Adopting a Fuzzy C-Means Clustering Method to Evaluate Soil Liquefaction from SPT Measurements

Iman Ashayeri*, Mahnoosh Biglari and Vahid Veysinezhad

Civil Engineering Department, School of Engineering, Razi University, Kermanshah, Iran

Abstract: Soil liquefaction occurs during major earthquakes in saturated, relatively shallow soils and is the primary cause of building foundation failures. Hence, assessing the likelihood of soil liquefaction has become one of the most important areas of study in geotechnical and earthquake engineering. However, the complex nonlinear behavior of liquefiable soils and the nondeterministic nature of earthquakes make liquefaction analysis difficult and uncertain. In the present study, we address this uncertainty by applying fuzzy set theory and adapting the SPT-based cyclic stress approach to the Fuzzy C-means Clustering Method (FCM). In this context, several liquefaction records are compiled from the cited literature and used to assess the possibility of liquefaction occurrence. The effectiveness of this proposed method in handling the vagueness of liquefaction occurrence is compared with the conventional deterministic method and a probabilistic method of liquefaction assessment. The results show the new method's clear advantages over both the deterministic and probabilistic methods, particularly as the fine content of the liquefiable soil increases.

Keywords: Soil liquefaction, Earthquake hazard, Fuzzy C-means clustering, Reliability, Liquefaction possibility, Computer coding, Data mining.

1. INTRODUCTION

Soil liquefaction is a major cause of foundation failure, lifeline damage, and economic loss during and after a major earthquake. It typically occurs in saturated, relatively shallow soils because undissipated excess pore-water pressure during ground shaking significantly reduces shear stiffness. Consequently, assessing the likelihood of liquefaction has become a central focus in geotechnical and earthquake engineering and a key step in the design and construction of buildings and structures in earthquake-prone areas. However, the complex nonlinear behavior of the liquefiable soils and the nondeterministic nature of earthquakes make liquefaction analysis difficult and uncertain. Two major tools are used to evaluate liquefaction initiation: laboratory testing and in-situ tests. The former is more complex and expensive, requiring a representative soil sample to be tested under appropriate stress conditions and equivalent earthquake loading. The observed behavior of the soil element can then be used in complex numerical modeling for future analyses. The latter is simpler and less expensive, but evaluates liquefaction indirectly from earthquake loading. In the in-situ method, liquefaction analysis is performed by comparing the soil's state with known past liquefied cases. Among in-situ tests, the standard penetration test (SPT), cone penetration test (CPT), and measurement of the shear wave velocity (V_S) are the most commonly used.

Consequently, a capacity-demand analysis is performed. The seismic capacity is evaluated from the in-situ test, and the seismic demand is calculated from a site-specific seismic hazard analysis or approximated by the equivalent cyclic stress ratio (CSR_{eq}). Practically, the equivalent cyclic stress ratio is defined as a function of peak horizontal acceleration at the ground surface, soil element depth, initial shear and overburden stresses, and earthquake magnitude. To evaluate seismic capacity, screen analysis is performed based on available liquefaction events. Empirical relations have been developed to estimate the soil's seismic strength using the cyclic resistance ratio (*CRR*). The factor of safety against liquefaction is then calculated by dividing the seismic capacity by the seismic demand, i.e., CRR/CSR_{eq}. This deterministic approach is known as the cyclic stress approach for liquefaction assessment. The cyclic stress method was developed by Seed and Idriss [1] after earthquakes in Alaska and Niigata and evolved over more than twenty-five years into standard procedures, informed by the knowledge and experiences of more than 20 experts, as summarized in the NCEER (1996) and NCEER/NSF (1998) workshops [2]. Such a vast input of experts' influences and judgments, which has not been recorded in any other geotechnical engineering problem, has made the cyclic stress approach the most accepted and widely used method for liquefaction assessment in practice.

However, the vagueness of empirical relations and the simplification of the actual problem, on the one hand, and the value of machine intelligence in data mining and extracting information from numerical records, on the other, have led to the application of artificial intelligence in liquefaction assessment. This has been the subject of several investigations in recent years. The literature reports numerous applications of artificial neural networks (ANN) [3 to 8], genetic

*Address correspondence to this author at the Civil Engineering Department, School of Engineering, Razi University, Kermanshah, Iran;
E-mail: i.ashayeri@razi.ac.ir

Received on 16-09-2026 Accepted on 28-10-2026 Published on 02-11-2026

E-ISSN: 3083-3590/26

algorithms (GA), genetic programming, and gene expression programming (GP/GEP) [9], as well as hybrid systems such as the adaptive neuro-fuzzy inference system (ANFIS) [10, 11]. A state-of-the-art review presented liquefaction potential assessment of soils using machine learning techniques [12].

In this study, we attempt to address this vagueness using fuzzy set theory and adapt the cyclic stress approach to the Fuzzy C-means Clustering Method (FCM). We compile liquefaction records from the cited literature and use them. The effectiveness of this proposed method in handling the vagueness of liquefaction occurrence is compared with the conventional deterministic method recommended by [2] and a well-known probabilistic liquefaction assessment method developed by [13]. First, a brief review of the mathematical basis of FCM is presented, and secondly, the database of liquefaction events is presented and analyzed statistically. The third part introduces the Fuzzy C-means Clustering Method for Liquefaction Assessment (FCM-LA), and the final section discusses and compares the results with deterministic and probabilistic methods.

2. FUZZY C-MEANS CLUSTERING (FCM) AND COMPUTER CODE

Generally, clustering methods divide a dataset into a specified number of subsets, or clusters, so that each data point belongs to exactly one cluster. Following the pioneering work of L. A. Zadeh [14] and the introduction of fuzzy set theory, clustering has found many applications in different fields of engineering. This theory has provided a new foundation for classification and clustering techniques, in which an individual data point can be a member of more than one cluster, and the sum of its membership across all clusters equals unity. This idea was first introduced by Bezdek [15], and is briefly explained in the following.

