Convolutional Neural Network—Optimized Moth Flame Algorithm for Shallow Landslide Susceptible Analysis

VU DONG PHAM¹, QUOC-HUY NGUYEN²¹, HUU-DUY NGUYEN², VAN-MANH PHAM², VAN MANH VU³, AND QUANG-THANH BUI¹

¹Center for Applied Research in Remote Sensing and GIS (CARGIS), Faculty of Geography, VNU University of Science, Hanoi 120041, Vietnam
²Faculty of Geography, VNU University of Science, Hanoi 120041, Vietnam
³Faculty of Environmental Science, VNU University of Science, Hanoi 120041, Vietnam

Corresponding author: Quang-Thanh Bui (qthanh.bui@gmail.com)

This work was supported by the Asia Research Center, Vietnam National University - Hanoi and the Korea Foundation for Advanced Studies under Grant CA.19.8A.

ABSTRACT Convolutional neural network (CNN) is a widely used method in solving classification and regression applications in industries, engineering, and science. This study investigates the optimizing capability of a swarm intelligence algorithm named moth flame optimizer (MFO) for the optimal search of a CNN hyper-parameters (values of filters) and weights of fully connected layers. The proposed model was run with a 3-dimensional dataset (7 width × 7 height × 12 depth), which was constructed through including seven neighbor pixels (vertically and horizontally) from landslide location and 12 predictor variables. Muong Te district, Lai Chau province, Vietnam was selected as the case study, as it had recently undergone severe impacts of landslides and flash floods. The performance of this proposed model was compared with conventional classifiers, i.e., Random forest, Random subspace, and CNN-optimized Adaptive gradient descend, by using standard metrics. The results showed that the CNN-optimized MFO (Root mean square error = 0.3685, Mean absolute error = 0.2888, Area under Receiver characteristic curve = 0.889 and Overall accuracy = 80.1056%) outperformed the benchmarked methods in all comparing indicators. Besides, the statistical test of difference was also carried out by using the Wilcoxon signed ranked test for non-parametric variables. With these statistical measurements, the proposed model could be used as an alternative solution for landslide susceptibility mapping to support local disaster preparedness plans.

INDEX TERMS Convolutional neural network, meta-heuristic algorithm, moth flame optimization algorithm, landslide susceptibility.

I. INTRODUCTION Landslide is among the deadliest natural hazards that cause human loss, severe damages to infrastructures, and have negative impacts on social-economic development [1]. So, it is vital to map vulnerable areas to landslide to eliminate potential damages and to restrict growth in hazardous locations. The landslide susceptibility measures the probability of how likely landslide occurs in an area from small to large scales, and it has been implemented by different techniques in numerous case studies across the mountainous region of tropical countries. Examples of previously published works could be found in the literature with statistical analysis methods [2], analytical hierarchical process [3], probabilistic approach [4], [5], weight of evidence [6], [7], evidential belief function [2], [8] and multivariate regression [9].

Recently, the information mining techniques, geographic information systems are considered as useful methods for natural hazard analysis as well as for landslide susceptibility assessment and prediction [10]–[13], in which non-linear relations between historical landslides and physical, climatic and anthropogenic factors were successfully modeled. Besides, various studies had investigated hybrid methods to assess susceptibility to landslides [14]–[18], in which the parameters of classifiers were fine-tuned by meta-heuristic optimizers.

Convolutional neural network (CNN) is a widely used method in handling images for object detection, pattern recognition, land cover classification, and natural
hazard analysis. For landslide, CNN can be used for either the detection of historical landslide locations by using air-born, space-born, Unmanned aerial vehicle (UAV) data, or landslide susceptibility mapping. In both cases, CNN was found robust as it outperformed several benchmarked methods, as discussed in the studies of [19]–[21]. The differences of these works relied on the formulation of CNN structure (precisely the depth of CNN, number of feature maps, filter dimensions, filter values and depths of fully connected layers), and resampling methods for training and validation datasets. The diversities of actual problems and complexity of physical and human-made environments of landslide events make it hard to define a conventional CNN model for all cases. Therefore, new investigations of CNN structures and optimization algorithms are useful contributions to general knowledge in the field of landslide studies as well as natural hazard analysis. To the best experience of authors in this field, the implementation of CNN in landslide susceptible mapping is few, and the integration of population-based optimization algorithms into CNN is still limited.

Population intelligence-inspired algorithms had been successfully investigated in a broad range of research topics such as land classification [22]–[24], forest fire [25], flood analysis [18] as well as in landslide applications [26]. However, as the applications are so broad and no optimization algorithm fits all problems [27], therefore the verification of new optimizers is necessary. The purpose of this study is to investigate the potential use of a meta-heuristic algorithm in tuning the CNN for landslide applications. This paper designed a CNN structure to include a swarm-intelligent optimization algorithm in searching for optimal hyper-parameters of filters and weights of the fully connected layers. Moth Flame Optimization (MFO) algorithm was used as a replacement for conventional gradient descent algorithm, which had been commonly implemented in previous CNN studies. The proposed model (CNNFMO) was verified with a trainset, which was collected from Lai Chau province of Vietnam. The study area is a mountainous province which had been frequently impacted by landslides and other natural disasters. From the study area, regarding previous studies, twelve predictor variables, namely Digital elevation model (DEM), Aspect, Slope, Compound topographic index (CTI), Stream power index (SPI), Normalized difference vegetation index (NDVI), Normalized difference water index (NDWI), Normalized difference build-up index (NDBI), Distance from river, River density, curvature, and accumulated precipitation were selected. The data was processed by QGIS, and the proposed model was built by the authors in Matlab R2018.

