Query Details

Back to Main Page

1. Please note that the equations and citations are renumbered to ensure sequential order . Please check and confirm the change.

Research

Integrating Harris Hawks optimization and TensorFlow deep learning for flash flood susceptibility mapping using geospatial data

Le Duc Tinh
Email : leductinh@humg.edu.vn
Affiliationids : Aff1

Do Thi Phuong Thao
Email : dothiphuongthao@humg.edu.vn
Affiliationids : Aff1

Dieu Tien Bui
Email : Dieu.T.Bui@usn.no
Affiliationids : Aff2

Nguyen Gia Trong✉
Email : nguyengiatrong@humg.edu.vn
Affiliationids : Aff1 Aff3, Correspondingaffiliationid : Aff3

Aff1 Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, No. 18 Pho Vien, Duc Thang, Bac Tu Liem, Hanoi, 10000, Vietnam

Aff2 GIS Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, Bø i Telemark, 3800, Norway

Aff3 Geodesy and Environment Research Group, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, 10000, Vietnam

Received: 12 February 2024 / Accepted: 30 May 2024

Abstract

Flash floods are recognized as some of the most devastating natural disasters globally, causing significant damage to socio-economic infrastructures, ecosystems, and human lives, thus highlighting the critical need for accurately identifying areas at risk. In order to address this challenge, our study introduces a novel approach by integrating Harris Hawks Optimization (HHO) with the TensorFlow Deep Neural Network (TFDNN), termed HHO-TFDNN, for assessing flash flood susceptibility. The innovation of HHO-TFDNN resides in its dual structure: TFDNN is employed to develop flash flood prediction models, while HHO is utilized to optimize their parameters. This methodology was applied to a region in northern Vietnam, frequently impacted by flash floods. A detailed flash flood database was assembled using various geospatial data sources for the model's training and validation. The results underscore the model's exceptional predictive accuracy, demonstrated by a high F-score of 0.913, a Kappa statistic of 0.825, and an overall accuracy of 91.2%. These findings establish HHO-TFDNN as a highly effective tool for predictive modeling in flash flood management.

Keywords

TensorFlow
Deep neural network
Harris Hawks optimization
Flash flood
Geospatial modeling

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12145-024-01351-1.

Communicated by Hassan Babaie.

Introduction

Flash floods remain among the most catastrophic natural phenomena globally (Diakakis et al. 2020), resulting in significant fatalities in numerous countries, especially those along the coast of the Pacific Ocean (Hu et al. 2018). A defining characteristic of flash floods is their rapid onset, often occurring within a brief timeframe of less than six hours (Gourley et al. 2012) following intense rainfall, especially when associated with troughs or tropical storms. Consequently, the swift progression of flash floods severely restricts the opportunity for rapid and effective decision-making, leading to the highest average fatality rate compared to other flood types (Laudan et al. 2020). It is projected that the likelihood of extreme rainfall events will significantly escalate in the future due to global warming (Myhre et al. 2019); thus, the number of individuals affected by flash floods is expected to increase continuously. Therefore, allocating resources towards research and innovation focused on improving flash flood prediction is crucial in developing robust flood management strategies.

In Vietnam, natural disasters' intensity, frequency, and unpredictability have escalated markedly (MONRE 2017). Over the past two decades, these calamities have resulted in approximately 13,000 fatalities and inflicted property damage exceeding $6.4 billion (Baca and Nguyen 2017), an annual economic toll estimated to approximate 1.5% of the country's GDP (MONRE 2017). Among various natural disasters, flash flooding stands out as a particularly severe phenomenon in the small, steep basins of northern mountainous areas as well as central Vietnam. This type of disaster tends to recur and intensify during the peak of the rainy season, typically from June to August annually. A notable example includes the occurrence of heavy rains, which precipitated landslides and widespread flash floods in the northern mountainous regions of Vietnam in early August 2017. A significant flash flood event in the Nam Pam commune, Son La province, on August 2nd and 3rd, 2017, resulted in 15 fatalities, 15 injuries, and the destruction of 279 homes. Concurrently, a separate flash flood in the Mu Cang Chai district, the focus of this study, claimed the lives of 14 individuals and led to the destruction of 156 houses. Hence, identifying and predicting flash flood-prone areas is critical in aiding governmental bodies with the prevention, mitigation, and adaptation strategies for natural disasters and safeguarding livelihoods (Tien Bui et al. 2020).

Reviews of the literature indicate that research on flash flood modeling and prediction has a variety of methodological approaches. Among these, hydrological approaches that utilize rainfall-runoff modeling are adept at delivering both temporal and spatial probabilities with a high degree of accuracy, i.e., TOPMODEL (Beven and Kirkby 1979), HiResFlood-UCI (Nguyen et al. 2016), HEC-HMS (Zema et al. 2017), HEC-RAS (Munna et al. 2021). The main feature of these models is that flash flood propagation is established based on the rainfall and runoff relationship through simulations using a series of complex mathematical equations. Utilizing Geographic Information Systems (GIS), various maps, such as those representing soil type, land use/land cover (LULC), and terrain, can be integrated to quantify surface runoff. Combined with DEM, areas affected by flash floods, according to timelines, will be estimated and identified on the map. Typically, rainfall-runoff models necessitate extensive time series data from monitoring stations. Thus, the scarcity of such stations, i.e., in developing countries, may pose a significant challenge to developing precise hydrological models (Tien Bui and Hoang 2017).

The evolution of GIS technology, combined with the availability of multi-sourced geospatial data and advancements in machine learning and deep learning, has led to the consideration of new "on-off" modeling approaches (Tien Bui et al. 2016), which do not necessitate the use of long-period time series data, including, random forests (Wang et al. 2015), Gaussian Mixture (Tien Bui and Hoang 2017), Decision Trees (Khosravi et al. 2018), Multilayer Neural Networks (Ngo et al. 2018a), support vector machines (Choubin et al. 2019), Regression Splines (Tien Bui et al. 2019), Ensembles (Band et al. 2020), Deep neural networks (Tien Bui et al. 2020), Deep Learning (Costache et al. 2020a), convolutional neural networks (Panahi et al. 2021), deep belief networks (Shahabi et al. 2021), boosted regression trees (Abedi et al. 2022), Fuzzy multicriteria decision making(Costache et al. 2022), hybrid machine ensembles (Habibi et al. 2023), blending machine learning (Yin et al. 2023). Fundamentally, the "on-off" modeling approaches require a comprehensive collection of input map data as well as maps that detail the locations of past and present flash floods. Research findings suggest that these models excel at identifying significant correlations between flash flood data and input maps, thus facilitating the identification of areas susceptible to future flash floods.

Among various models, deep learning has proven to be a powerful method for mapping susceptibility to flash floods. This effectiveness stems from its capacity to identify complex flood patterns and focus on local contextual details (Trong et al. 2023). However, research on the application and optimization of deep learning models for flash flood modeling and prediction is currently limited. Consequently, conducting new studies across diverse geo-environmental conditions is crucial to expanding the existing knowledge base and deriving more reliable conclusions. This study aims to partially address the above gap in the literature by presenting and validating a novel integrated deep-learning model named HHO-TFDNN for flash flood susceptibility mapping. In the proposed model, the TensorFlow-based deep learning (TFDNN) technique is used to generate an initial flash flood model, whereas the Harris Hawks Optimization algorithm (HHO) (Heidari et al. 2019) is used to optimize the model. To the best of our knowledge, the use of the Harris Hawks Optimizer (HHO) for optimizing TensorFlow Deep Neural Networks (TFDNN) in flash flood modeling has seldom been considered. The focus of this study is the mountainous region of Yen Bai province in northwest Vietnam, an area recognized for its vulnerability to flash floods induced by heavy rainfall during tropical storms. This setting offers a relevant backdrop for implementing and assessing the efficacy of the proposed predictive model.

Background of the utilized algorithms

TensorFlow deep neural networks

Deep learning represents an advanced segment of machine learning, characterized by its use of sophisticated architectures composed of multiple layers of processing (LeCun et al. 2015). These multilayered structures are adept at mining complex patterns and representing data nuancedly, enabling the algorithm to learn from vast amounts of information with minimal human intervention. This approach allows

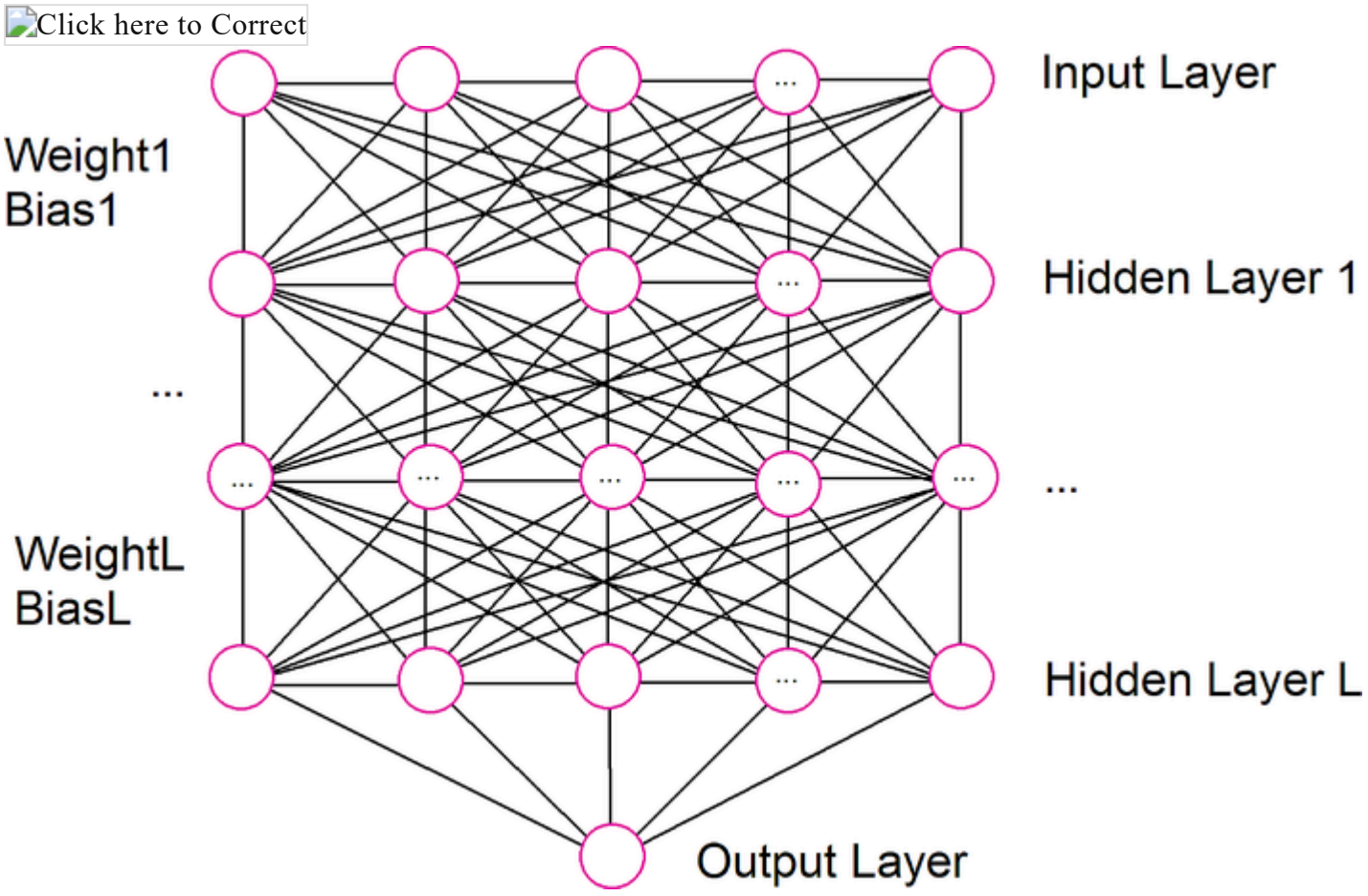
for the extraction of high-level features from raw data, making deep learning exceptionally effective for a wide range of applications, including flash flood modeling and prediction (Bui et al. 2020).

In the last decade, deep learning with various frameworks has been developed, including TensorFlow from Google (Abadi et al. 2015), PyTorch from Facebook (Paszke et al. 2019), and CNTK from Microsoft (Seide and Agarwal 2016), made a significant impact in various domains. This study utilized the TensorFlow deep learning framework due to its capability to integrate with ArcGIS Pro software through Python programming. This integration facilitates the handling of multi-sourced geospatial data and allows for the compilation of modeling results into flash flood susceptibility maps.

Figure 1 illustrates a typical architecture of a deep neural network comprising an input layer, an output layer, and several hidden layers. In the context of a flash flood dataset (FFDS), represented as FFDS = (X, y), X denotes a vector consisting of 10 factors related to flash floods, while y represents the flash flood index, with values ranging from 0 to 1.

Fig. 1

A typical architecture of a deep neural network



Training a deep neural network model involves fine-tuning the network's parameters, including weights and biases, to minimize the prediction error. For this particular analysis, the optimization of the network's parameters was achieved using Harris Hawks optimization (HHO), which is explained in the section below.

Harris Hawks optimization (HHO)

As explained above, this study employed the Harris Hawks Optimizer (HHO) introduced by Heidari et al. (2019) to fine-tune the weights and biases of the TensorFlow Deep Neural Network (TFDNN) model dedicated to mapping flash flood susceptibility. The HHO represents a probabilistic, gradient-free global optimization algorithm that mimics the collaborative hunting tactics and pursuit patterns of Harris' Hawks in nature. It mathematically models the cunning approach a congregation of Harris Hawks employs as they converge on a target from various angles. The HHO was chosen for this study because of its ability to achieve quick convergence and deliver improved accuracy across different fields (Alabool et al. 2021). In the context of this research, the HHO algorithm was employed to fine-tune the weights and biases of the TFDNN model through the following four steps (Nhu et al. 2023):

Step 1: The first step of this metaheuristic approach involves creating an initial population, N, within a feasible search space delineated by a lower boundary (LB) and an upper boundary (UB). Within this search domain, the location of each Harris Hawk is identified by its coordinates, which correspond to the weights and biases of the TFDNN model, making each hawk a potential solution for the model. In order to determine the best position of Harris Hawk for the model, a cost function was devised to evaluate the efficacy of each solution.

Step 2: The fitness is calculated for each hawk in the initial population to identify the most effective Harris Hawk solution among them. Following this, the exploration phase commences, during which the hawks' positions are adjusted according to Eq. 1 (Heidari et al. 2019; Nhu et al. 2023), and then, the best position of Harris Hawk with respect to the TFDNN model is determined.

$$\mathrm{PO}^{\mathrm{t}+1}=\begin{cases}\mathrm{PO}_{Rh}^{t}-\mathrm{r}_1\left|\mathrm{PO}_{R}^{t}-2\mathrm{r}_2\mathrm{PO}^{\mathrm{t}}\right|; & \mathrm{q}\geq 0.5\\ \left|\mathrm{PO}_{\mathrm{B}}^{t}-\mathrm{PO}_{m}^{t}\right|-\mathrm{r}_3\left(\mathrm{LB}+\mathrm{r}_4\left(\mathrm{UB}-\mathrm{LB}\right)\right); & \mathrm{q}<0.5\end{cases} \quad (1)$$

where $\mathrm{PO}^{\mathrm{t}+1}$ and PO^{t} signify the positions of the hawks at iterations t+1and t, respectively. $\mathrm{PO}_{\mathrm{B}}^{t}$ presents the position of the individual exhibiting the highest fitness level, also known as the prey's position at iteration t. PO_{R}^{t} is identified as a randomly selected hawk, PO_{Mean}^{t} denotes the mean position of the individuals, which is determined using Eq. 2. The parameters $\mathrm{r}_1,\mathrm{r}_2,\mathrm{r}_3,\mathrm{r}_4$, and q are variables generated randomly within the range of 0 to 1.

$$\mathrm{PO}_{Mean}^{t} = \frac{1}{\mathrm{N}}\mathrm{PO}^{\mathrm{t}} \quad (2)$$

Step 3. Based on the escape energy (ES), which is computed using Eq. 3,

$$\mathrm{ES} = 2\mathrm{ES}_0\left(1 - \mathrm{t}/\mathrm{It}_{max}\right) \quad (3)$$

where ES denotes the escaping energy of a prey, ES_0 represents an initial value of the energy and It_{max} is the maximum number of iterations used.

The position of each Harris Hawk will be updated based on the prey's escaping energy as follows: When |E| is greater than or equal to 1, the position update for each hawk is carried out according to Eq. 1. Conversely, if |E|<1, the positions of the Harris Hawks are adjusted using Eqs. 4, 5, and 6.

$$\mathrm{PO}^{\mathrm{t}+1} = \begin{cases} \Delta\mathrm{PO}^{\mathrm{t}} - \mathrm{ES}\left|\mathrm{JPO}_B^t - \mathrm{PO}^{\mathrm{t}}\right|; 0.5 \le |\mathrm{ES}| < 1 \,\&\, r \ge 0.5 \\ \mathrm{PO}_B^t - \mathrm{ES}\left|\Delta\mathrm{PO}^{\mathrm{t}}\right|; \qquad |\mathrm{ES}| < 0.5 \,\&\, r \ge 0.5 \end{cases} \quad (4)$$

$$\mathrm{PO}^{\mathrm{t}+1} = \begin{cases} \mathrm{PO}_B^t - \mathrm{ES}\left|\mathrm{JPO}_B^t - \mathrm{PO}^{\mathrm{t}}\right|; 0.5 \le |\mathrm{ES}| < 1 \& r < 0.5; \mathrm{F1} \\ \mathrm{PO}_B^t - \mathrm{ES}\left|\mathrm{JPO}_B^t - \mathrm{PO}^{\mathrm{t}}\right| + \mathrm{S} + \mathrm{Levy}; 0.5 \le |\mathrm{ES}| < 1 \& r < 0.5; \mathrm{F2} \end{cases} \quad (5)$$

$$\mathrm{PO}^{\mathrm{t}+1} = \begin{cases} \mathrm{PO}_B^t - \mathrm{ES}\left|\mathrm{JP}_{Best}^t - \mathrm{P}_m^t\right|; |\mathrm{ES}| < 0.5 \& r < 0.5; \mathrm{F1} \\ \mathrm{PO}_B^t - \mathrm{ES}\left|\mathrm{JPO}_B^t - \mathrm{P}_m^t\right| + \mathrm{S} + \mathrm{Levy}; |\mathrm{ES}| < 0.5 \& r < 0.5; \mathrm{F2} \end{cases} \quad (6)$$

wherein, ES_0 denotes the prey's initial energy, which is allocated a random value between [-1, 1]. T indicates the maximum iteration count; *r* is a number produced randomly within the interval [0,1]; *J* is a randomly determined number within the range [0,2]; F1 and F2 represent specific fitness conditions; S is a stochastic vector; and Levy pertains to the Levy flight function. $\Delta\mathrm{PO}^{\mathrm{t}}$ signifies the discrepancy between the optimal position, PO_B^t, and the present location of the hawk, PO^{t}.

Step 4: The process concludes once a specific termination condition is met, which may be either the completion of the predetermined maximum iteration count or the achievement of the targeted fitness threshold.

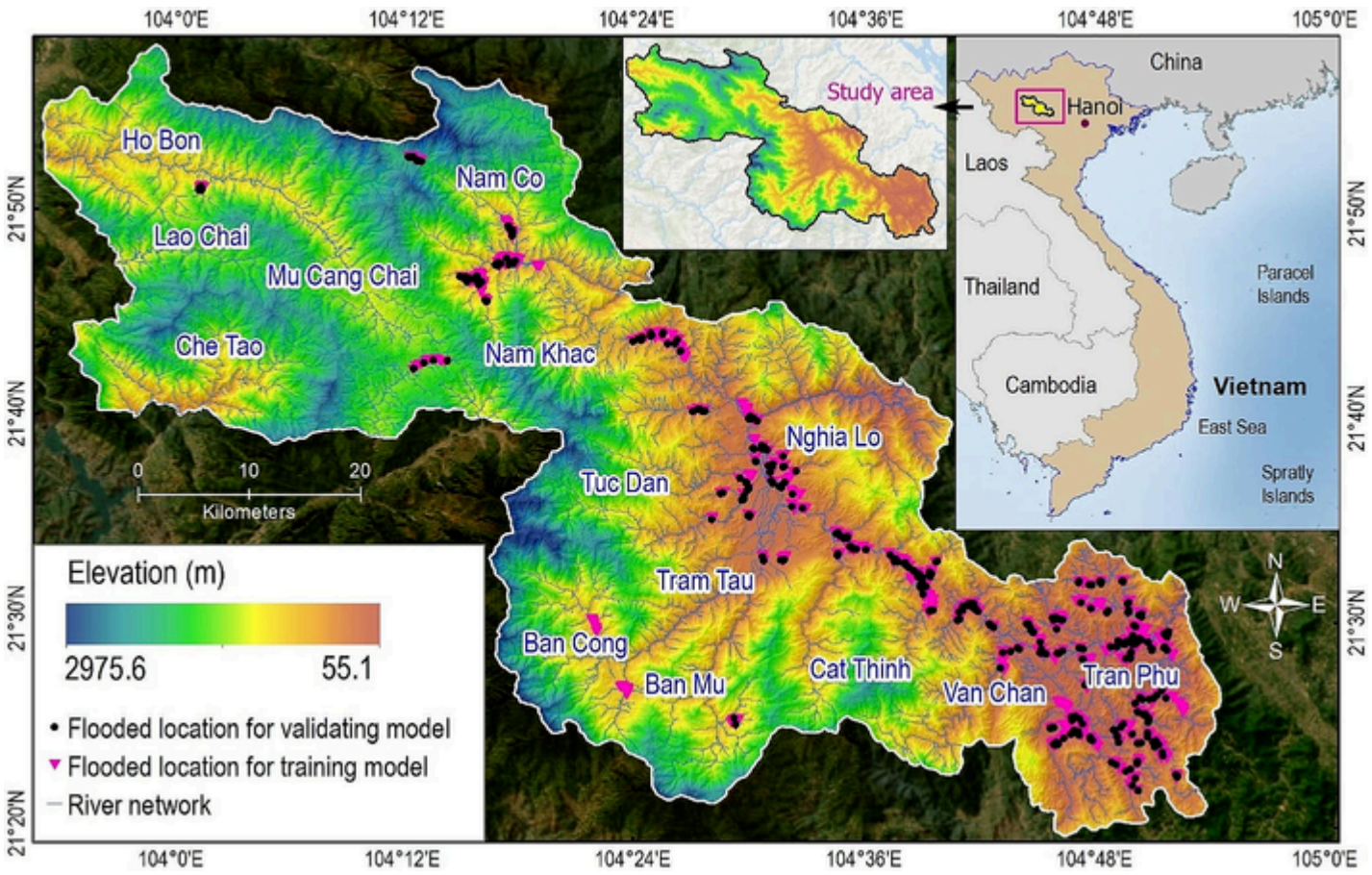
Study area and data

Description of the study area

The study area spans the Nghia Lo town, the Mu Cang Chai, and Van Chan districts, which belong to the Yen Bai province, covering an area of 3173.4 km^2. Geographically, its coordinates stretch from 20º20'to 21º45' North latitude from 104º20' to 104º53' East longitude (Fig. 2). Morphometrically, the study area has typical mountainous characteristics with complex features of the terrain with high forest coverage, mountains, caves, streams, interlaced, and flat valleys. In this area, attitude varies widely from 55.1 m to 2975.6 m with an average of 400 m, and the high slope areas (> 30º) occupy more than 50% of the total area, leading to high probability of flash flood generation. A complex and dense stream network characterizes the hydrological system. Most of the main streams originate from the high mountains (> 2000 m), providing good conditions to generate hydroelectricity and are the main water supply source in the area.

Fig. 2

The studied area (Nghia Lo town, the Mu Cang Chai and Van Chan districts) and flash-flood locations



From a geological perspective, the study area comprises a total of 19 formations and complex outcrops, which are distributed unevenly. Notably, two formations account for more than 73% of the total study area. The Van Chan formation, representing 50.3% of the area, is characterized by its main lithologies, including gravelstone, conglomerate, clay shale, and tuffaceous siltstone. Meanwhile, the Ngoi Thia formation, constituting 23.5% of the area, is distinguished by the presence of rhyolite, porphyric rhyolite, and tuffaceous sandstone as its principal lithologies. Detailed descriptions of the remaining geological units are presented in Table 1.

Table 1

The geological units in the Yen Bai province

No	Geological units	Area (%)	Main lithologies	
1	Van Chan formation	50.3	Gravestone, conglomerate, clay shale, tuffaceous siltstone	
2	Ngoi Thia formation	23.5	Rhyolite, porphyric rhyolite, tuffaceous sandstone	
3	Sinh Quyen formation	4.2	Two-mica plagio-gneiss, quartz two-mica schist, two-mica-garnet schist	
4	Suoi Bang formation	3.1	Marl, siltstone, quinoid siltstone, sandstone, coquina	
5	Cam Duong formation	2.3	Black schist, chlorite-sericite schist, apatite bearing schist, quartzite	
6	Quaternary	2.1	Pebble, granule, clayey sand	
7	Phu Sa Phin complex	1.9	Syenite granosyenite, subalkaline granite	
8	Sa Pa formation	0.1	Marbleized limestone, white-gray dolomitic limestone, schist	
9	Xom Giau complex	0.1	Kalifeldspath-rich biotite granite, pegmatite granite	
10	Others	12.4	sandstone, siltstone, calcareous siltstone, gabbro, diabase, granite, aplitic	

In terms of land use/land cover (LULC), our analysis points out that forests dominate 60.6% of the total study area, followed by grasslands, which cover 19.0%. Bareland constitutes 7.5% of the area, while orchard and crop areas each account for 4.3%. Paddy rice fields represent 3.2% of the study area, and urban and built-up areas comprise approximately 1.0%. In the context of soil distribution within the study area, 87.1% is occupied by three primary soil types: HYR (46.2%), YCMR (27.9%), and HLY (13.1%). Following these, YMR covers 4.8% of the area, LYS accounts for 2.6%, ALU and rocky mountain areas each represent 1.3%, and BOR constitutes 1.2%.

The study area is located in a tropical monsoon region, distinguished by two distinct seasons: a dry season extending from November to April and a rainy season from May to October. The dry season is characterized by reduced rainfall and cooler temperatures, particularly from December to February, when conditions can become notably cold, especially in the highland areas. In these high elevations, such as around Mu Cang Chai, mornings may be foggy with occasional frost, and nighttime temperatures can plummet to between − 2 °C and − 3 °C. Conversely, the rainy season brings substantial rainfall to the province, with the annual average reaching up to 5490 mm (Pham et al. 2018). This rainfall is often intense, particularly in August and September, leading to landslides, floods, and specifically flash floods, posing significant challenges to the region.

Data

Historical flash-flooded location

In the development of prediction models, the historical data of flash-flooded locations and their influencing factors are crucial. For this research, we compiled a flash-flood inventory map using 1,374 flash flood locations (Fig. 2), derived from the state-funded project No. 03/HD-KHCN-NTM 2020, under the auspices of the Ministry of Agriculture and Rural Development of Vietnam (Nguyen et al. 2020).

These polygons, representing flash flood events over the five years from 2015 to 2019, were identified through the analysis of Sentinel-1 SAR imagery employing the Change Detection Technique (Ngo et al. 2018b).

Notably, the Sentinel-1 mission encompasses a leading pair of satellites equipped with C-band Synthetic Aperture Radar (SAR) technology, and thus, these two polar-orbiting satellites are each equipped with a C-band SAR sensor, which can capture images with a spatial resolution of 10 m (Amitrano et al. 2018). Operating in tandem as a constellation ensures a high revisit frequency, providing fresh imagery every six days, thereby offering consistent and reliable data coverage.

Additionally, we utilized information on flash floods reported by local media and imagery from Google Earth to interpret flash flood extents. Our field investigations and subsequent analyses indicate that flash floods in the study area typically occur abruptly following heavy and intense rainfall, with rates of up to 40 mm per hour (Thao and Viet 2023). The steep terrain and the area's recent economic development exacerbate these events. Key factors contributing to the severity of flash floods include the degradation of watershed forests, primarily due to unplanned logging and the practice of slash-and-burn agriculture. These activities lead to rapid and concentrated water runoff during heavy rain events. Moreover, the recent economic development has spurred infrastructure projects, including the construction of bridges and culverts, which often do not have adequate drainage capacity, thereby increasing the risk of flash flooding.

Flash-flood indicators

The construction of a flash flood susceptibility map involves utilizing the statistical relationship between historical flood events and various influencing factors. The careful selection of these factors is essential for accurately predicting future flash flood occurrences. To ensure the relevance and effectiveness of these factors, we based our selection on an analysis of catchment characteristics, informed by a comprehensive review of the literature, including works (Duong Thi et al. 2020; Hapuarachchi et al. 2011; Luu et al. 2023; Ngo et al. 2021b; Nhu et al. 2020a; Tien Bui et al. 2020). For this research, ten influencing factors were identified: rainfall, river density, land cover, elevation, slope, curvature, topographic wetness index (TWI), aspect, soil type, and geology.

Rainfall: Rainfall is a critical factor in shaping the formation and dynamics of water flow within a watershed, significantly influencing the magnitude, velocity, and behavior of flash flood flows, as evidenced by research (Bryndal et al. 2017; Fang et al. 2015; Papagiannaki et al. 2015). Consequently, this study has selected it as a primary variable for analysis. To assess the impact of rainfall, we initially compiled statistics on all flash floods that occurred over the past five years, as reported by the Natural Disaster Prevention Committee of Yen Bai Province. From the analysis of these events, we identified the heaviest total rainfall over three days that led to the most flash floods and used this data to construct a rainfall map. This map was developed using the Inverse Distance Weighting (IDW) interpolation method (Bui et al. 2011). As depicted in Fig. 3 a, rainfall in the study area exhibits considerable variation, ranging from 349.2 mm to 602.4 mm, and is characterized by an uneven spatial distribution.

River density: River density reflects the extent of development within a drainage network and is widely acknowledged as a key factor influencing flow processes within a watershed (Pallard et al. 2009). Given its significance, river density was selected as a parameter for this research. For the purposes of this analysis, river density was calculated based on the river network extracted from topographic maps at a 1:50,000 scale (Do et al. 2020). The computed river density values range from 0 to 10.3, indicating higher river density in mountainous regions and lower river density in the lowland areas along valleys, as illustrated in Fig. 3 b.

Landcover: Land cover (LC) significantly influences the dynamics of flash flood processes due to its profound effects on infiltration and runoff generation, as highlighted in studies by (Jodar-Abellan et al. 2019; Nie et al. 2011). For instance, a decrease in forested areas can lead to increased surface runoff and stream discharge, whereas a higher density of forests can mitigate erosion and reduce discharge into streams and rivers (Guzha et al. 2018). In this study, we utilized the LC map developed by Truong et al. (2019), which was created using ALOS-2/PALSAR-2 imagery, Landsat 8 OLI images, and geospatial data. This map identified eight LC types within the study area, as shown in Fig. 3 c.

Elevation: Elevation is a critical factor in the modeling and prediction of flash floods, as it significantly influences the gravitational force acting on flowing waters (Borga et al. 2014). This influence directly affects both the direction and velocity of water flow across various landscapes. Due to the force of gravity, water invariably moves from regions of higher elevation to those of lower elevation. For this project, we utilized a Digital Elevation Model (DEM) derived from national topographic maps at a 1:10,000 scale (Do et al. 2020) to create an elevation map with a grid cell resolution of 20 m. As illustrated in Fig. 3 d, the elevation within the study area exhibits significant variation, ranging from 55.1 m to 2975.6 m. Areas of high elevation (> 500 m) are primarily found in the Northwest and Southwest regions, whereas the lowland areas (< 100 m) are located in the Southeast, which coincides with zones of frequent flash flood events.

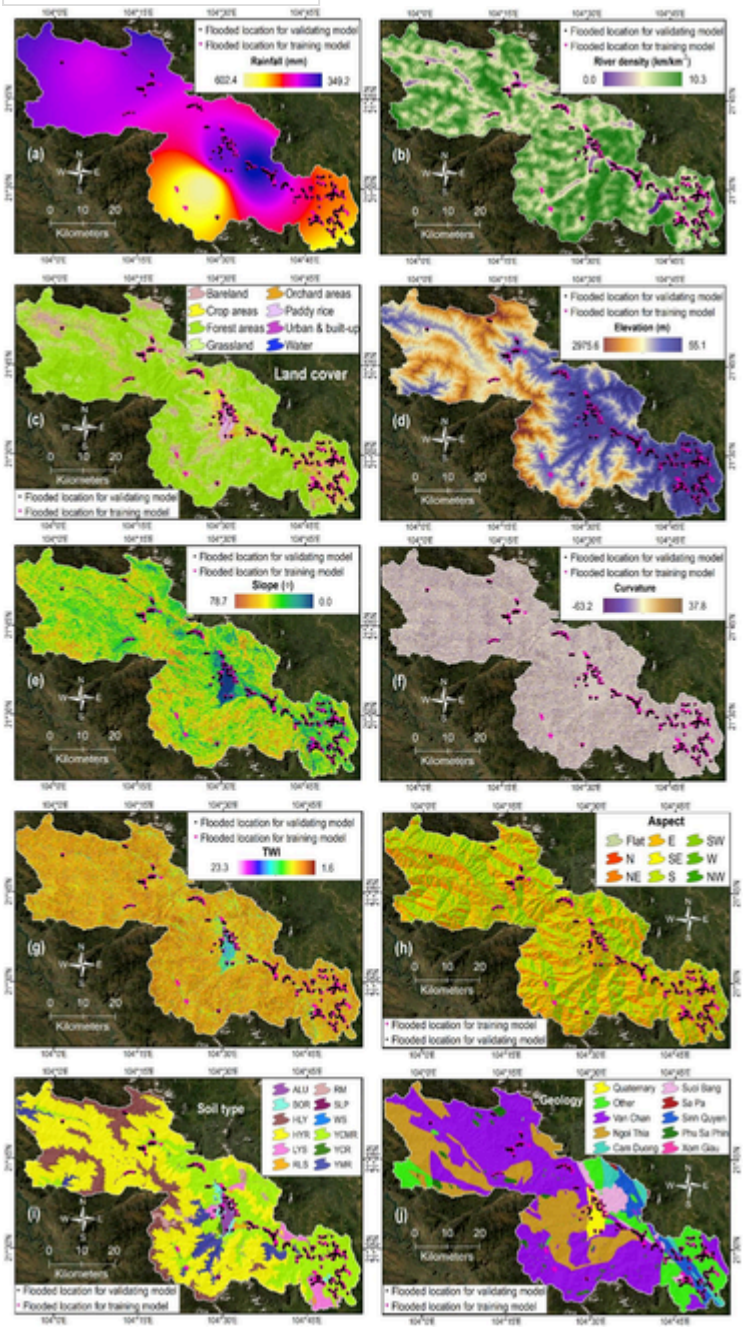
Slope: The slope is widely acknowledged as a crucial factor that can either accelerate or decelerate the flow processes occurring within a watershed (Li et al. 2020). It directly impacts the velocity at which water moves across the surface. In areas with steep slopes, water tends to flow more rapidly, which can lead to quicker runoff and potentially increase flash flooding. Conversely, in areas with gentle slopes, the flow of water is slower, allowing more time for infiltration into the soil, which can reduce runoff and the likelihood of flooding. In this analysis, the slop map for the study area was derived from the DEM. The slope varies from 0.0 to 78.7 degrees, with the mountainous regions exhibiting a significant proportion of steep slopes—more than 35% of these areas have slopes exceeding 50 degrees, as depicted in Fig. 3 e. This variation in slope underscores the diverse topographical characteristics of the region and highlights the presence of extensive steep terrain, particularly in the mountainous areas.

Curvature: Curvature plays a pivotal role in determining the flow paths of water, influencing both convergence/divergence and deceleration/acceleration, which in turn affect infiltration processes and the dynamics of runoff flow (Papanicolaou et al. 2018). Generally, curvatures are categorized into three types: concavity (positive values), convexity (negative values), and flat (zero values). In this study, a curvature map was developed using the Digital Elevation Model (DEM) with a resolution of 20 × 20 m. The curvature within the study area exhibits significant variation, ranging from − 63.2 to 37.8. Analysis of previous flash flood incidents reveals that

they predominantly occurred in areas with concave curvature, indicating characteristics of flash flood divergence and acceleration, as shown in Fig. 3 f.

Fig. 3

Flash-flood influencing factors in this analysis: (**a**) Rainfall, (**b**) River density, (**c**) Landcover, (**d**) Elevation, (**e**) Slope, (**f**) Curvature, (**g**) Topographic wetness index, (**h**) Aspect, (**i**) Soil, and (**j**) Geology. For the soil type, ALU: Alluvial soils; BOR: Brown-yellowish soils on old alluvium; HLY: Humus-light-yellow soil on high mountains; HYR: Humus-yellow-red soils; LYS: Light-yellowish soils on sandstone; RLS: Red soil on limestone; RM: Rocky mountain, SLP: Sloping soils; WS: Water surface; YCMR: Yellowish-red soils on claystone and metamorphic rock; YCR: Yellowish-red soil on cultivated rice; YMR: Yellowish red soil on acid magmatic rocks



Topographic wetness index (TWI): The Topographic Wetness Index (TWI) quantifies the spatial relationship between the formation and variations of runoff flow, reflecting its impact on flash flood dynamics (Costache et al. 2020b). Therefore, it is selected for this analysis. TWI was extracted from the DEM in the studied area, where values range from 1.6 to 23.3, with higher values observed along valleys. These elevated TWI values indicate a tendency for significant water flow accumulation in lowland areas. Consequently, this pattern suggests that flash flood events are more likely to occur along valleys where conditions favor the concentration and rapid movement of runoff, highlighting the importance of TWI as a predictive indicator for flash flood susceptibility in these regions.

Aspect: Aspect, the direction a slope faces, is a critical factor that directly and indirectly influences runoff processes due to its significant impact on soil moisture levels, vegetation development, and the direction of water flow (Hofierka et al. 2009). In the study area, slope aspect is categorized into nine classes, as illustrated in Fig. 3 h. analysis of flash flood records reveals a higher frequency of events in the southwest (SW), West (W), and northwest (NW) directions.

Soil: Soil types should be considered for flash flood modeling due to their significant influence on water infiltration, retention, and runoff processes. The physical properties of soil, including texture, structure, and organic matter content, determine how quickly rainwater can infiltrate the ground versus how much will become surface runoff (Sangati et al. 2009). Soils with high infiltration rates, such as sandy soils, can reduce the risk of flash floods by absorbing more rainfall. In contrast, compacted or clay-rich soils with low infiltration rates lead to increased runoff (Bagarello et al. 2013), heightening flash flood potential. Furthermore, soil moisture content and hydraulic conductivity affect the speed and volume of water flow, making soil characteristics crucial data points for accurately predicting flash flood events and mitigating their impacts. For this research, a soil map was compiled based on the national soil map with a scale of 100,000 (Ngo et al. 2018c). Accordingly, twelve types were identified in this area (Fig. 3 i).

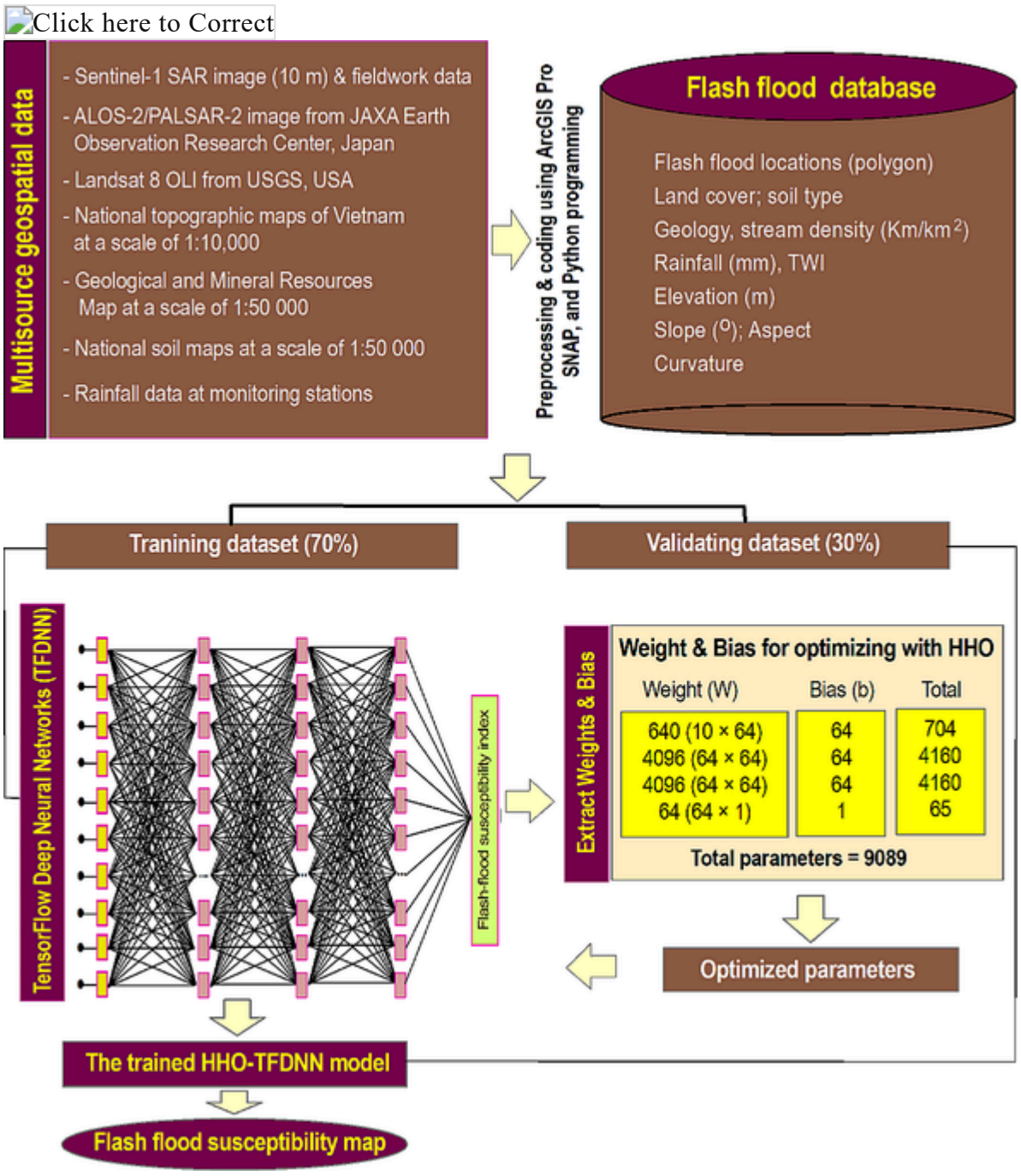
Geology: Geology plays a crucial role in accurately predicting flash flood susceptibility, as it influences the ground's permeability and porosity (Bryant et al. 1993). These factors determine the rate of water infiltration versus surface runoff, directly affecting the probability and severity of flash floods. In this study, the geological map of the study area with ten groups was generated using the national geological and mineral resources map of Vietnam at a scale of 1:50,000.

The proposed HHO-TFDNN framework for flash-flood susceptibility assessment

The description of the proposed HHO-TFDNN employed for flash-flood susceptible assessment is presented in Fig. 4. In this research, the SNAP toolbox and ArcGIS Pro 3.0 were utilized for processing the Sentinel-1 SAR imagery and flash flood indicators, respectively. The HHO algorithm codes can be found in Heidari et al. (2019); whereas the authors wrote the HHO-TFDNN with the help of built-in functions provided by the Keras Deep Learning API of TensorFlow 2.0 (https://www.tensorflow.org) in the Python environment.

Fig. 4

The flowchart of the proposed HHO-TFDNN for flash flood susceptible mapping in this research



Building the flash-flood database

We employed the ESRI-geodatabase format (Zeiler 1999) to build the flash-flood database for this project. The ESRI geodatabase is capable of organizing geospatial data from different sources. As a result, the flash-flood database contains 1374 flash-flood polygons and ten flash-flood indicators mentioned in Sect. 3.2.2. All indicators were resampled to a raster format with a resolution of 20 × 20 m using the Nearest Neighbor resampling technique in ArcGIS Pro 2.6. Because the HHO-TFDNN could process values in the rank from 0 to 1, we normalized the ten indicators into the rank 0.01 and 0.99 (Ngo et al. 2021a). In the subsequent phase, we randomly generated 1,374 sample points in non-flood areas. The sampling tool in ArcGIS Pro 2.6 was then employed to extract values for all samples. Ultimately, the data was divided into a 70/30 ratio (Ngo et al. 2021b; Rana and Mahanta 2023) for the HHO-TFDNN model's training and validating, respectively.

The HHO-TFDNN structure and its trainable parameters

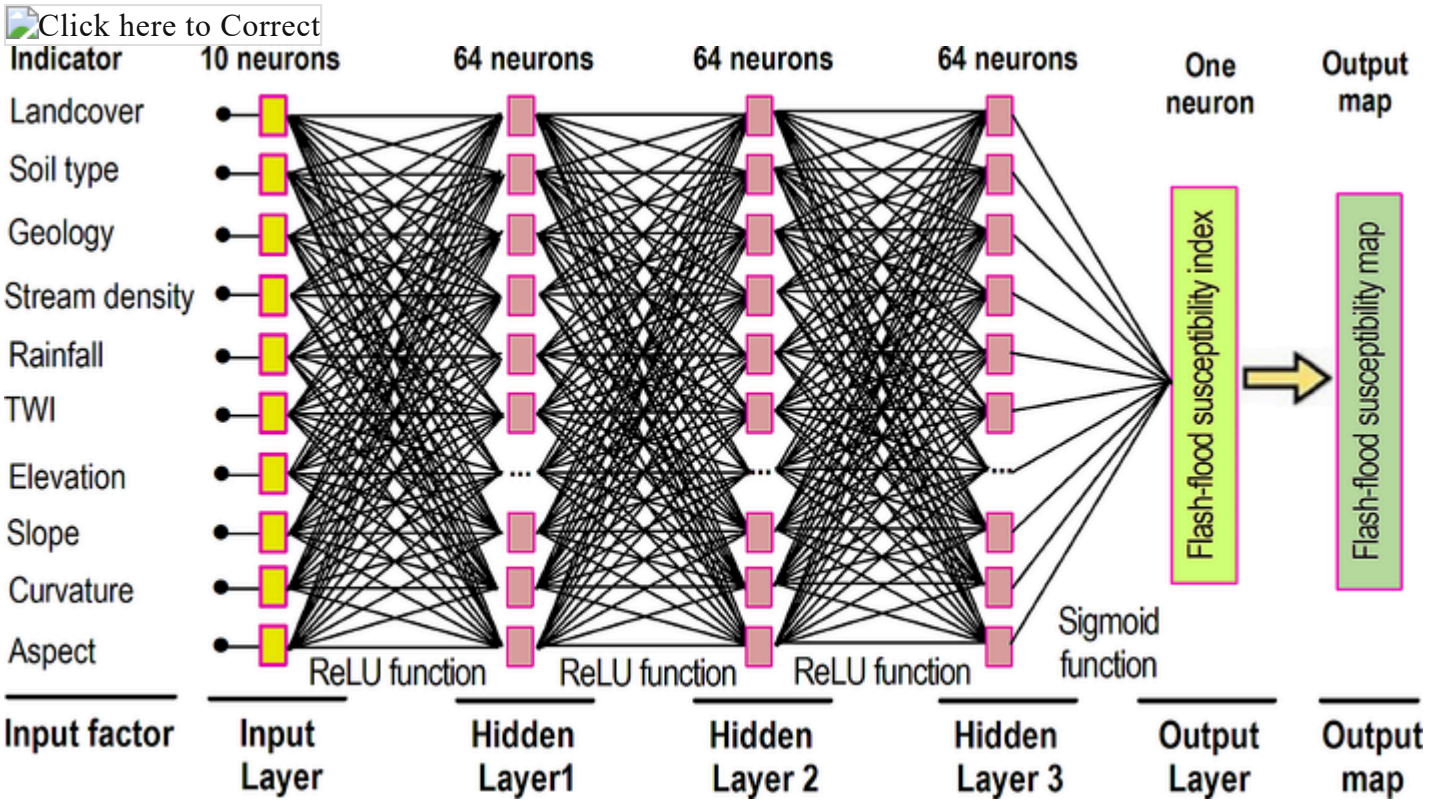
At the core of the proposed HHO-TFDNN model, a deep neural network (DNN) could be employed as a pattern recognizer to infer the input indicators to flash-flood or non-flood classes. However, we formulate the flash flood modeling as a regression problem in this research. The structure of the employed TFDNN in this project is shown in Fig. 5. In the first layer (also called the input layer), information regarding the ten indicators is received. This information is then analyzed in the hidden layers to reveal meaningful patterns hidden in the collected data. Useful knowledge can be extracted to generate regression values in the range of [0,1]. The final layer is used to produce numerical values. A threshold of 0.5 was adopted to separate output indices into two classes for the model performance assessment.

DNN models are highly used for flash-flood susceptibility assessment because the problem of interest entails the analysis of nonlinear functional mapping and multivariate data. Therefore, the employed machine learning approach must be capable of mining complex patterns from such nonlinear and multivariate datasets. Since a DNN is equipped with multiple hidden layers, an enriched hyper-plane set

can better separate the learning space. Therefore, a DNN possesses a high chance of capturing and inferring the sophisticated concept used for pattern classification.

Fig. 5

Structure of the proposed HHO-TFDNN model



Cost function

During the training phase of a DNN model, the weights of the HHO-TFDNN model must be adapted to identify the most appropriate functional mapping between the actual and the predicted class labels of flood and non-flood. To adjust such connecting weights, we employed Mean Absolute Error (MAE) as the cost function (Ngo et al. 2021b) to measure the performance of the flash flood models as below:

$$\mathbf{MAE} = \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} |\mathbf{TFF_i} - \mathbf{OFF_i}| \quad (7)$$

where $\mathbf{TFF_i}$ denotes the target values; $\mathbf{OFF_i}$ is the predicted flash-flood values yielded by the HHO-TFDNN model, and n denotes the size of the training dataset. **AQ1**

Prediction performance evaluation

HHO-TFDNN model has been trained to fit the records stored in the training dataset, and a set of performance measurement metrics was used to quantify the established model's quality. In this project, we used TP- true positive, FP- false positive, FN- false negative, TN- true negative as described in (Nhu et al. 2020b). Based on the result of these four basic indices, PPV- Positive predictive value, NPV- Negative predictive Value, Sens-Sensitivity, Spec- Specificity, Accuracy, F-Score were further computed (Hoang 2020; López et al. 2013). In addition to the aforementioned metrics, the ROC curve and AUC (van Erkel and Pattynama 1998) are also widely employed to assess the overall generalization capability of the HHO-TFDNN model. Moreover, the Kappa index (McHugh 2012) can also be computed to quantify a binary classification model's predictive accuracy. Therefore, this index is used in this study for model assessment and result comparison.

Results and analysis

Model fitting and validation

In this study, the HHO metaheuristic is integrated to optimize the HHO-TFDNN's weights for flash-flood susceptibility assessment. This is a challenging task since the number of optimized parameters is large, with 9089 trainable weights (as demonstrated in Table 2). Via the exploration and exploitation processes of a population of hawks, the HHO algorithm gradually identifies better candidate solutions and discards inferior ones.

Table 2

Summary of the HHO-TFDNN model with 9089 trainable weight parameters

No.	Layer	Activate function	Neurons	Parameters		
				Weight (w)	Bias (b)	Total
1	Dense 1	ReLU	64	640 (10 × 64)	64	704
2	Dense 2	ReLU	64	4096 (64 × 64)	64	4160
3	Dense 3	ReLU	64	4096 (64 × 64)	64	4160

No.	Layer	Activate function	Neurons	Parameters		
				Weight (w)	Bias (b)	Total
4	Dense 4	Sigmoid	1	64 (64 × 1)	1	65

When the optimization process terminates, a set of appropriate connecting weights, representing a model with a high degree of fitness, were further used to generate the flash-flood susceptibility model. The training and validating performance of the DNN model optimized by the HHO metaheuristic is presented in Table 3.

Table 3

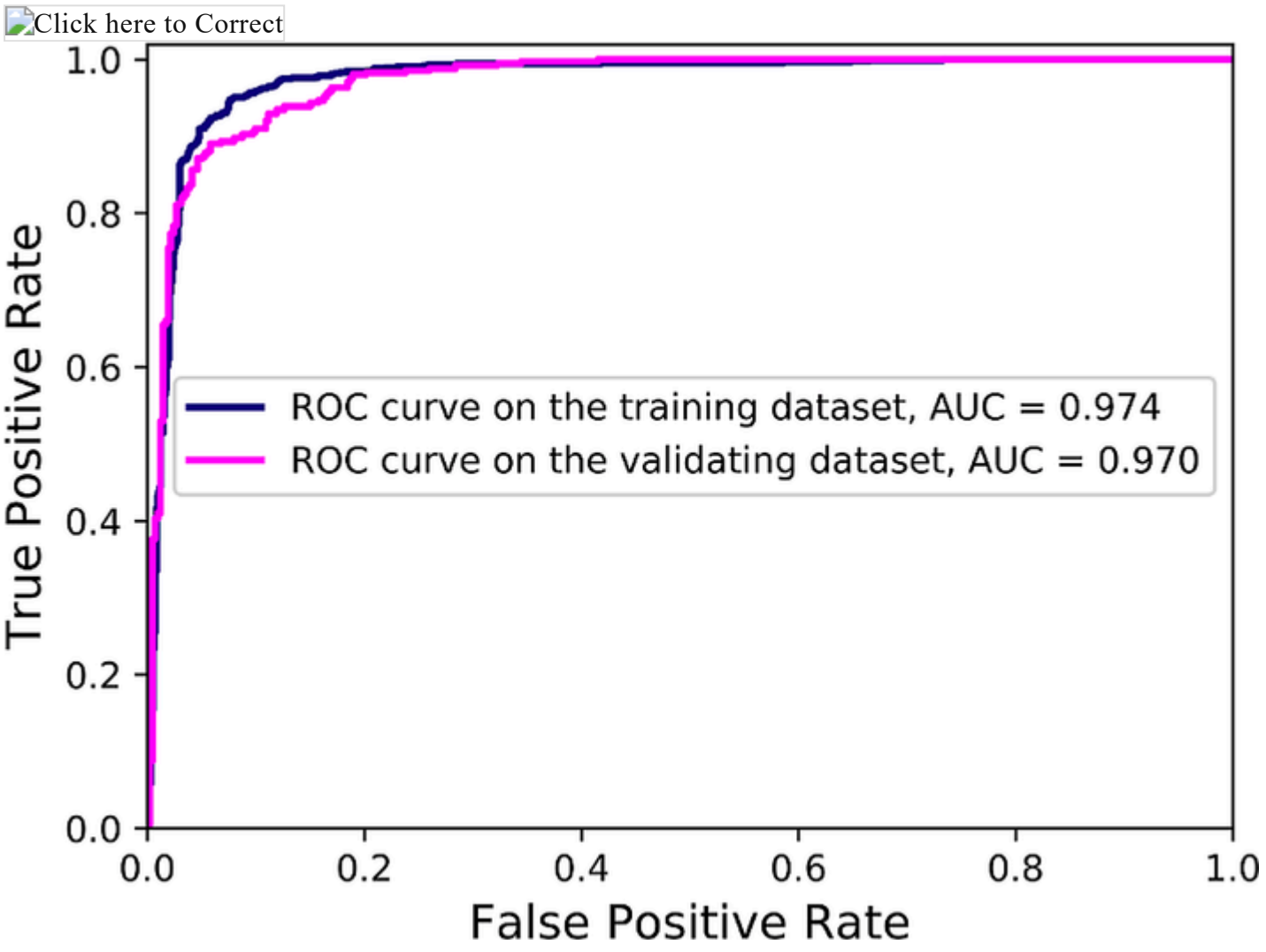
Statistical performance of the HHO-TFDNN model

Modeling phase	Statistical performance of the HHO-TFDNNs model									
	TP	TN	FP	FN	Sens(%)	Spec(%)	Accuracy	F-score	Kappa	AUC
Training	898	890	64	72	92.6	93.3	92.9	0.930	0.859	0.974
Validating	376	373	36	36	91.3	91.2	91.2	0.913	0.825	0.970

We observe from Table 3 that the HHO-TFDNN model has achieved good performances in both training (Sen = 92.6%, Spec = 93.3%, accuracy = 92.9%, F-score = 0.930, Kappa = 0.858, and AUC = 0.974) and validating phase (Sen = 91.3%, Spec = 91.2%, accuracy = 91.2%, F-score = 0.913, Kappa = 0.825, and AUC = 0.970). In addition, the proposed model has also yielded relatively balanced outcomes of true positive and true negative prediction in both training (TP = 898 and TN = 890) and validating phases (TP = 376 and TN = 373). The accuracy index above 90% indicates that the integrated DNN and HHO is highly suitable for flash-flood susceptibility assessment in the selected study area. Furthermore, the ROC curves associated with AUC values are demonstrated in Fig. 6.

Fig. 6

The obtained ROC curve and AUC values of the HHO-TFDNN model



Flash flood factor ranking

After training and validation, results are obtained, and it is useful to examine the impact of the collected influencing factors on the flash-flood susceptibility prediction process. This study uses the wrapper-based algoritm (Colaco, et al. 2019) with 5-fold cross-validation to rank the flash flood indicators. Herein, this algoritm utilizes a comprehensive search strategy that examines every potential subset combination of the ten flash flood factors, assessing them based on the model's performance. In this process, MAE and AUC were employed to identify the optimal combination of flash flood factors for the study area. The computational results are reported in Table 4. As can be seen from this table, the factors of land use/land cover (1st rank), slope (2nd rank), and elevation (3rd rank) have shown a degree of high relevancy. Moderately relevant factors include curvature, soil, TWI, river density, rainfall, and lithology. The factor of aspect (10th rank) receives a comparatively low degree of relevancy with AUC = 0.052. Nevertheless, since all factors' AUC values are positive, all factors should be included in the flash-flood susceptibility prediction process.

Table 4

Ranking of the factors in this project using the 5-fold cross-validation-based wrapper measure

No	Factor	The wrapper measure		Ranked factor
		MAE	AUC	
1	Landcover	0.376	0.461	1
2	Slope (°)	0.360	0.429	2
3	Elevation (m)	0.286	0.380	3
4	Curvature	0.220	0.325	4
5	Soil	0.216	0.355	5
6	TWI	0.211	0.334	6
7	River Density (km/km^2)	0.167	0.269	7
8	Rainfall (mm)	0.148	0.242	8
9	Lithology	0.105	0.236	9
10	Aspect	0.007	0.052	10

Flash-flood susceptibility map

The proposed HHO-TFDNN model has been proven to be the best-suited model for flash-flood susceptibility assessment in this project by analyzing the experimental results. Thus, this newly developed deep learning model is further employed to compute the studied area's flash-flood susceptibility map.

The computational results yielded by the integrated HHO-DNN model are subsequently converted to raster format and opened in the ArcGIS software package. Based on susceptibility indices, the flood susceptibility map (refer to Fig. 7) can be obtained and displays five susceptible zones: high (5%), moderate (5%), low (5%), very low (5%), and no (80%). The threshold values (0.973, 0.413,0.139, and 0.073) for classifying these zones were established by superimposing historical flash flood locations on the flash-flood susceptibility map (Tien Bui et al. 2016). Moreover, a graphical presentation, Fig. 8, demonstrates the distribution of pixels in the five classes above computed from the HHO-TFDNN model. Table 5 further reports the five susceptibility classes' analysis results regarding the attributes of map coverage percentage, susceptibility zone, flooded location percentage, and area.

Fig. 7

The HHO-TFDNN derived flash-flood susceptibility map for the studied area

Click here to Correct

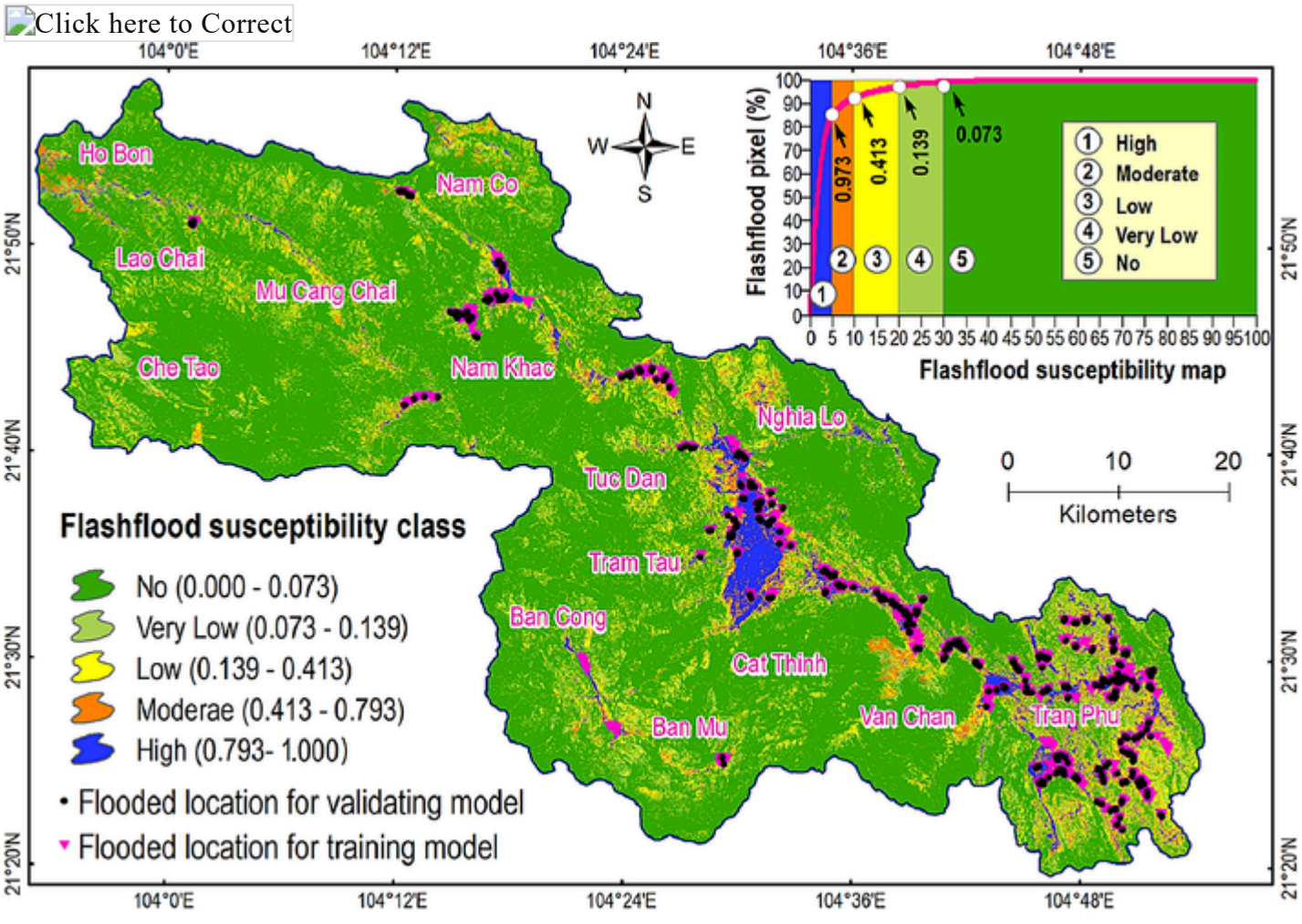


Table 5

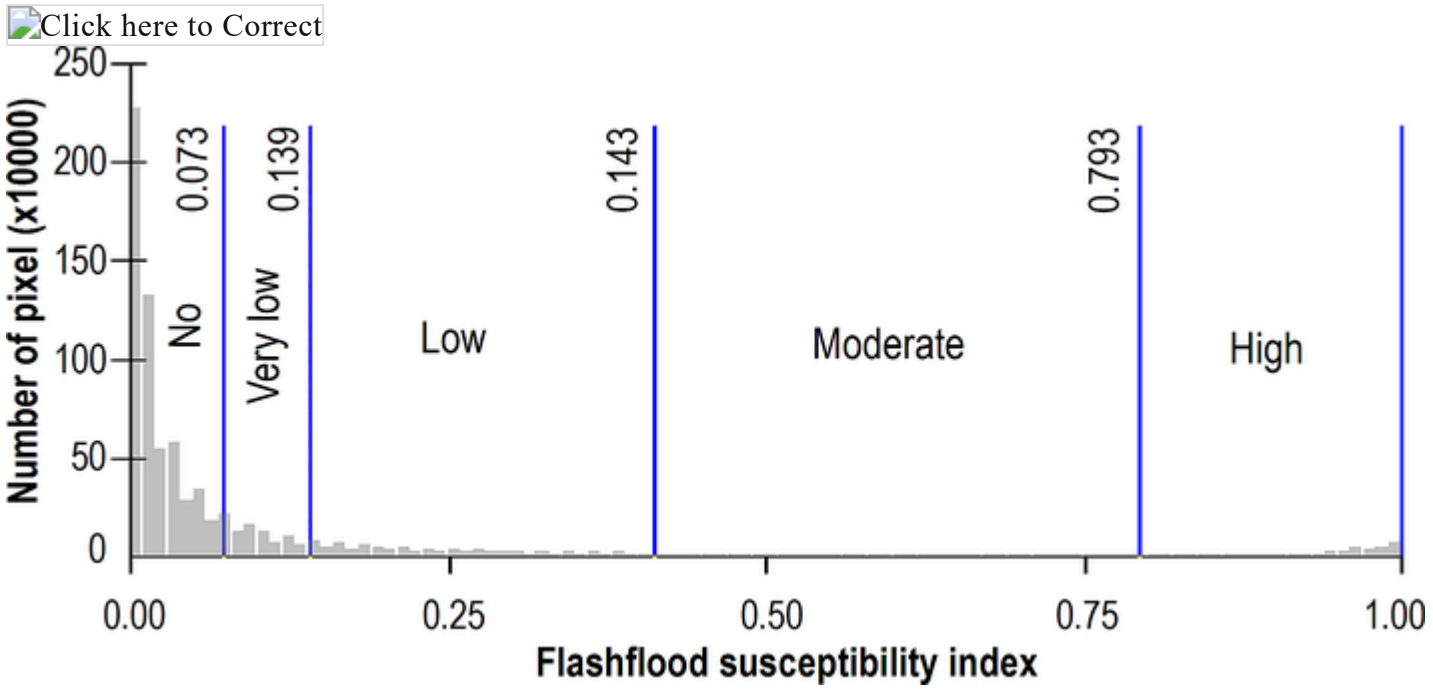
Five susceptibility classes of the HHO-TFDNN model

No	Index value	Map coverage (%)	Susceptibility zone	Flooded location (%)	Areas (km^2)
1	0.000-0.072	70	No	0.00	2221.4

No	Index value	Map coverage (%)	Susceptibility zone	Flooded location (%)	Areas (km^2)
2	0.072–0.140	10	Very low	2.94	317.3
3	0.140–0.413	10	Low	4.81	317.3
4	0.413–0.793	5	Moderate	7.60	158.7
5	0.793-1.000	5	High	84.65	158.7

Fig. 8

Distribution of pixels in the five classes derived from the HHO-TFDNN model



Discussion

Flash flooding continues to be a dangerous natural hazard event, causing severe damage to infrastructure and natural and built environments. While flash flood forecasting is still challenging (Jain et al. 2018), flash flood susceptibility identified in advance-prone areas may be a good solution for risk reduction and mitigation. In this research, we have proposed a novel framework for flash flood susceptibility modeling. The framework was tested and deployed at a heavily suffered flood area in northern Vietnam. The advantage of the TFDNN framework is that it is an open-source platform developed by Google with various available ecosystem tools suitable for environmental modelings, such as flash floods, in this research. During the last five years, TensorFlow deep learning has successfully demonstrated overwhelming applications in various domains (Pang et al. 2020).

The HHO-TFDNN is an ensemble framework that integrates TensorFlow deep learning with the Harris Hawks Optimizer (HHO) to to build a powerful flash flood model. In this framework, the TFDNN is employed initially to develop an initial flash flood model, while the HHO is utilized to search for and optimize the weights of the TFDNN model. A significant advantage of this approach is its effective capability to find and optimize the model's weights. This analysis also indicates that the weights strongly influence the prediction power of the TFDNN for flash-flood modeling. Herein, 9089 weights of the TFDNN were needed to be trained and optimized. Thus, a search space with 9089 dimensions was established. In other words, the position of a Harris' hawk in the searching space has 9089 parameters, and each Harris' hawk was a solution of the HHO-TFDNN model.

With the HHO's population of 35 Harris' hawks and 1000 iterations, 35000 TFDNN models were tested and checked for the flash flood data at hand to find the best one. Consequently, the final model's high performance indicates that the HHO can optimize the TFDNN model with good results. Besides, the proposed HHO-TFDNN model has achieved a relative balance between training (Accuracy = 92.9%) and validating accuracy (Accuracy = 91.2%). Therefore, it can be confirmed that the newly developed model has successfully tackled the overfitting issue, which is often encountered in GIS data modeling (Ahn et al. 2020; Ibrahim and Bennett 2014).

The prediction capability of the TFDNN is strongly influenced by its structure; however, no guideline is available. Therefore, in this research, we designed the deep learning model with a structure of ten input neurons, 64 neurons for each of three hidden layers, and one output neuron based on (Tien Bui et al. 2020). It is still difficult to conclude that this structure is the best for modeling in this research. Thus, further work should be carried out to determine the best structure for autonomous flash flood modeling.

Regarding the input factors, ten indicators have contributed to flash flood modeling determined by the wrapper algorithm; however, landcover and slope are the most important in this study area. This result makes sense because deforestation and steeply mountainous are the main issues for the occurrence of natural disasters in the province (Pham et al. 2020).

Concluding remarks

Flash-flood susceptibility assessment is an essential and urgent task in the study area due to increasing inclement climate conditions and land space modification driven by economic developments. This study has proposed a novel solution for flash-flood susceptibility assessment by integrating metaheuristic optimization and a deep learning framework. The HHO metaheuristic has been used as a global optimizer to adapt the employed TFDNN parameters automatically. Besides, a flash-flood database with 1866 flash flood polygons and ten indicators has been utilized for training and validating the HHO-TFDNN model. Based on the obtained findings, we have some conclusions as follows:

- The HHO-TFDNN combines open-source TensorFlow deep learning with the Harris Hawks Optimizer, presenting a novel framework suitable for flash-flood susceptibility assessment.
- Land cover and slope are the factors most influencing flash flood occurrence in this study area.
- Future extension of this study should consider other advanced metaheuristic algorithms to train deep learning models. Also, new methods and techniques for autonomously determining the deep learning structure are worth investigating.

Publisher's Note

Author contributions

Conceptualization: Le Duc Tinh, Do Thi Phuong Thao, Dieu Tien Bui, and Nguyen Gia Trong; methodology: Le Duc Tinh, Do Thi Phuong Thao, Dieu Tien Bui, and Nguyen Gia Trong; validation, Le Duc Tinh, Do Thi Phuong Thao, Dieu Tien Bui, and Nguyen Gia Trong; writing-original draft preparation: Le Duc Tinh, Do Thi Phuong Thao, Dieu Tien Bui, and Nguyen Gia Trong; writing review and editing: Dieu Tien Bui and Nguyen Gia Trong. All authors have read and agreed to the published version of the manuscript.

Funding

The study was funded by the Ministry of Education and Training in Vietnam under grant number B2022-MDA-09.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Conflict of interest

The authors declare no conflict of interest.

Electronic supplementary material

Below is the link to the electronic supplementary material.

Supplementary Material 1

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M (2015) TensorFlow: large-scale mac hine learning on heterogeneous systems

Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB (2022) Flash-flood susceptibility mapping based on XGBoost, random forest a nd boosted regression trees. Geocarto Int 37(19):5479–5496

Ahn S, Ryu D-W, Lee S (2020) A machine learning-based Approach for spatial estimation using the spatial features of Coordinate Inform ation. ISPRS Int J Geo-Information 9(10):587

Alabool HM, Alarabiat D, Abualigah L, Heidari AA (2021) Harris hawks optimization: a comprehensive review of recent variants and app lications. Neural Comput Appl 33:8939–8980

Amitrano D, Di Martino G, Iodice A, Riccio D, Ruello G (2018) Unsupervised rapid flood mapping using Sentinel-1 GRD SAR images. I EEE Trans Geosci Remote Sens 56(6):3290–3299

Baca AC, Nguyen DH (2017) Toward integrated disaster risk management in Vietnam: recommendations based on the drought and saltwat er intrusion crisis and the case for investing in longer-term resilience. The World Bank.

Bagarello V, Di Stefano C, Iovino M, Sgroi A (2013) Using a transient infiltrometric technique for intensively sampling field-saturated hy draulic conductivity of a clay soil in two runoff plots. Hydrol Process 27(24):3415–3423

Band SS, Janizadeh S, Pal SC, Saha A, Chakrabortty R, Melesse AM, Mosavi A (2020) Flash flood susceptibility modeling using new app roaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sens 12(21):3568

Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de L'hydrologie Du bassin versant. Hydrol Sci J 24(1):43–69

Borga M, Stoffel M, Marchi L, Marra F, Jakob M (2014) Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. J Hydrol 518:194–205

Bryant S, Cade C, Mellor D (1993) Permeability prediction from geologic models. AAPG Bull 77(8):1338–1350

Bryndal T, Franczak P, Kroczak R, Cabaj W, Kołodziej A (2017) The impact of extreme rainfall and flash floods on the flood risk management process and geomorphological changes in small Carpathian catchments: a case study of the Kasiniczanka river (outer carpathians, Poland). Nat Hazards 88(1):95–120

Bui DT, Lofman O, Revhaug I, Dick O (2011) Landslide susceptibility analysis in the Hoa Binh province of Vietnam using statistical index and logistic regression. Nat Hazards 59:1413–1444

Bui DT, Hoang N-D, Martínez-Álvarez F, Ngo P-TT, Hoa PV, Pham TD, Samui P, Costache R (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. Sci Total Environ 701:134413

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci Total Environ 651:2087–2096

Costache R, Ngo PTT, Bui DT (2020a) Novel ensembles of deep learning neural network and statistical learning for flash-flood susceptibility mapping. Water 12(6):1549

Costache R, Pham QB, Sharifi E, Linh NT, Abba SI, Vojtek M, Vojteková J, Nhi PT, Khoi DN (2020b) Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. Remote Sens 12(1)

Costache R, Tin TT, Arabameri A, Crăciun A, Ajin R, Costache I, Islam ARMT, Abba S, Sahana M, Avand M (2022) Flash-flood hazard using deep learning based on H2O R package and fuzzy-multicriteria decision-making analysis. J Hydrol 609:127747

Diakakis M, Deligiannakis G, Antoniadis Z, Melaki M, Katsetsiadou NK, Andreadakis E, Spyrou NI, Gogou M (2020) Proposal of a flash flood impact severity scale for the classification and mapping of flash flood impacts. J Hydrol 590:125452

Do HM, Yin KL, Guo ZZ (2020) A comparative study on the integrative ability of the analytical hierarchy process, weights of evidence and logistic regression methods with the Flow-R model for landslide susceptibility assessment. Geomatics Nat Hazards Risk 11(1):2449–2485

Duong Thi L, Do Van T, Van HL (2020) Detection of flash-flood potential areas using watershed characteristics: application to Cau River watershed in Vietnam. J Earth Syst Sci 129:1–16

Fang H, Sun L, Tang Z (2015) Effects of rainfall and slope on runoff, soil erosion and rill development: an experimental study using two loess soils. Hydrol Process 29(11):2649–2658

Gourley JJ, Erlingis JM, Hong Y, Wells EB (2012) Evaluation of tools used for monitoring and forecasting flash floods in the United States. Weather Forecast 27(1):158–173

Guzha AC, Rufino MC, Okoth S, Jacobs S, Nóbrega RLB (2018) Impacts of land use and land cover change on surface runoff, discharge and low flows: evidence from East Africa. J Hydrol Reg Stud 15:49–67

Habibi A, Delavar MR, Sadeghian MS, Nazari B, Pirasteh S (2023) A hybrid of ensemble machine learning models with RFE and Boruta wrapper-based algorithms for flash flood susceptibility assessment. Int J Appl Earth Obs Geoinf 122:103401

Hapuarachchi H, Wang Q, Pagano T (2011) A review of advances in flash flood forecasting. Hydrol Process 25(18):2771–2784

Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: Algorithm and applications. Future Generation Comput Syst 97:849–872

Hoang N-D (2020) Image processing-based spall object detection using Gabor filter, texture analysis, and adaptive moment estimation (Adam) optimized logistic regression models. Adv Civil Eng: 8829715

Hofierka J, Mitášová H, Neteler M (2009) Chapter 17 Geomorphometry in GRASS GIS. In: Hengl T, Reuter HI (eds) Developments in so il science. Elsevier, pp 387–410

Hu P, Zhang Q, Shi P, Chen B, Fang J (2018) Flood-induced mortality across the globe: spatiotemporal pattern and influencing factors. Sc i Total Environ 643:171–182

Ibrahim AM, Bennett B (2014) The assessment of machine learning model performance for predicting alluvial deposits distribution. Proce dia Comput Sci 36:637–642

Jain SK, Mani P, Jain SK, Prakash P, Singh VP, Tullos D, Kumar S, Agarwal S, Dimri A (2018) A brief review of flood forecasting techni ques and their applications. Int J River Basin Manage 16(3):329–344

Jodar-Abellan A, Valdes-Abellan J, Pla C, Gomariz-Castillo F (2019) Impact of land use changes on flash flood prediction using a sub-dai ly SWAT model in five Mediterranean ungauged watersheds (SE Spain). Sci Total Environ 657:1578–1591

Khosravi K, Pham BT, Chapi K, Shirzadi A, Shahabi H, Revhaug I, Prakash I, Tien Bui D (2018) A comparative assessment of decision tr ees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Sci Total Environ 627:744–755

Laudan J, Zöller G, Thieken AH (2020) Flash floods versus river floods – a comparison of psychological impacts and implications for pre cautionary behaviour. Nat Hazards Earth Syst Sci 20(4):999–1023

LeCun Y, Bengio Y, Hinton G (2015) Deep Learn Nat 521(7553):436–444

Li X, Gao J, Guo Z, Yin Y, Zhang X, Sun P, Gao Z (2020) A study of rainfall-runoff movement process on high and steep slopes affected by double turbulence sources. Sci Rep 10(1):9001

López V, Fernández A, García S, Palade V, Herrera F (2013) An insight into classification with imbalanced data: empirical results and cur rent trends on using data intrinsic characteristics. Inf Sci 250:113–141

Luu C, Ha H, Bui QD, Luong N-D, Khuc DT, Vu H, Nguyen DQ (2023) Flash flood and landslide susceptibility analysis for a mountaino us roadway in Vietnam using spatial modeling. Quaternary Sci Adv 11:100083

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 22(3):276–282

MONRE (2017) National disaster risk in Viet nam in the period 2006–2016 and forecasting and warning system (Issue February). https://www.apec-epwg.org/media/2309/f15e3a390421e8a5719bb2c859049604.pdf

Munna GM, Alam MJB, Uddin MM, Islam N, Orthee AA, Hasan K (2021) Runoff prediction of Surma basin by curve number (CN) meth od using ARC-GIS and HEC-RAS. Environ Sustain Indic 11:100129

Myhre G, Alterskjær K, Stjern CW, Hodnebrog Ø, Marelle L, Samset BH, Sillmann J, Schaller N, Fischer E, Schulz M (2019) Frequency of extreme precipitation increases extensively with event rareness under global warming. Sci Rep 9(1):16063

Ngo P-TT, Hoang N-D, Pradhan B, Nguyen QK, Tran XT, Nguyen QM, Nguyen VN, Samui P, Tien Bui D (2018a) A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using Sentinel-1 SAR imagery and geospatial d ata. Sens (Basel) 18(11):3704

Ngo P-TT, Hoang N-D, Pradhan B, Nguyen QK, Tran XT, Nguyen QM, Nguyen VN, Samui P, Tien Bui D (2018b) A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using sentinel-1 SAR imagery and geospatial d ata. Sensors 18(11):3704

Ngo TP-T, Hoang N-D, Pradhan B, Nguyen KQ, Tran TX, Nguyen MQ, Nguyen NV, Samui P, Tien Bui D (2018c) A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using Sentinel-1 SAR imagery and geospatial d ata. Sensors, 18(11)

Ngo P-TT, Pham TD, Hoang N-D, Tran DA, Amiri M, Le TT, Hoa PV, Van Bui P, Nhu V-H, Bui DT (2021a) A new hybrid equilibrium op timized SysFor based geospatial data mining for tropical storm-induced flash flood susceptible mapping. J Environ Manage 280:111858

Ngo P-TT, Pham TD, Nhu V-H, Le TT, Tran DA, Phan DC, Hoa PV, Amaro-Mellado JL, Bui DT (2021b) A novel hybrid quantum-PSO an d credal decision tree ensemble for tropical cyclone induced flash flood susceptibility mapping with geospatial data. J Hydrol 596:125682

Nguyen P, Thorstensen A, Sorooshian S, Hsu K, AghaKouchak A, Sanders B, Koren V, Cui Z, Smith M (2016) A high resolution coupled hydrologic–hydraulic model (HiResFlood-UCI) for flash flood modeling. J Hydrol 541:401–420

Nguyen V-N, Yariyan P, Amiri M, Dang Tran A, Pham TD, Do MP, Thi Ngo PT, Nhu V-H, Long NQ, Tien Bui D (2020) A new modeling approach for spatial prediction of flash flood with biogeography optimized CHAID tree ensemble and remote sensing data. Remote Sens 12(9):1373

Nhu V-H, Thi Ngo P-T, Pham TD, Dou J, Song X, Hoang N-D, Tran DA, Cao DP, Aydilek IB, Amiri M (2020a) A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. Remote Sens 12(17):2688

Nhu V-H, Thi Ngo P-T, Pham TD, Dou J, Song X, Hoang N-D, Tran DA, Cao DP, Aydilek İB, Amiri M, Costache R, Hoa PV, Tien D, Bui (2020b) A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible map ping. Remote Sens 12(17):2688

Nhu V-H, Hoa PV, Melgar-García L, Tien Bui D (2023) Comparative analysis of deep learning and swarm-optimized random forest for gr oundwater spring potential identification in tropical regions. Remote Sens 15(19):4761

Nie W, Yuan Y, Kepner W, Nash MS, Jackson M, Erickson C (2011) Assessing impacts of landuse and landcover changes on hydrology fo r the upper San Pedro watershed. J Hydrol 407(1):105–114

Pallard B, Castellarin A, Montanari A (2009) A look at the links between drainage density and flood statistics. Hydrol Earth Syst Sci 13 (7):1019–1029

Panahi M, Jaafari A, Shirzadi A, Shahabi H, Rahmati O, Omidvar E, Lee S, Bui DT (2021) Deep learning neural networks for spatially ex plicit prediction of flash flood probability. Geosci Front 12(3):101076

Pang B, Nijkamp E, Wu YN (2020) Deep learning with TensorFlow: a review. J Educ Behav Stat 45(2):227–248

Papagiannaki K, Lagouvardos K, Kotroni V, Bezes A (2015) Flash flood occurrence and relation to the rainfall hazard in a highly urbanize d area. Nat Hazards Earth Syst Sci 15(8):1859–1871

Papanicolaou AN, Abban BKB, Dermisis DC, Giannopoulos CP, Flanagan DC, Frankenberger JR, Wacha KM (2018) Flow Resistance int eractions on Hillslopes with heterogeneous attributes: effects on Runoff Hydrograph characteristics. Water Resour Res 54(1):359–380

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z N Gimelshein, and L Antiga (2019) Pytorch: an imperative style, high-performance deep learning library. Proc Adv Neural inf Process Syst: 8026–8037

Pham BT, Bui DT, Prakash I (2018) Bagging based support Vector machines for spatial prediction of landslides. Environ Earth Sci 77(4): 146

Pham NTT, Nong D, Garschagen M (2020) Natural hazard's effect and farmers' perception: perspectives from flash floods and landslides in remotely mountainous regions of Vietnam. Sci Total Environ: 142656

Rana MS, Mahanta C (2023) Flash-flood susceptibility modelling in a data-scarce region using a novel hybrid approach and trend analysi s of precipitation. Hydrol Sci J 68(16):2336–2356

Sangati M, Borga M, Rabuffetti D, Bechini R (2009) Influence of rainfall and soil properties spatial aggregation on extreme flash flood re sponse modelling: an evaluation based on the Sesia river basin, North Western Italy. Adv Water Resour 32(7):1090–1106

Seide F, Agarwal A (2016) CNTK: Microsoft's open-source deep-learning toolkit. Proceedings of the 22nd ACM SIGKDD International C onference on Knowledge Discovery and Data Mining, pp 2135–2135

Shahabi H, Shirzadi A, Ronoud S, Asadi S, Pham BT, Mansouripour F, Geertsema M, Clague JJ, Bui DT (2021) Flash flood susceptibility mapping using a novel deep learning model based on deep belief network, back propagation and genetic algorithm. Geosci Front 12(3):10 1100

Thao VB, Viet BX (2023) Rainfall threshold analysis for flash floods and debris flows in Lai Chau, Dien Bien, Yen Bai, and Son La provi nces. J Hydrometeorology (Vietnam) 749:96–110

Tien Bui D, Hoang N-D (2017) A bayesian framework based on a gaussian mixture model and radial-basis-function Fisher discriminant a nalysis (BayGmmKda V1.1) for spatial prediction of floods. Geosci Model Dev 10(9):3391–3409

Tien Bui D, Pradhan B, Nampak H, Bui Q-T, Tran Q-A, Nguyen Q-P (2016) Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. J Hydrol 540:317–330

Tien Bui D, Hoang N-D, Pham T-D, Ngo P-TT, Hoa PV, Minh NQ, Tran X-T, Samui P (2019) A new intelligence approach based on GIS-based Multivariate Adaptive Regression splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. J Hydrol 575:314–326

Tien Bui D, Hoang N-D, Martínez-Álvarez F, Ngo P-TT, Hoa PV, Pham TD, Samui P, Costache R (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. Sci Total Environ 701:134413

Trong NG, Quang PN, Cuong NV, Le HA, Nguyen HL, Tien Bui D (2023) Spatial prediction of fluvial flood in high-frequency tropical cyclone area using TensorFlow 1D-convolution neural networks and geospatial data. Remote Sens 15(22):5429

Truong VT, Hoang TT, Cao DP, Hayashi M, Tadono T, Nasahara KN (2019) JAXA Annual Forest Cover maps for Vietnam during 2015–2018 using ALOS-2/PALSAR-2 and Auxiliary Data. Remote Sens 11(20):2412

van Erkel AR, Pattynama PMT (1998) Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. Eur J Radiol 27(2):88–94

Wang Z, Lai C, Chen X, Yang B, Zhao S, Bai X (2015) Flood hazard risk assessment model based on random forest. J Hydrol 527:1130–1141

Yin Y, Zhang X, Guan Z, Chen Y, Liu C, Yang T (2023) Flash flood susceptibility mapping based on catchments using an improved blending machine learning approach. Hydrol Res 54(4):557–579

Zeiler M (1999) Modeling our world: the ESRI guide to geodatabase design. ESRI, Inc.

Zema DA, Labate A, Martino D, Zimbone SM (2017) Comparing different infiltration methods of the HEC-HMS model: the case study of the Mésima Torrent (Southern Italy). Land Degrad Dev 28(1):294–308