Let *Y* be a set of *N* real *n*-dimensional vectors, that is $Y = \{y_1, y_2, \dots, y_N\} \subset R^n$, and $U=[u_{ik}]$ be a real $c \times N$ matrix where $2 \le c < N$ is the number of clusters. The element u_{ik} represents the membership of y_k in the i^{th} cluster and satisfies Equations (1 and 2).

$$u_{ik} \in [0,1] \tag{1}$$

$$\sum_{i=1}^{c} u_{ik} = 1 \quad for\ all\ k = 1\ to\ N \tag{2}$$

FCM aims to find vectors of *c* cluster centers $V = \{v_1, v_2, \dots, v_c\} \subset R^n$, so that the generalized least-squared errors (J_m) of Eq. (3) are minimized.

$$J_m(U,V) = \sum_{k=1}^{N} \sum_{i=1}^{c} \{(u_{ik})^m \|y_k - v_i\|_A^2\} \tag{3}$$

where *m* is the weighting exponent and theoretically can take any value in [1,∞); [15] stated that as *m* tends to unity, the clusters become increasingly nonfuzzy (hard), and conversely, each element of *U* tends to *1/c* as *m* tends to infinity. Meanwhile, no theoretical or computational evidence specifies the optimal value for *m*; [15] mentioned that taking $1.5 \le m \le 3$ gives reasonable results for *U* and *V*. Therefore, the value of *m* is set equal to 2 for this study, and the possible effects of other values of *m* on the final results are reserved for future investigations.

The parameter J_m is also a function of the squared distance between y_k and v_i, expressed by the A-norm on R^n that is $\| \quad \|_A$ defined in Eq. (4).

$$\|y_k - v_i\|_A^2 = (y_k - v_i)^T A (y_k - v_i) \tag{4}$$

Among the infinitely many A-norms available to be used in Eq. (4), three are most commonly used: the Euclidean, Diagonal, and Mahalanobis norms [16]. The Euclidean A-norm corresponds to $A = I$, where *I* is the $n \times n$ identity matrix, and identifies hyperspherical clusters [15]. Because geotechnical data are often well-suited to this cluster shape, the Euclidean A-norm is used in the present study.

The algorithm used in this study was presented and implemented in the FORTRAN-IV environment by [16]. The basic steps of the algorithm are as follows:

Step 1: Start with specific *c*, *m*, and a randomly valued temporal $\hat{U}$.

Step 2: Compute $\hat{V}$, the temporal cluster centers from Eq. (5).

$$\hat{v}_i = \frac{\sum_{k=1}^{N} (\hat{u}_{ik})^m y_k}{\sum_{k=1}^{N} (\hat{u}_{ik})^m} \quad for\ i = 1\ to\ c \tag{5}$$

Step 3: Compute the updated temporal $\hat{U}$ from Eq. (6).

$$\hat{u}_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\|y_k - \hat{v}_i\|_A^2}{\|y_k - \hat{v}_j\|_A^2} \right)^{\frac{1}{m-1}} \right)^{-1} \quad for\ k = 1\ to\ N\ and\ i = 1\ to\ c \tag{6}$$

Step 4: Compare the updated temporal $\hat{U}$ with the previous value. If the error is below an acceptable threshold (e.g., 0.01), stop.

Step 5: Repeat from Step 2 using the updated temporal $\hat{U}$.

This algorithm was translated, refined, and implemented in the Microsoft FORTRAN Power Station 4.0 environment for use in current and future studies; it is available in Appendix A for readers' interest.

3. EARTHQUAKE AND LIQUEFACTION EVENTS

The foundation of any data mining technique is a database that holds embedded information to be extracted for the desired application. In this study, we gather liquefaction records from well-cited literature. The database contains 697 liquefaction records from 14 earthquake events between 1891 and 1999 in Central America, North America, Japan, Turkey, and Taiwan. A total of 697 liquefaction records are used to build four sub-databases screened by the contents of fine materials. In the first sub-database (i.e., DBFC15), only records of samples with fine content of less than or equal to 15% are included. In the second sub-database (i.e., DBFC35), the fine content of soil samples is restricted to 35 percent. The third sub-database (i.e., DBFC35100) contains all records with fine content greater than 35 percent, and the last sub-database (i.e., DBFC100) contains all 697 data, in which the fine content varies from zero to 100 percent. The intention was to evaluate the performance and reliability of all investigated methods (i.e., FCM-LA, deterministic, and probabilistic) by varying the fine content as a key liquefaction parameter. Table **1** provides more detail on each sub-database population. Note that the maximum value for $(N_1)_{60cs}$ was capped at 50. As shown in Table **1**, the ratios of liquefied records across all four sub-databases are relatively similar and slightly below 0.5. This ratio will be used later to scale the final results of FCM-LA. Further statistical indices of the whole database are available in Table **B.1** of Appendix B.

4. FCM'S CONTRIBUTION TO LIQUEFACTION ASSESSMENT (FCM-LA)

This new approach is based on the cyclic stress approach. Preprocessing is required for all data to make them comparable in cyclic stress ratio and SPT number. Starting from the SPT number, it must be corrected for the effective overburden pressure, driving energy, and fine content to be presented as $(N_1)_{60cs}$ for all samples. This approach avoids selecting any correction relations. However, for the current database, the SPT numbers were already corrected to $(N_1)_{60}$ (Table **B.1**) and only require a fine-content correction to reach $(N_1)_{60cs}$. Accordingly, NCEER's recommendation [2], as in Eqs. (7, 7a, and 7b), is used.

$$(N_1)_{60cs} = \alpha + \beta (N_1)_{60} \tag{7}$$

$$\alpha = \begin{cases} 0 & FC \le 5\% \\ \exp\left[1.76 - \left(190/FC^2\right)\right] & 5\% < FC < 35\% \\ 5 & FC \ge 35\% \end{cases} \tag{7a}$$

$$\beta = \begin{cases} 1.0 & FC \le 5\% \\ \left[0.99 + \left(FC^{1.5}/1000\right)\right] & 5\% < FC < 35\% \\ 1.2 & FC \ge 35\% \end{cases} \tag{7b}$$

