Vietnam Journal of Earth Sciences, 45(3), 338-356, https://doi.org/10.15625/2615-9783/18500

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

A novel HHO-RSCDT ensemble learning approach for forest fire danger mapping using GIS

Tran Xuan Truong[1], Doan Thi Nam Phuong[1], Le Thanh Nghi[1], Viet-Ha Nhu[2,*], Dieu Tien Bui[3]

[1]*Department of Photogrammetry and Remote Sensing, Faculty of Geomatics and Land Administration, Hanoi University of Mining and Geology, Hanoi, Vietnam*
[2]*Department of Geological-Geotechnical Engineering, Hanoi University of Mining and Geology, Hanoi, Vietnam*
[3]*GIS Group, Department of Business and IT, University of South-Eastern Norway, Gullbringvegen 36, 3800 Bø i Telemark, Norway*

Received 4 May 2023; Received in revised form 22 June 2023; Accepted 10 July 2023

ABSTRACT

Accurate prediction models for spatial prediction of forest fire danger play a vital role in predicting forest fires, which can help prevent and mitigate the detrimental effects of such disasters. This research aims to develop a new ensemble learning model, HHO-RSCDT, capable of accurately predicting spatial patterns of forest fire danger. The HHO-RSCDT method combines three distinct components, namely Random Subspace (RS), Credal Decision Tree (CDT), and Harris Hawks Optimizer (HHO). Herein, RS generates a series of subspace datasets, which are subsequently utilized to produce individual CDT classifiers. Then, HHO optimizes the ensemble model, enabling the model to achieve higher predictive performance. The model was trained and validated using a forest fire dataset at Phu Yen province, Vietnam. The dataset includes 306 forest fire locations and ten influencing factors from the study province. The results showed the capability of the HHO-RSCDT model in predicting forest fire danger, with an accuracy rate of 83.7%, a kappa statistic of 0.674, and an AUC of 0.911. A comparison between the HHO-RSCDT model and two state-of-the-art machine learning methods, i.e., support vector machine (SVM) and random forest (RF), indicated that the HHO-RSCDT performs better, making it a valuable tool for modeling forest fire danger. The forest fire danger map produced using this novel model could be a new tool for local authorities in the Phu Yen province, assisting them in managing and protecting the forest ecosystem. By providing a detailed overview of the areas most susceptible to forest fires, the map can help authorities to develop targeted and effective forest management strategies, such as focusing on areas with high fuel loads or implementing controlled burning programs.

Keywords: Forest fire; Random Subspace; Credal Decision Tree; Harris Hawks Optimizer; GIS; Vietnam.

1. Introduction

Forest fire is one of the most significant ecological and environmental problems in many countries (Michael et al., 2021), posing a severe threat to forest ecosystems, human life, and property (McWethy et al., 2019; Agbeshie et al., 2022). The risk of forest fires is increasing, especially in areas with long dry seasons, high temperatures, and strong winds (Leigh et al., 2015; Tavakol, 2020). One of the most effective strategies to mitigate forest fires' adverse impacts is to predict the risk of

*Corresponding author, Email: nhuvietha@humg.edu.vn

forest fires accurately. Thus, forest fire danger mapping is an essential tool for forest fire prevention and management, enabling early warning and timely decision-making to minimize the impact of forest fires (Van Le et al., 2021). Therefore, developing accurate forest fire danger prediction models has become increasingly important for forest management and protection (Tehrany et al., 2019; Oliveira, 2021; Naderpour et al., 2019).

Forest fire danger can be defined as the probability of a fire breaking out and rapidly spreading in a particular geographic region. This probability is determined by several factors, including but not limited to prevailing weather patterns, the presence of combustible materials, and the topographical characteristics of the area. The spatial prediction of forest fire danger offers insight into the areas where fires are prone to occur but does not provide a precise forecast for when a fire event might happen (Tehrany et al., 2019). Literature review shows numerous studies have been conducted on this topic, and various models have been proposed for spatial prediction of forest fire danger (Chicas and Østergaard Nielsen, 2022). These models employ various methods, including numeric simulations (Lattimer et al., 2020; Mell et al., 2007; Moinuddin and Sutherland, 2020; Zhu et al., 2018; Pimont et al., 2012; Iudin et al., 2015), statistical analysis (de Santana, 2021; Dos Reis et al., 2021; Sivrikaya and Küçük, 2022; Storey and Price, 2022a, 2022b), machine learning (Pourghasemi et al., 2020; Jain et al., 2020; Ngoc Thach et al., 2018; Bui et al., 2017; Sachdeva 2018), deep learning (Van Le et al., 2021; Rihan et al., 2023; Naderpour et al., 2021), and ensemble learning (Tehrany et al., 2019; Moayedi et al., 2020; Bjånes et al., 2021; Xie and Peng, 2019; Tuyen et al., 2021). Reviews of these methods can be found in (Naderpour et al., 2019; Jain et al., 2020; Abid, 2021).

Among machine learning approaches, it has been observed that ensemble learning methods tend to provide higher levels of accuracy (Naderpour et al., 2019). Ensemble learning combines multiple models or algorithms to achieve better predictive power than any single model or algorithm can provide alone. This technique can improve the accuracy of machine learning models, particularly in complex and large-scale problems, i.e., forest fire modeling, where single models may not be sufficient to capture all of the relevant geo-environmental information. Thus, ensemble methods are increasingly recognized for their potential to enhance the accuracy of predictions in a wide range of applications (Ganaie et al., 2022).

Despite the recognized benefits of ensemble learning in enhancing prediction accuracy (Sagi and Rokach, 2018), no single method or technique is universally the best for spatially predicting forest fire danger in all regions. Thus, the study of new ensemble models to improve the accuracy of spatial prediction for forest fire danger is still essential, as the need for accurate and timely forest fire risk assessment continues to grow, especially when new geospatial data sources i.e., Sentinel 2 imagery, Landsat 8 OLI, and Landcover data (Phan et al., 2021) are available.

This study addresses the research mentioned above gap by introducing and validating a novel ensemble learning method, namely HHO-RSCDT. This method is designed to accurately predict forest fire danger spatial patterns, specifically focusing on a case study conducted in Phu Yen province. The HHO-RSCDT approach involves combining the strengths of Random Subspace (RS) (Ho, 1998), Credal Decision Tree (CDT) (Mantas 2016), and Harris Hawks Optimizer (HHO) (Heidari et al., 2019) to achieve improved prediction performance. RS is utilized to generate multiple subsets of the input data, each of which is then used to train a separate CDT classifier. The HHO algorithm is then employed to optimize the ensemble model and obtain the best possible combination of the trained classifiers. This approach aims to enhance the accuracy of forest fire danger prediction by leveraging each component algorithm's strengths and optimizing their collective performance.

2. Background of the algorithms used

2.1. Credal Decision Tree

The Credal Decision Tree (CDT) introduced by Mantas and Abellán, (2014a) is considered for fire danger modeling in this research because it allows for the consideration of imprecise and uncertain information (Mantas and Abellán, 2014b), which is often present in multisource geospatial derived input data, i.e., weather conditions, vegetation, topography, and human activities.

Compared to other decision tree algorithms, the CDT algorithm's main distinction lies in the method it employs to choose features, i.e., factors that affect forest fires, for the division at each node of the tree. In this case, the selection criteria are based on the imprecise probabilities and uncertainties of the credal sets (CS) (Mantas and Abellán, 2014b), which enables the algorithm to manage noisy data effectively (Mantas, 2016).

Suppose we have a training dataset of forest fire Z, where Z consists of pairs of inputs and outputs, represented by $x_i \in R^d$ and $y_i \in (0,1)$, respectively. Here, i ranges from 1 to N, which denotes the total number of samples, and d represents the dimensionality of the input samples. In this research, d equals ten, indicating ten forest fire influencing factors. The output variable y_i takes values from 0 to 1, corresponding to the two possible classes: fire and non-fire. The CDT aims to construct a set of classification models f: $R^d \rightarrow [0,1]$ that can predict the likelihood of forest fire occurrence based on ten driving factors. The classification models are constructed using a set of credal sets that describe the uncertainty and imprecision in the input data, CS^Z, which are subsets of Z using Eq.1 (Michael et al., 2021) below:

$$\mathrm{CS^Z} = \left\{\mathrm{p/p(FF)} \in \left[\frac{\mathrm{n(FF)}}{\mathrm{N+s}}, \frac{\mathrm{n(FF)+s}}{\mathrm{N+s}}\right]\right\} \quad (1)$$

where p(FF) is the probability distribution, FF is a forest fire influencing factor, N is the sample size, n(FL) is the frequent value, and s is a hyperparameter of the CDT algorithm with a default value of 1.

