Soil Dynamics and Earthquake Engineering 130 (2020) 105988

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: http://www.elsevier.com/locate/soildyn

Evaluation of soil liquefaction using AI technology incorporating a coupled ENN / t-SNE model

Pierre Guy Atangana Njock [a], Shui-Long Shen [b,c,*], Annan Zhou [d], Hai-Min Lyu [e]

[a] Department of Civil Engineering, School of Naval Architecture, Ocean, and Civil Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Minhang District, Shanghai, 200240, China
[b] Department of Civil and Environmental Engineering, College of Engineering, Shantou University, Shantou, Guangdong, 515063, China
[c] Key Laboratory of Intelligent Manufacturing Technology (Shantou University), Ministry of Education, Shantou, Guangdong, 515063, China
[d] Discipline of Civil and Infrastructure, School of Engineering, Royal Melbourne Institute of Technology (RMIT), Victoria, 3001, Australia
[e] State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau S.A.R., China

ARTICLE INFO

Keywords:
Differential evolution
Optimisation
t-SNE
Neural network
Liquefaction

ABSTRACT

This paper presents a new evolutionary neural network (ENN) algorithm coupled with the dimensionality reduction technique 't-distributed stochastic neighbour embedding' (t-SNE). The ENN model features the crossbreeding of a differential evolution method and a stochastic gradient optimisation algorithm. The t-SNE is used to visualise the training and testing datasets and the ENN model performance. The proposed ENN model is applied to a relatively large soil liquefaction database. The good convergence and generalisation ability of the proposed model and the negligible misclassification results demonstrate that the proposed ENN model can provide accurate, efficient, and flexible results. The prominent and practical abilities of t-SNE to recover the structure of the initial conditions and to demonstrate the ENN model performance are discussed. This coupled approach simplifies the analysis and/or prediction of hazards for which large quantities of data are required.

1. Introduction

One of the focuses of civil engineers in the past three decades has been to tackle the soil liquefaction phenomenon efficiently, given the random nature of its occurrence. Soil liquefaction occurs when a saturated or partially saturated soil substantially loses strength and stiffness in response to an applied stress (such as shaking during an earthquake or other sudden changes in its stress condition), in which a material that is ordinarily a solid behaves like a liquid [1,2]. Soil liquefaction occurs when the effective stress (shear strength) of the soil is reduced to essentially zero and imposes a real risk on the surrounding structures [3–8]. The dedication of engineers and researchers has thus been triggered and further reinforced by a plethora of casualties [9] and the severity of structural damages [10–13] caused by this phenomenon. However, the prediction of soil liquefaction has remained a formidable task owing to the nonlinearity of soil behaviour [14–22] and the characteristics of the seismic power dissipation. The traditional methods for predicting soil liquefaction [23–26] still have numerous limitations, which have initiated the proposals of more powerful (efficient, flexible, and accurate) strategies. Among them, data mining (DM) has attracted particular attention recently. Data mining (DM) is a multidisciplinary subfield of computer science that includes statistics, database technology, and machine learning for the analysis of previously unknown or unsuspected relationships buried in large datasets. The DM algorithms are increasingly accepted in geotechnical engineering and have generally shown an excellent ability to solve multi-dimensional and complex problems [27].

DM techniques have been employed for regression and data classification when describing and analysing soil liquefaction. For example, researchers have taken advantage of the various activation functions and nonlinear mapping aptitude of artificial neural networks (ANNs) to assess the liquefaction potential of soil [28–33]. According to the literature, ANNs remain the leading DM systems applied in soil liquefaction research owing to their strong capacities regarding complicated nonlinear problems. Some researchers have focused on the attractiveness of the 'kernel' used in support vector machine (SVM) algorithms to determine prediction models for soil liquefaction [34–38]. Some other attempts of using DM techniques in soil liquefaction include the

* Corresponding author. Department of Civil and Environmental Engineering, College of Engineering, Shantou University, Shantou, Guangdong, 515063, China.
E-mail addresses: atangana@sjtu.edu.cn (P.G. Atangana Njock), shensl@stu.edu.cn (S.-L. Shen), annan.zhou@rmit.edu.au (A. Zhou), haiminlyu@um.edu.mo (H.-M. Lyu).

https://doi.org/10.1016/j.soildyn.2019.105988
Received 11 March 2019; Received in revised form 24 November 2019; Accepted 24 November 2019
Available online 3 December 2019

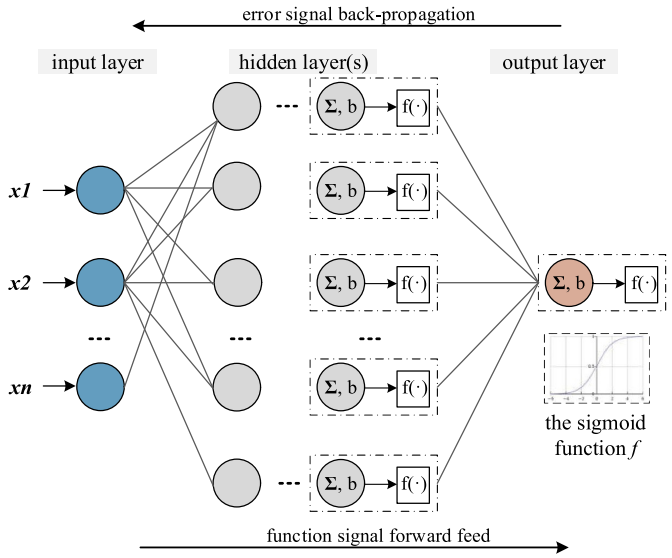

Fig. 1. Illustration of typical back-propagation artificial neural network (BP-ANN) architecture.

relevance vector machine [39], adaptive neuro fuzzy inference system [40,41], and decision tree [42].

Nevertheless, the DM techniques have caused concern, the major one being the well-known problem of convergence to local minima [43]. Consequently, researchers have attempted to adopt new computation paradigms for developing neural networks with enhanced predictability and computation performances. This recent practice has mainly focused on improving the capacity of DM approaches with two popular optimisation algorithms: particle swarm optimisation (PSO), and the genetic algorithm (GA) [37,44–46]. However, these optimisation methods remain limited in terms of efficiency for certain applications. The interpretability of the results is another important problem associated with the application of models based on DM techniques. The difficulty in interpreting the results hampers the definition of the effect of given input parameters on an objective output.

An additional problem related to the interpretability of the data is their high dimensionality. Whereas processing high-dimensional data can be easy for computers, the same task remains difficult for the human brain, which is limited to three dimensions. The effective visualisation of this data type in a low-intrinsic dimensionality is therefore important and still presents a challenge [47–49]. A low-intrinsic dimensionality is usually embedded within a high-dimensional space in a nonlinear and complex way. Moreover, analysing soil liquefaction data involves the management of several variables (or dimensions) that are routinely tabulated in a multi-entry table. These data become extremely difficult to interpret when many observational data are involved. However, the quantity of observational data is crucial for soil liquefaction studies. The accuracy of the results predicted by the DM algorithms is in general proportional to the quantity of observation data. To the best of the authors' knowledge, the applications of dimensionality reduction techniques to geotechnical engineering problems such as soil liquefaction are very limited.

In this paper, the existing optimised prediction models are briefly reviewed. A hybrid model combining the differential evolution algorithm (DEA) and stochastic optimisation approach is proposed. In addition, a novel dimensionality reduction technique is adopted to visualise the initial dataset and to recover the structure of the trained data to prove the performance of the proposed model. Moreover, the proposed evolutionary neural network (ENN) is validated with a value-added soil liquefaction database.

2. Brief review of existing methods

2.1. Dependency on weights and biases

Artificial neural networks (ANNs) are powerful analytical tools for multifaceted problems. This innovation in the AI field was achieved by mimicking some of the fundamental features of the human brain such as self-organisation, adaptation, and fault tolerance. Since the pioneering work of McCulloch and Pitts [50], the principle of ANNs has remained the same: to learn the behaviour of a given system based on physical observations, and to predict future occurrences.

