Geomechanics and Engineering, *Vol. 18, No. 5 (2019) 493-502*
DOI: https://doi.org/10.12989/gae.2019.18.5.493

Application of a support vector machine for prediction of piping and internal stability of soils

Xinhua Xue*

State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, No.24 South Section 1, Yihuan Road, Chengdu, 610065, P.R. China

(Received April 13, 2018, Revised July 11, 2019, Accepted July 18, 2019)

Abstract. Internal stability is an important safety issue for levees, embankments, and other earthen structures. Since a large part of the world's population lives near oceans, lakes and rivers, floods resulting from breaching of dams can lead to devastating disasters with tremendous loss of life and property, especially in densely populated areas. There are some main factors that affect the internal stability of dams, levees and other earthen structures, such as the erodibility of the soil, the water velocity inside the soil mass and the geometry of the earthen structure, etc. Thus, the mechanism of internal erosion and stability of soils is very complicated and it is vital to investigate the assessment methods of internal stability of soils in embankment dams and their foundations. This paper presents an improved support vector machine (SVM) model to predict the internal stability of soils. The grid search algorithm (GSA) is employed to find the optimal parameters of SVM firstly, and then the cross – validation (CV) method is employed to estimate the classification accuracy of the GSA-SVM model. Two examples of internal stability of soils are presented to validate the predictive capability of the proposed GSA-SVM model. In addition to verify the effectiveness of the proposed GSA-SVM model, the predictions from the proposed GSA-SVM model were compared with those from the traditional back propagation neural network (BPNN) model. The results showed that the proposed GSA-SVM model is a feasible and efficient tool for assessing the internal stability of soils with high accuracy.

Keywords: internal stability; piping; support vector machine; grid search algorithm; cross-validation

1. Introduction

Seepage-induced internal erosion presents a significant threat to the stability of levees, embankments, and other earthen structures (Chang and Zhang 2013). The case histories of earthen dam failures in the nineteenth and twentieth centuries have documented the loss of human lives and damages to properties. For instance, the 1998 Great Flood in the Yangtze River basin in China is considered to be one of the most devastating floods in the Chinese history. The peak flow rate of flood, volume of flow, and duration were all record-breaking. About 5 million hectares of croplands were inundated, the official death toll was as high as 3600, and 13.2 million people lost their homes. The flood also destroyed or damaged much infrastructure and many services facilities. Economic losses were estimated at over $36 billion. Fig. 1 is the illustration of piping leakage of the Yangtze River dike. As can be seen from Fig. 1, the failure of the Yangtze River dike was mainly caused by internal erosion in the form of piping, which is the process whereby internal voids are created by seepage flow. Approximately half of all the world's dam failures are attributed to piping phenomena (Foster *et al*. 2000, Fell *et al*. 2005, Zhang *et al*. 2009).Since a large part of the world's population lives near oceans, lakes and rivers

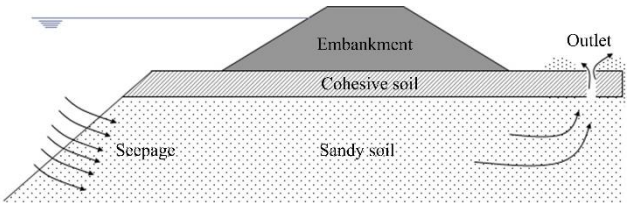

Fig.1 Illustration of piping leakage of the Yangtze River dike

(Song *et al*. 2011), floods resulting from breaching of dams can lead to devastating disasters with tremendous loss of life and property, especially in densely populated areas. Therefore, it is vital to investigate the assessment methods of internal stability of soils in embankment dams and their foundations.

Recently, many research works have been carried out to investigate the mechanism of internal stability of soils. For instance, Foster *et al*. (2000) evaluated the probability of dams by internal erosion and defined a four-stage process consisting of initiation of erosion, continuation of erosion, progression to form a pipe or occasionally cause surface instability and initiation of a breach. Garner and Fannin (2010) considered that the combination of material susceptibility, critical hydraulic load and critical stress condition can govern the initiation of internal erosion, as shown in Fig. 2. Chang and Zhang (2013) investigated the control variables for internal stability of soils and extended the internal stability criteria for well-graded and gap-graded

*Corresponding author, Professor
E-mail: xuexinhua@scu.edu.cn

ISSN: 2005-307X (Print), 2092-6219 (Online)

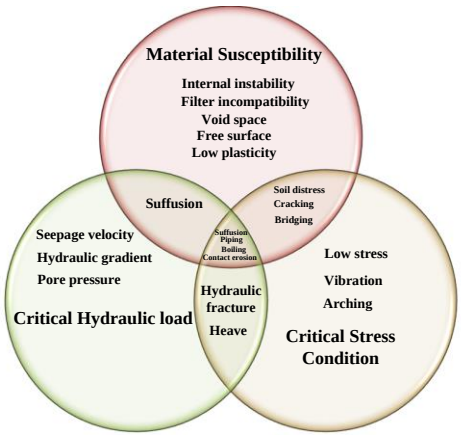

Fig. 2 Venn diagram illustrating interactions of geometric, hydraulic and mechanical susceptibility of soils to initiation of internal erosion (Garner and Fannin 2010)

soils based on laboratory tests. Atallah *et al*. (2015) investigated the piping potential of lake-bottom sediment and its role in seepage and lake-level fluctuations. They concluded that soil piping is the main reason for periodic water level fluctuations at Mountain Lake, Giles Country, Virginia, USA.

There are many factors can affect the internal stability of soils, such as geometric, hydraulic and mechanical conditions (Chang and Zhang 2013).These factors inherently exhibit the characteristics of uncertainty, fuzziness and spatial variability. Furthermore, it is difficult to have a definitive consideration of all the pertinent parameters. In this context, some new and innovative technologies are needed to assess the internal stability of soils.