2.2. Random Subspace Ensemble

Random Subspace (RS) is a machine learning ensemble technique proposed by Ho (Ho, 1998) for constructing multiple decision trees, and then, these decision trees are aggregated to form an ensemble model. This technique helps to reduce the risk of overfitting and can improve the overall accuracy and robustness of the prediction model. Based on the training dataset Z described in Section 2.1, the RS technique creates k subsets from the dataset, where each subset is referred to as a subspace of Z. Herein, each subspace is created by selecting a random subset of m (where $m \leq d$) forest fire influencing factors from the full set of factors available in the dataset, where d is the total number of available influencing factors in the dataset. Then, each k is used to generate a CDT classifier, and finally, the ensemble model is derived by combining all CDT classifiers. The global performance of RS is influenced by k, m, the maximum depth of the CDT tree (mCDTree), and the minimum total weight of the instances in a leaf (minNum), and Harris Hawks Optimizer can optimize these parameters in the next section.

2.3. Harris Hawks Optimizer

Harris Hawks Optimizer (HHO) proposed by Heidari et al., (2019) is a metaheuristic optimization algorithm inspired by the hunting behavior of Harris hawks in nature. The algorithm simulates Harris hawks' social behavior and hunting strategies to solve optimization problems. HHO has shown promising results in solving various optimization problems in different domains, including engineering, finance, and landslides (Alabool et al., 2021). The algorithm has also demonstrated faster convergence and better accuracy than other optimization algorithms, making it a popular choice for many researchers and practitioners.

In this Research, HHO is used to optimize four parameters (k, m, mCDTree, and minNum) of the Random Subspace based Credal Decision Tree (RSCDT) model with the steps below:

Step 1: Firstly, the searching space for four parameters, namely k, m, mCDTree, and minNum, is defined. Then, a population of hawks (nHawks) is initialized, and these hawks have four coordinates each, which correspond to the values of the four parameters (k, m, mCDTree, and minNum) and represent a possible solution for the forest fire danger model.

Step 2: Evaluate the fitness of each hawk in the population based on an objective function (see Section 4.2). Then, update the hawks' position based on their fitness using Eq. 2; this is called the search phase.

$$\mathrm{Pos}(t+1) = \begin{cases} \mathrm{Pos_{rand}}(t) - r_1|\mathrm{Pos_{rand}}(t) - 2r_2\mathrm{Pos}\ (t)| & \text{with } q \geq 0.5 \\ |\mathrm{Pos_{Pr}}(t) - \mathrm{Pos_m}(t)| - r_3\big(LB + r_4(UB - LB)\big) & \text{with } q < 0.5 \end{cases} \quad (2)$$

where $\mathrm{Pos}(t)$ and $\mathrm{Pos}(t+1)$ are the position of the hawks at the iteration t and the iteration t +1, respectively; $\mathrm{Pos_{Pr}}(t)$ is the individual position with the best fitness, also called the location of the prey at iteration t; r_1, r_2, r_3, r_4, and q are random numbers between 0 and 1; $\mathrm{Pos_m}(t)$ is the average position of individuals.

Step 3: $\mathrm{Pos}(t+1)$ is updated using Eq. 2 if $|E| \geq 1$ or the following equations when $|E| < 1$ (Peng, L., et al., 2023).

$$\mathrm{Pos}(t+1) = \begin{cases} \Delta\mathrm{Pos}(t) - E|J\mathrm{Pos_{Pr}}(t) - \mathrm{Pos}(t)|; & 0.5 \leq |E| < 1 \ \&\ r \geq 0.5 \\ \mathrm{Pos_{Pr}}(t) - E|\Delta\mathrm{Pos}(t)| & ; |E| < 0.5 \ \&\ r \geq 0.5 \end{cases} \quad (3)$$

$$\mathrm{Pos}(t+1) = \begin{cases} \mathrm{Pos_{Pr}}(t) - E|J\mathrm{Pos_{Pr}}(t) - \mathrm{Pos}(t)|; & 0.5 \leq |E| < 1 \ \&\ r < 0.5; F1 \\ \mathrm{Pos_{Pr}}(t) - E|J\mathrm{Pos_{Pr}}(t) - \mathrm{Pos}(t)| + S + \mathrm{Levy}; & 0.5 \leq |E| < 1 \ \&\ r < 0.5; F2 \end{cases} \quad (4)$$

$$\mathrm{Pos}(t+1) = \begin{cases} \mathrm{Pos}_{Pr}(t) - E|J\mathrm{Pos}_{Pr}(t) - \mathrm{Pos}_m(t)|; & |E| < 0.5 \ \&\ r < 0.5;\ F1 \\ \mathrm{Pos}_{Pr}(t) - E|J\mathrm{Pos}_{Pr}(t) - \mathrm{Pos}_m(t)| + S + \mathrm{Levy}; & |E| < 0.5 \ \&\ r < 0.5; F2 \end{cases} \quad (5)$$

where E is the escape energy of prey; r is the random number between 0 and 1; J is the random number between 0 and 2; F1 and F2 are the fitness conditions; S is a random vector; and Levy is the Levi's flight function.

Step 4 - Termination: Stop the algorithm when a stopping criterion is met, such as reaching the maximum number of iterations or the desired fitness level.

3. Study area and data

3.1. Study area

The study area is Phu Yen, a coastal province in the south-central region of Vietnam (see Fig. 1). It lies between longitudes 108°41’E and 109°28’E and latitudes 12°42’N and 13°42’N, covering an area of 5,049.6 km^2. The province boasts a diverse topography encompassing mountains, hills, plains, and coastal areas. Mountains and hills comprise 70% of the land area, while the plain is narrow and heavily dissected. In contrast, the coastal areas are relatively flat and sandy. The elevation ranges from –46.1 m to 1706.3 m. The slope of the province is from 0 to 59.9° with a mean of 14.41 m and a standard deviation of 9.43 m.

As of 2022, the population of the Phu Yen province is 876,619 people, and the population density is 174 people/km^2 (www.phuyen.gov.vn, accessed on 28 March 2023), and about 70% of the population lives in rural areas, where agriculture, forestry, aquaculture, and fisheries are the primary likelihood (Truong and Tri, 2021). Phu Yen province boasts a rich and diverse range of vegetation types well-suited to its varied geography and climate (Hoi and Dung, 2022). Coastal areas are characterized by mangroves, which are well-adapted to saline conditions and play a critical role in protecting the coast from storms and erosion. In addition, the province's coastal sand dunes provide a unique habitat for specialized plant species adapted to these environments' harsh conditions. Mountainous forests dominate the landscape at medium elevations, featuring a mix of evergreen and deciduous trees and planted forests (Tri et al., 2019). The mix of the evergreen forest is dominant and accounts for more than 95% of the natural forest that provide essential ecosystem services, such as carbon sequestration, soil stabilization, and water regulation. At higher elevations, mountainous forests take over, characterized by a mix of broadleaf that provide habitat for a wide range of wildlife (Hoi and Dung, 2022; Lung et al., 2011).

