

## Article

# A Machine Learning Approach to Evaluating the Damage Level of Tooth-Shape Spur Dikes

Xianglong Wei <sup>1,2</sup>, Yongjun Lu <sup>1,\*</sup>, Zhili Wang <sup>1</sup>, Xingnian Liu <sup>2</sup> and Siping Mo <sup>1</sup>

<sup>1</sup> State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China; sevenup2018@163.com (X.W.); zlwang@nhri.cn (Z.W.); spmo@nhri.cn (S.M.)

<sup>2</sup> State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China; liuxingnian@126.com

\* Correspondence: yjlu@nhri.cn; Tel.: +86-130-5768-6402

Received: 11 October 2018; Accepted: 14 November 2018; Published: 17 November 2018



**Abstract:** Little research has been done on the application of machine learning approaches to evaluating the damage level of river training structures on the Yangtze River. In this paper, two machine learning approaches to evaluating the damage level of spur dikes with tooth-shaped structures are proposed: a supervised support vector machine (SVM) model and an unsupervised model combining a Kohonen neural network with an SVM model (KNN-SVM). It was found that the supervised SVM model predicted the damage level of the validation samples with high accuracy, and the unsupervised data-mining KNN-SVM model agreed well with the empirical evaluation result. It is shown that both machine learning approaches could become effective tools to evaluate the damage level of spur dikes and other river training structures.

**Keywords:** spur dikes; SVM; KNN; damage level

## 1. Introduction

The Yangtze River is the largest river, and the most important navigation channel, in China. In recent decades, systematic river training works have been carried out on this waterway. Various channel training structures, such as spur dikes and revetments, have been built, which play important roles in protecting key shorelines, controlling unfavorable riverbed evolution, and increasing the navigability of the channel. However, as time has passed, water and sand scour have caused the channel training structures to sustain various levels of damage, impairing their ability to regulate the channel.

At present, research on channel training structures is mostly focused on the damage mechanism, such as local scour in the vicinity of the structures [1–3] and turbulent flow around the structures [4–6], and how changes in structure design parameters affect the water flow and sediment transport characteristics [7,8]. Little research about evaluating the damage level of the channel training structures has been carried out so far. It is very important to evaluate the damage level of these structures in order to maintain them over time. There are many factors that could affect the running condition of the training structures. These factors are difficult to describe quantitatively. Some researchers use a number to describe the stability of hydraulic engineering structures; for example,  $N_s$  in Equation (1). This equation was proposed by Hudson (1959) [9] to calculate the mass of stones in a rubble mound breakwater.  $N_s$  describes the stability of the structures, and is a function of the slope angle  $\alpha$  and the stability coefficient  $K_D$  as shown in Equation (2).  $K_D$  is related to the damage ratio ( $D$ ) of the rubble mound breakwater.

$$W = \frac{\rho_\gamma H^3}{N_s^3 \times (Sr - 1)^3} \quad (1)$$

$$N_s^3 = K_D \times \cot \alpha \quad (2)$$

where  $S_\gamma = \frac{\rho_\gamma}{\rho}$ ,  $\rho_\gamma$  is the mass density of the unit,  $\rho$  is the mass density of the water.  $H$  is the wave height.

Since some factors, such as the size of the structures and the duration of a wave attack, are not considered in the Hudson's formula, researchers have further developed this formula and given various definitions for the damage parameter (damage level), such as  $N_\Delta$ ,  $N_0$ , and  $S$  shown in Table 1 [10–14].

**Table 1.** Various definitions for the damage parameter (damage level).

Definition	Formula	Researcher
Damage parameter	$N_\Delta = \frac{A\rho_b^9 D_{50}}{\rho_a D_{50}^3 \frac{\pi}{6}}$	Thompson and Shuttler (1975) [10]
Damage parameter	$N_0 = \left( \frac{H_{1/3}/\Delta Dn - 1.33}{2.32} \right)^2 N^{0.5}$	Hanzawa et al. (1996) [11]
Damage level	$S = \frac{A}{D_{50}^3}$	Vander Meer (1988) [12]
Damage level	$S = S_e (1 - e^{-KN})$	Kaku (1991) [13]
Damage level	$S = 0.6D'$	Kajima (1994) [14]

Notes:  $A$  is the erosion area in a cross-section,  $\rho_b$  is bulk density of material as laid on the slope,  $\rho_a$  is mass density of stone,  $D_{50}$  is the diameter of stone which exceeds the 50% value of the sieve curve,  $\Delta$  is  $S_r - 1$ ,  $Dn$  is the nominal diameter of the unit.

With regard to the channel training structures in the Yangtze River, Han proposed Equation (3) to calculate the volume damage ratio of a spur dike and Equation (4) to calculate the area damage ratio of a flexible mattress for beach protection [15,16]. The damage levels are evaluated based on the damage ratios. Besides this, the strength reduction method can also be used to evaluate the damage level. Niu [17] used this method to obtain the stability number of the slope.

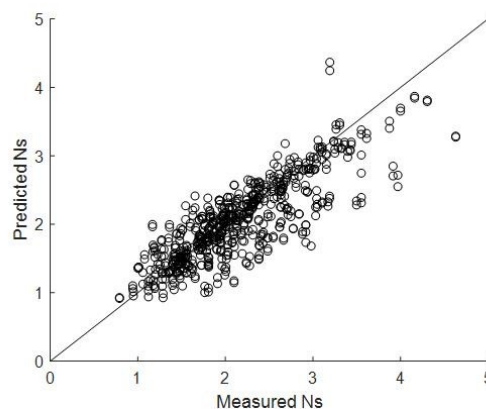
$$\frac{V_d}{V_t} = 0.000069 \left( \frac{V_m - V_c}{\sqrt{gH}} \right)^{1.1} \left( \frac{L_D}{H} \right)^{-0.25} \left( \frac{d_{50}}{H} \right)^{-0.7} \left( \frac{D}{H} \right)^{-0.9} \left( \frac{180 - \theta}{90} \right)^{0.11} \quad (3)$$

$$\frac{S_d}{S_t} = 2.31 \left( \frac{v}{\sqrt{gH}} \right)^{1.4} \left( \frac{b_{max}}{B} \right)^{0.8} \exp \left[ -7.2 \left( \frac{H - d}{d} - 0.3 \right)^2 \right] \quad (4)$$

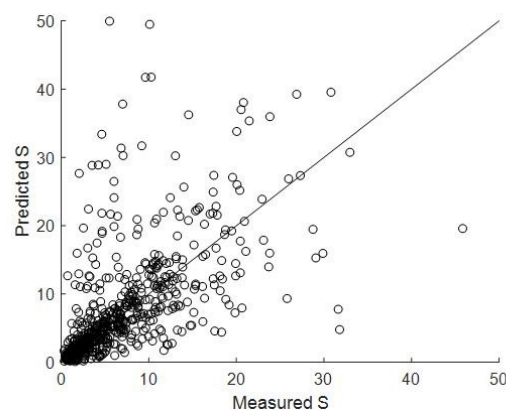
where  $V_d$  is the damage volume of the spur dike,  $V_t$  is the total volume of the spur dike,  $L_d$  is the length of the spur dike,  $V_m$  is the maximum velocity,  $V_c$  is the incipient velocity,  $\theta$  is the angle of the spur dike.  $S_d$  is the damage area of beach protection flexible mattress,  $S_t$  is the total area of beach protection flexible mattress,  $b_{max}$  is the maximum of beach width,  $l_{max}$  is the maximum of beach length,  $B$  is the width of the river,  $d$  is average water depth,  $v$  is the water velocity.

As can be seen in Figures 1 and 2, the stability numbers and the damage levels that are calculated by the empirical formulae do not agree well with those observed in the experiments, so a neural network was proposed to analyze the stability number and the damage level of the structures. Mase [18] proposed a back-propagation neural network model to analyze the stability of the rubble-mound breakwater based on the data from Van der Meer's 1988 experiment. The stability numbers and damage levels predicted by the trained model are more accurate than those given by the formulae. Dong et al. [19] established five different neural network models to predict the stability numbers and damage levels, and compared the structure and accuracy of the five neural network models. Besides

the neural network, a fuzzy logic approach [20,21] and a support vector machine [22,23] can also be effective tools to predict the stability number and the damage level.



**Figure 1.** A comparison between the measured and the predicted stability numbers by Van der Meer's formula.



**Figure 2.** A comparison between the measured and the predicted damage levels by Van der Meer's formula.

Little research has been done on the methods for evaluating the damage level of river training structures on the Yangtze River. It is hard to evaluate the damage level of river training structures quantitatively, since there are so many interrelated factors that affect the operating condition of the river training structures. Currently, the most common method to evaluate the damage level of the river training structures on the Yangtze River channel is to classify the damage level empirically according to measured data and the damage condition as observed on site. Damage to the river training structures can be classified into four levels, as shown in Table 2.

**Table 2.** The empirical criteria for a damage level evaluation of the river training structures on the Yangtze River.

Damage Level	Evaluation Criterion
Level 1	In good condition; operating normally
Level 2	Slight deformation; normal operation not affected
Level 3	Large deformation; operation can be sustained but reparation is needed immediately
Level 4	Severe damage; function lost or severely impaired

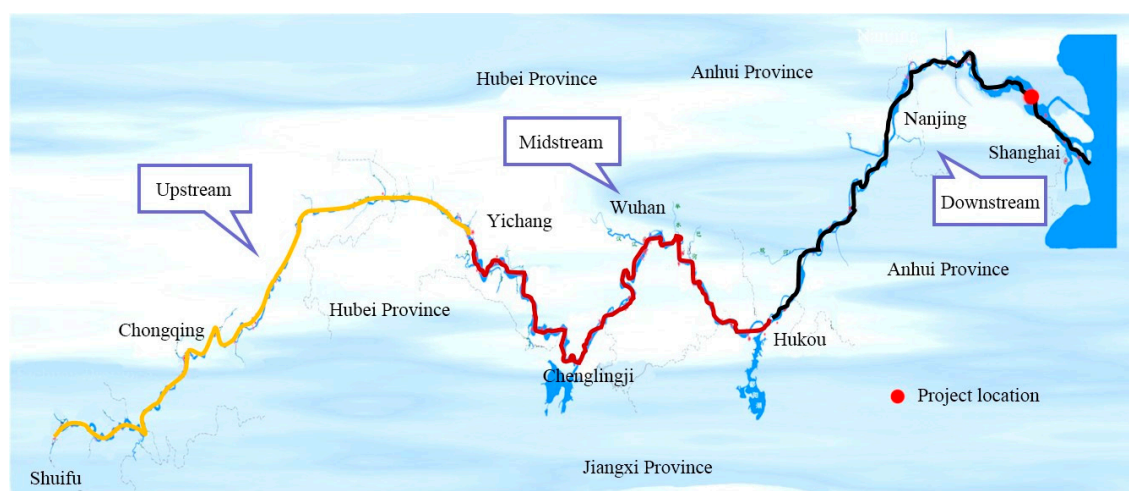
Previous studies have shown that supervised machine learning approaches could be effective tools to predict the damage level of hydraulic structures. During the early stages of study and application of a new river training structure, the empirical classification results of the damage level will not be reliable enough, but would form the training labels of the supervised machine learning algorithm.

Then, the use of a data-mining algorithm would be required. The primary purpose of this paper is to present a data-driven model without involving the hydrodynamic factors to evaluate the damage level of the river training structures. In the following sections, a machine learning approach is proposed to evaluate the damage level of a spur dike with tooth-shaped structures, which is a new type of river training structure that has been built near the estuary of the Yangtze River. Two machine learning models were built based on the experimental data in Fei (2017) [24]. First, we used the support vector machine (SVM) algorithm to build a supervised training model to predict the damage level of the validation samples. Then, we combined a Kohonen neural network with the support vector machine to build an unsupervised evaluation model (KNN-SVM), which could be seen as a novel approach to predict the water damage level of spur dikes.

## 2. Materials and Methods

### 2.1. The 12.5 m Deep-Water Channel Training Project on the Yangtze River below Nanjing

The Yangtze River is a major navigation channel in China, and is known as the “Golden Channel”. The length of the main stem of the Yangtze River is 6300 km. The stream above Yichang is the upper reach of the Yangtze River, which is 4505 km long. The stream between Yichang and Hukou is the middle reach, which is 955 km long. The stream below Hukou is the lower reach, which is 938 km long, as shown in Figure 3. Since 2012, Phase 1 of the 12.5 m Deep-Water Channel Training Project has been carried out downstream of Nanjing to regulate the 56-km-long channel between Taicang and Nantong in order to increase the depth of the channel to 12.5 m.



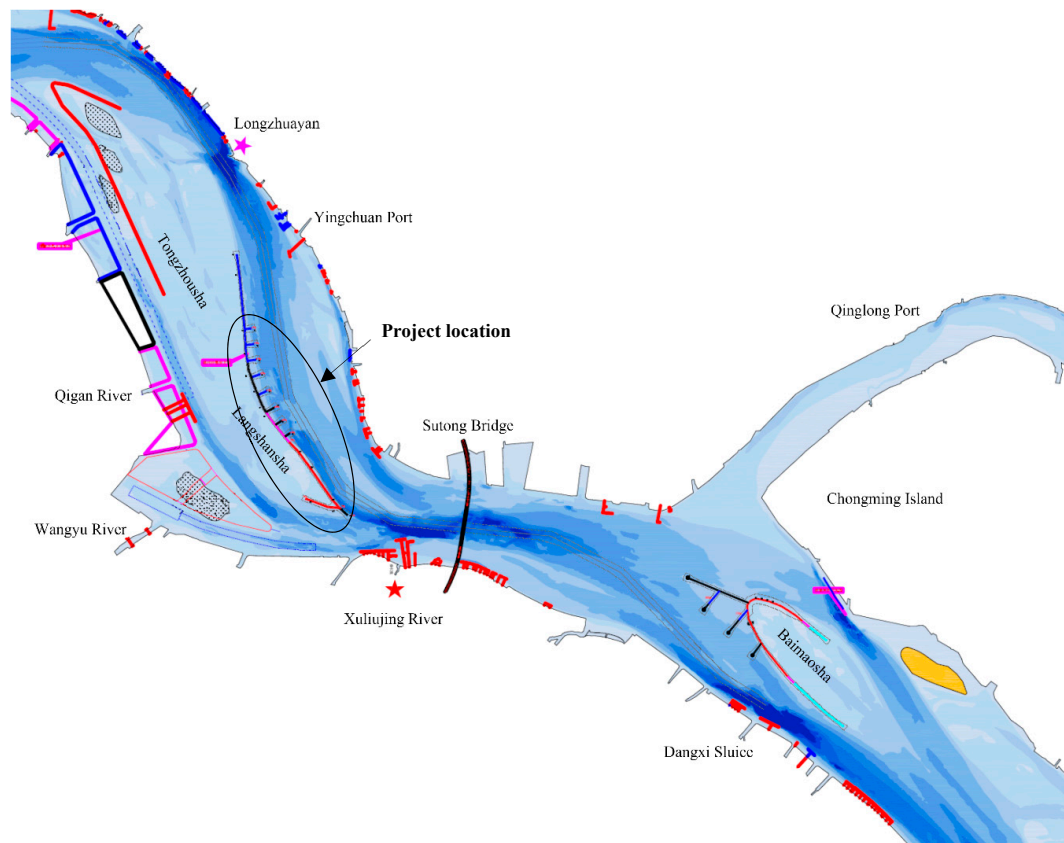
**Figure 3.** The main channel of the Yangtze River (from Shuifu to the estuary of the Yangtze River).

The regulated channel is near the estuary; therefore, it is influenced by both the river flow and the tidal flow. During the flood season, the tide could reach the Jiangyin–Jiulong reach; during the drought season, the tide could reach the Yizheng–Zhenjiang reach. When an extraordinary flood occurs ( $82,300 \text{ m}^3/\text{s}$ ), the tide can only affect the estuary reach. Sediment transport is complicated in a tidal estuary. The medium size of the bed load is  $0.15\text{--}0.25 \text{ mm}$ , and the medium size of the suspended load is  $0.005\text{--}0.02 \text{ mm}$ . The sediment concentration is influenced by sediment from upstream that is provided by the river and sediment from downstream that is provided by the tide. Since the Three Gorges Reservoir started operating, the sediment concentration in this section has decreased to  $0.1\text{--}0.3 \text{ kg/m}^3$ .

The project aims at protecting the beach, stabilizing the channel, guiding the flow, and increasing the channel’s depth. A 34.95-km-long submerged dike, 11 spur dikes, and four revetments have been built in the Tongzhousha–Baimaoshan reach.

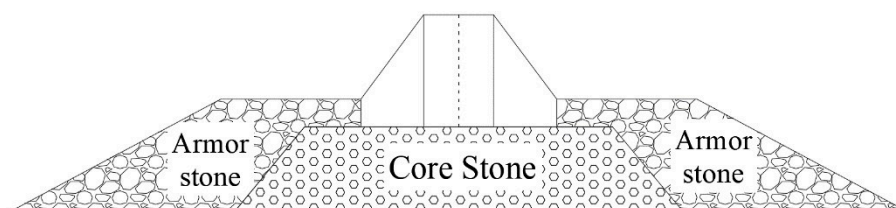
## 2.2. The New Type of River Training Structure on the Yangtze River

A spur dike with tooth-shaped structures is a new type of river training structure that was first used near the estuary of the Yangtze River in the Tongshazhou reach (shown in Figure 4). Spur dikes with tooth-shape structures were built perpendicular to the revetments. They were expected to control unfavorable riverbed evolution under the wave action and the river current and provide a habitat for the aquatic organisms. Their research and application are still at an early stage. The characteristics of the flow field have recently been researched using a numerical simulation [25,26]. Tong et al. [27] carried out a series of experiments to study the stability of spur dikes with tooth-shaped structures.



**Figure 4.** A location map of the channel training project in the estuary of the Yangtze River.

Figure 5 shows the profile of a spur dike with tooth-shaped structures. The spur dike consists of two parts: the upper part is made of precast concrete, and the lower part of the dike is core stone covered by armor stones. The size of each part is marked in Figure 5. It is a submerged dike. Each spur dike block is 6.2 m long, 7 m wide, and 4 m high. A three-dimensional (3D) view of the spur dike array can be seen in Figure 6. The weight of each block is 155.16 t, and the volume of each block is 64.65 m<sup>3</sup>. The weight of the core stone ranges from 1 kg to 100 kg, and the thickness of the core stone layer is 3 m. The weight of the armor stone ranges from 200 kg to 300 kg, and the armor stone layer is 4 m high.



**Figure 5.** The cross section of a spur dike with tooth-shaped structures (cm) [27].





**Figure 6.** A row of a spur dike with tooth-shaped structures.

### 2.3. Data Sets

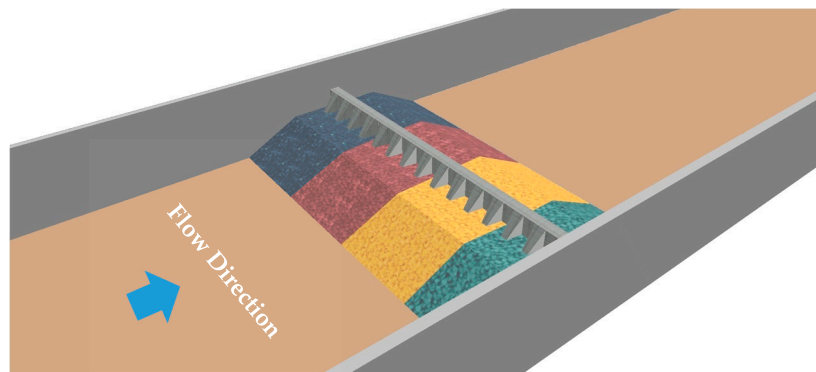
All data used to train the model in this paper were from flume experiments that were conducted by Fei [24]. And the data sets were obtained from the doctoral thesis of Fei (2017). In this paper, we might not have the right to share all the experiment data, but the data was public and available. This section provides a brief introduction of the experiment in order to clarify the source and meaning of the data.

The experiment was set up in a flume that was 40 m long and 2 m wide. A 15-meter-long movable bed section in the middle part of the flume was the main experimental area. A 15-cm-deep layer of uniform sediment ( $D_{50} = 0.075$  mm) covered the movable bed reach. Eight dike blocks (shown in Figure 7) stood in a line in the flume, and the central axis of the spur dike was perpendicular to the flow direction. The block models were made of concrete. The core stones were placed under the spur dike. The armor stones were divided equally into four parts in the transverse direction and dyed with four different colors in order to trace them (Figure 8). Both the horizontal scale and the vertical scale of the model were 1:25.

The experimental conditions were based on the actual hydrological condition of the 12.5 m Deep-Water Channel Regulation Project on the Yangtze River. The water depth ranged from 26 cm to 36.5 cm (corresponding to 6.5~9.2 m on site) based on the Froude similarity, and, similarly, the velocity upstream of the spur dike ranged from 15 cm/s to 33 cm/s (corresponding to 0.75~1.65 m/s on site). Each experiment lasted for 3 h.

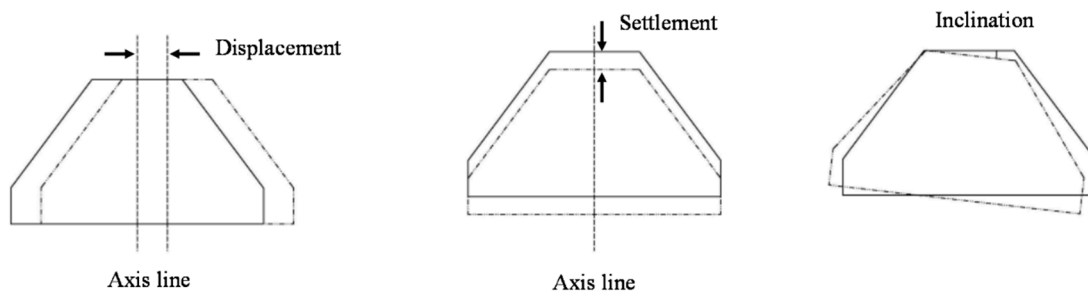


**Figure 7.** A single block model of a spur dike with tooth-shaped structures.



**Figure 8.** The layout of the spur dike with tooth-shaped structures in the flume [24].

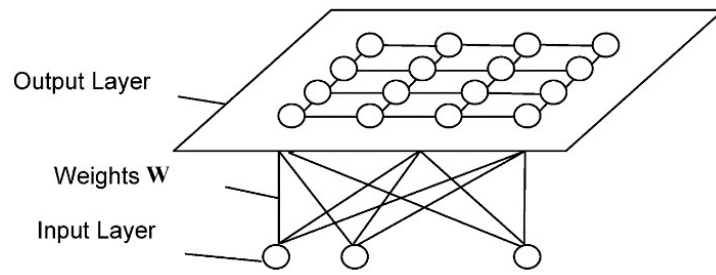
The flow drag and local scour in the vicinity of the structure were the main reasons for the damage to the spur dike [24]. Displacement, settlement, and inclination were the three major damage patterns. Figure 9 shows the schematic diagram of these damage patterns. After a 3-h flow attack experiment, the values of the displacement, settlement, and inclination were measured. Twenty-one (21) measurement points were placed on the axis line of the spur dikes at an equal distance. Each set of data included the values of displacement, settlement, and inclination at a measurement point. Sixteen (16) groups of data were measured, and 336 sets of experimental data were collected in total, which are used in this paper to train the machine learning model.



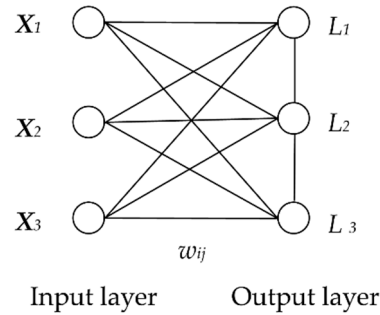
**Figure 9.** The three damage patterns of spur dikes with tooth-shaped structures [24].

#### 2.4. The Kohonen Neural Network

Self-organized Feature Maps (SOM) was proposed by Kohonen [28], and is known as a Kohonen neural network (KNN). KNN is an unsupervised learning model for cluster analysis. Figure 10 shows the usual structure of KNN [29]. It has two layers, the input layer and the output layer. The number of neurons in the input layer is determined by the input vector dimension. The neurons of the output layer are placed on a two-dimensional matrix. There are weights between the neurons in the input layer and the neurons in the output layer. There are also weights on the output layers between the output neurons. For this paper, a one-dimensional KNN was used, whose structure is shown in Figure 11. There were three neurons ( $X_1$ ,  $X_2$  and  $X_3$ ) in the input layer, standing for the input data of displacement, settlement, and inclination. Damage to the spur dike in the flume experiment was classified into three levels ( $L_1$ ,  $L_2$  and  $L_3$ ), so three neurons were set in the output layer.



**Figure 10.** The normal structure of a Kohonen neural network (KNN).



**Figure 11.** The KNN structure used in this paper.

Assuming that there are  $N$  neurons in the input layer and  $M$  neurons in the output layer, here is a brief introduction of the training process of the network. More detailed information about KNN can be found in Kohonen [30] and Waller et al. [31]:

Step 1: Initialize the weights  $W_{ij}$  randomly ( $i = 1, 2, 3, \dots, n; j = 1, 2, 3, \dots, m$ ).  $W_{ij}$  are the weights between the input layer and the output layer.

Step 2: Calculate the Euclidean distance between the randomly selected input  $X_i$  and weight vectors  $W_{ij}$ .

$$d_j = \left[ \sum_{i=1}^n (X_i - \omega_{ij})^2 \right]^{0.5} \quad (5)$$

Step 3: The neuron that has the minimum Euclidean distance to the input vector  $X$  is defined as the winner and marked as  $c$ . The winner has a neighborhood of  $N_c(t)$  whose radius is  $r$ . The neighborhood radius size ( $r$ ) changes linearly in the training process as shown in Equation (6).

$$r_s = r_{max} - \frac{s}{T}(r_{max} - r_{min}) \quad (6)$$

where  $T$  is the total number of training loops,  $s$  is the current training loop,  $r_{max}$  is the maximum neighborhood radius size,  $r_{max}$  is referred to as the initial neighborhood size, and  $r_{min}$  is the minimum neighborhood radius size.

Step 4: The weight of the winner neuron and the weights of other neurons within the neighborhood of  $N_c(t)$  are adjusted using Equation (8). Meanwhile, the other weights retain the original value.

$$N_c(t) = \{t | find(norm(pos_t, pos_c)) < r\} \quad t = 1, 2, \dots, n \quad (7)$$

$$\omega_{ij} = \omega_{ij} + \eta \times (X_i - \omega_{ij}) \quad (8)$$

where  $\eta$  is the learning rate, which changes linearly in the training process just like the neighborhood size,  $pos_t$ ,  $pos_c$  are the positions of the neurons  $c$  and  $t$ , respectively, and “norm” stands for the Euclidean distance between two neurons as shown in Equation (9).

$$d_{norm} = \left[ (pos_{tx} - pos_{cx})^2 + (pos_{ty} - pos_{cy})^2 \right]^{0.5} \quad (9)$$



where  $pos_{tx}$  and  $pos_{ty}$  are the  $x$  coordinate value and the  $y$  coordinate value of neuron  $t$ , and  $pos_{cx}$  and  $pos_{cy}$  are the  $x$  coordinate value and the  $y$  coordinate value of neuron  $c$ .

Step 5: Adjust the learning rate and neighborhood radius. If  $s = T$ , then the algorithm ends, if not, go back to step 2, and new data are used to repeat the training steps.

The code for Kohonen neural networks can be downloaded on the website of <http://www.buaapress.com.cn/mzs/file/detail/id/442/key/037fdd19774875e6e8a9843838a62197>. The parameters set in the models were clarified in the paper. The structure of Kohonen neural network was changed in this paper shown in Figure 11, according to the feature of this case.

## 2.5. Support Vector Machine

Support vector machine (SVM) is a frequently used data-driven model for prediction and classification. It was developed by Vapnik [32]. SVM is formulated based on the structure risk minimization principle. It has a great advantage in solving nonlinear problems with a small sample. SVM has been widely used in hydrological prediction [33,34] and anomaly detection [23,35]. Figure 12 shows the structure of support vector machine.

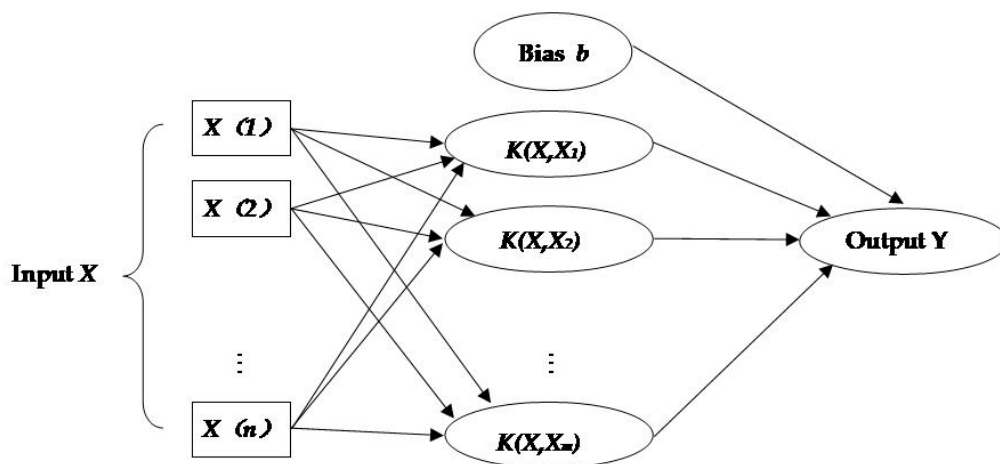


Figure 12. The normal structure of support vector machine.

The C-SVC model was selected in this study, which is a widely used two-class SVM model. The training process of C-SVC is as follows:

Consider a training set:

$$T = (x_1, y_1), \dots, (x_l, y_l) \in (X \times Y)^l \quad (10)$$

where  $x_i \in X = R^n$ ,  $y_i \in Y = \{-1, 1\}$ .  $x_i$  is the input vector, and  $y_i$  is the training label of the input vector.

A nonlinear (linear) function  $\phi(x)$  is used to map the input vector  $x_i$  from a  $d$ -dimensional input space to a  $d_h$ -dimensional feature space. The optimal hyperplane in the feature space is then defined as:

$$f(x) = w \times \phi(x) + b \quad (11)$$

where  $w$  is the weight vector and  $b$  is the bias. The optimal hyper-plane must obey the following rule:

$$y_i[w \times \phi(x_i) + b] \geq 1 \quad i = 1, 2, \dots, l. \quad (12)$$

The classifier can be written as:

$$f(x) = \text{sgn}(w \times \phi(x) + b). \quad (13)$$

The support vector machine aims at constructing a hyperplane with the maximal distance between the two classes. It has been proved that finding the maximum distance is equivalent to finding the minimum of  $\phi(w)$ :

$$\min \phi(w) = \frac{1}{2} \|w\|^2. \quad (14)$$

A slack variable  $\xi_i$  and a positive regularization constant  $C$  are introduced into Equation (15). The greater the value of  $C$  is, the less error can be tolerated. The problem of finding the optimal hyperplane can be generalized as follows:

$$\min \phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (15)$$

$$\text{s.t. } y_i[w \times \phi(x_i) + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, l. \quad (16)$$

The penalty magnitude of the classifier function is determined by the positive regularization constant  $C$ . If the value of  $C$  is too small, the accuracy of the prediction of the model will be too low. On the other hand, if the value of  $C$  is too large, the training model may be over-fit to the training data [36].

The function  $\phi(x_i)$  in the classifier function is given by the kernel function  $K(x_i, x_j)$ , and the most frequently used kernel functions are shown in Table 3, where  $g$  determines the spread of the kernel functions.

**Table 3.** A list of kernel functions.

Kernel Function	Formula
Linear	$K(x_i, x_j) = x_i^T x_j$
Polynomial	$K(x_i, x_j) = [g(x_i, x_j)]^3$
Radial basis function	$K(x_i, x_j) = \exp(-g \ x_i - x_j\ ^2)$
Sigmoid	$K(x_i, x_j) = \tanh(g(x_i, x_j) + r)$

### 3. Results and Discussion

Not every instance of damage to a spur dike can be detected [37], and slight damage may not affect the spur dike's regular operation. The damage to the spur dike was classified into three levels. The classification criteria were set empirically according to the damage condition of the spur dike in the flume experiment and the field observation as shown in Table 4. The 336 experiment data sets were categorized into these three damage levels.

**Table 4.** The empirical criteria for the damage level evaluation of the spur dike [24].

Damage level	3 General Damage	2 Critical Damage	1 Severe Damage
Inclination (°)	2–5	5–10	>10
Displacement (cm)	3–6	6–8	>8
Settlement (cm)	1–2.4	2.4–5	>5

A supervised training model (SVM) and an unsupervised KNN-SVM model were built to predict the damage levels. For the SVM model, 100 sets of data were selected randomly as training data, and the remaining 30 sets of data were used as a validation sample, as shown in Table 5. All of the 336 sets of experimental data were used to train the unsupervised Kohonen model.

**Table 5.** The data sets for validation [24].

Inclination (°)	Displacement (cm)	Settlement (cm)	Damage Level	Inclination (°)	Displacement (cm)	Settlement (cm)	Damage Level
16	5	6.4	1	15	4.9	6.4	1
15	5.2	7.7	1	18.3	5.8	8.1	1
17.5	5.6	7.6	1	3.2	1	3.8	3
2.2	0.5	4.4	3	3	0.5	3.1	3
10	3	4.9	2	19	6	7.1	1
19.1	6	6.9	1	19	6.5	7.3	1
17.8	6.2	7.5	1	7.5	2.1	5.5	2
17.8	6.2	7.7	1	8	2.4	5.6	2
8.5	2.9	5.7	1	15.3	4.7	7.8	1
2.2	0.4	3	3	3	0.5	3	3
7	2.5	5.5	1	10	3	5.7	1
18.1	6	7.8	1	19	6.5	7.5	1
17	5.6	7.3	1	5.3	1.2	4	2
15.4	5	6.1	1	16	4	7.9	1
19	5.9	8	1	5.8	1.4	3.9	2

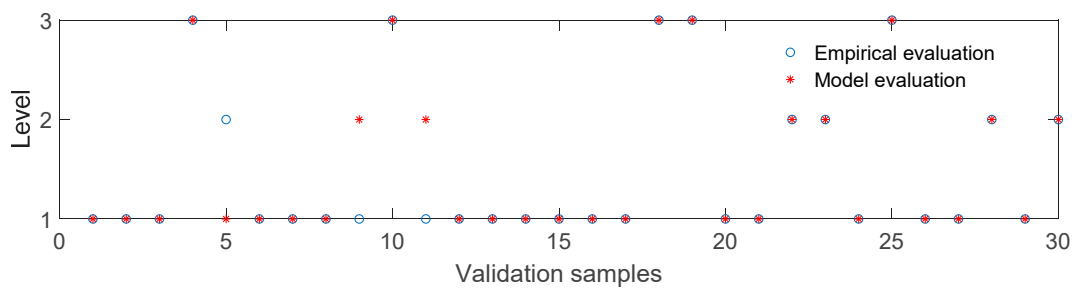
### 3.1. The Supervised Training Model

Choosing appropriate values for the positive regularization constant  $C$  and the spread  $g$  in the kernel function is the key to building an SVM model with high accuracy. In this paper, the values of  $C$  and  $g$  were selected using  $k$ -fold cross validation [38] (3-fold cross validation was used in this paper). The optimal parameters that generate the highest regression accuracy for different kernel functions are shown in Table 6. As we can see in the table, the models with a radial basis function (RBF) and a sigmoid function have higher regression accuracy for this research. If the value of  $C$  is too large, the training model may be over-fit to the training data [36]. So, the RBF kernel function and the corresponding optimal parameters were chosen to train the SVM model. In this paper, “libsvm”, which was an open source software developed by Chang and Lin [39], was selected to build the SVM model. The open source code for the SVM model can be downloaded on the web site of <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

**Table 6.** The model regression accuracies of different kernel functions.

Kernel Function	Accuracy	Parameters
Polynomial	95% (95/100)	-s 0 -v 3 -c 0.00098 -g 0.435275
Radial basis function	6% (96/100)	-s 0 -v 3 -c 0.435275 -g 0.25
Sigmoid	96% (96/100)	-s 0 -v 3 -c 21.1121 -g 0.0017

After the training process, the model was used to predict the damage levels of the validation sample. Figure 13 shows the comparison between the damage levels of the validation sample predicted by the model and evaluated according to the empirical criteria. Twenty-seven (27) samples were predicted correctly by the model, so the accuracy was 90%. This suggests that the support vector machine model can be an effective tool to predict the damage level of a spur dike. The damage levels of one data set predicted by the model were higher than the ones evaluated by the empirical criteria, and another two data sets were in the opposite situation. When the model is used for extrapolating the damage level of real spur dikes on the Yangtze River, the number of training data sets should be enough. No less than 50 sets of training data should be provided by the empirical criteria before the model training process.

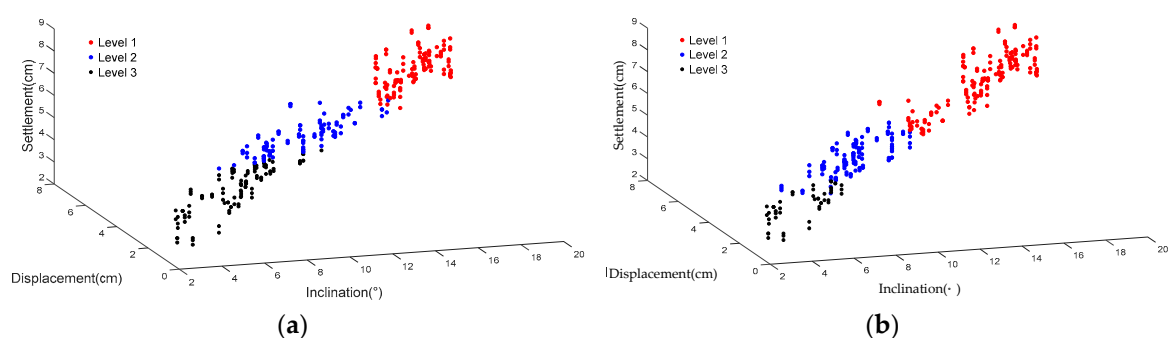


**Figure 13.** A comparison between the model prediction result and the empirical evaluation result.

### 3.2. The Unsupervised Model

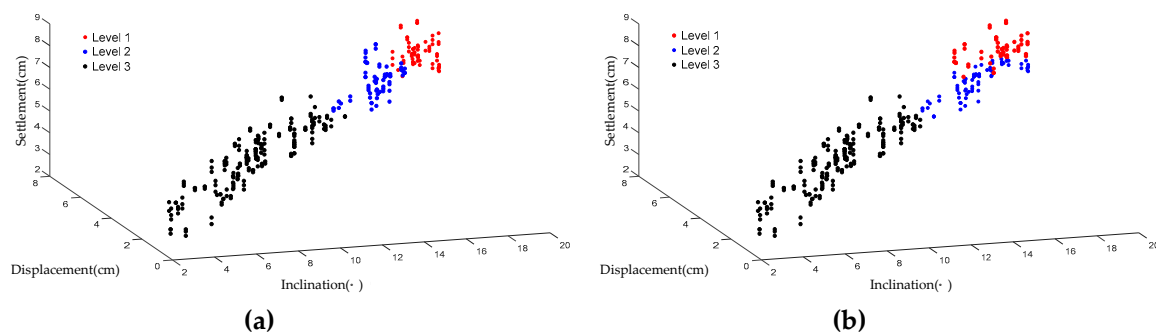
In most cases, measurement data are available but the classification criteria of the damage levels are unknown, so no training data are available to build the SVM prediction model. To solve this kind of problem, an unsupervised model was proposed by combining KNN and SVM. KNN is a data-mining algorithm that can classify data with a common feature through observing, analyzing, and comparing the data. In the KNN training process, the weights  $W_{ij}$  are randomly initialized, so the winner neuron (damage level) might be different for the same data set in different simulation processes. Then, the SVM model was introduced to mitigate the problem of the unsteady output of the KNN model. To obtain a steady result, the predicted results of KNN were selected as the training data of the SVM model in combination with the corresponding measured data (inclination, displacement, settlement); then, a steady damage level evaluation result was given by SVM.

During the training process of the KNN model, the learning rate  $\eta$  was reduced from 0.1 to 0.01, and the neighborhood radius was reduced from 1.5 to 0.4. Two thousand (2000) loops were run to adjust the learning rate and the neighborhood radius until the network converged. The number of training loops for initial covering of the input space was determined by the neighborhood radius and the learning rate. In this case, about after 200 training loops, a satisfactory result can be given when compared with the empirical damage level classification. If more training loops are set, then a more precise result will be given. After several training processes, the measurement data were classified by the KNN model. The results show that the damage level classification of the KNN model demonstrated two main classification patterns, as shown in Figures 14 and 15.



**Figure 14.** Two different predicted results of the damage level classification pattern (I): (a) and (b) were two different predicted results in different simulation processes.

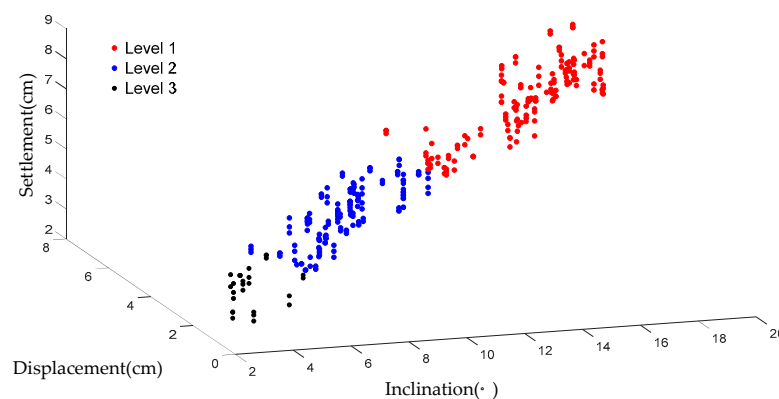
In the KNN training process, the weights  $W_{ij}$  were randomly initialized, so the winner neuron might be different for the same data set in different simulation processes, as can be seen in Figure 14a,b and Figure 15a,b. In the classification pattern (I), about half of the data sets were evaluated as Level 1, while in pattern (II), about half of the data sets were evaluated as Level 3. That indicates that the classification criteria in pattern (I) were much stricter than in pattern (II). For spur dikes, relatively strict classification criteria are helpful for maintaining the river training structures; therefore, the classification pattern (I) and its classification criteria were preferred for this research.



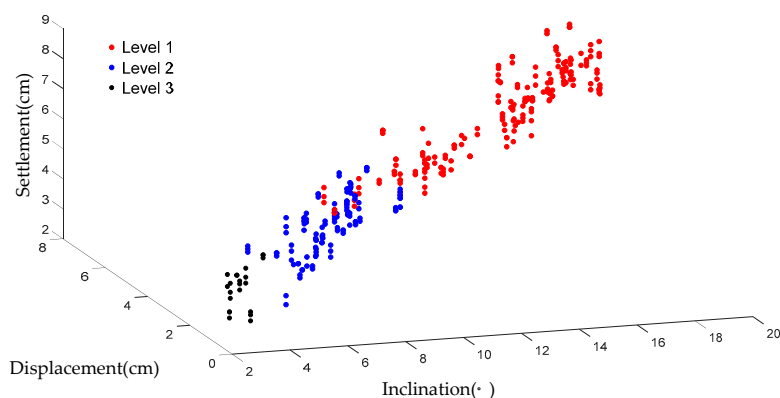
**Figure 15.** Two different predicted results of the damage level classification pattern (II): (a) and (b) were two different predicted results in different simulation processes.

With regard to the classification pattern (I), the results of different training processes with the same model parameters were not exactly the same, as shown in Figure 14a,b, especially for the data sets near the critical values of damage between different damage levels. A stable output of damage level prediction is necessary in practical engineering, so the predicted damage levels of the KNN model together with the corresponding experiment data were input to the SVM model to conduct supervised training. To train the SVM model, 50 sets of the classification pattern (I) results together with the corresponding experiment data were used as the training data of the SVM model. After the learning process, the damage levels were classified by the SVM model and are shown in Figure 16. The damage levels classified according to the empirical criteria are shown in Figure 17. It can be seen from the comparison between Figures 16 and 17 that most of the damage levels classified by the model and the empirical approach were the same. The data sets whose damage levels were evaluated differently by the model and the empirical approach lay in the critical range between different damage levels. This indicates that the classification criteria of the model were slightly different from the empirical classification criteria.

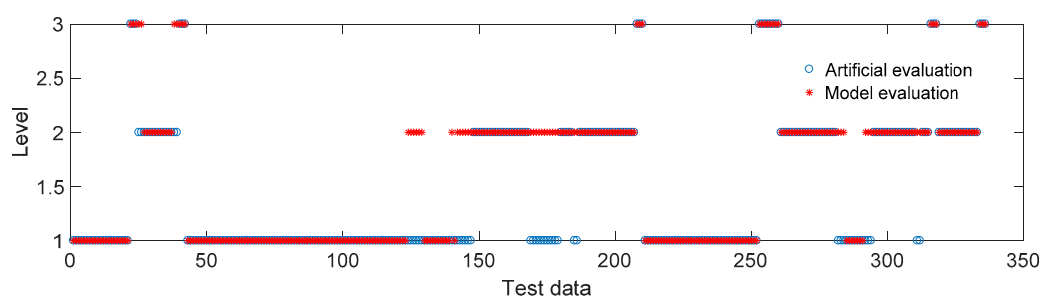
The comparison between the damage levels classified by the unsupervised model and the empirical approach is shown in Figure 18. The damage levels evaluated by the model and the empirical approach of 298 data sets, which was 88.7% of the 336 data sets in total, were the same. As for the remaining 38 data sets, most of the damage levels evaluated by the empirical approach were higher than the damage levels evaluated by the model as shown in Figure 18. This indicates that the empirical classification criteria (Table 4) were slightly stricter than the classification criteria of the unsupervised model. The potential risk of using the KNN-SVM model for extrapolating the damage level of other structures is the uncertain winner neuron for the same training data in different training processes, and this would lead to different classification results. Then, an empirical evaluation result needs to be provided as a reference.



**Figure 16.** The damage level classification result predicted by KNN-SVM.



**Figure 17.** The damage level classification result evaluated by the empirical criteria.



**Figure 18.** A comparison between the damage level classification result evaluated by the empirical approach and that predicted by the KNN-SVM model.

#### 4. Conclusions

A supervised support vector machine model was built based on the empirical damage level evaluation results. The high accuracy achieved in predicting the damage levels of the validation data sets indicated that the trained SVM model can be an effective tool for evaluating the practical damage level of spur dikes with tooth-shaped structures.

An advanced unsupervised evaluation method, namely the KNN-SVM model, was proposed. The data-mining algorithm Kohonen neural network was selected to initially classify the damage levels. Then, the SVM model was adopted to learn from the results predicted by KNN, and stable evaluation results were achieved. The prediction of the model agreed well with the empirical evaluation results. The KNN-SVM model can be an effective method to evaluate the damage level of spur dikes and other similar structures. Meanwhile, the result predicted by the unsupervised data-mining model could be a reference for the stability study of spur dikes with tooth-shaped structures.

In this paper, the model inputs only included the displacement, settlement, and inclination of the spur dike. In a further study, more parameters of the spur dike will be considered to develop a more practical KNN-SVM model to evaluate the damage level of a spur dike.

**Author Contributions:** Conceptualization, X.W. and Y.L.; methodology and software, X.W.; writing—original draft preparation, X.W.; writing—review and editing, Y.L., Z.W., S.M., and X.L.; supervision, Y.L.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant No. 51520105014) and the National Key R&D Program of China (Grant No. 2016YFC0402108 & Grant No. 2017YFC0405900).

**Acknowledgments:** We are particularly grateful to Prof. Xingnong Zhang and Dr. Xiaoxin Fei for providing all of the training data and the validation data for the machine learning models.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Jamieson, E.C.; Rennie, C.D.; Townsend, R.D. 3D Flow and Sediment Dynamics in a Laboratory Channel Bend with and without Stream Barbs. *J. Hydraul. Eng.* **2013**, *139*, 154–166. [[CrossRef](#)]
2. Vaghefi, M.; Ghodsian, M. Experimental study on scour around a T-shaped spur dike in a channel bend. *J. Hydraul. Eng.* **2012**, *138*, 471–474. [[CrossRef](#)]
3. Zhang, L.; Wang, P.; Yang, W.; Zuo, W.; Gu, X.; Yang, X. Geometric characteristics of spur dike scour under clear-water scour conditions. *Water* **2018**, *10*, 680. [[CrossRef](#)]
4. Koken, M.; Constantinescu, G. Flow and turbulence structure around a spur dike in a channel with a large scour hole. *Water Resour. Res.* **2011**, *47*, 3960–3976. [[CrossRef](#)]
5. Duan, J.G.; He, L.; Fu, X.; Wang, Q. Mean flow and turbulence around experimental spur dike. *Adv. Water Resour.* **2009**, *32*, 1717–1725. [[CrossRef](#)]
6. Nguyen, T.H.T.; Ahn, J.; Park, S.W. Numerical and physical investigation of the performance of turbulence modeling schemes around a scour hole downstream of a fixed bed protection. *Water* **2018**, *10*, 103. [[CrossRef](#)]
7. Bassar, H.; Karami, H.; Shamshirband, S.; Jahangirzadeh, A.; Akib, S.; Saboohi, H. Predicting optimum parameters of a protective spur dike using soft computing methodologies—A comparative study. *Comput. Fluids* **2014**, *97*, 168–176. [[CrossRef](#)]
8. Jiao, Z.; Dou, X.; Zheng, J.; Zhang, X.; Gao, X. Influence of spur dike on hydrodynamic exchange between channel and shoal of generalization estuary in physical model test. *China Ocean Eng.* **2017**, *31*, 624–630. [[CrossRef](#)]
9. Hudson, Y. Laboratory investigation of rubble-mound breakwater. *J. Waterw. Harbors Div. ASCE* **1959**, *85*, 93–122.
10. Thompson, D.M.; Shuttler, R.M. *Riprap Design for Wind-Wave Attack, A Laboratory Study in Random Waves*; HR Wallingford: Oxfordshire, UK, 1975.
11. Hanzawa, M.; Sato, H.; Takahashi, S.; Shimosako, K.; Takayama, T.; Tanimoto, K. New stability formula for wave-dissipating concrete blocks covering horizontally composite breakwaters. *Coast. Eng.* **1996**, *1997*, 1665–1678.
12. Meer, J.W.V.D. *Rock Slopes and Gravel Beaches Under Wave Attack*; Delft Hydraulics: Delft, The Netherlands, 1988.
13. Kajima, R. A new method of structurally resistive design of very important seawalls against wave action. In Proceedings of the International Workshop on Wave Barriers in Deepwaters, Yokosuka, Japan, 10–14 January 1994.
14. Kaku, S.; Kobayashi, N.; Ryu, C.R. Design formulas for hydraulic stability of rock slopes under irregular wave attack. In Proceedings of the 38th Japanese Conference Coastal Engineering, Tokyo, Japan, 1991; pp. 661–665.
15. Han, L.; Wang, P.; Liu, H.; He, Q. Discriminatory analysis of rockfill groin's safety based on water-destruction volumetric ratio. *Eng. J. Wuhan Univ.* **2014**, *47*, 201–206. (In Chinese)
16. Han, L.; Wang, P.; Wei, S. Quantitative method and regulating effect evaluation of washout for X-shaped beach protection flexible mattress. *Eng. J. Wuhan Univ.* **2017**, *50*, 487–499. (In Chinese)
17. Niu, Z.; You, R.; Lu, J. The Dike stability analysis based on strength reduction method. *Procedia Eng.* **2012**, *28*, 560–563.
18. Mase, H.; Sakamoto, M.; Sakai, T. Neural network for stability analysis of rubble-mound breakwaters. *J. Waterw. Port Coast. Ocean Eng.* **1995**, *121*, 294–299. [[CrossRef](#)]
19. Dong, H.; Park, W.S. Neural network for design and reliability analysis of rubble mound breakwaters. *Ocean Eng.* **2005**, *32*, 1332–1349.
20. Erdik, T. Fuzzy logic approach to conventional rubble mound structures design. *Expert Syst. Appl.* **2009**, *36*, 4162–4170. [[CrossRef](#)]
21. Mandal, S.; Rao, S.; Harish, N. Damage level prediction of non-reshaped berm breakwater using ANN, SVM and ANFIS models. *Int. J. Nav. Archit. Ocean Eng.* **2012**, *4*, 112–122. [[CrossRef](#)]
22. Harish, N.; Mandal, S.; Rao, S.; Patil, S.G. Particle swarm optimization based support vector machine for damage level prediction of non-reshaped berm breakwater. *Appl. Soft Comput.* **2015**, *27*, 313–321. [[CrossRef](#)]
23. Fisher, W.D.; Camp, T.K.; Krzhizhanovskaya, V.V. Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection. *J. Comput. Sci.* **2017**, *20*, 143–153. [[CrossRef](#)]

24. Fei, X. *Stability Mechanism and Criterion of Typical Regulation Structures in the Lower-Middle Reaches of the Yangtze River*; Nanjing Hydraulic Research Institute: Nanjing, China, 2017. (In Chinese)
25. Yan, J.; Xia, Y.; Zhang, W.; Hua, X.; Zhao, Z. The numerically simulation on drag force of the orthogonal spur dike with toothed mixture embankment. *Port Waterw. Eng.* **2017**, *4*, 1–7. (In Chinese)
26. Cai, Z.; Xia, Y.; Hua, X.; Yan, J. Flow force characteristics of new-type structure submerged spur dike during deep waterway regulation. *Hydro-Sci. Eng.* **2018**, *3*, 16–23. (In Chinese)
27. Tong, D.S.; Lin, B.I.; Hou, S.Q.; Wang, H.L. Experimental study on new type component of dike body of regulating structure in phase I of the Yangtze River 12.5 m deep-water channel project. *Chin. Harbour Eng.* **2014**, *35*, 39–43. (In Chinese)
28. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [[CrossRef](#)]
29. Hagan, M.T.; Demuth, M.H.; Beale, M.H.; Jesús, O.D. *Neural Network Design*; China Machine Press: Beijing, China, 2002; p. 357.
30. Kohonen, T. *Self-Organizing Maps*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 266–270.
31. Waller, N.G.; Kaiser, H.A.; Illian, J.B.; Manry, M. A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. *Psychometrika* **1998**, *63*, 5–22. [[CrossRef](#)]
32. Vapnik, V.N. The Nature of statistical learning theory. *Technometrics* **1997**, *38*, 409.
33. Yoon, H.; Jun, S.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [[CrossRef](#)]
34. Gong, Y.; Wang, Z.; Xu, G.; Zhang, Z. A comparative study of groundwater level forecasting using data-driven models based on ensemble empirical mode decomposition. *Water* **2018**, *10*, 730. [[CrossRef](#)]
35. Fisher, W.; Camp, T.; Krzhizhanovskaya, V. crack detection in earth dam and levee passive seismic data using support vector machines. *Procedia Comput. Sci.* **2016**, *80*, 577–586. [[CrossRef](#)]
36. Belousov, A.I.; Frese, S.A.V.V. A flexible classification approach with optimal generalization performance: support vector machines. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 15–25. [[CrossRef](#)]
37. Broderick, L.L. *Riprap Stability Versus Monochromatic and Irregular Waves*; Oregon State University: Corvallis, OR, USA, 1984.
38. Anguita, D.; Ghio, A.; Ridella, S.; Sterpi, D. K-Fold Cross Validation for Error Rate Estimate in Support Vector Machines. In Proceedings of the International Conference on Data Mining, Las Vegas, NV, USA, 13–16 July 2009; pp. 291–297.
39. Chang, C.; Lin, C. LIBSVM: A library for support vector machines. *TIST* **2011**, *2*, 1–27. [[CrossRef](#)]