The cyclic stress ratio must be evaluated consistently across all records. Per NCEER recommendations [2], calculate the equivalent cyclic stress ratio for magnitude 7.5 earthquakes (i.e., $CSR_{eq\text{-}7.5}$) using Eq. (8).

$$CSR_{eq-7.5} = 0.65 \left(\frac{a_{max}}{g}\right) \left(\frac{\sigma_{v0}}{\acute{\sigma}_{v0}}\right) r_d \left(\frac{1}{MSF}\right) \tag{8}$$

where a_{max} is the peak horizontal acceleration at the ground surface, and σ_{v0} and $\acute{\sigma}_{v0}$ are the total and effective vertical stresses, respectively. Any stress reduction coefficient (r_d) and magnitude scaling factor (*MSF*) can be used. Meanwhile, the NCEER-recommended relationships [2] in Eqs. (9 and 10) are used here.

$$r_d = \begin{cases} 1.0 - 0.00765z & z \le 9.15m \\ 1.174 - 0.0267z & 9.15 < z \le 23m \\ 0.744 - 0.008z & 23 < z \le 30m \\ 0.5 & z > 30m \end{cases} \tag{9}$$

$$MSF = \frac{10^{2.24}}{M_w^{2.56}} \tag{10}$$

where M_w is the earthquake moment magnitude.

Figure **1** plots $CSR_{eq\text{-}7.5}$ versus $(N_1)_{60cs}$ for all 697 records, without distinguishing between liquefied and non-liquefied records. We also plot the NCEER-recommended [2] Cyclic Resistance Ratio $(CRR_{7.5})$, Eq. (11), as a dashed line for comparison.

$$CRR_{7.5} = \frac{1}{34-(N_1)_{60cs}} + \frac{(N_1)_{60cs}}{135} + \frac{50}{[10(N_1)_{60cs}+45]^2} - \frac{1}{200} \quad for\ (N_1)_{60cs} < 30 \tag{11}$$

With a data point's coordinates defined by $(N_1)_{60cs}$ and $CSR_{eq\text{-}7.5}$, it is time to assign the third attribute or any point in Figure **1**, namely the possibility of liquefaction. Accordingly, FCM-LA aims to cluster the

Table 1: The Population of Each Sub-Database

Database Name	Characteristic	Total number of records	Number of liquefied records	Liquefaction population ratio (LPR)
DBFC15	Fine content ≤ 15% $(N_1)_{60cs} \le 50$	177	80	0.4520
DBFC35	Fine content ≤ 35% $(N_1)_{60cs} \le 50$	281	129	0.4591
DBFC35100	35% < Fine content ≤ 100% $(N_1)_{60cs} \le 50$	416	180	0.4327
DBFC100	Fine content ≤ 100% $(N_1)_{60cs} \le 50$	697	309	0.4433

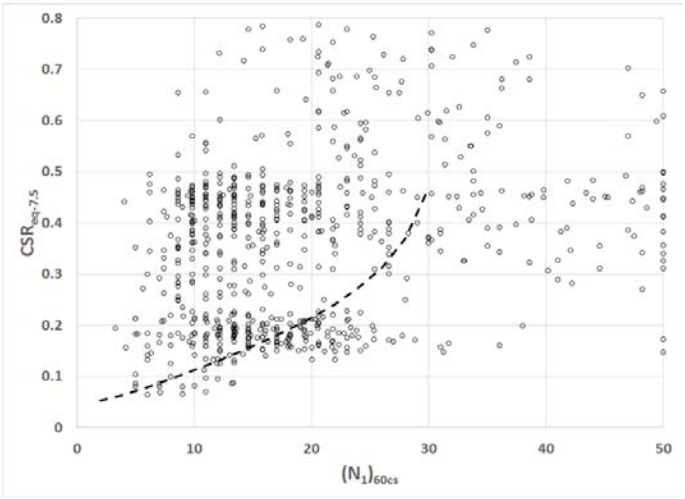


Figure 1: Demonstration of all 697 liquefaction records along with the NCEER's CRR.

two-dimensional space of $\{(N_1)_{60cs}, CSR_{eq-7.5}\}$ so that it can calculate a point's possibility of liquefaction based on its position in $\{(N_1)_{60cs}, CSR_{eq-7.5}\}$ coordinates. The proposed procedure is similar to [17], but has been simplified and adapted for the database of standard penetration tests of the current study. To achieve this goal, we applied the following procedure. Assume the coordinates of *N* records are available, with liquefaction attributes equal to 1 for liquefied and 0 for non-liquefied cases. In addition, we have a record with a specific $\{(N_1)_{60cs}, CSR_{eq-7.5}\}$ and an unknown liquefaction attribute. The FCM-LA algorithm takes the following steps to estimate the possibility of liquefaction for this case.

Step 1: Add the coordinates of the case with an unknown liquefaction attribute to the database of the other *N* records. This makes it the $N+1^{th}$ record in the database.

Step 2: For a desired number of clusters (*c*), run FCM-LA to obtain the coordinates of the cluster centers (*V*) and the membership values of the *N+1* records to those clusters (*U*).

Step 3: Calculate the liquefaction tendency of each cluster center (LTc_i) using Eq. (12).

$$LTc_i = \frac{\sum_{i=1}^{N} u_{ik} L_k}{\sum_{k=1}^{N} u_{ik}} \times \frac{0.5}{LPR} \quad (12)$$

where, L_k is the known liquefaction attribute for *N* records, and *LPR* is presented in Table **1** for each database in this study.

The numerator of Eq. (12) counts the fuzzy population of liquefied cases in the i^{th} cluster, and the denominator counts the fuzzy population of all cases in that cluster. Values of LTc_i closer to unity indicate that most of the population in the i^{th} cluster has $L_k=1$; conversely, values closer to zero indicate that most of the population in the i^{th} cluster has $L_k=0$. From a statistical perspective, the weighting factor (*0.5/LPR*) corrects for the slight non-uniformity in the populations of liquefied and non-liquefied records in each database.

