Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(6): 9233–9252.
DOI: 10.3934/mbe.2021454
Received: 14 July 2021
Accepted: 18 October 2021
Published: 26 October 2021

Research article

Probabilistic evaluation of CPT-based seismic soil liquefaction potential: towards the integration of interpretive structural modeling and bayesian belief network

Mahmood Ahmad[1], Feezan Ahmad[2], Jiandong Huang[3,*], Muhammad Junaid Iqbal[1], Muhammad Safdar[4] and Nima Pirhadi[5]

[1] Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu 28100, Pakistan
[2] State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China
[3] School of Mines, China University of Mining and Technology, Xuzhou 221116, China
[4] Earthquake Engineering Center, University of Engineering and Technology Peshawar, Peshawar 25000, Pakistan
[5] School of Civil Engineering and Geomatics, Southwest Petroleum University, Chengdu 610513, Sichuan, China

* **Correspondence:** Email: huang@cumt.edu.cn (J.H.); Tel: +8617853959072.

Abstract: This paper proposes a probabilistic graphical model that integrates interpretive structural modeling (ISM) and Bayesian belief network (BBN) approaches to predict cone penetration test (CPT)-based soil liquefaction potential. In this study, an ISM approach was employed to identify relationships between influence factors, whereas BBN approach was used to describe the quantitative strength of their relationships using conditional and marginal probabilities. The proposed model combines major causes, such as soil, seismic and site conditions, of seismic soil liquefaction at once. To demonstrate the application of the propose framework, the paper elaborates on each phase of the BBN framework, which is then validated with historical empirical data. In context of the rate of successful prediction of liquefaction and non-liquefaction events, the proposed probabilistic graphical model is proven to be more effective, compared to logistic regression, support vector machine, random forest and naive Bayes methods. This research also interprets sensitivity analysis and the most probable explanation of seismic soil liquefaction appertaining to engineering perspective.

Keywords: bayesian belief network; cone penetration test; liquefaction potential; interpretive structural modeling; sensitivity analysis

1. Introduction

Determination of soil liquefaction potential is a fundamental step for seismic-induced hazard mitigation. In the last few decades, numerous researchers have attempted to present different methods that are based on in situ tests to predict the soil liquefaction potential, e.g., Seed and Idriss [1,2]; Robertson and Wride [3]; Youd and Idriss [4]; Juang et al. [5]; Moss et al. [6]; Idriss and Boulanger [7]; such as standard penetration test (SPT), cone penetration test (CPT), and techniques for shear wave velocity (V_s). The findings of the cone penetration test (CPT) have been adapted by many researchers from in situ tests as the basis for evaluating the liquefaction potential of the test method (e.g., Youd and Idriss [4]; Juang et al. [5]). The CPT is being used increasingly in geotechnical investigations owing to its simplicity, accuracy, continuous soil profile and repeatability than other types of in-situ test methods.

Artificial intelligence (AI) techniques as for example random forest [8], adaptive neuro-fuzzy inference system (ANFIS) [9], relevance vector machine (RVM) [10], artificial neural network (ANN) [5,11,12], genetic programming (GP) [13–16] and support vector machine (SVM) [12,17–19] models were developed to predict liquefaction potential based on in situ test database. Over conventional modeling techniques, the primary strength of AI techniques is their process of capturing nonlinear and complex correlation between system variables without having to presume the correlations between different variables of input and output. In the scope of assessing the occurrence of liquefaction, these techniques may be trained to learn the relationship between soil, site, and earthquake characteristics with the potential for liquefaction, needing no prior knowledge of the form of the relation. Mostly models are black box owing to the fact that the relationship between the system's inputs and output parameters is represented in terms of a weight matrix and biases that are not visible to the user.

The Bayesian Belief Network (BBN) is a graphical model that enables a set of variables to be probabilistically connected [20]. To address cause-effect relationships and complexities, BBN may provide an effective structure. BBN not only provides sequential inference (from causes to results) but also reverse inference (from results to causes). The benefits of BBNs include the following compared to other methods: 1) BBN achieves a combination of qualitative and quantitative analysis; 2) BBN allows reversal inference (from results to causes) and it is simple to obtain the ranking of factors affecting the casualties; 3) BBN has a good learning ability; 4) allows data to be combined with domain knowledge; and 5) Even with very limited sample sizes, BBN can demonstrate good prediction accuracy. Furthermore, its application in seismic liquefaction potential on CPT-based in-situ tests data is found comparatively less e.g., Ahmad et al. [21–23].

The contributions of this paper are fourfold:

a) This study integrates ISM and BBN to assess CPT-based seismic soil liquefaction potential that uses conditional and marginal probabilities to describe the quantitative strength of their relationships;

b) The performance of the proposed model is comparatively assessed with four traditional seismic soil liquefaction modeling algorithms (Logistic regression, SVM, RF, and Naive Bayes);

c) The sensitivity analysis of predictor variables is presented owing to know the effect of input factors on the liquefaction potential; and

d) The most probable explanation (MPE) of seismic soil liquefaction with reference to engineering perspective is presented.

This article consists of six major sections. Next section presents research methodology, Section 3 is devoted to the probabilistic graphical model development. Section 4 presents comparison and evaluation measures with the widely used prediction methods. Results and discussion are presented in Section 5. Finally, in the last part, conclusions and future work are set out.

2. Research methodology

This study's working approach was divided into two parts (Figure 1): interpretive structural modelling and Bayesian belief networks.

2.1. Interpretive structural modeling

Interpretive structural modeling (ISM) is a well-established technique that describes a situation or a problem to classify the relationships between particular issues. A collection of different elements that are directly and indirectly connected are organized into a structured comprehensive model in this approach [24,25]. The model thus created depicts the structure of a complex problem or issue in a carefully constructed pattern that implies graphics and words [26–28]. Different researchers have increasingly used this technique to depict the interrelationships between various elements relevant to the issues. The ISM approach includes the identification of variables that are important to the issue or problem. Then a contextually relevant subordinate relationship is identified. On the basis of a pair wise comparison of variables, after the contextual relationship has been determined, a structural self-interaction matrix (SSIM) is defined. After this, SSIM is converted into a reachability matrix (RM) and its transitivity is examined. A matrix model is obtained after transitivity embedding is complete. Then, the element partitioning and a structural model extraction called ISM are derived. The development of the ISM model is explained in further depth by Ahmad et al. [29,30].

2.2. Bayesian belief network

Bayesian Belief Networks (BBN) is a graphical network of causal connections between different nodes. In BBN models, the network structure is a directed acyclic graph (DAG) that graphically represents the logical relationship between nodes, and the conditional probability of quantifying the strength of this relationship is the network parameter [31–33]. The network structure and network parameter can be obtained via expert knowledge [34,35] or training from data [36].