II. STUDY AREA AND DATA USED

A. STUDY AREA AND HISTORICAL LANDSLIDE OCCURRENCES

Lai Chau is a mountainous province in the Northwest region of Vietnam with complex topography such as profound and narrow valleys, dense rivers, and streams (Figure 1). This topography shapes the moisture circulation and has effects on the micro-weather. In the rainy season, the maximum rainfall is between June to August. Annual values range from 2,000 mm to 2,500 mm, and the dry season starts from October to March next year. Forests cover the study area. However, deforestation, which is related to hydropower development, agriculture, and illegal logging, also contributes to the frequent occurs of landslides. This province was selected as the study area because of its high vulnerability to landslide hazards. Recently, the province had undergone severe impacts of landslides and flash floods that caused significant loss of humans and massive loss of economic properties. Therefore, susceptible maps of natural hazards are necessary.
Technically, the historical landslide occurrences can be detected through analyzing high-resolution satellite imageries, airborne, or UAV data by comparing reflectance differences and surface changes between two dates of the dataset. This method is used in conjunction with field surveys to identify areas that are not detectable from images or the areas that had been recovered under human intervention. In this study, the landslide areas were interpreted from satellite, airborne data in combination with field surveys for validation. The historical landslide occurrences were represented by polygons, as shown in Figure 1 (a) with a boxplot describing the variation of polygon sizes, in which the smallest landslide area was around 10 m², the largest was around 51,000 m², and the mean area is about 250 m².

In numerous previous studies on landslide susceptibility [7], [13], [28], [29], the landslide locations were defined as points, which were measured from field surveys with x and y coordinates. Then the predictor variables were extracted from raster layers, in which each landslide point receives attribute values from the pixels, sharing the same locations. At the screening stage of input data, landslide polygons were pre-processed to filter out small polygons (smaller than 100 m²). This threshold was chosen because the area value is more than half of the DEM size (12.5m DEM was used as the spatial reference for data resampling and conversion). The landslide polygons (larger than 100 m²) were converted to a raster layer, and landslide locations were extracted as the centers of the pixels in this raster. In fact, the center points are good representations of raster pixels in spatial analysis. Besides, the proposed model is a binary classification method (either landslide or non-landslide). Therefore, a similar number of non-landslide points were randomly identified across the study area to build up the trainset. The final training dataset consisted of 2374 points (historical landslides and randomly selected non-landslides) with associated twelve attributes and was randomly divided into a training set (70% or 1662 samples) and validation set (30% or 712 samples) (Figure 1 b). This 70/30 ratio has been proposed in numerous landslide susceptibility studies [30] and was also applied in this work.

**B. PREDICTOR VARIABLES**

The determination of the resolution of predictor variables must be coherent with this variation of occurrence areas. However, the selection of training datasets is also subject to data availability and accessibility. The 12.5m DEM from the ALOS-PALSAR, which is downloadable from https://asf.alaska.edu/, is a suitable choice because of its free downloads and spatially satisfactory.

With the open development of spatial data, and data sharing policies of several data providers such as NASA, ESA, or GEOSS portal, the accessibility to spatial data is more comfortable. In susceptibility mapping studies, input data are essential factors for the classification accuracy of the landslide model. In general, from the literature review, these data can be divided into three groups: physical geography, climatic, and anthropic activities. In this study, twelve factors were chosen, including elevation, slope, aspect, curvature, SPI, distance to river, river density, NDVI, NDBI, NDWI, CTI, and accumulated rainfall. Lithology, which is measured as nominal data, was not considered in this study, because of the requirement of continuous input data from the proposed CNN model. Since the original measurements of twelve layers were in different units, they were normalized to [0 1] range before feeding into the CNN.

The study area is covered by a large part of the forest, which is an essential factor in soil stabilization to fight against landslide. Satellite data plays a crucial role in monitoring surface cover and land use using several indices such as NDVI to monitor the vegetation covers. It is considered as the influential factor in many landslide studies [13], [31]. NDWI is known to be strongly related to plant water content. It is, therefore, a good proxy for plant water stress and is also associated with the soil condition. Besides, landslide occurrence is not only influenced by natural causes but also by human intervention such as road construction and settlement. These land cover types are impervious areas, which affect water runoff and might have specific impacts on locations of landslides. NDBI has been useful for mapping built-up areas and other land cover types and was considered as a potential variable in this study. These indices were calculated from the Landsat 8 OLI image.

In the study area, surface water and groundwater are the main hydrological factor that causes landslides through eroding slopes. The fluctuation of groundwater causes changes in the water pressure in the soil and the stability of the slope. They are represented by CTI, SPI which were derived from the DEM. Distance to the river measures the proximity to rivers or streams, as the closer the area to rivers or streams, the soils become more saturated. River density measures how well or how poorly an area is drained by streams. These two factors are commonly used in previously published works and were included as input variables for mapping the landslide susceptibility. Finally, rainfall is considered either long-term or short-term effects on triggering a soil volume to slide down. In this work, accumulated rainfall was calculated by summing precipitations of four months in the rainy season of 2018.