Step 4: Calculate the unknown liquefaction possibility of the $N+1^{th}$ record (LP_{N+1}) using Eq. (13).

$$LP_{N+1} = \sum_{i=1}^{c} u_{iN+1} LTc_i \quad (13)$$

Theoretically, this procedure can be used to calculate the liquefaction possibility for $M > 1$ unknown records in a single run. However, the user must keep in mind that *M* must be significantly smaller than *N*. Although the authors know of no theoretical or computational method to determine the maximum acceptable *M/N* ratio, they find no significant benefit to running the procedure once for *M* unknowns rather than repeating it *M* times, one by one.

As illustrated in more detail in the following section, information extracted from the known liquefaction records in each database is stored in the cluster centers' LTc_i, which is used to evaluate the liquefaction possibility of the unknown case (LP_{N+1}). Therefore, one may ask what the suitable number of clusters is for liquefaction assessment. No explicit relation exists for the suitable number of clusters because it depends on the number of records in a database and their liquefaction attributes. However, we present a conceptual criterion here and apply it in the next section. Consider that the LTc_i for *c* clusters is determined by FCM-LA. It is helpful to count the number of cluster centers with LTc_i falling into 10 intervals, i.e., $LTc_i \in \{[0, 0.1), [0.1, 0.2), [0.2, 0.3) \ldots [0.9, 1]\}$. Then calculate the normalized ratio for each interval by dividing the number of cluster centers in each interval by the number of clusters. The standard deviation of the normalized ratio indicates how evenly clusters are distributed over LTc_i = [0, 1]. The FCM-LA

may start with fewer clusters and be repeated increasing the number of clusters. The suitable number of clusters is the one that yields the minimum standard deviation of the normalized ratio, or where further increases in the number of clusters do not change it considerably. This criterion ensures the clusters are uniformly distributed across LTc_i and avoids overpopulating the number of clusters.According to the explanations above, Table **2** presents the standard deviation of the normalized ratio across these 10 intervals of LTc_i, along with the selected values for suitable cluster numbers indicated in bold, for each sub-database.

5. EVALUATION OF FCM-LA

This section presents FCM-LA results for four sub-databases in Table **1**. First, we determine the optimal number of clusters for each sub-database. Second, we evaluate FCM-LA's information-extraction capability by showing the distribution of cluster centers in the $\{(N_1)_{60cs}, CSR_{eq\text{-}7.5}\}$ coordinate and corresponding LTc_i values for each cluster center. Finally, we compare FCM-LA's performance with NCEER's deterministic method for calculating the liquefaction factor of safety [2] and with a probabilistic approach for determining liquefaction probability using SPT tests [13].

5.1. Liquefaction Tendency of the Cluster Centers

The liquefaction tendencies of cluster centers (LTc_i) are key to evaluating each record's liquefaction possibilities (Eq. 13). Thus, Figure **2** to **5** show the distribution of cluster centers and their LTc_i for each sub-database at the cluster numbers selected in Table **2**. The spatial distribution of cluster centers reflects the spatial distribution of records. Consequently, different cluster centers appear in Figure **2** and **3** and in Figure **4** and **5**, even though each pair of sub-databases has the same number of clusters. The LTc_i values are shown as data labels in Figure **2** to **5** and clearly indicate that increasing $(N_1)_{60cs}$ reduces the liquefaction tendency of

Table 2: The Standard Deviation of the Normalized Population of Clusters for Different Cluster Numbers at Each Sub-Database

Database							
DBFC15		**DBFC35**		**DBFC35100**		**DBFC100**	
No. of clusters	**St. Dev.**	**No. of clusters**	**St. Dev.**	**No. of clusters**	**St. Dev.**	**No. of clusters**	**St. Dev.**
10	0.094	10	0.115	35	0.085	35	0.105
15	0.057	15	0.085	40	0.085	40	0.091
20	0.062	20	0.047	50	0.076	50	0.072
25	**0.043**	**25**	**0.066**	60	0.074	60	0.083
30	0.052	30	0.068	70	0.065	70	0.073
35	0.047	35	0.065	**80**	**0.061**	**80**	**0.066**
-	-	40	0.076	90	0.067	90	0.068
-	-	-	-	100	0.061	100	0.064

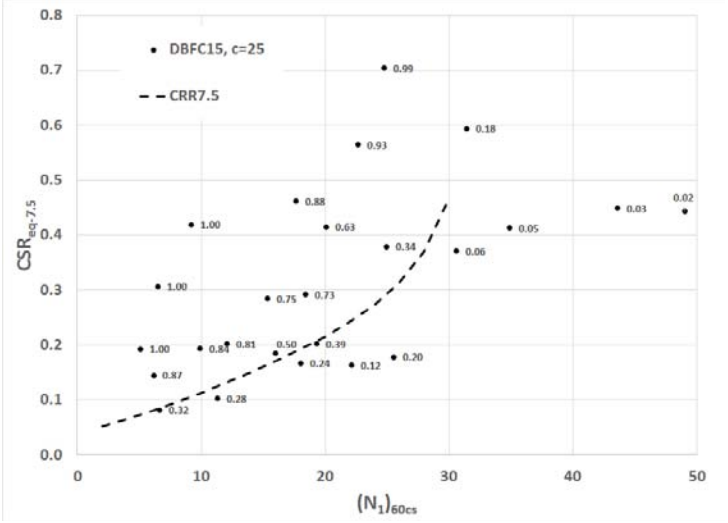


Figure 2: Distribution of cluster centers for DBFC15 with 25 clusters and LTc_i of each cluster.

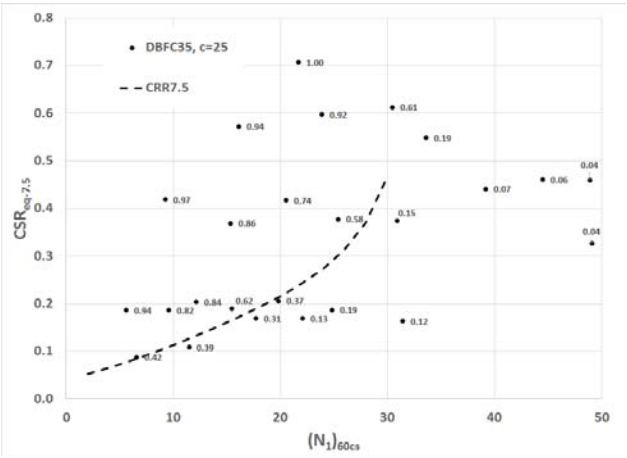


Figure 3: Distribution of cluster centers for DBFC35 with 25 clusters and LTc_i of each cluster.