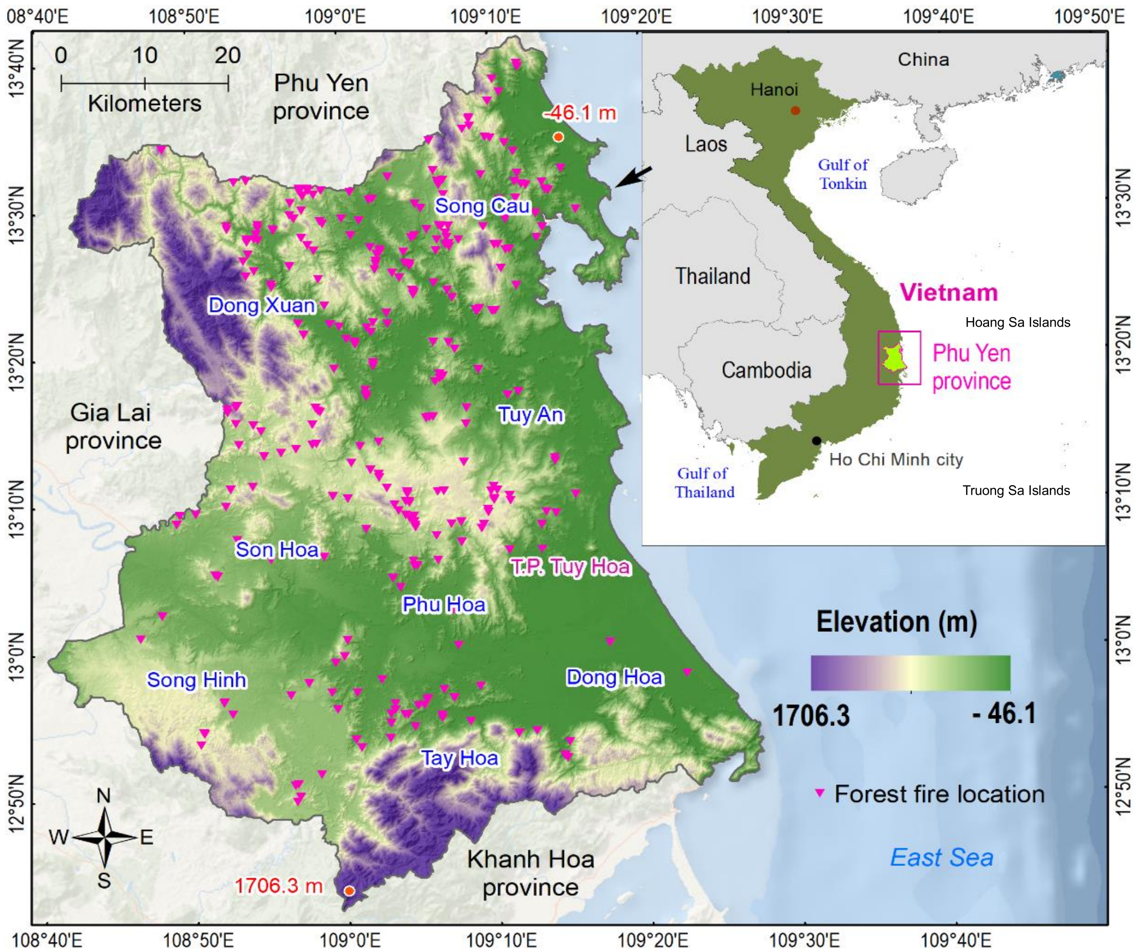

Figure 1. Study area and forest fire locations

The climate of the Phu Yen province features tropical monsoon, hot and humid, and is influenced by oceanic climate (Lee and Dang, 2018). The average temperature is around 26.5°C. There are two distinct seasons: the rainy season from September to December and the dry season from January to August. In recent years, Phu Yen province has experienced several El Niño-related droughts, which led to significant crop losses and water shortages in the region (Duong et al., 2020).

Hot weather and prolonged drought on a large scale are one of the leading causes of forest fires in the province. For example, on June 27, 2019, at 4 pm, a fire broke out in the Ro Huou mountain region, likely caused by the scorching sun and high temperatures. The fire quickly spread to the adjacent forests, burning down a significant portion of the planted forest aged between 3 to 7 years old. The location of the fire was a steep hill with gusty winds exacerbating the situation. Other forest fires broke out in Dinh Thai village on September 10, 2019. The fire escalated rapidly due to the challenging terrain and strong winds resulting in the burning of 10 hectares of planted forest belonging to the residents

3.2. Data used

3.2.1. Forest fire inventory

Historical forest fire data are crucial to develop accurate models for predicting forest fire danger. This data provides valuable information on where fires have occurred and the conditions that contributed to their ignition and spread. As a result, it is essential to create a detailed forest fire inventory map as the first step in the modeling process. In this research, the forest fire inventory map with 306 fire locations from 2019-2023 in a national project B2021-MDA-13 prepared by Truong et al., 2023a was used (Fig. 1). This project was funded by the Ministry of Education and Training (MoET) of Vietnam. These forest fires were compiled from several sources, including the forest fire prevention database of the Ministry of Agriculture and Rural Development of Vietnam (www.watch.pcccr.vn, accessed on 28 March 2023), the Fire Information for Resource Management System (FIRMS) project of NASA (www.firms.modaps.eosdis.nasa.gov, accessed on 28 March 2023), and fieldwork with handheld GPS. According to statistical analysis, most forest fires (around 76%) occurred during June, July, August, and September. Additionally, it was found that most of these fires occurred during the daytime.

3.2.2. Influencing factors

Identifying the influencing factors for forest fires in the Phu Yen province is essential for predicting the spatial prediction of fire danger in forests. Based on our examination of the historical data on forest fires in the province, it is evident that climate conditions, vegetation type, topography, and human activities were the primary factors that led to forest fires. Therefore, this research, 10 ignition factors were considered, including elevation, slope, aspect, distance to road, NDVI, NDWI, relative humidity, temperature, and rainfall.

In this research, a digital elevation model (DEM) 30 m resolution for the study area was derived from the ALOS DEM of JAXA (available at www.eorc.jaxa.jp, accessed on 28 March 2023). The DEM was produced using data captured by the Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM), which was installed on the Advanced Land Observing Satellite (ALOS) (Takaku and Tadono, 2009). Then, the DEM was utilized to obtain three influencing factors, namely elevation (Fig. 2a), slope (Fig. 2b), and aspect (Fig. 2c), using the Spatial Analysis tool in ArcGIS Pro.

The elevation is often considered a critical factor in fire regimes as it influences an area's weather and climatic conditions, impacting fuel moisture content and availability (Miller and Urban, 2000). While slope should be used because it affects the rate and direction of forest fire spread, making it an essential input for modeling and predicting fire behavior (Zheng et al., 2017). In the case of aspect, this factor influences the exposure of vegetation to sunlight and wind, affecting its moisture content and potential to fuel fires, making it relevant in forest fire modeling (Slijepcevic et al., 2018).

Many previous studies considered the distance to the road in predicting the potential spread and impact of forest fires (Ngoc Thach et al., 2018; Robinne et al., 2016; Gonzalez-Olabarria et al., 2019) because it can influence the likelihood of a fire occurring in a particular area, as humans are often the cause of forest fires. Roads provide easier access for people to enter remote forested areas. In this research, we generated a distance to road map (shown in Fig. 2d) for the study area by buffering the road networks and then classified the distance into five categories: 0–120, 120–240, 240–480, 480–900, and >900 m. The road networks used for the map were sourced from the national topographic map of Vietnam at a scale of 1:50,000 and the Open Street Map (www.openstreetmap.org, accessed on 28 March 2023).

Satellite-derived indices, Normalized Difference Vegetation Index (NDVI), and Normalized Difference Water Index (NDWI), are important influencing factors for forest fire prediction as they can provide information about the vegetation and moisture conditions in the Phu Yen province. Herein NDVI provides information on vegetation health and density (Gouveia et al., 2012), while NDWI indicates the presence of water (Teng et al., 2021). Both these indices can indicate the flammability of vegetation, with low NDVI and high NDWI values indicating high moisture content and lower flammability. Conversely, high NDVI and low NDWI values indicate that drier vegetation is more susceptible to fire.

The study computed NDVI and NDWI (as shown in Figs. 2e and 2f, respectively) for the study area using the reflectance values of Landsat 8 OLI imagery with a 30 m resolution. Two images captured on April 10, 2022, were used, and Eqs. 6 and 7 (Ke et al., 2015; Gao, 1996) were employed to calculate NDVI and NDWI. The Landsat 8 OLI imagery was obtained from the USGS archive at www.earthexplorer.usgs.gov (accessed on 28 March 2023).

$$NDVI = (Band\ 5 - Band\ 4)/(Band\ 5 + Band\ 4) \quad (6)$$

$$NDWI = (Band\ 5 - Band\ 6)/(Band\ 5 + Band\ 6) \quad (7)$$

Relative humidity should be selected because it measures the amount of moisture in the air relative to its capacity and affects the moisture content of fuel materials in the forest (Matthews et al., 2012). Herein, when the air is dry, the fuel is also likely to dry, making it more susceptible to catching fire. On the other hand, if the air is moist, the fuel retains more moisture, reducing its flammability. Surface temperature is another critical climatic factor influencing forest fire behavior (Liu, 2014). High surface temperatures can lead to high fuel temperatures, increasing the likelihood of ignition and fire spread. In addition, high temperatures can increase plants' evapotranspiration rate, which reduces fuel materials' moisture content and makes them more susceptible to catching fire.

Rainfall is also a critical factor that affects forest fire prediction because it can significantly impact the moisture content of fuels in the forest (Flannigan et al., 2016). Adequate rainfall can help reduce the risk of forest fires by increasing fuel materials' moisture content, making them less likely to catch fire. Conversely, drought conditions can lead to dry fuels that are more susceptible to fire. In this analysis, relative humidity (Fig. 2g), surface temperature (Fig. 2h), and rainfall (Fig. 2i) were obtained for the study area by utilizing the climatic data provided by NASA (USA), which is accessible through www.power.larc.nasa.gov (accessed on 28 March 2023). Herein, the climate data from June, July, August, and September of 2018-2020 were utilized since these months have been observed to have a high incidence of forest fires.

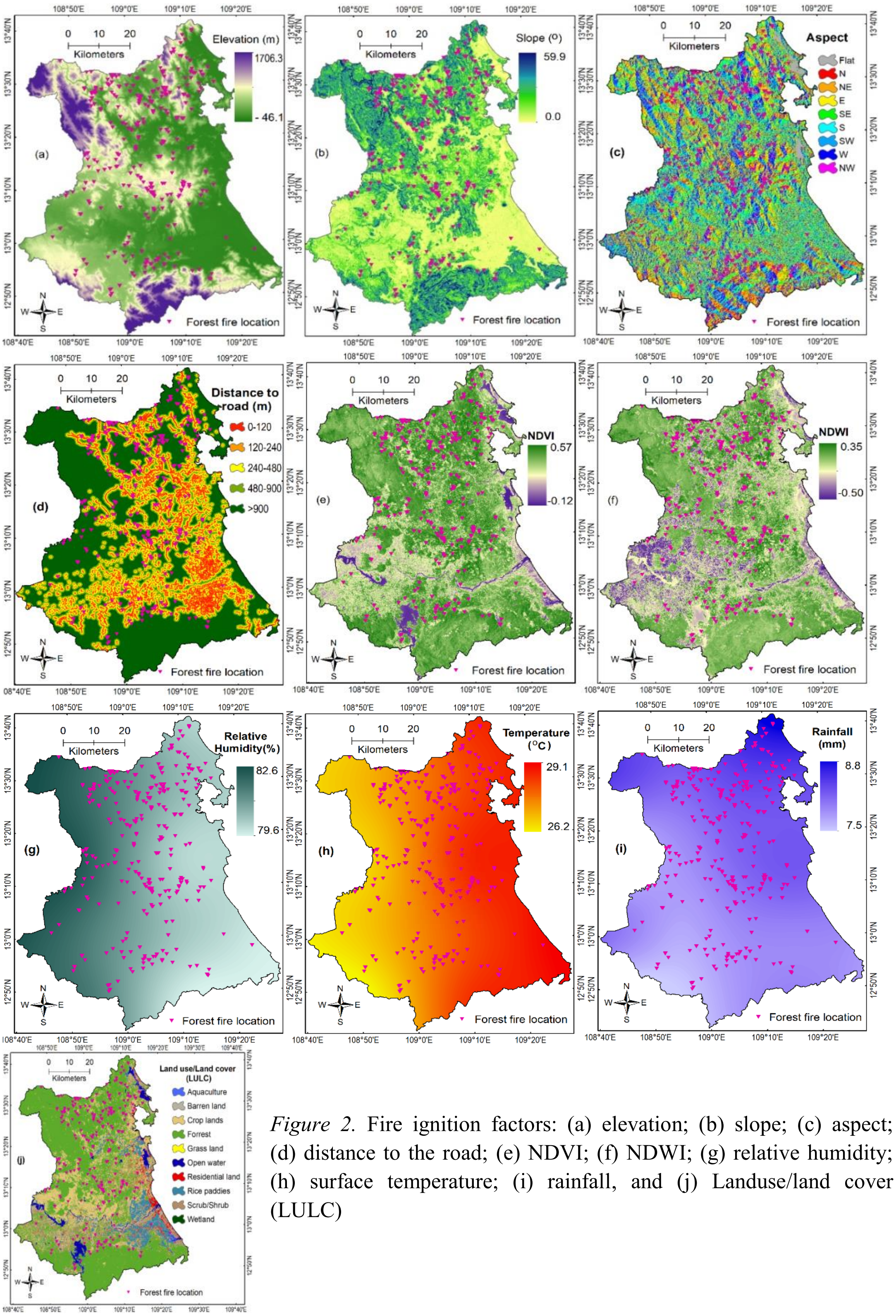

Figure 2. Fire ignition factors: (a) elevation; (b) slope; (c) aspect; (d) distance to the road; (e) NDVI; (f) NDWI; (g) relative humidity; (h) surface temperature; (i) rainfall, and (j) Landuse/land cover (LULC)

Finally, land use/land cover (LULC) is essential for forest fire prediction because it may influence the fuel load and flammability of different land cover types (Calviño-Cancela et al., 2016). Land cover types, such as forests, grasslands, and shrublands, have different vegetation densities and fuel loads (Rouet-Leduc et al., 2021), which can affect the spread and intensity of a fire. Moreover, land cover types can have varying moisture content and flammability, which can be influenced by weather conditions such as rainfall and temperature (Littell et al., 2016). Hence, incorporating LULC data may provide critical information for modeling and predicting forest fires. In this research, the study area's LULC map with 30 m resolution for the Phu Yen province with 13 different classes (shown in Fig. 3j) was created by compiling the 30-meter resolution LULC map products of 2020 produced by the Japan Aerospace Exploration Agency (JAXA), which can be downloaded at www.eorc.jaxa.jp, accessed on 28 March 2023.

4. Proposed HHO-RSCDT for improving the accuracy of forest fire danger mapping

This section describes the proposed HHO-RSCDT ensemble learning for forest fire danger mapping (Fig. 3). We process the multisource geospatial data using ArcGIS Pro 2.8.0 software. The Matlab code for the Harris Hawks optimization (HHO) algorithm can be accessed from (Heidari et al., 2019). At the same time, the HHO-RSCDT was implemented by the researchers in the Matlab platform using the Matlab Weka Classifiers tool provided by should be Dunham, (2023). In addition, a Python script was also programmed to code the ten influencing factors for the HHO-RSCDT model and then convert the result into the final forest fire danger map. The conceptual framework of the HHO-RSCDT proposed in this research is demonstrated in Fig. 3.

4.1. Data processing and building a GIS database

The first step is data processing which processes the multisource geospatial data into ten forest fire-influencing factors using ArcGIS Pro software to build a GIS database (Fig. 3). Then, all the factor maps were converted into a raster format 30 m resolution. Next, eight continuous factors (elevation, slope, NDVI, NDWI, relative humidity, temperature, and rainfall) were normalized (Eq. 8) (Bui et al., 2017) in a range [0.01−0.99] using the Spatial Analysis tool in ArcGIS Pro. For the remaining factors (aspect, distance to road, and LULC), each category was assigned an integer attribute (Bui et al., 2012) and then normalized in the range [0.01−0.99] above using Eq. 8.

$$\mathrm{Nv} = \frac{\mathrm{Fa_i} - \mathrm{Min(Fa)}}{\mathrm{Max(Fa)} - \mathrm{Min(Fa)}}[0.99 - 0.01] + 0.01 \quad (8)$$

where $\mathrm{Fa_i}$ is the value of the considered factor. $\mathrm{Min(Fa)}$ and $\mathrm{Max(Fa)}$ are the minimum value and the maximum value of the considered factor, and Nv is the newly computed value for the considered factor.

The subsequent stage involved the creation of 306 non-fire locations chosen randomly from areas in the Phu Yen province that were unaffected by forest fires, as determined by an NDVI score of less than 1.5. Following this, a "1" was attributed to each of the 306 fire locations, while a "0" was assigned to each of the 306 non-fire locations. Afterward, the locations were randomly divided into two sets. The first subset, which comprised 428 locations (214 forest fire locations and 214 non-fire locations), was utilized to construct a training dataset, while the remaining 184 locations (92 forest fire locations and 92 non-fire locations) were reserved for model validation purposes. A sampling procedure was executed to extract the values of the ten factors influencing the data for both the training and validation datasets.

In forest fire modeling, assessing the role of influencing factors is crucial for the prediction model's accuracy and reliability. The role of the influencing factor was assessed using the Wrapper Subset Evaluator method (Abawajy and Kelarev, 2017).

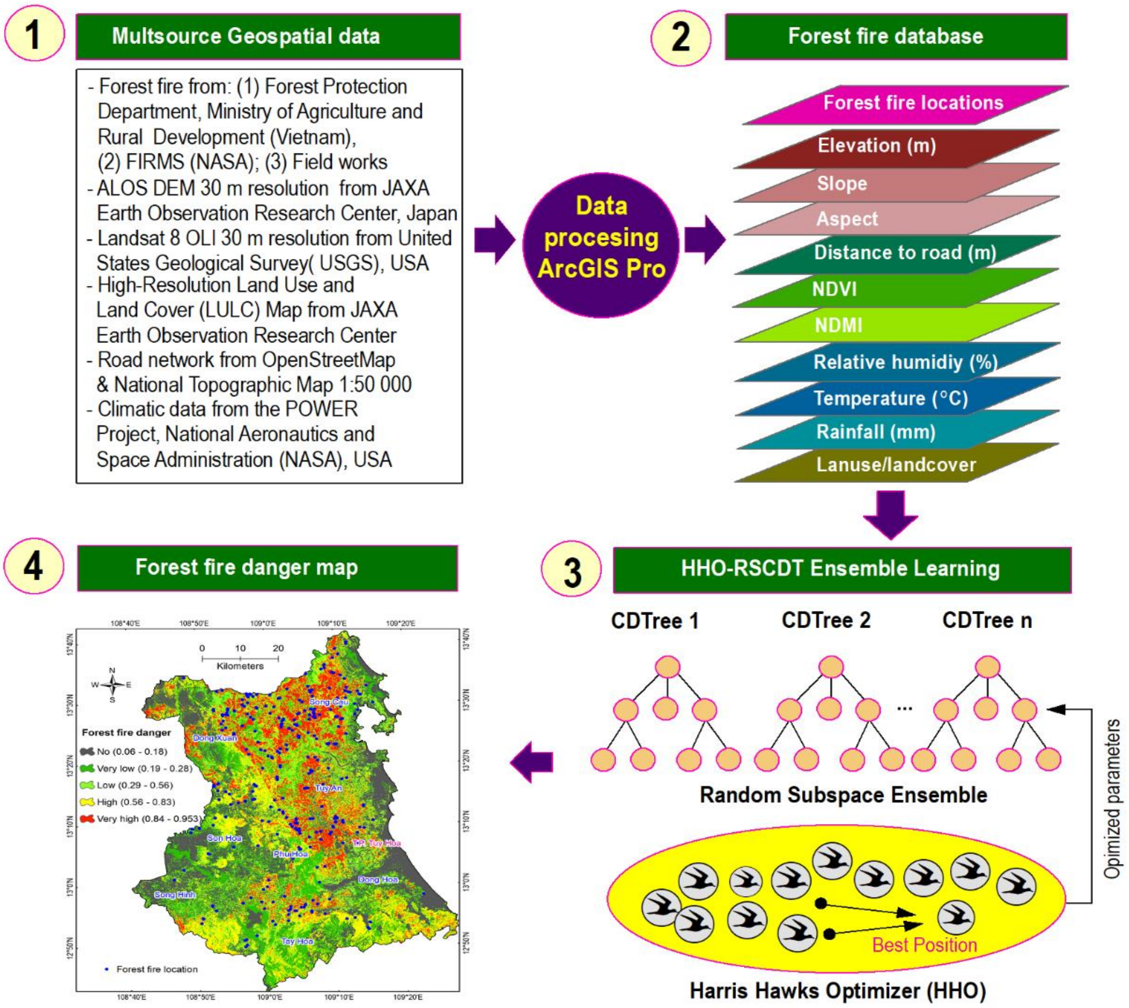

Figure 3. The proposed HHO-RSCDT Ensemble learning for forest fire danger mapping

4.2 Configuration of the HHO-RSCDT Ensemble model

The process of configuring the HHO-RSCDT ensemble model involves two stages: the first is integrating the Credal Decision Tree into the Random Subspace ensemble, and then, four parameters (k, m, mCDTree, and minNum) were extracted and connected to the HHO for searching and optimizing them. Subsequently, a four-dimensional search space was established, as explained in Section 2.3. Herein, the coordinates of each hawk in the four-dimensional searching space corresponded to four parameters (k, m, mCDTree, and minNum). Therefore, each hawk was a potential solution for the HHO-RSCDT ensemble model and was measured by an objective function in Eq. 9.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{FFI_i} - \mathrm{FFO_i})^2 \qquad (9)$$

where MSE is the Mean Squared Error; $\mathrm{FFI_i}$ is the fire danger value in the training dataset Z (see Section 2.1); $\mathrm{FFO_i}$ is the fire danger output from the HHO-RSCDT model; N is the total number of the training samples used.

In the optimization process using HHO, a population of 30 hawks was selected, and

1000 iterations were employed. The search space was established as 4-dimensional, with k ∈ [1–500], ensuring a maximum of 500 trees were used to maintain the diversity of the ensemble. We employed 10 forest fire influencing factors, so m ∈ [1–10]. Additionally, mCDTree was set to a maximum of 100, limiting the depth of the tree to 100, mCDTree ∈ [1–100], and minNum ∈ [1–10].

4.3. Evaluation metrics

The performance of the HHO-RSCDT model was evaluated using standard statistical metrics for two-class pattern classification, fire, and non-fire, including classification accuracy (Acc), Kappa, Receiver Operating Characteristic (ROC) curve, and Area under the curve (AUC). Because the literature extensively covers these metrics, commonly used in forest fire danger modeling (Chicas and Østergaard Nielsen, 2022). Hence, we refrain from providing an elaborate explanation of these metrics, and thus, interested readers may refer to relevant articles for more information, i.e., (Tehrany et al., 2019; Ngoc Thach et al., 2018).

4.4. Forest fire danger map

Once the HHO-RSCDT model is successfully trained and validated, the model was then used to compute the forest fire danger map for all pixels of the study area. In the next step, five classes, very high, high, low, very low, and no (Bui et al., 2017) were used. The values for separating these classes were determined using a plot graph (Bui et al., 2017) generated by crossing the forest fire inventory map with the forest fire danger.

5. Results and analysis

5.1. Training and validating result

The training phase of the HHO-RSCDT model was conducted over 1000 iterations, and the optimal combination of the four parameters was determined to be: k = 10, m = 5, mCDTree = 29, and minNum = 3.0. The performance of the HHO-RSCDT model on the training dataset with 5-fold cross-validation is shown in Table 1 and Fig. 4. Our analysis shows that the model has a high level of goodness-of-fit, with an accuracy (Acc) of 80.6%. The Kappa statistic, which measures the agreement between predicted and observed values, is also 0.612, indicating a satisfactory result. The model's overall performance, as measured by AUC with 5-fold cross-validation, is 0.884, and the AUC increased to 0.952 (Fig. 4) using the whole training dataset, indicating a high level of performance. The other performance metrics are presented in Table 1.

Table 1. Performance metrics of the HHO -RSCDT model

HHO-RSCDT ensemble model	Statistical metrics										
	TP	TN	FP	FN	PPV	NPV	Sens	Spec	Acc	Kappa	AUC
Training dataset	149	196	65	18	69.6	91.6	89.2	75.1	80.6	0.612	0.884
Validating dataset	73	81	19	11	79.3	88.0	86.9	81.0	83.7	0.674	0.911

To assess the HHO-RSCDT model's performance on new data, we utilized the validation dataset, and the result is also shown in Table 1. The HHO-RSCDT model's accuracy (Acc) was determined to be 83.7%, indicating that 83.7% of the validation dataset samples were classified correctly, which is a good result. Moreover, the Kappa statistic of 0.674 indicates a satisfactory result. The global performance of the model measured by AUC is 0.911 (Table 1 and Fig. 4), indicating a high prediction power of the model. The other validated metrics are presented in Table 1.

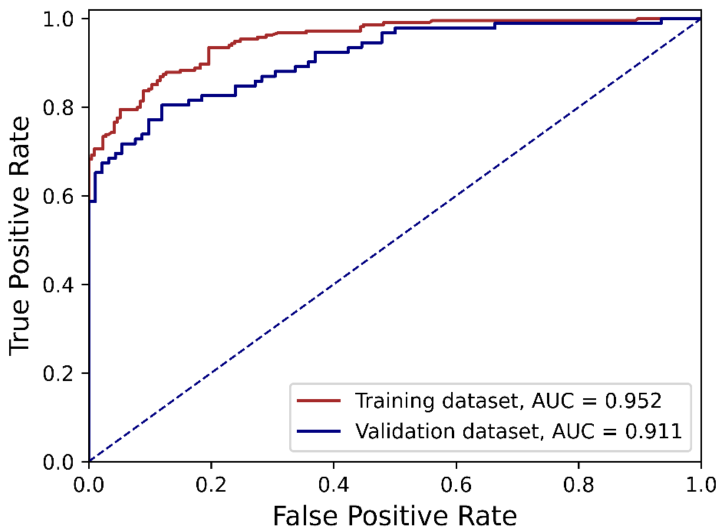

Figure 4. The ROC curve and the Area under the curve (AUC) of the HHO-RSCDT model

5.2. Role of fire ignition factor contribute to the HHO-RSCDT Ensemble Model

The role of influencing factors results is shown in Table 2.

Table 2. Role of the influencing factors contributing to the HHO-RSCDT ensemble model

Influencing factor	Merit value	Ranking
NDVI	0.304	1
Relative humidity (%)	0.171	2
Elevation (m)	0.145	3
Surface Temperature (°C)	0.133	4
Rainfall (mm)	0.127	5
NDWI	0.122	6
Distance to road (m)	0.079	7
Slope (°)	0.067	8
Aspect	0.012	9
LULC	0.007	10

It can be seen that NDVI, with a merit value of 0.304, is the most important, followed by relative humidity (0.171), elevation (0.145), surface temperature (0.133), rainfall (0.127), and NDWI (0.122). In contrast, LULC has the lowest contribution to the HHO-RSCDT model; the merit value is 0.007.

5.3. Model comparison

To evaluate the effectiveness of the proposed HHO-RSCDT model, a comparison was made between HHO-RSCDT and the benchmark models Random Forest (RF) and Support Vector Machines (SVM), which were built in our previous work (Truong et al., 2023b). Herein, the performance of the RF model (Acc = 78.8%, Kappa = 0.576, and AUC = 0.865) and the SVM model (Acc = 76.1%, Kappa = 0.522, and AUC = 0.851) model in the validation dataset is lower than the HHO-RSCDT model (Acc = 83.7%, Kappa = 0.674, and AUC = 0.911).

5.4. Forest fire danger map

To construct the forest fire danger map, the HHO-RSCDT model was used to compute the forest fire danger index for each of all pixels of the study area, and subsequently, five classes, very high, high, low, very low, and no, were obtained (Fig. 5).

Herein, the four different values for separating these classes are 0.84, 0.56, 0.29, and 0.18, which were determined by assigning the top 15% of the study area with the highest fire danger values to the 'very high' class. For the 'high', 'low' and 'very low' classes, 20% of the study area each was allocated, and the remaining 25% was assigned to the 'no class' classes (Figs. 6 and 7).

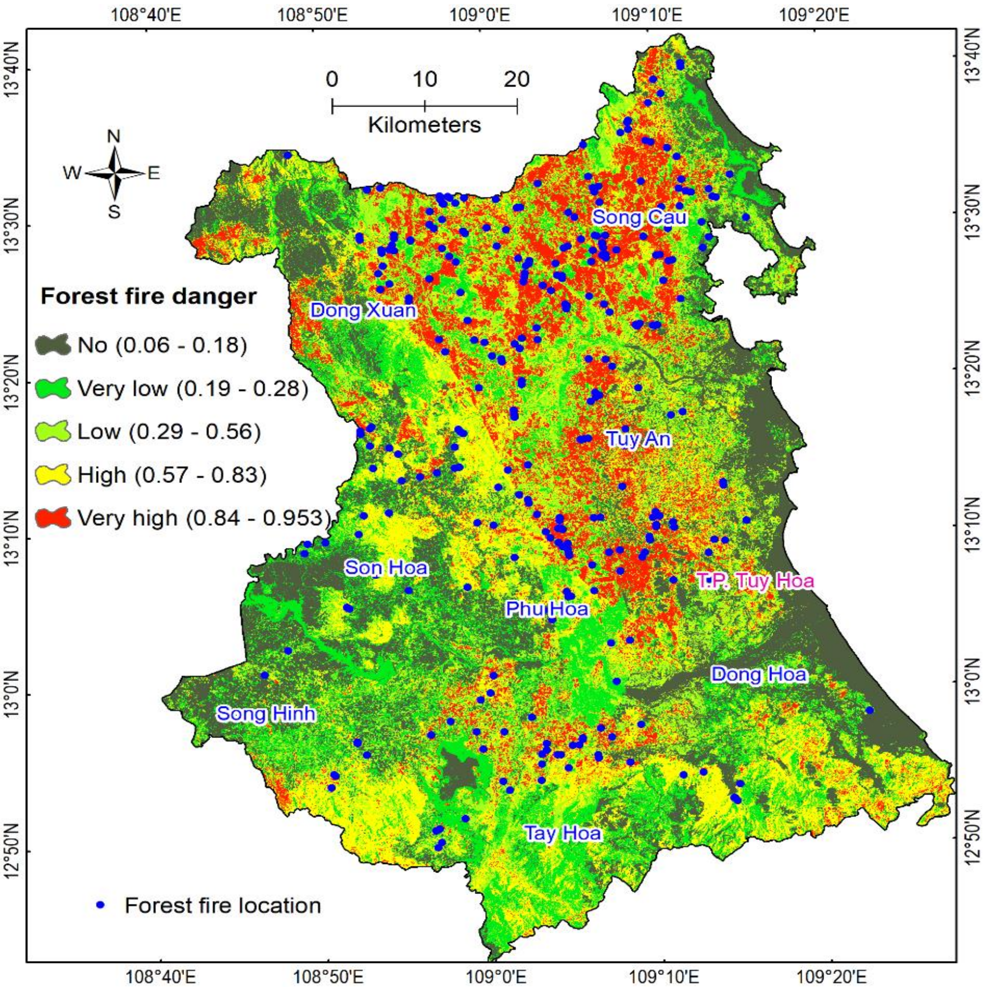

Figure 5. Forest fire danger map using the HHO-RSCDT model for the study area

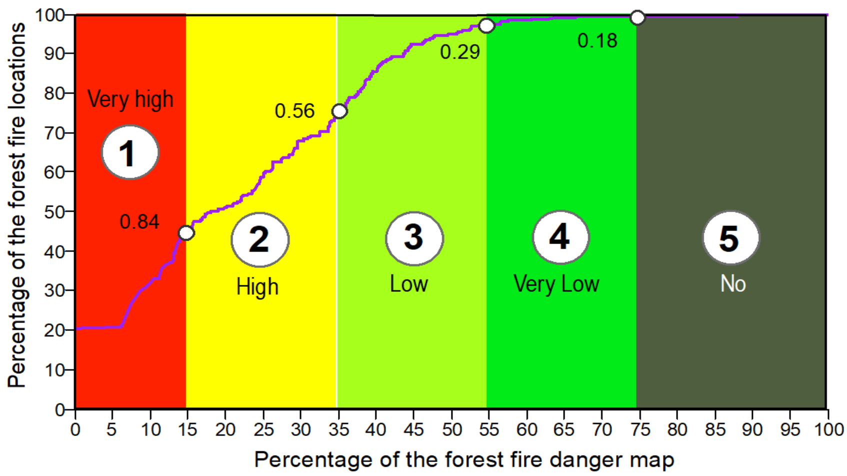

Figure 6. Plot graph for determining the separated values of the forest fire danger classes for the Phu Yen province

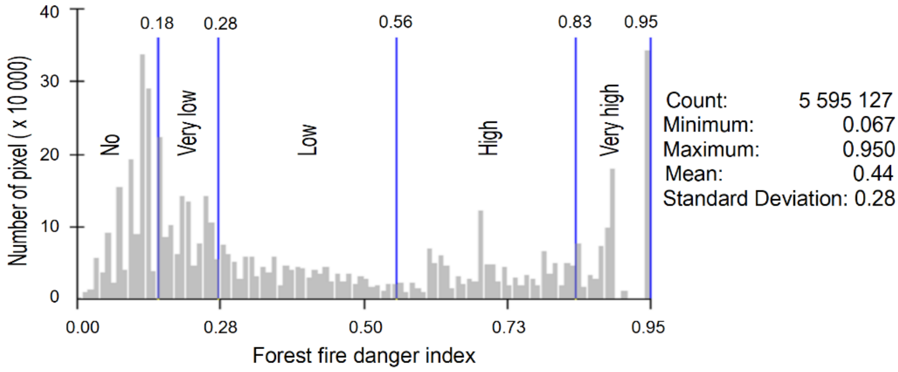

Figure 7. Five forest fire classes for the study area

Properties of the five forest fire danger classes derived by the HHO-RSCDT model for the Phu Yen province are shown in Table 3. It could be seen that 44.4% and 29.7% of the forest fire locations were located in the 'very high' class (757.5 km^2) and the 'high' class (1009.9 km^2), respectively. In contrast, 2.7% and 0.0% of the forest fire locations are located in the 'very low' class and the 'no' classes, respectively (Table 3).

Table 3. Properties of the five forest fire danger classes derived by the HHO-RSCDT model

No	Forest fire danger index	Forest fire location (%)	Description	Forest fire danger map (%)	Areas (km^2)
1	0.84–0.95	44.4	Very High	15	757.5
2	0.57–0.83	29.7	High	20	1009.9
3	0.29–0.56	23.2	Low	20	1009.9
4	0.19–0.28	2.7	Very Low	20	1009.9
5	0.06–0.18	0.0	No	25	1262.4

6. Discussions

Forest fires remain a significant problem in Vietnam (Pham et al., 2021) and Phu Yen province due to human activities, such as burning agricultural land and illegal logging. In addition, climate change has led to more frequent and severe droughts (Nguyen-Thi-Lan et al., 2021; Nguyen and Hoang, 2022). This study introduces and validates a novel ensemble modeling, HHO-RSCDT, for predicting the spatial patterns of forest fire danger, with a specific focus on a case study in Phu Yen province. The HHO-RSCDT is an ensemble of Random Subspace (RS), Credal Decision Tree (CDT), and Harris Hawks Optimizer (HHO). RS is used to generate various subspace datasets, and then, each dataset is adopted to generate a CDT classifier, whereas HHO is integrated to optimize the ensemble model. The HHO-RSCDT model demonstrates a high predictive ability, indicating that this novel ensemble learning approach can accurately predict the probability of forest fire danger, particularly in the 'high' and 'very high' areas.

The findings of this study suggest that the four parameters (k, m, mCDTree, and minNum) significantly impact the performance of the HHO-RSCDT model, underscoring the need for careful parameter selection. As a result, the high predictive performance of the HHO-RSCDT model suggests that the HHO algorithm has effectively searched and optimized the four parameters (k, m, mCDTree, and minNum) autonomously. Moreover, the HHO-RSCDT model outperforms the benchmark models, such as the RF and the SVM regarding prediction performance. This is because the CDT algorithm effectively manages noisy geospatial data in this work. Additionally, allowing RS to establish a forest of 10 CDT trees ensures diversity. In conjunction with the optimization phase of the HHO-RSCDT model, which utilizes 1000 iterations and 30 hawks, a comprehensive exploration of 30,000 potential combinations of k, m, mCDTree, and minNum was conducted to determine the optimal configuration. These factors contribute to the high performance of the HHO-RSCDT model. Consequently, the HHO-RSCDT model can be regarded as a promising tool for forest fire danger modeling. This finding supports recent research in the literature indicating that ensemble learning is a practical approach for natural hazard modeling (Naderpour et al., 2019; Ganaie et al., 2022; Ado et al., 2022), as it leverages diverse data and processes complex geospatial information.

Forest fires are complex processes; therefore, identifying the essential factors that contribute to forest fires can help improve the prediction model's accuracy and provide valuable insights for forest management and fire prevention strategies. This research indicates that NDVI is the most critical factor in the HHO-RSCDT model. This is a good result since NDVI is widely recognized as a proxy for available fuel for forest fires (Michael et al., 2021), as it provides information on vegetation health and moisture content - both critical factors in determining the likelihood of a forest fire occurrence. In contrast, the contribution of LULC to the HHO-RSCDT model was found to be the lowest, possibly because the forest fire locations were distributed relatively evenly across some LULC classes.

7. Concluding remarks

In this research, the effectiveness of a new ensemble modeling, HHO-RSCDT, is proposed and verified for forest fire danger mapping with a case study in Phu Yen province of Vietnam. Following the findings of this study, we have drawn some conclusions presented below.

- The HHO-RSCDT, an integration of HHO, RS, and CDT, has demonstrated its capability to predict forest fires with high accuracy.
- HHO is an effective optimization algorithm capable of autonomously searching and optimizing four parameters - k, m, mCDTree, and minNum.
- The most significant contributing factor to the HHO-RSCDT model is NDVI.
- The HHO-RSCDT demonstrates better prediction power than the RF and SVM benchmarks, indicating its potential as a valuable tool for modeling forest fire danger.
- A limitation of this work is the absence of an uncertainty analysis for the HHO-RSCDT model. Therefore, future research should consider conducting such an analysis to enhance the model's transparency, credibility, and applicability. By addressing this limitation, the model can become a more reliable tool for decision support in forest fire management.
- Finally, despite the above limitation, the forest fire danger map generated through this study could serve as a valuable resource for the local authorities in Phu Yen province, aiding them in forest management and protection efforts.

Acknowledgments

This work was financially supported by the Ministry of Education and Training (MoET) in Vietnam under grant number B2021-MDA-13.

Conflict of interest

The authors declare no conflict of interest.

References

Abawajy J.H., Kelarev A., 2017. Iterative classifier fusion system for the detection of Android malware. IEEE Transactions on Big Data, 5(3), 282-292.

Abid F., 2021. A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. Fire Technology, 57(2), 559-590.

Ado M., Amitab K., Maji A.K., Jasińska E., Gono R., Leonowicz Z., Jasiński M., 2022. Landslide susceptibility mapping using machine learning: A literature survey. Remote Sensing, 14(13), 3029.

Agbeshie A.A., Abugre S., Atta-Darkwa T., Awuah R., 2022. A review of the effects of forest fire on soil properties. Journal of Forestry Research, 33(5), 1419-1441.

Alabool H.M., Alarabiat D., Abualigah L., Heidari A.A., 2021. Harris hawks optimization: a comprehensive review of recent variants and applications. Neural Computing and Applications, 33(15), 8939-8980.

Bjånes A., De La Fuente R., Mena P., 2021. A deep learning ensemble model for wildfire susceptibility mapping. Ecological Informatics, 65, 101397.

Bui D.T., Bui Q.-T., Nguyen Q.-P., Pradhan B., Nampak H., Trinh P.T., 2017. A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agricultural and Forest Meteorology, 233, 32-44. Doi: 10.1016/j.agrformet.2016.11.002.

Bui D.T., Pradhan B., Lofman O., Revhaug I., Dick O.B., 2012. Landslide susceptibility assessment in the Hoa Binh province of Vietnam: A comparison of the Levenberg-Marquardt and Bayesian regularized neural networks. Geomorphology, 171-172, 12-29.

Calviño-Cancela M., Chas-Amil M.L., García-Martínez E.D., Touza J., 2016. Wildfire risk associated with different vegetation types within and outside wildland-urban interfaces. Forest Ecology and Management, 372, 1-9.

Chicas S.D., Østergaard Nielsen J., 2022. Who are the actors and what are the factors that are used in models to map forest fire susceptibility? A systematic review. Natural Hazards, 114(3), 2417-2434.

de Santana R.O., Delgado R.C., Schiavetti A., 2021.. Modeling susceptibility to forest fires in the Central Corridor of the Atlantic Forest using the frequency ratio method. Journal of Environmental Management, 296, 113343.

Dos Reis M., de Alencastro Graça P.M.L., Yanai A.M., Ramos C.J.P., Fearnside P.M., 2021. Forest fires and deforestation in the central Amazon: Effects of landscape and climate on spatial and temporal dynamics. Journal of Environmental Management, 288, 112310.

Dunham M., 2023. Matlab Weka Interface (https://www.mathworks.com/matlabcentral/fileexchange/21204-matlab-weka-interface), MATLAB Central File Exchange. Retrieved April 10, 2023.

Duong P.X., Du H.T., Linh V.T.T., Thai T.D., Minh-Thu P., 2020. The Effect of ENSO on Hydrological Structure and Environment in the South Central Coast-Vietnam. Journal of Marine Science, 2(1), 10-16.

Flannigan M., Wotton B., Marshall G., De Groot W., Johnston J., Jurko N., Cantin A., 2016. Fuel moisture sensitivity to temperature and precipitation: climate change implications. Climatic Change, 134, 59-71.

Ganaie M.A., Hu M., Malik A., Tanveer M., Suganthan P., 2022. Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.

Gao B.C., 1996. NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. Remote sensing of environment, 58(3), 257-266.

Gonzalez-Olabarria J.R., Reynolds K.M., Larrañaga A., Garcia-Gonzalo J., Busquets E., Pique M., 2019. Strategic and tactical planning to improve suppression efforts against large forest fires in the Catalonia region of Spain. Forest ecology and management, 432, 612-622.

Gouveia C., Bastos A., Trigo R., DaCamara C., 2012. Drought impacts on vegetation in the pre-and post-fire events over Iberian Peninsula. Natural Hazards and Earth System Sciences, 12(10), 123-3137.

Heidari A.A., Mirjalili S., Faris H., Aljarah I., Mafarja M., Chen H., 2019. Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems, 97, 849-872.

Ho T.K., 1998. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence, 20(8), 832-844.

Hoi N.D., Dung N.T., 2022. Structure, diversity and dynamics of natural-anthropogenic landscapes of Phu Yen province. Indian Journal of Ecology, 49(4), 1225-1233.

Iudin D., Sergeyev Y.D., Hayakawa M., 2015. Infinity computations in cellular automaton forest-fire model. Communications in Nonlinear Science and Numerical Simulation, 20(3), 861-870.

Jain P., Coogan S.C., Subramanian S.G., Crowley M., Taylor S., Flannigan M.D., 2020. A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4), 478-505.

Ke Y., Im J., Lee J., Gong H., Ryu Y., 2015. Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations. Remote sensing of environment, 164, 298-313.

Lattimer B.Y., Hodges J.L., Lattimer A.M., 2020. Using machine learning in physics-based simulation of fire. Fire Safety Journal, 114, 102991.

Lee S.K., Dang T.A., 2018. Application of AquaCrop model to predict sugarcane yield under the climate change impact: A case study of Son Hoa district, Phu Yen province in Vietnam. Research on Crops, 19(2), 310-314.

Leigh C., Bush A., Harrison E.T., Ho S.S., Luke L., Rolls R.J., Ledger M.E., 2015. Ecological effects of extreme climatic events on riverine ecosystems: Insights from A ustralia. Freshwater Biology, 60(12), 2620-2638.

Littell J.S., Peterson D.L., Riley K.L., Liu Y., Luce C.H., 2016. A review of the relationships between drought and forest fire in the United States. Global change biology, 22(7), 2353-2369.

Liu Y., Goodrick S., Heilman W., 2014. Wildland fire emissions, carbon, and climate: Wildfire-climate interactions. Forest Ecology and Management, 317, 80-96.

Lung N.N., Quat N.X., Lien A., Que A., Van Con A., Ky A., Cam L.V., 2011. Final report on forest ecological stratification in Vietnam. UN-REDD Program Vietnam: Hanoi, Vietnam.

Mantas C.J., Abellán J., 2014a. Credal-C4.5: Decision tree based on imprecise probabilities to classify noisy data. Expert Systems with Applications, 41(10), 4625-4637.

Mantas C.J., Abellán J., 2014b. Credal Decision Trees to Classify Noisy Data Sets, In: Polycarpou M., de Carvalho A.C.P.L.F., Pan J.-S., Woźniak M., Quintian H., Corchado E. (Polycarpou M., de Carvalho A.C.P.L.F., Pan J.-S., Woźniak M., Quintian H., Corchado E. (Polycarpou M., de Carvalho A.C.P.L.F., Pan J.-S., Woźniak M., Quintian H., Corchado E.s), Hybrid Artificial Intelligence Systems. Springer International Publishing. Cham, 689-696.

Mantas C.J., Abellán J., Castellano J.G., 2016. Analysis of Credal-C4.5 for classification in noisy domains. Expert Systems with Applications, 61, 314-326.

Matthews S., Sullivan A.L., Watson P., Williams R.J., 2012. Climate change, fuel and fire behaviour in a eucalypt forest. Global Change Biology, 18(10), 3212-3223.

McWethy D.B., Schoennagel T., Higuera P.E., Krawchuk M., Harvey B.J., Metcalf E.C., Schultz C., Miller C., Metcalf A.L., Buma B., Virapongse A., Kulig J.C., Stedman R.C., Ratajczak Z., Nelson C.R., Kolden C., 2019. Rethinking resilience to wildfire. Nature Sustainability, 2(9), 797-804.

Mell W., Jenkins M.A., Gould J., Cheney P., 2007. A physics-based approach to modelling grassland fires. International Journal of Wildland Fire, 16(1), 1-22.

Michael Y., Helman D., Glickman O., Gabay D., Brenner S., Lensky I.M., 2021. Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series. Science of the Total Environment, 764, 142844.

Miller C., Urban D.L., 2000. Connectivity of forest fuels and surface fire regimes. Landscape Ecology, 15, 145-154.

Moayedi H., Mehrabi M., Bui D.T., Pradhan B., Foong L.K., 2020. Fuzzy-metaheuristic ensembles for spatial assessment of forest fire susceptibility. Journal of Environmental Management, 260, 109867.

Moinuddin K.A.M., Sutherland D., 2020. Modelling of tree fires and fires transitioning from the forest floor to the canopy with a physics-based model. Mathematics and Computers in Simulation, 175, 81-95.

Naderpour M., Rizeei H.M., Khakzad N., Pradhan B., 2019. Forest fire induced Natech risk assessment: A survey of geospatial technologies. Reliability Engineering & System Safety, 191, 106558.

Naderpour M., Rizeei H.M., Ramezani F., 2021. Forest Fire Risk Prediction: A Spatial Deep Neural Network-Based Framework. Remote Sensing, 13(13), 2513.

Nguyen H.N., Hoang T.B.H., 2022. Vietnam: Vietnam in a Time of El Niño, in El Niño Ready Nations and Disaster Risk Reduction: 19 Countries in Perspective, M.H. Glantz, Editor., Springer International Publishing. Cham, 159-179.

Nguyen-Thi-Lan H., Fahad S., Nguyen-Anh T., Tran-Thi-Thu H., Nguyen-Hong C., To-The N., 2021. Assessment of farm households' perception, beliefs and attitude toward climatic risks: A case study of rural Vietnam. Plos one, 16(12), e0258598.

Oliveira S., Rocha J., Sá A., 2021. Wildfire risk modeling. Current Opinion in Environmental Science & Health, 23, 100274.

Peng L., Cai Z., Heidari A.A., Zhang L., Chen H., 2023. Hierarchical Harris hawks optimizer for feature selection. Journal of Advanced Research. Doi: 10.1016/j.jare.2023.01.014.

Pham T.T., Nguyen T.-D., Tham H.T., Truong T.N., Lam-Dao N., Nguyen-Huy T., 2021. Specifying the relationship between land use/land cover change and dryness in central Vietnam from 2000 to 2019 using Google Earth Engine. Journal of Applied Remote Sensing, 15(2), 024503-024503.

Phan D.C., Trung T.H., Truong V.T., Sasagawa T., Vu T.P.T., Bui D.T., Hayashi M., Tadono T., Nasahara K.N., 2021. First comprehensive quantification of annual land use/cover from 1990 to 2020 across mainland Vietnam. Scientific reports, 11(1), 1-20.

Pimont F., Dupuy J.-L., Linn R., 2012. Coupled slope and wind effects on fire spread with influences of fire size: a numerical study using FIRETEC. International journal of wildland fire, 21(7), 828-842.

Pourghasemi H.R., Gayen A., Lasaponara R., Tiefenbacher J.P., 2020. Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling. Environmental Research, 184, 109321.

Rihan M., Bindajam A.A., Talukdar S., Naikoo M.W., Mallick J., Rahman A., 2023. Forest Fire Susceptibility Mapping with Sensitivity and Uncertainty Analysis Using Machine Learning and Deep Learning Algorithms. Advances in Space Research, 72(2), 426-443.

Robinne F.-N., Parisien M.-A., Flannigan M., 2016. Anthropogenic influence on wildfire activity in Alberta, Canada. International Journal of Wildland Fire, 25(11), 1131-1143.

Rouet-Leduc J., Pe'er G., Moreira F., Bonn A., Helmer W., Shahsavan Zadeh S.A., Zizka A., van der Plas F., 2021. Effects of large herbivores on fire regimes and wildfire mitigation. Journal of Applied Ecology, 58(12), 2690-2702.

Sachdeva S., Bhatia T., Verma A., 2018. GIS-based evolutionary optimized Gradient Boosted Decision Trees for forest fire susceptibility mapping. Natural Hazards, 92, 1399-1418.

Sagi O., Rokach L., 2018. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249.

Sivrikaya F., Küçük Ö., 2022. Modeling forest fire risk based on GIS-based analytical hierarchy process and statistical analysis in Mediterranean region. Ecological Informatics, 68, 101537.

Slijepcevic A., Anderson W.R., Matthews S., Anderson D.H., 2018. An analysis of the effect of aspect and vegetation type on fine fuel moisture content in eucalypt forest. International Journal of Wildland Fire, 27(3), 190-202.

Storey M.A., Price O.F., 2022a. Prediction of air quality in Sydney, Australia as a function of forest fire load and weather using Bayesian statistics. Plos one, 17(8), e0272774.

Storey M.A., Price O.F., 2022b. Statistical modelling of air quality impacts from individual forest fires in New South Wales, Australia. Natural Hazards and Earth System Sciences, 22(12), 4039-4062.

Takaku J., Tadono T., 2009. PRISM on-orbit geometric calibration and DSM performance. IEEE Transactions on Geoscience and remote Sensing, 47(12), 4060-4073.

Tavakol A., Rahmani V., Harrington J., 2020. Temporal and spatial variations in the frequency of compound hot, dry, and windy events in the central United States. Scientific reports, 10(1), 1-13.

Tehrany M.S., Jones S., Shabani F., Martínez-Álvarez F., Tien Bui D., 2019. A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multisource geospatial data. Theoretical and Applied Climatology, 137, 637-653.

Teng J., Xia S., Liu Y., Yu X., Duan H., Xiao H., Zhao C., 2021. Assessing habitat suitability for wintering geese by using Normalized Difference Water Index (NDWI) in a large floodplain wetland, China. Ecological Indicators, 122, 107260.

Thach N.N., Dang B.T.N., Pham X.C., Nguyen H.T., Bui T.H., Hoang N.D., Bui D.T., 2018. Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study. Ecological Informatics, 46, 74-85.

Tri D.Q., Tran T.D., Truong D.D., 2019. Application of meteorological and hydrological drought indices to establish drought classification maps of the Ba River Basin in Vietnam. Hydrology, 6(2), 49.

Truong D.D., Tri D.Q., 2021. Evaluation of Economic Damage Caused by Drought in Central Region Vietnam: A Case Study of Phu Yen Province. Economic and Environmental Geology, 54(6), 649-657.

Truong T.X., Nhu V.-H., Phuong D.T.N., Nghi L.T., Hung N.N., Hoa P.V., Tien Bui D., 2023a. A new approach based on Tensorflow Deep Neural Networks and GIS for spatial prediction of forest fire danger in tropical areas. Remote Sens., 15(14), 3458.

Truong T.X., Trung N.V., Anh T.V., Khoa P.V., Hanh T.H., Viet-Ha N., Trong N.G., Nghi L.T., Phuong D.T.N., Diem Hang C.T., 2023b. Research and build new deep learning models using remote sensing and geoinformatics data for detecting and predicting forest fire risk. Hanoi University of Mining and Geology, Hanoi, Vietnam. Project Report B2021-MDA-13, 120p.

Tuyen T.T., Jaafari A., Yen H.P.H., Nguyen-Thoi T., Van Phong T., Nguyen H.D., Van Le H., Phuong T.T.M., Nguyen S.H., Prakash I., 2021. Mapping forest fire susceptibility using spatially explicit ensemble models based on the locally weighted learning algorithm. Ecological Informatics, 63, 101292.

Van Le H., Hoang D.A., Tran C.T., Nguyen P.Q., Hoang N.D., Amiri M., Ngo T.P.T., Nhu H.V., Van Hoang T., Bui D.T., 2021. A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. Ecological Informatics, 63, 101300.

Xie Y., Peng M., 2019. Forest fire forecasting using ensemble learning approaches. Neural Computing and Applications, 31(9), 4541-4550.

Zheng Z., Huang W., Li S., Zeng Y., 2017. Forest fire spread simulating model using cellular automaton with extreme learning machine. Ecological Modelling, 348, 33-43.

Zhu Q., Liu Y., Jia R., Hua S., Shao T., Wang B., 2018. A numerical simulation study on the impact of smoke aerosols from Russian forest fires on the air pollution over Asia. Atmospheric Environment, 182, 263-274.