Since the predictor variables were extracted from satellite data with different spatial resolutions, therefore data processing is necessary to resample those into similar data formats. DEM from the ALOS-PALSAR source was used as a target raster format (12.5m × 12.5m) for the conversions of all variables. All layers were prepared in WGS84/UTM zone 48. In preparing the input datasets, the training points were assigned attribute values from twelve variables through extracting values of the layers which have similar spatial locations. The specification summary of predictor variables and historic landslide occurrences are shown in Table 1.
TABLE 1. Sources and specification of training data.

| ID | Variables          | Resolution       | Source                                      |
|----|--------------------|------------------|---------------------------------------------|
| 1  | DEM                | 12.5m x 12.5m    | ALOS – PALSAR DEM, which is downloadable from [9] |
| 2  | Aspect             | 12.5m x 12.5m    | Calculated from DEM                         |
| 3  | Slope              | 12.5m x 12.5m    | Calculated from DEM                         |
| 4  | CTI                | 12.5m x 12.5m    | Calculated from DEM                         |
| 5  | SPI                | 12.5m x 12.5m    | Calculated from DEM                         |
| 6  | Curvature          | 12.5m x 12.5m    | Calculated from DEM                         |
| 7  | NDVI               | 12.5m x 12.5m    | These variables were calculated from Landsat 8 OLI (30m) and resampled to 12.5m by using DEM as the reference |
| 8  | NDWI               | 12.5m x 12.5m    | These variables were calculated from Landsat 8 OLI (30m) and resampled to 12.5m by using DEM as the reference |
| 9  | NDBI               | 12.5m x 12.5m    | River network was extracted from the Topographic map at 1:50,000. And these variables were calculated from the river network by using DEM as the spatial reference |
| 10 | Distance to river  | 12.5m x 12.5m    | River network was extracted from the Topographic map at 1:50,000. And these variables were calculated from the river network by using DEM as the spatial reference |
| 11 | River density      | 12.5m x 12.5m    | Rain raster was interpolated from the National Meteo stations by using DEM as the spatial reference |
| 12 | Rain               | 12.5m x 12.5m    | Rain raster was interpolated from the National Meteo stations by using DEM as the spatial reference |
| 13 | Historical landslides occurrences | Points | The landslide polygons were collected through field surveys and satellite analysis. They were converted to rasters, and the cell centers were created to represent the locations of landslide occurrences |

III. METHODS

A. CONVOLUTATIONAL NEURAL NETWORK

The Convolutional neural network (CNN) is a type of artificial neural networks (ANNs) and has good performances in many computer vision tasks such as image classification [32], real-time object detection [33], object segmentation [34]. The idea behind the high performance of CNN architectures is by using the ability of convolutional layers for detecting local conjunctions of features, and pooling layers for merging semantically similar features into one [35]. This approach shows that CNN is not only much less computationally expensive than the fully connected structure of ANNs but also performing more effectively in computer vision tasks. In general, a CNN architecture consists of a sequence of layers, namely Convolutional layers, Pooling layers, and Fully connected layers (like regular layers in ANNs). Every layer of CNN transforms one volume of activations to another through a differentiable function. These layers are computed following bellow formulas:

Convolutional layer:

\[ h_{ij} = \sum_{k=1}^{m} \sum_{l=1}^{n} w_{k,l} x_{i+k-1,j+l-1} \]

Pooling layer (max pooling):

\[ h_{ij} = \max \{ x_{i+k-1,j+l-1} \mid 1 \leq k \leq \text{mand}, 1 \leq l \leq m \} \]

Fully connected layer:

\[ h = \sum_{i} w_{i} x_{i} \]

With \( h_{ij} \) is the output result at the location \((i, j)\) of the next layer with input \( x \) and filter \( w, m \) is the filter width and height.

Commonly, a non-linear function will be applied for convolutional and fully connected layers. For example, some popular functions such as ReLU (Rectified Linear Unit), Sigmoid, Tanh are applied to the output layer to convert all negative values into zeros.

\[ \text{ReLU}(h) = \max(0, h) \]

\[ \text{Sigmoid}(h) = \frac{1}{1 + e^{-h}} \]

\[ \text{Tanh}(h) = \frac{2}{1 + e^{-2h}} - 1 \]

B. MOTH FLAME OPTIMIZATION ALGORITHM

Nature-inspired optimization algorithms are fast-growing disciplines through mimicking behaviors of animals in shaping their movements, searching for foods, and forming swarm social structures. The number of published works in this field has dramatically increased to find suitable solutions for diverse specific problems. MFO is a novel algorithm that has been proposed and developed by [36] to mathematically model the navigation of moths in referencing to light sources (the moon, artificial lights). With artificial lights, moths tend to navigate in a spiral path towards to light source. Mathematically, moths move in a swarm, and their positions are modeled in a swarm, and their positions if a better solution is found by using the logarithmic spiral, as shown in Eq 1. The selection of spiral functions is discussed in more detail [36].

\[ S(X_{i}, F_{i}) = \text{Distance}_{i} \times e^{bt} \times \cos(2\pi t) + F_{j} \]  

(1)

where \( \text{Distance}_{i} \) is the Euclidean distance in \( D \) dimensional space from the \( i^{th} \) moth to the \( j^{th} \) flame; \( b \) is a constant, and \( t \) is a random value between \([-1, 1]\).