In recent years, due to vast developments in computational software and hardware, several alternative artificial intelligence approaches such as artificial neural networks (ANNs) and genetic algorithm (GA) have emerged (e.g., Zhang and Goh 2016, Balendra and Ravi 2017, Xu *et al*. 2017, Wei and He 2017, Kaveh *et al*. 2018, Massimina *et al.* 2018). Although the ANNs and GA are found to be more efficient compared to statistical methods, they may have some inherent drawbacks such as difficulty in convergence. Except for ANNs and GA, the support vector machine (SVM) is also one of the widely used machine learning techniques and is gaining more and more attention due to its good performance and attractive features (Vladimir and Vapnik 2000). Presently, the SVM has been successfully used in various fields such as meteorology (Osowski and Garanty 2007), finance (Shin *et al*. 2005), bioinformatics (Lee and Lee 2003), pavement engineering (Maalouf *et al*. 2008), optimal control (Suykens *et al*. 2001), model induction (Dibike *et al*. 2001), text categorization (Wang and Chiang 2007) and time series analysis (Lau and Wu 2008).

When using SVM, one main issue is to effectively choose the kernel function and the optimal parameters. At present, the methods for optimizing SVM parameters can be classified into four types: 1) direct setting of parameters using empirical formulas (Cristianini *et al*. 2006); 2) metaheuristic algorithms, such as particle swarm optimization (Huang and Dun 2008) and the genetic algorithm (Huang and Wang 2006); 3) grid search algorithm (GSA); and 4) other methods, such as linear search and the gradient descent method. The direct-setting method performs quickly but has low accuracy. Metaheuristic algorithms are general-purpose algorithms, which do not depend on the problem and can offer good results, but it is still necessary to choose their own proper parameters (e.g., crossover rate, mutation rate and initial population) for the algorithms in optimizing SVM parameters. Compared to other optimization methods, the GSA can obtain the optimal parameters through exhaustive search and apply parallel computing, and thus saving the time required for parameter optimization. Therefore, the GSA is employed in this study to find the optimal parameters of SVM. Then, the cross - validation (CV) method is employed to estimate the classification accuracy of the GSA-SVM model. Two examples of internal stability of soils are presented to validate the predictive capabilities of the proposed GSA-SVM model. In addition, the predictions from the proposed GSA-SVM model were compared with those from the traditional back propagation neural network (BPNN) model.

2. Methodology

2.1 Support vector machines (SVM)

Consider a given training dataset $(x_i, y_i), x_i \in R^n, y_i \in (-1,+1), i = 1,2,\cdots,n$, where each $x_i \in R^n$ shows the input space of the sample, y_i is a label that determines the class of x_i. If data are linearly separable, the hyper-plane that separates the given data can be expressed as

$$f(x) = \sum_{i=1}^{n} w_i x_i + b = 0 \tag{1}$$

where b is the bias, w is an adjustable weight.

The following optimization problem is formulated

$$\text{Minimize:} \quad \frac{1}{2}\|w\|^2 \tag{2a}$$

$$\text{Subjected to:} \begin{cases} y_i - w \cdot x_i - b \le \varepsilon \\ w \cdot x_i + b - y_i \le \varepsilon \end{cases} \tag{2b}$$

where ε denotes the precision parameter.

It can be concluded from Eq. (2) that the convex optimization problem is feasible. Sometimes, however, this may not be the case. Hence, we can introduce slack variables ξ_i and ξ^*_i to cope with infeasible solutions. Therefore, Eq. (2) becomes

$$\text{Minimize:} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi^*_i\right) \tag{3a}$$

$$\text{Subjected to:} \begin{cases} y_i - w \cdot x_i - b \le \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \le \varepsilon + \xi^*_i \end{cases} \quad \xi_i, \xi^*_i \ge 0 \tag{3b}$$

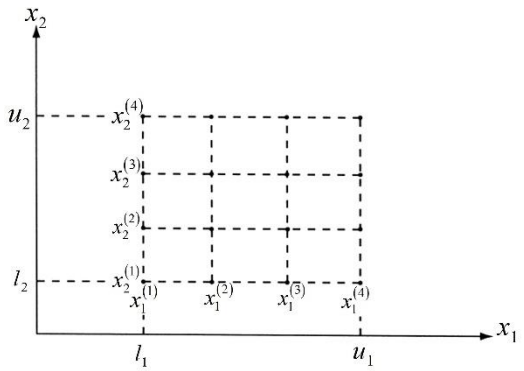

Fig. 3 Grid with p_i=4 (Rao 2009)

where C = penalty factor.

Solving the above optimal problems, the final form of SVM can be expressed as follows

$$f(x)=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)K\left(x_i,x_j\right)+b \tag{4}$$

where α_i and α^*_i are Lagrange multipliers. $K(x_i,x_j)$ is a kernel function. In this study, the radial basis function (RBF) kernel is employed and given as follows

$$K\left(x_i,x_j\right)=\exp\left(-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}\right) \tag{5}$$

where σ= the kernel parameter.

2.2 Grid search algorithm (GSA)

Grid search is the process of performing hyper-parameter tuning in order to determine the optimal values for a given model. In this method, a suitable grid in the design space is generated firstly, then the objective function is evaluated at all the grid points, eventually the grid point corresponding to the lowest function value is selected. For instance, if l_i and u_i are known as the lower and upper bounds on the ith design variable, respectively, the range (l_i, u_i) can be divided into p_i−1 equal parts so that $x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(p_i)}$ denote the grid points along the x_i axis (i=1,2…,n). Fig.3 shows a grid with p_i=4 in a two-dimensional design space. It can be seen that the grid method can be employed to find an approximate minimum for problems with a small number of design variables (Rao 2009).

2.3 Cross-validation (CV)

Cross-validation (CV) is a resampling procedure used to evaluate machine learning models on a limited data sample. It is primarily used in situations where one wants to estimate how accurately the machine learning model will perform for a given predictive modeling problem (Lee and Chern 2013).