The primary idea of a BBN is based on Bayes' theorem, which states that the relationship between two nodes, hypothesis H (parent) and evidence E (child), is represented as:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \tag{1}$$

where $P(H|E)$ is one's belief in hypothesis H after observing evidence E, $P(E|H)$ is the chance that E will be observed if H is true, $P(H)$ is the probability that the hypothesis will hold true, and $P(E)$ is the probability that the evidence will occur. The posterior probability is $P(H|E)$, and the prior probability is $P(H)$. In a BBN analysis, the updated probability is derived by expanding the $P(E)$ in Eq (1) for n number of mutually exclusive hypotheses H_i (i=1,..., n) and a given evidence E [37] as:

$$P\left(H_j|E\right)=\frac{P\left(E|H_j\right)\times P\left(H_j\right)}{\sum_{i=1}^{n}P\left(E|H_i\right)\times P\left(H_i\right)} \tag{2}$$

A BBN is used to update probability as new information becomes available. The network allows researchers to compute the probabilities of any subset of variables given evidence about another subset.

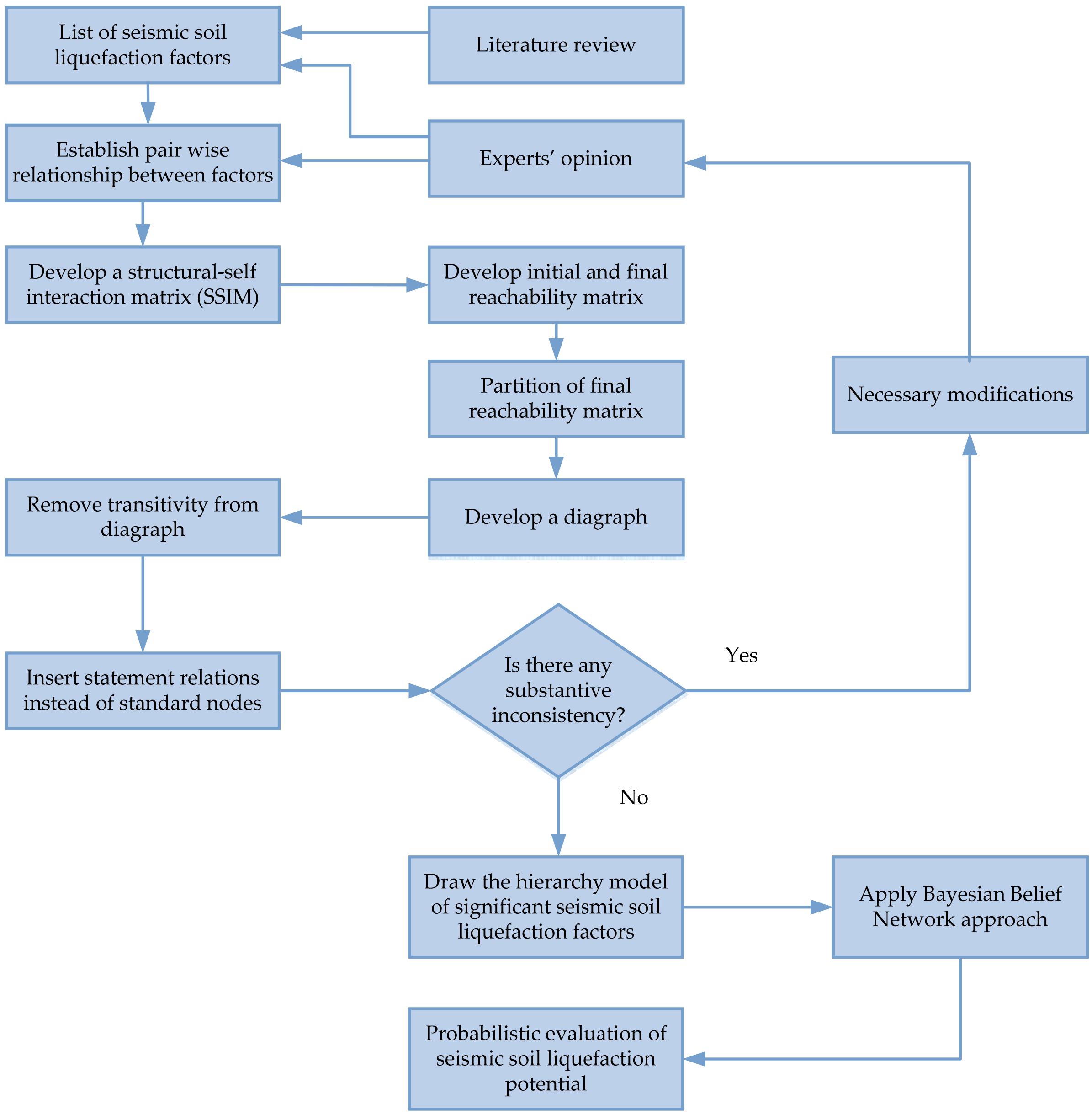

Figure 1. The process outline of the methodology of research used in the present study.

3. Development of probabilistic graphical model

3.1. Data set and predictor variables

The data set used in this analysis was on the basis of the revised version of the CPT case history records collected by Boulanger and Idriss [38]. The entire data set consists of 253 cases with a soil behaviour type index, $I_c < 2.6$, of which 180 are liquefied cases, another 71 are non-liquefied cases and the remaining 2 are doubtful cases (marginal between liquefied and non-liquefied) in this research work. These case histories are derived from CPT measurements of 17 sites and field performance reports of major earthquakes (the complete database is available in Table S1). Liquefaction is caused by seismic parameters, site conditions, and soil properties that include a varied range of factors. So nine critical factors or variables used for the possible evaluation of liquefaction are chosen, namely earthquake magnitude (M) V_1, peak ground acceleration (a_{max}, g) V_2, fines content (FC, %) V_3, equivalent clean sand penetration resistance (q_{c1Ncs}) V_4, soil behaviour type index (I_c) V_5, vertical effective stress (σ'_v, kPa) V_6, groundwater table depth (D_w, m) V_7, depth of soil deposit (D_s, m) V_8, thickness of soil layer (T_s, m) V_9, and output is liquefaction potential, V_{10} in this paper according to Okoli and Schabram [39] and Tranfield et al. [40]. For more details of CPT case histories, viewers may refer to the Boulanger and Idriss reference [38]. The statistical characteristics of the data set used in this study, such as minimum (Min.), maximum (Max.), mean, standard deviation (SD) and coefficient of variation (COV), are shown in Table 1.

Table 1. Statistical aspects of the dataset.