- The moths move around the flames and update their positions if a better solution is found by using the logarithmic spiral, as shown in Eq 1. The selection of spiral functions discussed in more detail [36].

\[ S(X_{i}, F_{i}) = \text{Distance}_{i} \times e^{bt} \times \cos(2\pi t) + F_{j} \]  

(1)

where \( \text{Distance}_{i} \) is the Euclidean distance in \( D \) dimensional space from the \( i^{th} \) moth to the \( j^{th} \) flame; \( b \) is a constant, and \( t \) is a random value between \([-1, 1]\).

- Moths are sorted in name order. After each iteration, flames are sorted by their fitness values, and moths move around the new flames in the next iteration. This mechanism ensures the exploration and exploitation cover the search space and avoid local minimum stagnation.

- The number of flames is adaptively decreased to maintain the exploitation of the best promising positions by the following equation

\[ \text{Number of flames} = \text{round}(N - l \times (N - 1)/T) \]  

(2)

where \( l \) is current iteration, \( N \) is the maximum number of flames and \( T \) is the maximum number of iterations.

C. BENCHMARKED ALGORITHMS

The proposed model was run with the training set and assessed by validation set as previously described, and the
performance was compared to several benchmarked methods. From the literature review, several ensemble techniques such as Random Forest (RF), Random subspace (RS) were also used as benchmarked functions as they had been successfully implemented for landslide studies [1]. Besides, the CNNFMO was also trained by the conventional Adagrad algorithm, in which the learning rate is adaptively changed after each iteration, performing smaller updates (i.e., low learning rates) for parameters associated with frequently occurring features, and larger updates (i.e., high learning rates) for parameters related to the infrequent feature. The configurations of these methods were determined through a trial-and-error process.

### D. PROTOTYPE MODEL AND CONFIGURATION

The study of [31] validated the potential uses of CNN in landslide susceptible mapping by using three different 1D, 2D, and 3D trainsets. This study proposed and investigated a CNN, which is similar to LeNet-5 [8], with 3-dimensional input instances and MFO as a replacement for conventional gradient descent algorithms. The illustration of this model is shown in (Figure 3). At first, the training datasets were prepared, and each landslide location was used to construct the 3-dimensional instance of $7 \times 7 \times 12$ through taking 49 neighbor pixels, and twelve predictor variables. Secondly, the input datasets were fed into convolutional/ReLU layers by using 16 cube filters, in which the dimension of each was defined by $3 \times 3 \times 3$. Sixteen feature maps were further processed with Max-pooling layers of $2 \times 2 \times 2$. The output was flattened before feeding to the fully connected layer, which has one hidden layer of 50 neurons. Depending on the predefined threshold, the output of this model could be identified as landslide or non-landslide.

From the configuration of the CNN (16@3×3×3 filters, one hidden layer in the fully connected layer with 50 neurons respectively), which has 16482 adaptive variables, the adaptive variables are encoded as dimensions of the search space for FMO. RMSE is used as the cost function (Eq3) and the algorithm iterates 1000 by using a training dataset and validated by using the validation dataset. The termination point could be either RMSE (as the lost function) is achieved, or the maximum number of iterations is reached. The optimal position of the first moth (represented by dimensional values in the search space) is used as optimal values for weights and filters of the proposed CNN, which in turn is used to generate landslide susceptibility map for the entire study area.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{predicted}_i - \text{observed}_i)^2}
\] (3)

One of the aspects that need proper configuration (usually through the trial-error process) includes the setting of
moth flame optimization algorithm and benchmark functions. Apart from already defined parameters of the proposed CNN structure, the selection of parameters for MFO also contributed to the performance of the model, such as controlling convergence speed and exploitation or exploration searches. Trial and errors were the only way to determine the most suitable parameters for the current dataset of this study. The maximum number of iterations was set at 1000; The searchable moths and flames were 30 for each group; Lower and upper bounds $[-1, 1]$.

### E. ACCURACY ASSESSMENT

Apart from RMSE, which is used as the objective function (loss function) in for the search of weights of the CNN, other statistical measurements such as Mean absolute error (MAE), Overall accuracy (OA) and Area under Receiver Operating Characteristics (AUC) are also useful indicators for model comparison. MAE measures fitting error and has a certain level of similarity to RMSE. OA is a standard indicator for classification application, which measures the percentage of correctly classified pixels over entire training or validation datasets. AUC plots the false position rate in the x-axis against true positive rate in the y-axis to measure the areas under ROC per the variation of true-false thresholds. This indicator is typically used to compare the performances of different models.