As shown in Fig. 1, a conventional BP-ANN can be discretised into a sequence of three layers depending on the transformations performed by the neurons on the signal path. A typical neural network consists of simple information-processing units (or neurons) and fully weighted

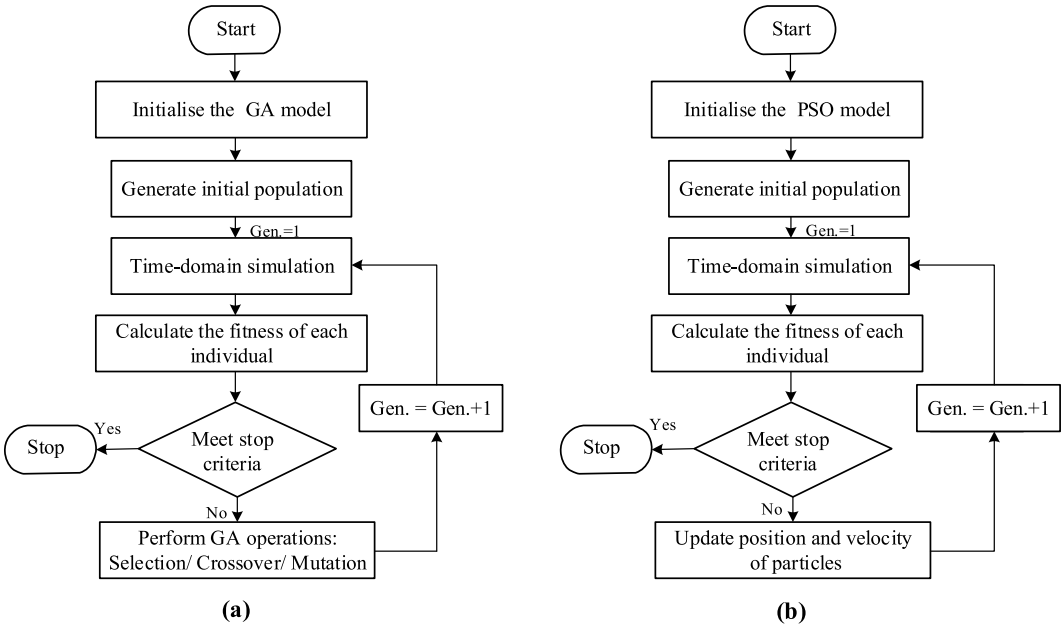

Fig. 2. Flowcharts of (a) genetic algorithm (GA) and (b) particle swarm optimisation (PSO).

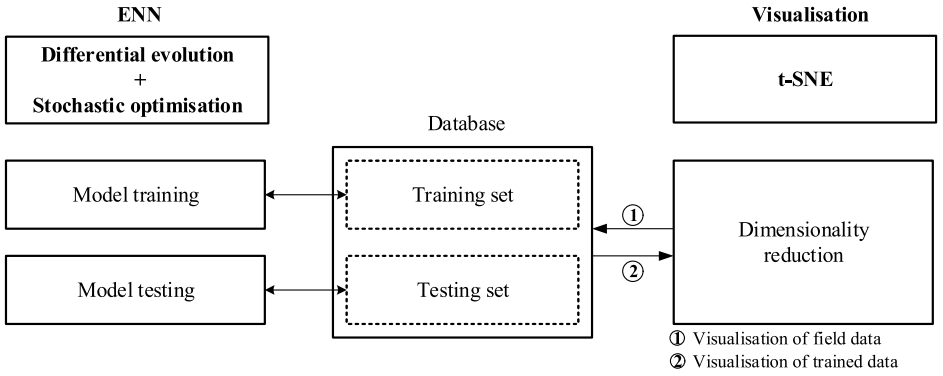

Fig. 3. Simplistic illustration of coupled t-SNE/ENN approach.

connections between these neurons. The strength of the connection between two neurons i and j is usually defined by a synaptic weight w^{ji}, which is eventually regulated by a bias b^{ij} throughout the learning process. To be more specific, bias units are used to scale the inputs to a certain range to enhance the convergence properties of the network. In general, an accurate estimation of the target output is rather impossible with the initially selected weights and biases. Hence, these two parameters must be constantly modified throughout the training process to achieve a minimal error value. In this process, the signal forward-feed is first achieved with an activation function. Next, the error signal back-propagation (BP) is executed from the output to the input layers by adjusting the weights and biases. Kolen and Pollack [51] demonstrated the strong reliance of the BP learning scheme on the initial weights.

Following this procedure, the updated value of the weights w^{ji} and biases b^{ij} can be expressed as follows [51]:

$$\begin{cases} w_l^{ji} = w_l^{ji} - \alpha \dfrac{\partial L}{\partial w_l^{ji}} \\ b_l^{j} = b_l^{j} - \alpha \dfrac{\partial L}{\partial b_l^{j}} \end{cases}, \tag{1}$$

where the loss L is defined using the following equation::

$$L(\mathbf{w}, b) = \frac{1}{M} \sum_{n=1}^{M} \log(\cosh(\widehat{k}_n - k_n)). \tag{2}$$

In Eqs. (1)–(6), α is the learning rate that governs the parameter updating; k_n and $\widehat{k}_n$ represent the predicted and target values of the nth sample, respectively. With the repeated arrival of feedback of results in the same training dataset, the loss L decreases gradually. However, in certain circumstances, the performance of the model might worsen drastically from the initial run (initial training data) to the subsequent ones (new data): the so-called 'overfitting phenomenon' occurs. This issue is critical for ANNs because the 'over-fitted' models cannot often provide reliable generalisations. Thus, optimisation algorithms are employed to improve the efficiency of BP–ANN algorithms.

2.2. Optimisation algorithms applied to liquefaction studies

PSO [52] and the GA [53,54] are two well-known optimisation techniques that are commonly implemented to enhance the capabilities of machine learning models. Fig. 2 provides an epigrammatic comparison of these two methods. In GA, fitness proportional selection and genetic recombination are conducted. The respective counterparts in PSO are best selection and the self-governing movement of particles.

Xue and Liu [44] employed the GA and PSO approaches to improve the capacity of their ANN models for assessing seismic liquefaction potential. They consecutively assessed and compared the performances of GA- and PSO-improved models, which achieved fairly satisfactory results. However, these processes required thorough parametric studies to determine the optimal GA and PSO parameters. Moreover, a relatively small quantity of data was used to train their optimised neural networks. However, when the quantity of data increases, the network structure might become more complex and involve multiple hidden layers. Under such conditions, the computation efficiency will be severely hampered. Xue et al. [46] adopted a different machine learning technique—the SVM approach—to evaluate the seismic liquefaction potential based on cone penetration test (CPT) data. The GA was adopted as optimisation approach to determine the optimal values of the penalty parameter and Kernel function in the SVM model. A similar study was conducted in Ref. [37]; the PSO approach was employed to optimise the selection of SVM parameters. Although these models generate acceptable results, the implementation procedure remains less efficient because any data updating requires the retuning of optimisation parameters. It should be noted that the absence of crossovers and mutation operators in the PSO can sometimes be problematic owing to the issue of local minima. To solve this issue, Rahbarzare and Azadi [45] recently proposed the crossbreeding of GA and PSO and integrated it into a fuzzy SVM model. In their model, the GA operators were incorporated in the PSO algorithm to determine the best classifier parameters. This astute strategy allowed them to considerably increase the accuracy of the classifier and, hence, to avoid the recourse to the laborious 'trial and error' procedure. In summation, both GA and PSO must be tailored in advance to be able to work in specific problem domains. This issue makes both approaches arguably less effective and prone to suboptimal solutions; particularly, when the data quantity increases.

3. Methodological approach

To overcome the aforementioned limitations, an ENN algorithm is proposed in this paper, in which the search of the optimal solutions relies on the optimisation of the neural network architecture. Moreover, a popular dimensionality reduction technique—t-distributed stochastic neighbour embedding (t-SNE)—is used to map the training and testing datasets. Then, the t-SNE technique is re-employed to recover the structure of the trained data and to demonstrate the potential of the proposed evolutionary algorithm. Reliability and consistency of the database will be ensured in this coupled process. The overall methodological approach is illustrated in Fig. 3.

3.1. Proposed ENN model

In general, the optimisation of neural networks comprises two main aspects: (i) the efficiency of computations, which involves the good convergence of the optimisation strategy with respect to potential variabilities in the implementation conditions; and (ii) the accuracy of the model, which is the ability to achieve a good prediction performance [55–57]. Satisfying these two facets simultaneously is challenging for many existing approaches [29–33,44].

In the proposed model, the optimal amount of regularisation is a

trade-off between the model simplicity and capacity, which is controlled through a set of defined hyper-parameters. A differential evolution technique is used to define the neural network architecture and to optimise the hyper-parameters in lieu of biases and weights. Moreover, a stochastic optimisation method is adopted to ensure an adaptive learning rate of the neural network and to reduce the computation time. Thus, the proposed ENN model aims to increase the efficiency and reliability of computations by optimising the convergence speed and structural design of the neural network during the learning process.