Statistical parameters	M	a_{max} (g)	FC (%)	q_{c1Ncs}	I_c	σ'_v (kPa)	D_w (m)	D_s (m)	T_s (m)
Min.	5.9	0.09	0	16.1	1.16	19	0.2	1.4	0.3
Max.	9	0.84	85	311.9	2.59	147	7.2	11.8	6.5
Mean	6.98	0.32	17.71	93.89	1.96	57.62	2.04	4.44	1.83
SD	0.55	0.15	19.27	38.06	0.29	24.55	1.21	1.96	1.22
COV	0.08	0.46	1.09	0.41	0.15	0.43	0.59	0.44	0.67

Previous studies [23,41–43] showed detail understanding about the variables' selection and discretization. BBN has a good capability to deal with discrete variables, but is weak in continuous variables processing, so the nine significant factors and output (liquefaction potential, V_{10}) require to be transformed into discrete values before the propose model is constructed accordance to the possible factor range and expert knowledge, as shown in Table 2.

3.2. Probabilistic graphical model for CPT-based seismic soil liquefaction potential

The data set has been divided into training and testing datasets according to statistical aspects for example mean, maximum, minimum, etc. to build the models:

- A training data set is required to build the models. The authors used 201 (80%) CPT case history data for the training set in this study.

- A testing data set is required to predict the performance of the established models. The remaining 50 (20%) CPT case history data is used as a testing data set in this study.

ISM technique suggests the use of domain or expert knowledge in the creation of the contextual relationships between the nine significant variables and contextual relationships are ultimately analyzed by field experts who have approved and represented by SSIM (see Table 3).

Table 2. Grading specifications for seismic soil liquefaction variables.

Category	Variable	Number of grade	Explanation	Range
Seismic parameter	Earthquake magnitude, M	4	Super	$8 \leq M$
			Big	$7 \leq M < 8$
			Strong	$6 \leq M < 7$
			Medium	$4.5 \leq M < 6$
	Peak ground acceleration, a_{max} (g)	4	Super	$0.40 \leq a_{max}$
			High	$0.30 \leq a_{max} < 0.40$
			Medium	$0.15 \leq a_{max} < 0.30$
			Low	$0 \leq a_{max} < 0.15$
Soil parameter	Fines content, FC (%)	3	Many	$50 < FC$
			Medium	$30 < FC \leq 50$
			Less	$0 \leq FC \leq 30$
	Equivalent clean sand penetration resistance, q_{c1Ncs}	4	Super	$135 \leq q_{c1Ncs}$
			Big	$90 \leq q_{c1Ncs} < 135$
			Medium	$45 \leq q_{c1Ncs} < 90$
			Small	$0 \leq q_{c1Ncs} < 45$
	Soil behaviour type index, I_c	4	Gravelly sand to dense sand	$I_c < 1.31$
				$1.31 \leq I_c < 1.61$
			Clean sand	$1.61 \leq I_c < 2.40$
			Silty sand or Sand with silt	$2.40 \leq I_c < 2.60$
			Sandy silt	
Site condition	Vertical effective stress, σ'_v (kPa)	4	Super	$150 \leq \sigma'_v$
			Big	$100 \leq \sigma'_v < 150$
			Medium	$50 \leq \sigma'_v < 100$
			Small	$0 \leq \sigma'_v < 50$
	Groundwater table depth, D_w (m)	3	Deep	$4 \leq D_w$
			Medium	$2 < D_w < 4$
			Shallow	$D_w \leq 2$
	Depth of soil deposit, D_s (m)	3	Deep	$10 \leq D_s < 20$
			Medium	$5 \leq D_s < 10$
			Shallow	$0 \leq D_s < 5$
	Thickness of soil layer, T_s (m)	3	Thick	$10 \leq T_s$
			Medium	$5 \leq T_s < 10$
			Thin	$0 < T_s < 5$

Table 3. SSIM for seismic soil liquefaction variables.

V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_i
	V	O	O	O	O	O	O	O	V	V_1
		O	O	A	O	O	O	O	V	V_2
			V	V	O	O	O	O	V	V_3
				A	A	A	A	O	V	V_4
					A	O	O	O	V	V_5
						A	A	O	V	V_6
							O	O	V	V_7
								O	V	V_8
									V	V_9
										V_{10}

In the next phase, the SSIM is changed to a binary matrix for seismic soil liquefaction factors, called the initial reachability matrix (IRM), by exchanging the original symbols with 1 or 0, as shown in Table A1 in Appendix A. When the IRM is obtained, the transitivity property is verified to get the final matrix of reachability (FRM). The transitivity check is the basic principle of the ISM methodology that if the '*a*' variable is related to the '*b*' variable and the '*b*' variable is related to the '*c*' variable, the '*a*' variable is ultimately correlated with the '*c*' variable. The new entries that are labeled as '1*' are implied after transitivity checking. The FRM with rank, driving and dependence powers is shown in Appendix A of Table A2. The variables used to derive multilevel hierarchy structure levels, along with their reachability set (S_r), antecedent set (S_a), and intersection set (S_i), are shown in Table A3–A7 in Appendix A. The findings showed that there are five partition levels which are as follows:

$$L_1 = \{V_{10}\};\ L_2 = \{V_2, V_4, V_9\};\ L_3 = \{V_1, V_5\};\ L_4 = \{V_3, V_6\};\ L_5 = \{V_7, V_8\}.$$

The liquefaction potential multilevel hierarchy structure is formed from the FRM. The transitivity relations between two variables, such as the direct links between the D_s and the D_w with liquefaction potential, are eliminated because the D_s and the D_w will influence the liquefaction potential through vertical effective stress. In the next phase, there is no conceptual inconsistency in the structural model so the ISM is developed for the soil liquefaction potential (see Figure 2). There is a restriction of no links between skipping-level nodes in the ISM model (for example, FC and liquefaction potential).

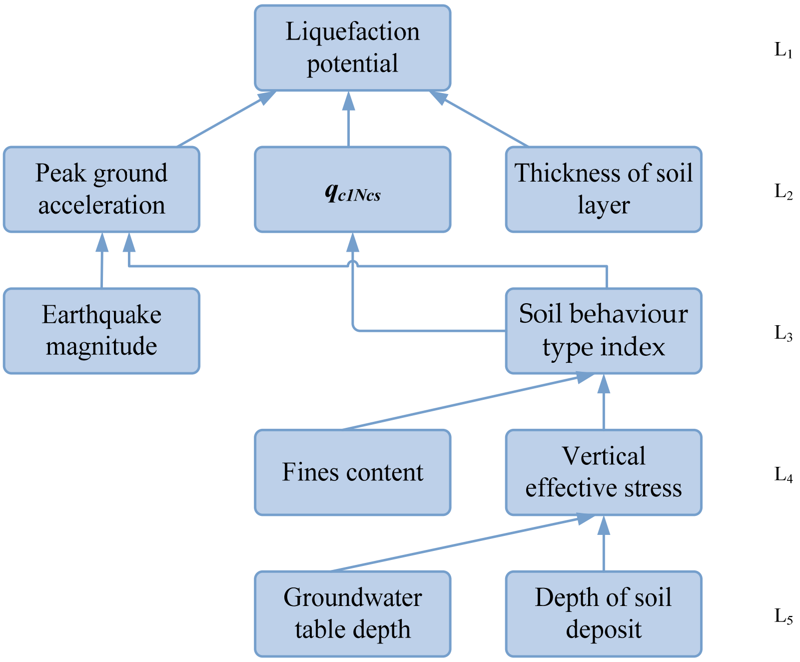

Figure 2. Interpretive structural modeling of liquefaction potential.