### IV. RESULTS AND DISCUSSION

#### A. DATA PRE-PROCESSING AND EXPLORATION

The preprocessing of train datasets plays a crucial role in machine learning applications, aiming at improving the performance of machine learning models. Typically, this process starts with a pair-wise analysis of predictor variables and determines the most productive features in referencing the outputs of the training dataset. Even though the proposed CNN performed feature selection or learned the features as well during the tuning operation, the pre-processing was the very first step to filter out feature redundancy and to ensure the consistency of the overall performance of the machine learning model. Figure 4 shows the rank of all predictor variables which were implemented by using Random Forest algorithms. It could be seen that Elevation and rivers have the strongest influences on the landslide occurrences, and vegetation condition was ranked in the successional order in both Mean Decrease Accuracy and Mean Decrease Gini. The ranking orders were also found similar in previously published works [14], [17], [26].

The approach of this paper took into consideration the spatial dependency of historic landslide locations and their surrounding areas by including neighbor pixels of all predictor variables into the training dataset. This method had been discussed in the study of [31], and a neighbor area of $7 \times 7$ pixels was proposed for each training point. Technically, the dimension of the input dataset was subject to the nature of study areas and type, resolution of remotely sensed data. In this study, the spatial resolution of elevation derived variables and indexes from ALOS PALSAR were defined by $12.5m \times 12.5m$ so that 49 neighbor pixels ($7 \times 7$) would cover an area of around 7600 square meters. This specific area was twice larger than almost all historic landslide locations, which had been presented in the previous section. Therefore the $7 \times 7 \times 12$ training sets would be appropriate for the verification of the proposed CNN in this study.
The training dataset included more than 1662 points, and their associated values, which were extracted from 12 predictor variables, were fed into the proposed CNN model. The hyper-parameters of filters and weights of the fully connected were considered dimensions of moths, and the model searched for the most optimal moth with the smallest RMSE. These values were randomly set during the initiation phase and were updated after each iteration. After 1000 iteration, the RMSE for the validation dataset ended at 0.3685, and the shape of the variation curve was shown in Figure 5 in which horizontal lines represented the exploitation search, and the vertical jumps show new positions of the exploration search. It could be seen from the figure that RMSE curves for validation data and training datasets follow a similar pattern, and the over-fitting problem does not seem to exist.

This study verifies the potential use of the combination between the meta-heuristic algorithm and CNN for landslide susceptibility mapping. The performance of this method was compared to several benchmarked methods which had been commonly used in landslide studies. Two ensemble methods such as RF, RS, and the CNN optimized Adagrad were used for benchmarking. Table 2 shows several statistical measures, in which the proposed method outperformed RF (RMSE = 0.4142, AUC = 0.831, OA = 75.9259%), RS (RMSE = 0.3847, AUC = 0.870, OA = 78.3333%) and CNN-Adagrad (RMSE = 0.3847, AUC = 0.870, OA = 78.3333%). Figure 6 visualizes the changes of False Positive Rate in the horizontal axis against True Positive Rate as described in the vertical axis. In the case of RF and RS, the curves crossed over each other in some points, but the red curve was higher than the remaining for all threshold values, which were used to determine landslide or non-landslide status of instances. The ROC variations are another sign to ensure the robustness of the proposed methods. It could be seen that the proposed CNN outperformed the other and could be considered as an alternative solution for landslide susceptibility analysis in similar areas.

An additional test was carried out to determine whether the performance of models was statistically different. Non-parametric Wilcoxon signed-rank test was chosen because of the unknown distribution of variables, and no assumption is required. Each pair of two models was tested by using AUC values with the Null hypothesis stating that there was no difference between the predictive capabilities of the two selected models. The results in (Table 3) show that all p-values are smaller than 0.05, that the null hypothesis would be rejected, and the differences were statistically significant.

Finally, the landslide susceptibility map was created by estimating the susceptible level of all locations for the entire study area. This step was implemented through feeding all locations with associated twelve predictor variables into the proposed model, which was fine-tuned and validated as described in the previous section. For each location, the algorithm constructed a 3D (7 × 7 × 12) sample and outputted susceptible value. The values ranged from 0 - 1 and were qualitatively broken down into 5 classes, namely Very Slow [0 - 0.18], Low [0.19 - 0.36], Moderate [0.37 - 0.54], High [0.55 - 0.73], and Very High [0.74 - 1]. It could be seen that the most susceptible area was clustered around the rivers and suited to the observed historical landslides. This map provided useful information for local authorities for coordinating preparedness plans for mitigation of negative impacts.

1) MODEL PERFORMANCE COMPARISON AND DISCUSSION

The proposed CNN resulted in RMSE = 0.3685, MAE = 0.2888, AUC = 0.889 and OA = 80.1056% with data in the mountainous area of Vietnam, together with results from several benchmark functions (TABLE 2). The significance
test was also implemented to show that the proposed CNN was statistically different from the other, or by another mean, this model outperforms the other with such a training dataset. The results are satisfactory in comparison to previously published works in landslide susceptibility mapping such as Ada and San [37] (AUC from 0.82 and 0.87), Nguyen, et al. [28] (highest AUC = 0.826), [38] by using Random Subspace and Classification And Regression Trees (AUC = 0.841), [31] with 2D CNN structure (AUC = 0.813), [29] by ensemble methods (AUC = 0.865). The AUC values of this model and previously published works should not be used for direct comparison because of the differences in training data and because machine learning models might perform differently from case to case [27]. Technically, the approach of this study takes into consideration the inclusion of neighbor pixels in assessing the susceptibility level of a location. This approach is different from almost all previously published works when landslide locations were defined as point data, no matter what spatial resolution of input rasters was used. The determination of $7 \times 7$ pixels in the training dataset had been successfully validated in the works of [31], and this resampling method was also proved to be useful in this case study (through comparing to benchmark functions). These comparisons provide firms evidence that the training of the proposed model is successful, and the generated results are useful to help decision-makers of the land use planning, and risk management of landslide in mountainous areas in Vietnam as well as in other countries.