3.1.1. Parametrisation

To solve the overfitting problem, a '*regulariser*' is added to Eq. (2). Consequently, the evolved expressions of the loss L and weights w^{ji} can be written as in Eq. (5) and (6).

$$L(\mathbf{w},b)=\frac{1}{M}\sum_{n=1}^{M}\log(\cosh(\widehat{k}_n-k_n))+\underbrace{\psi(\theta)}_{regularizer} \tag{3}$$

Because the model should remain sufficiently stable whenever the data distribution changes, a '$L2$ regularisation' approach is adopted to favour hypotheses with which small norms of weights are obtained. The 'L^2 regularisation' is defined as follows:

$$\psi_{L^2}(w)=\frac{\lambda}{2M}\sum_{n=1}^{M}\left\|w_l^{ji}\right\|^2. \tag{4}$$

Thus,

$$\tilde{L}(\mathbf{w},b)=L(\mathbf{w},b)+\frac{\lambda}{2M}\sum\left\|w_l^{ji}\right\|^2, \tag{5}$$

$$w_l^{ji}=\left(1-\frac{\lambda\alpha}{M}\right)w_l^{ji}-\alpha\frac{\partial L}{\partial w_l^{ji}}, \tag{6}$$

where λ denotes a hyper-parameter that defines the regularisation degree, and M is the sample size. It should be noted that the regularisation quantity includes no bias parameters, which ensures consistency between the initial and updated values of the biases. In addition, this regularisation approach promotes an efficient generalisation ability of the neural network.

3.1.2. Differential evolution approach

In this study, the differential evolution method suggested by Storn and Price [58] is adopted. It is a heuristic technique applied for minimising possibly non-differentiable and nonlinear continuous space functions. This paradigm is applied in this study to determine the best hyper-parameters (*i.e.* epoch size and regularisation parameter) as well as the optimal neural network architecture (*i.e.* number of neurons in a hidden layer and number of hidden layers). The method assumes that the population for each generation g consists of NP D-dimensional parameter vectors, which can be written as

$$\mathbf{x}_{i,k}, \qquad i=1,2,...,NP, \qquad k=1,2,...,K, \tag{7}$$

where K is the total number of generations, and NP remains invariant throughout the minimisation process. The initial set is distributed over the entire parameter space. Furthermore, the DEA is applied to determine the optimal individual $\mathbf{x}$. The four main parameters representing an individual $\mathbf{x}$ are: the regularisation parameter $\boldsymbol{\lambda}$, epoch size **EP**, number of neurons in a hidden layer $\mathbf{m}$, and number of hidden layers $\mathbf{n}$. More importantly, three consecutive steps are vital for the successful implementation of this method: the mutation, crossover, and selection processes.

3.1.2.1. Mutation.

For each population defined by Eq. (5), the mutation operation is defined by

$$\mathbf{v}_{i,k}=\mathbf{x}_{\mathbf{r1},k}+F(\mathbf{x}_{\mathbf{r2},k}-\mathbf{x}_{\mathbf{r3},k}), \tag{8}$$

where $\mathbf{v}_{i,k}$ is the mutant vector; r_1, r_2, and $r_3 \in \{1,2,...,NP\}$ are distinct and randomly selected integers other than the index i; F is a constant factor ($F\in[0, 2]$) that guarantees the optimal scaling of the differential variation $(\mathbf{x}_{\mathbf{r2},k}-\mathbf{x}_{\mathbf{r3},k})$.

3.1.2.2. Crossover.

The crossover operation is implemented to enhance the diversification of the parameter vectors obtained from the mutation process. The trial vector $\mathbf{u}_{i,k}$ can be defined as

$$u_{i,k}^{j}=\begin{cases} v_{i,k}^{j}, & \text{if } (\text{rand}(0,1)\le \text{CR}) \text{ or } (j=j_{rand}) \\ x_{i,k}^{j}, & \text{otherwise} \end{cases}, \tag{9}$$

where $u_{i,k}^{j}$ is the jth evaluation of $\mathbf{u}_{i,k}$; j_{rand} is an randomly chosen integer within the range $[1,D]$; CR is the crossover rate ($CR\in[0, 1]$) that governs the fraction of individual components calculated from the mutant vector.

3.1.2.3. Selection.

The trial vector $\mathbf{u}_{i,k}$ is weighed against the objective vector $x_{i,k}$ to decide whether or not it should be incorporated in the next generation $g+1$. To this end, a greedy criterion is adopted as follows:

$$\mathbf{x}_{i,k+1}=\begin{cases} \mathbf{u}_{i,k}, & \text{if } f(\mathbf{u}_{i,k})>f(\mathbf{x}_{i,k}) \\ \mathbf{x}_{i,k}, & \text{otherwise} \end{cases}, \tag{10}$$

where f is the objective function and $\mathbf{x}_{i,k+1}$ the individual vector of generation $k+1$.

3.1.3. Stochastic gradient optimisation algorithm

It is acknowledged that the efficient training of a neural network requires significant computing resources; particularly, when a significant quantity of data is concerned. The hyper-parameters and learning optimising are the typical 'meta-problems' that hamper the training efficiency of neural networks. Moreover, the learning rate is of pivotal importance for the model performance. In general, at a low learning rate, the training process is more reliable, but the optimisation tends to be time-consuming. By contrast, at a high learning rate, there is a high probability that the training does not converge. According to the literature, learning rate (or step size) is one of the most arduous parameters to tailor throughout a neural network training process [59,60]. Several stochastic gradient-based optimisation techniques have been developed to alleviate this issue, including *Adam* [61], *Adagrad* [62], and *RMSprop*

Table 1
Algorithm of stochastic gradient-based optimisation method (*Adamax*).

Algorithm *Adamax*. Please note: satisfactorily settings for machine learning problems are: $\alpha=0.002$, $\omega_1=0.9$, $\omega_2=0.999$, and $\delta=10^{-8}$; ω_1^t denotes the parameter ω_1 with respect to power t; the learning rate with the bias correction term for the first moment is $\alpha/(1-\omega_1^t)$.

Required: α: step size
Required: δ: small constant used for numerical stabilisation
Required: $\omega_1, \omega_2 \in [0,1]$: exponential decay rates for moment estimates
Required: $\boldsymbol{\theta}$: initial parameter vector ($\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{w},\mathbf{b})$, $\mathbf{w}$: weights, $\mathbf{b}$: biases)
  $s_0 \leftarrow 0$ (Initialisation of first moment vector)
  $r_0 \leftarrow 0$ (Initialisation of exponentially weighted infinity norm)
  $t \leftarrow 0$ (Initialisation of time step)
  While $\boldsymbol{\theta}_t$ is not converging, execute:
    $t \leftarrow t+1 \mathbf{g}_t \leftarrow \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_{t-1})$ (Obtaining gradients)
    $\mathbf{s}_t \leftarrow \omega_1 \mathbf{s}_{t-1}+(1-\omega_1)\cdot \mathbf{g}_t$ (Updating biased first moment estimate)
    $\mathbf{r}_t \leftarrow \max(\omega_2 \cdot u_{t-1}, \ |\mathbf{g}_t|)$ (Updating exponentially weighted infinity norm)
    $\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t-1}-(\alpha/(1-\omega_1^t))\cdot \mathbf{m}_t/u_t$ (Updating parameters)
  End while
Return $\boldsymbol{\theta}$ (resulting parameters)

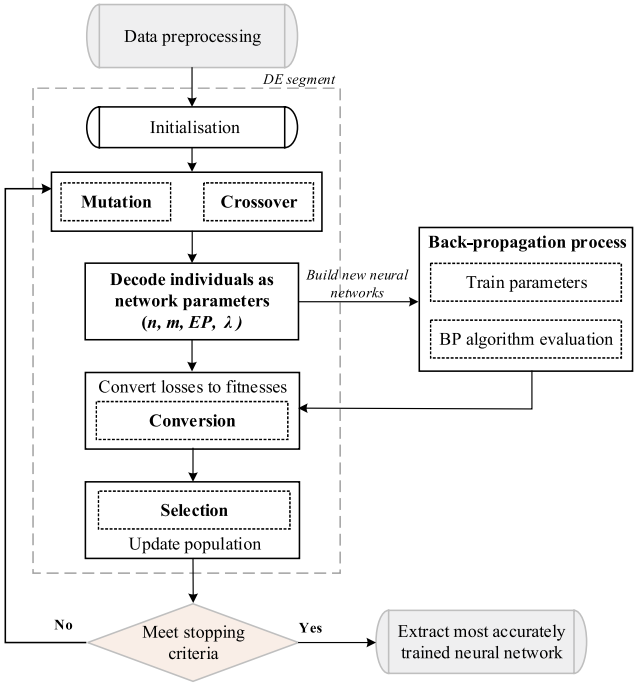

Fig. 4. Flowchart of proposed evolutionary neural network (ENN) model.

Table 2
Pseudo-code of proposed evolutionary neural network (ENN) model.

Step 1: Preprocess data and initialise network
Step 2: Perform DE operations: mutation and crossover
Step 3: Transfer hyper-parameters to BP algorithm ***n*** (number of hidden layers), ***m*** (number of neurons in hidden layer), ***EP*** (epoch size), and λ (regularisation parameter)
Step 4: Convert losses of networks to fitness
Step 5: Carry out selection operation (update the population)
Step 6: Check stopping criteria
Step 7: Export optimal hyper-parameters (best neural network)
Step 8: End

[63]. These algorithms are decisive for improving the convergence performance of a model, without penalising the search process. In this study, the authors adopt the optimisation algorithm *Adamax*, which is a variation of *Adam* established based on the infinity norm [61]. This adaptive learning rate approach is suited for problems dealing with a large database like a soil liquefaction analysis. *Adamax* is a hybrid method that integrates the advantages of *Adagrad* and *RMSprop*. The basic form of the *Adamax* algorithm is presented in Table 1.