It can be seen that peak ground acceleration (PGA), q_{c1Ncs} and soil layer thickness (T_s) in the second level are directly influencing factors of liquefaction potential, while D_s and the D_w in the last level are the most vibrant factors that form the basis of the ISM hierarchy. In other levels, fines content, vertical effective stress, soil behaviour type index and earthquake magnitude are the indirect factors influencing the liquefaction potential.

A network model with an unknown structure or insufficient knowledge can be hard to create directly. To fix this issue, Liao et al. [37] used ISM to develop a network diagram, which they specifically used as a BBN for evaluating outsourcing risk. This approach effectively processes the relationships between variables by splitting the problem into different levels, making the overall structure clear and easy to understand and ensuring a deeper understanding of the problem. In order to facilitate constructing a BBN diagram, the final network diagram obtained from ISM defines the interdependent relationships between factors at the same level or between two levels. The model system is built directly into Netica software distributed by Norsys Software Corp to define the quantitative intensity of their relationships. The graphical presentation is shown in Figure 3.

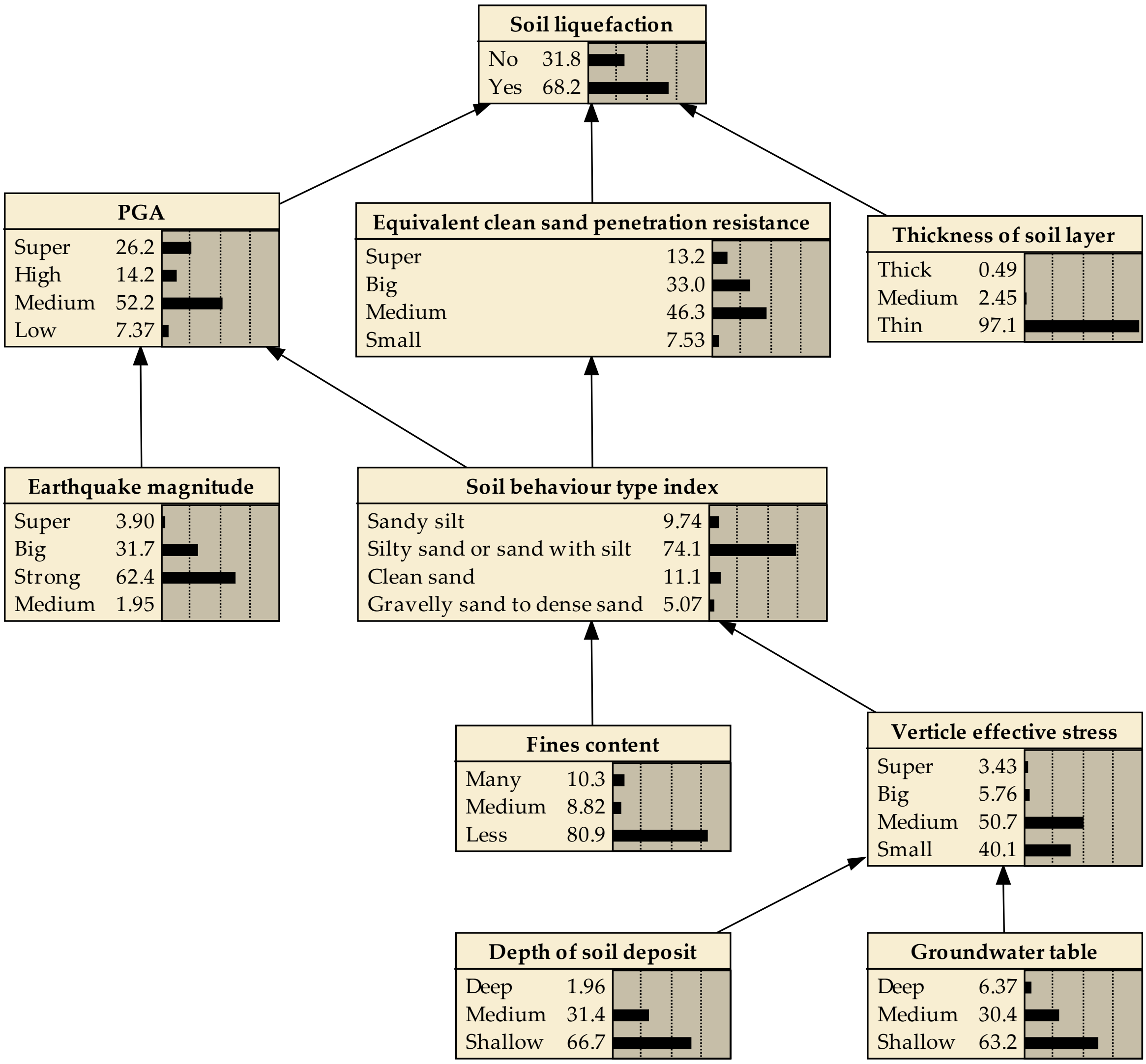

Figure 3. Soil liquefaction potential graphical model.

4. Evaluation and prediction

To assess the proposed model, it was compared to various other well-known methods using scalar performance measurements.

4.1. Compared methods

The proposed model was compared with other widely used prediction methods such as logistic regression, support vector machine, random forest, and naive Bayes. Table 4 provides a brief overview of these methods along with tuning parameters. For more information, readers can consult the corresponding reference materials.

Table 4. Prediction methods compared.

Method	Description	Tuning parameters
Logistic Regression (LR)	LR is a probability evaluation process focused on the calculation of maximum probability [44].	$\{C_1, L\}$
Support Vector Machine (SVM)	SVM, based on mathematical learning models is one of the most robust prediction methods [45]. SVM training method computes a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier, given a set of training examples, each marked as belonging to one of two categories.	$\{C_2, \varepsilon\}$
Random Forest (RF)	RF [46] is a meta-learning scheme that integrates many independently developed base classifiers and participates in a voting process to obtain a prediction for the final class.	$\{n_{tree}, n_s\}$
Naive Bayes (NB)	NB [47] assumes that the predictive variables, provided the target/dependent variable, are conditionally independent.	None

Note: C_1 = cost strength; L = regularization type (either L_1 or L_2); n_{tree} = number of trees; n_s = split subsets; C_2 = Cost; ε = regression loss epsilon