In machine learning, generalization is a challenge, in which either under-fitting or over-fitting results in poor performance of classification models. By far for deep learning, over-fitting is critical and more challenging to solve than under-fitting as models are designed with a high level of complexity that leads to the learning of background noise or random fluctuations. To eliminate this problem, the gathering of more training data is the first choice. Still, it might not be achievable due to the limited accessibility of data such as physical constraint in approaching the entire study area, difficulty in collecting historical datasets. In this study, the regularization technique was applied through defining the search boundary (lower bound and upper bound) of MFO to prevent the model from taking extreme values for connecting weights. The hold-out method was used for training and validation of the model, and the variations of RMSEs in (Figure 5) show that the overfitting problems do not exist since the changes of both curves follow similar patterns after 1000 iterations.

The selection of landslide conditioning factors is the essential element that should be taken into consideration because it influences the quality of classification models. The selection depends on the characteristics of the study area, type of landslides such as rainfall or earthquake-induced landslides, and modeling methods. In this study, twelve predictor variables were proposed as they have been used in several previously published works, and they are globally collectible. Besides, the implementation of $[0 - 1]$ normalization would keep the original distribution of data and prevent the potential intervention of humans during the preparation of the input dataset.

Spatial prediction for landslides is considered a challenging task for risk management. That’s why the exploitation of new methods and techniques with high accuracy is essential. This paper addresses this problem by using the convolutional neural network – optimized moth flame algorithm.
This algorithm differs from other meta-heuristic methods through behaviors of moths in flying to light sources and has a certain level of balancing exploitation and exploration processes. The search capability of FMO, as well as other algorithms, are subject to the dataset, specifically to dimensions of data, data size, normalization techniques, and even to similar subjects but in different geographic locations.

V. CONCLUSION AND FUTURE REMARKS

This study is among the few first works on the customization of CNNs for landslide susceptible analysis with a case study in a tropical mountainous area of Vietnam. The CNNFMO outperforms two ensemble methods and Adagrad optimizer in all statistical measurements, specifically it ends up with RMSE = 0.3685, MAE = 0.2888, AUC = 0.889 and OA = 80.1056. Besides, the tradeoff between model complexity and over-fitting problems was also taken into consideration by introducing several tactics such as normalization of input data, setting a boundary for search space, random dropout at the fully connected layer. The neighborhood effect was also considered by including surrounding pixels of landslide locations for twelve predictor variables that result in a significant increase in the number of pixels to improve the quality of the classification model. Precise detection of landslide areas or landslide susceptibility mapping is two conventional research approaches that take advantage of machine learning development and geospatial information technology. The applications of these techniques in hazard-prone areas are growing when spatial data are widely accessible in both space and time scales and the steady introduction of new meta-heuristic optimization algorithms. The potential uses of meta-heuristic optimization algorithms are promising in combination with machine learning and spatial data for natural hazard studies. Reportedly, more than 40,000 insect swarm had been investigated, and the mimic potentiality of their behaviors in solving engineering, industrial, and scientific problems is under investigation. It had been mentioned in many previously published works that the accuracy of machine learning models are subject to the training dataset, and the performance varies cases by cases. Therefore, the investigation of potential uses of such algorithms with different training datasets is essential to improve the knowledge in the field of natural hazard management.

CONFLICT DECLARATION

Authors have no conflict of interest.

REFERENCES

[1] W. Chen, M. Panahi, and H. R. Pourghasemi, “Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling,” Catena, vol. 157, pp. 310–324, Oct. 2017.

[2] M. Juliev, M. Mergili, I. Mondal, B. Nurtayev, A. Pulatov, and J. Hübl, “Comparative analysis of statistical methods for landslide susceptibility mapping in the Bostanlik District, Uzbekistan,” Sci. Total Environ., vol. 655, pp. 801–814, Feb. 2019.

[3] I. Milevski and S. Dragičević, “Landslides susceptibility zonation of the territory of north macedonia using analytical hierarchy process approach,” Contributions, vol. 40, no. 1, p. 115, Jun. 2019.

[4] S. Lari, F. Frattini, and G. B. Crosta, “A probabilistic approach for landslide hazard analysis,” Eng. Geol., vol. 182, pp. 3–14, Nov. 2014.

[5] S. Mondal and S. Mondal, “Probabilistic approaches and landslide susceptibility,” in Geoinformatics and Modelling of Landslide Susceptibility and Risk, Cham, Switzerland: Springer, 2019, pp. 145–163.