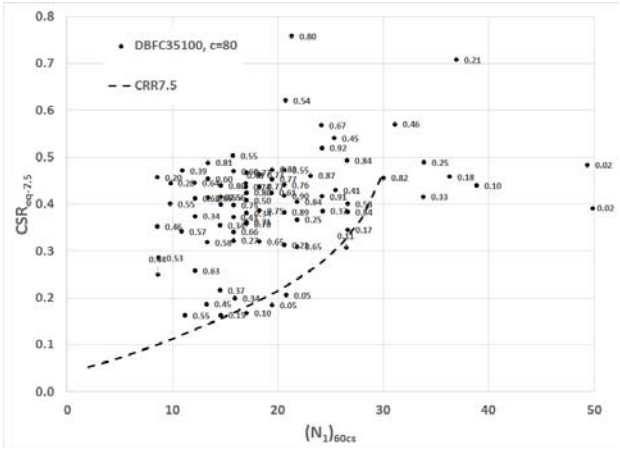


Figure 4: Distribution of cluster centers for DBFC35100 with 80 clusters and LTc_i of each cluster.

cluster centers, whereas decreasing $CSR_{eq-7.5}$ increases it. The NCEER-recommended $CRR_{7.5}$ curve is plotted in Figure **2** to **5** for comparison. In Figure **2** and **3**, the $CRR_{7.5}$ curve clearly separates $LTc_i > 0.5$, and all cluster centers with $LTc_i > 0.5$ lie above the curve, except for one center in each figure. However, in Figure **4** and **5**, some cluster centers with $LTc_i < 0.5$ lie above the $CRR_{7.5}$ curve. This indicates that FCM-LA and deterministic approaches diverge more as the fine content exceeds 35 percent. This will be illustrated in more detail in the next section.

5.2. Comparison of FCM-LA with the Deterministic and Probabilistic Methods

FCM-LA aims to better address the vagueness in liquefaction assessment. This section presents the liquefaction possibility for each record across four sun-databases using FCM-LA. To provide a basis for comparison, we calculated the factor of safety against liquefaction (e.g., Eq. 14), as recommended by NCEER [2], and the probability of liquefaction (P_L) [13] for all records.

$$F.S. = \frac{CRR_{7.5}}{CSR_{eq-7.5}} \tag{14}$$

The liquefaction records for Loma Prieta (1989) from [18] lack values for total and effective overburden stresses (Table **B.1**); therefore, only their factor of safety against liquefaction could be calculated. Hence, all comparisons with the probabilistic method exclude records from [18]. The results from all methods are presented for each record in Table **B.1**.

We evaluate each method's reliability in liquefaction assessment using the liquefaction victory ratio, defined as the number of correctly recognized liquefied cases divided by the total number of cases evaluated as liquefied. This ratio quantifies how reliable each method's liquefied classification is. Thus, criteria for liquefaction initiation must be defined. The criteria are defined in terms of the liquefaction possibility (LP_i), liquefaction factor of safety ($F.S.$), and liquefaction probability (P_L). Table **3** defines these criteria and presents the number of successful assessments of liquefied cases, the number of cases indicated as liquefied, and their ratio, the liquefaction victory ratio.

The difference between FCM-LA and the probabilistic approach of [13] is clear from the results in Table **3**. In all cases, FCM-LA's liquefaction victory ratio is much higher than that of the probabilistic approach.

Table 3: Statistical Analyses of the Reliability of the Liquefaction Assessment Method for Four Sub-Databases

Assessment method	Database	Criteria for liquefaction	No. of liquefied records*	No. of correct liquefaction recognition*	Liquefaction victory ratio*
FCM-LA	DBFC15	$LP_i > 0.5$	81 (72)	70 (63)	0.864 (0.875)
Cetin *et al*. (2004)		$P_L > 0.5$	103	67	0.650
F.S. (NCEER)		$F.S. < 1$	93 (83)	73 (66)	0.785 (0.795)
FCM-LA		$LP_i > 0.6$	72 (65)	66 (60)	0.917 (0.923)
Cetin *et al*. (2004)		$P_L > 0.6$	100	67	0.670
F.S. (NCEER)		$F.S. < 0.75$	74 (68)	66 (61)	0.892 (0.897)
FCM-LA		$LP_i > 0.7$	61 (56)	58 (54)	0.951 (0.964)
Cetin *et al*. (2004)		$P_L > 0.7$	95	66	0.695
F.S. (NCEER)		$F.S. < 0.5$	39 (38)	38 (37)	0.974 (0.974)
FCM-LA		$LP_i > 0.8$	43 (40)	42 (40)	0.977 (1)
Cetin *et al*. (2004)		$P_L > 0.8$	88	66	0.750
FCM-LA		$LP_i > 0.9$	21 (20)	21 (20)	1 (1)
Cetin *et al*. (2004)		$P_L > 0.9$	81	65	0.802
FCM-LA	DBFC35	$LP_i > 0.5$	139 (129)	115 (107)	0.827 (0.829)
Cetin *et al*. (2004)		$P_L > 0.5$	174	114	0.655
F.S. (NCEER)		$F.S. < 1$	144 (133)	118 (110)	0.819 (0.827)
FCM-LA		$LP_i > 0.6$	126 (117)	111 (103)	0.881 (0.880)
Cetin *et al*. (2004)		$P_L > 0.6$	169	113	0.669
F.S. (NCEER)		$F.S. < 0.75$	116 (110)	104 (99)	0.897 (0.900)
FCM-LA		$LP_i > 0.7$	100 (95)	91 (87)	0.910 (0.916)
Cetin *et al*. (2004)		$P_L > 0.7$	162	112	0.691
F.S. (NCEER)		$F.S. < 0.5$	62 (61)	58 (57)	0.935 (0.934)
FCM-LA		$LP_i > 0.8$	40 (38)	39 (37)	0.975 (0.974)
Cetin *et al*. (2004)		$P_L > 0.8$	151	112	0.742
FCM-LA		$LP_i > 0.9$	17 (16)	17 (16)	1 (1)
Cetin *et al*. (2004)		$P_L > 0.9$	143	111	0.776
FCM-LA	DBFC35100	$LP_i > 0.5$	(210)	(148)	(0.705)
Cetin *et al*. (2004)		$P_L > 0.5$	341	148	0.434
F.S. (NCEER)		$F.S. < 1$	(305)	(132)	(0.433)
FCM-LA		$LP_i > 0.6$	(123)	(103)	(0.837)
Cetin *et al*. (2004)		$P_L > 0.6$	338	146	0.432
F.S. (NCEER)		$F.S. < 0.75$	(289)	(127)	(0.439)
FCM-LA		$LP_i > 0.7$	(71)	(67)	(0.944)
Cetin *et al*. (2004)		$P_L > 0.7$	331	143	0.432
F.S. (NCEER)		$F.S. < 0.5$	(236)	(102)	(0.432)
FCM-LA		$LP_i > 0.8$	(29)	(29)	(1)
Cetin *et al*. (2004)		$P_L > 0.8$	323	141	0.437
FCM-LA		$LP_i > 0.9$	(5)	(5)	(1)
Cetin *et al*. (2004)		$P_L > 0.9$	319	140	0.439
FCM-LA	DBFC100	$LP_i > 0.5$	412 (405)	275 (269)	0.667 (0.664)
Cetin *et al*. (2004)		$P_L > 0.5$	565	294	0.520
F.S. (NCEER)		$F.S. < 1$	484 (473)	289 (281)	0.597 (0.594)
FCM-LA		$LP_i > 0.6$	262 (256)	198 (193)	0.756 (0.754)