4.2. Evaluation measures

Several measure indexes are used in order to comprehensively evaluate the performances of the developed models for seismic liquefaction. There are four possible outcomes for a single prediction in the binary class scenario, i.e., liquefaction and non-liquefaction. The correct classification is true negative (TN) and true positive (TP). If the output is incorrectly predicted as negative, a false positive (FP) occurs, If the result is wrongly labelled as negative, a false negative (FN) occurs. The confusion matrix (see Figure 4) can be used to evaluate these as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

$$\text{Precision}^{+} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Precision}^{-} = \frac{TN}{FN + TN} \tag{5}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{7}$$

$$\text{F-score}^{+} = \frac{2 \times \text{Precision}^{+} \times \text{Sensitivity}}{\text{Precision}^{+} + \text{Sensitivity}} \tag{8}$$

$$\text{F-score}^{-} = \frac{2 \times \text{Precision}^{-} \times \text{Specificity}}{\text{Precision}^{-} + \text{Specificity}} \tag{9}$$

The total number of correct predictions is calculated as accuracy (*Acc*). Precision evaluates the accuracy of a single class's predictions, whereas sensitivity is concerned with true positives and false negatives, and specificity deals with false positives and true negatives. F-measure combines precision and sensitivity or specificity values to achieve a harmonic mean. The best F-score is 1, while the worst F-score is 0. A good classifier model has close sensitivity and specificity values, which can be challenging to achieve with AI algorithms [48]. Additionally, another performance metric that is computationally efficient can be used, namely G_{mean} [49]. G_{mean} is the geometric mean of each class instance's individual accuracies, and it's typically employed when each class's performance is significant and expected to be high at the same time [50,51]. In case of seismic soil liquefaction, liquefaction cases are often more than non-liquefaction instances as a result, when a data set contains a class imbalance, the *Acc* alone can be misleading. Therefore, in addition to F-score, G_{mean} in terms of error rate has been utilised in various studies to assess the performance of the classifier model [48], and is defined as:

$$G_{(mean)error} = 1 - G_{mean} \tag{10}$$

A single performance metric G_{mean} can be defined as:

$$G_{mean} = \sqrt{\text{Sensitivity} \times \text{Specificty}} \tag{11}$$

The performance metric $G_{mean(error)}$ used for binary classification on liquefaction susceptibility of soil ranges from 0 to 1, 0 indicating a completely correct classifier model and 1 suggesting a classifier model with no predictive power.

Actual	Predicted	
	Liquefied (+)	Non-liquefied (-)
Liquefied (+)	TP	FN
Non-liquefied (-)	FP	TN

Figure 4. Typical confusion matrix for 2 × 2 classification problem.

5. Evaluation and prediction

5.1. Comparative performance of multiple learners based on test dataset

The prediction results of the proposed models, i.e., BBN-ISM, LR, SVM, RF and NB were obtained on the test set. Subsequently, as shown in Table 5, each model's confusion matrix was calculated. The values on the main diagonal indicated the correctly predicted number of samples. The performance metrics were determined on the basis of Eqs (3)–(10) mentioned in Table 6, based on Figure 4. The results in Table 6 show that the developed model gave the best predictive performance, with much higher *Acc* than other models (from 4 to 16 % improvement over other models). The performance of the RF and LR models is at par and just found secondary to the proposed model. In addition, the accuracy degrees of BBN-ISM were found highest and up to 78%, followed by 74% accuracy of the RF and LR models. As indicated in Table 6, the values of $G_{mean(error)}$ of all the models (RF and L = 36.7%; SVM = 37.8%; and NB = 46.6%) are less than the BBN-ISM (26.8%). The obtained results for the specificity of the proposed BBN-ISM model (0.643) are better than the RF, LR, SVM, and NB models. Similarly, F-score for liquefied cases and non-liquefied cases for the BBN-ISM model is better than the rest of models. Comparing their values of performance measures, BBN-ISM model performed better than RF, LR, SVM, and NB models. From these comparisons, we can state that although the ML models have good accuracy prediction for soil liquefaction potential, however, their predictive capability is different from case to case, which depends on the data used for each model. In addition, the performance of ML models is determined by the tuning parameters used to train them. In this work, the tuning parameters of the applied models are selected by the trial-error process as presented in Table 7.

Table 5. Confusion matrices results based on test data of seismic soil liquefaction.

	Model									
Actual	BBN-ISM		RF		LR		SVM		NB	
	Predicted									
	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
No	**9**	5	**7**	8	**7**	8	**7**	8	**6**	9
Yes	6	**30**	5	**30**	5	**30**	6	**29**	10	**25**

Note: The diagonal elements (correct decisions) are marked in bold.

Table 6. Performance evaluation of testing dataset.

Model	*Acc* (%)	$G_{mean(error)}$ (%)	Specificity	Precision^{+}	F-Score^{+}	Sensitivity	Precision^{-}	F-score^{-}
BBN-ISM	78	26.8	0.643	0.857	0.845	0.833	0.600	0.621
RF	74	36.7	0.467	0.789	0.822	0.857	0.583	0.519
LR	74	36.7	0.467	0.789	0.822	0.857	0.583	0.519
SVM	72	37.8	0.467	0.784	0.806	0.829	0.538	0.500
NB	62	46.6	0.400	0.735	0.725	0.714	0.375	0.387

Table 7. Tuning parameters of each model for optimal values.

Model	Hyperparameters	Optimal values
RF	$\{n_{tree}, n_s\}$	{5,5}
LR	$\{C_1, L\}$	$\{1, L_2\}$
SVM	$\{C_2, \varepsilon\}$	{1.00, 0.10}
NB	None	None

5.2. Most probable explanation

The most probable explanation (MPE) is drawn from the established model to decide which situation is most probable to cause soil liquefaction potential. For instance, if the soil liquefaction is "yes" as shown in Figure 5, the set that is most probable to cause "soil liquefaction" which is [peak ground acceleration = medium, equivalent clean sand penetration resistance = medium, thickness of soil layer = thin, earthquake magnitude = strong, soil behaviour type index = silty sand or sand with silt, fines content = less, vertical effective stress = small, groundwater table = shallow, and depth of soil deposit = shallow]. This shows explicitly that the set is indeed well associated with the judgment of engineering.

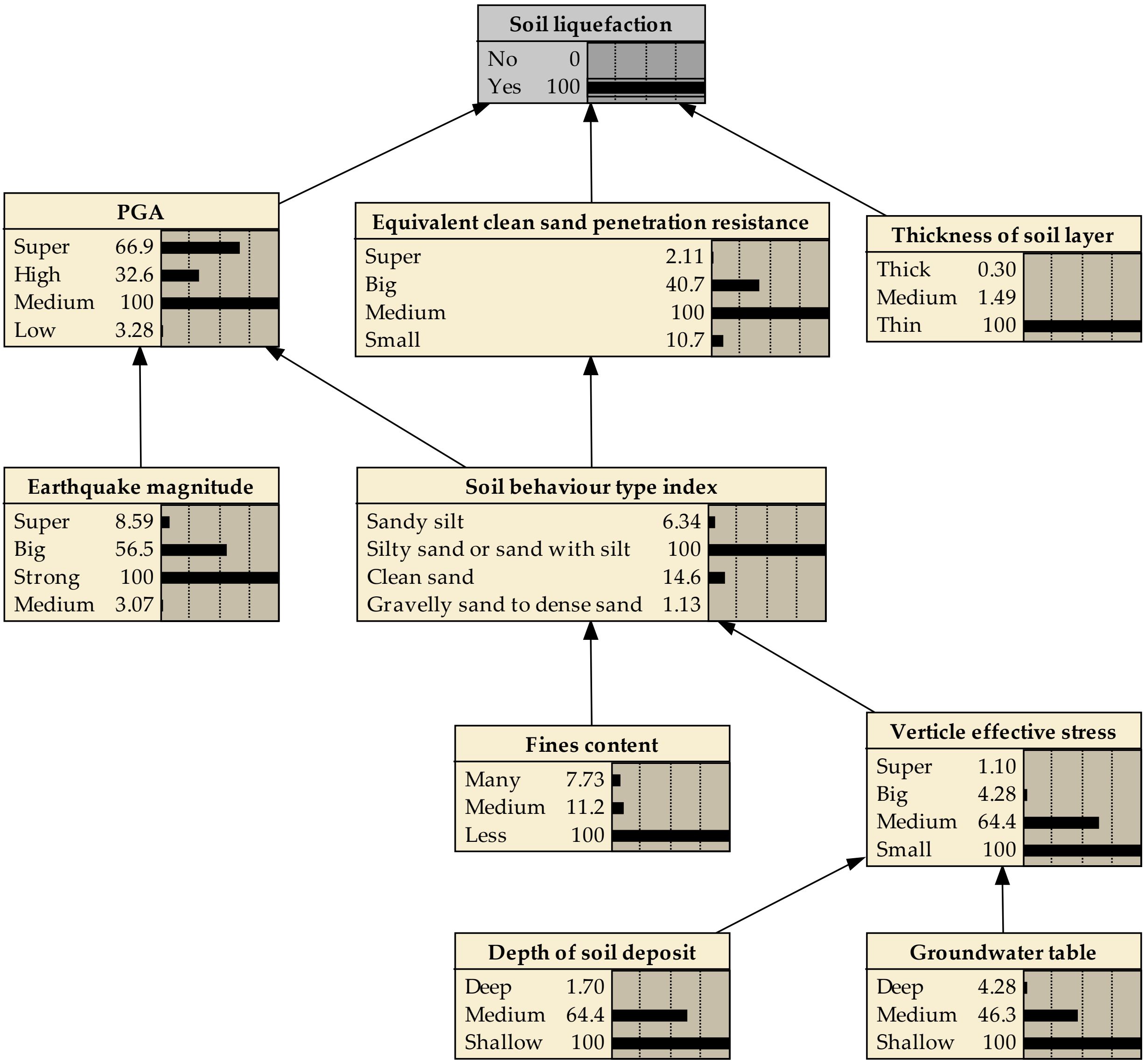

Figure 5. Most probable explanation of seismic soil liquefaction potential when the evidence state is "Yes".

5.3. Sensitivity analysis

In this study, to determine the impact of each factor on the liquefaction potential, a sensitivity analysis was performed on nine input factors with variance of beliefs. Based on the sensitivity analysis, a basic event that has a relatively large contribution to the probability of a resulting event makes it easier to reduce the probability of these basic events by considering effective measurements, thereby reducing the probability of a resulting event. The target node "soil liquefaction" is selected for sensitivity analysis, and the results are shown in Table 8. Table 8 presents that the mutual info of the "equivalent clean sand penetration resistance" node is the greatest, i.e., 0.13920, which indicates that it has the strongest influence on "soil liquefaction," potential followed by "peak ground acceleration," "soil behaviour type index," and so on which have mutual info equal to 0.04439 and 0.03655 respectively, whereas the "depth of soil deposit" is bared minimum sensitive factor with a mutual info equal to 0.00004; those findings are strongly consistent with the literature.

Table 8. Sensitivity analysis of "soil liquefaction" node.

Node	q_{c1Ncs}	a_{max} **(g)**	I_c	T_s **(m)**	σ'_v **(kPa)**	FC **(%)**	M	D_w **(m)**	D_s **(m)**
Mutual info	0.13920	0.04439	0.03655	0.00334	0.00135	0.00021	0.00019	0.00009	0.00004
Percent	15.4000	4.9200	4.0500	0.3700	0.1490	0.0234	0.0212	0.0101	0.0047
Variance of beliefs	0.0423618	0.0133434	0.0117117	0.0010781	0.0004219	0.0000640	0.0000581	0.0000275	0.0000128

6. Conclusions and future prospect

In this paper, probabilistic evaluation of CPT-based seismic soil liquefaction was carried by systematically integrating ISM and the BBN. The models were trained and tested based on Boulanger and Idriss database compiles from various soil liquefaction in different countries. The proposed model predicts the seismic soil liquefaction using major contributing factors on soil liquefaction. The most important conclusions of the present research work are as follows:

1) The accuracy of the proposed model on testing dataset is 78% and the F-score is 0.845 for liquefaction data and 0.621 for non-liquefaction data. The proposed model has better prediction ability than the RF, LR, SVM, and NB models, and its implementation is simpler due to a simple graphical result.

2) The MPE of seismic soil liquefaction is that the peak ground acceleration = medium, equivalent clean sand penetration resistance = medium, thickness of soil layer = thin, earthquake magnitude = strong, soil behaviour type index = silty sand or sand with silt, fines content = less, vertical effective stress = small, groundwater table = shallow, and depth of soil deposit = shallow, which suits well in accordance with engineering practice.

3) Sensitivity analysis results revealed that q_{c1Ncs} and PGA are the strongest influencing parameters, followed by I_c, T_s, σ'_v, FC, M, D_w, and D_s that affecting soil liquefaction.

Since the CPT case histories database have class imbalanced and the sampling biased in training and testing data set may lead anecdotal results to some degree. Nevertheless, these anecdotal findings regarding seismic soil liquefaction potential evaluation are greatly insightful from a preliminary viewpoint. In addition, owing to the ISM shortcomings, such as ignoring relationships between the nodes of the skipping-level, there is no feedback circuit between any two levels, and additionally some significant node relationships are ignored. Therefore, in the future, the causal mapping approach should be employed to change the structure and to refine the prediction performance results, taking into account the ISM shortcomings.

Acknowledgments

The authors would like to thank the experts who participated in the modeling process.