One common type of CV is the k-fold cross validation. It is performed as per the following steps:

1) Split the original training data set into k equal folds f_1, f_2, ⋯, f_k.

2) For i=1 to i=k

a) Keep the fold f_i as validation set and keep all the remaining k -1 folds in the cross validation training data set.

b) Train the predictive model by use of the CV training set, and calculate the accuracy of the model by validating the predicted results against the validation set.

3) Estimate the accuracy of the predictive model by averaging the accuracies derived in all the k cases of CV.

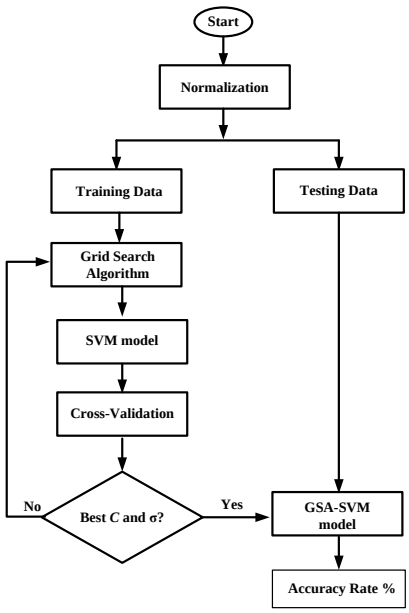

Fig. 4 Flow chart of the GSA-SVM model

2.4 GSA-SVM model

From the above it can be seen that the penalty factor C and kernel parameter σ need to be determined for a given problem. Herein the GSA is employed to find the optimal values of SVM. The steps of GSA-SVM model proposed in this study can be described as follows:

Step 1: Collect sample data and normalize them.

Step 2: Divide the data set into training set and testing set.

Step 3: Initialize the parameters' search ranges and look for the first optimal parameter by GSA in the training set. In this study, the 10-fold cross validation was employed to calculate the classification accuracy. That is, repeat the CV process 10 times and each of the 10 subsamples was regarded as validation samples.

Step 4: Set the reasonable neighborhood ranges of the first optimal solution and the suboptimal solutions, respectively. Then reduce the step length and subdivide the grid to perform parameter search. The CV method is applied again to calculate classification accuracy.

Step 5: Train the SVM using every optimal parameter solution of each group, which is supplied by Step 4. Then, calculate the classification accuracy of the training set by each optimal solution, respectively.

Step 6: Select the highest accuracy solution as the final solution; if solutions' accuracies are the same, select the solution whose penalty factor C is the smallest as the final

Table 1 Practical project data in case study 1 (1: piping; 0: no piping) (data from Zhang *et al*. (2004))

No.	H/m	H_w/m	λ	c'/kPa	φ'/°	γ_{sat}/kN/m^3	k/cm/s	d_b/cm	δ/°	Actual	GSA-SVM	BPNN
1	133	123	0.455	40	27.0	21.3	3.0×10^{-7}	0.009	79.0	1	1	0.804
2	87.5	80	0.4	12	28.9	21.2	3.5×10^{-7}	0.008	75.0	1	1	0.901
3	35.5	31	0.295	60	15.6	20.5	1.0×10^{-7}	0.008	43.0	1	1	0.856
4	31	29	0.249	20	26.7	20.8	4.0×10^{-4}	0.01	14.0	0	0	0.091
5	31	29	0.249	15	26.7	20.1	7.8×10^{-5}	0.009	14.0	0	0	0.070
6	29	25	0.435	30	31.6	20.8	2.0×10^{-5}	0.01	23.5	0	0	0.024
7	39	35.5	0.466	109	21.2	20.7	5.1×10^{-6}	0.04	29.0	0	0	0.029
8	39	35.5	0.466	76	13.8	20.7	5.1×10^{-6}	0.04	25.0	0	0	0.044
9	28	25	0.286	157	13.2	20.3	3.6×10^{-7}	0.009	60.0	1	1	0.911
10	28	25	0.286	153	24.8	21.2	4.8×10^{-8}	0.009	60.0	1	1	0.902
11	96	90	0.417	20	26.0	21.0	3.5×10^{-8}	0.012	65.0	1	1	0.863
12	56	49	0.364	30	29.0	19.5	2.0×10^{-7}	0.013	59.0	1	1	0.806
13	51	47	0.308	42	34.5	21.2	2.2×10^{-8}	0.012	66.0	1	1	0.907
14	133.1	126	0.476	41	32.0	21.7	1.3×10^{-6}	0.004	76.0	1	1	0.788
15	13.0	10.5	0.364	44	38.4	22.9	1.0×10^{-2}	0.02	26.6	0	0	0.007
16	6.7	5.5	0.4	109	21.2	20.7	5.1×10^{-6}	0.004	67.5	1	1	0.916
17	6.0	4.75	0.5	51	38.5	20.9	7.0×10^{-3}	0.024	26.6	0	0	0.001
18	87.5	80	0.256	14	27.0	21.2	4.0×10^{-6}	0.008	28.0	0	0	0.681
19	51.5	46	0.455	100	19.3	21.0	2.9×10^{-6}	0.007	45.0	1	1	0.852
20	39.5	33	0.347	32	27.2	20.2	5.5×10^{-8}	0.01	67.0	1	1	0.909
21	29.0	26	0.315	26	27.8	20.8	2.0×10^{-5}	0.01	65.0	1	1	0.915
22	42.5	39	0.361	84	32.2	20.6	1.3×10^{-8}	0.004	53.1	1	1	0.829
23	7.0	5.6	0.4	20	30.0	19.5	3.8×10^{-5}	0.017	26.6	0	0	0.006

solution. Then the optimal parameter group of SVM is determined.

Step 7: Perform model prediction using the testing set and the optimal parameters to test the performance of the model.