Table 3 (Continue)

Assessment method	Database	Criteria for liquefaction	No. of liquefied records*	No. of correct liquefaction recognition*	Liquefaction victory ratio*
Cetin *et al*. (2004)		P_L > 0.6	558	293	0.525
F.S. (NCEER)		*F.S.* < 0.75	410 (404)	(253)	0.629 (0.626)
FCM-LA		LP_i > 0.7	105 (102)	98 (95)	0.933 (0.931)
Cetin *et al*. (2004)		P_L > 0.7	545	291	0.534
F.S. (NCEER)		*F.S.* < 0.5	289 (288)	173 (172)	0.599 (0.597)
FCM-LA		LP_i > 0.8	26 (24)	26 (24)	1 (1)
Cetin *et al*. (2004)		P_L > 0.8	533	291	0.546
FCM-LA		LP_i > 0.9	4 (3)	4 (3)	1 (1)
Cetin *et al*. (2004)		P_L > 0.9	519	289	0.557

* Values in parentheses must be compared with the probabilistic method of Cetin *et al*. (2004) in [13].

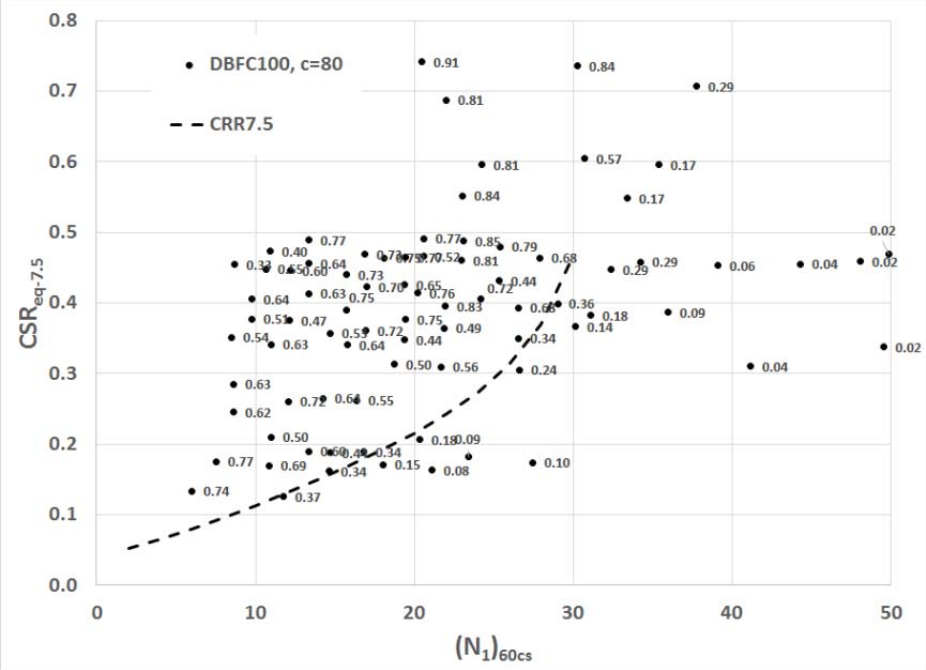


Figure 5: Distribution of cluster centers for DBFC100 with 80 clusters and LTc_i of each cluster.

For instance, for records in DBFC 15, any LP_i greater than 0.7 indicates liquefaction in about 95 percent of cases. Meanwhile, any P_L greater than 0.7 for the same records indicates liquefaction in about 70 percent of cases. Moreover, if LP_i > 0.9, the record is identified as liquefied with certainty. However, the best liquefaction victory ratio for P_L > 0.9 is about 0.8.

The results also show FCM-LA's advantages over the deterministic approach. Although the liquefaction victory ratio for the deterministic approach is only slightly smaller than FCM-LA for the records in DBFC15 and DBFC35, this trend does not hold for records with fine content above 35 percent. Figures **6** to **8** show that, as fine content exceeds 35 percent, both the deterministic and probabilistic methods lose reliability more significantly than FCM-LA. Notably, reducing the factor-of-safety criterion for liquefaction initiation from 1 to 0.5 does not improve the liquefaction victory ratio in the deterministic method, and similarly, increasing the probability criterion for liquefaction initiation from 0.5 to 0.9 does not significantly increase the liquefaction victory ratio for the DBFC35100 and DBFC100 records.