[6] A. Gadtaula and S. Dhakal, “Landslide susceptibility mapping using weight of evidence method in Haku, Rasuwa District, Nepal,” J. Nepal Geol. Soc., vol. 58, pp. 163–171, Jul. 2019.

[7] V. Kumar and R. Anbalagan, “Landslide susceptibility mapping of the Tehri reservoir rim area using the weights of evidence method,” J. Earth Syst. Sci., vol. 128, p. 153, 2019.

[8] Z. Chen, S. Liang, Y. Ke, Z. Yang, and H. Zhao, “Landslide susceptibility assessment using different slope units based on the evidential belief function model,” Geocarto Int., pp. 1–24, Mar. 2019.

[9] C.-J. F. Chung, A. G. Fabbi, and C. J. Van Westen, “Multivariate regression analysis for landslide hazard zonation,” in Geographical Information Systems in Assessing Natural Hazards. Dordrecht, The Netherlands: Springer, 1995, pp. 107–133.

[10] T. Kavzoglu, I. Colkkesen, and E. K. Sahin, “Machine learning techniques in landslide susceptibility mapping: A review and a case study,” in Landslides: Theory, Practice and Modelling, Cham, Switzerland: Springer, 2019, pp. 283–301.

[11] Q. He, H. Shahabi, A. Shirzadi, S. Li, W. Chen, N. Wang, H. Chai, H. Bian, J. Ma, Y. Chen, X. Wang, K. Chapi, and B. B. Ahmad, “Landslide spatial modelling using novel bivariate statistical based Naïve Bayes, RBF Classifier, and RBF Network machine learning algorithms,” Sci. Total Environ., vol. 663, pp. 1–15, May 2019.

[12] C. N. Madawala, B. T. G. S. Kumara, and L. Indrathilaka, “Novel machine learning ensemble approach for landslide prediction,” in Proc. Int. Res. Conf. Smart Comput. Syst. Eng. (ICSCSE), Mar. 2019, pp. 78–84.

[13] H.-D. Nguyen, V.-D. Pham, Q.-H. Nguyen, V.-M. Pham, M. H. Pham, V. M. Vu, and T.-Q. Bui, “An optimal search for neural network parameters using the Salp swarm optimization algorithm: A landslide application,” Remote Sens. Lett., vol. 11, no. 4, pp. 353–362, Apr. 2020.

[14] 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,” Bull. Eng. Geol. Environ., vol. 78, no. 4, pp. 2865–2886, Jun. 2019.

[15] H. Hong, H. Shahabi, A. Shirzadi, W. Chen, K. Chapi, B. B. Ahmad, M. S. Roodposhti, A. Y. Hesar, Y. Tian, and D. T. Bui, “Landslide susceptibility assessment at the Wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods,” Natural Hazards, vol. 96, no. 1, pp. 173–212, Mar. 2019.

[16] F. Yan, Q. Zhang, S. Ye, and B. Ren, “A novel hybrid approach for landslide susceptibility mapping integrating analytical hierarchy process and normalized frequency ratio methods with the cloud model,” Geomorphology, vol. 327, pp. 170–187, Feb. 2019.

[17] A. Shirzadi, D. T. Bui, B. T. Pham, K. Solaimani, K. Chapi, and A. Kavian, “Shallow landslide susceptibility assessment using a novel hybrid intelligence approach,” Environ. Earth Sci., vol. 76, p. 60, Jun. 2019.

[18] Q.-T. Bui, Q.-H. Nguyen, X. L. Nguyen, V. D. Pham, H. D. Nguyen, and V.-M. Pham, “Verification of novel integrations of swarm intelligence algorithms into deep learning neural network for flood susceptibility mapping,” J. Hydrod., vol. 581, Feb. 2020, Art. no. 124379.

[19] H. Yu, Y. Ma, L. Wang, Y. Zhai, and X. Wang, “A landslide intelligent detection method based on CNN and RSG_R,” in Proc. IEEE Int. Conf. Mechatronics Automat. (ICMA), Aug. 2017, pp. 40–44.

[20] S. L. Ullo, M. S. Langenkamp, T. P. Okarinen, M. P. Del Rosso, A. Sebastianelli, F. Piccirillo, and S. Sica, “Landslide geohazard assessment with convolutional neural networks using sentinel-2 imagery data,” 2019, arXiv:1906.06151. [Online]. Available: http://arxiv.org/abs/1906.06151.

[21] O. Ghorbanazadeh, T. Blaschke, K. Gholamnia, S. Meena, D. Tiede, and I. Aryal, “Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection,” Remote Sens., vol. 11, no. 2, p. 196, Jan. 2019.

[22] Q.-T. Bui, M. P. Van, N. T. T. Hang, Q.-H. Nguyen, N. X. Linh, P. M. Hai, T. A. Tuan, and P. Van Cu, “Hybrid model to optimize object-based land cover classification by meta-heuristic algorithm: An example for supporting urban management in Ha Noi, Viet Nam,” Int. J. Digit. Earth, vol. 12, no. 10, pp. 1118–1132, Oct. 2019.
[23] Q.-T. Bui, M. V. Pham, Q.-H. Nguyen, L. X. Nguyen, and H. M. Pham, “Whale optimization algorithm and adaptive neuro-fuzzy inference system: A hybrid method for feature selection and land pattern classification,” Int. J. Remote Sens., vol. 40, no. 13, pp. 5078–5093, Jul. 2019.

[24] Q.-T. Bui, Q.-H. Nguyen, V. M. Pham, V. D. Pham, M. H. Tran, T. T. H. Tran, H. D. Nguyen, X. L. Nguyen, and H. M. Pham, “A novel method for multispectral image classification by using social spder optimization algorithm integrated to fuzzy C-mean clustering,” Can. J. Remote Sens., vol. 45, no. 1, pp. 42–53, Jan. 2019.

[25] Q.-T. Bui, “Metaheuristic algorithms in optimizing neural network: A comparative study for forest fire susceptibility mapping in Dak Nong, Vietnam.” Geomatics, Natural Hazards Risk, vol. 10, no. 1, pp. 136–150, Jan. 2019.

[26] D. T. Bui, H. Shahabi, A. Shirzadi, K. Chapi, N.-D. Hoang, B. Pham, Q.-T. Bui, C.-T. Tran, M. Panahi, B. B. Ahmad, and L. Saro, “A novel integrated approach of relevance vector machine optimized by imperialist competitive algorithm for spatial modeling of shallow landslides,” Remote Sens., vol. 10, no. 10, p. 1538, Sep. 2018.

[27] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, Apr. 1997.

[28] V. Nguyen, B. Pham, B. Vu, I. Prakash, S. Jha, H. Shahabi, A. Shirzadi, D. Ba, R. Kumar, J. Chatterjee, and D. T. Bui, “Hybrid machine learning approaches for landslide susceptibility modeling,” Forests, vol. 10, no. 2, p. 157, Feb. 2019.

[29] S. Park and J. Kim, “Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance,” Appl. Sci., vol. 9, no. 5, p. 942, Mar. 2019.

[30] A. Shirzadi, K. Solaimani, M. H. Roshan, A. Kavian, K. Chapi, H. Shahabi, S. Keesstra, B. B. Ahmad, and D. T. Bui, “Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution,” Catena, vol. 178, pp. 172–188, Jul. 2019.

[31] Y. Wang, Z. Fang, and H. Hong, “Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China,” Sci. Total Environ., vol. 666, pp. 975–993, May 2019.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017.

[33] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[34] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Med. Image Comput. Comput. Assisted Intervent. (MICCAI), vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[35] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 5/2015 2015.

[36] S. Mirjalili, “Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm,” Knowl.-Based Syst., vol. 89, pp. 228–249, Nov. 2015.

[37] M. Ada and B. T. San, “Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey,” Natural Hazards, vol. 90, no. 1, pp. 237–263, Jan. 2018.

[38] B. T. Pham, I. Prakash, and D. T. Bui, “Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees,” Geomorphology, vol. 303, pp. 256–270, Feb. 2018.

VU DONG PHAM received the bachelor’s degree in land management from the VNU University of Science, Hanoi, Vietnam, in 2018. He is currently pursuing the master’s degree in remote sensing and GIS, VNU University of Science. His research interests include big data analysis, machine learning, and GIS applications.

QUOC-HUY NGUYEN received the bachelor’s degree in applied mathematics and informatics from the VNU University of Science, in 2011 and the master’s degree in information technology from the VNU University of Engineering and Technology, Hanoi, Vietnam, in 2016. He is currently pursuing the Ph.D. degree in applied GIS and remote sensing for disaster early warning system with the GIS Research Center of Civil and Hydraulic Engineering-International Program of GeoInformatics, Feng Chia University, Taichung, Taiwan. He is also a member of the Center for Applied Research in Remote Sensing and GIS (CARGIS), Faculty of Geography, VNU University of Science. His research interests include machine learning, wireless sensor networks, and big data analysis.

HUU-DUY NGUYEN received the degree in geography from the VNU University of Science, in 2011, and the master’s degree in geography and the Ph.D. degree from the University of Orléans, France, in 2014 and 2018, respectively. He is currently involved in floods, landuse/landcover change, urban morphology, machine learning for sustainable use of natural resources and mitigation to impacts of climate change-induced disasters, and food security.

VAN-MANH PHAM received the bachelor’s and master’s degrees in cartography, remote sensing, and GIS from the VNU University of Science, Hanoi, Vietnam, in 2007 and 2013, respectively. He is currently pursuing the Ph.D. degree in cartography, remote sensing, and GIS with the VNU University of Science. His research interests include remote sensing, spatial analysis, satellite image analysis, land-cover/land-use, and geoinformatics.

VAN MANH VU received the Ph.D. degree in geography from Vietnam National University, in 2004. He has been an Associate Professor of earth science with the VNU University of Science, Hanoi, Vietnam, since 2012. He is also with the Department of Environmental Management, Faculty of Environmental Sciences, VNU University of Science. His research interests include environmental monitoring and modeling, geo-informatics, and environmental management.

QUANG-TANH BUI received the Ph.D. degree in GIS and geography from Greifswald University, Germany, in 2007. He is currently working with the Center for Applied Research in Remote Sensing and GIS (CARGIS), Faculty of Geography, VNU University of Science, Hanoi, Vietnam. His research interests include geospatial technology, big spatial data, and machine learning.

***