In this method, each pair of GSA-CV is used to reasonably divide the training set and make the training equalization. The GSA can search all nodes while CV can obtain training sample equilibrium. Besides, the GSA changes the searching tradition through searching the range of suboptimal solutions to find real optimal solution. This approach has the advantages of both GSA and CV methods and improves the search accuracy. A flowchart of the GSA-SVM algorithm is illustrated in Fig. 4.

3. Model training and validation using case studies

In this section, two case studies of internal stability of soils are presented to validate the predictive capability of the GSA-SVM model.

3.1 Case study No. 1 using field data

In this study, the following nine factors were taken into account as the input parameters for the GSA-SVM model: dam's height H, the level of water H_w, ratio of downstream slope λ, angle of effective internal friction φ', effective cohesion c', saturated unit weight γ_{sat}, permeability coefficient k, effective particle-size d_b and the angle of inclination of filter δ. The database used in this study was selected from Zhang *et al*. (2004). Among the 23 case records, the first 17 case records were used for training of the model. The remaining 6 case records were used as testing set for model validation. For each case, "1" represents piping and "0" represents no piping. In addition to verify the effectiveness of the proposed GSA-SVM model, this case study also adopts the traditional back propagation neural network (BPNN) to predict the internal stability of soils. In this case study, the structure of BPNN is designed as 9-5-1. That is, the number of neurons in input,

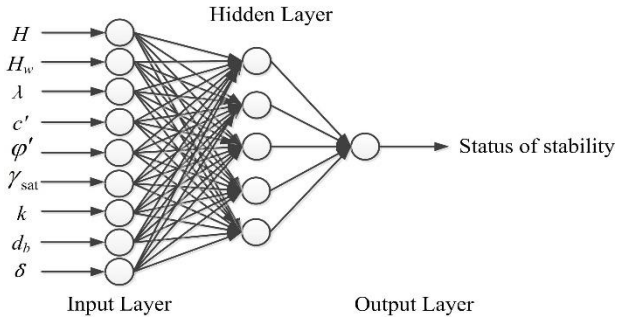

Fig. 5 Structure of BPNN in case study 1

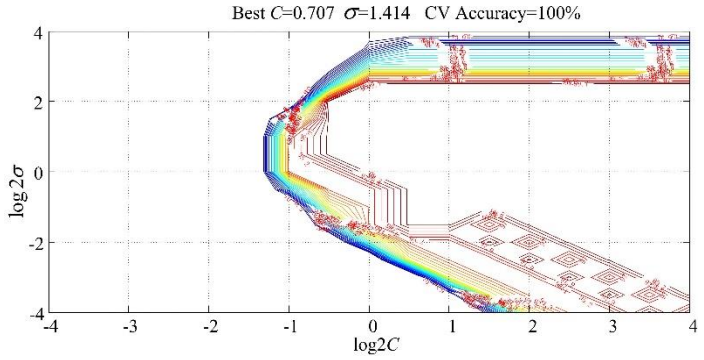

Fig. 6 Parameters C and σ versus the accuracy rate in two dimensions in case study 1

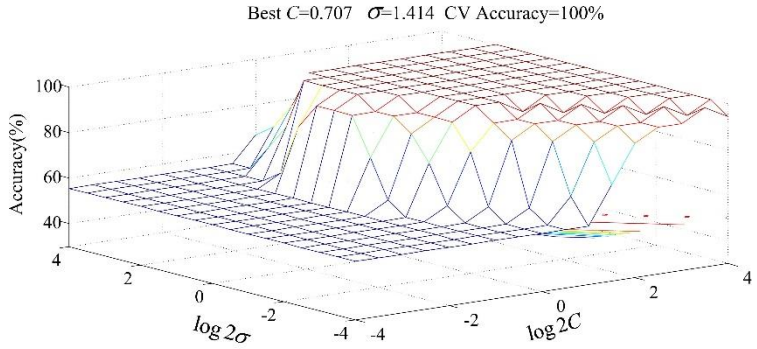

Fig. 7 Parameters C and σ versus the accuracy rate in three dimensions in case study 1

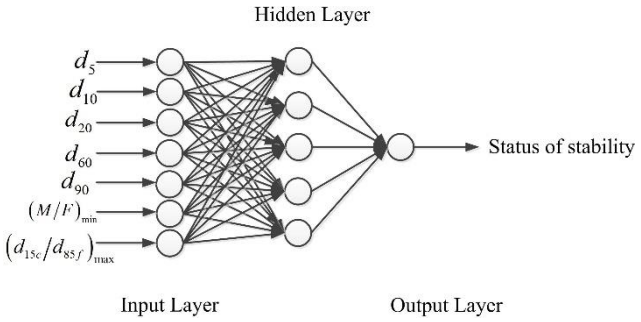

Fig. 8 Structure of BPNN in case study 2

hidden and output layers of BPNN is 9, 5 and 1, respectively. The structure of BPNN is shown in Fig.5. The detailed dataset is summarized in Table 1.

Before the dataset was used to train the GSA-SVM model, it was preprocessed using Eq. (6). Each parameter is normalized between 0 and 1, with

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{6}$$

where y is a normalized input parameter, x is the original input parameter, x_{max} and x_{min} are the maximum and minimum input parameters, respectively.

After performing the GSA-SVM procedure, the optimal parameters of SVM, i.e., C and σ, were selected and given as follows: the penalty factor C=0.707, the kernel parameter σ =1.414. Figs. 6 and 7 show the contour maps of parameters C and σ versus the accuracy rate in two and three dimensions, respectively. From Figs. 6 and 7 it can be seen that the CV accuracy is 100%. Then the proposed GSA-SVM model is employed to predict the internal stability of soils. From Table 1 it can be seen that the predictions from the proposed GSA-SVM model agree well with the actual results and it achieves a high accuracy rate (100%). However, there are one case was misclassified by BPNN and the overall classification accuracy rate is 95.65%.