6. SUMMARY AND CONCLUSION

The cyclic stress approach is widely used in practice to evaluate the liquefaction potential of saturated soils largely because of its simplicity and close connection to in-situ tests. This study addresses uncertainties in liquefaction assessment using fuzzy set theory and derives liquefaction possibility from certified experiments cited in the literature. First, the FCM-LA was used to determine the membership values of liquefaction records to clusters optimally

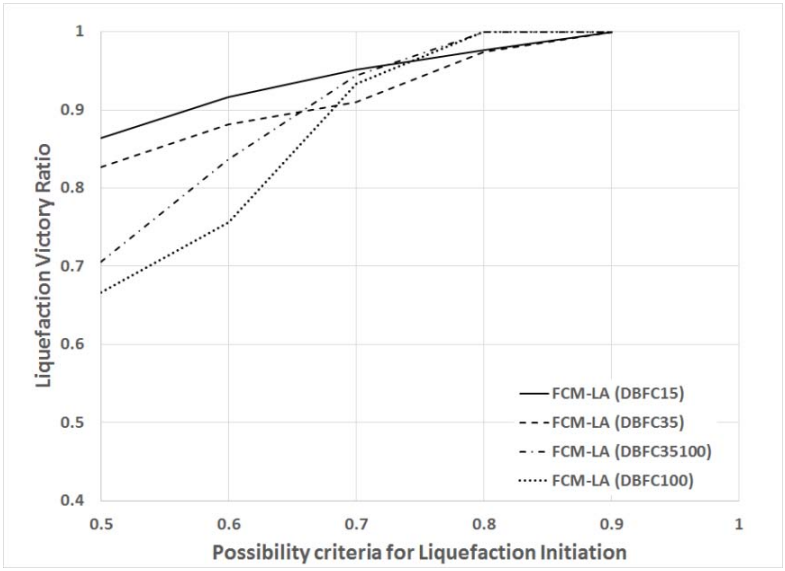


Figure 6: Liquefaction victory ratio of FCM-LA at different liquefaction initiation criteria for each sub-database.

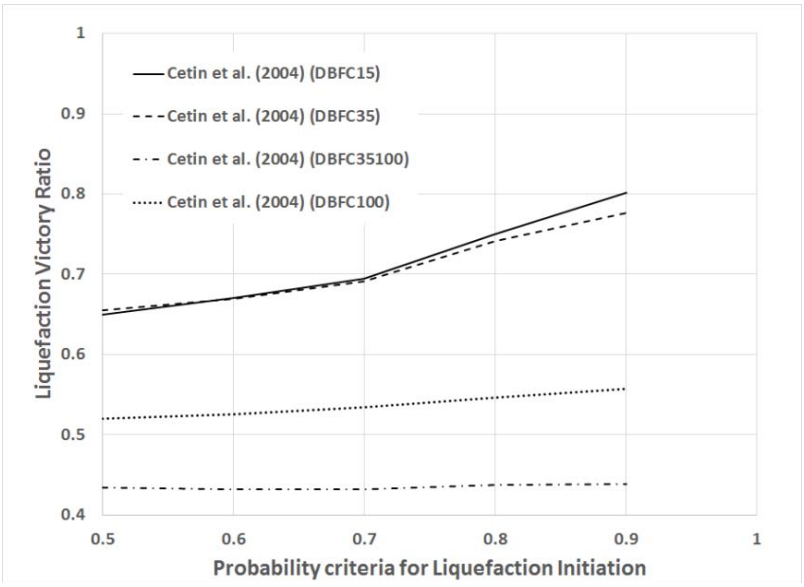


Figure 7: Liquefaction victory ratio of the probabilistic method at different liquefaction initiation criteria for each sub-database.

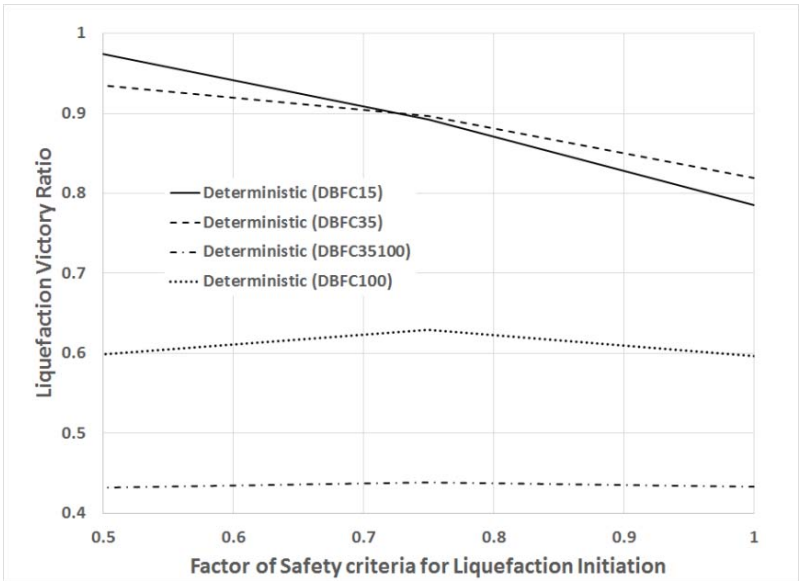


Figure 8: Liquefaction victory ratio of deterministic method at different liquefaction initiation criteria for each sub-database.

distributed in the $\{(N_1)_{60cs}, CSR_{eq\text{-}7.5}\}$ coordinates. Second, a new procedure was proposed to determine the liquefaction tendency of the cluster centers, which were then used to calculate the liquefaction possibility of an unknown record. The performance of the FCM-LA procedure was compared with the deterministic method that gives the factor of safety against liquefaction, as well as the probabilistic approach developed by [13], which calculates the probability of liquefaction. The outcomes of the analyses are summarized as follows:

1. The liquefaction tendency of cluster centers shows that increasing $(N_1)_{60cs}$ and decreasing $CSR_{eq\text{-}7.5}$ reduce liquefaction possibility as a general rule, particularly for soils with less than 35 percent fine materials.