3.1.4. Implementation procedure of proposed model

The flow chart (see Fig. 4) and pseudo-code (see Table 2) of the proposed ENN model are presented here. The relevant data are initially fed into the network. Then, for each individual, the DE operations are carried out to define the generation hyper-parameters, including **n** (number of hidden layers), **m** (number of neurons in a hidden layer), **EP** (epoch size), and λ (regularisation parameter). The neural network is built with the previously obtained hyper-parameters. Then, the model is trained and run to see whether it satisfies the stopping criteria (convergence criterion and/or maximal number of generations). If the result is flagged as 'No', the 'DE segment' is repeated. If it is 'Yes', the most accurately trained neural network is extracted.

3.2. t-Stochastic neighbour embedding (t-SNE)

3.2.1. Backgrounds

The t-SNE technique developed by Maaten and Hinton [64] has a strong mapping ability; hence, it has become prevalent in the field of machine learning. It is an enhanced version of the SNE method [65]. The difference between t-SNE and SNE can be captured through a comparison of the gradients, as shown in Fig. 5. In Fig. 5, negative gradients represent repulsion between two points, whereas positive values represent an attraction among low-dimensional datapoints. The repulsion effect of dissimilar datapoints of SNE (in comparison with that of t-SNE) is minimal in the low-dimensional representation, whereas a stronger attraction is noticeable elsewhere in the gradients. Researchers have confirmed that the t-SNE is one of the best-performing dimensionality reduction techniques, and it has been successfully applied to many real-world datasets including imagery studies [66], genetics [67, 68], geology [69], and materials science [70]. The prominent idea behind t-SNE is to embed high-dimensional data in low dimensions to preserve the likenesses between points. Thus, points that are far away in a high-dimensional space correspond to distant points in the low-dimensional space. In addition, the points that are close to each other in the high-dimensional space are consistent with the nearby embedded low-dimensional points. Moreover, as underlined by Maaten and Hinton [64], the configuration of an optimal solution differs slightly from one run to another. This type of difference is common when using t-SNE, and will not significantly affect the evaluation results.

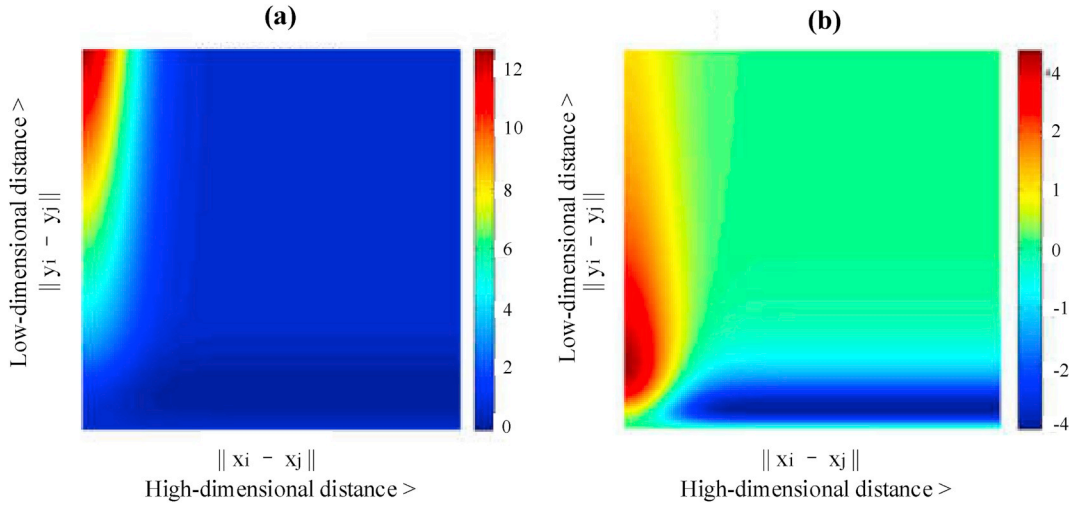

Fig. 5. Comparison of gradients of loss function of: a) SNE and b) t-SNE (after [64]).

Table 3
Range of input variables used in liquefaction analysis.

No	Input variables (units)	Min. value	Max. value	Comments
1	Earthquake magnitude $\mathbf{M}_w$	5.9	9	Characterise intensity of ground shaking
2	Maximal ground surface acceleration $\mathbf{a}_{max}$(g)	0.09	0.84	
3	Depth **d** (m)	1.4	11.8	Sampling depth for CPT test
4	Water depth $\mathbf{d}_w$ (m)	0.2	7.2	Determine effect of water depth
5	Total overburden stress $\boldsymbol{\sigma}_v$ (kPa)	24	210	Related to susceptibility of soils to liquefaction, which increases with increasing overburden pressures [24]
6	Effective overburden stress $\boldsymbol{\sigma}_v^{'}$ (kPa)	19	147	
7	CPT cone tip resistance $\mathbf{q}_c$(MPa)	0.94	45	Describes liquefaction resistance of soil by considering its theoretical correlation to undrained shear strength
8	CPT friction ratio $\mathbf{R}_f$	0.03	2.91	Classifies soil by its behaviour (low friction ratios indicate sandy soils; high friction ratios represent clayey soils)
9	Fines content **FC** (%)	0	85	Represents how fines affects soil behaviour (usually sand)
10	Shear stress ratio $\boldsymbol{\tau}_{av}/\boldsymbol{\sigma}_v^{'}$	0.071	0.695	Represents severity or level of earthquake loading

3.2.2. Mathematical framework

A set of high-dimensional points $X = \{x_1, \ x_2, \ldots, \ x_n\}$ ($x \in R^p$ and $p > 3$) is assumed, which should be mapped into a low-dimensional space $Y = \{y_1, \ y_2, \ldots, \ y_n\}$ ($y \in R^2$). The t-SNE method first computes the similarity of datapoints x_i and x_j, which is denoted by $p_{i|j}$. The parameter $p_{i|j}$ is the conditional probability that x_i picks x_j as neighbour in the case that neighbours are picked in proportion to a Gaussian density centred at x_i, as shown in Eq. (9).

$$\boldsymbol{p}_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / \ 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / \ 2\sigma_i^2\right)} \tag{11}$$

where σ_i denotes the variance of the Gaussian function, which is centred at the data point x_i. This value is set by ensuring that the effective number of local neighbours of each observation (termed 'perplexity') remains optimal.

The similarity is defined as a symmetrised version of the conditional similarity following Eq. (10):

$$\boldsymbol{p}_{i,j} = \frac{\boldsymbol{p}_{j|i} + \boldsymbol{p}_{i|j}}{2N} \tag{12}$$

For any $y_i, y_j \in R^2$, the previous definition is extended for the mapping into a lower dimension:

$$\boldsymbol{q}_{i,j} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k \neq i} \left(1 + \left\|y_k - y_j\right\|^2\right)^{-1}} \tag{13}$$

Subsequently, the t-SNE algorithm applies a heavy-tailed distribution to the embedded low-dimensional data points to overcome the issue of crowding, *i.e.* to promote the displaying of dissimilar points far apart in the map. Furthermore, Kullback–Leibler divergences between Q and P are calculated with the gradient descent method (Eq. (14)). More details can be found in Ref. [64].

$$KL(P||Q) = \sum_{i \neq j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \tag{14}$$

Table 4
Parameters adopted for simulating the proposed approach.

	Parameters	Values
differential evolution algorithm	Population size	40
	Number of generations	50
Neural network	Number of hidden layers (**n**)	[1, 4]
	Number of neurons in a hidden layer (**m**)	[3, 30]
	Regularisation parameter (λ)	[0.0, 0.4]
	Epoch size (**EP**)	[500, 2000]
t-SNE	Exaggeration	100
	NumDimensions	2
	NumPCAComponents	10
	Perplexity	15
	LearnRate	500

4. Case study: liquefaction potential assessment

4.1. Soil liquefaction database

The field records are prominently important for validating the DM models. However, field records of soil liquefactions are very rarely available in the literature. Nevertheless, to illustrate the robustness of the proposed model, a larger quantity (compared with those of the majority of studies that adopted DM techniques to predict liquefaction [37,44–46]) of field records is gathered in this study. The database exploited in this study includes case records originally compiled by Boulanger et al. [71] based on the field performances of 219 sites obtained from various earthquakes in the US, Japan, Turkey, China, and New Zealand. This database consists of 253 CPT field records, among which 181 are liquefaction cases and 72 non-liquefaction cases. These data are randomly divided into a training set of 200 cases and a testing set of 53 cases. The following 10 variables (see Table 3) are considered in the proposed model to describe the site liquefaction susceptibility, dynamic loading conditions, geological profile, and bearing capacity of the soil. The variables reflect the nonlinearity of the soil behaviour [72–74] and the characteristics of the seismic power dissipation.