3.2 Case study No. 2 using laboratory experimental data

In this case study, the following factors including various characteristic particle sizes, such as d_5, d_{10}, d_{20}, d_{60}, d_{90}, $(M/F)_{\min}$ and $(d_{15c}/d_{85f})_{\max}$ were taken into account as the input parameters of the GSA-SVM model, where M = mass fraction between grain size d and $4d$; F = mass fraction at any grain size d; d_{15c}= grain diameter corresponding to 15% mass passing in the coarse - grained portion (e.g., sand and gravel); d_{85f} = diameter corresponding to 85% mass passing in the fine - grained portion (e.g., silt and clay); d_5, d_{10}, d_{20},

Table 2 Training dataset in case study 2 (1: stable; 0:unstable) (data from Chang and Zhang (2013))

No.	d_5/mm	d_{10}/mm	d_{20}/mm	d_{60}/mm	d_{90}/mm	$(M/F)_{min}$	$(d_{15c}/d_{85f})_{max}$	Experimental	GSA-SVM	BPNN
1	0.195	0.3	0.52	7.0	14.8	0.53	3.6	1	1	0.386
2	0.134	0.23	0.4	2.5	12.3	1.7	3.0	1	1	0.999
3	0.12	0.191	0.325	1.18	5.4	1.4	2.6	1	1	0.999
4	0.175	0.42	1.8	10.0	16.8	0.73	8.1	0	0	0.111
5	0.58	1.23	2.62	14.5	21.5	1.3	4.9	1	1	0.995
6	0.58	1.26	2.61	18.7	66.0	1.3	4.8	1	1	0.914
7	0.3	0.89	2.67	14.0	22.0	1.3	10.5	1	1	0.876
8	0.65	1.13	1.9	8.4	18.1	1.9	1.4	1	1	0.999
9	0.68	0.98	1.4	3.6	7.0	2.1	1.7	1	1	0.999
10	2.05	3.6	4.15	10.3	19.3	2.1	2.1	1	1	0.999
11	0.36	0.51	1.01	13.6	27.4	0.5	3.5	0	0	0.289
12	0.401	0.64	1.41	11.2	20.5	0.89	3.7	0	0	0.453
13	0.23	0.39	1.19	9.10	15.8	0.49	6.7	0	0	0.267
14	1.8	3.3	5.58	10.9	20.6	2.8	4.0	1	1	0.999
15	0.56	1.03	1.75	3.9	5.68	3.7	3.2	1	1	0.999
16	0.181	1.25	7.9	37.6	68.0	0.62	51.0	0	1	0.000
17	0.278	0.53	1.78	37.5	88.8	0.65	6.5	0	0	0.000
18	0.467	1.06	3.29	41.6	89.8	1.0	8.0	0	0	0.000
19	0.161	0.502	6.78	28.6	44.2	0.25	41.0	0	0	0.000
20	0.89	2.0	5.81	33.8	69.0	1.3	6.2	1	1	0.978
21	0.191	0.235	0.331	4.51	15.0	0.35	3.5	1	1	0.439
22	0.198	0.36	1.19	6.82	15.6	0.65	6.3	0	0	0.213
23	0.221	0.298	0.458	2.09	9.21	1.8	1.7	1	1	0.999
24	0.002	0.01	0.052	0.091	0.129	0.55	33.0	1	1	0.699
25	0.002	0.009	0.152	0.176	0.196	0.4	13.0	0	0	0.412
26	0.008	0.098	0.192	0.51	1.34	0.17	17.0	0	0	0.300
27	0.002	0.01	0.132	0.491	1.3	0.17	16.0	1	1	0.327
28	0.002	0.009	0.168	0.452	1.42	0.04	20.0	0	0	0.065
29	0.008	0.17	0.248	0.5	1.59	0	16.3	0	0	0.021
30	0.762	0.851	1.04	2.45	8.0	1.7	1.4	1	1	0.999
31	0.589	0.702	0.98	3.89	11.1	1.3	1.7	1	1	0.988
32	0.47	0.642	1.12	6.98	13.2	1.4	2.3	1	1	0.998
33	0.03	0.052	0.12	1.18	6.0	0.35	4.1	0	0	0.438
34	0.14	0.454	1.28	7.38	13.6	1.4	8.7	1	1	0.989
35	1.3	1.68	2.49	7.96	13.5	2.1	1.9	1	1	0.999
36	0.078	0.16	0.432	4.25	12.1	1.1	5.4	1	1	0.908
37	0.142	0.163	0.228	4.98	12.4	1.0	4.7	0	0	0.487
38	1.85	11.5	20.5	34.8	50.4	0.5	13.8	0	0	0.008
39	0.201	1.35	2.8	16.0	43.0	1.5	15.0	1	1	0.974
40	0.122	0.179	1.68	4.28	6.35	0.14	10.9	0	0	0.051
41	0.25	0.45	1.68	4.28	6.35	0.98	9.7	0	0	0.141
42	0.389	0.672	1.68	4.28	6.35	1.7	4.7	1	1	0.999
43	0.61	1.06	2.18	4.5	6.35	2.7	4.0	1	1	0.999
44	0.104	0.174	0.397	4.98	12.8	1.0	3.6	1	1	0.928

Table 2 Continued

No.	d_5/mm	d_{10}/mm	d_{20}/mm	d_{60}/mm	d_{90}/mm	$(M/F)_{min}$	$(d_{15c}/d_{85f})_{max}$	Experimental	GSA-SVM	BPNN
45	0.13	0.37	1.42	11.9	17.9	0.88	10.6	0	0	0.026
46	0.092	0.125	0.212	2.10	11.6	1.2	2.3	1	1	0.998
47	0.165	0.6	0.68	1.06	1.47	1.5	3.7	1	1	0.999
48	0.136	0.158	0.6	1.0	1.45	1.4	3.7	1	1	0.999
49	0.127	0.142	0.177	0.94	1.45	0.8	3.7	1	1	0.697
50	0.13	0.155	0.42	0.708	1.04	2.8	2.6	1	1	0.999
51	0.122	0.135	0.168	0.658	1.02	1.6	2.6	1	1	0.999
52	0.128	0.151	0.83	1.33	1.9	0	5.2	0	0	0.046
53	0.121	0.135	0.167	1.24	1.87	0	5.2	0	0	0.081
54	0.131	0.156	1.24	2.0	2.85	0	7.4	0	0	0.023
55	0.122	0.14	0.18	1.85	2.74	0	7.4	0	0	0.062
56	0.362	1.67	14.2	100	167	0.44	35.0	0	1	0.000
57	0.224	0.67	3.62	80.2	160	0.5	18.0	0	0	0.000
58	0.098	0.26	1.08	35.4	98.0	0.67	35.0	0	1	0.000
59	0.009	0.46	2.8	33.5	81.0	0.17	9.9	0	0	0.000
60	0.25	0.495	5.0	76.5	130	0.22	25.0	0	0	0.000
61	0.23	0.9	2.8	16.2	42.1	0.88	13.0	0	0	0.011
62	0.14	0.26	0.84	37.5	46.1	0.5	9.0	0	0	0.000
63	0.145	0.174	0.298	3.8	6.89	0.67	6.0	0	0	0.239
64	0.124	0.218	0.605	7.02	20.1	0.72	5.0	0	0	0.178
65	0.362	0.526	2.31	10.3	16.2	0.83	7.0	0	0	0.184
66	0.105	0.145	0.254	11.9	50.2	0.56	13.7	0	0	0.021
67	0.06	0.134	0.242	11.9	50.2	0.88	14.3	0	0	0.004
68	0.224	0.42	1.25	8.73	14.9	0.72	5.5	1	1	0.266
69	0.632	0.952	1.67	3.52	5.35	2.2	2.4	1	1	0.999
70	0.015	0.108	0.215	0.37	0.44	0.25	14.5	0	0	0.448

Table 3 Testing dataset in case study 2 (1: stable; 0: unstable) (data from Chang and Zhang (2013))

No.	d_5/mm	d_{10}/mm	d_{20}/mm	d_{60}/mm	d_{90}/mm	$(M/F)_{min}$	$(d_{15c}/d_{85f})_{max}$	Experimental	GSA-SVM	BPNN
71	0.019	0.045	0.112	0.52	20.2	1.2	12.8	1	1	0.010
72	0.011	0.025	0.085	0.448	5.56	1.3	2.7	1	1	0.999
73	0.007	0.018	0.048	0.415	8.35	1.2	6.7	1	1	0.934
74	0.002	0.010	0.085	0.402	0.62	0	19.3	0	0	0.038
75	0.001	0.002	0.02	0.336	0.6	0.01	19.3	0	0	0.050
76	0.001	0.002	0.003	0.26	0.59	0.43	19.3	1	1	0.538
77	0.116	0.126	0.156	0.775	1.56	0.56	2.8	1	1	0.665
78	0.112	0.119	0.135	1.33	1.78	0	7.1	0	0	0.075
79	0.112	0.119	0.135	1.34	1.81	0	7.9	0	0	0.073
80	0.135	0.178	1.58	4.32	8.0	0.14	11.0	0	0	0.052
81	0.112	0.15	0.214	2.42	6.98	0.33	4.9	0	0	0.472
82	0.01	0.027	0.07	0.601	5.2	0.45	5.5	0	0	0.356
83	0.05	0.185	0.78	6.64	14.8	0.5	19	0	0	0.217
84	0.023	0.058	0.268	6.15	18.5	0.42	12.6	1	1	0.155
85	0.011	0.028	0.072	5.49	14.9	0.4	8.5	1	1	0.176

Table 3 Continued

No.	d_5/mm	d_{10}/mm	d_{20}/mm	d_{60}/mm	d_{90}/mm	$(M/F)_{min}$	$(d_{15c}/d_{85f})_{max}$	Experimental	GSA-SVM	BPNN
86	0.005	0.012	0.025	0.475	8.4	0.75	5.0	1	1	0.388
87	0.082	0.258	0.61	5.42	12.9	0.62	8.1	1	1	0.138
88	0.006	0.095	0.39	4.75	9.9	0.4	98.0	1	0	0.436
89	0.026	0.066	3.42	7.56	15.2	0.05	36.0	1	1	0.000
90	0.011	0.022	0.055	6.61	12.0	0.02	71.0	0	0	0.002
91	0.004	0.065	3.78	7.18	13.3	0.08	5.6	1	1	0.000
92	0.032	0.078	0.87	27.1	57.4	0.30	38.0	0	0	0.000
93	0.017	0.041	0.235	31.0	57.6	0.15	58.0	0	0	0.000
94	0.037	0.128	2.42	30.5	57.1	0.24	66.0	0	0	0.000
95	0.024	0.056	0.64	24.1	57.6	0.23	26.0	0	0	0.000
96	0.023	0.065	0.64	38.6	60.1	0.25	35.0	0	0	0.000
97	0.053	0.318	4.62	28.5	58.6	0.31	99.0	0	1	0.000
98	0.003	0.009	0.032	0.29	0.575	0.56	8.6	1	1	0.221
99	0.098	0.105	0.122	2.05	4.0	0.75	10.0	0	0	0.101
100	0.005	0.015	0.18	0.21	0.27	0	18.9	0	0	0.039

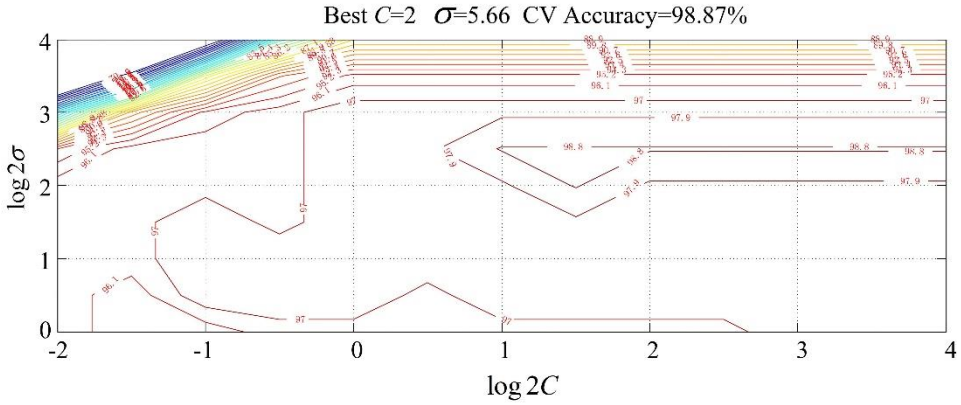

Fig. 9 Parameters C and σ versus the accuracy rate in two dimensions

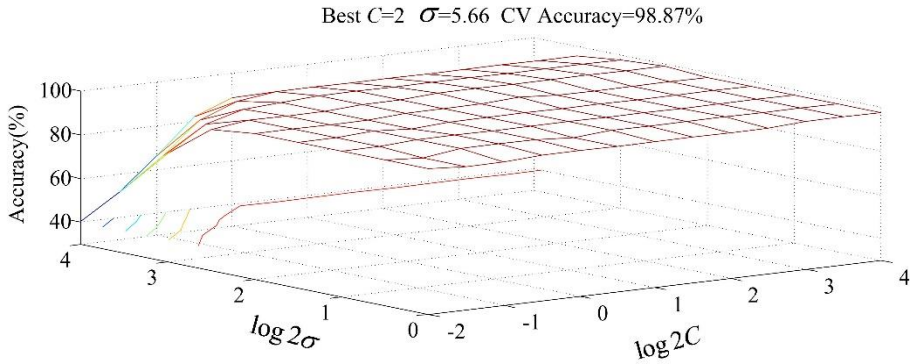

Fig. 10 Parameters C and σ versus the accuracy rate in three dimensions

d_{60} and d_{90} = diameters corresponding to 5%, 10%, 20%, 60% and 90% mass passing, respectively. The database used in this study was selected from Chang and Zhang (2013). 70% of the total data were used for training, and the remaining 30% of the data were used for testing. In this case study, "1" represents stable and "0" denotes unstable. In addition to verify the effectiveness of the proposed GSA-SVM model, this case study also adopts the traditional BPNN model to predict the internal stability of soils. Herein, the structure of BPNN is designed as 7-5-1. That is, the number of neurons in input, hidden and output layers of BPNN is 7, 5 and 1, respectively. The structure of BPNN is shown in Fig. 8. The detailed datasets are summarized in Tables 2 and 3, respectively.

After performing the GSA-SVM procedure, the optimal parameters of SVM, i.e., C and σ, were selected and given as follows: the penalty factor C=2, the kernel parameter σ =5.66. Figs. 9 and 10 show the contour maps of parameters C and σ versus the accuracy rate in two and three dimensions, respectively. From Figs. 9 and 10 it can be seen that the CV accuracy is 98.87%. Then the proposed GSA-SVM model is employed to predict the internal stability of soils. The predicted results are listed in Tables 2 and 3. As can be seen from Tables 2 and 3, 95 out of the 100 datasets were correctly classified by GSA-SVM, achieving an overall classification accuracy rate of 95%. However, there are 13 cases were misclassified by BPNN and the overall classification accuracy rate is 87%. The results show that the proposed GSA-SVM model is an effective tool for predicting internal stability of soils

4. Conclusions

In this study, an improved support vector machine (SVM) model is developed for the assessment of internal stability of soils. The grid search algorithm (GSA) is employed to find the optimal parameters of SVM firstly, and then the cross - validation (CV) method is employed to estimate the classification accuracy of the GSA-SVM model. In addition to verify the effectiveness of the proposed GSA-SVM model, this study also adopts the traditional back propagation neural network (BPNN) model to predict the internal stability of soils. The following conclusions may be drawn from this study.

1) When using SVM, one main issue is to effectively choose the kernel function and the optimal parameters. Two examples of internal stability of soils confirmed that the combination of GSA and CV technologies can effectively improve the whole search accuracy of the SVM model.

2) The developed GSA-SVM model is a feasible, efficient and accurate tool for predicting the internal stability of soils. GSA-SVM model may be one of the most competent artificial intelligence subsystems to evaluate the internal erosion potential and stability of soils.

3) There are some main factors that affect the internal stability of dams, levees and other earthen structures, such as the erodibility of the soil, the water velocity inside the soil mass and the geometry of the earthen structure, etc. Thus, the mechanism of internal erosion and stability of soils is very complicated and needs to be further investigated through laboratory and field tests.

References

Atallah, N., Shakoor, A. and Watts, C.F. (2015), "Investigating the potential and mechanism of soil piping causing water-level drops in Mountain Lake, Giles Country, Virginia", *Eng. Geol.*, **195**, 282-291. https://doi.org/10.1016/j.enggeo.2015.06.001.

Balendra, M.M. and Ravi, S.S. (2017), "Assessment of slope stability using multiple regression analysis", *Geomech. Eng.*, **13**(2), 237-254. https://doi.org/10.12989/gae.2017.13.2.237.

Castiglia, M., de Magistris, F.S. and Napolitano, A. (2018), "Stability of onshore pipelines in liquefied soils: Overview of computational methods", *Geomech. Eng.*, **14**(4), 355-366. http://doi.org/10.12989/gae.2018.14.4.355.

Chang, D.S. and Zhang, L.M. (2013). "Extended internal stability criteria for soils under seepage." *Soils Found.*, **53**(4), 569-583. https://doi.org/10.1016/j.sandf.2013.06.008.

Cristianini, N., Kandola, J., Elisseeff, A. and Shawe-Taylor, J. (2006), *On Kernel Target Alignment*, in *Innovations in Machine Learning: Theory and Applications*, Springer, 205-256.

Dibike, Y.B., Velickov, S., Solomatine, D.P. and Abbott, M. (2001), "Model induction with support vector machines: Introduction and applications", *J. Comput. Civ. Eng.*, **15**(3), 208. https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208).

Fell, R., MacGregor, P., Stapledon, D. and Bell, G. (2005), *Geotechnical Engineering of Dams*, Balkema, Leiden, The Netherlands.

Foster, M. and Fell, R. (1999), "A framework for estimating the probability of failure of embankment dams by internal erosion and piping using event tree methods", UNICIV R-377, University of New South Wales, Australia.

Foster, M., Fell, R. and Spannagle, M. (2000), "The statistics of embankment dam failures and accidents", *Can. Geotech. J.*, **37**(5), 1000-1024. https://doi.org/10.1139/t00-030.

Garner, S.J. and Fannin, R.J. (2010), "Understanding internal erosion: a decade of research following a sinkhole event", *Int. J. Hydropower Dams*, **17**(3), 93-98.

Huang, C.L. and Dun, J.F. (2008), "A distributed PSO-SVM hybrid system with feature selection and parameter optimization", *Appl. Soft Comput.*, **8**(4), 1381-1391. https://doi.org/10.1016/j.asoc.2007.10.007.

Huang, C.L. and Wang, C.J. (2006), "A GA-based feature selection and parameters optimization for support vector machines", *Expert Syst. Appl.*, **31**(2), 231-240. https://doi.org/10.1016/j.eswa.2005.09.024.

Kaveh, A., Hamze-Ziabari, S.M. and Bakhshpoori, T. (2018), "Soft computing-based slope stability assessment: A comparative study", *Geomech. Eng.*, **14**(3), 257-269. http://doi.org/10.12989/gae.2018.14.3.257.

Lau, K.W. and Wu, Q.H. (2008), "Local prediction of non-linear time series using support vector regression", *Pattern Recogn.*, **41**(5), 1539-1547. https://doi.org/10.1016/j.patcog.2007.08.013.

Lee, C.Y. and Chern S.G. (2013), "Application of a support vector machine for liquefaction assessment", *J. Mar. Sci. Technol.*, **21**(3), 318-324. https://doi.org/10.6119/JMST-012-0518-3.

Lee, Y. and Lee, C. (2003), "Classification of multiple cancer types by multicategory support vector machines using gene expression data", *Bioinformatics*, **19**, 1132-1139. https://doi.org/10.1093/bioinformatics/btg102.

Maalouf, M., Khoury, N. and Trafalis, T.B. (2008), "Support vector regression to predict asphalt mix performance", *Int. J. Numer. Anal. Meth. Geomech.*, **32**(16), 1989-1996. https://doi.org/10.1002/nag.718.

Osowski, S. and Garanty, K. (2007), "Forecasting of the daily meteorological pollution using wavelets and support vector machine", *Eng. Appl. Artif. Intel.*, **20**(6), 745-755. https://doi.org/10.1016/j.engappai.2006.10.008.

Pham-Van, S., Hinkelmann, R., Nehrig, M. and Martinez, I. (2011), "A comparison of model concepts and experiments for seepage processes through a dike with a fault zone", *Eng. Appl. Comput. Fluid Mech.*, **5**(1), 149-158. https://doi.org/10.1080/19942060.2011.11015359.

Rao, S.S. (2009), *Engineering Optimization: Theory and Practice*, John Wiley & Sons, Inc., Hoboken, New Jersey, U.S.A.

Shin, K.S., Lee, T.S. and Kim, H.J. (2005), "An application of support vector machines in bankruptcy prediction model", *Expert Syst. Appl.*, **28**(1), 127-135. https://doi.org/10.1016/j.eswa.2004.08.009.

Suykens, J.A.K., Vandewalle, J. and de Moor, B. (2001), "Optimal control by least squares support vector machines", *Neural*

Networks, **14**(1), 23-35. https://doi.org/10.1016/S0893-6080(00)00077-0.

Vladimir, N. and Vapnik, V. (2000), *The Nature of Statistical Learning Theory*, Springer Verlag, New York, U.S.A.

Wang, T.Y. and Chiang, H.M. (2007), "Fuzzy support vector machine for multi-class text categorization", *Inform. Process. Manage.*, **43**(4), 914-929. https://doi.org/10.1016/j.ipm.2006.09.011.

Wei, G. and He, T.Y. (2017), "Displacement prediction in geotechnical engineering based on evolutionary neural network", *Geomech. Eng.*, **13**(5), 845-860. http://doi.org/10.12989/gae.2017.13.5.845.

Xu, J.C., Ren, Q.W. and Shen, Z.Z. (2017), "Sensitivity analysis of the influencing factors of slope stability based on LS-SVM", *Geomech. Eng.*, **13**(3), 447-458. http://doi.org/10.12989/gae.2017.13.4.447.

Zhang, L.M., Xu, Y. and Jia, J.S. (2009), "Analysis of earth dam failures-A database approach", *Georisk*, **3**(3), 184-189. https://doi.org/10.1080/17499510902831759.

Zhang, W.G. and Goh, A.T.C. (2016), "Evaluating seismic liquefaction potential using multivariate adaptive regression spines and logistic regression", *Geomech. Eng.*, **10**(3), 269-284. http://doi.org/10.12989/gae.2016.10.3.269.

Zhang, W.H., Yu, G.S. and Cai, Y.Q. (2004), "Mechanism model and artificial intelligence method for prediction and judgment of piping occurring in embankment", *J. Zhejiang Univ. Eng. Sci.*, **38**(7), 902-908. (in Chinese).

Zhao, Z.X., Chen, J.S. and Chen L. (2008), "Application of BP neural network to assessment of noncohesive piping-typed soils", *Chin. J. Geotech. Eng.*, **30**(4), 536-540. (in Chinese).

CC