Author Contributions

Conceptualization, Mahmood Ahmad and Feezan Ahmad; Methodology, Mahmood Ahmad, Nima Pirhadi and Feezan Ahmad; Software, Mahmood Ahmad and Feezan Ahmad; Validation, Mahmood Ahmad and Nima Pirhadi; Formal analysis, Mahmood Ahmad, Feezan Ahmad and Muhammad Safdar; Investigation, Mahmood Ahmad, Feezan Ahmad and Muhammad Junaid Iqbal; Data curation, Muhammad Junaid Iqbal; Writing—original draft preparation, Mahmood Ahmad; Writing—review and editing, Jiandong Huang; Article processing charge (APC), Jiandong Huang. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. H. B. Seed, I. M. Idriss, Simplified procedure for evaluating soil liquefaction potential, *J. Soil Mech. Found. Div.*, **97** (1971), 1249–1273.
2. H. B. Seed, I. M. Idriss, Evaluation of liquefaction potential sand deposits based on observation of performance in previous earthquakes, in *Proceedings of ASCE national convention (MO)*, 481–544.
3. P. K. Robertson, C. Wride, Evaluating cyclic liquefaction potential using the cone penetration test, *Can. Geotech. J.*, **35** (1998), 442–459.
4. T. L. Youd, I. M. Idriss, Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils, *J. Geotech. Geoenviron.*, **127** (2001), 297–313.
5. C. H. Juang, H. Yuan, D. H. Lee, P. S. Lin, Simplified cone penetration test-based method for evaluating liquefaction resistance of soils, *J. Geotech. Geoenviron.*, **129** (2003), 66–80.
6. R. Moss, R. B. Seed, R. E. Kayen, J. P. Stewart, A. Der Kiureghian, K. O. Cetin, CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential, *J. Geotech. Geoenviron.*, **132** (2006), 1032–1051.
7. I. Idriss, R. Boulanger, Semi-empirical procedures for evaluating liquefaction potential during earthquakes, *Soil Dyn. Earthquake Eng.*, **26** (2006), 115–130.

8. V. Kohestani, M. Hassanlourad, A. Ardakani, Evaluation of liquefaction potential based on CPT data using random forest, *Nat. Hazards*, **79** (2015), 1079–1089.
9. X. Xue, X. Yang, Application of the adaptive neuro-fuzzy inference system for prediction of soil liquefaction, *Nat. Hazards*, **67** (2013), 901–917.
10. P. Samui, Seismic liquefaction potential assessment by using relevance vector machine, *Earthquake Eng. Eng. Vib.*, **6** (2007), 331–336.
11. A. T. Goh, Neural-network modeling of CPT seismic liquefaction data, *J. Geotech. Eng.*, **122** (1996), 70–73.
12. P. Samui, T. Sitharam, M. Contadakis, Machine learning modelling for predicting soil liquefaction susceptibility, *Nat. Hazards Earth Syst. Sci.*, **11** (2011), 1–9.
13. A. H. Gandomi, A. H. Alavi, A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems, *Neural Comput. Appl.*, **21** (2012), 189–201.
14. P. K. Muduli, S. K. Das, CPT-based seismic liquefaction potential evaluation using multi-gene genetic programming approach, *Indian Geotech. J.*, **44** (2014), 86–93.
15. P. K. Muduli, S. K. Das, S. Bhattacharya, CPT-based probabilistic evaluation of seismic soil liquefaction potential using multi-gene genetic programming, *Georisk: Assess. Manage. Risk Eng. Syst. Geohazards*, **8** (2014), 14–28.
16. P. K. Muduli, S. K. Das, First-order reliability method for probabilistic evaluation of liquefaction potential of soil using genetic programming, *Int. J. Geomech.*, **15** (2015), 04014052.
17. A. T. Goh, S. Goh, Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data, *Comput. Geotech.*, **34** (2007), 410–421.
18. T. Oommen, L. G. Baise, R. Vogel, Validation and application of empirical liquefaction models, *J. Geotech. Geoenviron.*, **136** (2010), 1618–1633.
19. M. Pal, Support vector machines-based modelling of seismic liquefaction potential, *Int. J. Numer. Anal. Methods Geomech.*, **30** (2006), 983–996.
20. J. Pearl, *Probabilistic reasoning in intelligent systems: Representation & reasoning*, Morgan Kaufmann Publishers, San Mateo, 1988.
21. M. Ahmad, X. W. Tang, J. N. Qiu, W. J. Gu, F. Ahmad, A hybrid approach for evaluating CPT-based seismic soil liquefaction potential using Bayesian belief networks, *J. Cent. South Univ.*, **27** (2020), 500–516.
22. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, W. J. Gu, A step forward towards a comprehensive framework for assessing liquefaction land damage vulnerability: Exploration from historical data, *Front. Struct. Civ. Eng.*, **14** (2020), 1476–1491.
23. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, Evaluating Seismic Soil Liquefaction Potential Using Bayesian Belief Network and C4. 5 Decision Tree Approaches, *Appl. Sci.*, **9** (2019), 4226.
24. A. P. Sage, *Methodology for Large-Scale Systems*, New York, 1977.
25. J. N. Warfield, Developing interconnection matrices in structural modeling, *IEEE Trans. Syst. Man Cybern.*, **4** (1974), 81–87.
26. V. Ravi, R. Shankar, Analysis of interactions among the barriers of reverse logistics, *Technol. Forecast. Soc. Change.*, **72** (2005), 1011–1029.
27. R. Shankar, R. Narain, A. Agarwal, An interpretive structural modeling of knowledge management in engineering industries, *J. Adv. Manag. Res.*, **1** (2003), 28–40.

28. T. Raj, R. Attri, Identification and modelling of barriers in the implementation of TQM, *Int. J. Product. Qual.*, **8** (2011), 153–179.
29. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, Evaluation of liquefaction-induced lateral displacement using Bayesian belief networks, *Front. Struct. Civ. Eng.*, **15 (**2021), 80–98,
30. M. Ahmad, X. W. Tang, F. Ahmad, M. Hadzima-Nyarko, A. Nawaz, A. Farooq, Elucidation of Seismic Soil Liquefaction Significant Factors. in *Earthquakes-From Tectonics to Buildings*, 2021.
31. K. Masmoudi, L. Abid, A. Masmoudi, Credit risk modeling using Bayesian network with a latent variable, *Expert Syst. Appl.*,**127** (2019), 157–166.
32. A. Castelletti, R. Soncini-Sessa, Bayesian Networks and participatory modelling in water resource management, *Environ. Model. Software*, **22** (2007), 1075–1088.
33. A. Ghribi, A. Masmoudi, A compound poisson model for learning discrete Bayesian networks, *Acta Math. Sci.*, **33** (2013), 1767–1784.
34. S. A. Joseph, B. J. Adams, B. McCabe, Methodology for Bayesian belief network development to facilitate compliance with water quality regulations, *J. Infrastruct. Syst.*, **16** (2010), 58–65.
35. S. Nadkarni, P. P. Shenoy, A Bayesian network approach to making inferences in causal maps, *Eur. J. Oper. Res.*, **128** (2001), 479–498.
36. G. Kabir, S. Tesfamariam, A. Francisque, R. Sadiq, Evaluating risk of water mains failure using a Bayesian belief network model, *Eur. J. Oper. Res.*, **240** (2015), 220–234.
37. S. Tesfamariam, Z. Liu, Seismic risk analysis using Bayesian belief networks, in *Handbook of seismic risk analysis and management of civil infrastructure systems*, 2013, 175–208.
38. R. Boulanger, I. Idriss, *CPT and SPT based liquefaction triggering procedures*, 2014.
39. C. Okoli, K. Schabram, A guide to conducting a systematic literature review of information systems research, *Inf. Syst.*, **10** (2010), 1–51.
40. D. Tranfield, D. Denyer, P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review, *Br. J. Manage.*, **14** (2003), 207–222.
41. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, Interpretive Structural Modeling and MICMAC Analysis for Identifying and Benchmarking Significant Factors of Seismic Soil Liquefaction, *Appl. Sci.*, **9** (2019), 233.
42. L. Zhang, Predicting seismic liquefaction potential of sands by optimum seeking method, *Soil Dyn. Earthquake Eng.*, **17** (1998)**,** 219–226.
43. J. L. Hu, X. W. Tang, J. N. Qiu, A Bayesian network approach for predicting seismic liquefaction based on interpretive structural modeling, *Georisk: Assess. Manage. Risk Eng. Syst. Geohazards*, **9** (2015), 200–217.
44. D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, 2^{nd} edition, John Wiley & Sons, New York, 2000.
45. V. N. Vapnik, The nature of statistical learning, *Theory*, 1995.
46. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32.
47. G. John, P. Langley, Estimating Continuous Distributions in Bayesian Classifiers, in *proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 1995.
48. S. K. Das, R. Mohanty, M. Mohanty, M. Mahamaya, Multi-objective feature selection (MOFS) algorithms for prediction of liquefaction susceptibility of soil based on in situ test methods, *Nat. Hazards*, **103** (2020), 2371–2393.
49. M. Kubat, S. Matwin, Addressing the curse of imbalanced training sets: one-sided selection, in *Proceedings of the Fourteenth International Conference on Machine Learning*, (1997), 179–186.

50. B. Yuan, W. Liu, A measure oriented training scheme for imbalanced classification problems, in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, (2001), 293–303.
51. Y. Sun, M. S. Kamel, Y. Wang, Boosting for learning multiple classes with imbalanced class distribution, in *Proceedings of Sixth international conference on data mining (ICDM06)*, (2006), 592–602.

Appendix A

Table A1. Initial reachability matrix.

V_i	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}
V_1	1	1	0	0	0	0	0	0	0	1
V_2	0	1	0	0	0	0	0	0	0	1
V_3	0	0	1	1	1	0	0	0	0	1
V_4	0	0	0	1	0	0	0	0	0	1
V_5	0	1	0	1	1	0	0	0	0	1
V_6	0	0	0	1	1	1	0	0	0	1
V_7	0	0	0	1	0	1	1	0	0	1
V_8	0	0	0	1	0	1	0	1	0	1
V_9	0	0	0	0	0	0	0	0	1	1
V_{10}	0	0	0	0	0	0	0	0	0	1

Table A2. Final reachability matrix.

V_i	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	Dri.	Rank
V_1	1	1	0	0	0	0	0	0	0	1	3	III
V_2	0	1	0	0	0	0	0	0	0	1	2	IV
V_3	0	1*	1	1	1	0	0	0	0	1	5	I
V_4	0	0	0	1	0	0	0	0	0	1	2	IV
V_5	0	1	0	1	1	0	0	0	0	1	4	II
V_6	0	1*	0	1	1	1	0	0	0	1	5	I
V_7	0	0	0	1	1*	1	1	0	0	1	5	I
V_8	0	0	0	1	1*	1	0	1	0	1	5	I
V_9	0	0	0	0	0	0	0	0	1	1	2	IV
V_{10}	0	0	0	0	0	0	0	0	0	1	1	V
Dep.	1	5	1	6	5	3	1	1	1	10	34/34	
Rank	V	III	V	II	III	IV	V	V	V	I		**V/V**

Note: Dri.: driving power; Dep.: dependence power

Table A3. Level partition–Iteration 1.

V_i	S_r	S_a	S_i	L_i
V_1	V_1, V_2, V_{10}	V_1	V_1	
V_2	V_2, V_{10}	V_1, V_2, V_3, V_5, V_6	V_2	
V_3	V_2, V_3, V_4, V_5, V_{10}	V_3	V_3	
V_4	V_4, V_{10}	V_3, V_4, V_5, V_6, V_7, V_8	V_4	
V_5	V_2, V_4, V_5, V_{10}	V_3, V_5, V_6, V_7, V_8	V_5	
V_6	V_2, V_4, V_5, V_6, V_{10}	V_6, V_7, V_8	V_6	
V_7	V_4, V_5, V_6, V_7, V_{10}	V_7	V_7	
V_8	V_4, V_5, V_6, V_8, V_{10}	V_8	V_8	
V_9	V_9, V_{10}	V_9	V_9	
V_{10}	V_{10}	V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}	V_{10}	L_1

Table A4. Level partition–Iteration 2.

V_i	S_r	S_a	S_i	L_i
V_1	V_1, V_2	V_1	V_1	
V_2	V_2	V_1, V_2, V_3, V_5, V_6	V_2	L_2
V_3	V_2, V_3, V_4, V_5	V_3	V_3	
V_4	V_4	V_3, V_4, V_5, V_6, V_7, V_8	V_4	L_2
V_5	V_2, V_4, V_5	V_3, V_5, V_6, V_7, V_8	V_5	
V_6	V_2, V_4, V_5, V_6	V_6, V_7, V_8	V_6	
V_7	V_4, V_5, V_6, V_7	V_7	V_7	
V_8	V_4, V_5, V_6, V_8	V_8	V_8	
V_9	V_9	V_9	V_9	L_2

Table A5. Level partition–Iteration 3.

V_i	S_r	S_a	S_i	L_i
V_1	V_1	V_1	V_1	L_3
V_3	V_3, V_5	V_3	V_3	
V_5	V_5	V_3, V_5, V_6, V_7, V_8	V_5	L_3
V_6	V_5, V_6	V_6, V_7, V_8	V_6	
V_7	V_5, V_6, V_7	V_7	V_7	
V_8	V_5, V_6, V_8	V_8	V_8	

Table A6. Level partition–Iteration 4.

V_i	S_r	S_a	S_i	L_i
V_3	V_3	V_3	V_3	L_4
V_6	V_6	V_6, V_7, V_8	V_6	L_4
V_7	V_6, V_7	V_7	V_7	
V_8	V_6, V_8	V_8	V_8	

Table A7. Level partition–Iteration 5.

V_i	S_r	S_a	S_i	L_i
V_7	V_7	V_7	V_7	L_5
V_8	V_8	V_8	V_8	L_5