A comprehensive summary of the data set used in this study is provided in the '*Data-in-brief paper*' linked to this paper. It should be noted that there is a negligible degree of discrepancy in the data owing to their variability and the updating from one source to another. Nevertheless, in addition to the high predictability, the proposed ENN model is designed to be flexible and consider newly available data [75,76].

4.2. Simulation setup

In this simulation, the *log-cosh loss* function (Eq. (4)) was adopted as objective function throughout the evolutionary process, owing to its good smoothing ability. Basically, this approach minimizes the margin between predicted and true values. Furthermore, Table 4 show a synoptic of the initial parameters adopted for simulating the proposed evolutionary algorithm as well as that used for the t-SNE approach. Specifically for t-SNE, the MATLAB implementation framework [77] was espoused in this study and the parameters tuning was conducted according to Ref. [78].

4.3. Results and discussion

In this section, the results obtained with the proposed coupled ENN/t-SNE method for the prediction of liquefaction field data are presented and discussed. In particular, the performance (accuracy and efficiency) of the proposed crossbred 'differential evolution–stochastic optimisation' model is assessed with soil liquefaction data. Subsequently, the dimensionality reduction ability of t-SNE is evaluated; in particular, its

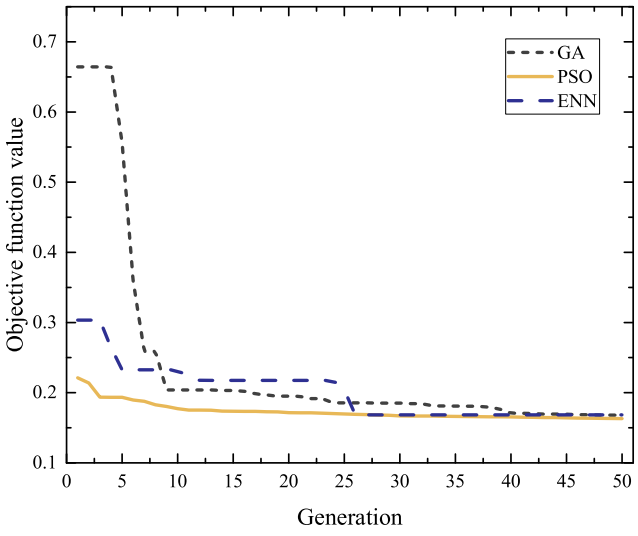

Fig. 6. Objective function value versus generation.

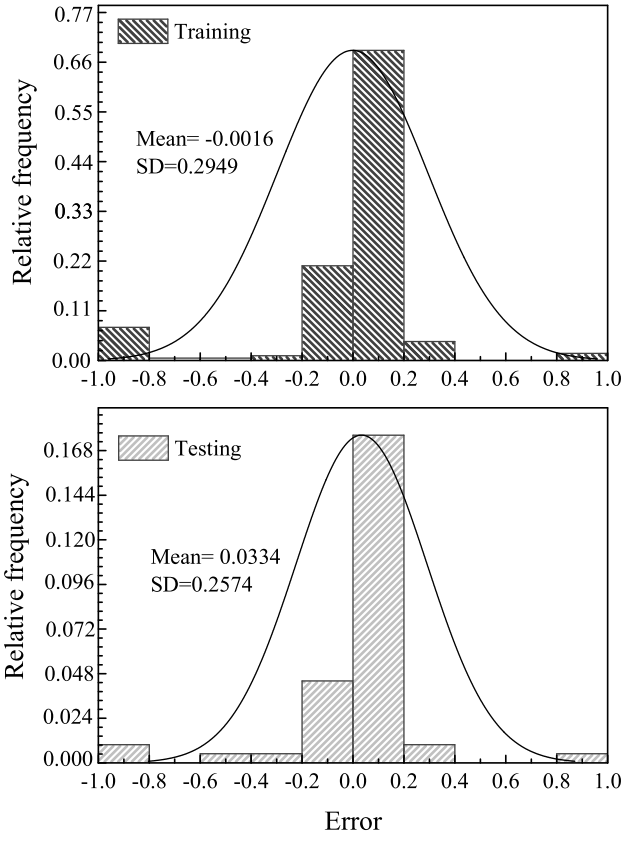

Fig. 7. Error histogram of trained and tested data.

capacity to recover the initial structure of the initial high-dimensional dataset.

4.3.1. *Performance of proposed model*

The optimum neural ENN architecture was obtained straightaway by picking the best individual out of the different generations. Apart from some few changes of the population size, the proposed approach was found rather stable with regard to the number of runs. This is certainly

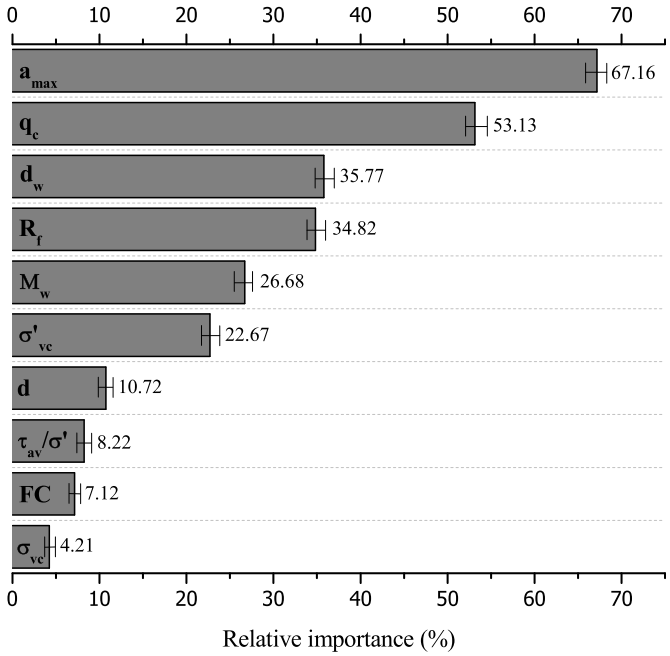

Fig. 8. Relative importance of input variables for occurrence of liquefaction.

one of the main advantage of the proposed approach. In particular, the optimum hyper-parameters resulting from our simulation were found to be: epoch size **EP** = 1410, regularisation parameter λ = 0.20928, number of hidden layers **n** = 4, and number of neurons in hidden layer **m** = 3. Subsequently, the optimum parameters of ENN were used to trained PSO and GA-optimised models. Indeed, as can be seen in Fig. 6 below, the evolutionary process curve of the proposed model has been compared to that of PSO and GA-optimised models. It is observed that the three algorithms converge suitably and can reach a satisfactory solution after a relatively early iteration stage. Although the PSO presents a slightly better capability to minimize the cost function, the ENN could achieve the optimum convergence faster than the two other approaches. This demonstrates the efficiency of the proposed ENN in optimising the hyper-parameters and providing a viable neural network architecture. No overfitting was experienced during the learning process owing to the applied regularisation strategy in the neural network.

The statistical distributions of the error values obtained from the training and testing phases are presented in Fig. 7. The error is determined as the difference between the measured and predicted values (absolute error). In the training and testing phases, the maximal frequency is obtained for relatively small error values varying from 0 to 0.2. Thus, the majority of the trained and tested outputs are distributed around the zero-error line. Moreover, the relative percentage errors are rather centralised, which proves the good generalisation ability of the proposed model. The fitting curve of the relative percentage errors of the training and testing phases exhibit a typical Gaussian distribution; the standard deviation of the testing set (0.2574) is slightly smaller than that of the training set (0.2949). Furthermore, prediction performances of approximately 97% and 94% are obtained for the testing and training phases, respectively. The detailed results are listed in the tables provided in the '*Data-in-brief paper*' linked to this article.

4.3.2. *Effect of liquefaction parameters on occurrence of soil liquefaction*

Fig. 8 presents the sensitivity analysis results to provide a detailed interpretation of the proposed model. The sensitivity of the proposed ENN model is expressed as the first-order partial derivative of the objective output (liquefaction occurrence) over the input variables (liquefaction parameters) [79,80]. This procedure allows for a quantification of the importance level of each liquefaction parameter.

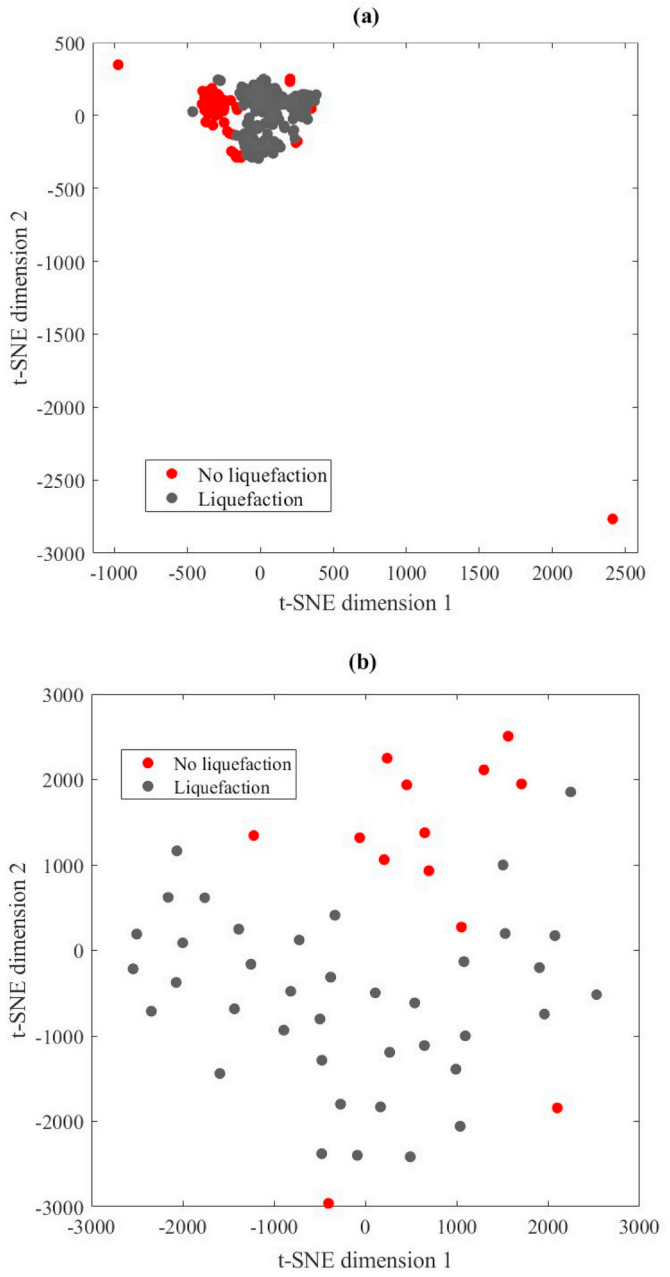

Fig. 9. Field observation data of liquefaction used for: a) training and b) testing.

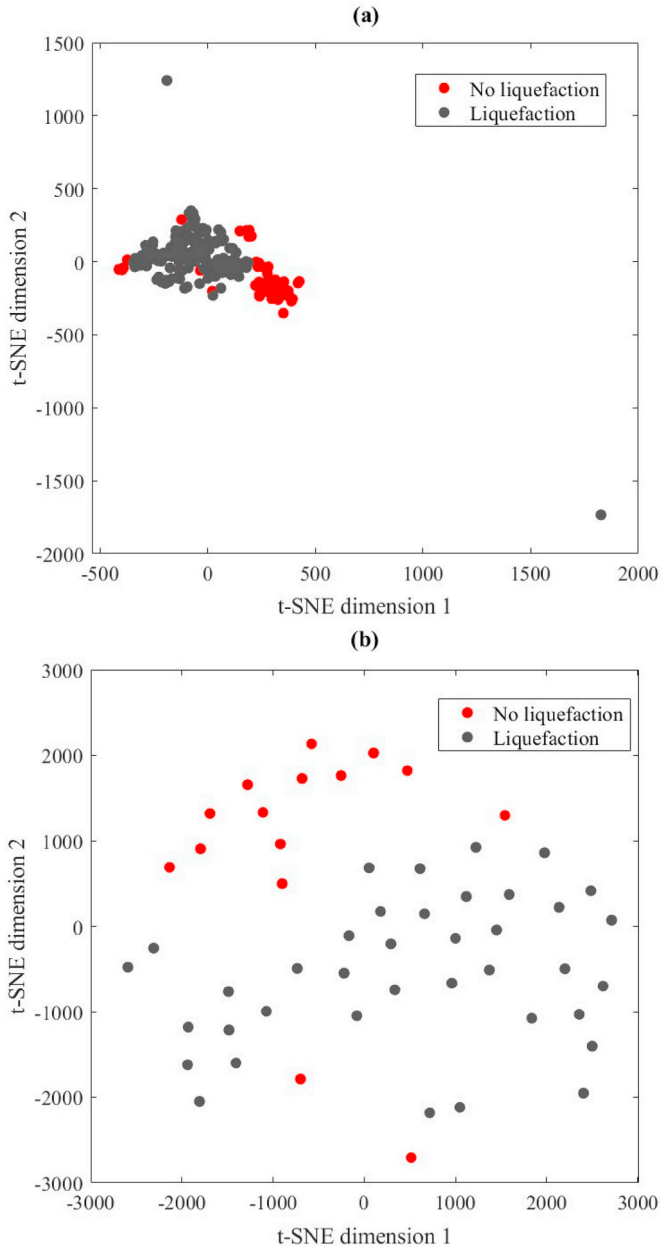

Fig. 10. Proportion of predicted values of ENN model in: a) training phase, b) testing phase.

According to the results, the first key parameter responsible for the soil liquefaction is the maximal acceleration $\mathbf{a_{max}}$ at the ground surface, which essentially characterises earthquake severity. The second key parameter is the CPT cone tip resistance q_c, which typically reflects soil behaviour. Among the five most prominent variables that govern the occurrence of soil liquefaction, two variables are found to be related to the intensity of ground shaking ($\mathbf{a_{max}}$, $\mathbf{M_w}$), and the others are linked to soil behaviour ($\mathbf{q_c}$, $\mathbf{d_w}$, $\mathbf{R_f}$). These results are generally consistent with the deterministic triggering correlation proposed by Boulanger and Idriss [71] for the same case data.

4.3.3. Visualising the liquefaction potential and ENN model performance

The t-SNE approach is employed to reduce the high-dimensional liquefaction data to two-dimensional data before and after the process with the proposed ENN. These values are provided in the linked 'Data-in-brief paper'. Following the description provided in Section 3.2.2, the low-dimensional space in this case corresponds to the liquefaction occurrence. Fig. 9 depicts the trend of the liquefaction occurrence based on field observations. These data are divided into training and testing sets. In spite of the relatively high dimensionality, the t-SNE-based visualisation enables recovering the structure of the initial conditions. Apart from a few exceptions, all cases in which liquefaction was observed cluster together owing to their equal attributes. This observation is also true for non-liquefaction cases. The t-SNE approach visualises (within a reasonable runtime) high-dimensional points by allocating each data point to a position in a two-dimensional space.

Fig. 10 shows the proportions of predicted values of the liquefaction occurrence assessed by the proposed ENN. The results verify the ability of the t-SNE approach to recover the initial structure. A comparison between Figs. 9 and 10 further proves the potential of the proposed ENN. The clusters in Figs. 9a and 10a present quasi-identical densities. Besides, the quantity of misclassifications is consistent with the results listed in the result table (see '*Data-in-brief paper*'). The mapping ability of

t-SNE confirms the good performance of the proposed model during the training and testing phases.

5. Conclusions

This study presents a coupled approach that combines a new ENN algorithm and the dimensionality reduction technique t-SNE. The proposed approach predicts the soil liquefaction potential and visualises the prediction results. The following conclusions are drawn:

(1) Based on the evaluation of the loss values with respect to the number of generations, it was observed that the proposed model converges very satisfactorily, which confirms the high efficiency of this 'crossbreed' model in optimising the hyper-parameters and neural network architecture.
(2) The statistics of the error values exhibit a centralised relative percentage error, which demonstrates the good generalisation ability of the proposed model.
(3) Based on the nonlinear dimensionality reduction method t-SNE, relevant large datasets were visualised before and after training with the proposed model. The prominent and practical abilities of t-SNE to recover the structure of the initial conditions and to demonstrate the performance of the proposed model were observed.
(4) The quantification of the importance levels between liquefaction variables revealed that the maximal acceleration $\mathbf{a}_{\mathbf{max}}$ at the ground surface (which essentially characterises the severity of an earthquake) is the first key variable responsible for soil liquefaction.

Author statement

The present statement is to certify that all authors have seen and approved the final version of the manuscript being resubmitted. They warrant that the article is the authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research work described herein was funded by the Research Funding of Shantou University for New Faculty Member (Grant No. NTF19024-2019), and the National Nature Science Foundation of China (NSFC) (Grant No. 41372283). The source of financial supports is gratefully acknowledged.

References

[1] Stewart JP, Chu DB, Seed RB, Ju JW, Perkins WJ, Boulanger RW, Chen YC, Ou CY, Sun J, Yu MS. Incidents of soil liquefaction from the 921 Chi Chi (Taiwan) earthquake. In: Loh CH, Liao WI, editors. Proceedings, International Workshop on Annual Commemoration of Chi-Chi Earthquake, III. Taipei, Taiwan: National Center for Research on Earthquake Engineering; 2000. p. 119–30 (Also appeared in Uzarski J, Arnold C, editors. Chi-Chi, Taiwan, earthquake of September 21, 1999 reconnaissance report. Earthquake spectra 2001, vol. 17, Supplement A, Chapter 4. Oakland, CA: Earthquake Engineering Research Institute.).
[2] Boulanger RW, Idriss IM. Soil liquefaction during earthquake. Oakland, CA: Engineering monograph, Earthquake Engineering Research Institute; 2008. p. 266.
[3] Rahman MA, Arulrajah A, Piratheepan J, Bo MW, Imteaz MA. Resilient modulus and permanent deformation responses of geogrid-reinforced construction and demolition materials. J Mater Civ Eng 2014;26(3):512–9. https://doi.org/10.1061/(asce)mt.1943-5533.0000824.
[4] Arulrajah A, Yaghoubi E, Wong YC, Horpibulsuk S. Recycled plastic granules and demolition wastes as construction materials: resilient moduli and strength characteristics. Constr Build Mater 2017;147:639–47. https://doi.org/10.1016/j.conbuildmat.2017.04.178.
[5] Arulrajah A, Piratheepan J, Disfani MM, Bo MW. Resilient moduli response of recycled construction and demolition materials in pavement subbase applications. J Mater Civ Eng 2013;25(12):1920–8. https://doi.org/10.1061/(asce)mt.1943-5533.0000766.
[6] Haghighi H, Arulrajah A, Mohammadinia A, Horpibulsuk S. A new approach for determining resilient moduli of marginal pavement base materials using the staged repeated load CBR test method. Road Mater Pavement Des 2018;19(8):1848–67. https://doi.org/10.1080/14680629.2017.1352532.
[7] Disfani MM, Arulrajah A, Haghighi H, Mohammadinia A, Horpibulsuk S. Flexural beam fatigue strength evaluation of crushed brick as a supplementary material in cement stabilized recycled concrete aggregates. Constr Build Mater 2014;68:667–76. https://doi.org/10.1016/j.conbuildmat.2014.07.007.
[8] Arulrajah A, Disfani MM, Haghighi H, Mohammadinia A, Horpibulsuk S. Modulus of rupture evaluation of cement stabilized recycled glass/recycled concrete aggregate blends. Constr Build Mater 2015;84:146–55. https://doi.org/10.1016/J.CONBUILDMAT.2015.03.048.
[9] Marano KD, Wald DJ, Allen TI. Global earthquake casualties due to secondary effects: a quantitative analysis for improving rapid loss analyses. Nat Hazards 2010;52:319–28. https://doi.org/10.1007/s11069-009-9372-5.
[10] Chian SC, Tokimatsu K, Madabhushi SPG. Soil liquefaction-induced uplift of underground structures: physical and numerical modeling. J Geotech Geoenviron Eng 2014;140(10):04014057. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001159.
[11] Priestley MJN, Seible FG, Calvi GM. Seismic design and retrofit of bridges. John Wiley & Sons, Ltd.; 1996. p. 679. https://doi.org/10.1002/9780470172858.
[12] Bird JF, Bommer JJ, Crowley H, Pinho R. Modelling liquefaction-induced building damage in earthquake loss estimation. Soil Dyn Earthq Eng 2006;26(1):15–30. https://doi.org/10.1016/j.soildyn.2005.10.002.
[13] Grasso S, Maugeri M. The road map for seismic risk analysis in a Mediterranean city. Soil Dyn Earthq Eng 2009;29(6):1034–45. https://doi.org/10.1016/j.soildyn.2008.12.001.
[14] Ren D-J, Shen S-L, Arulrajah A, Cheng W-C. Prediction model of TBM disc cutter wear during tunnelling in heterogeneous ground. Rock Mech Rock Eng 2018;51:3599–611. https://doi.org/10.1007/s00603-018-1549-3.
[15] Ren D-J, Shen S-L, Arulrajah A, Wu H-N. Evaluation of ground loss ratio with moving trajectories induced in double-O-tube (DOT) tunnelling. Can Geotech J 2018;55:894–902. https://doi.org/10.1139/cgj-2017-0355.
[16] Shen S-L, Wu Y-X, Xu Y-S, Hino T, Wu H-N. Evaluation of hydraulic parameters from pumping tests in multi-aquifers with vertical leakage in Tianjin. Comput Geotech 2015;68:196–207. https://doi.org/10.1016/J.COMPGEO.2015.03.011.
[17] Shen S-L, Wu Y-X, Misra A. Calculation of head difference at two sides of a cut-off barrier during excavation dewatering. Comput Geotech 2017;91:192–202. https://doi.org/10.1016/J.COMPGEO.2017.07.014.
[18] Tan Y, Lu Y, Wang D. Deep excavation of the gate of the orient in Suzhou stiff clay: composite earth-retaining systems and dewatering plans. J Geotech Geoenviron Eng 2018;144:05017009.
[19] Tan Y, Wei B, Lu Y, Yang B. Is basal reinforcement essential for long and narrow subway excavation bottoming out in shanghai soft clay? J Geotech Geoenviron Eng 2019;145:05019002.
[20] Tan Y, Lu Y. Forensic diagnosis of a leaking accident during excavation. J Perform Constr Facil 2017;31:04017061.
[21] Xu YS, Yan XX, Shen SL, Zhou AN. Experimental investigation on the blocking of groundwater seepage from a waterproof curtain during pumped dewatering in an excavation. Hydrogeol J 2019;27(7):2659–72. https://doi.org/10.1007/s10040-019-01992-3.
[22] Xu Y-S, Shen S-L, Lai Y, Zhou A-N. Design of sponge city: lessons learnt from an ancient drainage system in Ganzhou, China. J Hydrol 2018;563:900–8. https://doi.org/10.1016/J.JHYDROL.2018.06.075.
[23] Idriss IM, Boulanger RW. Semi-empirical procedures for evaluating liquefaction potential during earthquakes. Soil Dyn Earthq Eng 2006;26(2–4):115–30. https://doi.org/10.1016/j.soildyn.2004.11.023.
[24] Seed HB, De Alba P. Use of SPT and CPT tests for evaluating the liquefaction resistance of sands. In: Bolton Seed H, De Alba Pedro, editors. Useof In Situ Tests inGeotechnical Engineering, 6. ASCE, Geotechnical Special Publication (GSP); 1986. p. 281–302.
[25] Kokusho T. Liquefaction potential evaluations: energy-based method versus stress-based method. Can Geotech J 2013;50:1088–99. https://doi.org/10.1139/cgj-2012-0456.
[26] Ishac MF, Heidebrecht AC. Energy dissipation and seismic liquefaction in sands. Earthq Eng Struct Dyn 1982;10:59–68. https://doi.org/10.1002/eqe.4290100105.
[27] Freeman JA, Skapura DM. Neural networks algorithms, applications and programming techniques. Addison-Wesley; 1991.
[28] Goh ATC. Predicting seismic liquefaction using neural networks. In: Third International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics, 1; April2-7, 1995. p. 261–4. St.Louis, Missouri.
[29] Hanna AM, Ural D, Saygili G. Evaluation of liquefaction potential of soil deposits using artificial neural networks. Eng Comput 2007;24:5–16. https://doi.org/10.1108/02644400710718547.
[30] Ramakrishnan D, Singh TN, Purwar N, Barde KS, Gulati A, Gupta S. Artificial neural network and liquefaction susceptibility assessment: a case study using the 2001 Bhuj earthquake data, Gujarat, India. Comput Geosci 2008;12:491–501. https://doi.org/10.1007/s10596-008-9088-8.

[31] Goh AT. Probabilistic neural network for evaluating seismic liquefaction potential. Can Geotech J 2002;39(1):219–32. https://doi.org/10.1139/t01-073.

[32] Kim YS, Kim BT. Use of artificial neural networks in the prediction of liquefaction resistance of sands. J Geotech Geoenviron Eng 2006;132(11):1502–4. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:11(1502).

[33] 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.

[34] Goh ATC, Goh SH. Support vector machines: their use in geotechnical engineering as illustrated using seismic liquefaction data. Comput Geotech 2007;34:410–21. https://doi.org/10.1016/j.compgeo.2007.06.001.

[35] Samui P, Sitharam TG. Machine learning modelling for predicting soil liquefaction susceptibility. Nat Hazards Earth Syst Sci 2011;11:1–9. https://doi.org/10.5194/nhess-11-1-2011.

[36] Samui P, Kim D, Sitharam TG. Support vector machine for evaluating seismic-liquefaction potential using shear wave velocity. J Appl Geophys 2011;73(1):8–15. https://doi.org/10.1016/j.jappgeo.2010.10.005.

[37] Xue X, Yang X. Seismic liquefaction potential assessed by support vector machines approaches. Bull Eng Geol Environ 2016;75:153–62. https://doi.org/10.1007/s10064-015-0741-x.

[38] Pal M. Support vector machines-based modelling of seismic liquefaction potential. Int J Numer Anal Methods Geomech 2006;30:983–96. https://doi.org/10.1002/nag.509.

[39] Samui P. Seismic liquefaction potential assessment by using Relevance Vector Machine. Earthq Eng Eng Vib 2007;6:331–6. https://doi.org/10.1007/s11803-007-0766-7.

[40] Cabalar AF, Cevik A, Gokceoglu C. Some applications of adaptive neuro-fuzzy inference system (ANFIS) in geotechnical engineering. Comput Geotech 2012;40:14–33. https://doi.org/10.1016/j.compgeo.2011.09.008.

[41] Xue X, Yang X. Application of the adaptive neuro-fuzzy inference system for prediction of soil liquefaction. Nat Hazards 2013;67:901–17. https://doi.org/10.1007/s11069-013-0615-0.

[42] Gandomi AH, Fridline MM, Roke DA. Decision tree approach for soil liquefaction assessment. Sci World J 2013:346285. https://doi.org/10.1155/2013/346285.

[43] Gori M, Tesi A. On the problem of local minima in backpropagation. IEEE Trans Pattern Anal Mach Intell 1992;14(1):76–86. https://doi.org/10.1109/34.107014.

[44] Xue X, Liu E. Seismic liquefaction potential assessed by neural networks. Environ. Earth Sci. 2017;76:1–15. https://doi.org/10.1007/s12665-017-6523-y.

[45] Rahbarzare A, Azadi M. Improving prediction of soil liquefaction using hybrid optimization algorithms and a fuzzy support vector machine. Bull Eng Geol Environ 2019;78(7):4977–87. https://doi.org/10.1007/s10064-018-01445-3.

[46] Xue X, Xiao M. Application of genetic algorithm-based support vector machines for prediction of soil liquefaction. Environ. Earth Sci. 2016;75:874. https://doi.org/10.1007/s12665-016-5673-7.

[47] Lyu HM, Shen SL, Zhou AN, Zhou WH. Flood risk assessment of metro systems in a subsiding environment using the interval FAHP-FCA approach. Sustain. Cities Soc. 2019;50:101682. https://doi.org/10.1016/j.scs.2019.101682. 2019.

[48] Lyu H-M, Shen S-L, Zhou A, Yang J. Perspectives for flood risk assessment and management for mega-city metro system. Tunn Undergr Space Technol 2019;84:31–44. https://doi.org/10.1016/J.TUST.2018.10.019.

[49] Lyu HM, Sun WJ, Shen SL, Zhou AN. Risk assessment using a new consulting process in fuzzy AHP. J Constr Eng Manag 2019. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001757. ASCE. In press.

[50] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5(4):115–33. https://doi.org/10.1007/BF02478259.

[51] Kolen JF, Pollack JB. Back propagation is sensitive to initial conditions. In: Lippmann RP, Moody JE, Touretzky DS, editors. Advances in neural information processing systems 3. Morgan-Kaufmann; 1991. p. 860–7.

[52] Kennedy J, Eberhart R. Particle swarm optimization. Proc ICNN'95 - Int Conf Neural Netw 1995;4:1942–8. https://doi.org/10.1109/ICNN.1995.488968. 4.

[53] Booker LB, Goldberg DE, Holland JH. Classifier systems and genetic algorithms. Artif Intell 1989;40(1–3):235–82. https://doi.org/10.1016/0004-3702(89)90050-7.

[54] Holland JH, Chu EC, Beasley JE. Genetic Algorithms - computer programs that "evolve" in ways that resemble natural selection can solve complex problems even their creators do not fully understand. Sci Am 1992. https://doi.org/10.1137/0202009.

[55] Jin Y-F, Yin Z-Y, Zhou W-H, Huang H-W. Multi-objective optimization-based updating of predictions during excavation. Eng Appl Artif Intell 2019;78:102–23. https://doi.org/10.1016/J.ENGAPPAI.2018.11.002.

[56] Yin ZY, Jin YF, Shen JS, Hicher PY. Optimization techniques for identifying soil parameters in geotechnical engineering: comparative study and enhancement. Int J Numer Anal Methods Geomech 2018;42:70–94. https://doi.org/10.1002/nag.2714.

[57] Lyu HM, Shen SL, Yang J, Yin ZY. Inundation analysis of metro systems with the storm water management model incorporated into a geographical information system: a case study in Shanghai. Hydrol Earth Sys Sci 2018;23(10):4293–307. https://doi.org/10.5194/hess-23-4293-2019.

[58] Storn R, Price K. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 1997;11(4):341–59. https://doi.org/10.1023/A:1008202821328.

[59] Salimans T, Kingma DP. Weight Normalization : a simple reparameterization to accelerate training of deep neural networks. 30th conference on neural information processing systems. 2016. Barcelona, Spain.

[60] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016. https://doi.org/10.1038/nmeth.3707. Deep Learning - MIT. Nature 2016.

[61] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: 3rd international conference for learning representations, San Diego; 2015. 2015.

[62] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 2011;12:2121–59.

[63] Tieleman T, Hinton GE, Srivastava N, Swersky K. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Netw Mach Learn 2012;4:26–31.

[64] Van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008;9:2579–605.

[65] Bunte K, Haase S, Biehl M, Villmann T. Stochastic neighbor embedding (SNE) for dimension reduction and visualization using arbitrary divergences. Neurocomputing 2012;90:23–45. https://doi.org/10.1016/j.neucom.2012.02.034.

[66] Li M, Luo X, Yang J. Extracting the nonlinear features of motor imagery EEG using parametric t-SNE. Neurocomputing 2016;218:371–81. https://doi.org/10.1016/J.NEUCOM.2016.08.083.

[67] Li W, Cerise JE, Yang Y, Han H. Application of t-SNE to human genetic data. J Bioinform Comput Biol 2017:1750017. https://doi.org/10.1142/S0219720017500172.

[68] Xu W, Jiang X, Hu X, Li G. Visualization of genetic disease-phenotype similarities by multiple maps t-SNE with Laplacian regularization. BMC Med Genomics 2014;7(Suppl 2):S1. https://doi.org/10.1186/1755-8794-7-S2-S1.

[69] Balamurali M, Melkumyan A. t-SNE based visualisation and clustering of geological domain. In: Hirose A, Ozawa S, Doya K, Ikeda K, Lee M, Liu D, editors. Neural information processing. Cham: Springer International Publishing; 2016. p. 565–72.

[70] Wu D, Huang Y, Chen H, He Y, Chen S. VPPAW penetration monitoring based on fusion of visual and acoustic signals using t-SNE and DBN model. Mater Des 2017;123:1–14. https://doi.org/10.1016/J.MATDES.2017.03.033.

[71] Boulanger RW, Idriss IM. CPT and SPT based liquefaction triggering procedures. 2014. p. 134 (Report No. UCD/CGM-14/01).

[72] Wu H-N, Shen S-L, Yang J. Identification of tunnel settlement caused by land subsidence in soft deposit of Shanghai. J Perform Constr Facil 2017;31:04017092.

[73] Wu Y-X, Shen JS, Cheng W-C, Hino T. Semi-analytical solution to pumping test data with barrier, wellbore storage, and partial penetration effects. Eng Geol 2017;226:44–51. https://doi.org/10.1016/J.ENGGEO.2017.05.011.

[74] Xu YS, Wang XW, Yang TL, Shen SL. Evaluation of optimized depth of waterproof curtain to mitigate negative impacts during dewatering. J Hydrol 2019;577:123969. https://doi.org/10.1016/j.jhydrol.2019.123969.

[75] Elbaz K, Shen S-L, Zhou A, Yuan D-J, Xu Y-S. Optimization of EPB shield performance with adaptive neuro-fuzzy inference system and genetic algorithm. Appl Sci 2019;9. https://doi.org/10.3390/app9040780.

[76] Wu YX, Lyu HM, Han J, Shen SL. Dewatering-induced building settlement around a deep excavation in soft deposit in Tianjin, China. J Geotech Geoenviron Eng ASCE 2019;145(5):05019003. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002045. 2019.

[77] Maaten L Van Der, Hinton G. User's guide for t-SNE software 2012:1–4. https://lvdmaaten.github.io/tsne/User_guide (Accessed on February 11th, 2019).

[78] Wattenberg M. How to use t-SNE effectively. Www.Distill.Pub; 2016 (Accessed on February 11th, 2019).

[79] Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process Lett 1995;2:1–4. https://doi.org/10.1007/BF02309007.

[80] Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 2003;160:249–64. https://doi.org/10.1016/S0304-3800(02)00257-0.