induced Permanent Ground Displacements*, Association for the Development of Earthquake Prediction Japan (1986).
22. Youd, T.L. and Perkins, D.M. “Mapping of liquefaction severity index”, *J. Geotech. Eng.*, **113**(11), pp. 1374–1392 (1987).
23. Bartlett, S.F. and Youd, T.L. “Empirical prediction of liquefaction-induced lateral spread”, *J. Geotech. Eng.*, **121**(4), pp. 316–329 (1995).
24. Bardet, J.P., Mace, N., and Tobita, T., *Liquefaction-Induced Ground Deformation and Failure*, Pacific Earthquake Engineering Research Center (PEER) (1999).
25. Bardet, J.P., Tobita, T., Mace, N., et al. “Regional modeling of liquefaction-induced ground deformation”, *Earthq. Spectra*, **18**(1), pp. 19–46 (2002). <https://doi.org/10.1193/1.1463409>
26. Rauch, A.F. and Martin, J.R. “EPOLLS model for predicting average displacements on lateral spreads”, *J. Geotech. Geoenvironmental Eng.*, **126**(4), pp. 360–371 (2000).
27. Youd, T.L., Hansen, C.M., and Bartlett, S.F. “Revised multilinear regression equations for prediction of lateral spread displacement”, *J. Geotech. Geoenvironmental Eng.*, **128**(12), pp. 1007–1017 (2002). [https://doi.org/10.1061/\(ASCE\)1090-0241\(2002\)128:12\(1007\)](https://doi.org/10.1061/(ASCE)1090-0241(2002)128:12(1007))
28. Zhang, J. and Zhao, J.X. “Empirical models for estimating liquefaction-induced lateral spread displacement”, *Soil Dyn. Earthq. Eng.*, **25**(6), pp. 439–450 (2005). <https://doi.org/10.1016/j.soildyn.2005.04.002>
29. Zhang, J., Yang, C., Zhao, J.X., et al. “Empirical models for predicting lateral spreading considering the effect of regional seismicity”, *Earthq. Eng. Eng. Vib.*, **11**(1), pp. 121–131 (2012). <https://doi.org/10.1007/s11803-012-0103-7>
30. Kalantary, F., MolaAbasi, H., Salahi, M., et al. “Prediction of liquefaction induced lateral displacements using robust optimization model”, *Sci. Iran.*, **20**(2), pp. 242–250 (2013). <https://doi.org/10.1016/j.scient.2012.12.025>
31. Goh, A.T.C. and Zhang, W.G. “An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines”, *Eng. Geol.*, **170**, pp. 1–10 (2014). <https://doi.org/10.1016/j.enggeo.2013.12.003>
32. Khoshnevisan, S., Juang, H., Zhou, Y.-G., et al. “Probabilistic assessment of liquefaction-induced lateral spreads using CPT – Focusing on the 2010–2011 Canterbury earthquake sequence”, *Eng. Geol.*, **192**, pp. 113–128 (2015). <https://doi.org/10.1016/J.ENGGEOL.2015.04.001>
33. Hasaengebi, N., Ulusay, R., and Önder Çetin, K. “A new empirical method to predict liquefaction-induced lateral spread”, *Eng. Geol. Soc. Territ.*, **5**, Springer International Publishing, Cham., pp. 1071–1075 (2015). https://doi.org/10.1007/978-3-319-09048-1_203

34. Baziar, M.H. and Ghorbani, A. “Evaluation of lateral spreading using artificial neural networks”, *Soil Dyn. Earthq. Eng.*, **25**(1), pp. 1–9 (2005). <https://doi.org/10.1016/j.soildyn.2004.09.001>
35. Kaya, Z. “Predicting liquefaction-induced lateral spreading by using neural network and neuro-fuzzy techniques”, *Int. J. Geomech.*, **16**(4), pp. 04015095 (2016). [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0000607](https://doi.org/10.1061/(ASCE)GM.1943-5622.0000607)
36. Goharriz, M. and Marandi, S.M. “An optimized neuro-fuzzy group method of data handling system based on gravitational search algorithm for evaluation of lateral ground displacements”, *Int. J. Optim. Civ. Eng.*, **6**(3), pp. 385–403 (2016).
37. Rezania, M., Faramarzi, A., and Javadi, A.A. “An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement”, *Eng. Appl. Artif. Intell.*, **24**(1), pp. 142–153 (2011). <https://doi.org/10.1016/j.engappai.2010.09.010>
38. Mola-Abasi, H. and Shooshpasha, I. “Prediction of liquefaction induced lateral displacements using polynomial neural networks and genetic algorithms”, *15th World Conf. Earthq. Eng.*, Lisbon, Portugal (2012).
39. Johari, A., Habibagahi, G., and Nakhaee, M. “Prediction of unsaturated soils effective stress parameter using gene expression programming”, *Sci. Iran.*, **20**(5), pp. 1433–1444 (2013).
40. Keshavarz, A. and Mehramiri, M. “New gene expression programming models for normalized shear modulus and damping ratio of sands”, *Eng. Appl. Artif. Intell.*, **45**, pp. 464–472 (2015). <https://doi.org/10.1016/j.engappai.2015.07.022>
41. Johari, A., Javadi, A.A., and Kajafi, H. “A genetic-based model to predict maximum lateral displacement of retaining wall in granular soil”, *Sci. Iran.*, **23**(1), pp. 54–65 (2016).
42. Johari, A. and Nejad, A.H. “Prediction of soil-water characteristic curve using gene expression programming”, *Iran. J. Sci. Technol. Trans. Civ. Eng.*, **39**(C1), pp. 143–165 (2015).
43. Ferreira, C. “Gene expression programming: A new adaptive algorithm for solving problems”, *Complex Syst.*, **13**(2), pp. 87–129 (2001).
44. Ferreira, C., *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, Stud. Comput. Intell. Springer, 2nd, Revis Ed., New York (2006). <https://doi.org/10.1007/3-540-32849-1>
45. Ferreira, C. “Mutation, transposition, and recombination: An analysis of the evolutionary dynamics”, *6th Jt. Conf. Inf. Sci. 4th Int. Work. Front. Evol. Algorithms*, North Carolina, USA, pp. 614–617 (2002).

Biographies

Amin Keshavarz is currently an Associate Professor of Civil Engineering in the School of Engineering at Persian Gulf University, Iran. He received his BSc degree in Civil Engineering from Persian Gulf University in 1997. He also received his MSc and PhD degrees in Civil Engineering (Soil Mechanics and Foundations) from Shiraz University, Iran in 2000 and 2007, respectively. His research interests are stress characteristics and ZEL methods, soil dynamics and geotechnical earthquake engineering, and stability analysis of reinforced and unreinforced soil slopes, and retaining walls.

Hojjat Tofghi was born in Kazerun, Iran in 1984. He received his BS degree in Civil Engineering from Persian Gulf University in 2009 and his MS degree in Earthquake Engineering from Islamic Azad University, Bushehr, Iran in 2017. His area of interest encompasses geotechnical engineering particularly liquefaction and lateral spreading as well as application of intelligent systems to geotechnical and other numerical issues in civil engineering.