2. The newly introduced liquefaction possibility handles liquefaction vagueness better than the liquefaction factor of safety and the probability of liquefaction used in the literature. This was more pronounced in soils with more than 35 percent fine content.

3. The predominant information about liquefaction occurrence differs across sub-databases categorized by fine content. Fine content strongly influenced the performance of both the deterministic and probabilistic methods. In other words, modifications to the deterministic method and the adopted probability distribution function used in the probabilistic method cannot adequately model liquefaction occurrence in soils with fine contents above 35 percent. However, FCM-LA provides more reliable results because it can store the predominant information of each sub-database independently. Thus, it is recommended to use the matching sub-database to calculate liquefaction possibility because unrelated records introduce noisy information.

4. The FCM-LA is reasonably founded and easy to implement in computer codes. Furthermore, the FCM-LA is implemented independently of the database, and future liquefaction records will not undermine its reliability; instead, they will confirm its abilities and advantages over other liquefaction assessment approaches, which must be modified to deal with new events and observations, meaning updating the criteria or probability density functions in the deterministic and probabilistic methods, respectively. This can be seen as an early step toward using data mining techniques globally, through information science and technology tools like the web, for liquefaction assessment.

APPENDIX MATERIALS

The appendix materials can be downloaded from the journal website along with the article.

REFERENCE

[1] Seed HB, Idriss IM. Simplified procedure for evaluating soil liquefaction potential. J Soil Mech Found Div 1971; 97(9): 1249-73. https://doi.org/10.1061/JSFEAQ.0001662

[2] Youd TL, Idriss IM, Andrus RD, Arango I, Castro G, Christian JT, Dobry R, Finn WDL, Harder LF, Hynes ME, Ishihara K, Koester JP, Liao ASC, Marcuson WF, Martin R, Mitchell JK, Moriwaki Y, Power MS, Robertson PK, Seed RB, Stokoe KH. 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 Eng 2001; 127(10): 817-33. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:10(817)

[3] Tung ATY, Wang YY, Wong FS. Assessment of liquefaction potential using neural networks. Soil Dynamics and Earthquake Engineering 1993; 12: 325-35. https://doi.org/10.1016/0267-7261(93)90035-P

[4] Goh ATC. Seismic liquefaction potential assessed by neural networks. J Geotech Geoenviron Eng Div 1994; 120(9): 1467-80. https://doi.org/10.1061/(ASCE)0733-9410(1994)120:9(1467)

[5] Goh ATC. Neural network modeling of CPT seismic liquefaction data. J Geotech Eng 1996; 122(1): 70-3. https://doi.org/10.1061/(ASCE)0733-9410(1996)122:1(70)

[6] Juang CH, Chen CJ, Tien Y-M. Appraising cone penetration test based liquefaction resistance evaluation methods: artificial neural network approach. Can Geotech J 1999; 36: 443-54. https://doi.org/10.1139/t99-011

[7] Baziar MH, Nilipour N. Evaluation of liquefaction using neural networks and CPT results. Soil Dynamics and Earthquake Engineering 2003; 23: 631-6. https://doi.org/10.1016/S0267-7261(03)00068-X

[8] Hanna AM, Ural D, Saygili G. Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data. Soil Dynamics and Earthquake Engineering 2007; 27: 521-40. https://doi.org/10.1016/j.soildyn.2006.11.001

[9] Kayadelen C. Soil liquefaction modeling by Genetic Expression Programming and Neuro-Fuzzy. Expert Systems with Applications 2011; 38: 4080-7.[10] Chern SG, Lee CY, Wang CC. CPT-Based Liquefaction Assessment by Using Fuzzy-Neural Network. J Marine Science Tech 2008; 16(2): 139-48. https://doi.org/10.1016/j.eswa.2010.09.071

[10] Yazdi JS, Kalantary F, Yazdi HS. Prediction of liquefaction potential based on CPT up-sampling. Computers & Geotechnics 2012; 44: 10-23. https://doi.org/10.1016/j.cageo.2012.03.025

[11] Jas K, Dodagoudar GR. Liquefaction Potential Assessment of Soils Using Machine Learning Techniques: A State-of-the-Art Review from 1994-2021. Int J Geomech 2023; 23(7). https://doi.org/10.1061/IJGNAI.GMENG-7788

[12] Cetin KO, Seed RB, Kiureghian AD, Tokimatsu K, Harder LF, Kayen RE, Moss RES. Standard Penetration Test-Based Probabilistic and Deterministic Assessment of Seismic Soil Liquefaction Potential. J Geotech Geoenviron Eng 2004; 130(12): 1314-40. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:12(1314)

[13] Zadeh LA. Fuzzy sets. Inf and Cont 1965; 8: 338-53. https://doi.org/10.1016/S0019-9958(65)90241-X

[14] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. New York: Plenum; 1981. https://doi.org/10.1007/978-1-4757-0450-1

[15] Bezdek JC, Ehrlich R, Full W. FCM: The Fuzzy c-Means Clustering Algorithm. Computers and Geosciences 1984; 10(2-3): 191-203. https://doi.org/10.1016/0098-3004(84)90020-7

[16] Mohammadikish, S., Ashayeri, I., Biglari, M. *et al*. Soil Liquefaction Assessment by CPT and VS Data and Incomplete-Fuzzy C-Means Clustering. Geotech Geol Eng 42, 2205-2220 (2024). https://doi.org/10.1007/s10706-023-02669-1

[17] Boulanger RW, Mejia LH, Idriss IM. Liquefaction at Moss Landing During Loma Prieta Earthquake. J Geotech Geoenviron Eng 1997; 123(5): 453-67. https://doi.org/10.1061/(ASCE)1090-0241(1997)123:5(453)

https://doi.org/10.65904/3083-3590.2026.02.11

© 2026 Ashayeri *et al*.
This is an open access article licensed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited.