

## Research Article

# Predicting Rainfall-Induced Soil Erosion Based on a Hybridization of Adaptive Differential Evolution and Support Vector Machine Classification

Tuan Vu Dinh <sup>1,2</sup>, Hieu Nguyen <sup>2,3</sup>, Xuan-Linh Tran <sup>2,4</sup> and Nhat-Duc Hoang <sup>2,4</sup>

<sup>1</sup>Faculty of Environmental Sciences, VNU University of Science, Vietnam National University, Hanoi, Vietnam

<sup>2</sup>Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

<sup>3</sup>Faculty of Natural Sciences, Duy Tan University, Da Nang 550000, Vietnam

<sup>4</sup>Faculty of Civil Engineering, Duy Tan University, Da Nang 550000, Vietnam

Correspondence should be addressed to Tuan Vu Dinh; [vudinhtuan@vnu.edu.vn](mailto:vudinhtuan@vnu.edu.vn)

Received 18 December 2020; Revised 21 January 2021; Accepted 28 January 2021; Published 20 February 2021

Academic Editor: Essam Houssein

Copyright © 2021 Tuan Vu Dinh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soil erosion induced by rainfall is a critical problem in many regions in the world, particularly in tropical areas where the annual rainfall amount often exceeds 2000 mm. Predicting soil erosion is a challenging task, subjecting to variation of soil characteristics, slope, vegetation cover, land management, and weather condition. Conventional models based on the mechanism of soil erosion processes generally provide good results but are time-consuming due to calibration and validation. The goal of this study is to develop a machine learning model based on support vector machine (SVM) for soil erosion prediction. The SVM serves as the main prediction machinery establishing a nonlinear function that maps considered influencing factors to accurate predictions. In addition, in order to improve the accuracy of the model, the history-based adaptive differential evolution with linear population size reduction and population-wide inertia term (L-SHADE-PWI) is employed to find an optimal set of parameters for SVM. Thus, the proposed method, named L-SHADE-PWI-SVM, is an integration of machine learning and metaheuristic optimization. For the purpose of training and testing the method, a dataset consisting of 236 samples of soil erosion in Northwest Vietnam is collected with 10 influencing factors. The training set includes 90% of the original dataset; the rest of the dataset is reserved for assessing the generalization capability of the model. The experimental results indicate that the newly developed L-SHADE-PWI-SVM method is a competitive soil erosion predictor with superior performance statistics. Most importantly, L-SHADE-PWI-SVM can achieve a high classification accuracy rate of 92%, which is much better than that of backpropagation artificial neural network (87%) and radial basis function artificial neural network (78%).

## 1. Introduction

Soil erosion induced by water is the main culprit of the degradation of upland and mountain ecosystems [1]. The erosion process poses a threat to the capacity of the land to provide ecosystem services that are needed to reach the Sustainable Development Goal (SDG) target 15.3 [2]. Currently, soil erosion is a critical concern in many regions; this issue must be prevented to achieve sustainable development goals [3, 4]. Soil erosion is characterized by a complex and dynamic process involving detachment, transport, and deposition of soil material. There are many

factors affecting erosion magnitude, namely, climate, soil type, soil structure, vegetation, and cropping on top and especially land management [5].

In tropical areas, such as Northwest Vietnam, soil erosion potential is high due to heavy rainfall and is currently accelerating under maize monocropping in the uplands [6, 7]. A recent increase in maize monocropping in the region is caused by market demand for the rapid growing livestock and poultry industry. The majority of the maize area expansion has been achieved by encroachment into steep forested mountain watersheds. Land preparation for a maize cropping season employed a weeding prior to

burning, followed by several ploughs before sowing seeds. The practice is carried out at the onset of the monsoon rains; by the time, bare fields are exposed to high-intensity rains. This often results in high runoff leading to large erosion and longer-term degradation with declining crop production on-site and strong environmental impacts off-site [8, 9]. Erosion control measures are based on the following: covering the soil to protect it from raindrop impact; increasing the infiltration capacity of the soil to reduce runoff; improving the aggregate stability of the soil; and increasing surface roughness to reduce the velocity of runoff [10–12]. Many attempts have been tested worldwide: contour ploughing or contour farming using stone, wood, or vegetation barriers/hedgerow; cover crop; minimum tillage or zero tillage; and mulching. Such measures can be effective in different geographic and climatic conditions under various soil characteristics and land management systems. Among these, land management is crucial in controlling erosion [11, 13–16].

Cerdà et al. [17] investigate the hydrological and erosional impact as well as farmer's perception on catch crops and weeds in citrus organic farming; this study recommends that farmers should be informed and given initial subsidies for implementing new measures for improving soil quality and preventing soil erosion. In the study of Tuan et al. [18], kinetic energy of rainfall is found to be one of the most important factors that cause soil erosion, particularly when heavy rains coincide with poor ground cover at the beginning of the cropping season. However, it should be noted that the importance of a factor varies from case to case and other comprehensive variables affecting erosion should also be studied.

Soil loss studies at the plot scale have been of crucial importance to identify the mechanism of the processes. The erosion plot experiments can help to introduce new erosion prevention as it provides access to dependable and faithful erosion measurements and large numbers of data necessary to test new models [19]. A number of empirical erosion models such as USLE and RUSLE [20], SWAT [21], and WEPP [22] used data from these types of studies for prediction. However, the performance and accuracy are not as good as expected [23].

In the last decades, machine learning-based models have proved to be a helpful alternative to deal with the multivariate and complex nature of the phenomena in various disciplines of engineering [24–38]. Optimized kernel logistic regression models were employed in [39] for landslide susceptibility assessment. The evidential belief function-based function tree (FT), logistic regression, and logistic model tree were utilized for predicting landslide occurrences in [40]. Ensemble learning methods used for natural hazard risk assessment have been proposed in [41–44].

In addition, ANN is used to estimate the soil erodibility in Malaysia [45] and to conclude that rainfall and runoff are important factors affecting the amount of soil eroded [46]. In [47], Kohonen neural networks (KNNs) are utilized to model runoff erosion with a better outcome than that of the conventional multiple linear regression models. It was also shown that ANN is generally more competitive than the

WEPP model in quantitative prediction of soil loss [48]. Albaradeia et al. [49] confirm that while WEPP underestimated the soil loss, ANN provided results that are in line with observation. Recently, Vu et al. [50] successfully constructed a machine learning model based on multivariate adaptive regression splines for predicting soil loss occurrences.

Based on the literature review, it can be seen that advanced machine learning methods can be a good option for soil erosion modeling. Even though previous research works extensively relied on ANN, other advanced machine learning algorithms are worth explored because the prediction of soil erosion is both complex and dynamic. We are particularly interested in the support vector machine (SVM) [51–54] because it usually has comparable results with ANN and has a lower risk of overfitting. In addition, there has been a great success in using SVM for different but closely related hydrological and geological problems [55–61].

Furthermore, to enhance the predictive performance of the SVM, this study relies on the approach of using metaheuristic approaches. It is noted that the process of fine-tuning a machine learning model can be formulated as a global optimization problem. In recent years, various metaheuristic approaches have been successfully employed, including monarch butterfly optimization [62, 63], slime mould algorithm [64], moth search algorithm [65, 66], Harris hawks optimization [67–70], differential flower pollination [71], symbiotic organisms search [72], Henry gas solubility optimization [73], and satin bowerbird optimizer [74]. As can be seen from the literature, there is an increasing trend of hybridizing metaheuristics and machine learning to tackle complex problems in the field of engineering.

Therefore, the current study aims at extending the body of knowledge by establishing soil erosion prediction models for tropical hilly regions based on an integration of the history-based adaptive differential evolution with linear population size reduction and population-wide inertia term (L-SHADE-PWI) metaheuristic and the support vector machine pattern classification method. A dataset, featuring ten explanatory variables, collected from plot experiments in Son La province (Vietnam) is employed to construct and verify the prediction models. In this study, the problem of erosion status prediction is formulated as a pattern classification task within which a vector of explanatory variables is assigned to either “erosion” or “nonerosion.” We hypothesize that the structure of the L-SHADE-PWI-SVM, an integration of machine learning and metaheuristic optimization, is capable of predicting rainfall-induced soil erosion.

In summary, the innovative points of the current study can be stated as follows:

- (i) A novel integration of the L-SHADE-PWI metaheuristic and the SVM pattern classifier is first proposed to predict the complex phenomenon of soil erosion.
- (ii) The effort to fine-tune the machine learning model is eliminated via the utilization of metaheuristic optimization.

- (iii) The model construction phase of the proposed L-SHADE-PWI-SVM is entirely data-driven. Therefore, it is convenient to be used by land-use planners and hazard prevention agencies without much domain knowledge in machine learning.
- (iv) The superiority of the proposed hybrid approach is validated using experiments based on a repeated subsampling of the collected data.

The rest of the paper is organized as follows. In Section 2, we review the related research methods. In Section 3, we present the proposed L-SHADE-PWI-SVM model. Section 4 is devoted to experiments and result comparisons. We end with the conclusions in Section 5.

## 2. Material and Methods

*2.1. General Description of the Study Area and the Collected Dataset.* For this study, data in Son La province (Northwest Vietnam) were collected from 2009 to 2011 at two experimental sites, which are two catchments with bounded plots. Figure 1 shows the location of the study area in the maps of Son La and Vietnam. The area is governed by tropical monsoon climate with mild and generally warm temperature (the annual average is 21°C, and the lowest and highest average temperatures are 16°C in February and 27°C in August, respectively [75]). The winter (from November to March) is relatively dry while the summer (from May to October) can have a lot of rain.

At the experimental sites, erosion plots of 72 m<sup>2</sup> (4 m × 18 m) with boundaries were installed to avoid the exchange of water runoff from outside. There were 24 of them in total: four treatments, three replicates arranged in a completely randomized block design. In order to gather deposited sediment from soil erosion at the plots, a system of buckets was employed. Erosion data were collected on storm basic during 3 years from 2009 to 2011. Detailed information on soil, crop, and field managements can be found in the previous study of Tuan et al. [18].

There are many factors contributing towards rill and interrill soil erosion induced by raindrop impact and surface runoff. They include climate, soil, topography, and land use. The amount and intensity of rainfall affect erosivity, while soil properties, land use, and topography (described via slope length and steepness) can have a big influence on the degree of erodibility. To represent these factors, a set of ten explanatory variables has been chosen. In Table 1, these variables are listed together with their statistics.

First,  $I_{30}$  is the prolonged peak rates of detachment and runoff in 30 minutes (30 min intensity). The storm energy  $E$  is calculated according to the following formula [76]:

$$E = 1099[1 - 0.72 \exp(-1.27i)], \quad (1)$$

where  $i$  is the maximum intensity of 30 minutes. The product of  $E$  and  $I_{30}$  gives us the kinetic rainfall energy ( $EI$ ) which is the combined interaction of total energy and peak intensity in each particular storm. This quantity represents how particle detachment is combined with transport capacity [77].

The second and third variables are topographic factors, namely, slope length and steepness. They all have a great influence on soil loss. The bigger the slope is, the higher potential of soil loss will be. Similarly, the larger the slope length is, the more likely there will be soil erosion as there is a greater accumulation of runoff. In this study, a Nikon Forestry 550 Inclinometer was used to measure the slope of the plots.

Soil erodibility is heavily influenced by texture, but permeability, structure, and organic matter also play a role. In this study, wet analysis [78] was used to analyze soil texture with a two pseudoreplicate plot. In addition, it is known that aggregate stability is the index that is closely related to soil erodibility. However, there is currently no efficient method to calculate this index. Therefore, OC and pH, two simple characteristics of the soil, were used instead. Total OC was obtained using a C/N analyzer with carbonate content removed by HCl. pH was measured using a glass electrode with a soil-to-water ratio of 1 : 2.5.

The final variable chosen is ground cover rate as this is one of the most influential factors of soil erosion rate in tropical areas [18]. For more information on how this variable is measured, we refer the reader to Tuan et al. [18].

In our model to predict soil erosion, we aim to classify each sample, which is associated with a vector of values for the explanatory variables, into two classes: “erosion” or “nonerosion.” In order to provide ground truth classification to train the model, we follow the same criterion on soil loss as in [79]: “erosion” label is assigned to any sample with more than 3 tons per hectare soil loss, and other records are classified as “nonerosion.” A total of 236 data samples have been collected within which 118 records were classified as “erosion.”

### *2.2. Support Vector Machine (SVM) for Pattern Classification.*

The original SVM algorithm was first introduced in [80] for linear binary classification. The algorithm builds a hyperplane to separate positive and negative samples with the margin as large as possible. However, in practice, it is often the case that samples are not linearly separable and such hyperplane does not exist. This can lead to the poor performance of the algorithm. Accordingly, the original SVM algorithm is extended for nonlinear classification via the use of kernel functions [81]. Kernel functions, which are often nonlinear, map the space of input variables into a much higher-dimensional space, where separable hyperplanes are more likely to exist [82, 83]. In addition, soft margin is proposed to handle the case the hyperplane does not exist even in the higher-dimensional space [53]. With these improvements, the SVM has become one of the most popular and successful classification algorithms [84–86].

Given a training set consists of  $N$  samples  $\{x_k, y_k\}_{k=1}^N$ , where  $x_k \in R^n$  is a feature or input vector and  $y_k \in \{-1, +1\}$  is the corresponding response or class label. The training phase of the SVM algorithm can be mathematically formulated as the following optimization problem [53, 87].

Find  $w, b, e$  to minimize

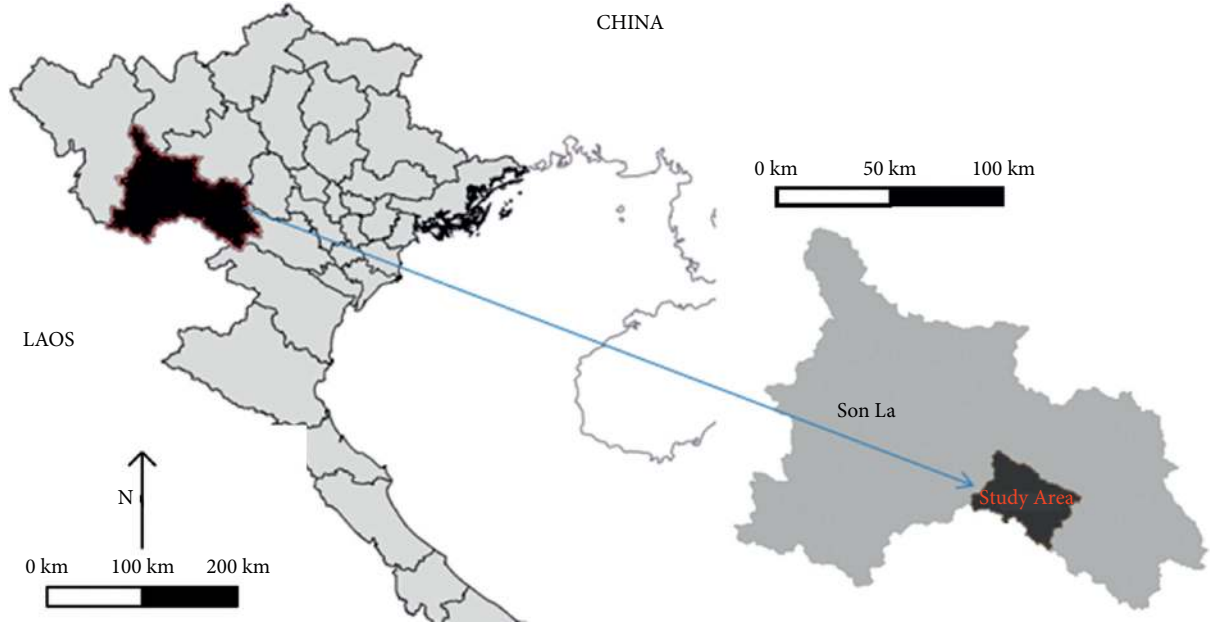


FIGURE 1: Location of the study area.

TABLE 1: Statistical descriptions of soil erosion influencing factors.

| Influencing factors                     | Notation | Min   | Mean   | Std.   | Max     |
|---|----------|-------|--------|--------|---------|
| EI30                                    | $X_1$    | 0.00  | 134.77 | 385.10 | 3008.93 |
| Slope degree                            | $X_2$    | 24.83 | 28.85  | 2.39   | 34.77   |
| OC topsoil (%)                          | $X_3$    | 0.89  | 1.86   | 0.56   | 2.79    |
| pH topsoil                              | $X_4$    | 5.13  | 5.75   | 0.56   | 7.06    |
| Bulk density ( $\text{g}/\text{cm}^3$ ) | $X_5$    | 1.23  | 1.39   | 0.08   | 1.58    |
| Topsoil porosity (%)                    | $X_6$    | 46.34 | 52.36  | 3.15   | 59.48   |
| Topsoil texture (silt fraction %)       | $X_7$    | 31.35 | 33.82  | 1.48   | 37.71   |
| Topsoil texture (clay fraction %)       | $X_8$    | 18.61 | 30.03  | 4.67   | 38.35   |
| Topsoil texture (sand fraction %)       | $X_9$    | 29.66 | 36.15  | 4.12   | 46.51   |
| Soil cover rate (%)                     | $X_{10}$ | 1.05  | 53.51  | 25.36  | 99.71   |

Note: EI denotes energy time intensity. Topsoil refers to the upper layer 0–20 cm of soil, and its characters are indicated by asterisk “\*”. The soil bulk density (BD), also known as dry bulk density, is the weight of dry soil ( $M_{\text{solids}}$ ) divided by the total soil volume ( $V_{\text{soil}}$ ). Porosity is a measure of the total pore space in the soil.

$$J_p(w, e) = \frac{1}{2}w^T w + C \frac{1}{2} \sum_{k=1}^N e_k^2, \quad (2)$$

subjected to

$$y_k(w^T \varphi(x_k) + b) \geq 1 - e_k, \quad k = 1, \dots, N, e_k \geq 0, \quad (3)$$

where the first two outcomes  $w \in R^n$  and  $b \in R$  define the classification hyperplane (they are a normal vector and the corresponding offset, respectively); the third outcome  $e$  is the vector of slack variables introduced to handle the case data cannot be separated without error;  $C$  is the penalty constant decides how much weight should be put on

classification error; and  $\varphi(x)$  denotes the nonlinear mapping from the input space to a much higher-dimensional space.

The explicit formula of  $\varphi(x)$  is not normally required. However, one needs to know the dot product of  $\varphi(x)$  in the higher-dimensional space. It is referred to as the kernel function given by

$$K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l), \quad x_k, x_l \in R^n. \quad (4)$$

There are different choices of kernel functions, such as linear, polynomial, sigmoid, and radial basis functions. In this work, radial basis function (RBF) is employed due to its good performance in pattern recognition tasks [88]. It is defined as

$$K(x_k, x_l) = \exp\left(-\frac{\|x_k - x_l\|^2}{2\sigma^2}\right), \quad (5)$$

where  $x_k, x_l \in R^n$ , and  $\sigma$  is a parameter that can be tuned.

Once training is finished, the final classification is performed as follows:

$$y(x_l) = \text{sign}\left(\sum_{k=1}^{SV} \alpha_k y_k K(x_k, x_l) + b\right), \quad (6)$$

where  $\alpha_k$  is the solution of the dual problem of equations (2) and (3) and  $SV$  is the number of support vectors (the number of  $\alpha_k > 0$ ).

*2.3. The History-Based Adaptive Differential Evolution with Linear Population Size Reduction and Population-Wide Inertia (L-SHADE-PWI).* From Section 2.2, it is obvious that, before using the SVM algorithm, appropriate values of the penalty coefficient ( $C$ ) and the parameter in the kernel function ( $\sigma$ ) need to be determined. These hyperparameters play an important role in the learning phase of the model. As a consequence, they can have essential influences on the predictive capability of the soil erosion prediction model. The step of selecting these hyperparameters is usually referred to as model selection. It is problem-dependent. It is also very challenging because the hyperparameters must be selected from continuous domains, where there is an infinite number of possible choices [33, 89–92].

Traditional approaches for model selection include grid search and random search. In grid search, hyperparameters are exhaustively searched from a specified subset of the space of hyperparameters. Random search, on the other hand, randomly select combinations of hyperparameters for consideration. Both grid search and random search are simple and can be done in parallel. However, they cannot deliver desired predictive accuracy since hyperparameter selection is highly data-dependent and complex. For the case of an SVM model, the complexity coefficient ( $C$ ) and the kernel function parameter ( $\gamma$ ) are both real numbers. Hence, there are an infinite number of possible combinations of  $C$  and  $\gamma$ . This means that an exhaustive grid search is infeasible and researchers have turned to metaheuristic methods to address the model selection problem [72, 91, 93, 99].

In this work, we propose to use the history-based adaptive differential evolution with linear population size reduction and population-wide inertia (L-SHADE-PWI) algorithm to select the hyperparameters for the SVM algorithm. The L-SHADE-PWI [100] is a newly proposed metaheuristic with promising optimization capability. With the L-SHADE-PWI algorithm, the model selection of an SVM model is formulated as a global optimization problem where the objective function measures how well the model fits the training and validating sets using a set of the hyperparameters of the penalty coefficient ( $C$ ) and the kernel function parameter ( $\gamma$ ).

The L-SHADE-PWI is essentially an extension of the history-based adaptive differential evolution with linear population size reduction (L-SHADE) algorithm proposed

in [101]. This newly developed method inherits the effective mutation-crossover operator of the standard differential evolution (DE), the adaptive tuning strategy of the L-SHADE, the DE/current-to-pbest/1 mutation strategy described in [102], the linear population size reduction proposed in [101], and a population-wide inertia term (PWI) incorporated in the mutation operation [100].

A brief summary of the L-SHADE-PWI algorithm is presented in Algorithm 1. In the heart of the algorithm, there are three main steps, including mutation, crossover, and selection. The mutation step generates a trial vector. The mutation operator requires the computing of the population-wide inertia term  $PWI_{g-1}$ . The PWI term indicates an averaged direction and size of successful moves that lead to better solutions in the preceding generation. The incorporation of the population-wide inertia aims at enhancing the variations of population members in the potential direction that on average brought about cost function's reduction [100]. The PWI term is calculated as follows [100]:

$$PWI_{g-1} = \frac{\sum_{i=1}^{PS_{g-1}} mv_{i,g-1}}{NI_{g-1}}, \quad (7)$$

$$mv_{i,g-1} = u_{i,g-1} - x_{i,g-1}, \quad \text{if } f(u_{i,g-1}) \leq f(x_{i,g-1}), \quad (8)$$

$$mv_{i,g-1} = 0, \quad \text{if } f(u_{i,g-1}) > f(x_{i,g-1}), \quad (9)$$

where  $mv_{i,g-1}$  denotes the move of a population member in the search space occurred in generation  $g-1$ . It is worth noticing that individual moves that do not lead to better cost function values are not taken into account. Therefore, the sum described in equation (7) is divided by  $NI_{g-1}$  which represents the number of beneficial moves [100].

Accordingly, this trial vector  $v$  is used in the crossover step to create a mutated vector  $u$ . The final step compares the mutated vector with its parent and replaces them in the next generation if better target is achieved. In Algorithm 1,  $PS$  is the population size;  $D$  represents the number of the searched variables (two in our case);  $CF(x)$  is the cost function; and  $x$  denotes the vector of the searched variables. The two vectors of length  $H$  MF and MCR are archives containing the mean values of the mutation scale ( $F$ ) and the crossover probability ( $CR$ ). The two vectors SF and SCR keep  $CR$  and  $F$  values delivering offspring better than the parent.  $G_{max}$  is the limit (maximum number) of generations. Finally, a scheme for population size reduction is implemented to facilitate the convergence rate of the metaheuristic algorithm.

### 3. The Proposed L-SHADE-PWI Optimized SVM for Rainfall-Induced Soil Erosion Prediction

This section aims at describing the structure of the L-SHADE-PWI-SVM model used for rainfall-induced soil erosion. The newly developed model is an integration of machine learning and metaheuristic optimization. In detail, the SVM machine learning is employed to generalize a decision boundary that

```

Define PS, D, CF(x), Gmax, an archive A = ∅
Define the two archives of MF and MCR and the two sets of SF and SCR
Create a population P with PS members randomly
Identify and record xbest,g which is currently the best solution
For each generation g
  For each individual xi,g
    Select MFk and CRk randomly from MF and MCR
    Calculate Fi = Gaussian(MFk, 0.1)
    Calculate CRi = Gaussian(CRk, 0.1)
    Select a xpbest,g randomly from A
    Select two random members xr1,g and xr2,g
    Calculate the population-wide inertia term PWIg-1
    Create a trial vector according to [100]:
    vi,g+1 = xi,g + Fi(xr1,g - xr2,g) + Fi(xpbest,g - xi,g) + Fi × PWIg-1
    Perform crossover according to [103]:
    uj,i,g+1 = { vj,i,g+1, if randj ≤ Cr or j = rnb(i)
                xj,i,g, if randj > Cr and j ≠ rnb(i)
    Carry out the selection operator [103]
  End For
  Collect successful Fi and CRi
  Update SF, SCR, MF, MCR, and A [104]
  Update PS [101]
End For
Return xbest,g

```

ALGORITHM 1: Pseudocode of the L-SHADE-PWI algorithm.

separates the input space into two distinctive domains: non-erosion (the negative class) and erosion (the positive class). The factors of EI30, slope, OC topsoil, pH topsoil, bulk density, total pore volume, texture-silt, texture-clay, texture-sand, and soil cover rate are employed as influencing factors for pattern classification. Because the model establishment phase of the SVM necessitates an appropriate specification of the penalty coefficient and the kernel function parameter, the proposed model relies on the L-SHADE-PWI to optimize the hyperparameter selection phase.

Figure 2 demonstrates the overall structure of the proposed L-SHADE-PWI-SVM model. A software program based on the model structure has been developed by the authors in Visual C#.NET (framework 4.6.2) environment and the Accord.NET Framework [105]. The newly established model has been tested on the ASUS FX705GE-EW165T (Core i7 8750H and 8 GB RAM) platform.

It is proper to note that the influencing factors of the collected dataset have been randomly separated into a training (90%) dataset and a testing dataset (10%). The first set is used for model construction and the second set is reserved for model validation. Moreover, the Z-score equation has been used to standardize the data range since it may enhance the classification performance [106]. The Z-score equation is given by

$$X_Z = \frac{X_D - M_X}{STD_X}, \quad (10)$$

where  $X_Z$  and  $X_D$  are the normalized and the original variables, respectively.  $M_X$  and  $STD_X$  denote the mean value and the standard deviation of the soil erosion influencing factors, respectively.

As described earlier, the model training and soil erosion prediction phases of the SVM necessitate an appropriate set of the penalty coefficient and the kernel function parameter. The first hyperparameter specifies how the SVM's loss function increases due to misclassified data samples. The second hyperparameter influences the smoothness of the classification boundary. Thus, these two hyperparameters strongly affect the generalization and predictive accuracy of the SVM-based soil erosion prediction model. This research proposes to use the L-SHADE-PWI metaheuristic for optimizing the performance of the SVM model. At the first generation, the L-SHADE-PWI initializes a population of hyperparameters randomly. In each subsequent generation, this metaheuristic approach gradually explores and exploits the search space to identify potential regions which contains high-quality sets of the SVM model's hyperparameters.

Relying on the hyperparameters found by the L-SHADE-PWI metaheuristic, the SVM model analyzes the training dataset and generalizes a decision boundary that separates input samples associated with nonerosion and erosion classes. It is noted that this study employs the Accord.NET Framework [105] to carry out the SVM model's training and prediction phases. Furthermore, to optimize the SVM-based soil erosion prediction model, this research relies on a  $K$ -fold cross-validation with  $K = 5$ . Based on the cross-validation framework, the dataset is separated into 5 mutually exclusive sets. In each of the five runs, one set is utilized for model verification and the other four sets are used as training samples. The average predictive performance obtained from the 5 folds is used to express the generalization capability of the rainfall-induced soil erosion prediction model.

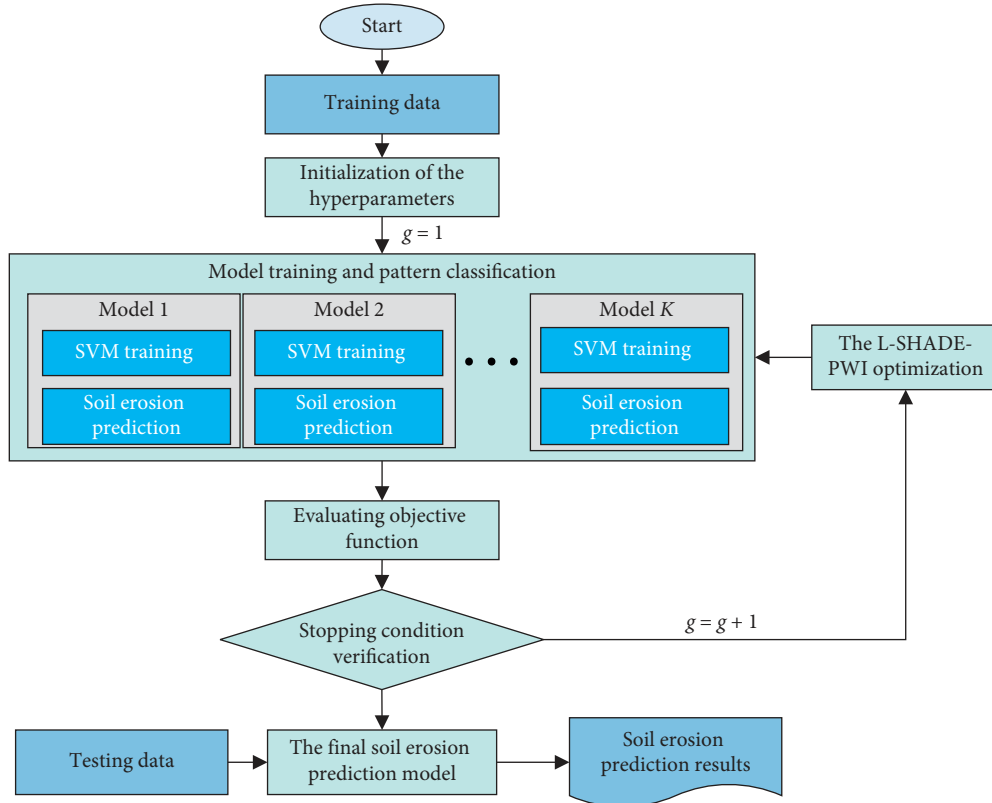


FIGURE 2: The proposed L-SHADE-PWI-SVM model for soil erosion prediction.

Accordingly, the objective function ( $F$ ) used for the L-SHADE-PWI metaheuristic-based optimization is given by

$$F = \frac{\sum_{k=1}^K (FNR_k + FPR_k)}{K}, \quad (11)$$

where  $FNR_k$  and  $FPR_k$  are the false-negative rate (FNR) and the false-positive rate (FPR) obtained from  $k^{\text{th}}$  run, respectively.

The FNR and FPR are calculated according to the following equations:

$$FNR = \frac{FN}{FN + TP}, \quad (12)$$

$$FPR = \frac{FP}{FP + TN}, \quad (13)$$

where FN, FP, TP, and TN represent the false-negative, false-positive, true-positive, and true-negative data samples, respectively.

## 4. Model Results and Discussions

**4.1. Preliminary Feature Importance Investigation.** For the purpose of constructing a robust soil prediction model, a preliminary assessment of the relevance of input features (EI30, slope, OC, pH, bulk density, soil porosity, soil texture (silt, clay, and sand fractions), and soil cover rate) should be carried out. This study relies on the well-known ReliefF method [107] to

investigate the importance of each aforementioned soil erosion influencing factor. This step of the study aims at identifying irrelevant influencing factors. Using the ReliefF method, a weighting value is computed for each input factor. A large weight indicates a strong relevance between an influencing factor and the soil erosion status (either erosion or nonerosion). The preliminary assessment of the relevancy of input features is reported in Figure 3. In this figure, the ReliefF weight values of all input variables are shown. It is observed that the input factor  $X_1$  (EI30) receives the highest weight value, followed by the input factor  $X_{10}$  (soil cover rate),  $X_3$  (OC topsoil), and  $X_9$  (sand fraction). The factors  $X_7$  (silt fraction) and  $X_8$  (clay fraction) have low weight values. However, these weights are still significantly larger than zero. Therefore, the L-SHADE-PWI metaheuristic-optimized SVM model employs all of the ten factors.

**4.2. Model Performance Evaluation Metrics.** To quantify and compare performances of soil erosion prediction models, the classification accuracy rate (CAR), which is the percentage of correctly classified cases, is often employed:

$$CAR = \frac{N_c}{N_a} \times 100\%, \quad (14)$$

where  $N_c$  and  $N_a$  denote the numbers of correctly classified samples and the total number of data samples, respectively.

Besides CAR, the true-positive rate (TPR) (the percentage of positive instances correctly classified), false-

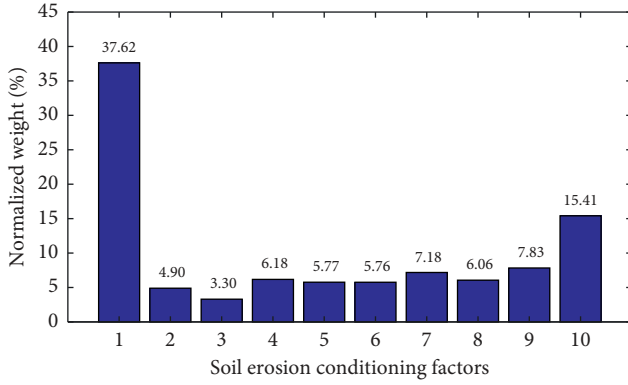


FIGURE 3: ReliefF-based inspection of variable importance.

positive rate (FPR) (the percentage of negative instances misclassified), false-negative rate (FNR) (the percentage of positive instances misclassified), and true-negative rate (TNR) (the percentage of negative instances correctly classified) are also employed [56, 108]. The FPR and FNR indices have been defined in the previous section of the article; the TPR and TNR indices are given by

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (15)$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (16)$$

where TP, TN, FP, and FN represent the true-positive, true-negative, false-positive, and false-negative data samples, respectively.

In addition, based on the results of the TP, TN, FP, and FN, indices of precision, recall, negative predictive value (NPV), and  $F1$  score are also useful for expressing the model predictive capability [109]:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (17)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (18)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}, \quad (19)$$

$$F1 \text{ - score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \quad (20)$$

**4.3. Experimental Results and Comparison.** To evaluate the model performance, the original dataset has been randomly divided into two mutually exclusive sets. Since the size of the collected dataset is moderate, the training/testing data ratio is selected to be 9/1 [110]. This means that 90% of the original dataset is used for model construction; the rest of the dataset is reserved for the model testing phase. The testing set is employed as novel data instance to verify the generalization of the constructed L-SHADE-PWI-SVM

model used for rainfall-induced soil erosion. Illustrations of training and testing datasets are provided in Table 2.

In addition, since the model construction phase of the SVM-based model requires the setting of the penalty coefficient and kernel function parameters, the L-SHADE-PWI has been used to automatically fine-tune these hyper-parameters. The L-SHADE-PWI-based optimization process is illustrated in Figure 4. It is noted that the metaheuristic population consists of 20 members and the maximum number of generations is 100. The outcomes of the optimization process are as follows:

- (i) The penalty coefficient  $C = 958.84$
- (ii) The kernel function parameter  $\sigma = 27.22$

To demonstrate the capability of the proposed L-SHADE-PWI-SVM, the backpropagation artificial neural network (BPANN) [106, 111] and radial basis function artificial neural network (RBFANN) models are employed as benchmark methods. The BPANN and RBFANN models have been constructed in MATLAB environment with the help of built-in functions provided in the Statistics and Machine Learning Toolbox [112].

It is noted that the number of neurons in the hidden layer of the BPANN model is selected to be  $(2/3)D_X + C_N$  based on the suggestion in [113], where  $D_X$  and  $C_N$  denote the numbers of input features and class outputs, respectively. Herein, the numbers of input features and class outputs are 10 and 2, respectively. Moreover, the BPANN is trained using the Levenberg–Marquardt (LM) algorithms [111, 114]. The sigmoidal activation function is used for the BPANN to train with the LM algorithm (denoted as LM-BPANN). The BPANN model is trained with the maximum number of epochs = 1000. Moreover, based on several trial-and-error experiments with the collected dataset, the appropriate RBFANN model [115] consists of 30 neurons and has been trained with a basic width of 1.2.

The L-SHADE-PWI-SVM as well as the used benchmark models is trained with a training set (90%) and a testing set (10%) randomly drawn from the original datasets. It is noted that since one-time evaluation is not sufficient to verify the model capability due to the effect of the random data sampling process, this study has repeated the random splitting of the dataset into model training and testing phases 20 times to negate such an undesired effect. Accordingly, the prediction results of the L-SHADE-PWI-SVM and benchmark models obtained from 20 runs are reported in Table 3 and Figure 5.

It can be seen that the proposed hybrid machine learning model has achieved the most desired accuracy on soil erosion detection with  $\text{CAR} = 92.292\%$  while LM-BPANN ranks second with  $\text{CAR} = 87.083\%$  and RBFANN ranks third with  $\text{CAR} = 77.917\%$ . L-SHADE-PWI-SVM is also the most reliable predictor with a smaller CAR deviation (5.1 versus 6.6 and 9.9). This is clearly demonstrated in the box plots in Figure 5.

The proposed model is also superior in other metrics such as precision, recall, NPV, and  $F1$ -score with better measurement value and less variation (smaller deviation). More specifically, L-SHADE-PWI-SVM has precision = 0.944, recall = 0.900, NPV = 0.904, and  $F1$ -score = 0.919, and LM-



TABLE 2: Illustration of the training and testing datasets.

| Dataset  | Sample | $X_1$   | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |     |
|----------|--------|---------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|-----|
| Training | 1      | 2044.01 | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 42.09    | 1            |     |
|          | 2      | 89.52   | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 6.37     | 1            |     |
|          | 3      | 550.22  | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 26.25    | 1            |     |
|          | 4      | 973.58  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 10.45    | 1            |     |
|          | 5      | 192.8   | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 15.06    | 1            |     |
|          | ...    | ...     | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...      | ...          | ... |
|          | 218    | 136.44  | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 59.12    | -1           |     |
|          | 219    | 6.82    | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 88.05    | -1           |     |
|          | 220    | 0.01    | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 1.76     | -1           |     |
|          | 221    | 0.01    | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 60.71    | -1           |     |
| 222      | 6.42   | 29.8    | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 69.61 | -1       |              |     |
| Testing  | 1      | 75.09   | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 2.63     | 1            |     |
|          | 2      | 3008.93 | 27.9  | 2.34  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 61.5     | 1            |     |
|          | 3      | 973.58  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 30.48    | 1            |     |
|          | 4      | 862.96  | 28.47 | 2.35  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 46.59    | 1            |     |
|          | 5      | 388.7   | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 86.43    | 1            |     |
|          | ...    | ...     | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...   | ...      | ...          | ... |
|          | 20     | 0.01    | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 80.55    | -1           |     |
|          | 21     | 131.01  | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 58.84    | -1           |     |
|          | 22     | 230.39  | 26.5  | 2.3   | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 81.91    | -1           |     |
|          | 23     | 15.87   | 28.37 | 2.07  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 65.08    | -1           |     |
| 24       | 682.46 | 26.33   | 2.38  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 61.97 | -1       |              |     |

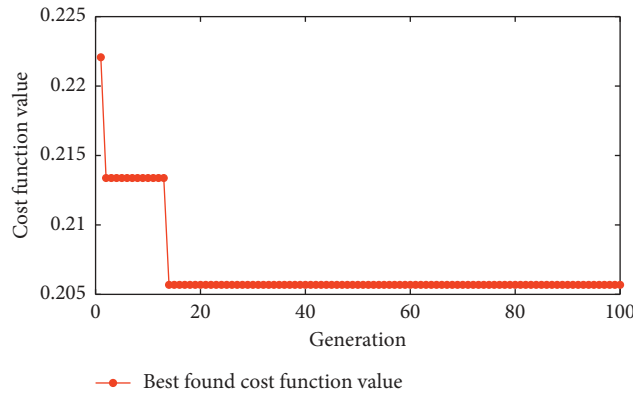


FIGURE 4: Evolutionary progress of L-SHADE-PWI.

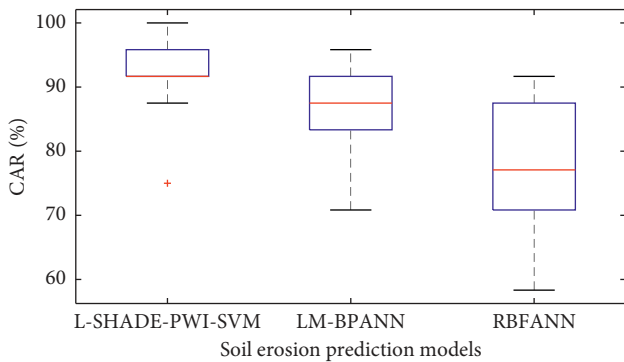


FIGURE 5: Box plots of the model performances.

BPANN is the second best prediction approach with precision = 0.901, recall = 0.854, NPV = 0.868, and F1-score = 0.869, followed by RBFANN with precision = 0.804, recall = 0.758, NPV = 0.769, and F1-score = 0.776.

Moreover, the receiver operating characteristic curve (ROC) and the area under the curve (AUC) [116] are also employed for assessing the prediction model performance [56, 117, 118]. It is because the AUC can express the overall predictive accuracy of the employed classifiers used for soil erosion prediction. This study also computes the AUC to indicate the generalization capability of the proposed L-SHADE-PWI-SVM as well as the benchmark models. The AUC results are reported in Table 4. As can be seen from the experimental results, the AUC value in the testing phase of the proposed L-SHADE-PWI-SVM (0.908) is higher than those of the LM-BPANN (0.898) and RBFANN (0.787). In addition, the ROCs of the proposed model are illustrated in Figure 6.

4.4. Discussion. To further confirm the superiority of the proposed hybrid L-SHADE-PWI-SVM model, the Wilcoxon signed-rank test [119] with the significant level

TABLE 3: Result comparison.

| Phase    | Indices   | L-SHADE-PWI-SVM |       | LM-BPANN |       | RBFANN |       |
|----------|-----------|-----------------|-------|----------|-------|--------|-------|
|          |           | Mean            | Std.  | Mean     | Std.  | Mean   | Std.  |
| Training | CAR       | 92.052          | 0.738 | 92.759   | 2.431 | 77.712 | 1.855 |
|          | TP        | 93.650          | 3.150 | 95.150   | 5.594 | 83.550 | 3.069 |
|          | TN        | 101.500         | 2.351 | 101.500  | 3.547 | 81.200 | 3.139 |
|          | FP        | 4.250           | 1.209 | 4.500    | 3.547 | 24.800 | 3.139 |
|          | FN        | 12.600          | 1.847 | 10.850   | 5.594 | 22.450 | 3.069 |
|          | Precision | 0.957           | 0.011 | 0.957    | 0.031 | 0.772  | 0.022 |
|          | Recall    | 0.881           | 0.018 | 0.898    | 0.053 | 0.788  | 0.029 |
|          | NPV       | 0.890           | 0.013 | 0.906    | 0.044 | 0.784  | 0.022 |
|          | F1-score  | 0.917           | 0.009 | 0.925    | 0.027 | 0.780  | 0.019 |
| Testing  | CAR       | 92.292          | 5.108 | 87.083   | 6.609 | 77.917 | 9.943 |
|          | TP        | 10.550          | 1.877 | 10.250   | 1.164 | 9.100  | 1.252 |
|          | TN        | 11.600          | 2.010 | 10.650   | 1.631 | 9.600  | 1.759 |
|          | FP        | 0.650           | 0.813 | 1.350    | 1.631 | 2.400  | 1.759 |
|          | FN        | 1.200           | 0.951 | 1.750    | 1.164 | 2.900  | 1.252 |
|          | Precision | 0.944           | 0.069 | 0.901    | 0.106 | 0.804  | 0.127 |
|          | Recall    | 0.900           | 0.081 | 0.854    | 0.097 | 0.758  | 0.104 |
|          | NPV       | 0.904           | 0.078 | 0.868    | 0.083 | 0.769  | 0.094 |
|          | F1-score  | 0.919           | 0.053 | 0.869    | 0.060 | 0.776  | 0.096 |

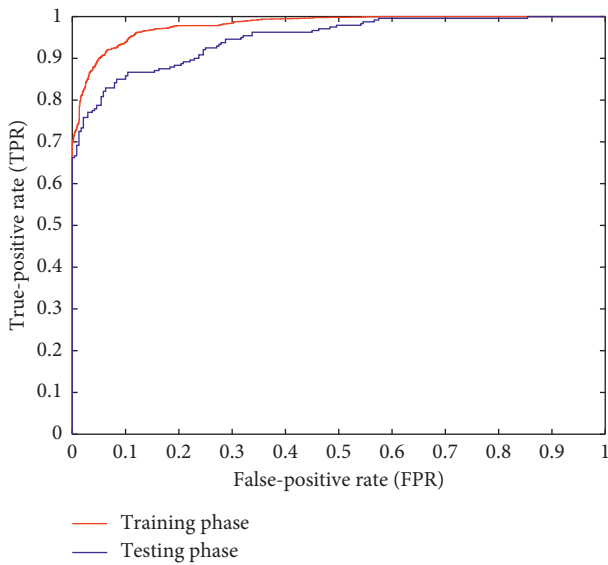


FIGURE 6: ROCs of the proposed L-SHADE-PWI-SVM.

( $p$  value) = 0.05 is also employed in this study to demonstrate the statistical significance of the difference in model results. This nonparametric hypothesis testing method is widely employed for comparing classification models [120]. The test outcomes of pairwise model comparison are reported in Table 5. Observably, with  $p$  values < 0.05, the null hypothesis of equal means is rejected.

Moreover, to assess the reliability of the L-SHADE-PWI-SVM-based soil erosion prediction model, the coefficient of variation (COV) [121, 122] is employed in this section of the study. The COV (%) is computed as follows [121, 122]:

$$\text{COV} = 100 \times \frac{\sigma}{\mu} \quad (21)$$

TABLE 4: AUC result comparison.

| Phase    | L-SHADE-PWI-SVM |       | LM-BPANN |       | RBFANN |       |
|----------|-----------------|-------|----------|-------|--------|-------|
|          | Mean            | Std.  | Mean     | Std.  | Mean   | Std.  |
| Training | 0.927           | 0.011 | 0.942    | 0.025 | 0.795  | 0.020 |
| Testing  | 0.908           | 0.065 | 0.898    | 0.044 | 0.787  | 0.094 |

The COV calculation results of the proposed L-SHADE-PWI-SVM and the benchmark approaches used for rainfall-induced erosion susceptibility prediction are reported in Table 6. It is noted that the COV is the ratio of the standard deviation to the mean and is used to quantify the dispersion of a model prediction outcomes obtained from a repeated data sampling process [123]. In the particular case of rainfall-induced soil erosion susceptibility evaluation, a small value of COV is desirable since it indicates a stable data-driven model used for predicting such phenomenon. As shown in Table 6, the proposed L-SHADE-PWI-SVM has achieved the lowest COV in all of the performance measurement metrics (5.535% for CAR, 7.309% for precision, 9.000% for recall, 8.628% for NPV, and 5.767% for F1-score). These outcomes point out that the newly developed model is the most reliable tool for rainfall-induced soil erosion susceptibility assessment.

In conclusion, L-SHADE-PWI-SVM is more accurate and more reliable model compared to LM-BPANN ( $p$  values = 0.009816) and RBFANN ( $p$  values = 0.000631). Statistical tests also confirm that the superior of L-SHADE-PWI-SVM in terms of accuracy is also significant with  $p$  values < 0.05. This means that the newly constructed L-SHADE-PWI-SVM is highly appropriate for soil erosion prediction in the study area.

Compared to the conventional approaches for soil erosion prediction such as the Revised Universal Soil Loss Equation (RUSLE) [124, 125] which requires significant

TABLE 5: Results of the Wilcoxon signed-rank test.

| Model comparison             | Test outcome | <i>p</i> value |
|------------------------------|--------------|----------------|
| L-SHADE-PWI-SVM vs. LM-BPANN | Significant  | 0.009816       |
| L-SHADE-PWI-SVM vs. RBFANN   | Significant  | 0.000631       |

TABLE 6: Calculations of the coefficient of variation (COV).

| Indices   | L-SHADE-PWI-SVM | LM-BPANN | RBFANN |
|-----------|-----------------|----------|--------|
| CAR (%)   | 5.535           | 7.589    | 12.761 |
| Precision | 7.309           | 11.765   | 15.796 |
| Recall    | 9.000           | 11.358   | 13.720 |
| NPV       | 8.628           | 9.562    | 12.224 |
| F1-score  | 5.767           | 6.904    | 12.371 |

TABLE 7: The Collected Dataset.

| Data sample | $X_1$   | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |
|-------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|
| 1           | 2044.01 | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 42.09    | 1            |
| 2           | 975.56  | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 25.91    | 1            |
| 3           | 2044.01 | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 60.19    | 1            |
| 4           | 975.56  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 8.98     | 1            |
| 5           | 2044.01 | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 54.96    | 1            |
| 6           | 1138.37 | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 29.74    | 1            |
| 7           | 975.56  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 18.01    | 1            |
| 8           | 2044.01 | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 51.12    | 1            |
| 9           | 2044.01 | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 44.35    | 1            |
| 10          | 1138.37 | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 23.59    | 1            |
| 11          | 975.56  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 9.65     | 1            |
| 12          | 976.92  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 5.99     | 1            |
| 13          | 975.56  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 29.9     | 1            |
| 14          | 975.56  | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 30.56    | 1            |
| 15          | 3008.93 | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 53.9     | 1            |
| 16          | 1270.46 | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 10.54    | 1            |
| 17          | 1472.5  | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 4.85     | 1            |
| 18          | 2044.01 | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 47.03    | 1            |
| 19          | 1138.37 | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 31.98    | 1            |
| 20          | 1270.46 | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 12.33    | 1            |
| 21          | 2044.01 | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 59.98    | 1            |
| 22          | 975.56  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 26.91    | 1            |
| 23          | 975.56  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 7.65     | 1            |
| 24          | 3008.93 | 26.27 | 2.18  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 60.88    | 1            |
| 25          | 1138.37 | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 27.69    | 1            |
| 26          | 2044.01 | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 65.48    | 1            |
| 27          | 3008.93 | 28.03 | 2.27  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 44.96    | 1            |
| 28          | 1270.46 | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 13.13    | 1            |
| 29          | 975.56  | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 22.07    | 1            |
| 30          | 1138.37 | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 44.29    | 1            |
| 31          | 685.09  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 14.97    | 1            |
| 32          | 1138.37 | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 21.1     | 1            |
| 33          | 3008.93 | 28.47 | 2.35  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 58.07    | 1            |
| 34          | 685.09  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 12.75    | 1            |
| 35          | 263.04  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 6.74     | 1            |
| 36          | 550.22  | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 24.66    | 1            |
| 37          | 2044.01 | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 65.79    | 1            |
| 38          | 1472.5  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 3.69     | 1            |
| 39          | 973.58  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 9.73     | 1            |
| 40          | 192.8   | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 18.75    | 1            |
| 41          | 2503.7  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 86.7     | 1            |
| 42          | 685.09  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 19       | 1            |

TABLE 7: Continued.

| Data sample | $X_1$   | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |
|-------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|
| 43          | 1138.37 | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 35.06    | 1            |
| 44          | 88      | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 12.06    | 1            |
| 45          | 550.22  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 21.08    | 1            |
| 46          | 2044.01 | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 61.2     | 1            |
| 47          | 976.92  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 6.43     | 1            |
| 48          | 1138.37 | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 26.56    | 1            |
| 49          | 1138.37 | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 27.73    | 1            |
| 50          | 2503.7  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 82.31    | 1            |
| 51          | 973.58  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 18.13    | 1            |
| 52          | 973.58  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 8.29     | 1            |
| 53          | 976.92  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 17.52    | 1            |
| 54          | 2503.7  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 86.5     | 1            |
| 55          | 3008.93 | 28.37 | 1.95  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 66.18    | 1            |
| 56          | 550.22  | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 26.25    | 1            |
| 57          | 976.92  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 5.1      | 1            |
| 58          | 2503.7  | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 82.9     | 1            |
| 59          | 263.04  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 5.74     | 1            |
| 60          | 306.15  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 11.06    | 1            |
| 61          | 180.82  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 16.56    | 1            |
| 62          | 1969.39 | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 87.25    | 1            |
| 63          | 685.09  | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 29.65    | 1            |
| 64          | 3008.93 | 24.83 | 2.2   | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 59.03    | 1            |
| 65          | 976.92  | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 24.94    | 1            |
| 66          | 263.04  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 7.23     | 1            |
| 67          | 973.58  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 10.45    | 1            |
| 68          | 152.99  | 27.57 | 2.07  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 70.78    | 1            |
| 69          | 2503.7  | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 89.3     | 1            |
| 70          | 89.52   | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 6.37     | 1            |
| 71          | 3008.93 | 28    | 1.95  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 64.44    | 1            |
| 72          | 192.8   | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 15.06    | 1            |
| 73          | 973.58  | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 26.16    | 1            |
| 74          | 3008.93 | 27.9  | 2.34  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 61.5     | 1            |
| 75          | 685.09  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 16.08    | 1            |
| 76          | 184.93  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 18.07    | 1            |
| 77          | 89.52   | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 8.04     | 1            |
| 78          | 973.58  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 27.24    | 1            |
| 79          | 263.04  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 17.64    | 1            |
| 80          | 2503.7  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 87       | 1            |
| 81          | 685.09  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 29.52    | 1            |
| 82          | 976.92  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 25.61    | 1            |
| 83          | 973.58  | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 30.44    | 1            |
| 84          | 3008.93 | 27.57 | 2.07  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 58.33    | 1            |
| 85          | 550.22  | 24.83 | 2.66  | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 38.42    | 1            |
| 86          | 973.58  | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 22.25    | 1            |
| 87          | 1270.46 | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 30.71    | 1            |
| 88          | 682.46  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 52.7     | 1            |
| 89          | 685.09  | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 27.86    | 1            |
| 90          | 85.74   | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 13.55    | 1            |
| 91          | 1472.5  | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 8.73     | 1            |
| 92          | 685.09  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 34.51    | 1            |
| 93          | 82.99   | 27.57 | 2.08  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 41.51    | 1            |
| 94          | 75.09   | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 2.63     | 1            |
| 95          | 976.92  | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 21.36    | 1            |
| 96          | 88      | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 9.56     | 1            |
| 97          | 550.22  | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 38.42    | 1            |
| 98          | 3008.93 | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 66.21    | 1            |
| 99          | 184.93  | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 22.5     | 1            |
| 100         | 89.52   | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 7.48     | 1            |
| 101         | 192.8   | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 34.01    | 1            |
| 102         | 550.22  | 28    | 2.14  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 29.34    | 1            |

TABLE 7: Continued.

| Data sample | $X_1$  | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |
|-------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|
| 103         | 82.99  | 28.47 | 2.15  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 30.78    | 1            |
| 104         | 973.58 | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 30.48    | 1            |
| 105         | 88     | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 27.89    | 1            |
| 106         | 88     | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 11.23    | 1            |
| 107         | 270.98 | 28.47 | 2.35  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 54.14    | 1            |
| 108         | 88     | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 26.64    | 1            |
| 109         | 862.96 | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 41.05    | 1            |
| 110         | 388.7  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 86.91    | 1            |
| 111         | 263.04 | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 25.93    | 1            |
| 112         | 89.52  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 28.75    | 1            |
| 113         | 88     | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 31.63    | 1            |
| 114         | 685.09 | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 23.51    | 1            |
| 115         | 263.04 | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 25.19    | 1            |
| 116         | 862.96 | 28.47 | 2.35  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 46.59    | 1            |
| 117         | 85.74  | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 16.88    | 1            |
| 118         | 388.7  | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 86.43    | 1            |
| 119         | 136.44 | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 59.12    | -1           |
| 120         | 0.13   | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 66.79    | -1           |
| 121         | 0.07   | 26.27 | 2.39  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 92.23    | -1           |
| 122         | 51.22  | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 32.17    | -1           |
| 123         | 152.99 | 32.33 | 2.64  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 78.9     | -1           |
| 124         | 0.01   | 26.5  | 2.3   | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 76.61    | -1           |
| 125         | 248.37 | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 67.52    | -1           |
| 126         | 0.07   | 26.5  | 2.3   | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 89.2     | -1           |
| 127         | 0.01   | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 78.39    | -1           |
| 128         | 0.06   | 27.9  | 2.34  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 78.29    | -1           |
| 129         | 1.59   | 28.03 | 2.25  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 42.49    | -1           |
| 130         | 0.02   | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 12.47    | -1           |
| 131         | 2.09   | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 78.27    | -1           |
| 132         | 6.82   | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 88.05    | -1           |
| 133         | 0.01   | 27.57 | 2.08  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 54.61    | -1           |
| 134         | 4.61   | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 74.21    | -1           |
| 135         | 41.92  | 27.9  | 2.34  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 94.48    | -1           |
| 136         | 0.13   | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 68.43    | -1           |
| 137         | 0.53   | 27.9  | 2.36  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 22.47    | -1           |
| 138         | 0.13   | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 13.56    | -1           |
| 139         | 0.01   | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 1.76     | -1           |
| 140         | 299.78 | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 87.63    | -1           |
| 141         | 0.01   | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 60.71    | -1           |
| 142         | 6.79   | 26.33 | 2.36  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 68.94    | -1           |
| 143         | 46.34  | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 29.89    | -1           |
| 144         | 0.01   | 26.5  | 2.33  | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 17.34    | -1           |
| 145         | 0.27   | 28.47 | 2.15  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 50.56    | -1           |
| 146         | 10.42  | 32.33 | 2.29  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 12.61    | -1           |
| 147         | 0.27   | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 79.4     | -1           |
| 148         | 0.36   | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 31.35    | -1           |
| 149         | 6.67   | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 45.21    | -1           |
| 150         | 299.78 | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 85.91    | -1           |
| 151         | 9.59   | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 33.71    | -1           |
| 152         | 2.59   | 24.83 | 2.2   | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 90.72    | -1           |
| 153         | 6.42   | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 69.61    | -1           |
| 154         | 25.4   | 30.27 | 1.24  | 5.72  | 1.45  | 54.87 | 33.43 | 27.21 | 39.36 | 68.61    | -1           |
| 155         | 3.38   | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 49.22    | -1           |
| 156         | 12.41  | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 28.42    | -1           |
| 157         | 89.52  | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 26.26    | -1           |
| 158         | 152.99 | 27.9  | 2.34  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 73.49    | -1           |
| 159         | 15.22  | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 94.25    | -1           |
| 160         | 202.32 | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 25.35    | -1           |
| 161         | 115.94 | 30.27 | 1.24  | 5.72  | 1.45  | 54.87 | 33.43 | 27.21 | 39.36 | 64.78    | -1           |
| 162         | 3.91   | 28.03 | 2.25  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 66.79    | -1           |

TABLE 7: Continued.

| Data sample | $X_1$  | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |
|-------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|
| 163         | 61.6   | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 40.05    | -1           |
| 164         | 0.24   | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 28.69    | -1           |
| 165         | 0      | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 24.7     | -1           |
| 166         | 0.07   | 28.63 | 2.36  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 80.3     | -1           |
| 167         | 0.03   | 24.83 | 2.2   | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 95.27    | -1           |
| 168         | 21.94  | 28.47 | 2.08  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 63.41    | -1           |
| 169         | 51.22  | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 28.54    | -1           |
| 170         | 550.22 | 28.37 | 2.05  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 40.79    | -1           |
| 171         | 4.21   | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 19.05    | -1           |
| 172         | 64.15  | 30.27 | 1.24  | 5.72  | 1.45  | 54.87 | 33.43 | 27.21 | 39.36 | 70.53    | -1           |
| 173         | 54.17  | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 59.86    | -1           |
| 174         | 385.11 | 27.57 | 2.11  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 71.11    | -1           |
| 175         | 0.01   | 28    | 1.95  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 48.2     | -1           |
| 176         | 0.03   | 27.57 | 2.11  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 28.77    | -1           |
| 177         | 0.01   | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 80.55    | -1           |
| 178         | 6.42   | 29.57 | 1     | 5.8   | 1.46  | 55.26 | 31.35 | 30.43 | 38.21 | 47.99    | -1           |
| 179         | 0.01   | 28.47 | 2.35  | 5.36  | 1.25  | 47.09 | 31.47 | 34.59 | 33.93 | 86.16    | -1           |
| 180         | 0.02   | 27.93 | 0.89  | 5.83  | 1.47  | 55.47 | 33.47 | 30.69 | 35.84 | 22.84    | -1           |
| 181         | 248.37 | 24.83 | 2.66  | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 68.24    | -1           |
| 182         | 0.13   | 33.73 | 1.06  | 6.95  | 1.57  | 59.38 | 37.71 | 19.51 | 42.77 | 47.47    | -1           |
| 183         | 1.96   | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 10.46    | -1           |
| 184         | 295.82 | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 89.38    | -1           |
| 185         | 22.45  | 28.03 | 2.27  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 57.02    | -1           |
| 186         | 0.01   | 26.5  | 2.33  | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 45.64    | -1           |
| 187         | 131.01 | 32.33 | 2.64  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 62.69    | -1           |
| 188         | 25.82  | 26.33 | 2.38  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 43.64    | -1           |
| 189         | 0.13   | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 48.41    | -1           |
| 190         | 0.01   | 32.33 | 2.3   | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 57.92    | -1           |
| 191         | 0.03   | 27.57 | 2.07  | 5.13  | 1.48  | 55.87 | 33.29 | 32.17 | 34.54 | 81.94    | -1           |
| 192         | 0.14   | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 47.9     | -1           |
| 193         | 0.72   | 26.27 | 2.41  | 5.23  | 1.32  | 49.92 | 32.41 | 31.09 | 36.49 | 24.38    | -1           |
| 194         | 0.01   | 28.03 | 2.27  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 70.92    | -1           |
| 195         | 0.01   | 32.33 | 2.29  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 80.56    | -1           |
| 196         | 0.38   | 27.9  | 2.18  | 5.5   | 1.34  | 50.69 | 35.81 | 29.23 | 34.95 | 66.79    | -1           |
| 197         | 0.01   | 26.33 | 2.36  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 63.38    | -1           |
| 198         | 57.23  | 29.8  | 1.13  | 5.7   | 1.34  | 50.58 | 31.99 | 38.35 | 29.66 | 82.15    | -1           |
| 199         | 0.53   | 26.5  | 2.33  | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 25.72    | -1           |
| 200         | 6.74   | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 17.21    | -1           |
| 201         | 1.36   | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 41.61    | -1           |
| 202         | 0.01   | 28.37 | 2.05  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 83.99    | -1           |
| 203         | 0.03   | 30.27 | 1.24  | 5.72  | 1.45  | 54.87 | 33.43 | 27.21 | 39.36 | 81.04    | -1           |
| 204         | 0.42   | 32.33 | 2.64  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 52.59    | -1           |
| 205         | 6.82   | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 95.3     | -1           |
| 206         | 180.82 | 28.03 | 2.38  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 35.11    | -1           |
| 207         | 131.01 | 26.33 | 2.79  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 58.84    | -1           |
| 208         | 0.01   | 28.37 | 2.05  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 17.44    | -1           |
| 209         | 0.79   | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 19.32    | -1           |
| 210         | 766.63 | 26.5  | 2.3   | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 74.62    | -1           |
| 211         | 34.74  | 24.83 | 2.66  | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 31.81    | -1           |
| 212         | 17.58  | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 11.52    | -1           |
| 213         | 3.32   | 28.37 | 1.95  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 67.05    | -1           |
| 214         | 14.81  | 28    | 2.14  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 33.43    | -1           |
| 215         | 152.99 | 28.63 | 2.39  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 65.34    | -1           |
| 216         | 0.03   | 28.17 | 1.33  | 6.38  | 1.48  | 55.8  | 34.99 | 24.95 | 40.05 | 63.29    | -1           |
| 217         | 0.1    | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 94.6     | -1           |
| 218         | 6.16   | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 21.81    | -1           |
| 219         | 4.35   | 29.6  | 1.46  | 6.68  | 1.41  | 53.12 | 33.19 | 20.29 | 46.51 | 76.4     | -1           |
| 220         | 0.01   | 28.17 | 1.2   | 6.59  | 1.37  | 51.75 | 34.93 | 27.57 | 37.5  | 14.91    | -1           |
| 221         | 230.39 | 26.5  | 2.3   | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 81.91    | -1           |
| 222         | 0.01   | 24.83 | 2.66  | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 59.49    | -1           |

TABLE 7: Continued.

| Data sample | $X_1$  | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | $X_{10}$ | Class output |
|-------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|----------|--------------|
| 223         | 0.03   | 24.83 | 2.2   | 5.97  | 1.37  | 51.79 | 34.35 | 34.25 | 31.4  | 84.78    | -1           |
| 224         | 0.03   | 26.5  | 2.07  | 5.39  | 1.38  | 52.05 | 32.65 | 29.37 | 37.98 | 96.62    | -1           |
| 225         | 6.82   | 32.33 | 2.64  | 5.6   | 1.41  | 53.11 | 33.21 | 35.47 | 31.32 | 97.64    | -1           |
| 226         | 61.6   | 34.77 | 1.42  | 6.72  | 1.37  | 51.87 | 32.73 | 24.33 | 42.93 | 36.87    | -1           |
| 227         | 6.16   | 34.17 | 0.99  | 5.89  | 1.34  | 50.62 | 34.31 | 35.33 | 30.35 | 68.71    | -1           |
| 228         | 9.8    | 30.2  | 1.28  | 5.85  | 1.4   | 52.69 | 36.19 | 27.95 | 35.86 | 39.13    | -1           |
| 229         | 1.79   | 28.03 | 2.25  | 5.54  | 1.3   | 49.07 | 33.93 | 30.15 | 35.92 | 44.54    | -1           |
| 230         | 10.42  | 28.63 | 2.32  | 5.2   | 1.44  | 54.49 | 34.73 | 28.89 | 36.37 | 1.05     | -1           |
| 231         | 15.87  | 28.37 | 2.07  | 5.15  | 1.23  | 46.34 | 33.07 | 32.09 | 34.84 | 65.08    | -1           |
| 232         | 0.42   | 28    | 1.95  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 48.18    | -1           |
| 233         | 25.4   | 29.8  | 1.53  | 7.06  | 1.58  | 59.48 | 36.31 | 18.61 | 45.08 | 66.82    | -1           |
| 234         | 21.94  | 28    | 2.14  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 64.13    | -1           |
| 235         | 682.46 | 26.33 | 2.38  | 5.91  | 1.37  | 51.75 | 34.69 | 34.29 | 31.01 | 61.97    | -1           |
| 236         | 0.38   | 28    | 2.17  | 5.13  | 1.37  | 51.53 | 33.07 | 31.51 | 35.42 | 39.53    | -1           |

efforts on parameter calibration, the newly proposed method is entirely data-driven in which all of the model parameters are determined via the model training process. In addition, recently proposed machine learning methods for soil erosion susceptibility prediction have dominantly relied on individual or ensemble of models [126–128]. Therefore, trial-and-error processes and modeling experience are required for constructing such machine learning models. An integration of machine learning and metaheuristic used for soil erosion prediction is rarely investigated. Thus, the current study is an attempt to fill this gap in the literature and show the great potentiality of this hybrid framework for tackling the problem at hand.

Nevertheless, one disadvantage of the proposed approach is that the optimization process to determine an optimal set of parameters of the SVM model might be costly, especially when the size of the collected dataset is large. It is because the model training and prediction processes of the SVM are embedded into the cost function calculation phase of the employed metaheuristic. Another limitation of the current approach is that automatic feature selection has not been integrated into the hybridization of the L-SHADE-PWI and SVM. Such drawbacks should be resolved with in future extensions of the study.

## 5. Conclusions

In tropical regions, soil erosion is a natural hazard that causes various harmful effects on the land including loss of soil, soil structure breakdown, and the decline of organic matter as well as nutrients within the soil. These ultimately lead to a critical economic loss for landowners. This study has developed a hybrid intelligent method, named as L-SHADE-PWI-SVM, for predicting the status of soil erosion based on an integration of machine learning and metaheuristic. The SVM pattern recognition method is employed to generalize a decision boundary that separates input data into two categories of erosion and nonerosion.

The ten variables of EI30, slope, OC topsoil, pH topsoil, bulk density, soil porosity, soil texture (silt, clay, and sand fractions), and soil cover rate are used as erosion

conditioning factors. In addition, to optimize the SVM model performance, the state-of-the-art L-SHADE-PWI metaheuristic is employed. The newly developed L-SHADE-PWI-SVM has achieved a good predictive performance (CAR = 92.292%, F1 score = 0.919, and AUC = 0.908) obtained from a repeated data sampling process. The experimental results supported by the Wilcoxon signed-rank test demonstrate that the proposed hybrid model is superior to the benchmark methods including the LM-BPANN (CAR = 87.083%, F1-score = 0.869, and AUC = 0.898) and the RBFANN (CAR = 77.917%, F1-score = 0.776, and AUC = 0.787). The proposed L-SHADE-PWI-SVM has outperformed the benchmark approaches in all of the evaluation metrics. These facts strongly confirm the efficacy of applying the proposed hybrid machine learning method for solving the problem of interest. The L-SHADE-PWI-SVM can be a very promising tool to assist landowners and managers to quickly identify the potential soil erosion areas and develop preventive measures (Table 7).

Future developments of the current study may include the following:

- (i) The applications of other advanced metaheuristics in optimizing the machine learning model used for soil erosion prediction
- (ii) The integration of state-of-the-art feature selection into the machine learning structure to further enhance the prediction accuracy
- (iii) The investigation of advanced kernel functions for better dealing with nonlinearity in soil erosion data classification

## Data Availability

The data used to support the findings of this study can be found in Table 7.

## Conflicts of Interest

The authors confirm that there are no conflicts of interest regarding the publication.

## Acknowledgments

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant no. 105.08-2017.302.

## References

- [1] J. Jiao, H. Zou, Y. Jia, and N. Wang, "Research progress on the effects of soil erosion on vegetation," *Acta Ecologica Sinica*, vol. 29, no. 2, pp. 85–91, 2009.
- [2] S. Visser, S. Keesstra, G. Maas, and M. De Cleen, "Soil as a basis to create enabling conditions for transitions towards sustainable land management as a key to achieve the SDGs by 2030," *Sustainability*, vol. 11, no. 23, p. 6792, 2019.
- [3] S. Molenaar, G. Mol, J. De Leeuw et al., "Soil-related sustainable development goals: four concepts to make land degradation neutrality and restoration work," *Land*, vol. 7, no. 4, p. 133, 2018.
- [4] S. D. Keesstra, J. Bouma, J. Wallinga et al., "The significance of soils and soil science towards realization of the United Nations sustainable development goals," *Soil*, vol. 2, no. 2, pp. 111–128, 2016.
- [5] S. A. El-Swaify, "Factors affecting soil erosion hazards and conservation needs for tropical steeplands," *Soil Technology*, vol. 11, no. 1, pp. 3–16, 1997.
- [6] G. Clemens, S. Fiedler, N. D. Cong, N. Van Dung, U. Schuler, and K. Stahr, "Soil fertility affected by land use history, relief position, and parent material under a tropical climate in NW-Vietnam," *Catena*, vol. 81, no. 2, pp. 87–96, 2010.
- [7] P. Schmitter, H. L. Fröhlich, G. Dercon et al., "Redistribution of carbon and nitrogen through irrigation in intensively cultivated tropical mountainous watersheds," *Biogeochemistry*, vol. 109, no. 1–3, pp. 133–150, 2012.
- [8] F. Turkelboom, J. Poesen, and G. Trébuil, "The multiple land degradation effects caused by land-use intensification in tropical steeplands: a catchment study from northern Thailand," *Catena*, vol. 75, no. 1, pp. 102–116, 2008.
- [9] P. Schmitter, G. Dercon, T. Hilger et al., "Sediment induced soil spatial variation in paddy fields of Northwest Vietnam," *Geoderma*, vol. 155, no. 3–4, pp. 298–307, 2010.
- [10] R. P. C. Morgan, *Soil Erosion and Conservation*, Blackwell Science Ltd, Oxford, UK, 3rd edition, 2005.
- [11] M. López-Vicente, H. Kramer, and S. Keesstra, "Effectiveness of soil erosion barriers to reduce sediment connectivity at small basin scale in a fire-affected forest," *Journal of Environmental Management*, vol. 278, Article ID 111510, 2021.
- [12] A. Cerdà, A. Novara, P. Dlapa et al., "Rainfall and water yield in Macizo del Caroig, Eastern Iberian Peninsula. Event runoff at plot scale during a rare flash flood at the Barranco de Benacancel," *Cuadernos de Investigación Geográfica Geographical Research Letters*, 2021, In press.
- [13] A. Novara, A. Cerda, E. Barone, and L. Gristina, "Cover crop management and water conservation in vineyard and olive orchards," *Soil and Tillage Research*, vol. 208, Article ID 104896, 2021.
- [14] J. Rodrigo-Comino, E. Terol, G. Mora, A. Giménez-Morera, and A. Cerdà, "Vicia sativa roth. can reduce soil and water losses in recently planted vineyards (vitis vinifera L.)," *Earth Systems and Environment*, vol. 4, no. 4, pp. 827–842, 2020.
- [15] S. D. Keesstra, J. Rodrigo-Comino, A. Novara et al., "Straw mulch as a sustainable solution to decrease runoff and erosion in glyphosate-treated clementine plantations in Eastern Spain. An assessment using rainfall simulation experiments," *Catena*, vol. 174, pp. 95–103, 2019.
- [16] A. Cerdà, J. Rodrigo-Comino, A. Novara et al., "Long-term impact of rainfed agricultural land abandonment on soil erosion in the Western Mediterranean basin," *Progress in Physical Geography: Earth and Environment*, vol. 42, no. 2, pp. 202–219, 2018.
- [17] A. Cerdà, J. Rodrigo-Comino, A. Giménez-Morera, and S. D. Keesstra, "Hydrological and erosional impact and farmer's perception on catch crops and weeds in citrus organic farming in Canyoles river watershed, Eastern Spain," *Agriculture, Ecosystems & Environment*, vol. 258, pp. 49–58, 2018.
- [18] V. D. Tuan, T. Hilger, L. MacDonald et al., "Mitigation potential of soil conservation in maize cropping on steep slopes," *Field Crops Research*, vol. 156, pp. 91–102, 2014.
- [19] M. A. Nearing, G. Govers, and L. D. Norton, "Variability in soil erosion data from replicated plots," *Soil Science Society of America Journal*, vol. 63, no. 6, pp. 1829–1835, 1999.
- [20] K. G. Renard and J. R. Freimund, "Using monthly precipitation data to estimate the R-factor in the revised USLE," *Journal of Hydrology*, vol. 157, no. 1–4, pp. 287–306, 1994.
- [21] Z. Kliment, J. Kadlec, and J. Langhammer, "Evaluation of suspended load changes using AnnAGNPS and SWAT semi-empirical erosion models," *Catena*, vol. 73, no. 3, pp. 286–299, 2008.
- [22] J. M. Lafren, L. J. Lane, and G. R. Foster, "WEPP: a new generation of erosion prediction technology," *Journal of Soil and Water Conservation*, vol. 46, pp. 34–38, 1991.
- [23] W. S. Merritt, R. A. Letcher, and A. J. Jakeman, "A review of erosion and sediment transport models," *Environmental Modelling & Software*, vol. 18, no. 8–9, pp. 761–799, 2003.
- [24] Ł. Sadowski, M. Nikoo, M. Shariq, E. Joker, and S. Czarnecki, "The nature-inspired metaheuristic method for predicting the creep strain of green concrete containing ground granulated blast furnace slag," *Materials*, vol. 12, no. 2, p. 293, 2019.
- [25] A. Goetzke-Pala, A. Hoła, and Ł. Sadowski, "A non-destructive method of the evaluation of the moisture in saline brick walls using artificial neural networks," *Archives of Civil and Mechanical Engineering*, vol. 18, no. 4, pp. 1729–1742, 2018.
- [26] A. Ashrafian, F. Shokri, M. J. Taheri Amiri, Z. M. Yaseen, and M. Rezaie-Balf, "Compressive strength of foamed cellular lightweight concrete simulation: new development of hybrid artificial intelligence model," *Construction and Building Materials*, vol. 230, Article ID 117048, 2020.
- [27] A. Malik, A. Kumar, S. Kim et al., "Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 323–338, 2020.
- [28] K. Khosravi, P. Daggupati, M. T. Alami et al., "Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: a case study in Iraq," *Computers and Electronics in Agriculture*, vol. 167, Article ID 105041, 2019.
- [29] M. Abedini, B. Ghasemian, A. Shirzadi et al., "A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment," *Geocarto International*, vol. 34, no. 13, pp. 1427–1457, 2019.
- [30] M. A. Shahin, "State-of-the-art review of some artificial intelligence applications in pile foundations," *Geoscience Frontiers*, vol. 7, no. 1, pp. 33–44, 2016.



- [31] M. A. Shahin, "A review of artificial intelligence applications in shallow foundations," *International Journal of Geotechnical Engineering*, vol. 9, no. 1, pp. 49–60, 2015.
- [32] R. Mohanty, S. Suman, and S. K. Das, "Prediction of vertical pile capacity of driven pile in cohesionless soil using artificial intelligence techniques," *International Journal of Geotechnical Engineering*, vol. 12, no. 2, pp. 209–216, 2018.
- [33] X. Ding, M. Hasanipannah, H. Nikafshan Rad, and W. Zhou, "Predicting the blast-induced vibration velocity using a bagged support vector regression optimized with firefly algorithm," *Engineering with Computers*, 2020.
- [34] Q. Fang, B. Yazdani Bejarbaneh, M. Vatandoust, D. Jahed Armaghani, B. Ramesh Murlidhar, and E. Tonnizam Mohamad, "Strength evaluation of granite block samples with different predictive models," *Engineering with Computers*, 2019.
- [35] H. Luo and S. G. Paal, "A locally weighted machine learning model for generalized prediction of drift capacity in seismic vulnerability assessments," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 11, pp. 935–950, 2019.
- [36] S. I. Abba, S. J. Hadi, S. S. Sammen et al., "Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination," *Journal of Hydrology*, vol. 587, Article ID 124974, 2020.
- [37] O. Kisi, M. Alizamir, S. Trajkovic, J. Shiri, and S. Kim, "Solar radiation estimation in mediterranean climate by weather variables using a novel bayesian model averaging and machine learning methods," *Neural Processing Letters*, vol. 52, no. 3, pp. 2297–2318, 2020.
- [38] S. Q. Salih, A. Sharafati, K. Khosravi et al., "River suspended sediment load prediction based on river discharge information: application of newly developed data mining models," *Hydrological Sciences Journal*, vol. 65, no. 4, pp. 624–637, 2020.
- [39] X. Chen and W. Chen, "GIS-based landslide susceptibility assessment using optimized hybrid machine learning methods," *CATENA*, vol. 196, Article ID 104833, 2021.
- [40] X. Zhao and W. Chen, "Optimization of computational intelligence models for landslide susceptibility evaluation," *Remote Sensing*, vol. 12, no. 14, p. 2180, 2020.
- [41] W. Chen and Y. Li, "GIS-based evaluation of landslide susceptibility using hybrid computational intelligence models," *CATENA*, vol. 195, Article ID 104777, 2020.
- [42] K. Pham, D. Kim, S. Park, and H. Choi, "Ensemble learning-based classification models for slope stability analysis," *CATENA*, vol. 196, Article ID 104886, 2021.
- [43] P.-T. T. Ngo, T. D. Pham, V.-H. Nhu et al., "A novel hybrid quantum-PSO and credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data," *Journal of Hydrology*, Article ID 125682, 2020.
- [44] L. Bragagnolo, R. V. D. Silva, and J. M. V. Grzybowski, "Artificial neural network ensembles applied to the mapping of landslide susceptibility," *CATENA*, vol. 184, Article ID 104240, 2020.
- [45] M. F. Yusof, H. M. Azamathulla, and R. Abdullah, "Prediction of soil erodibility factor for Peninsular Malaysia soil series using ANN," *Neural Computing and Applications*, vol. 24, no. 2, pp. 383–389, 2014.
- [46] M. Kim and J. E. Gilley, "artificial neural network estimation of soil erosion and nutrient concentrations in runoff from land application areas," *Computers and Electronics in Agriculture*, vol. 64, no. 2, pp. 268–275, 2008.
- [47] C. A. S. de Farias and C. A. G. Santos, "The use of Kohonen neural networks for runoff-erosion modeling," *Journal of Soils and Sediments*, vol. 14, no. 7, pp. 1242–1250, 2014.
- [48] P. Licznar and M. A. Nearing, "Artificial neural networks of soil erosion and runoff prediction at the plot scale," *CATENA*, vol. 51, no. 2, pp. 89–114, 2003.
- [49] I. Albaradeya, A. Hani, and I. Shahrour, "WEPP and ANN models for simulating soil loss and runoff in a semi-arid Mediterranean region," *Environmental Monitoring and Assessment*, vol. 180, no. 1–4, pp. 537–556, 2011.
- [50] D. T. Vu, X.-L. Tran, M.-T. Cao, T. C. Tran, and N.-D. Hoang, "Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline," *Measurement*, vol. 164, Article ID 108066, 2020.
- [51] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, Hoboken, NJ, USA, 1998.
- [52] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Denver, CO, USA, December 1996.
- [53] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [54] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, PA, USA, July 1992.
- [55] D. Tien Bui, N.-D. Hoang, H. Nguyen, and X.-L. Tran, "Spatial prediction of shallow landslide using Bat algorithm optimized machine learning approach: a case study in Lang Son Province, Vietnam," *Advanced Engineering Informatics*, vol. 42, Article ID 100978, 2019.
- [56] B. T. Pham, D. Tien Bui, and I. Prakash, "bagging based support vector machines for spatial prediction of landslides," *Environmental Earth Sciences*, vol. 77, p. 146, 2018.
- [57] Y. Huang and L. Zhao, "Review on landslide susceptibility mapping using support vector machines," *CATENA*, vol. 165, pp. 520–529, 2018.
- [58] W. Chen, H. R. Pourghasemi, and S. A. Naghibi, "A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China," *Bulletin of Engineering Geology and the Environment*, vol. 77, no. 2, pp. 647–664, 2018.
- [59] L. Y. Zhou, F. P. Shan, K. Shimizu, T. Imoto, H. Lateh, and K. S. Peng, "A comparative study of slope failure prediction using logistic regression, support vector machine and least square support vector machine models," *AIP Conference Proceedings*, vol. 1870, Article ID 060012, 2017.
- [60] D. Tien Bui, T. A. Tuan, N.-D. Hoang et al., "Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization," *Landslides*, vol. 14, no. 2, pp. 447–458, 2017.
- [61] D. Tien Bui, B. Pradhan, O. Lofman, and I. Revhaug, "Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naïve bayes models," *Mathematical Problems in Engineering*, vol. 2012, Article ID 974638, 26 pages, 2012.
- [62] H. Faris, I. Aljarah, and S. Mirjalili, "Improved monarch butterfly optimization for unconstrained global search and

- neural network training,” *Applied Intelligence*, vol. 48, no. 2, pp. 445–464, 2018.
- [63] G.-G. Wang, S. Deb, and Z. Cui, “Monarch butterfly optimization,” *Neural Computing and Applications*, vol. 31, no. 7, pp. 1995–2014, 2019.
- [64] S. L. Zubaidi, I. H. Abdulkareem, K. S. Hashim et al., “Hybridised artificial neural network model with slime mould algorithm: a novel methodology for prediction of urban stochastic water demand,” *Water*, vol. 12, no. 10, p. 2692, 2020.
- [65] A. M. Hussein, M. Abd Elaziz, M. S. M. Abdel Wahed, and M. Sillanpää, “A new approach to predict the missing values of algae during water quality monitoring programs based on a hybrid moth search algorithm and the random vector functional link network,” *Journal of Hydrology*, vol. 575, pp. 852–863, 2019.
- [66] D. J. Kalita, V. P. Singh, and V. Kumar, “A dynamic framework for tuning SVM hyper parameters based on moth-flame optimization and knowledge-based-search,” *Expert Systems with Applications*, vol. 168, Article ID 114139, 2020.
- [67] Z. Yu, X. Shi, J. Zhou, X. Chen, and X. Qiu, “Effective assessment of blast-induced ground vibration using an optimized random forest model based on a Harris hawks optimization algorithm,” *Applied Sciences*, vol. 10, no. 4, p. 1403, 2020.
- [68] H. Moayedi, M. Gör, Z. Lyu, and D. T. Bui, “Herding Behaviors of grasshopper and Harris hawk for hybridizing the neural network in predicting the soil compression coefficient,” *Measurement*, vol. 152, Article ID 107389, 2020.
- [69] E. H. Houssein, M. E. Hosney, M. Elhoseny, D. Oliva, W. M. Mohamed, and M. Hassaballah, “Hybrid Harris hawks optimization with cuckoo search for drug design and discovery in chemoinformatics,” *Scientific Reports*, vol. 10, Article ID 14439, 2020.
- [70] E. H. Houssein, M. E. Hosney, D. Oliva, W. M. Mohamed, and M. Hassaballah, “A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery,” *Computers & Chemical Engineering*, vol. 133, Article ID 106656, 2020.
- [71] D. Tien Bui, N.-D. Hoang, and P. Samui, “Spatial pattern analysis and prediction of forest fire using new machine learning approach of multivariate adaptive regression splines and differential flower pollination optimization: a case study at Lao Cai province (Viet Nam),” *Journal of Environmental Management*, vol. 237, pp. 476–487, 2019.
- [72] M.-Y. Cheng, D. Prayogo, and Y.-W. Wu, “Prediction of permanent deformation in asphalt pavements using a novel symbiotic organisms search-least squares support vector regression,” *Neural Computing and Applications*, vol. 31, no. 10, pp. 6261–6273, 2018.
- [73] N. Neggaz, E. H. Houssein, and K. Hussain, “An efficient henry gas solubility optimization for feature selection,” *Expert Systems with Applications*, vol. 152, Article ID 113364, 2020.
- [74] W. Chen, X. Chen, J. Peng, M. Panahi, and S. Lee, “Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer,” *Geoscience Frontiers*, vol. 12, no. 1, pp. 93–107, 2021.
- [75] L. B. Thao, *Vietnam, the Country and Its Geographical Regions*, The Gioi Publishers, Hanoi, Vietnam, 1997.
- [76] K. C. McGregor, R. L. Bingner, A. J. Bowie, and G. R. Foster, “Erosivity index values for Northern Mississippi,” *Transactions of the ASAE*, vol. 38, no. 4, pp. 1039–1047, 1995.
- [77] K. G. Renard, G. R. Foster, G. A. Weesies, D. K. McCool, and D. C. Yoder, *Predicting Soil Erosion by Water: a Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)*, *Agricultural Handbook 703*, US Department of Agriculture, Washington, DC, USA, 1997.
- [78] A. Klute, *Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods*, Soil Science Society of America, American Society of Agronomy, Madison, WI, USA, 1986.
- [79] C. Valentin, F. Agus, R. Alamban et al., “Runoff and sediment losses from 27 upland catchments in Southeast Asia: impact of rapid land use changes and conservation practices,” *Agriculture, Ecosystems & Environment*, vol. 128, no. 4, pp. 225–238, 2008.
- [80] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, Inc., Printed in the United States of America, Hoboken, NY, USA, 1998.
- [81] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, Berlin, Germany, 2011.
- [82] C. Qi, A. Fourie, G. Ma, X. Tang, and X. Du, “Comparative study of hybrid artificial intelligence approaches for predicting hangingwall stability,” *Journal of Computing in Civil Engineering*, vol. 32, Article ID 04017086, 2018.
- [83] A. Malik, Y. Tikhamarine, D. Souag-Gamane, O. Kisi, and Q. B. Pham, “Support vector regression optimized by meta-heuristic algorithms for daily streamflow prediction,” *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 11, pp. 1755–1773, 2020.
- [84] B. T. Pham, A. Jaafari, I. Prakash, and D. T. Bui, “A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling,” *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 4, pp. 2865–2886, 2018.
- [85] S. A. Naghibi, H. R. Pourghasemi, and K. Abbaspour, “A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS,” *Theoretical and Applied Climatology*, vol. 131, no. 3–4, pp. 967–984, 2018.
- [86] G. M. Hadjidemetriou, P. A. Vela, and S. E. Christodoulou, “Automated pavement patch detection and quantification using support vector machines,” *Journal of Computing in Civil Engineering*, vol. 32, Article ID 04017073, 2018.
- [87] N.-D. Hoang, Q.-L. Nguyen, and D. T. Bui, “Image processing-based classification of asphalt pavement cracks using support vector machine optimized by artificial bee colony,” *Journal of Computing in Civil Engineering*, vol. 32, Article ID 04018037, 2018.
- [88] M. S. Tehrany, B. Pradhan, S. Mansor, and N. Ahmad, “Flood susceptibility assessment using GIS-based support vector machine model with different kernel types,” *CATENA*, vol. 125, pp. 91–101, 2015.
- [89] A. Sina and D. Kaur, “Short term load forecasting model based on kernel-support vector regression with social spider optimization algorithm,” *Journal of Electrical Engineering & Technology*, vol. 15, no. 1, pp. 393–402, 2020.
- [90] H. Nguyen, Y. Choi, X.-N. Bui, and T. Nguyen-Thoi, “Predicting blast-induced ground vibration in open-pit mines using vibration sensors and support vector regression-based optimization algorithms,” *Sensors*, vol. 20, p. 132, 2020.
- [91] N.-D. Hoang and Q.-L. Nguyen, “A novel approach for automatic detection of concrete surface voids using image texture analysis and history-based adaptive differential evolution optimized support vector machine,” *Advances in*

- Civil Engineering*, vol. 2020, Article ID 4190682, 15 pages, 2020.
- [92] Z. Yu, X. Shi, J. Zhou et al., "Prediction of blast-induced rock movement during bench blasting: use of gray wolf optimizer and support vector regression," *Natural Resources Research*, vol. 29, no. 2, pp. 843–865, 2020.
- [93] A. Jaafari, M. Panahi, B. T. Pham et al., "Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility," *CATENA*, vol. 175, pp. 430–445, 2019.
- [94] S. Deng, X. Wang, Y. Zhu, F. Lv, and J. Wang, "Hybrid grey wolf optimization algorithm based support vector machine for groutability prediction of fractured rock mass," *Journal of Computing in Civil Engineering*, vol. 33, Article ID 04018065, 2019.
- [95] X.-N. Bui, P. Jaroonpattanapong, H. Nguyen, Q.-H. Tran, and N. Q. Long, "A novel hybrid model for predicting blast-induced ground vibration based on k-nearest neighbors and particle swarm optimization," *Scientific Reports*, vol. 9, Article ID 13971, 2019.
- [96] D. Prayogo and Y. T. T. Susanto, "Optimizing the prediction accuracy of friction capacity of driven piles in cohesive soil using a novel self-tuning least squares support vector machine," *Advances in Civil Engineering*, vol. 2018, Article ID 6490169, 9 pages, 2018.
- [97] Z. M. Yaseen, H. Faris, and N. Al-Ansari, "Hybridized extreme learning machine model with salp swarm algorithm: a novel predictive model for hydrological application," *Complexity*, vol. 2020, Article ID 8206245, 14 pages, 2020.
- [98] H. A. Afan, M. F. Allawi, A. El-Shafie et al., "Input attributes optimization using the feasibility of genetic nature inspired algorithm: application of river flow forecasting," *Scientific Reports*, vol. 10, p. 4684, 2020.
- [99] V.-H. Nhu, P.-T. Thi Ngo, T. D. Pham et al., "A new hybrid firefly-PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping," *Remote Sensing*, vol. 12, no. 17, p. 2688, 2020.
- [100] A. P. Piotrowski, "L-SHADE optimization algorithms with population-wide inertia," *Information Sciences*, vol. 468, pp. 117–141, 2018.
- [101] R. Tanabe and A. S. Fukunaga, "Improving the search performance of SHADE using linear population size reduction," in *Proceedings of the 2014 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1658–1665, Beijing, China, July 2014.
- [102] J. Zhang and A. C. Sanderson, "JADE: adaptive differential evolution with optional external archive," *IEEE Transactions on Evolutionary Computation*, vol. 13, pp. 945–958, 2009.
- [103] K. Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution - A Practical Approach to Global Optimization*, Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2005.
- [104] R. Tanabe and A. Fukunaga, "Success-history based parameter adaptation for Differential Evolution," in *Proceedings of the 2013 IEEE Congress on Evolutionary Computation*, pp. 71–78, Cancun, Mexico, June 2013.
- [105] Accord, "Accord.NET framework," 2019, <http://accord-framework.net/>.
- [106] G. Montavon, G. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2012.
- [107] I. Kononenko, *Estimating Attributes: Analysis and Extensions of RELIEF*, Springer, Berlin, Heidelberg, 1994.
- [108] R. C. Prati, G. E. A. P. A. Batista, and D. F. Silva, "Class imbalance revisited: a new experimental setup to assess the performance of treatment methods," *Knowledge and Information Systems*, vol. 45, no. 1, pp. 247–270, 2015.
- [109] D. Tien Bui, N.-D. Hoang, F. Martínez-Álvarez et al., "A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area," *Science of The Total Environment*, vol. 701, Article ID 134413, 2020.
- [110] T.-D. Nguyen, T.-H. Tran, and N.-D. Hoang, "Prediction of interface yield stress and plastic viscosity of fresh concrete using a hybrid machine learning approach," *Advanced Engineering Informatics*, vol. 44, Article ID 101057, 2020.
- [111] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. D. Jesús, *Neural Network Design*, Vikas Publishing House, Noida, India, 2nd edition, 2014.
- [112] Matwork, "Statistics and machine learning toolbox user's guide: Matwork Inc.," 2017, [https://www.mathworks.com/help/pdf\\_doc/stats/stats.pdf](https://www.mathworks.com/help/pdf_doc/stats/stats.pdf).
- [113] J. Heaton, *Artificial Intelligence for Humans, Volume 3 Deep Learning and Neural Networks*, Heaton Research, Inc., St. Louis, MO, USA, 2015.
- [114] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
- [115] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.
- [116] A. R. van Erkel and P. M. T. Pattynama, "Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology," *European Journal of Radiology*, vol. 27, no. 2, pp. 88–94, 1998.
- [117] W. Chen, Y. Chen, P. Tsangaratos, I. Ilia, and X. Wang, "Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments," *Remote Sensing*, vol. 12, no. 23, p. 3854, 2020.
- [118] W. Chen, X. Zhao, P. Tsangaratos et al., "Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping," *Journal of Hydrology*, vol. 583, Article ID 124602, 2020.
- [119] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, John Wiley & Sons, Hoboken, NJ, USA, 1999.
- [120] V.-H. Dang, N.-D. Hoang, L.-M.-D. Nguyen, D. T. Bui, and P. Samui, "A novel GIS-based random forest machine algorithm for the spatial prediction of shallow landslide susceptibility," *Forests*, vol. 11, no. 1, p. 118, 2020.
- [121] B. Everitt, *The Cambridge Dictionary of Statistics*, Cambridge University Press, Cambridge, UK, 1998.
- [122] T. Whelen and P. Siqueira, "Coefficient of variation for use in crop area classification across multiple climates," *International Journal of Applied Earth Observation and Geoinformation*, vol. 67, pp. 114–122, 2018.
- [123] N.-D. Hoang, "Image processing-based pitting corrosion detection using metaheuristic optimized multilevel image thresholding and machine-learning approaches," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6765274, 19 pages, 2020.
- [124] M. Kouli, P. Soupios, and F. Vallianatos, "Soil erosion prediction using the revised universal soil loss equation (RUSLE) in a GIS framework, chania, northwestern crete, Greece," *Environmental Geology*, vol. 57, no. 3, pp. 483–497, 2009.

- [125] I. Gaubi, A. Chaabani, A. Ben Mammou, and M. H. Hamza, "A GIS-based soil erosion prediction using the revised universal soil loss equation (RUSLE) (lebna watershed, cap bon, Tunisia)," *Natural Hazards*, vol. 86, no. 1, pp. 219–239, 2017.
- [126] H. R. Pourghasemi, N. Sadhasivam, N. Kariminejad, and A. L. Collins, "Gully erosion spatial modelling: role of machine learning algorithms in selection of the best controlling factors and modelling process," *Geoscience Frontiers*, vol. 11, no. 6, pp. 2207–2219, 2020.
- [127] A. Arabameri, W. Chen, M. Loche et al., "Comparison of machine learning models for gully erosion susceptibility mapping," *Geoscience Frontiers*, vol. 11, no. 5, pp. 1609–1620, 2020.
- [128] O. Ghorbanzadeh, H. Shahabi, F. Mirchooli et al., "Gully erosion susceptibility mapping (GESM) using machine learning methods optimized by the multi collinearity analysis and K-fold cross-validation," *Geomatics, Natural Hazards and Risk*, vol. 11, no. 1, pp. 1653–1678, 2020.